UNIVERSITEIT GENT

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

Unemployment: Scars and Preferences

Corinna Ghirelli

Supervisor: Prof. Dr. Bart Cockx (Ghent University)

Submitted at Ghent University,
To the Faculty of Economics and Business Administration,
In fulfillment of the requirements for the degree of Doctor in Economics

UNIVERSITEIT GENT

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

Doctoral committee:	Prof. Dr.	Marc De Clercq (Ghent University, Dean)
	Prof. Dr.	Patrick Van Kenhove (Ghent University, Academic Secretary)
	Prof. Dr.	Bart Cockx (Ghent University, Supervisor)
	Prof. Dr.	William Parienté (Université catholique de Louvain)
	Prof. Dr.	Matteo Picchio (Marche Polytechnic University)
	Prof. Dr.	Dieter Verhaest (University of Leuven)
	Prof. Dr.	Bruno Van der Linden (Université catholique de Louvain)

Submitted at Ghent University,
To the Faculty of Economics and Business Administration,
In fulfillment of the requirements for the degree of Doctor in Economics

Acknowledgments

I have elaborated this dissertation during the last - very intense - four years. It did not turn out to be an easy task in many occasions. I certainly would not have managed to overcome many of the impasses I came across without the continuous active support of my supervisor. More in general, I like to see this dissertation more as the result of a teamwork which benefitted from the contribution of different people, which here I want to thank.

First, I am really grateful to my supervisor Prof. Dr. Bart Cockx, for all the time he dedicated to my research and his unlimited and precious guidance. Every day, he was available to discuss any detail of my research. He was there to support me when my motivation was low and to help me when I could not find a way out. He took a hard look at my work on a daily basis. Not only has this improved the quality of my research, but also my problem-solving skills. It was a privilege to work together with him during these four years.

Moreover, I want also to thank Prof. Dr. Bruno Van der Linden, my supervisor in the Research Master at UCLouvain and a member of my PhD guidance committee, for having continued to support me after the Research Master throughout the PhD. Both my supervisor and Bruno contributed very much to write the third chapter of my thesis, which has recently been published in an economic journal.

My thanks also go out to the other members of the jury, Prof. Dr. William Parienté, Prof. Dr. Matteo Picchio and Prof. Dr. Dieter Verhaest, for all the time they have invested in reading my dissertation and their valuable comments. The discussions that took place during the private defence have been very useful to improve the overall dissertation and, in particular, the single authored chapter.

I gratefully acknowledge the financial support from the Special Research Fund ("BOF") of Ghent University, as well as from the National Bank of Belgium (project code 05Y00110) that allowed me to conduct this research.

Moreover, I want to thank the secretaries of the department of social economics, Nathalie and Erna, for their availability during these four years of the PhD as well as for their administrative support in organizing the PhD defence.

A deep thank goes also to all my colleagues at the departments of general, financial and social economics. In particular, my officemates Stijn, Eva, Yannick, Andrea and Jose Luis, my "Gent–family": always supportive and ready to help. Last by not least, I would like to thank all my friends, those that are far for making me feel close, and those who are in Belgium for all the good moments spent together.

Grazie ai miei fratelli per tener alta la bandiera, ai miei genitori per il loro incondizionato supporto, e alle Amigas perché che mondo sarebbe senza le Amigas... A Mikel, eskerrik asko denagatik!

Ghent, December 2014
Corinna Ghirelli

Table of Contents

List of Figures

List of Tables

Overall Introduction

The context

In the last years, the Great Recession in 2008 and its devastating impact on youth unemployment triggered the debate on the long-term effects of economic downturns on the labor market career of young graduates. Long-term penalties may arise because youth unemployment not only brings the obvious loss of current income during the period of inactivity, but it may also inflict a longer-term scar through the increased future incidence of unemployment and lower subsequent earnings in employment. Moreover, scarring effects of unemployment may not be limited to the people who experience it. That is, a higher risk of unemployment in economic downswings may force workers to accept lower-level or lower-paying jobs than the ones they would have aimed to during better labor market conditions. This may induce long-lasting effects if the specific human capital acquired in these mismatched jobs is useless to climb the ladder towards higher-level positions that the worker would have targeted in the first place. Moreover, the opportunity costs of changing jobs increases with the experience acquired in the mismatched positions, potentially prolonging the scar. But is this view evidence-based? The first two chapters of this dissertation explore the long-term impacts of economic downturns on the labor market outcomes of young male graduates in Flanders, the most prosperous of the three regions in Belgium.

Moreover, the increased unemployment rate and unemployment duration since the Great Recession have fueled the discussion on the efficiency-insurance trade-off of the unemployment insurance system - a fundamental safety net that provides for the unemployed. That is, the provision of unemployment insurance raises moral hazard problems, i.e. the more generous the unemployment insurance, the lower the search incentives for the unemployed. To cope with this efficiency loss many European countries have imposed job search requirements on benefit recipients. To verify compliance, job search effort is monitored and, in the case of non-compliance, benefit recipients are sanctioned. Despite its wide implementation, the implications of this policy measure are not clear-cut. In the third chapter we evaluate, from a theoretical point of view, the social welfare implications of job search monitoring under a more realistic framework than what is assumed in the existing literature: first, we allow the unemployed job-seekers to procrastinate in job search and second, we allow for imperfections in the job search monitoring technology.[1]

[1] These assumptions have been explored separately in the literature, but not jointly.

The Belgian case

The case of Belgium is particularly interesting, because labor market rigidity, as measured by flows in and out of unemployment, is among the highest in the OECD countries. First, the literature suggests that the effects of graduating during downswings are much more persistent in a rigid labor market than in a flexible one. Secondly, in Belgium the sources of this rigidity differ according to workers' skill levels. Hence, we are able to obtain some insights in how distinct sources of rigidities drive different hysteresis mechanisms. For instance, until the end of 2013, white and blue collar workers were facing different levels of employment protection legislation (EPL), i.e. a flexible EPL for the latter while a very rigid EPL for the former.[2] For blue collar workers the sources of rigidity are different. They are supported by a short-time work compensation program that subsidies temporary reduction of labor force during downturns. This policy introduces rigidity by tying employers to employees. Moreover, minimum wages - among the highest of the OECD countries - are likely to be binding for blue collar youth. Thus, minimum wages represent an additional source of rigidity for this group together with a quite generous unemployment insurance system.

The generosity of the Belgian unemployment insurance system is mostly due to the absence of time limit on the payment of unemployment benefits, as opposed to many other countries. Moreover, unemployment benefits are provided not only to unemployed workers who are involuntary dismissed and with sufficient employment records, but also to unemployed school-leavers with no employment records, conditional on a waiting period of 12 months.[3] Since 2004, a job search monitoring scheme has been introduced targeting long-term (13 months or more) unemployed. The unemployed who are subject to monitoring are required to collect written proofs of their search activity, which is evaluated by case-workers in face-to-face interviews. This scheme is quite lenient compared to many other OECD countries: the frequency of the evaluations is lower, the search requirement imprecise, and sanctions can be imposed only after a second negative evaluation. However, sanctions are typically tougher than in many other countries.[4] Cockx and Dejemeppe (2012)[5] have shown that this scheme is effective in speeding up the transition into employment of the long-term unemployed after the notification and before the first evaluation, i.e. the sanctions are effective threats.

[2]Since the beginning of 2014 a single employment contract has been introduced in Belgium stipulating the same EPL for white and blue collar workers.

[3]Before (Since) 2012 the waiting period is 9 (12) months after registration to the Public Employment Service.

[4]The first evaluation of job search activity does not take place before 8 months from a notification letter which communicates to the long-term unemployed that he will be subject to monitoring. No quantitative targets are set on the number of job applications. No sanction is imposed after a first evaluation. After a second negative evaluation, a temporary (for 4 months) complete withdrawal of entitlement to unemployment benefits can be imposed. After a third negative evaluation the withdrawal can be definitive.

[5]Bart Cockx, Muriel Dejemeppe, "Monitoring job search effort: An evaluation based on a regression discontinuity design", Labour Economics, Volume 19, Issue 5, October 2012, Pages 729–737.

The Structure of the Dissertation

In the first two chapters of this dissertation we use the same data to answer to distinct but related research questions. In the first chapter we study the scars of experiencing a higher risk of unemployment at graduation on the labor market outcomes of Flemish male youth. By contrast, in the second chapter we aim to quantify the long-term effects of experiencing individual non-employment at the start of the career on the labor market outcomes of these youth. Throughout the analyses we distinguish between low and high educated school-leavers because, as mentioned above, labor market institutions in Belgium create different sources of rigidities for the low and the high educated. The third chapter is distinct from the previous two but still related to the experience of unemployment. It is a theoretical analysis on the social efficiency of job search monitoring – a policy measure that has been widely implemented to reduce unemployment – under more realistic time-preferences. The contribution of each chapter and the links relating them are outlined hereafter.

The first chapter studies the short- and long-run impacts of graduating in economic downturns in Flanders. This research question is estimated with a reduced form model, where the provincial unemployment rate at graduation, as a measure of the labor market conditions at graduation, is regressed on subsequent labor market outcomes of the school graduates. By graduating in adverse labor markets, the new graduates incur higher risks of experiencing unemployment. Such risks may entail long-lasting consequences on the career of the youth which may be not limited to actual individual unemployment. In this chapter we identify these scars and relate them to the rigidities characterising the Belgian labor market. Namely, we find that the low and the high educated undergo different long-term penalties from graduating in downturns: while a downturn at graduation is found to have a persistent negative effect on earnings for both groups, the high educated experience a persistent negative effect on the hourly wage and not on the annual hours worked, and the reverse holds for the low educated. This is explained by the presence of minimum wages which, at the start of the career, are likely to be binding for the low educated but not for the high educated, and by the strict EPL for white collar workers, which pushes the high educated to downgrade towards lower-paying jobs. Moreover, we provide evidence on the hypothesis that the effects of graduating during downswings are much more persistent in a rigid labor market than in a flexible one.

In the second chapter instead we narrow the scope of the analysis and focus on the long-term impacts of experiencing non-employment early in the career. We examine non-employment rather than unemployment to take into account the period of time that school-leavers may spend out of the labor market before first entry. We hypothesize that for low educated youth the scars of graduating during downturns arise through the loss of early work experience. We test this by means of an instrumental variable approach, where the unemployment rate at graduation is used as an instrument for the proportion of time spent in early non-employment, which in turn is used to explain the subsequent labor market career of the low educated, evaluated six and eight years after graduation. The results of the second chapter corroborate the hypotheses of the first chapter, according to which the long-term impacts of graduating in a downturn materialize through the loss of early work experience for the low educated but not for the high educated.

The first two chapters are related in a number of dimensions. First, both these chapters exploit a unique employer-employee dataset that matches survey and administrative data on a random sample of male Flemish youth. The administrative data contain detailed and comprehensive information on individual labor market performance up to twelve years since graduation. The survey data follow closely the educational path of the youth from the beginning of secondary education and allow us to precisely define the moment of graduation, which is crucial in our analysis and typically not observed in administrative data. Second, their identification strategies are related. In the second chapter, the unemployment rate at graduation is used as an instrument to identify the long-term effects of early non-employment on the labor market career of the low educated evaluated six or eight years after graduation. In the first chapter, the unemployment rate at graduation (i.e. the instrument) is directly regressed on the outcomes of interest measured in the first twelve and ten years after graduation for the low and high educated, respectively. Thus, the first chapter estimates a reduced form version of the equation of interest in the second chapter. However, the two chapters use different frameworks. The first chapter exploits the panel structure of the data: namely, the unemployment rate at graduation is interacted with experience dummies so that the results yield a vector of effects of interest, each representing the effect of the labor market conditions at graduation over potential experience. By contrast, in the second chapter the outcomes are evaluated at a given point in time (six or eight years after graduation). Hence, the analysis restricts each time to the cross-section variation. Of course, precision is reduced as a consequence of this decision. We choose to fix each time the dependent variable for simplicity since we are interested not only in the scar of early non-employment but also in its persistence: exploring the effect of early non-employment at different cross-sections seemed to us the easiest way to implement the instrumental variable approach and at explore the persistence of the scar. The use of Panel instrumental variables is left for future research.

While the first two chapters are focused on the long-term penalties of early unemployment (the risk and the actual experience of unemployment, respectively), in the third chapter we study how efficiently, from a welfare perspective, job search monitoring scheme shortens unemployment. This chapter is linked to the previous chapters to the extent that the moral hazard induced by unemployment insurance can exacerbate the persistence of the scars of experiencing early unemployment for the school-leavers. Note that in Belgium this is especially relevant because also school-leavers with no employment records are eligible to unemployment benefits, conditional on a waiting period of one year.

Job search monitoring has been widely introduced in many countries as a way to cope with the moral hazard problem induced by the unemployment insurance scheme. The provision of unemployment insurance benefits is conditioned to specific requirements. To verify compliance, the job search of the unemployed is monitored, eventually imposing sanctions in case of non-compliance. In Belgium job search monitoring has been introduced in 2004 for the long-term unemployed. Despite the popularity of this measure, the efficiency of job search monitoring is not yet clear-cut since empirical evaluations have provided mixed results. From the theoretical side, social welfare evaluations of this measure have yielded contrasting results depending on the type of time preferences that are assumed to characterize the unemployed job-seekers. Under exponential time preferences, the expected lifetime utility of the unemployed is negatively affected because they are forced to search more than optimal by binding requirements. Nevertheless, exponential preferences provide an unrealistic characterization of search activity since job-seekers do not behave consistently over time but rather postpone search. A more realistic assumption is hyperbolic time-preferences, which allows to depict the behavior of procrastinating agents. This alternative assumption has offered more positive conclusions on the efficiency of *perfect* job search monitoring since a sophisticated hyperbolic agent is aware of his tendency to procrastinate and may benefit from a commitment device that forces him to be consistent with an optimal plan. In the third chapter, we reconcile this positive evaluation with the ambiguous empirical findings on the effectiveness of job search monitoring. That is, we show that if the unemployed have hyperbolic time preferences, job search monitoring is efficient only under limited conditions, and that efficiency may not be attainable if we allow (more realistically) for *imperfections* in the monitoring technology. We argue that, from a social welfare perspective, other policies such as job search assistance may be preferable.

1

Scars of Recessions in a Rigid Labor Market

1.1 Introduction

The Great Recession in 2008 and its devastating impact on youth unemployment spurred the debate on the long-term effects of economic downturns on the career prospects of young graduates. In the current public debate the dominant viewpoint is that the Great Recession creates a *lost generation*. But is this view evidence based? Economic research has only relatively recently started investigating this question, exploring data on past recessions in various countries. Existing evidence broadly confirms the conjecture that a recession[1] has persistent impact on labor market outcomes of young graduates, although the magnitude and persistence of these effects depend much on the considered outcome (employment, wage, earnings,...), the level of educational attainment, and the institutional environment (see Section 1.2). This study analyses various of these dimensions in a unique employer-employee dataset that matches survey and administrative data on a sample of male school graduates in Flanders, the most prosperous of three regions in Belgium.

The case of Belgium is particularly interesting, because labor market rigidity, as measured by flows in and out of unemployment, is among the highest in OECD. For instance, the average monthly job destruction rate between 1990 and 1999 was less than 0.5% of the labor force, while expected unemployment duration exceeded 20 months. These figures are comparable to the Italian, but contrast with those of the much more flexible US labor market, where 1.25% jobs were destroyed every month and the mean unemployment duration was as low as five months (Pérez and Yao, 2012). The research of Genda et al. (2010), and more recently, of Kawaguchi and Murao

[1] In line with the aforementioned literature, a “recession” refers to a situation of adverse labor market conditions as measured by a rise in the unemployment rate rather than to a decline of real GDP.

(2014) suggests that the effects of graduating during a recession are much more persistent in a rigid labor market than in a flexible one. Our study provides further valuable evidence on this hypothesis. Moreover, since in Belgium the sources of this rigidity differ according to skill level, we can obtain some insights in how distinct sources of rigidity drive different hysteresis mechanisms. For high skilled workers the rigidity is predominantly induced by the very strict employment protection legislation (EPL) for white collar workers. But this does not apply for low skilled workers, since they are usually employed as blue collar workers for whom, until very recently, EPL was very loose. For low skilled youth, (sectoral) minimum wages that are among the highest of OECD, and lenient short-time work compensation (STC) and unemployment insurance (UI) systems are therefore more relevant sources of rigidity.

In line with the institutional setting, we find that adverse labor market conditions hardly affect hourly wages of low educated youth and that nearly all the burden of the adjustment runs through the annual number of hours worked, predominantly induced by more unemployment experience. The negative impact matters more in the first years after graduation, but it persists up to 12 years later. This is in line with the ample evidence that experiencing unemployment (early in the career) inflicts long-term scars on labor market outcomes (e.g.,Arulampalam, 2001; Gregg, 2001; Gregg and Tominey, 2005; Schmillen and Umkehrer, 2013b).

For high educated youth, minimum wages are in general not binding, STC is not available, and replacement rates in UI are relatively low, so that a different hysteresis mechanism is operating. A recession may force high educated youth to downgrade and accept lower quality jobs paying lower wages. Those who refuse to downgrade become temporarily unemployed. The possibility of catching-up with a more fortunate generation that did not graduate in a recession is hampered by a slower pace of human capital accumulation in these lower quality jobs (Gibbons and Waldman, 2006), by rigidities induced by long-term contracting (Baker et al., 1994; Beaudry and DiNardo, 1991), or by search frictions that increase with age or job tenure, inducing workers to stop searching for a better paying job (Oreopoulos et al., 2012). The very strict EPL for white collar workers in Belgium increases the gradient of search frictions with age or job tenure. Hence, we find much more persistence in hourly wages and earnings of the high skilled in Belgium than in the more flexible North American labor markets.

Our study does not only provide insights in how distinct sources of rigidity drive different hysteresis mechanisms. It also contributes to the literature in other dimensions. First, we identify the long-run effects of recessions on labor market outcomes by exploiting the variation of the provincial unemployment rate between five provinces and eight graduation years (1994-2001 for the low educated and 1997-2004 for the high educated). This variation identifies the causal impact of recessions only to the extent that the provincial unemployment rate does not affect the composition of graduates in a province, either by changing the timing of graduation or by inducing inter-provincial mobility. We show that the latter is negligible and propose a new method for testing the former. The test is based on a discrete duration model relating the timing of graduation since the end of compulsory education to the provincial unemployment rate. We cannot reject the absence of such relationship.

Second, we have access to unusually rich data. Survey data are matched to comprehensive administrative data from the various institutions of the Belgian Social Insurance system. The survey data provides precise information on the timing of graduation, while the administrative data contain, up to twelve years after labor market entry, detailed information (hours worked, hourly wage and earnings) on salaried public and private sector employment, and also on time spent as self-employed worker. For salaried workers we have also information on the quality of the firm (as measured by the median wage). Existing research is often less comprehensive by focusing on particular labor market outcomes,[2] either wages or (salaried) employment, or by considering particular sub-populations (e.g. college graduates).

Third, we propose a new method of inference. Inference is complicated, since the regressor of interest, the provincial unemployment rate at graduation, is a serially correlated variable that is measured at a grouped level. Bertrand et al. (2004) demonstrate that in such circumstances the standard error can be severely downwards biased and that this bias can be very severe if the number of groups (clusters) is small. To address this problem, we propose a novel feasible generalized least squares (FGLS) method closely related to the efficient Minimum Distance (or Minimum Chi-square) estimator of Wooldridge (2006, 2010) for cross-section data and the FGLS recently introduced by Brewer et al. (2013) for difference-in-differences designs. We organize our discussion as follows. In the next section we briefly review the literature on the scarring effects of graduating in downturns. Section 1.3 summarizes the institutional setting. In Section 1.4 we describe the data. In Section 1.5 the estimation strategy is explained, including the way in which we deal with the problem of inference in the presence of a small number of clusters. We report the results in Section 1.6. Section 1.7 concludes. We refer the reader to the Supplementary Appendix in Chapter A for a detailed description of the variables construction, the sample selection, the complete estimation results, as well as a description and the results of the tests for the endogeneity of the timing of graduation and of the inter-provincial mobility.

1.2 Literature Review

Researchers studying the long-term effects of graduating during recessions in North-America report different findings for high than for low skilled youth. Graduating from college during a recession imposes a modest but long-lasting penalty on earnings that gradually fades away in about ten years (Genda et al., 2010 for US and Oreopoulos et al., 2012 for Canada).[3] These earnings losses are essentially due to a decline in hourly wages, not in employment or hours worked. During a recession college graduates are forced to accept lower quality jobs paying lower wages and/or offering less *career* perspectives and less opportunities for promotion and training, since high quality career jobs are then in reduced supply. Workers can react to this set-back and catch-up with the more lucky generations that graduated during booms by enhancing investments in human capital (Mroz and Savage, 2006), or by intensifying their search for higher paying jobs (Topel and Ward,

[2]Genda et al. (2010) is a notable exception.

[3]Mansour (2009) and Kahn (2010) found similar but more persistent effects for the US.

1992). However, the possibility of catching-up is hampered by hysteresis in career progression, both *within* and *between* firms. This can be due to imperfections in downward wage renegotiations of *lucky* generations that entered in a boom (Beaudry and DiNardo, 1991),[4] or to information imperfections (Baker et al., 1994), both shielding the internal labor market from the competitive forces in the external labor market. But also in the absence of long-term wage contracts, part of the acquired human capital in these lower quality jobs is task specific, leaving part of a worker's human capital unutilized when she gets promoted, and, hence, inducing lower wages many years later (Gibbons and Waldman, 2004, 2006). Moreover, search frictions that increase with age or job tenure can induce workers to stop searching for a better paying job (Oreopoulos et al., 2012, 2008).

For low skilled workers, by contrast, the effect of adverse labor market conditions at graduation has immediate important negative effects on wages and earnings, which fade away quickly after a couple of years, and a small and only marginally significant, though persistent, effect on employment. The number of hours worked is not affected (Genda et al., 2010).[5] This is because the labor market for low skilled workers involves less investments in human capital and long-term career contracts. It operates more like a spot market, in which wages rather than employment react rapidly to changes in the economic environment (see also e.g. Kilponen and Santavirta (2010) and Devereux and Hart (2007)).

These findings are specific, however, since, in contrast with many other countries, the US and Canada have very flexible labor markets. In a rigid labor market more persistent effects of a recession are expected. Employers then have more incentives to screen job applicants before hiring, because they are more forced into long-term relationships with their employees. If in an upturn screening is less costly among the pool of recent graduates than among a pool of job-seekers, unemployed because they graduated in a recession, then the latter group is more likely to be set back permanently. Genda et al. (2010) find supporting evidence of this hypothesis. In Japan the screening of recent graduates is indeed facilitated, both, because high schools are obliged by law to help firms in matching graduated students to jobs, and because social norm against dismissal[6] and resulting case law make dismissal of regular workers for economic reasons almost prohibitive. In contrast to their results for the US, these authors find strong and persistent (lasting more than 12 years) negative effects of a recession on both employment and earnings for low educated graduates in Japan. For the high educated, similar to the findings in the US, no significant effect on employment is found, while the effect on earnings also declines, but starts at a lower level and remains significantly negative after 12 years, rather than gradually fading away, as in the US. Recently, Kawaguchi and Murao (2014) find in a cross-country study more supporting evidence that recessions at labor market entry have more persistent adverse effects on the (un)employment

[4]Recently Hagedorn and Manovskii (2013) criticize this interpretation. They argue that wages are still determined by spot markets and not by long-term implicit contracts. They show that, once the current match quality is taken into account, past labor market conditions no longer play a role in the wage determination.

[5]Kondo (2007) reports similar findings for other disadvantaged groups, such as black men and women.

[6]By the 1970s high long-term unemployment created the perception that firms should be held responsible for their employees' job security.

rate[7] in countries with more labor market rigidity as measured by EPL strictness, union coverage, and benefit duration of UI.[8]

Evidence from European studies only partially support the conclusions of aforementioned studies. Within countries with moderate to high labor market rigidity, such as Norway, Sweden, Austria, and Germany, persistent effects roughly in line with aforementioned theory are found. Raaum and Røed (2006) find that in Norway a business cycle slump at ages 16 and 19 raises prime age unemployment rates by as much as one to two percentage points. These authors do not study the effect on wages. More recently, Liu et al. (2012) report similar persistent negative effects on employment for college graduates. The effect of graduating in a period of high unemployment on earnings is only significantly negative during the first three years, but could be underestimated as a consequence of the positive selection induced by the effect on employment (Heckman, 1974).[9] Kwon and Meyersson Milgrom (2007) study the effect of labor market entry conditions on wages of white collar workers in Sweden. If these workers enter the labor market in a boom they obtain faster promotions which leads to persistent wage premiums for such cohorts. Brunner and Kuhn (2014) report that in Austria a one percentage point increase in the unemployment rate decreases the daily wage by 0.9% and persists at least 20 years, but this is again a lower bound of the effect, since the employment rate is also persistently affected by the business cycle. Similar to high educated workers in the US, effects for white collar workers are smaller, and fade after five to ten years. However, in contrast to the US, and possibly related to labor market rigidity, blue collar workers suffer more importantly and more persistently from a recession. Stevens (2008) studying the effect for a population of low to medium skilled workers in Germany finds more persistent effects on employment and wages of a recession than for low educated workers in the US, but less persistent than in Austria and Japan for lower skilled workers. She reports a negative effect on employment during the first five years after graduation, which, as in the preceding study, leads to an underestimation of the effect on the wage in that period. The latter is important in the first four years after graduation, but fades away after seven years.

Two studies seem to contradict the theory that predicts more persistent effects in countries with strict labor market regulations. First, despite the high labor market rigidity in France, Gaini et al. (2012) report in this country lower employment rates for cohorts graduating in a recession only during the first two years and no wage penalty. The authors advance two potential explanations: (1) a high minimum wage dampens the effect on wages and (2) a persistently high unemployment rate, such that employers use unemployment less as a negative signal in the hiring process (Biewen and Steffes, 2010; Kroft et al., 2013a).

Second, according to Tumino and Taylor (2013b) graduating in a recession in Britain, a country with relatively limited labor market protection, very negatively affects the probability of em-

[7] The effect on wages or earnings is not studied.

[8] Belgium was the most rigid country according to their composite index.

[9] If among the pool of new graduates only the ones with best unobserved characteristics (e.g. ability, motivation) find a job in a downturn, the composition of workers is positively selected. Consequently, the effect on earnings for the subpopulation of workers is biased downwards, since it does not take into account that they represent disproportionately more productive individuals.

ployment and the level of wages of young men, and this effect, even if it declines, remains very negative even after ten years of potential experience. The authors do not discuss why the effects are more important and persistent in Britain than in North-America, but part of the explanation could be related to having a more vocational based educational system in the Britain and a more a more general one in North-America. Hanushek et al. (2011) argue that, although vocational based education is beneficial for low educated at the start of the career by easing the transition into the labor market, their skills can very quickly become obsolete. Consequently, in case of a recession at graduation, the low educated may have more difficulties in integrating into the labor market once the economy recovers. By contrast, general education provides broad knowledge and basic skills that makes graduates more adaptable to changes in labor demand and, hence, less vulnerable to persistent damage in case of a temporary slowdown of economic activity. Nevertheless, this distinguishing feature of the educational system cannot explain why recessions in Britain lead to equally or even more persistent adverse labor market outcomes than in other European countries, since most other European countries share this more vocationally based educational system or have even developed it more, such as in Germany or Austria. Another explanation is related to the higher incidence of over-education in the UK. Although there is not yet a clear-cut explanation of the determinants of scarring for UK, broad evidence has reported wage-penalties due to over-education in the last three-decades for this country (although it is not related to downturns) (e.g. Chevalier (2003); Dolton and Silles (2008)). Moreover, it has been shown that students stay longer in education in downturns (Taylor and Rampino, 2014; Tumino and Taylor, 2013a). Hence, it is possible that in downturns the mismatch between skills and jobs worsens, so that lower educated new graduates are left out of the labor force while mismatched higher educated risk long-term damages if they do not manage to climb the ladder towards higher-paying jobs.

1.3 Institutional Setting

Belgium is a federal state which has decentralized territorial competences to three Regions (Flanders, Brussels and Wallonia) and person-related issues to three language Communities (Flemish/-Dutch, French and German). Education is organized by the Communities, Unemployment Insurance (UI) and employment legislation by the federal authorities, and Active Labor Market Policies (ALMP) by the Regional employment offices. We briefly summarize the relevant institutional environment of the Flemish Region and Community for the period of analysis.

Since the beginning of the eighties education is compulsory up to age 18. At that age pupils have been six years in primary and six years in secondary school. If they do not repeat a grade, they complete secondary education at that age, except for those in a vocational track for whom secondary education ends after seven years. Secondary school is segmented in implicitly hierarchical tracks: (i) general, (ii) technical, (iii) vocational, and a smaller arts track. In addition, from age 15, pupils may choose a part-time vocational or apprenticeship track. *Downward* re-orientation and grade repetition is used more often as remediation policy than in other countries (OECD, 2012). The general and technical track directly (without a central entry exam) prepares for enrollment in higher education. Those in the vocational track must first complete a seventh specialization year. Our data refer to the pre-Bologna-reform period. At that time, three kinds of higher degrees could be obtained: (i) non-university of the *short type* (lasting three years), (ii) non-university of the *long type*, and (iii) university of the *long-type* (usually four or five years). According to the PISA studies assessing in a standardized way reading, mathematics and science skills at age 15 in large group of OECD countries, the Flemish pupils steadily performed at the very top of European countries. However, at the same time the spread of these scores is much higher than the OECD average, and the educational performance is highly segmented according to social background. School drop-out in secondary school is as high as 10% (Cockx (2013a); http://www.pisa.ugent.be/en/en).

UI in Belgium is very generous, not so much in terms of level, but of length of the entitlement. It is one of the only countries in the OECD in which no time limit is set to the entitlement and in which school graduates are entitled to (flat rate) unemployment benefits (UB) if they are still unemployed nine months after registration. Job search requirements are very loose according to international standards, and non-participation to active labor market programs (ALMP) is hardly sanctioned (Cockx and Dejemeppe, 2012a; OECD, 2007). Nevertheless, for youth UI may not be so large a disincentive for full-time work, since in Belgium the national minimum wage is among the highest in OECD (Cockx, 2013b). Moreover, in many sectors this national minimum is topped up, by 17% on average (Rycx and Kampelmann, 2013). Even if the minimum wage is found to have ambiguous effects on employment, at such high a level it is expected to reduce employment of low skilled youth, for whom it is most likely binding (Cahuc and Carcillo, 2014; Kramarz and Philippon, 2001; Neumark and Wascher, 2008).

Among the OECD countries, Belgium has close to the most stringent restrictions overall on individual and collective dismissals according to the OECD‘s EPL indicators (OECD, 2013). This is especially a consequence of very restrictive rules for collective dismissals. For individual dismissals in regular contracts EPL is less strict than the OECD average, but this conceals considerable heterogeneity in strictness according to the type of labor contract. In Belgium EPL for open-ended contracts differs between blue- and white-collar workers. For the latter group, it differs between those earning more than €32,254 a year and those earning less. The notice period for a blue-collar worker is generally less than one month for each five years of seniority, while for low-wage and high-wage white-collar workers it is three and five months respectively (Cockx, 2013b).

Even if EPL for blue collar workers is weak, a system of short-time work compensation (STC) specific for this group restrains firm mobility of blue collar workers. This STC compensates blue collar workers for temporary disruptions in business activity. These disruptions may last from maximum four weeks, in case of a complete temporary lay-off, up to maximum twelve months, in case of a work suspension of at most one week for every two weeks. Firms must justify the reason of the disruption to be able to invoke STC for their employees, but employer contributions financing the scheme are not experience rated. The replacement rate may reach nearly 100% (including employer top-ups), and no job search requirements or training are imposed on STC recipients, so that the scheme restrains mobility more strongly than in other countries. STC is extensively used in economic downturns and partly explains why the unemployment rate in Belgium only modestly increased during the Great Recession in 2008 (Hijzen and Venn, 2011; Høj, 2013).

Finally, in a nutshell, in Belgium most of the wages are usually determined according to wage scales by function or level of education, and for white collar workers also by seniority or age. Wages are automatically adjusted to the evolution of the consumer price index, and every two years the trade unions and employers' organizations bargain at the sectoral level on the extent of real wage increase. Wages are therefore downward rigid (Fuss, 2009).

We mentioned in the Introduction that Belgium has one of the most rigid labor markets in OECD in that inflow rates into unemployment are among the lowest, while expected unemployment duration is among the highest. The description in this section suggests that this is very much a consequence of institutions. Nevertheless, the sources of this rigidity are very different for the low skilled youth, prevalently blue collar workers, than for high skilled youth, prevalently white collar workers. This suggests that the mechanism underlying the long-run impact of adverse labor market conditions at graduation may be very different according to skill level. This is why we conduct a separate analysis for high and low educated youth.[10]

1.4 Data

For this analysis it is important to be able to accurately determine the timing of graduation, since measurement error in this timing may lead to important compositional biases. For instance, some studies assimilate graduation to the first registration as salaried worker according to administrative data. Such definition disregards those who transit from school to non-employment and, hence, disregards a group the size of which clearly varies with labor market conditions. This induces spurious correlation between labor market conditions at graduation and subsequent labor market outcomes.

To avoid this problem, we base our analysis on the Sonar surveys conducted on a representative sample of three birth cohorts, born in 1976, 1978 or 1980, and living in Flanders at age 23, and which contains information on the time at which school is left.[11] The surveys (as well as follow-up surveys at age 26 or 29) register retrospectively and on monthly basis the most important activity of the respondents, among which education. Based on this information, graduation is identified to occur in the first month that education has been interrupted for at least 4 months. We retain in the sample individuals leaving education between age 18, i.e. the compulsory schooling age in Belgium,[12] and age 24. Graduates after age 24 are not considered, because the small number of individuals involved would complicate inference dramatically (see Section 1.5.2). The surveys

[10]There is a very high correspondence between low (high) education and being employed as a blue (white) collar worker: see Table A.9 in Chapter A.

[11]For more details, see SONAR (2003, 2004a,b).

[12]We drop 0.17% of individuals for which graduation is reported to occur before 18.

also contain control variables for the analysis, which are measured at age 17, such that they are predetermined at graduation: father's and mother's education (years of completed education since age 12), the type of educational program (general, technical, vocational or a part-time vocational and apprenticeship program) in which the individual is enrolled at age 17, and the number of grade repetitions at age 17. The level of education is measured by the number of *completed* years of education: grade repetitions are not counted. This variable is used as a control in a sensitivity analysis.

The original Sonar sample contains about 3,000 individuals for each birth cohort, 9,000 in total. We only retain men to avoid dealing with sample selectivity induced by labor force withdrawal of women for reasons related to fertility and caring responsibilities.[13] Apart from the aforementioned selection criteria regarding school-leaving age, we drop the following individuals to enhance sample homogeneity: those who attended special needs and arts education, who did not speak Dutch at home, or who did not permanently reside in Flanders at graduation. After eliminating, in addition, individuals with missing or inconsistent values in variables, this leaves us with a final sample of 3,514 men. A more detailed description of the variable construction and the sample selection can be found in Sections A.1 and A.2 of Chapter A.

Our empirical analysis also hinges on having access to high quality information on labor market outcomes on a sufficiently long time span after graduation. We therefore matched the survey data to administrative data of Belgian Social Insurance institutions centralized at the Cross Roads Bank of Social Security. These data contain detailed quarterly information about labor market histories between 1998 and 2010.[14] For salaried workers we construct the following outcomes: log annual earnings, log annual hours and log average hourly wage. The hourly wage is obtained by dividing annual earnings by the total number of hours worked in a year.[15] We complement this information with three annual indicators of employment: self-employment (if registered as such part of the year, irrespectively of being a salaried worker in the same year), salaried employment (strictly positive earnings and not being self-employed), overall employment (either self- or salaried). Notice, by these very broad definitions of employment, the business cycle variation is more reflected in hours worked than in the employment rate. To get a sense of whether the variation in hours worked reflects fluctuations in the time spent unemployed or hours worked part-time,[16] we also distinguish between hours worked full-time and part-time.

The administrative data provide additional control variables measured at age 17 (living in

[13]We do not study the scarring effect of graduating in downturns for women, because other labor market institutions supporting mothering are likely to enter in the picture and complicate the interpretation. We therefore leave such an analysis for future research.

[14]Notice that part of the early labor market experience of cohorts graduating between 1994 and 1997 is missing.

[15]Taking log-transformations of these outcomes for salaried workers potentially leads to a sample selectivity problem, since for workers who are not in salaried employment this transformation is not defined and, hence, need to be dropped (Heckman, 1974). However, we will argue below that this selectivity is not a big concern in light of the evidence that graduating in downturns does not significantly affect the probability of being in salaried employment. Only for the high educated in the first few years after graduation there might be an issue, but this is discussed.

[16]The administrative data do not measure overtime for full-time workers, so this cannot be a source of variation in annual hours.

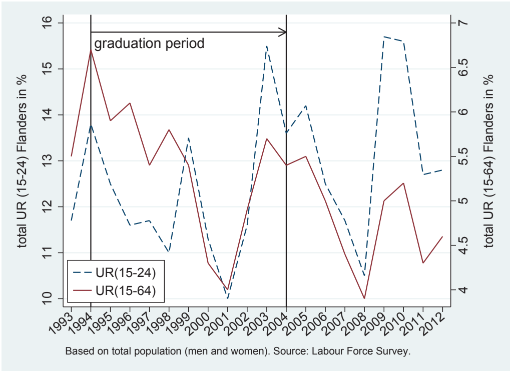


Figure 1.1: National Unemployment Rates (UR): overall (age 15-64) and youth (age 15-24).

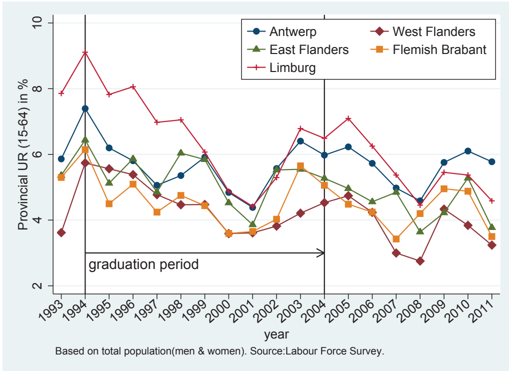


Figure 1.2: Provincial Unemployment Rates (UR): overall (age 15-64).

single parent household, not living together with either parents, the number of other household members by age class) and variables that help getting a better understanding of the mechanisms underlying the long-term negative impacts on individual outcomes of adverse labor market conditions at graduation (the median daily wage paid out on June 30 in recruiting firms and indicators of firm and provincial mobility[17]). To obtain a measure of *permanent* firm quality, we average the log median wage within recruiting firms over the observation period in a similar way as Oreopoulos et al. (2012) (see Section A.1.3 of this dissertation for details).

Final sources of information are the Labor Force Surveys (LFS). These provide long time series of the provincial and national[18] unemployment rates, used to characterize recessions in the labor market. Figure 1.1 plots both the youth (age 15-24) and overall unemployment rate (age 15-64) from 1993 to 2012. Observe, even if the level and the variability of youth unemployment is much higher than the overall rate, the time pattern of both series is very similar. The period of graduation 1994-2004 covered by our data captures a complete cycle, so that the main effects of interest can be identified on *national* data. However, in line with the literature, we aim at exploiting provincial specific time-shocks in the analysis. As Flanders is a relatively small region, one may question whether commuting and changing residence would not make it impossible to exploit the latter variation. That is, workers could offset the negative impact of graduating in provinces with few job opportunities by moving or commuting to provinces with a tighter labor market. However, the magnitude of the inter-provincial variation in the unemployment rate reported in Figure 1.2 demonstrates that mobility far from eliminates all inter-provincial variation.[19] A caveat is that

[17]An individual is defined to change firm in year t if he is observed in a different firm in at least two quarters of the year t, or if the first firm in which he was employed in year t differs from the last firm in year $t-1$. Transitions between self-employment and salaried employment are included in the definition of firm mobility. An individual is defined to move in year t, if he lives in another province at the end of year t than where he lived at the end of year $t-1$.

[18]*National* refers to "Flanders" and not to "Belgium".

[19]Notice that unemployment is measured in the province of residence and not of job location. Hence if workers commute to avoid the negative local labor market conditions, this evens out the provincial variation in the unemployment rate.

the high educated might commute more than the low educated, since they are (1) less liquidity constrained because of high expected wages or better working conditions and (2) more mobile due to higher motivation to find jobs that meet their expectations about wages/job profiles. Since provincial UR series by skill level are not available, we cannot check whether the provincial unemployment rates for the high educated vary less than for the low educated. However, if mobility would completely even out the inter-provincial variation of the unemployment rate for the high educated and the provincial variation would only reflect the variation for the low educated, then we should find for the high educated no relation between a higher provincial unemployment rate at graduation and subsequent labor market outcomes. This is not consistent with our findings reported below.[20]

In Table 1.1 in the Appendix 1.8 we report some descriptive statistics of the control variables in the retained sample. We distinguish between men with a degree not higher than secondary education, i.e."low educated", and those with a higher level of education, i.e. "high educated". We make this distinction throughout the analysis, because the minimum wage is more likely to be binding for low educated youth and because EPL-strictness varies significantly between white collar and blue collar workers (see Section 1.3). 69% of the low educated are prevalently (i.e. more than 50% of the time) employed as blue collar workers, while for the high educated this figure is only 11%.

1.5 Estimation Strategy

To identify the long-term effects of labor market conditions at entry on the mentioned labor market outcomes, we exploit, separately for low and high educated youth, the variation in the unemployment rate at graduation both at the national and at the provincial level over 8 years (1994-2001 for the low educated and 1997-2004 for the high educated). We operationalize this identification strategy by estimating, as in the existing literature (Section 1.2), a linear regression model of each outcome of interest on the unemployment rate at graduation and on a number of potentially confounding factors. This identification strategy requires that the composition of the graduating cohort to be unrelated to this unemployment rate. We show that the unemployment rate at labor market entry is not significantly related to (i) the timing of graduation, and (ii) the province of living. As mentioned in the Introduction, we test the former by demonstrating that the duration

[20] The Belgian Federal Public Service Economy (Service public fédéral Economie) published a report on inter-provincial commuting between residence and jobs for 2006 (Service public fédéral Economie, 2007). According to these data, 19.8% of Belgian workers were working in a different province than the province of residence in 2006. In Flanders, inter-provincial mobility varies across provinces: Flemish Brabant was the province with highest proportion commuters in another Flemish province, Brussels or Wallonia (42%); this figure is 10%, 9%, 24% and 14% for Antwerp, East Flanders, West Flanders and Limburg, respectively. Note that these figures do not make distinctions in terms of gender, age nor education of the workers. These figures only exist with regards inter-regional mobility: in 2006, mobility was higher among high educated workers (15.6%) compared to medium and low educated workers (9.2% and 7.6%, respectively). Moreover, young worker below 25 years old were less mobile than mature workers (7.4% compared to 11.1%, respectively). Men were more mobile than women (12% compared to 10%, respectively).

between the end of compulsory education at the age of 18 and each year of potential graduation is unrelated to the provincial unemployment rate in those years. To avoid breaking the flow of the arguments, we relegate the discussion of these tests to Section A.5 of this dissertation.

Since the provincial (national) unemployment rate is a serially correlated grouped regressor, it is well-known that standard inference is misleading in that it tends to over-reject null hypotheses (Bertrand et al., 2004). If, as in our case, the number of clusters is small (40 in the provincial model and 8 in the national model),[21] it is very difficult to obtain correctly sized tests (Angrist and Pischke, 2009; Cameron et al., 2008). In Section 1.5.2 we propose a new inference procedure that builds on the work of Brewer et al. (2013) and Wooldridge (2006, 2010), but that is modified to take the specificities of our data into account.

1.5.1 The Benchmark Linear Regression Model

The most general benchmark provincial regression model, which is separately estimated for low and for high educated men, is specified as follows:

$$y_{igpt} = f_g(e)u_{gp} + f_{gu}(e)u_{gp}\mathbf{1}[u_{gp} < u_{(g-1)p}] + \theta_e + \phi_t + f_t(e)u_{pt} + \eta_p + \omega_p t + f_0(g) + x_i'\delta + \varepsilon_{igpt} \tag{1.1}$$

where

- y_{itgp} is the labor market income of interest in calendar year t for individual i who graduated in year g while living in province p at the start of graduation year g;[22]

- u_{gp} is the unemployment rate in province p and graduation year g. $f_g(e)$ is a linear spline in potential experience,[23] so that the interaction with u_{gp} represents the effects of interest, the long-run effects from graduating in downturns;

- $\mathbf{1}[.]$ denotes the indicator function which is one if the expression between brackets holds and zero otherwise, so that $\mathbf{1}[u_{gp} < u_{(g-1)p}]$ is equal to one in case of an upturn in the graduation year. $f_{gu}(e)$ is again a linear spline in potential experience. This second term estimates the long-run effects from graduating in upturns;

- θ_e is a fixed effect for potential experience ($e \equiv t - g$), $f_0(g)$ is a linear spline in graduation year and ϕ_t is a calendar year fixed effect;[24]

[21] In the provincial model the number of clusters is equal to the 8 graduation years times the 5 provinces.

[22] To avoid cumbersome notation, we ignore the subscript in p_g.

[23] This terminology is borrowed from the literature. It counts all calendar years since graduation instead of actual years of employment experience.

[24] Even after dropping the fixed effect of the reference category, calendar time effects cannot be separately identified from the effect of potential experience and graduation year, because of the accounting identity $e \equiv t - g$. Since the calendar time effect is not our main interest, we follow Oreopoulos et al. (2012) and ensure identification by just dropping a second calendar time fixed effect rather than imposing that the cohort effects sum to zero, as in Deaton (1997).

- u_{pt} is the unemployment rate in calendar year t in the province p in which the individual resides at the start of the graduation year g. This controls for provincial specific shocks in the unemployment rate not captured by ϕ_t and potentially correlated with u_{gp}. It is interacted with $f_t(e)$, a linear spline in potential experience, which allows the current unemployment rate to have different effects over time;

- η_p is a fixed effect for the province of residence at graduation;

- ω_p is the provincial specific coefficient of a linear time trend. We include these provincial specific time trends, because the unemployment rates exhibit differential downward time trends (see Figure 1.2);

- x_i are individual control variables measured at age 17 reported in Table 1.1;

- ϵ_{igpt} are the remaining errors, which include random unobserved province-graduation cohort effects and measurement errors that may be serially correlated (see Section 1.5.2).

The linear splines are expressed as follows:

$$f_m(\tau_m) = \alpha_m + \sum_{j=0}^{J_{\tau_m}} \beta_{mj}.(\tau_m - 3j)\mathbf{1}[\tau_m \geq 3j] \;\; for\; m \in \{g, gu, t, 0\} \tag{1.2}$$

where, except for $m = g$, $\alpha_m \equiv 0$, $\tau_m = e$ if $m \in \{g, gu, t\}$, and $\tau_0 = g$. β_{m0} is the slope of the linear function in potential experience (if $m \in \{g, gu, t\}$) or graduation year (if $m = 0$), and β_{mj} is the magnitude by which this slope changes every three years (i.e. if $\tau_m \geq 3j$). The slope may change at most J_{τ_m} times. $J_e = 3$, but for the high educated sample we set $\beta_{m3} = 0$ for $m \in \{g, gu, t\}$, since at most 10 years of potential experience are considered instead of 12 years for the low educated.[25] $J_g = 2$, since only 8 graduation years are considered.

$f_g(e)$ is the function of interest, since it describes for each year e since graduation the effect of a percentage point increase of the unemployment rate at graduation on the outcome of interest. We allow for an asymmetric effect in an upturn which, similarly, is measured by a percentage point decrease of the unemployment rate at graduation.[26] The effect of a percentage point decrease of u_{gp} is $-(f_g(e) + f_{gu}(e))$. $f_t(e)$ allows for a differential effect of the current provincial unemployment rate from the common calendar year effect ϕ_t. $f_0(g)$ controls for common graduation year effects.

[25] We restrict the number of graduation and experience years in the analysis to avoid too small cell sizes, as this is problematic for inference (See Section 1.5.2). Beyond 10 (12) years of experience not all birth cohorts of the high (low) educated are observed any longer.

[26] This specification has been proposed by Genda et al. (2010). A simple rationale for asymmetric persistence of positive and negative shocks on employment and wage is offered by the insiders-outsiders theory, if newly hired workers in upturn gain influence in wage process gradually while insiders who are fired in downturns immediately loose it (Begg et al., 1989; Lindbeck and Snower, 2001). Higher asymmetries should be expected the higher the labor turnover costs and, in case of dismissal in downturns, the higher the insiders' transitions into unemployment compared to job-to-job movements and the lower the importance of firm specific skills compared to general (so that insider status is quickly lost).

For all outcomes we always start by estimating the most general specification as specified in (1.1). However, to enhance precision, we check each time whether we can impose the following restrictions: all splines for the upturn and the current unemployment rate u_{pt} jointly equal to zero; all θ_e and ϕ_t replaced by a linear spline as defined in (1.2); all ω_p jointly equal to zero. In Section 1.6 we only report the results of the estimations in which restrictions are imposed that could not jointly be rejected at the 5% level. We report in the tables which restrictions are imposed.

Finally, in a sensitivity analysis we also estimate a model in which we exploit the time variation of the national unemployment rate. In this national model the subscripts p disappear and also $f_t(e)u_{pt}$, η_p and $\omega_p t$, since these terms are then no longer identified.

1.5.2 Inference with a Small Number of Clusters

Brewer et al. (2013) recently proposed a straightforward method for inference in difference-in-differences (DiD) design with grouped errors. They demonstrate in Monte Carlo analysis that correctly sized tests can be obtained by using bias corrected clustered standard errors in an ordinary least squares (OLS) regression of the covariate-adjusted group-time means of the dependent variable on the covariates varying at the group-time level. The bias correction is simple to implement, because STATA correctly scales the standard errors by default. To enhance the power of this approach, the authors exploit the serial correlation in the grouped errors using the feasible generalized least squares (FGLS) estimator proposed by Hansen (2007)[27] that explicitly allows for a common autocorrelation pattern (e.g. AR(2)) across groups. To allow for misspecification of this autocorrelation process the aforementioned cluster robust inference is applied to this FGLS estimator. This delivers correctly sized tests and, if the number of time periods is sufficiently large (from about 10 time periods), yields substantial power gains.

Since our model can be seen as a generalized DiD setting, in which we have variables that vary at the group level (gp), i.e. each combination of graduation year (g) and province (p) is a cluster, at the time level ($t = g + e$), and at the group-time level (gpt), this approach can be applied to our analysis. However, in contrast to Brewer et al. (2013), group-time cells in our sample contain a relatively small number of observations, so that we cannot ignore measurement error in the covariate-adjusted group-time means of the dependent variables. To generalize their approach, we build on the work of Wooldridge (2006, 2010). Wooldridge proposes a FGLS estimator in case of cross-sectional data with only measurement error and no unobserved group effects. We adjust this method for panel data and show how, as in Brewer et al. (2013), autocorrelated unobserved group effects can be integrated in this approach.

In a first step, run a regression of y_{igpt} on x_i and group-time dummies using the micro-data on the individual level:

$$y_{igpt} = \mu_{gpt} + x_i'\delta + \epsilon_{igpt} \tag{1.3}$$

where μ_{gpt} are the group-time fixed effects, i.e. the covariate-adjusted group-time means, and ϵ_{igpt}

[27] Brewer et al. (2013) show that Hansen's *bias-corrected* FGLS delivers only little more power than the *ordinary* FGLS.

is the error term of this micro regression. In a second step, the estimated group-time fixed effects $\hat{\mu}_{gpt}$ are regressed on the group-time level covariates:

$$\hat{\mu}_{gpt} = f_g(e)u_{gp}+f_{gu}(e)u_{gp}\mathbf{1}[u_{gp} < u_{(g-1)p}]+\theta_e+\phi_t+f_t(e)u_{pt}+\eta_p+\omega_p t+f_0(g)+v_{gpt} \quad (1.4)$$

where $v_{gpt} = e_{gpt} + (\hat{\mu}_{gpt} - \mu_{gpt})$, e_{gpt} is the unobserved group-time shock measured at calendar time t and $(\hat{\mu}_{gpt} - \mu_{gpt})$ is the measurement error in the covariate-adjusted group-time means. Brewer et al. (2013) assume the latter to be zero. Consequently, even if cluster robust standard errors still result in correct inference, taking the (co-)variances of the measurement errors into account could enhance efficiency.

In the case of cross-sectional data, Wooldridge (2006, 2010) proposes implementing the *efficient* Minimum Distance (MD) estimator, also called the 'Minimum Chi-Square' estimator, of the covariate-adjusted group means on the group level explanatory variables. This consists in estimating (a cross-sectional) version of (1.4) by FGLS. If $e_{gpt} = 0$, the optimal weight in the FGLS is the inverse of the variance matrix of $(\hat{\mu}_{gpt} - \mu_{gpt})$ estimated in the first step. Since the efficiency of this procedure depends on whether unobserved group-time shocks e_{gpt} are indeed zero, it is useful to notice that this can be tested for. If the observed group level explanatory variables cannot fully explain the variation in $\hat{\mu}_{gpt}$, the regression model (1.4) is likely to be rejected against the saturated model, i.e. the weighted sum of squared residuals (WSSR), distributed χ^2 with degrees of freedom equal to the number of groups minus the number of estimated parameters, is larger than the conventional rejection level.

Generalizing Wooldridge (2006, 2010)'s approach to panel data requires accounting for the serial correlation in the error term ϵ_{igpt} of the first step regression. We do this by taking the individual i as clustering unit in the first step and use the conventional cluster-robust variance matrix of the $\hat{\mu}_{gpt}$ estimated in the first step as weighting matrix in the second step.[28] The χ^2 goodness-of-fit statistic allows testing for the presence of unobserved group-time shocks, i.e. $e_{gpt} \neq 0$. In case of no rejection, the conventional standard errors can be used for inference. In case of rejection,[29] the (bias-corrected) cluster robust standard errors allow for serial correlation in e_{gpt}.

In case of rejection, we attempt to increase power by explicitly allowing for the variance in e_{gpt} in addition to that of the measurement error, and for a particular serial correlation pattern in e_{gpt}, as in Brewer et al. (2013). To this purpose, we follow Amemiya and Nold (1975) by assuming that the measurement errors $(\hat{\mu}_{gpt} - \mu_{gpt})$ and the unobserved group-time shocks (e_{gpt}) are uncorrelated, and that the variance of e_{gpt} is homoskedastic across groups and time, so that we can estimate the latter by subtracting an estimate of the average variance of the measurement error $(\widehat{var}(\hat{\mu}_{gpt}) \equiv \hat{s}^2_{gpt})$ from an estimate of the average variance of the composite error term v_{gpt}:

$$\widehat{var}(e_{gpt}) \equiv \hat{\sigma}^2_e = \frac{1}{GPT}\sum_{g=1}^{G}\sum_{p=1}^{P}\sum_{t=1}^{T}\left(\hat{v}^2_{gpt} - \hat{s}^2_{gpt}\right) \quad (1.5)$$

[28]Since the individual is taken as clustering unit, the number of clusters is sufficiently large to implement conventional inference procedures.

[29]We use the conventional 5% as threshold for the size of the test.

where $\hat{v}_{gpt}$ denotes the residual of an OLS regression of (1.4), G the number of graduation years, P the number of provinces, and T the number of calendar years retained in the grouped regression.[30] Subsequently, this estimated variance can be added to the diagonal of the variance matrix of the measurement error to obtain an estimate of the (heteroskedastic) variance matrix of v_{gpt}: $\widehat{var}(v_{gpt}) = \hat{\sigma}_e^2 + \hat{s}_{gpt}^2$. The inverse of this variance is then used to estimate the parameters in (1.4) by FGLS.[31] Depending on whether the χ^2 goodness-of-fit statistic rejects the model or not, respectively conventional or cluster robust standard errors should be calculated after this modified FGLS.

In this empirical application we find for most outcomes that $\hat{\sigma}_e^2 < 0$,[32] even if the aforementioned χ^2 rejects the model and, hence, suggests that unobserved group-time shocks are present. This suggests that the imposed homoskedasticity assumption is not satisfied, and that it is thus difficult to enhance power by explicitly taking the unobserved group-time shocks, in addition to the measurement error, into account. The cluster robust standard errors calculated after the FGLS that just takes measurement error into account still provides correct inference, however. These are therefore the ones that we report in most of our estimations when the goodness-of-fit statistic rejects the model.

Finally, we explain how we deal with a number of practical issues encountered with the proposed inference methods. First, the benchmark outcomes must satisfy adding-up constraints: (i) the indicator of salaried employment and the one of self-employment sum to the indicator of overall employment; (ii) log hourly wages and log annual hours worked sum to log annual earnings; (iii) the sum of the annual number of hours worked full-time and part-time is equal to the total annual hours worked. These adding-up constraints are automatically satisfied if the first and second step regression models, (1.3) and (1.4), are estimated by OLS. However, this is no longer true if FGLS is applied in the second step on each outcome separately, since then the weighting matrices ignore the correlation that these constraints impose on these outcomes. To overcome this problem, we jointly estimate both the first and the second step in a *seemingly unrelated regression* (SUR), as proposed by Zellner (1962). Since the adding-up constraint makes the variance matrix of the three outcomes singular and hence non-invertible, we leave out one of the three outcomes and calculate the parameters and standard errors of the third model from the constraint.[33] An estimate of the variance matrix, the inverse of which is used as weight in the second step FGLS SUR, is obtained from the conventional cluster robust estimate of the variance matrix of the covariate adjusted means $\hat{\mu}_{gpt}$ calculated after a pooled OLS on the first step SUR. By clustering at the individual level in the first step, the variance matrix accounts not only for unrestricted serial

[30] In the data the number of calender years varies by group, i.e. by gp combination. Therefore T should be indexed by gp, but to avoid cumbersome notation we do not do this.

[31] Cockx and Dejemeppe (2005) show how an AR(1) process on e_{gpt} can be modeled.

[32] We only find $\hat{\sigma}_e^2 > 0$ for salaried employment rate in the national model, both the low and high educated group (see Section A.9 of this dissertation).

[33] Barten (1969) has shown that the parameter estimates are invariant to the equation deleted. However, Berndt and Savin (1975) have demonstrated that in case a model with autoregressive disturbances is modeled invariance requires restrictions on the parameters of the autoregressive process.

correlation in the outcomes, but also for unrestricted correlation across outcomes.

Second, in our data we find cases in which the employment status of all individuals belonging to a cluster gp does not vary over some calendar years t. This induces perfect serial correlation in the covariate-adjusted group-time means μ_{gpt} and, hence, the cluster robust variance matrix of these $\hat{\mu}_{gpt}$ is singular. We therefore use in these cases the Moore-Penrose generalized inverse of the variance matrix as weight in the second step FGLS. To avoid numerical imprecision, we manually set as many eigenvalues to zero as the number of times that the employment rate for particular groups is repeated over time. This accordingly reduces the number of degrees of freedom in the second step.

Finally, asymptotic inference for the Minimum Chi-Square estimator is only valid if groups are sufficiently large. In the statistical literature some rules of thumb are suggested for what is large enough for the Central Limit Theorem to apply. For continuous outcome variables (such as log hours, log wages or log earnings) a group size (N_{gpt}) of 30 observations is typically considered sufficient, while for dichotomous outcomes (such as the employment rate) the minimum of the expected number of *successes* and *failures* should be sufficiently large. A commonly accepted rule for the latter is that $\min\{N_{gpt}P_{gpt}, N_{gpt}(1 - P_{gpt})\} \geq 5$, where P_{gpt} denotes the probability of success and which can be estimated by aggregating the individual predictions of this probability in the first step OLS regression of (1.3) to the cluster-time level gpt. According to Cochran (1954) the approximation is, however, still acceptable if for less than 20% of the groups this expectation is smaller than 5 while remaining larger than 1.

For the national model these rules are satisfied if we restrict the analysis to graduation years 1994-2001 for the low educated and to 1997-2004 for the high educated. For the provincial model we must drop additional groups. For the continuous outcomes, applying the aforementioned rule reduces the sample size too much, so that we retain groups-time cells containing between 16 and 30 observations, which still delivers a reasonable approximation if the distribution of the underlying random variable does not differ too much from the Normal. For the dichotomous outcomes, we calculate for each group-time cell and outcome the aforementioned expectations, take the minimum of these expectations over the outcomes retained in the same SUR, and drop group-time cells with the smallest minimum until the aforementioned Cochran's rule is satisfied.

Dropping these cells introduces, however, a concern of selectivity. We therefore test for this. We construct for each outcome an indicator that is equal to one if the individual belongs to a group-time cell that is dropped according to the aforementioned rules and zero otherwise. Subsequently, we use these indicators as dependent variable in a one-step regression on model (1.1) in which we impose the same restrictions as the ones used for the corresponding outcome, and in which we cluster the standard errors by group gp. Finally, we test the null hypothesis that all the coefficients of the linear spline ($f_g(e)$) that interacts the unemployment at graduation (u_{gp}) are jointly significantly different from zero. Since the number of clusters is small, we tend to over-reject the null-hypothesis. But the null hypothesis is never rejected in any of the considered outcomes, so that we can therefore be confident that selectivity is not an issue.

In Table 1.2 in Appendix 1.8 we report for the benchmark continuous and dichotomous outcomes the number of cells that are dropped and retained, as well as the mean and maximum size of these cells. We also provide the aforementioned statistics for cells that are retained, but that do not satisfy the aforementioned stricter rules, i.e. for cell sizes between 16 and 30 if the outcome is continuous, and for cells for which the minimum of the aforementioned expectation is smaller than 5 in case of a dichotomous outcome. Finally, we also include the P-value of the joint test of selectivity mentioned in the previous paragraph.

1.6 Results

We now discuss our findings for the low educated and high educated sub-samples. We will report for each outcome the effect of a percentage point (*pp*) increase in the unemployment rate by the potential number of years of experience in the labor market since graduation. This corresponds to what is commonly reported in studies, but notice that it is difficult to use these effects to make cross-country comparisons. This is because a *pp* rise of the unemployment rate in a typical recession depends on the degree of labor market flexibility, and also on whether the overall, or youth unemployment rate is used in the analysis. In a rigid labor market the unemployment rate varies less over the business cycle, than in a flexible one, and the youth unemployment rate tends to vary more than the overall rate. The problem is that studies usually do not report by how many *pp*'s the unemployment rate increases in a typical recession. An exception is the study of Oreopoulos et al. (2012) for Canada, characterized by a flexible labor market. The authors report that the unemployment rate rises by five *pp*'s in a typical recession. To compare, in Flanders this figure was only 1.4 *pp*'s on average in the 1994-2010 period (and by 1.6 point only in the Great Recession in 2008).[34] Consequently, if we want to compare the effect of a typical recession in Canada to such one in Flanders, we must multiply the effect of a *pp* increase of the unemployment rate in Canada by a much larger factor than in Flanders. This can put the findings in a different perspective. For instance, despite the short-term effect of a one *pp* increase of the unemployment rate on earnings is much smaller for high educated Canadian graduates (-1.8%) than for Flemish (-5.8%), the effect for a typical recession is comparable or even slightly larger in Canada (-9% versus -8.1%).

1.6.1 The Findings for the Low Educated

In Figure 1.3 we graphically report for the low educated sample the effect of increasing the provincial unemployment rate at graduation by one *pp* on the main outcomes of interest. The reader can find in Table 1.3 in Appendix 1.9 the corresponding point estimates and standard errors, as well some information about the estimated model, among which the P-value of the goodness-of-fit statistic, whether cluster-robust or conventional standard errors are used (conventional standard errors are used if the aforementioned P-value exceeds 5%), and which restrictions were imposed on the benchmark specification (1.4). The complete estimation results (including the first step

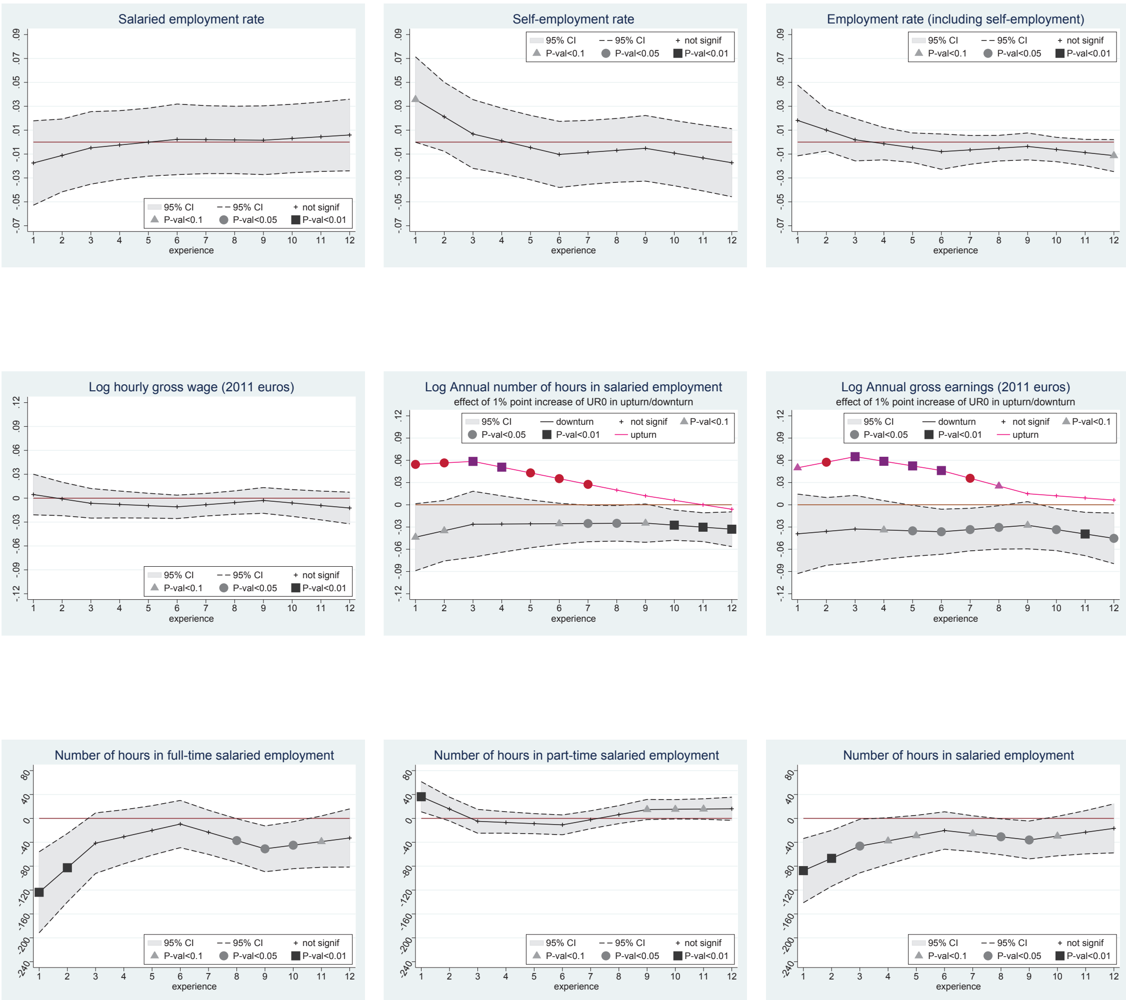
[34] A recession is defined to be a period during which the unemployment rate increases uninterruptedly. Hereafter, we will refer to a "typical recession" as the average *pp* increase of the provincial unemployment rate during such recessions in Flanders in the 1994-2010 period.

(1.3) and the second step OLS) are only reported in Section A.7 of this dissertation. Notice that for all provincial models (1.4) was estimated in a SUR by FGLS in which the estimated variance of the unobserved group effects was set to zero ($\widehat{var}(e_{gpt}) = 0$), since in all cases it was estimated to be strictly negative in (1.5). It is actually only found to be strictly positive for one outcome only in case that the variation of the national unemployment rate was exploited (see Section A.9 of this dissertation).

The top panel of Figure 1.3 reveals that a recession at graduation does not significantly (at 5%) affect employment, but confidence intervals are relatively wide, especially for salaried and self-employment. The point estimate suggests that a one *pp* increase of the unemployment rate at graduation increases the self-employment rate by 3.6 *pp*'s (significant at 10%), but this effect rapidly drops to zero after three years. This provides weak support that a recession *pushes* young low skilled graduates to self-employment by lack of salaried employment.[35] It also suggests that there might be some pressure from employers in economic downturns to take-up *pseudo* self-employment, as avoid the minimum wage restrictions and to be able to pay lower Social Insurance contributions that are due for self-employed workers (see e.g. European Commission (2010)). In line with this evidence, the salaried employment rate decreases, but to a lesser extent: according to the point estimates, -1.7 and -1.1 *pp* in the first and second year of potential experience, but these estimates are not significant. Since the point estimate of the effect on salaried employment is very close to zero beyond the first two years of potential experience, we should only be concerned that the effect of the unemployment rate at graduation on variables that are solely observed in case of salaried employment, such as wages, hours and earnings, is biased as a consequence of selectivity (Heckman, 1974) in the first few years after graduation, if at all.

[35]There are two competing hypotheses on the role of the business cycle on self-employment. Aside the mentioned *push* hypothesis, the *pull* hypothesis states that high unemployment negatively affects individual expectations about the success of self-employment, or reinforce credit constraints. Empirical evidence on the relative importance of these hypotheses is mixed. For recent evidence, see e.g. Fairlie (2013) and Yu et al. (2014).

Figure 1.3: The Effect of a One Percentage Point (pp) *Increase in the Provincial Unemployment Rate at Graduation: Low Educated.*



The figure displays for the low educated the effect of increasing (decreasing in case of an upturn) the provincial unemployment rate at graduation by one *pp* on the main outcomes of interest. The effects reported in top, middle and bottom panel result each time from two-step FGLS estimates of a SUR model on the first two outcomes described in detail in Section 1.5.2. The effects for the third outcome are obtained from the following adding up constraints: salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked. Table 1.3 in Appendix 1.9 reports the corresponding point estimates displayed in the figure.

In the middle panel of Figure 1.3 we observe that the average hourly wage is hardly affected by adverse conditions at graduation. At the same time the annual number of hours worked drops persistently during the 12 first years of potential experience, starting from a decrease of 4.4 *pp* in the first year after graduation, increasing slightly to a value that ranges between -2.5 and -3 *pp* from year three to twelve. From the bottom panel we can deduce that nearly all the effect on hours worked is induced by a reduction in hours worked full-time, and not by an increase in hours worked part-time. Only in the first year after a recession, the number of hours worked part-time increases significantly. These findings are in line with the hypothesis that the minimum wage is binding for this low skilled youth and, hence, induces a higher incidence of unemployment.[36] They are also in accordance with the literature, mentioned in the Introduction, that early career unemployment permanently increases the likelihood of unemployment later in the career (e.g. Schmillen and Umkehrer (2013b)). However, they contrasts starkly with the sharp temporary downward adjustment in wages, and marginal effects on employment and hours that Genda et al. (2010) report for low educated youth in the flexible US labor market.

The effect on earnings nearly mirrors that on hours worked. This is because the effect on wages is close to zero. We also find that the effect of an upturn at graduation on hours worked is not symmetric to that of a downturn. In an upturn the number of hours increase more importantly than they decrease in a downturn. Up to four years after an upturn, the hours worked increase by 5 to 6 % for each *pp* decrease of the unemployment rate at graduation. Thereafter the effect gradually fades to zero after ten years. This means that the costs of a recession are lower than the benefits of a boom, but also that the costs are more persistent.

Even if the wage is hardly affected by the business cycle at graduation, we find that an increase of one *pp* of the unemployment rate at graduation *persistently* decreases the median wage in the recruiting firms by roughly 1.5% over all 12 years following labor market entry and this effect is always significant, or close to (see Table 1.4). This suggests that scars of a recession are not only induced by a higher incidence of unemployment, but also by the lower quality of the first job, which is not directly apparent, because of the downward rigidity imposed by the minimum wage. This is in line with the observation in the middle panel of Figure 1.3 that some years after labor market entry a recession at graduation does seem to have a growing negative impact on the hourly wage. In the benchmark model this growing negative impact is not statistically significant. However, when in the sensitivity analysis we consider the second step OLS instead of FGLS estimates this impact steadily decreases and becomes significantly negative at the 5% level beyond seven years of experience (see Table A.22 of this dissertation). This is compatible with the hypothesis that lower quality jobs entered during a recession have a less steep wage profile with seniority than the higher quality jobs accessed during a tight labor market.

Since EPL is much stricter for white collar than for blue collar workers (see Section 1.3) and the high educated are predominantly employed as white collars and the low educated as blue collars (see Section 1.4), firm mobility is expected to be higher for the low than for the high educated. This is only to a minor extent reflected in the fraction that remains employed in the same firm. Five (ten) years after graduation this fraction is 26% (15%) for the low educated and 31% (17%) for the high educated.[37] This may be a consequence of the system of STC, which allows firms to retain blue collar workers over the business cycle (see Section 1.3). By contrast, 24% (18%) of the low educated changes jobs in more than two (four) out of the first (ten) years of potential experience against 12% (7%) of the high educated. This suggests that a lower EPL strictness does not so much enable more upward mobility, but that it rather leads to a more

[36]This is not incompatible with a zero effect on the employment rate, because the latter is only affected if the unemployment spell is not interrupted by any employment throughout a complete calendar year.

[37]For more detailed statistics, see Chapter A, Table A.11.

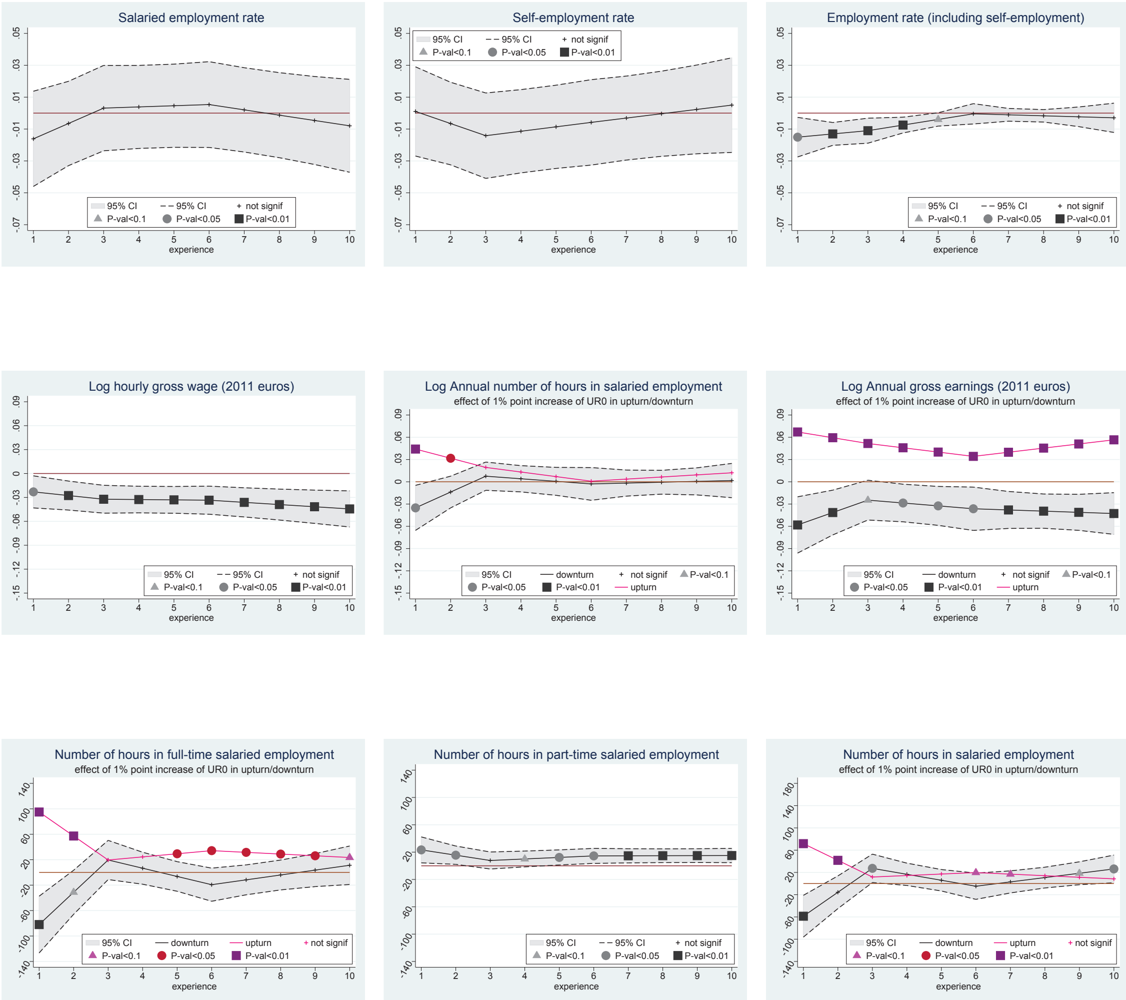
pronounced segmentation of the labor market in, on the one hand, a segment with relatively stable employment and, on the other hand, a segment in which workers cycle from one short tenured job to another. But this relationship is not necessarily causal, since this type of segmentation can be a characteristic of the labor market for low skilled workers. In any case, the less strict EPL does not allow the unfortunate low educated, entering the labor market during a recession, to catch up with the more fortunate group. The unemployment rate at graduation is unrelated to firm mobility (see Table 1.4).

Finally, geographic mobility is negligible. During the observation period each year on average only 1.5% of the low educated changes residence to another province. It is therefore not surprising to find that the unemployment rate at graduation is unrelated to geographic mobility (see Table 1.4).

1.6.2 The Findings for the High Educated

The findings for the high educated can be found in Figure 1.4 and Table 1.5. In the top panel of Figure 1.4 we see that the overall employment rate of the unfortunate cohorts is during the first five years after graduation slightly, but significantly lower than that of the more fortunate cohorts. The point estimates gradually rise from -1.5 *pp* to zero in these years. After five years the effects remains very closely to zero. The initial significant negative effect on employment for the high educated contrasts with the insignificant effect for the low educated. The high educated may be less liquidity constrained than low educated, and, hence, in face of a recession search longer for jobs matched to their qualifications, or may even start studying again. In view of the wide confidence intervals, it is difficult to assign this negative impact to a decrease in salaried or self-employment. According to the point estimates, the drop in the first year is completely explained by the lower salaried employment rate, while in the subsequent years it seems that the lower self-employment rate is more the driving force. In any case, the impact on the employment rate is relatively small, which is to be expected, because higher educated are more in demand, irrespectively of the business cycle. Hence, also the aforementioned selectivity issue for the labor market outcomes in salaried employment is likely to be negligible, or matters at most during the first two years after graduation.

Figure 1.4: Effect of one pp *Increase in the Provincial Unemployment Rate at Graduation: High Educated.*



The figure displays for the high educated the effect of increasing (decreasing in case of an upturn) the provincial unemployment rate at graduation by one *pp* on the main outcomes of interest. The effects reported in top, middle and bottom panel result each time from two-step FGLS estimates of a SUR model on the first two outcomes described in detail in Section 1.5.2. The effects for the third outcome are obtained from the following adding up constraints: salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked. Table 1.5 in Appendix 1.10 reports the corresponding point estimates displayed in the figure.

In contrast to our findings for the low educated, a recession does have a significant negative impact on the hourly wage. In the first two years it is slightly lower than in the following ones, but this could be a consequence of the aforementioned selectivity problem. Those who are unemployed in the first years after a recession are likely to be less productive and earning lower wages than the group that is employed in these years. This may bias the negative effect on wages slightly upwards. From the third year this bias should no longer play a role. We observe that the wage penalty rises steadily (and significantly) from then onwards, starting at -3.2% after three years and attaining -4.4% after ten years. The initial wage penalty suggests that high educated graduates are forced to accept lower quality jobs for which they are possibly overeducated (see Baert et al. (2013) for evidence on this using the same survey data) and which pay lower wages. As a consequence of the high labor market rigidity, in particular the strict EPL (see Section 1.3), job mobility is limited (see more evidence of this below). Hence, these unlucky generations are trapped in these lower quality jobs and definitely miss the opportunity to be hired in career jobs with a steeper seniority pay profile. In line with this hypothesis, the aforementioned study of Baert et al. (2013) demonstrates that young unemployed graduates who accept a job for which they are overeducated remain trapped in these jobs for many years. This explains why the wage penalty rises over time rather than gradually fades, as in more flexible labor markets (Genda et al., 2010; Oreopoulos et al., 2012).

The annual number of hours worked is 3.5% lower for each *pp* rise in the unemployment rate at graduation, but rapidly rises to zero after three years, level at which it remains subsequently. From the bottom panel of Figure 1.4 we can deduce that this is nearly entirely[38] due to a reduction in full-time hours. This suggests that those who refuse to downgrade become temporarily unemployed. But in contrast to lower educated workers, this does not permanently increase the likelihood of unemployment later in the career, but rather raises the likelihood of remaining employed in lower quality jobs.

The combined effect on wages and hours yields the effect on earnings. At the start of Section 1.6, we have already made a comparison of this effect with that of the more flexible labor market in Canada. Notice also, that the effect on hours worked is, as for the low educated, asymmetric for up- and downturn. In an upturn the annual number of hours worked increases slightly more and slightly longer.

In the previous section we already described firm mobility, and explained that the strict EPL for white collar workers lowers it for the high relative to low educated. Nevertheless, in contrast to the low skilled, we do find some evidence (see Table 1.6 in Appendix 1.10) that the high skilled who graduate during a downturn are more likely to move from one job to another in the following year. This may mitigate the initial wage penalty. However, in contrast to what is observed in more flexible labor markets (see e.g. Oreopoulos et al., 2012), in all subsequent years the unlucky graduates are not more (nor less) mobile than the lucky ones. In line with this and the aforementioned evidence, the median wage in the recruiting firm is during the first ten years of labor market experience persistently smaller (on average 3.3%) for each *pp* increase of the unemployment rate at graduation. Hence, the unfortunate workers remain trapped in lower quality jobs.

Finally, on average 3.2% of the high educated move to another province each year. This is more than double as much as for the low educated. Nevertheless, we cannot find any evidence that geographic mobility is significantly related to the unemployment rate at graduation (see Table 1.6).

[38] We also observe a small, but permanent increase in the number of hours worked part-time.

1.6.3 Sensitivity Analysis

In Section A.8 of this dissertation we report a number of sensitivity analyses for the six main outcomes that we considered in the analysis: the three indicators of employment, log hourly wage, log annual hours worked and log annual earnings. In a first sensitivity analysis we include dummy variables for each number of completed years of education as individual control variables in the first step regression. If the timing of education is unrelated to the unemployment rate, the inclusion of these variables should not affect the conclusions. This is what we find.

Second, we contrast the provincial to the national model. In the national model identification relies on the calender time variation of the unemployment rate at labor market entry in deviation from the parametrically specified (linear spline) dependence of graduation time. Because of this strong identifying assumption, we do not find a complete correspondence between the models, but overall the patterns of the effects are quite comparable, certainly if we account for the higher degree of imprecision in the national model. In particular, we find a larger negative effect on hours worked (and, hence in earnings) in the first year after graduation, and for the high educated also in years nine and ten. In addition, the effect on the self-employment rate of the high educated displays a strong positive trend from experience years six to ten, which we did not observe in the provincial model.

Third, for the dichotomous indicators we always estimated a linear probability model in the first step. As a sensitivity analysis, we estimate a probit model in the first step, and, subsequently, the covariate-adjusted group-time means on the aggregate regressors by FGLS. In Tables A.21 and A.23 of this dissertation we report the partial effects on the probability of employment for each year of potential experience, where the other aggregate regressors are evaluated at their sample mean. These partial effects are very similar to the ones that we found in the benchmark linear probability model. The main difference is that the estimates of probit model are much less precise.

The benchmark model was estimated by FGLS on the second step regression model (1.4). This should generally deliver more precise estimates than those obtained by OLS with cluster robust standard errors in this second step. For the continuous outcomes, FGLS results indeed generally in more precise estimates than OLS. Exceptions are the effects on the log hourly wage for the low educated (and only to slight extent for the high educated).[39] However, this can be a consequence of the log hourly wage being estimated jointly with log annual hours worked in a SUR regression. The larger standard errors in the wage equation are more than compensated by the higher precision of the effects in the hours equation. For the dichotomous variables, FGLS delivers in this particular application somewhat less precise estimates of the effects of interest. Only for the effect on the overall employment rate of the high educated the standard errors obtained by FGLS are consistently smaller.

Finally, we compare OLS applied to the one-step regression model (1.1) to OLS applied in the second step. In both cases we cluster standard errors by province and year of graduation combi-

[39]Notice, as mentioned in Section 2.5, the effect of a *pp* increase of the unemployment rate at graduation is now found to be significantly negative as from seven years after labor market entry for low educated.

nation (gp). Point estimates are very comparable, but because of the small number of clusters, we expect the standard errors of effects in the one-step OLS regression to be downward biased. This is what we find for the continuous outcomes. However, for the dichotomous outcomes this is only the case for the effect on the overall employment rate of the low educated. This suggests that for this particular application the downward bias is important for the continuous outcomes, but not for the dichotomous ones.

A last sensitivity analysis refers to the impact of the dot-com recession,[40] which occurred between March 2001 and November 2001 according to the official definition.[41] Compared to others, this recession was quite mild in terms of decline in GDP, but was followed by a "jobless recovery", i.e. the subsequent GDP growth was not accompanied by a rise in employment. In Flanders this recession caused the unemployment rate to rise by 1.7 *pp* in 2002-2003 (see Figure 1.1); the provincial unemployment rates remained at quite a high level also in 2004 and did not fall before 2005 (see Figure 1.2). Since the dot-com recession emerged during the graduation period of the high educated (i.e. the 1997-2004 period), it is important to ensure that the results for the high educated are not driven by this recession, but that they hold for the entire graduation period.[42] An additional argument for this robustness check is that this recession was largely unanticipated by forecasters Kliesen (2003). As a consequence, it is likely that students did not opportunistically decide the timing of or the province at graduation in view of the incoming recession: this runs in favor of our identifying assumption. We run this sensitivity by adding to equation (1.1) a dummy equal to one if graduation occurs in 2002-2004 (i.e. the jobless recovery following the dot-com recession) and with an interaction between this dummy and $f_g(e)u_{gp}$. This interaction identifies the differential short-run and long-run effects of increasing the provincial unemployment rate at graduation by one *pp* when graduation occurs in the period 2002-2004 versus the preceding period 1997-2001. Subsequently, we compare graphically the scarring effects of one *pp* increase in the unemployment rate when graduating in the period 1997-2004 - from the benchmark models in Figure 1.4 - with those of graduating in the 2002-2004 period - from the sensitivity analysis. The bottom line is that the results for the high educated are broadly the same when focusing on the dot-com recession and on the entire graduation period. In Appendix 1.11 we report a more detailed explanation of this sensitivity analysis and the aforementioned graphical comparison.

1.7 Conclusion

In this paper we studied the short- and long-run impacts of graduating in a recession in Flanders, the largest region in Belgium that is characterized by a very rigid labor market. We had access to unusually rich data: comprehensive administrative data from the various institutions of the Belgian Social Insurance system contain detailed information on individual labor market performance in salaried employment (earnings, wages, hours worked) and self-employment up to

[40]It followed a period of stock price rise of internet-based firms that culminated in the collapse of the bubble.
[41]In Europe, recessions are defined by two consecutive quarters of negative GDP growth.
[42]A similar sensitivity cannot be run for low educated as they graduate in 1994-2001, just before the recession.

twelve years since graduation. These data were matched with survey data that follow closely the educational path of a random sample of Flemish youth from the beginning of secondary education until the first years in the labor market. The surveys allow us to precisely define the moment of graduation, which is crucial in our analysis and typically not observed in administrative data. Moreover, we propose a new method of inference, which is complicated by the fact that since the regressor of interest, the provincial unemployment rate at graduation, is a serially correlated variable that is measured at a grouped level. We combine the efficient Minimum Chi-square estimator of Wooldridge (2006, 2010) for cross-section data with the FGLS of Brewer et al. (2013) for difference-in-differences approach and come up with a novel two-step FGLS estimator that accounts for serial correlation of the panel; whenever possible we improve the efficiency of the estimator by explicitly allowing for the variance in the grouped error. Our identifying assumption is that the composition of graduates by year and province is not affected by the provincial unemployment rate. This may be violated if students change the timing of - or the province of living at - graduation depending on the business cycle. We show that inter-provincial mobility is negligible and propose a new way to test that the timing of graduation is unrelated to the unemployment rate based on a discrete duration model. Due to the limited sample size, we tackle several difficulties arising from too small group size.

We analyzed long-run effects of graduating in a recession separately for high and low educated workers, because labor market institutions are different for white and blue collar workers, for which the aforementioned education types are very good proxies. While EPL is very strict for white collars and relatively weak for blue collars, the minimum wage is binding for the latter group and not for the former. Consequently, while a recession at graduation is found to have a persistent negative effect on earnings for both groups, the high educated experience a persistent negative effect on the hourly wage and not on the annual hours worked, and the reverse holds for the low educated. These results are corroborated by a comprehensive set of sensitivity analyses. Concretely, a typical recession, which increases the unemployment rate by 1.4 *pp*, would induce the annual hours worked (and, hence, earnings) of the low educated to decrease by about 4.5% up to twelve years after graduation, while the hourly wage (and, hence, earnings) of the high educated would be roughly 6% lower than in case of a status quo ten years after labor market entry. The Great Recession of 2008 resulted in an increase of the unemployment rate of 1.6 *pp*, so that we predict that the negative consequences for the cohorts that graduated during the Great Recession will be about 15% ($\cong 0.2/1.4$) higher than the aforementioned ones.

Our results add to the evidence that labor market rigidity leads to much more persistent negative effects of recessions at labor market entry (Genda et al., 2010; Kawaguchi and Murao, 2014) and to fundamentally unjust (since it is induced by just *bad luck*) inequality between generations. At the same time our findings for low educated demonstrates that relaxing EPL alone is not sufficient to avoid persistence. Broader structural policy reforms seem to be required, which are much more difficult to implement.

From the point of view of the labor market, a *flexicurity* system should be advocated, where workers are mostly protected by unemployment insurance, with limited employment protection and active labor market policies to offer employment experiences as quickly as possible to unemployed. In Belgium, the existing unemployment benefit system is already an important safety net that protects workers when unemployed in case of low demand; its monitoring system has been shown to reduce the associated moral hazard problem and raise the job finding rate (Cockx and Dejemeppe, 2012a). A second issue is the employment protection: this may be desirable to a certain extent, since it promotes stable employment relationships where workers invest in their firm-specific skills, which ultimately may foster productivity growth. However, excessive employment protection may hamper productivity growth if it prevents workers reallocation, or new hiring due to high expected future firing costs in future downswings. This may result in overall higher level of unemployment and longer unemployment duration. For the period under investigation, Belgium was characterized by a flexible EPL for blue collar workers and a rigid EPL for white collar workers: this asymmetry has caused substantial penalties for these groups of workers. On the one hand, when graduating in downturns the high educated found it more difficult to occupy white collar positions, because insiders were very well protected and firms reluctant to hiring. Thus, they accepted lower-paying jobs for which employment protection was absent, incurring a high risk of remaining trapped in mismatched positions. On the other hand, the low educated were marginalized from the labor market, as a consequence of the additional competition provided by the high educated, and experienced serious difficulties to find stable jobs. Note that since the beginning of 2014 a single employment contract has been introduced in Belgium, stipulating the same EPL for white and blue collar workers. Therefore, this controversial discrimination between blue and white collars has been finally removed. Lastly, the very high level of Belgian minimum wages should be moderated. This may facilitate the absorption of low educated new graduates, for which minimum wages are binding.

Acknowledgements

We wish to thank Matteo Picchio, William Parienté, Dieter Verhaest, Bruno Vanderlinden, and the participants to the seminars at Ghent University (internal seminar), IZA (Bonn), Department of Economics of the Marche Polytechnic University in Ancona, Louvain-la-Neuve (Doctoral Workshops), as well as the participants to the $XXVIII^{th}$ National Conference of Labour Economics in Rome (27-28 September 2013) and to the Annual Meeting of the Association of Southern European Economic Theorists (ASSET), Bilbao (7-9 November 2013) for their valuable comments and discussions. We are grateful to the Cross Roads Bank for Social Security and the “Sternpost Studied- en Schoolchildren" for the availability of the data. Corinna Ghirelli acknowledges financial support from the Special Research Fund (“BOF”) of Ghent University, as well as from the National Bank of Belgium (project code 05Y00110). The usual disclaimer applies.

1.8 Appendix A. Descriptive Statistics

Table 1.1: Descriptive Statistics of Individual Controls in the Retained Sample.

			Low Educated			
Variable	Obs	Mean	Std. Dev.	Min	Max	Label
live in single-parent	1885	0.119	0.324	0	1	1 if live with single parent at age17(Dec)[§]
not live with parents	1885	0.064	0.244	0	1	1 if not live with either parents at age17(Dec)
HH members aged 0-11[†]	1885	0.247	0.625	0	7	nr of other HH members aged0-11 at age17(Dec)
HH members aged 12-17	1885	0.507	0.687	0	7	nr of other HH members aged12-17 at age17(Dec)
HH members aged 18-29	1885	0.520	0.729	0	8	nr of other HH members aged18-29 at age17(Dec)
HH members aged 30-64	1885	1.890	0.400	0	5	nr of other HH members aged30-64 at age17(Dec)
HH members aged 65+	1885	0.037	0.207	0	2	nr of other HH members aged65+ at age17(Dec)
father education	1885	4.582	3.196	0	13	father completed education since age12[††]
mother education	1885	4.199	3.066	0	13	mother completed education since age12
years of delay in sec.edu[‡]	1885	0.825	0.843	-1	4	years of delay at age17(Aug)[§]
general edu[§§]	1885	0.108	0.310	0	1	1 if general edu at age17(Aug)
technical edu	1885	0.379	0.485	0	1	1 if technical edu at age17(Aug)
vocational edu	1885	0.412	0.492	0	1	1 if vocational edu at age17(Aug)
apprenticeship/PT voc	1885	0.101	0.302	0	1	1 if apprent./PT vocational at age17(Aug)
birth cohort76[§§]	1885	0.333	0.471	0	1	1 if born in 1976
birth cohort78	1885	0.333	0.471	0	1	1 if born in 1978
birth cohort80	1885	0.334	0.472	0	1	1 if born in 1980
			High Educated			
Variable	Obs	Mean	Std. Dev.	Min	Max	Label
live in single-parent	1629	0.079	0.270	0	1	1 if live with single parent at age17(Dec)
not live with parents	1629	0.027	0.162	0	1	1 if not live with either parents at age17(Dec)
HH members aged 0-11	1629	0.154	0.424	0	3	nr of other HH members aged0-11 at age17(Dec)
HH members aged 12-17	1629	0.578	0.687	0	4	nr of other HH members aged12-17 at age17(Dec)
HH members aged 18-29	1629	0.583	0.717	0	4	nr of other HH members aged18-29 at age17(Dec)
HH members aged 30-64	1629	1.926	0.299	1	4	nr of other HH members aged30-64 at age17(Dec)
HH members aged 65+	1629	0.029	0.185	0	2	nr of other HH members aged65+ at age17(Dec)
father education	1629	7.031	3.239	0	13	father completed education since age12
mother education	1629	6.382	2.935	0	13	mother completed education since age12
years of delay in sec.edu	1629	0.260	0.540	-1	3	years of delay at age17(Aug)
general edu	1629	0.652	0.477	0	1	1 if general edu at age17(Aug)
technical edu	1629	0.339	0.473	0	1	1 if technical edu at age17(Aug)
vocational edu	1629	0.009	0.092	0	1	1 if vocational edu at age17(Aug)
apprenticeship/PT voc	1629	0.001	0.025	0	1	1 if apprent./PT vocational at age17(Aug)
birth cohort76	1629	0.330	0.470	0	1	1 if born in 1976
birth cohort78	1629	0.341	0.474	0	1	1 if born in 1978
birth cohort80	1629	0.330	0.470	0	1	1 if born in 1980

§ “at age17(Dec)" (“at age17(Aug)") means that the variable is measured in December (August) of the year when the individual turns age 17.

† HH refers to household.

†† It measures the number of years of education successfully attained since age 12.

‡ The variable measures the educational progression at age 17: a value of 0 means that the student is on time; -1 means that the student has skipped one academic year; positive values indicate the number of repeated grades.

§§ In the first step of the main analysis, general education and birth cohort76 are the reference categories.

Table 1.2: Descriptive Statistics Regarding the Selection Rules for the Benchmark Outcomes in the Provincial Model*.

Graduation period:	Low educated 1994-2001		High educated 1997-2004	
Outcomes[§]	Continuous	Discrete	Continuous	Discrete
Number of cells (total)	420	420	350	350
Number of dropped cells	42	138	27	109
Number of retained cells	378	282	323	241
Statistics on dropped cells				
Mean size dropped cells	10.29	30.97	13.11	27.95
Max size dropped cells	15	111	15	66
Statistics on retained cells				
Mean size retained cells	45.19	57.56	37.64	47.48
Max size retained cells	104	111	79	89
Statistics on retained cells for which $16 \leq N_{gpt} < 30$ *(continuous) or* $EXP^{\dagger}_{gpt} < 5$ *(discrete)*[‡]				
Number of retained cells	94	38	123	46
Avg size retained cells	23.06	45.08	23.08	34.07
Max size retained cells	29	111	29	66
P-value joint test for selectivity[§§]				
Specification used for log hourly wage	0.322		0.637	
Specification used for log hours worked	0.091		0.105	
Specification used for all discrete outcomes		0.379		0.207

$*$ To avoid too small cell sizes, we impose the following selection rules. For continuous variables, drop cells gpt with size $N_{gpt} < 16$. For discrete variables, drop cells gpt with the smallest $EXP^{\dagger}_{gpt}$ until at most 20% of the retained cells are such that $EXP_{gpt} < 5$ (Cochran, 1954).

† $EXP_{gpt} = \min\{N_{gpt}P_{gpt}, N_{gpt}(1 - P_{gpt})\}$, where P_{gpt} denotes the probability of success and which can be estimated by aggregating the individual predictions of this probability in the first step OLS regression of (1.3) to the cell level gpt. Notice that the aforementioned minimum is calculated for each outcome in the SUR and that the selection rule is applied on the basis of the smallest minimum across these outcomes.

§ The benchmark continuous outcomes in the SUR are log hourly wage and log hours worked in salaried employment, and the benchmark discrete variables are salaried employment and self-employment. The statistics in the table refer to one outcome, since they are identical for each outcome retained in a SUR.

‡ These are groups that would have been dropped according to the more stringent selection rules, i.e. $N_{gpt} < 30$ for the continuous outcomes and $EXP_{gpt} < 5$ for the discrete outcomes.

§§ The test for selectivity is based on a one-step regression of an indicator that is set to one if the individual belongs to a cell that is dropped according to the selection rules mentioned in $*$. Standard errors are clustered by group gp, which therefore tends to over-reject the null hypothesis that all the coefficients of the linear spline that interacts the unemployment rate at graduation are jointly significantly different from zero. We impose the same restrictions on the regression model (1.1) as we do for each benchmark outcome that we retain in the SUR (see Table 1.3). Consequently, since different sets of restrictions are imposed on the regression of log hourly wage than on that of log hours worked, we report two P-values for the continuous outcomes. For the discrete outcomes the same restrictions are imposed on both outcomes, so that only one P-value is reported.

1.9 Appendix B. Results for the Main Outcomes of Interest of Low Educated

Table 1.3: Effect of a One pp *Increase of the Unemployment Rate at Graduation on Main Outcomes: Low Educated.*

Outcomes	salaried	self-empl.	overall empl.	log wage	log hours§	log earnings	FT hours	PT hours	Total hours
Imposed Restrictions:									
Effect URate at grad. symm. up/downturn	yes	yes	-	yes	no	-	yes	yes	-
Effect Current URate over exp=0	yes	yes	-	no	no	-	no	no	-
Spline for calendar year FE	yes	yes	-	no	no	-	no	no	-
Effect of prov-time trends=0	yes	yes	-	no	no	-	no	no	-
Test joint signif. all imposed restr.(P-val)	0.286		-	0.155		-	0.268		-
P-value of chi2 test	0.341			0.000		-	0.000		-
WSSR (2nd step)	331		-	1289		-	1519		-
Obs (2nd step)†	375		-	756		-	754		-
Parameters (2nd step)	54		-	88		-	86		-
Level of clustering	no	no	no	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$
Potential experience	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	-0.017	0.036	0.018	0.005	-0.044	-0.039	-123.741	36.280	-87.462
	(0.018)	(0.018)	(0.015)	(0.013)	(0.022)	(0.026)	(33.386)	(12.419)	(26.467)
2	-0.011	0.021	0.010	-0.001	-0.035	-0.036	-82.608	15.700	-66.908
	(0.015)	(0.015)	(0.009)	(0.010)	(0.020)	(0.023)	(28.233)	(9.828)	(23.051)
3	-0.005	0.007	0.002	-0.006	-0.026	-0.033	-41.474	-4.879	-46.354
	(0.015)	(0.015)	(0.009)	(0.009)	(0.022)	(0.022)	(24.951)	(9.819)	(22.119)
4	-0.002	0.001	-0.001	-0.008	-0.026	-0.034	-30.805	-6.854	-37.659
	(0.015)	(0.014)	(0.007)	(0.008)	(0.019)	(0.019)	(22.293)	(8.855)	(19.123)
5	0.000	-0.005	-0.005	-0.009	-0.026	-0.035	-20.135	-8.829	-28.964
	(0.014)	(0.014)	(0.006)	(0.008)	(0.016)	(0.017)	(20.394)	(8.282)	(16.793)
6	0.002	-0.010	-0.008	-0.011	-0.025	-0.036	-9.466	-10.804	-20.270
	(0.015)	(0.014)	(0.008)	(0.007)	(0.014)	(0.015)	(19.475)	(8.183)	(15.434)
7	0.002	-0.009	-0.007	-0.008	-0.025	-0.033	-23.298	-2.250	-25.548
	(0.014)	(0.014)	(0.006)	(0.007)	(0.012)	(0.014)	(18.219)	(7.363)	(14.743)
8	0.002	-0.007	-0.005	-0.006	-0.025	-0.031	-37.129	6.304	-30.826
	(0.014)	(0.014)	(0.005)	(0.007)	(0.012)	(0.014)	(18.010)	(7.409)	(14.801)
9	0.002	-0.005	-0.004	-0.003	-0.025	-0.028	-50.961	14.858	-36.103
	(0.015)	(0.014)	(0.006)	(0.008)	(0.013)	(0.016)	(18.883)	(8.307)	(15.598)
10	0.003	-0.009	-0.006	-0.006	-0.027	-0.033	-44.904	15.268	-29.637
	(0.015)	(0.014)	(0.005)	(0.008)	(0.010)	(0.014)	(19.371)	(7.981)	(16.251)
11	0.004	-0.013	-0.009	-0.009	-0.030	-0.039	-38.848	15.678	-23.170
	(0.015)	(0.014)	(0.006)	(0.009)	(0.010)	(0.014)	(21.167)	(8.434)	(17.862)
12	0.006	-0.017	-0.011	-0.012	-0.033	-0.045	-32.791	16.088	-16.703
	(0.015)	(0.014)	(0.007)	(0.010)	(0.012)	(0.017)	(23.977)	(9.556)	(20.202)

Standard errors between parentheses. The table shows the effect for low educated of increasing the provincial unemployment rate at graduation by one *pp* on the mentioned outcomes. The reported estimates result from predictions based on the estimates of the linear spline in potential experience $f_g(e)$ that multiplies the provincial unemployment rate at graduation u_{gp} in (1.4). Since the outcomes satisfy adding-up constraints (salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked), the estimates are each time obtained from a two-step FGLS SUR on the first two outcomes in the sum described in detail in Section 1.5.2 and briefly below. Effects on the third outcome (the sum) are then obtained from the adding-up constraints. First step: (1.3) is estimated by OLS. Standard errors are clustered at the individual level. Second step: (1.4) is estimated by FGLS, where the inverse of the variance matrix of the $\hat{\mu}_{gpt}$ in the first step is used as weight. Depending on the outcome, we impose restrictions which cannot be jointly rejected at the 5% level: these restrictions are listed in the top panel of the table. If the $\chi 2$ goodness-of-fit statistic rejects the model (P-value>0.05), standard errors clustered at the gp level are reported; otherwise conventional ones.

§ For log hours worked the following additional restriction (not mentioned in the table) is also imposed: $\beta_{g2} = 0$, i.e. the slope of the linear spline remains fixed after 6 years of experience. This restriction cannot be rejected.

† The lower number of observations for the discrete outcomes in the second step reflects both that a different number of cells was dropped to ensure large enough cell sizes and the fact that the variance matrix of the $\hat{\mu}_{gpt}$ in the first step is singular so that we had to use the generalized inverse of the variance as weight in the second step (see Section 1.5.2 for more details).

Column 5 and 6 report the effects of interests in downturn: the effects in case of an upturn are different, but not reported. All other columns report the effect of interest independently of the business cycle, since this restriction is not rejected.

Table 1.4: Effect on Mobility and Firm Quality: Low educated.

Outcomes	geographical mobility§	firm mobility	median daily wage in the firm
Imposed Restrictions:			
Effect URate at grad. symmetric up/downturn	no	yes	yes
Effect Current URate over exp=0	yes	yes	yes
Spline for calendar year FE	yes	no	no
Spline for exp FE	yes	no	no
Effect prov-time trends=0	yes	yes	no
Test joint signif. all imposed restr.(P-val)	0.478	0.252	0.854
P-value of chi2 test	-	0.001	0.001
WSSR (2nd step)	-	413	422
Obs (2nd step)	-	360	378
Parameters (2nd step)	-	35	39
Test joint signif. URate_grad (P-val)	0.126	0.720	-
Level of clustering	i	$g * p$	$g * p$
Estimation approach	one-step	two-step	two-step
Potential experience	(1)	(2)	(3)
1	0.000	0.008	-0.014
	(0.003)	(0.019)	(0.011)
2	-0.001	0.005	-0.016
	(0.002)	(0.012)	(0.010)
3	-0.001	0.001	-0.018
	(0.003)	(0.013)	(0.009)
4	-0.001	-0.001	-0.016
	(0.002)	(0.010)	(0.008)
5	0.000	-0.003	-0.015
	(0.003)	(0.008)	(0.008)
6	0.000	-0.005	-0.014
	(0.003)	(0.009)	(0.008)
7	0.002	0.001	-0.015
	(0.003)	(0.008)	(0.008)
8	0.003	0.006	-0.015
	(0.003)	(0.009)	(0.007)
9	0.004	0.012	-0.016
	(0.004)	(0.011)	(0.007)
10	0.003	0.009	-0.017
	(0.004)	(0.010)	(0.007)
11	0.002	0.006	-0.017
	(0.004)	(0.010)	(0.008)
12	0.001	0.003	-0.018
	(0.005)	(0.011)	(0.008)

Standard errors between parentheses. The table shows the effect for low educated of increasing the provincial unemployment rate at graduation by one *pp* on the mentioned outcomes. The reported estimates result from predictions based on the estimates of the linear spline in potential experience $f_g(e)$ that multiplies the provincial unemployment rate at graduation u_{gp} in (1.4). Since the outcomes need not satisfy adding-up constraints, except for geographic mobility (see §) the estimates are obtained from a two-step FGLS on each outcome separately. The estimation procedure is described in detail in Section 1.5.2 and briefly below. First step: (1.3) is estimated by OLS. Standard errors are clustered at the individual level. Second step: (1.4) is estimated by FGLS, where the inverse of the variance matrix of the $\hat{\mu}_{gpt}$ in the first step is used as weight. Depending on the outcome, we impose restrictions which cannot be jointly rejected at the 5% level: these restrictions are listed in the top panel of the table. If the $\chi 2$ goodness-of-fit statistic rejects the model (P-value>0.05), standard errors clustered at the *gp* level are reported; otherwise conventional ones.

§ Notice that the effect on geographic mobility between provinces is estimated in one step, since virtually no cell satisfies the rule of Cochran (1954) discussed in Section 1.5.2. In the one-step approach both clustering at the individual and the $g * p$-level tends to over-reject. Since the P-value of the joint test of significance of the unemployment rate at graduation is higher at the individual than at the $g * p$ level we report the former.

Column 1 reports the effects of interests in downturn: the effects in case of an upturn are different, but not reported. All other columns report the effect of interest independently of the business cycle, since this restriction is not rejected.

1.10 Appendix C. Results for the Main Outcomes of Interest of High Educated

Table 1.5: Effect of a One pp *Increase of the Unemployment Rate at Graduation on Main Outcomes: High Educated.*

Outcomes	salaried	self-empl.	overall empl.	log wage	log hours	log earnings	FT hours	PT hours	Total hours
Imposed Restrictions:									
Effect URate at grad. symm. up/downturn	yes	yes	-	yes	no		no	yes	-
Effect Current URate over exp=0	yes	yes	-	yes	yes		no	yes	-
Spline for calendar year FE	yes	yes	-	no	no		no	no	-
Effect of prov-time trends=0	yes	yes	-	no	no		no	no	-
Test joint signif. all imposed restr.(P-val)	0.494		-	0.309		-	0.390		-
P-value of chi2 test	0.329		-	2.89E-34		-	4.48E-32		-
WSSR (2nd step)	310		-	1084		-	1059		-
Obs (2nd step)†	262		-	646		-	646		-
Parameters (2nd step)	48		-	75		-	78		-
Level of clustering	no	no	no	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$
Potential experience	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	-0.016	0.001	-0.015	-0.023	-0.035	-0.058	-81.986	23.301	-58.685
	(0.015)	(0.014)	(0.006)	(0.010)	(0.015)	(0.019)	(22.004)	(9.359)	(18.501)
2	-0.006	-0.007	-0.013	-0.028	-0.014	-0.041	-31.209	15.540	-15.669
	(0.013)	(0.013)	(0.004)	(0.009)	(0.011)	(0.015)	(17.068)	(6.611)	(14.479)
3	0.003	-0.014	-0.011	-0.032	0.008	-0.025	19.569	7.778	27.347
	(0.014)	(0.014)	(0.004)	(0.009)	(0.009)	(0.013)	(15.191)	(6.132)	(12.584)
4	0.004	-0.011	-0.007	-0.033	0.004	-0.029	6.612	10.017	16.629
	(0.013)	(0.013)	(0.003)	(0.008)	(0.009)	(0.012)	(12.390)	(5.697)	(9.889)
5	0.005	-0.009	-0.004	-0.033	0.001	-0.032	-6.345	12.256	5.911
	(0.013)	(0.013)	(0.002)	(0.008)	(0.009)	(0.013)	(11.468)	(5.445)	(9.492)
6	0.005	-0.006	0.000	-0.033	-0.003	-0.036	-19.302	14.495	-4.807
	(0.014)	(0.014)	(0.003)	(0.009)	(0.011)	(0.014)	(12.835)	(5.403)	(11.630)
7	0.002	-0.003	-0.001	-0.036	-0.002	-0.038	-11.662	14.625	2.963
	(0.013)	(0.013)	(0.002)	(0.009)	(0.009)	(0.012)	(11.589)	(5.111)	(9.769)
8	-0.001	0.000	-0.002	-0.039	-0.001	-0.040	-4.022	14.755	10.733
	(0.014)	(0.014)	(0.002)	(0.010)	(0.008)	(0.011)	(11.567)	(4.964)	(9.198)
9	-0.005	0.002	-0.002	-0.042	0.001	-0.041	3.618	14.885	18.504
	(0.014)	(0.014)	(0.003)	(0.010)	(0.009)	(0.012)	(12.775)	(4.975)	(10.137)
10	-0.008	0.005	-0.003	-0.044	0.002	-0.043	11.258	15.016	26.274
	(0.015)	(0.015)	(0.005)	(0.011)	(0.011)	(0.014)	(14.919)	(5.142)	(12.244)

Standard errors between parentheses. The table shows the effect for low educated of increasing the provincial unemployment rate at graduation by one *pp* on the mentioned outcomes. The reported estimates result from predictions based on the estimates of the linear spline in potential experience $f_g(e)$ that multiplies the provincial unemployment rate at graduation u_{gp} in (1.4). Since the outcomes satisfy adding-up constraints (salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked), the estimates are each time obtained from a two-step FGLS SUR on the first two outcomes in the sum described in detail in Section 1.5.2 and briefly below. Effects on the third outcome (the sum) are then obtained from the adding-up constraints. First step: (1.3) is estimated by OLS. Standard errors are clustered at the individual level. Second step: (1.4) is estimated by FGLS, where the inverse of the variance matrix of the $\hat{\mu}_{gpt}$ in the first step is used as weight. Depending on the outcome, we impose restrictions which cannot be jointly rejected at the 5% level: these restrictions are listed in the top panel of the table. If the $\chi 2$ goodness-of-fit statistic rejects the model (P-value$>$0.05), standard errors clustered at the gp level are reported; otherwise conventional ones.

† The lower number of observations for the discrete outcomes in the second step reflects both that a different number of cells was dropped to ensure large enough cell sizes and the fact that the variance matrix of the $\hat{\mu}_{gpt}$ in the first step is singular so that we had to use the generalized inverse of the variance as weight in the second step (see Section 1.5.2 for more details).

Columns 5-7 and 9 report the effects of interests in downturn: the effects in case of an upturn are different, but not reported. All other columns report the effect of interest independently of the business cycle, since this restriction is not rejected.

Table 1.6: Effect on Mobility and Firm Quality: High educated.

Outcomes	geographical mobility§	firm mobility	median daily wage in the firm
Imposed Restrictions:			
Effect URate at grad. symmetric up/downturn	yes	yes	no
Effect Current URate over exp=0	yes	yes	yes
Spline for calendar year FE	no	no	no
Spline for exp FE	yes	no	no
Effect prov-time trends=0	yes	yes	yes
Test joint signif. all imposed restr.(P-val)	0.273	0.471	0.236
P-value of chi2 test	-	0.964	0.028
WSSR (2nd step)	-	214	336
Obs (2nd step)	-	285	323
Parameters (2nd step)	-	32	35
Test joint signif. URate_grad (P-val)	0.495	0.165	-
Level of clustering	i	no	$g*p$
Estimation approach	one-step	two-step	two-step
potential exp	(1)	(2)	(3)
1	0.009	0.026	-0.037
	(0.006)	(0.013)	(0.007)
2	0.005	0.008	-0.038
	(0.004)	(0.009)	(0.006)
3	0.001	-0.009	-0.039
	(0.005)	(0.011)	(0.007)
4	0.001	-0.007	-0.035
	(0.004)	(0.008)	(0.006)
5	0.001	-0.006	-0.031
	(0.004)	(0.008)	(0.006)
6	0.001	-0.004	-0.027
	(0.004)	(0.010)	(0.007)
7	0.002	-0.005	-0.029
	(0.004)	(0.008)	(0.007)
8	0.003	-0.006	-0.031
	(0.004)	(0.009)	(0.006)
9	0.005	-0.007	-0.033
	(0.004)	(0.011)	(0.007)
10	0.006	-0.008	-0.035
	(0.005)	(0.015)	(0.007)

Standard errors between parentheses. The table shows the effect for low educated of increasing the provincial unemployment rate at graduation by one *pp* on the mentioned outcomes. The reported estimates result from predictions based on the estimates of the linear spline in potential experience $f_g(e)$ that multiplies the provincial unemployment rate at graduation u_{gp} in (1.4). Since the outcomes need not satisfy adding-up constraints, except for geographic mobility (see §) the estimates are obtained from a two-step FGLS on each outcome separately. The estimation procedure is described in detail in Section 1.5.2 and briefly below. First step: (1.3) is estimated by OLS. Standard errors are clustered at the individual level. Second step: (1.4) is estimated by FGLS, where the inverse of the variance matrix of the $\hat{\mu}_{gpt}$ in the first step is used as weight. Depending on the outcome, we impose restrictions which cannot be jointly rejected at the 5% level: these restrictions are listed in the top panel of the table. If the $\chi 2$ goodness-of-fit statistic rejects the model (P-value>0.05), standard errors clustered at the gp level are reported; otherwise conventional ones.

§ Notice that the effect on geographic mobility between provinces is estimated in one step, since virtually no cell satisfies the rule of Cochran (1954) discussed in Section 1.5.2. In the one-step approach both clustering at the individual and the $g*p$-level tends to over-reject. Since the P-value of the joint test of significance of the unemployment rate at graduation is higher at the individual than at the $g*p$ level we report the former.

Column 3 reports the effects of interests in downturn: the effects in case of an upturn are different, but not reported. All other columns report the effect of interest independently of the business cycle, since this restriction is not rejected.

1.11 Appendix D. Additional Sensitivity: the Dot-com recession

The dot-com recession occurred in Belgium between March 2001 and November 2001 according to the official definition. In Flanders this caused the unemployment rate to rise by 1.7 *pp* in 2002-2003 (see Figure 1.1); the provincial unemployment rates remained at quite a high level also in 2004 and did not fall before 2005 (see Figure 1.2). Thus, this recession emerged during the graduation period considered for the high educated, i.e. the 1997-2004 period. It is therefore interesting to see to what extent the main results for the high educated are driven by the dot-com recession, or whether this recession implied different penalties for new graduates compared to those inflicted by one *pp* increase in the unemployment rate for the larger period 1997-2004. To do this, we augment equation (1.1) with a dummy equal to one if graduation occurs in 2002-2004 (i.e. the jobless recovery following the dot-com recession) and with an interaction between this dummy and $f_g(e)u_{gp}$.[43] This interaction identifies the differential long-run effects of increasing the provincial unemployment rate at graduation by one *pp* when graduation occurs in the 2002-2004 period versus the 1997-2001 period.

To facilitate the comparison with the main outcomes, Figure 1.5 plots the scarring effect of graduating in the recession period - the effect of one *pp* increase in the unemployment rate at graduation for the period 2002-2004 - against the benchmark, i.e. the effect of one *pp* increase in the unemployment rate at graduation for the entire period 1997-2004, shown in Figure 1.4 in the main text. To avoid clutter, confidence intervals are plotted only for the benchmark effects. Moreover, whenever the benchmark included asymmetric effects between graduating in downturns or upturns, we plot only the former and not the latter.

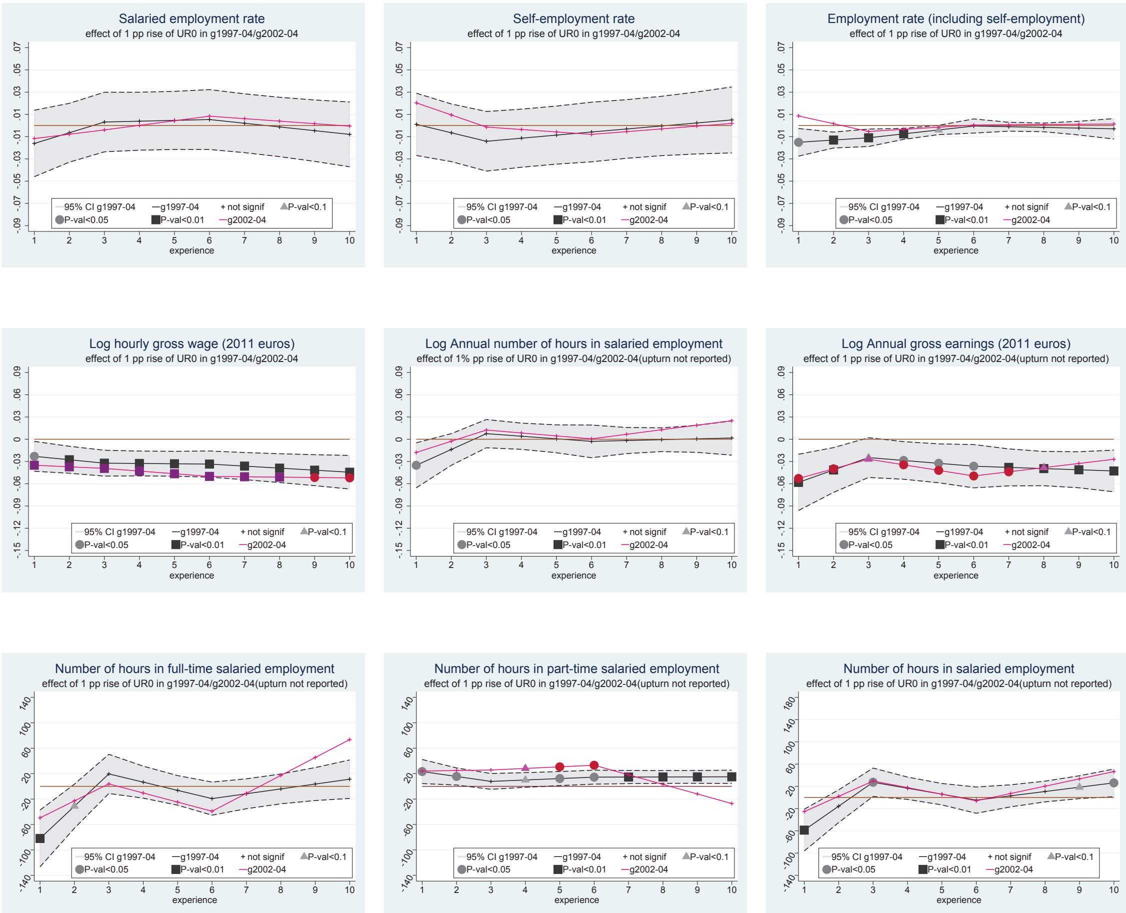
In the top panel, the effects on the employment indicators are not significant, but suggest that the "push-effect" into self-employment may be more important for graduating in the dot-com recession than in the benchmark. Importantly, the temporary negative effect on overall employment is significant in the benchmark, but it is absent in the recession period. In the middle panel, the scarring effect on wages is similar but more pronounced for the 2002-2004 graduation period than in the overall period 1997-2004. At the same time, the initial drop in (full-time) hours worked is significant in the benchmark but disappears for the cohorts graduating in the recession. The net effect on earnings is very similar in the two cases, but it suggests that graduating in the dot-com recession inflicts a slightly more severe but less persistent scar (until potential experience 8), compared to the benchmark. However, this may be also due to the loss of power resulting from the additional parameters included in the sensitivity analysis. Overall, these results are consistent with the idea that the high educated graduating in the dot-com recession preferred more often to downgrade and enter the labor market in a mismatched job rather than waiting for better offers - compared to the overall graduation period 1997-2004. In addition, the bottom panel shows that those graduating in the recession work slightly less hours in full-time salaried employment at

[43] In the main analysis, the graduation year linear spline $f_0(g)$ is allowed to change slope each three years, i.e. in 1997, 2000 and 2003; thus the period 2002-2004 is split into two splines. As a robustness check, we redefined the $f_0(g)$ to isolate the period 2002-2004 in the last spline: i.e. let the spline change slope in 1997, 2000 and 2002 and impose linear trends on sub-periods 1997-1999, 2000-2001, and 2002-2004. This does not affect the results.

potential experience 3-6 years, and work more hours in the subsequent period, compared to the benchmark (but this effects are not significant). This seems to be compensated by an increase (decrease) in hours worked in part-time employment in (after) the first 6 years of potential experience compared to the benchmark.

To sum up, this sensitivity analysis demonstrates that the results for the high educated are broadly the same when focusing on the dot-com recession. Note that by allowing for a specific scarring effect for the graduation cohorts 2002-2004, the confidence intervals widen: this indicates that we are loosing power in adding the interactions to isolate the penalties from graduating in the dot-com recession. Therefore, the analysis would gain from more compact specifications as those estimated in Figure 1.4 in the main text. However, the above comparison suggests that, if anything, the high educated graduating in the period 2002-2004 decreased their reservation wages more rapidly than usual, and consequently preferred more often to downgrade to lower-paying jobs rather than waiting for a better offer with no job.

Figure 1.5: Effect of one pp *Increase in the Provincial Unemployment Rate at Graduation for the High Educated: Comparison between the Benchmark (graduation period 1997-2004) and Graduating in the Dot-Com Recession (2002-2004).*



The figure compares for the high educated the effect of increasing the provincial unemployment rate at graduation by one pp *on the main outcomes of interest (a) when graduation occurs in the entire period 1997-2004 (i.e. the benchmark) versus (b) in the dot-com recession, i.e. in the period 2002-2004. The scarring effect in (a) correspond to ones shown in Figure 1.4. The scarring effects in (b) are computed as follows: for each outcome we estimate the benchmark specification - used to obtain Figure 1.4 - and allowed for a differential effect for graduating in the period 2002-2004. Then, in this figure, we plot the scarring effects from graduating in the period 2002-2004 (b) against the benchmark models (a). The confidence intervals are reported only for the benchmark models (a). Whenever the benchmark model allowed for asymmetric effects of graduating in downturns/upturns, the former but not the latter are reported. The effects reported in top, middle and bottom panel result each time from two-step FGLS estimates of a SUR model on the first two outcomes described in detail in Section 1.5.2. The effects for the third outcome are obtained from the following adding up constraints: salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked.*

Bibliography

Amemiya, T. and F. Nold (1975). A Modified Logit Model. *The Review of Economics and Statistics 57*(2), 255–257.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Arulampalam, W. (2001). Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages. *The Economic Journal 111*(475), 585–606.

Baert, S., B. Cockx, and D. Verhaest (2013). Overeducation at the Start of the Career: Stepping Stone or Trap? *Labour Economics 25*, 123–140.

Baker, G., M. Gibbs, and B. Holmstrom (1994). The Wage Policy of a Firm. *The Quarterly Journal of Economics 109*(4), 921–955.

Barten, A. P. (1969). Maximum Likelihood Estimation of a Complete System of Demand Equations. *European Economic Review 1*(1), 7–73.

Beaudry, P. and J. DiNardo (1991). The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data. *Journal of Political Economy 99*(4), 665–688.

Begg, D., A. Lindbeck, C. Martin, and D. Snower (1989). Symmetric and asymmetric persistence of labor market shocks. *Journal of the Japanese and International Economies 3*(4), 554–577.

Berndt, E. R. and N. E. Savin (1975). Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances. *Econometrica 43*(5/6), 937–958.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Biewen, M. and S. Steffes (2010). Unemployment persistence: Is there evidence for stigma effects? *Economics Letters 106*(3), 188–190.

Brewer, M., T. F. Crossley, and R. Joyce (2013). Inference with Difference-in-Differences Revisited. Working Paper 7742, IZA, Bonn.

Brunner, B. and A. Kuhn (2014). The impact of labor market entry conditions on initial job assignment and wages. *Journal of Population Economics 27*(3), 705–738.

Cahuc, P. and S. Carcillo (2014). Alléger le coût du travail pour augmenter l'emploi : les clés de la réussite. mimeo, Institut Montaigne.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics 90*(3), 414–427.

Chevalier, A. (2003). Measuring Over–education. *Economica 70*(279), 509–531.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics 10*(4), 417–451.

Cockx, B. (2013a). Youth Unemployment in Belgium: Diagnosis and Key Remedies. IZA Policy Papers 66, IZA, Bonn.

Cockx, B. (2013b). Youth Unemployment in Belgium: Diagnosis and Key Remedies. *Intereconomics 48*(4), 202–209.

Cockx, B. and M. Dejemeppe (2005). Duration Dependence in the Exit Rate out of Unemployment in Belgium: Is It True or Spurious? *Journal of Applied Econometrics 20*(1), 1–23.

Cockx, B. and M. Dejemeppe (2012). Monitoring job search effort: An evaluation based on a regression discontinuity design. *Labour Economics 19*(5), 729–737.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Chapter 2.7, pp. 101–109. Baltimore, Maryland: Johns Hopkins University Press.

Devereux, P. J. and R. A. Hart (2007). The Spot Market Matters: Evidence on Implicit Contracts from Britain. *Scottish Journal of Political Economy 54*(5), 661–683.

Dolton, P. J. and M. A. Silles (2008). The effects of over-education on earnings in the graduate labour market. *Economics of Education Review 27*(2), 125–139.

European Commission (2010). *Self-Employment in Europe 2010*. European Commission, Directorate-General for Employment, Social Affairs and Equal Opportunities, Unit D.2.

Fairlie, R. W. (2013). Entrepreneurship, Economic Conditions, and the Great Recession. *Journal of Economics & Management Strategy 22*(2), 207–231.

Fuss, C. (2009). What Is the Most Flexible Component of Wage Bill Adjustment? Evidence from Belgium. *Labour Economics 16*(3), 320–329.

Gaini, M., A. Leduc, and A. Vicard (2012). A scarred generation? french evidence on young people entering into a tough labour market. Working Paper G 2012 / 5, Institut National de la Statistique et des Etudes Economiques (INSEE), DESE.

Genda, Y., A. Kondo, and S. Ohta (2010). Long–Term Effects of a Recession at Labor Market Entry in Japan and the United States. *Journal of Human Resources 45*(1), 157–196.

Gibbons, R. and M. Waldman (2004). Task–Specific Human Capital. *The American Economic Review 94*(2), 203–207.

Gibbons, R. and M. Waldman (2006). Enriching a Theory of Wage and Promotion Dynamics inside Firms. *Journal of Labor Economics 24*(1), 59–107.

Gregg, P. (2001, November). The Impact of Youth Unemployment on Adult Unemployment in the NCDS. *The Economic Journal 111*(475), 626–563.

Gregg, P. and E. Tominey (2005). The wage scar from male youth unemployment. *Labour Economics 12*(4), 487–509.

Hagedorn, M. and I. Manovskii (2013). Job Selection and Wages over the Business Cycle. *American Economic Review 103*(2), 771–803.

Hansen, C. B. (2007). Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects. *Journal of Econometrics 140*(2), 670–694.

Hanushek, E. A., L. Woessmann, and L. Zhang (2011). General Education, Vocational Education, and Labor–Market Outcomes over the Life–Cycle. IZA Discussion Papers 6083, Institute for the Study of Labor (IZA).

Heckman, J. (1974). Shadow Prices, Market Wages, and Labor Supply. *Journal of Econometrics 42*(4), 679–694.

Hijzen, A. and D. Venn (2011). The Role of Short-Time Work Schemes during the 2008–09 Recession. OECD Social, Employment and Migration Working Papers No. 115, OECD Publishing.

Høj, J. (2013). Enhancing the Inclusiveness of the Labour Market in Belgium. OECD Economics Department Working Papers No. 1009, OECD Publishing.

Kahn, L. B. (2010). The Long–Term Labor Market Consequences of Graduating from College in a Bad Economy. *Labour Economics 17*(2), 303–316.

Kawaguchi, D. and T. Murao (2014). Labor Market Institutions and Long–Term Effects of Youth Unemployment. Discussion Paper 8156, IZA, Bonn.

Kilponen, J. and T. Santavirta (2010). New Evidence on Implicit Contracts from Linked Employer–Employee Data. *Scandinavian Journal of Economics 112*(4), 864–883.

Kliesen, K. L. (2003). The 2001 recession: how was it different and what developments may have caused it? *Federal Reserve Bank of St. Louis Review 85*(5), 23–38.

Kondo, A. (2007). Differential Effects of Graduating during a Recession across Race and Gender. Working Paper 07-05, ISERP, University of Columbia.

Kramarz, F. and T. Philippon (2001). The Impact of Differential Payroll Tax Subsidies on Minimum Wage Employment. *Journal of Public Economics 82*(1), 115–146.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013). Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment. *The Quarterly Journal of Economics 128*(3), 1123–1167.

Kwon, I. and E. Meyersson Milgrom (2007). Cohort Effects in Wages and Promotions. Discussion Paper 07-025, Stanford Institute for Economic Policy Research (SIEPR).

Lindbeck, A. and D. J. Snower (2001). Insiders versus outsiders. *Journal of Economic Perspectives 15*(1), 165–188.

Liu, K., K. G. Salvanes, and E. Sørensen (2012). Good Skills in Bad Times: Cyclical Skill Mismatch and the Long–Term Effects of Graduating in a Recession. Discussion Paper 6820, IZA, Bonn.

Mansour, H. (2009). The Career Effects of Graduating from College in a Bad Economy: The Role of WorkersÂŠ Ability. Mimeo, University of California, Santa Barbara.

Mroz, T. A. and T. H. Savage (2006). The Long–Term Effects of Youth Unemployment. *Journal of Human Resources 41*(2), 259–293.

Neumark, D. and W. L. Wascher (2008). *Minimum Wages*. Cambridge, MA: MIT Press.

OECD (2007). Activating the Unemployed: What Do Countries Do. In *Employment Outlook*, Chapter 5, pp. 207–242. Paris: OECD.

OECD (2012). *Equity and Quality in Education. Supporting Disadvantaged Pupils and Schools*. Paris: OECD.

OECD (2013). *Employment Outlook*. Paris: OECD.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012). The Short– and Long–Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Oreopoulos, P., T. v. Wachter, and A. Heisz (2008). The Short– and Long–Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates. Discussion Paper 3578, IZA, Bonn.

Pérez, E. and Y. Yao (2012). Can Institutional Reform Reduce Job Destruction and Unemployment Duration? Yes it Can. IMF Working Papers 12/54, International Monetary Fund.

Raaum, O. and K. Røed (2006). Do Business Cycle Conditions at the Time of Labor Market Entry Affect Future Employment Prospects? *The Review of Economics and Statistics 88*(2), 193–210.

Rycx, F. and S. Kampelmann (2013). Who Earns Minimum Wages in Europe? New Evidence Based on Household Surveys. Working Paper 13-01, Dulbea, Universite Libre de Bruxelles (ULB).

Schmillen, A. and M. Umkehrer (2013). The Scars of Youth: Effects of Early–Career Unemployment on Future Unemployment Experience. Discussion Paper 201306, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg].

Service public fédéral Economie, P.M.E., C. m. e. E. (2007). La navette entre Régions et provinces. Communique de Presse 19 juillet 2007, Direction General Statistique et Information Economique.

SONAR (2003). Hoe maken de jongeren de overgang van school naar werk? Basisrapportering Cohorte 1978 (eerste golf). Technical report, Leuven: Steunpunt WAV.

SONAR (2004a). Hoe maken de jongeren de overgang van school naar werk? Basisrapportering Cohorte 1976 (tweede golf). Technical report, Leuven: Steunpunt WAV.

SONAR (2004b). Hoe maken de jongeren de overgang van school naar werk? Cohorte 1980 (eerste golf). Technisch rapport. Technical report, Leuven: Steunpunt WAV.

Stevens, K. (2008). Adverse Economic Conditions at Labour Market Entry: Permanent Scars or Rapid Catch-up? Mimeo, University of Sydney.

Taylor, M. and T. Rampino (2014). Educational aspirations and attitudes over the business cycle. *Economica 81*(324), 649–673.

Topel, R. H. and M. P. Ward (1992). Job Mobility and the Careers of Young Men. *The Quarterly Journal of Economics 107*(2), 439–79.

Tumino, A. and M. Taylor (2013a). The impact of local labour market conditions on school leaving decisions. Mimeo, ISER, University of Essex.

Tumino, A. and M. Taylor (2013b). The labour market impacts of leaving education when unemployment is high: evidence from Britain. ISER Working Paper Series 12, ISER, University of Essex.

Wooldridge, J. M. (2006). Cluster-Sample Methods in Applied Econometrics: An Extended Analysis. Mimeo, Department of Economics, Michigan State University.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section And Panel Data (second ed.)*. The MIT Press, Cambridge, MA.

Yu, L., P. F. Orazem, and R. W. Jolly (2014). Entrepreneurship over the business cycle. *Economics Letters 122*(2), 105–110.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association 57*(298), 348–368.

2
The scarring effect of early non-employment

2.1 Introduction

High levels of youth unemployment are a great concern for policy makers, especially since the start of the Great Recession (Bell and Blanchflower, 2010). Historically, however, unemployment rates have been higher for young workers than for older workers. This is understandable, since younger workers have the least experience and hence are often the easiest to remove. Moreover, they lose or leave a job more often than older workers, because job shopping helps them to find a good match (Yedid-Levi et al., 2014). For instance, Topel and Ward (1992) find that two third of job changes and wage growth occur in the first ten years of workers' career. This initial high turnover may involve also short spells in unemployment. Therefore, youth unemployment need not to be necessarily detrimental to workers' career, if this is part of the process to find stable employment.[1] Other views predict that the experience of youth unemployment may entail long-term penalties in terms of reduced wages and persistent unemployment. These outcomes are explained by human capital loss (Pissarides, 1992), which may arise from the depreciation of existing capital - if the worker is not subject to any kind of training when not employed - as well as from forgone work experience (Ellwood, 1982). An alternative explanation comes from the signaling model, in which past unemployment records are interpreted by employers as signals of low productivity in a context

[1] According to the literature on job displacement, the costs of job loss for young workers are smaller and less persistent than for mature workers (e.g. Kletzer and Fairlie, 2003; Topel, 1990). This is because young workers, unlike mature workers, experience steep earnings growth in subsequent work experience, after a job loss. At the same time, earnings growth of displaced young workers is below the levels of young non-displaced workers: i.e. most of the earnings loss is due to foregone wage growth due to displacement.

of imperfect information (Lockwood, 1991).[2]

A broad evidence suggests that the consequences of experiencing youth unemployment are not just temporary. A number of studies provide evidence of persistence in youth unemployment: for UK, Gregg (2001) finds strong state dependence for new graduates experiencing early unemployment. Schmillen and Umkehrer (2013a) show similar results for students graduating from the German apprenticeship program. Based on a sample of young long-term unemployed in Belgium, Cockx and Picchio (2013) find that past unemployment duration negatively affects the probability to be employed. Other studies show that youth unemployment entails long-term penalties on earnings or wages: Gartell (2009) and Gregg and Tominey (2005) find persistent negative effects on earnings and wages for Sweden and UK, respectively. Results for the US show short-lived penalties of youth unemployment on subsequent employment, but long-run penalties on wages (e.g. Ellwood, 1982; Mroz and Savage, 2006). Moreover, Mroz and Savage (2006) find strong evidence of human capital "catch-up" response to youth unemployment. By investing in training when unemployed, youth mitigate the human capital losses induced by unemployment and, consequently, the earnings losses, which however persist 9 years after unemployment. This evidence is important from a policy point of view, as it provides grounds to advocate policies that aim to reduce the scars of youth unemployment: for instance, integrating youth in the labor market be means of wage subsidy programs; or, providing training schemes to young unemployed.

The goal of this study is to contribute to the existing empirical evidence on the long-term consequences of early labor market performances for Flanders, the most prosperous of three Belgian regions. Our analysis is based on very rich data combining survey data with administrative data. Therefore, we are able to evaluate the impact of early labor market outcomes on a range of labor market outcomes available for the first twelve years after graduation: hours worked and earnings for salaried public and private sector employment, salaried and self-employment. This gives us a comprehensive view of the long-term consequences of youth's early labor market outcomes.

The main identification problem is the presence of unobserved factors that may affect early as well as subsequent labor market performances, thereby introducing endogeneity. We address this problem by means of an Instrumental Variable (IV) approach, where the provincial unemployment rate at graduation is used as instrument for early labor market outcomes. An advantage of this methodology is that it exploits the *exogenous* variation of the instrument to disentangle the causality between early and subsequent labor market performances from the spurious correlation induced by unobserved individual characteristics.

Throughout the analysis we distinguish between low and high educated new graduates because in Belgium labor market institutions differ for blue and white collar workers, and this creates different sources of rigidities for the low and the high educated.[3] For instance, white collars are sheltered by a very strict employment protection legislation (EPL), which represents the main

[2] Kroft et al. (2013b) send fictitious resumes to real job postings in US and find that the likelihood of receiving callbacks for interviews significantly decreases with the length of the unemployment spells, mostly in the first 8 months. This is evidence that employer screening plays a primary role in generating duration dependence.

[3] In the data there is a clear correspondence between low (high) educated and blue (white) collar workers: see Table 2.4 in Appendix 2.7.

rigidity for high skilled workers. In contrast, blue collar workers have quite loose EPL,[4] but are supported by a short-time work compensation program (STC) that subsidies temporary reduction of labor force during downturns. This introduces rigidity in the labor market of blue-collar workers since it allows employers to tie their employees to them and has therefore similar consequences as EPL for white collar workers. Moreover, (sectoral) minimum wages - among the highest of OECD - are likely to be binding for low educated youth. Together with a quite generous unemployment insurance (UI) system,[5] minimum wages are therefore more relevant sources of rigidity for the low educated. Based on the same data of this study, Chapter 1 investigates the scars of graduating in downturns for Flanders and find that the low and the high educated undergo different hysteresis mechanisms, which are explained by the distinct labour market institutions designed for white and blue collars. While a downturn at graduation is found to inflict a persistent adverse effect on earnings for both low and high educated workers, the latter are penalized in terms of hourly wages and not of annual hours worked (because of strict EPL for white-collars and not binding minimum wages); the reverse holds for the low educated.

Hence, as a consequence of unfavorable labor market conditions at graduation, the high educated downgrade to lower-paying jobs whereas low educated workers remain unemployed more often. This suggests that, in Flanders, the long-term effects of graduating in downturns occur through the loss of early work experience for the latter and through the acceptance of lower-paying jobs for former. In particular, for the low educated the scar may persist since unemployment spells convey bad signals to the employers, or because of human capital depreciation which makes the unemployed less attractive to employers or, possibly, because of processes of discouragement or habituation (Clark, 2003; Clark et al., 2001). The high educated instead may find it difficult to upgrade to a position matching their initial aspirations as a consequence of the (useless) accumulation of human capital - which is specific to lower-paying position - and due to the high barriers to enter white-collar positions.

In this study we test the former hypothesis: that is, whether the low educated workers suffer from long-term penalties from experiencing early non-employment. We identify causality by means of an IV approach where the provincial unemployment rate at graduation is used as instrument for early non-employment. Unfortunately, we could not test the second aforementioned hypothesis - i.e. if, for the high educated, the average level of hourly wages earned in the first years after graduation has long-run repercussions on their subsequent career - as the instrument has not sufficient explanatory power to implement the IV approach. The interested reader can find this exercise in Appendix 2.14. Since we consider individuals graduating in the period 1994-2002 for the low educated in the 5 Flemish provinces, inference hinges at most on 45 clusters. This raises the possibility of underestimating standard errors due to few clusters. We tackle this problem

[4] Note that since the beginning of 2014 a single employment contract has been introduced in Belgium, stipulating the same EPL for white and blue collar workers.

[5] UI covers also unemployed school-graduates without employment records with no time limits and loose job search requirements. Until 2012, school graduates below age 26 were entitled to UI 9 months after registration at the Public Employment Service. Since 2012 the waiting period has been extended from 9 to 12 months.

applying wild bootstrap [6] to the IV approach, following Davidson and MacKinnon (2010).

The paper is organized as follows. In the next section we briefly review the literature. Section 2.3 describes the data. In Section 2.4 we explain the estimation strategy: we discuss the IV approach, including the way in which we deal with the problem of inference with few clusters. Section 2.5 discusses the results for the low educated and presents some sensitivity analysis. Section 3.7 concludes.

2.2 Literature Review

To our knowledge, two studies have investigated the long-term impacts of unemployment for Belgium. Gangji and Plasman (2007) study the adverse effect of the incidence of unemployment on re-entry wages considering a representative sample of Belgian workers aged 18-64 in the 1994-2002 period. They address the problem of endogeneity due to unobserved individual characteristics with a fixed effect (FE) estimation, thereby exploiting only the within-individual variation of the data.[7] They find that the incidence of unemployment is associated with a 5.1% penalty in hourly wages. Of course, this is an average of heterogenous effects comprising workers with different ages and skills. In Belgium, as argued in the Introduction, low and high skilled workers may undergo different wage penalties due to the different labor market institutions. Moreover, the incidence of unemployment for youth is likely to have different consequences than for mature workers.

The second study by Cockx and Picchio (2013) is focused on the long-run effects of youth long-term unemployment. It considers all Belgian school-graduates aged 18-25 who in 1998 were registered to the National Employment Office and remained unemployed for at least 9 months,[8] and follow them for each quarter until the end of 2002. They use a mixed proportional hazard model with competing risks where omitted heterogeneity is tackled as follows. The unobserved individual characteristics that are uncorrelated with the regressors are integrated out of the model. As for the unobserved characteristics that are correlated with the regressors, they impose a functional form restriction on the relationship between the former and the latter, thereby allowing for the dependence.[9] The causality of lagged endogenous unemployment duration, the parameter of interest, is identified by the exclusion restrictions generated by the time variation of the exogenous variables and by having multiple realizations per individual of the outcome variable. They find evidence of strong negative duration dependence in the job finding probability: further prolonging the unemployment spell by one year reduces the probability to find a job in the following 2 years from 60% to 16% for men. Of course these results apply to the specific sub-sample of long-term

[6] Wild bootstrap preserves the group structure of the data (Cameron et al., 2008).

[7] The sample is restricted to workers whose wage is observed at least twice in the period. Selectivity is addressed by Heckman's two-step estimator, with the number of children and house property as exclusion restrictions.

[8] In Belgium, before 2012 school-graduates were eligible to UI benefits if still unemployed 9 months after the registration to the National Employment Service.

[9] As a consequence, regressors' coefficients cannot be given a structural interpretation as cannot be separated out from the unobservables.

unemployed youth.

We contribute to this literature by providing new evidence on the long-term effects of early non-employment for low educated youth by means of IV approach, where the identification of the causality between early non-employment and subsequent labor market outcomes comes from the exogenous variation of the provincial unemployment rate at graduation. Note that our IV method shares some similarities with the aforementioned approaches - in the way they deal with the presence of omitted individual factors correlated with the variable of interest. The FE approach reaches consistency exploiting only the time-variation of the data, i.e. the variation within- and not across-individuals. Hence, the identification comes from the part of the variation that is uncorrelated with the (fixed) unobservables. The IV identification strategy is similar, as it exploits the (exogenous) variation of the instrument that is correlated with the endogenous regressor. Similarly, in mixed proportional hazard models, the time variation of the exogenous time-varying variables generate exclusion restrictions that are used for the identification. In sum, the difference between the methods implemented in the existing literature and our approach boils down to the type of *exogenous* variation that is exploited to identify causality.

For other countries, some studies have already applied IV to identify the long-term effects of youth unemployment exploiting the variation in the local unemployment rate before the first entry in the labor market.[10] Based on UK data, Gregg (2001) and Gregg and Tominey (2005) study the effect of youth unemployment on subsequent unemployment spells and wages, respectively: in both cases, the local unemployment rate at age 16 (the end of compulsory education) is used as instrument for early unemployment. The latter is defined as the cumulative months of unemployment or non-employment in the period of age 16-âĂŞ23, while the dependent variables are measured at age 33 and 42. The sample is restricted to students born in 1958 graduating up to age 21. The results show strong adverse effects of early unemployment on later wages and employment. For Germany, Schmillen and Umkehrer (2013a) focus on a sample of students graduating from the apprenticeship program between 1978 and 1980 and investigate the relationship between the cumulative days spent in unemployment in the first 8 years after graduation and the total days spent in unemployment in the subsequent 16 years. The regressor of interest is instrumented by the local unemployment rate at graduation. They find significant and long-lasting scarring effects, especially at the right tail of the unemployment distribution.

In a nutshell, the estimation strategy shared by these studies is based on the idea that the variation in the labor market conditions at school leaving is exogenous to the individual, and therefore generates a variation in the individual early unemployment that is unrelated to unobserved factors that may influence both early and adult performances. In particular, their identification relies on the exclusion restriction that the labor market conditions at graduation affect the dependent variables uniquely through early unemployment. Therefore, the underlying identifying assumption is that the scarring effect of graduating in downturns occurs entirely through the reduced early work

[10] A related study is Neumark (2002), who identifies the causality of early job stability on adult wages with IV, where indicators of job stability in the first 5 years since graduation are instrumented by labor market conditions faced in this early period.

experience. For Flemish low educated, Chapter 1 of this dissertation provides evidence of this pattern. However, the aforementioned exclusion restriction also rules out the possibility that this scar arises through the acceptance of lower-paying wages. This is precisely how the scar takes place for Flemish high educated (see Chapter 1). The reliability of this exclusion restriction is not discussed in this literature. In contrast, in this study we adapt the estimation strategy in such a way that the exclusion restriction is most likely valid in light of the aforementioned evidence: this means using for the low educated a measure of early non-employment as endogenous variable.[11]

2.3 Data

The analysis is based on the Sonar survey database, a representative sample of three birth-cohorts of Flemish youth - born in 1976, 1978 or 1980, which were interviewed at age 23, 26 and 29.[12] The surveys register retrospectively and on monthly basis the most important activity of the respondents, among which education. Based on this information, graduation is identified to occur in the first month that education has been interrupted for at least for 4 months. The surveys also contain control variables for the analysis, which are measured at age 17, such that they are predetermined at graduation: father's and mother's education (years of completed education since age 12), the type of educational program (general, technical, vocational, part-time vocational or apprenticeship) in which the individual is enrolled at age 17, and the number of repeated grades at age 17 since secondary education. From this database we calculate for each individual the number of completed years of education, i.e. the number of grades successfully passed from the start of secondary education until graduation. Based on the latter we divide the sample in low and high educated, with the former having completed at most secondary education and the latter having a higher level of education.[13] We make this distinction throughout the analysis, because, as mentioned in the Introduction, the minimum wage is more likely to be binding for low educated youth at the start of the career and because EPL-strictness varies significantly between white and blue collar workers.

The original Sonar sample contains about 9000 individuals, 3000 for each birth cohort. We restrict it as follows, to increase the homogeneity of the sample. We exclude few observations (0.17% of the sample) who dropped out from schooling before the end of compulsory education, which in Belgium is set at age 18. We focus on men since female labor supply is different from male labor supply due to mothering.[14] We drop individuals who attended special needs and arts education, who were not Belgian or did not speak Flemish at home, or who did not reside in Flanders at graduation. We retained individuals graduating from age 18 and 24, as students graduating with more than 24 years old are less than 5% of the sample. After eliminating, in addition, individ-

[11]For the high educated we use a measure of early wage as endogenous regressor. Since the instrument is weak in the IV approach, results are relegated to Appendix 2.14. Hereafter we will focus on the low educated.

[12]For more details, see SONAR (2003, 2004a,b).

[13]Low educated are those with at most 6 years of completed education (7 years if enrolled in vocational track at age 17). High educated are those with higher years of completed education.

[14]Long-term effects of early labor market outcomes for women are equally interesting and left for future research.

uals with missing or inconsistent values in variables, we are left with a final sample of 3586 low and high educated men. From this, we focus on 1902 low educated, who graduated in the period 1994-2002. Descriptive statistics of the final sample is in Appendix 2.7.

The survey data are matched to administrative data of Belgian Social Insurance institutions centralized at the Cross Roads Bank of Social Security, which give us access to high quality information on individuals' labor market outcomes for a sufficiently long time span after graduation. In particular, these data report quarterly information on the registration as self-employed, as well as earnings and time worked in dependent employment (for both public and private sector), between year 1998 and 2010. For salaried workers we construct log annual earnings and log annual hours worked. The log-transformation allows us to interpret the coefficients of interest as semi-elasticities. Note that we retain in the analysis also non-salaried employed by adding value one to the continuous variables before taking the logarithmic transformation. Hence, non-salaried employed have zero log-earnings and zero log-hours worked. As a consequence our estimates on continuous outcomes are unconditional on being salaried employed. This rules out the problem of selectivity due to restricting to the salaried employed, i.e. a potentially positively selected group. Yet it introduces another complication since the distribution of the outcome variables has a mass point at zero. The fraction of corner solutions is however quite small, as only 15% of the low educated are censored at zero at the moment of the evaluation. As a consequence, we will rely on OLS.[15] In addition, we construct three employment indicators: salaried employment, defined by positive earnings from salaried employment; self-employment, based on the registration as self-employed for at least one day during the calendar year; overall employment, which is the sum of self- and salaried employment. Note that salaried employed who are also registered as self-employed in the same calendar year are considered self-employed. These outcomes are measured 6 or 8 years after graduation. This choice is due to the availability of the administrative data (1998-2010) and by the fact that we want to measure the dependent variables as late as possible for all graduation cohorts. Since the low educated graduate in the period 1994-2002, the last graduation cohort is followed until potential experience 8. Later than that the sample gets smaller as the last graduation cohorts progressively drop out from the sample. At the same time, we want to get an idea of the persistence of the scar: hence we choose to evaluate the outcomes also two years earlier, i.e. at potential experience 6. Table 2.7 in Appendix 2.7 shows descriptive statistics of the outcome variables.

The administrative data also provide additional control variables measured at age 17: living in single parent household, not living together with either parents and the number of other household members by age class. Descriptive statistics of the control variables are reported in Table 2.6 of Appendix 2.7.[16] Finally, the administrative data give us access to yearly information on the province of residence between the year in which individuals turn age 18 and 2010.

From the year of graduation onwards, we associate each calendar year to a *potential* year

[15]OLS estimation provides approximations of the unconditional effects, as it does not take into account the corner solutions at zero. In principle, Tobit models would be more appropriate.

[16]For details on the construction of the control and outcome variables, see Section A.1 of this dissertation.

of labor market experience,[17] which corresponds to zero in the year of graduation. Potential experience 0 lasts from the month subsequent to graduation until December of that calendar year. Therefore, its length (measured in months) is computed as $12 - month\_of\_graduation$, i.e. it depends on the month of graduation: for a June graduate - which amounts to 90% of the sample - it lasts 6 months. All subsequent years of potential experience have a duration of 12 months. Our regressor of interest is a measure of the time spent in non-employment at potential experience 0-2, relative to the potential total hours if one would work full-time during the whole period. We express it as a proportion in order to take into account the fact that the reference period changes depending on the month of graduation, thereby ensuring that early non-employment is comparable across individuals. On average this period corresponds to 2.5 years after graduation (30 months), as 90% of the sample graduates in June. For simplicity hereafter we will refer to this reference period as to its average, i.e. 2.5 years after graduation.

This endogenous regressor can be measured precisely, by exploiting the administrative data on hours worked in salaried employment. However, the latter are available only since 1998, while students graduate since 1994 in our sample. Therefore, we base this variable on the administrative data for the individuals graduating since 1998 (68.5% of the sample), and exploit information from Sonar (survey data) whenever potential experience 0-2 occurs before 1998 (31.5% of sample). The reason why we combine administrative with survey data is to maximize the sample of study, thereby exploiting the variation of the instrument for the entire graduation period 1994-2002, rather than for the restricted period 1998-2002. Of course, the disadvantage is that the data on time worked in the Sonar database are less precise and hence we have to make some assumptions to convert this information into hours worked:[18] this certainly introduces measurement error in the endogenous regressor.

Briefly, the endogenous regressor is constructed as follows (for details, see Appendix 2.8): first, define the reference period as the entire calendar year for potential experience 1 and 2, and the part of calendar year following the month of graduation for potential experience 0; sum up all hours worked including self-employment in the reference period (a);[19] compute the potential total hours if one would work full-time during the whole period (b); express early non-employment as $100 * (b - a)/b$. As already mentioned, the denominator takes into account the fact that the reference period changes depending on the month of graduation and ensures comparability across individuals. This variable measures the intensity of early non-employment. It is equal to zero if in the reference period one has always worked as much as a full-time salaried employed, and above zero if one has worked less intensively than the full-time regime or if one has not worked for some time. Given the possibility of measurement error arising from the combination of survey and administrative data, Section 2.5.1 performs a sensitivity analysis for the restricted graduation period 1998-2002, where only administrative data are exploited to measure the endogenous regressor.

[17] This terminology is borrowed from the literature. "Potential" underlines that the variable counts all calendar years since graduation, as opposed to *actual* experience which endogenously considers only years in employment.

[18] These are imputed from monthly employment indicators assuming full-time employment (see Appendix 2.8).

[19] For self-employed we assume the working regime of full-time salaried employed (see Appendix 2.8).

A final source of information is the Labor Force Survey (LFS), which provides long time series of the provincial unemployment rates for Flanders. In our analysis, we use the provincial unemployment rate (15-64) at graduation as instrument for early non-employment. Figure 2.2 in Appendix A.1.5 plots this series from year 1993 until 2011. Note that the literature typically exploits more disaggregated unemployment rate series.[20] For Belgium, provincial unemployment rates are the most disaggregated data available for the period considered. The main drawback is that the inference relies on too few clusters, as the identification of the effects of interest comes from the variation of the unemployment rates by provinces and years. To the extent that we tackle this problem with wild bootstrap (see Section 2.4.4), provincial data are not much of a limitation. In contrast, more aggregated series provide the advantage of reducing the problem of endogenous migration, that would arise if new graduates offset the long-term effects of early non-employment by moving or commuting into provinces where there are more job opportunities. Our data suggest that in Flanders less than 2% of individuals change province of residence in the 1998-2010 period. However, as Flanders is a relatively small region, people could commute to work across provinces. In this case, we would underestimate long-term effects of early non-employment.[21] However, the magnitude of the inter-provincial variation in the unemployment rate reported in Figure 2.2 demonstrates that mobility and commuting are limited and far from eliminates all inter-provincial variation.[22]

2.4 Estimation Strategy

We are interested in the causal relationship of early non-employment, say Y^0, on subsequent labor market outcomes of interest Y for the low educated. Namely, we want to estimate an equation of the following type, where X is a vector of control variables that will be defined below and ϵ is an idiosyncratic error term:

$$Y = aY^0 + bX + e \text{ with } e = (\theta + \epsilon) \tag{2.1}$$

The main identification problem is the presence of some factors θ, unobserved to the researcher, that may affect both early non-employment and subsequent labor market performances, thereby introducing endogeneity. Therefore, OLS estimates will be biased as a consequence of these omitted factors. We remove this bias by means of a two-stage least squares (2SLS) estimator, where the provincial unemployment rate at graduation is used as instrument (Z) for early non-employment.[23] In practice, the identification strategy relies on the variation of the provincial

[20] At district level (Schmillen and Umkehrer, 2013a) and by wards (Gregg, 2001; Gregg and Tominey, 2005).

[21] If anything, this should be more worrying for the high educated, since they are (1) less liquidity constrained because of high expected wages or better working conditions and (2) more mobile due to higher motivation to find jobs that meet their expectations about wages/job profiles.

[22] This is because LFS series are based on the province of residence and not of job location. Hence if workers commute to avoid the adverse local labor market conditions, this evens out the provincial variation in the unemployment rate.

[23] A similar approach has been used by Neumark (2002), Schmillen and Umkehrer (2013a), Gregg (2001) and Gregg and Tominey (2005).

unemployment rate at graduation Z, which is exploited to generate an *exogenous* variation in early non-employment Y^0, which is then used to identify the causal relation of interest. In accordance with the traditional IV approach, we assume that the effect of interest is homogeneous.[24] In this framework, the 2SLS identifies the causal effect of interest under two conditions:

1. Z is uncorrelated with e. This implies that Z does not *directly* affect the outcome Y (exogeneity), and that any *indirect* effect of Z on Y occurs uniquely through the endogenous regressor Y^0 (exclusion restriction). This is an identifying assumption.

2. Z is correlated with Y^0, conditional on the controls X in (2.1) (strength). This condition can be tested by means of the F statistic of the excluded instrument in the first stage regression.

Note that in this framework the IV estimate refers to the entire population since the causal effect of interest is assumed to be homogeneous across individuals. Next section discusses in detail the identifying assumption 1. In particular, we will carefully examine which factors may violate the exclusion restriction and define the specification in such a way that the latter is most likely satisfied conditional on the covariates.

2.4.1 The Instrumental Variable Approach: identifying assumptions

Together, Condition 1 and 2 require that the instrument explains the endogenous regressor, but that at the same time it is exogenous in model (2.1). This has the following implications.

First, it amounts to rule out reverse causality between Z and Y^0, that is Z affects Y^0 but not the other way around. If it were the case, Z would be endogenous in (2.1) because of the correlation with Y^0, and as consequence it should be included as additional regressor in the specification. We exclude the possibility of reverse causality since the instrument and the endogenous regressor are measured at the provincial and individual level respectively, and an aggregate variable cannot be caused by an individual variable.

Second, the exogeneity assumption requires that the unemployment rate Z cannot affect the unobserved composition of new graduates by year and province. If this were the case, the relation between the instrument and early non-employment would spuriously reflect changes in the composition of graduates rather than causality, which would introduce selectivity. To rule this out, one needs to assume that students choose the moment of graduation independently of the business cycle (exogeneity of timing of graduation), and that before graduation they do not move to provinces where the unemployment rate is lower relatively to others (exogeneity of place). We test the former condition in Section A.5 of this dissertation, and demonstrate that the duration between the end of compulsory education at the age of 18 and each year of potential graduation is unrelated

[24]This is of course a restrictive assumption. In section 2.5.2 we discuss the interpretation of IV under heterogeneous effects, where the IV estimator identifies a weighted average of local average treatment effects (LATEs) under additional assumptions: Stable Unit Treatment Value Assumption (SUTVA) and Monotonicity.

to the provincial unemployment rate in those years.[25] As for mobility, almost nobody (0.44%) change residence between the first year that our data inform about the place of living, i.e. on December 31 of the year in which the individual turns 17, and the year of graduation. Therefore, the issue can be safely ignored. On this basis, we argue that in our sample the choice of graduation is independent from the labor market conditions. However, we cannot rule out endogeneity due to commuting, for instance if students enrolled into universities located in provinces where they expected to find more jobs in the future.[26]

Third, the exclusion restriction requires that that the instrument is not correlated with any of the omitted factors in model (2.1). If this holds, one can assume that the scars of graduating in downturns for the low educated are determined *exclusively* by the proportion of time spent in early non-employment. Accordingly, the persistence of these penalties should be rationalized entirely by the loss of human capital originated by early non-employment, or by the fact that the latter is perceived as a signal of bad quality by employers (see Introduction). This assumption is consistent with the evidence of Chapter 1, who found that adverse labor market conditions at graduation inflict to the low educated big initial penalties on earnings and hours worked, which fade away slowly. The absence of a similar impact on wages is due to the presence of minimum wages, which are likely to be binding for the low educated at the start of the career. Instead, the scars on earnings and hours worked are explained by the labor market rigidities that prevents workers reallocation: the short-term work compensation program that in hard times ties the employers to the employees, and the EPL (flexible for blue collars while rigid for white collars) which pushes the high educated to downgrade and hence increases the competition for low skilled positions. As a consequence, the low educated who graduate in downturns are rather likely to experience longer periods of non-unemployment at the start of the career. This has repercussions in the long-term, according to the aforementioned evidence.[27] Based on the latter, we argue that early non-employment is the relevant channel to explain long-term effects of adverse labor market conditions at graduation for the low educated.

Of course things can be a bit more blurry if we consider a wider definition of reservation wage which incorporates also the future wage growth linked with seniority in addition to the current wage. In this case, low educated graduating in a downturn may not only experience higher early non-employment, but could also accept lower-quality jobs, i.e. with less steep wage profile than the jobs accessed during a tight labor market. The unemployment rate at graduation would entail a growing negative impact on subsequent wages as a consequence of accepting this initial job, and this would represent a violation of the exclusion restriction when wages are the outcome of

[25] In a discrete duration model, an indicator of graduating since age 17 is regressed on birth cohort dummies, individual characteristics and the province of living measured at age 17, the elapsed duration in education since age 17, and the unemployment rate in each potential year of graduation (interacted with the elapsed duration), testing whether the coefficients of latter interactions are jointly significantly different from zero. The test deals with selectivity induced by unobserved heterogeneity. It uses the same sample as this study.

[26] We do not have information on the location of universities in which students graduate nor of subsequent jobs.

[27] Note that this scar may be nuanced by the extensive use of STC for blue collars: they will experience long periods of unemployment or reduced activity, but they are more likely to be called back.

interest. This possibility is discussed in Chapter 1, where the unemployment rate at graduation has a negative impact on wages starting from potential experience 6.[28] This evidence is compatible with the aforementioned hypothesis and hence represents a violation of the exclusion restriction when wages are evaluated. Thus, we restrict the analysis on hours worked and earnings, since the long-term penalties on the these outcomes are compatible with the idea that early non-employment is the main driver of the scars.

Moreover, other channels may as well contribute to explain the long-term penalties of labor market conditions at graduation: these channels would invalidate the exclusion restriction if not included in the specification. An example is the persistence of the unemployment rate series. If the current unemployment rate affects the outcomes, the correlation between the unemployment rate at graduation and the current unemployment rate violates the assumption that the instrument affects the outcomes only through early non-employment. To prevent that, it is important to additionally control for the current unemployment rate, as typically done in the literature. However, this may not be sufficient, as in principle one should control for all unemployment rate series up to the moment of evaluation (Oreopoulos et al., 2012, 2008). To keep a parsimonious specification, we add the average unemployment rate between the end of the early period and the moment of evaluation - between potential experience 3 and 6.[29]

More generally, the problem of the persistence of the unemployment rates refers to the literature on wage determination, which investigates how the sequence of labor market conditions experienced by a worker affects current wages (Beaudry and DiNardo, 1991). According to this view, labor markets operate as spot markets if current wages are affected by current unemployment rates and not by past ones. In contrast, wages result from long-term implicit contracts if past unemployment rates explain current wages despite current ones: with costless mobility, the minimum unemployment rate since hiring should matter the most, as workers are able to renegotiate the wage once better labor market conditions arise; if instead mobility is costly, the unemployment rate at hiring should be the relevant one. Beaudry and DiNardo (1991) found that, once the minimum unemployment rate since hiring is included together with the unemployment rate at hiring, the former but not the latter significantly explains current wages. This is consistent with the idea that wages are negotiated according to long-term implicit contracts with renewals.[30] In this case, the exclusion restriction may be violated if the unemployment rate at graduation mistakenly picks up the effect of the minimum unemployment rate since hiring, because of the persistence of the unemployment rates. To prevent that, we include the minimum unemployment rate since graduation in the specification.[31]

[28] In the benchmark model this growing negative impact is not statistically significant, but it is significant in the sensitivity analysis (see Table A.22 in Section A.9 of this dissertation).

[29] To rule out multicollineariy, we run a sensitivity analysis only including current unemployment rate (see Table 2.12 in Appendix 2.12). The stability of the results ensures that multicollinearity is not driving the results.

[30] Recently Hagedorn and Manovskii (2013) criticize this interpretation. They argue that wages are still determined by spot markets and not by long-term implicit contracts. They show that, once the current match quality is taken into account, past labor market conditions no longer play a role in the wage determination.

[31] Note that the hypothesis of long-term implicit contracts seems more likely for the high educated, for instance to

Finally, other violations of the exclusion restriction may be due to, for instance, differences in institutions that could be correlated both with the unemployment rate at graduation and with the outcomes. We therefore include province fixed effects to ensure that permanent differences across provinces violate the exclusion restriction. Similarly, we include province-specific time trend to capture whatever time-varying provincial heterogeneity, such as changes in legislations, that may be correlated with the instrument and the labor market outcomes at the moment of evaluation. Next section presents the equation of interest in light of all these arguments.

2.4.2 The Equation of Interest

To avoid clutter, we state the following definitions: t is the observation period, which runs from graduation until the moment of evaluation of the outcomes of interest T, i.e. 6 or 8 years after graduation; t_0 is the time of measurement of predetermined individual controls, which corresponds to the year in which individuals are aged 17; t_1 is the time window in which we measure early non-employment, i.e. on average the first 2.5 years after graduation[32]. We estimate the following equation, where subscript i indicates the individual, g the graduation year and p the province of residence at graduation:

$$y_{igpT} = \alpha + \beta y^0_{igpt_1} + \gamma_1 UR_{pT} + \gamma_2 \overline{UR}_p + x'_{it_0}\delta + \zeta minUR_{pt} + \eta_p + \omega_p T + f(g) + e_{igpT}$$
$$\text{with } e_{igpT} = \theta_i + \epsilon_{ipgT} \tag{2.2}$$

- y_{igpT} represents the following outcomes of interest measured in T, i.e. 6 or 8 years after graduation: three indicators of salaried, self- and overall employment, as well as log hours worked and log earnings in salaried employment. Before taking the logarithm of continuous variables we add value one, so that non-salaried employed at the moment of evaluations are included with value of zero after the logarithmic transformation. Therefore, the effects on continuous outcomes are unconditional on being salaried employed. The reason why do this is twofold: first, the instrument is not strong enough to estimate effects *conditional* on salaried employment, but it becomes relevant when non-salaried employed are also included in the sample.[33] Second, unconditional effects refer to the entire population of workers and avoid the problem of selectivity when focusing on the sub-population of salaried employed. We take the log of continuous outcomes to interpret the estimates as semi-elasticities.

- $y^0_{igpt_1}$ is the endogenous regressor representing early non-employment: it is expressed as the percentage of time spent in non-employment in period t_1, relative to potential total hours if one would work full-time during the whole period.

- UR_{pT} is the current unemployment rate in the province of graduation, i.e. measured at the moment of evaluation T. It ensures that the exclusion restriction is not violated by the

capture returns in human capital accumulation. In contrast, recent evidence has shown that labor markets operate like spot markets for the low educated (Devereux and Hart, 2007; Kilponen and Santavirta, 2010).

[32] This time window corresponds to potential experience 0-2, i.e. from the month after graduation until December of the second subsequent calendar year.

[33] In the first stage, the F statistic is about 4 in the conditional case and reaches 9 in the unconditional one.

correlation between current local labor market conditions and local labor market conditions at graduation.

- $\overline{UR}_p$ is the time average of the unemployment rate in the period subsequent to the measurement of early non-employment, i.e. from potential experience 3 to 6. Together with UR_{pT} , it controls for the persistence of the unemployment rate series.

- $minUR_{pt}$ is the minimum unemployment rate in the province of residence of graduation over the entire period t. It controls for the possibility that wages are determined by long-term contracts which are renegotiated by the workers during upturns. Under this assumption, the persistence of the unemployment rate series (i.e. the correlation between $minUR_{pt}$ and the instrument) and the correlation between the outcomes and the minimum unemployment rate could violate the exclusion restriction.

- x_i is a set of individual control variables, predetermined since measured in t_0: birth cohort dummies, family composition, parental education, repeated years since secondary education as well as the educational track at age 17.

- η_p is fixed effects for the province of living at graduation: it controls for time-fixed provincial heterogeneity, that is all differences across provinces that are constant over time: e.g. differences in institutions, or in the structure of the economy.

- $\omega_p T$ is the provincial specific linear time trends, which are included because the unemployment rates exhibit differential downward time trends (see Figure 2.2 in Appendix A.1.5). More generally, it controls for any time-varying provincial heterogeneity, for instance for changes in legislations or in the structure of the economy at the provincial level.

- $f(g)$ is a linear spline in the graduation year, which controls for aggregate shocks affecting all provinces over the graduation period. We impose a piece-wise linear specification because graduation year fixed effects absorb too much variation and as a consequence the instrument becomes weaker in the first stage. The spline is formulated as $f(g) = \alpha + \sum_{j=0}^{2} \beta_j.(g - 3j)\mathbf{1}[g \geq 3j]$ with $g = 1, .., 9$.

- ϵ_{igpT} is an i.i.d. error term, while θ_i represents unobserved individual factors correlated with $y^0_{it_1}$, thereby introducing endogeneity.

β is the coefficient of interest which represents the effect of one percentage point increase in the proportion of time spent in early non-employment on subsequent outcomes of interest (employment rates, hours worked and earnings) for the low educated: in presence of scarring we expect a negative β. The OLS estimate of β is biased due to the correlation between θ_i and $y^0_{igpt_1}$.

For all dependent variables, we estimate (2.2) by OLS or 2SLS. Thus, we estimate linear probability models for discrete labor market outcomes. For continuous variables we report unconditional OLS effects and hence do not take into account that these outcomes are left-censored at zero. However, we believe that this is a minor issue, since the fraction of corner solutions in the sample is quite small: only 15% of the low educated are not salaried employed at potential experience 6.[34] In any case, we provide heteroscedastic-robust standard errors in all estimations to account for the fact that the dependent variables are dichotomous or censored at zero.

2.4.3 The Bias and Its Direction

The aforementioned bias can go in both directions. To see this, consider a simplified version of (2.2): $y = \beta y^0 + \theta w + \epsilon$, where y^0 is individual early non-employment, y is an individual subsequent outcome of interest (i.e. employment rates, hours worked, and earnings) w is a fixed individual characteristic. By assumption $E(y^0\epsilon) = 0$ and $E(w\epsilon) = 0$, and we expect a negative β. Let $Cov(y^0, w) \neq 0$. If w is observed, the OLS give unbiased estimators of β and θ: $\beta_{OLS} = \frac{Cov(y^0,y)}{Var(y^0)}$ and $\theta_{OLS} = \frac{Cov(w,y)}{Var(w)}$. If instead w is omitted, the OLS estimator is biased since $\beta_{OLS} = \frac{E(y^0,y)}{E(y^0)^2} = \frac{E[y^0(\beta y^0+\theta w+\epsilon)]}{E(y^0)^2} = \beta + \theta\frac{Cov(y^0,w)}{Var(y^0)}$. The direction of the bias depends on the sign of θ - the relationship between early non-employment and the outcome of interest - and the correlation between the former and the omitted factor. Below we discuss four possible sources of bias and their corresponding sign.

- Ability and Motivation: Everything else equal, more able and motivated individuals are more likely to perform well in the labor market at any point in time: therefore these factors are negatively correlated with early non-employment and positively correlated with the outcomes of interest. The overall bias is negative, so that OLS overestimate the (negative) scarring effect of early non-employment.

- Returns to job search: heterogeneous returns may arise because of differences in the search intensity or in the methods of search chosen. Ceteris paribus, individuals with higher returns search more and more successfully (also on-the-job), and therefore perform better in the labor market. Hence, the outcomes of interest are positively correlated with returns to search. At the same time, in the first phase of job shopping they may alternate jobs with short spells in non-employment, if they find it optimal to consume leisure when young and their opportunity cost is lower.[35] This may generate a positive correlation between returns to job search and early non-employment (Neumark, 2002). Under these assumptions, the bias is positive and OLS underestimate scarring.

- Liquidity constraints: everything else equal, individuals with high liquidity constraints have low reservation wages because they need to earn a salary. Thus, we expect liquidity constraints to be negatively correlated with early non-employment. At the same time, these individuals are likely to accept low quality jobs because of their low reservation wage, which is likely to translate into worse labor market performances over time. We therefore expect

[34] In Schmillen and Umkehrer (2013a) the dependent variable, the number of days spent in unemployment in prime age, is censored at zero for almost 60% of cases. They use Tobit models.

[35] Neumark (2002) justify this assumption as follows: in a standard life cycle utility-maximization model individuals are more likely to consume leisure at the point in the life cycle when their wages are low.

also a negative correlation with the outcomes of interest. The resulting bias is positive so that OLS underestimate scarring.

- Measurement error: as explained in Section 2.3, we introduce measurement error in the construction of early non-employment, because we use information from the Sonar database to impute hours worked in the first 2.5 years since graduation for students graduating before 1998, which are not observed in the administrative data. Measurement error in the endogenous regressor reduces OLS estimates towards zero (Hausman, 2001), thereby underestimating the scarring effect of early non-employment.[36]

To recapitulate, we expect OLS to overestimate the negative effect of early non-employment on the outcomes of interest, if the bias comes from ability. In contrast, the OLS estimate will be overestimated if the bias is due to returns to job search, measurement errors in the endogenous regressor or liquidity constraints. The literature is in favour of the latter hypothesis (Gregg, 2001; Gregg and Tominey, 2005; Neumark, 2002; Schmillen and Umkehrer, 2013a).

2.4.4 Inference

It is well known that standard errors are underestimated in a micro-level regression with grouped covariates because it is assumed that each observation is independent of all others while the information of the grouped covariates is repeated within each cluster. Therefore, correct inference requires to take into account that the independent information of the grouped covariates is at the group level, which can be done with cluster-robust standard errors (Angrist and Pischke, 2008; Moulton, 1990, ch.8). In our 2SLS we use a grouped variable, the unemployment rate at graduation, as instrument for the endogenous regressor, early non-employment, which varies at the individual level. Therefore, the identification of causality comes from the variation by province and time of the provincial unemployment rates at graduation, which is exploited to construct the fitted values of the first stage.[37]

The clustered estimator is consistent provided that the number of clusters is large enough, as consistency is determined by the law of large numbers. This is because, given the grouped structure of the data, the relevant unit are clusters and not observations. Since we consider the low educated graduating in the 5 Flemish provinces in the period 1994-2002, inference hinges on 44 clusters.[38] This raises the possibility of underestimating standard errors due to few clusters. Empiricists tend to agree that 50 clusters is enough when clusters have roughly the same size, but that a higher number of clusters is required when clusters are unbalanced (Cameron et al., 2008; MacKinnon and Webb, 2014). Applying the clustered estimator when clusters are too few

[36] In linear models, OLS estimates of y^0 are underestimated due to measurement errors and the bias can be eliminated with IV. However, this does not holds for non-linear models (Amemiya, 1985; Hausman et al., 1995).

[37] In 2SLS, the bias of the conventional variance estimator with grouped data is determined by the intra-class correlation of the second stage residuals (ρ_e) and by the intra-class correlation of the first stage fitted values (ρ_x). ρ_x is highest with grouped regressors in the first stage. As for OLS, $\rho_e > 0$ does not matter for standard errors as long as ρ_x is zero, but also a small ρ_e can give important bias with $\rho_x > 0$ (Angrist and Pischke, 2008, ch.8).

[38] Clusters are 44 since $g2002p3$ is empty. Table 2.5 in Appendix 2.7 shows the distribution across clusters.

is likely to worsen the bias, as cluster robust standard errors may be even smaller than conventional ones. This is what we find by comparing conventional and cluster robust standard errors of 2SLS estimations (see Table 2.1), which suggests that we have too few clusters.

We tackle this problem with wild restricted efficient residual bootstrap (WRE bootstrap) proposed by Davidson and MacKinnon (2010), which are designed for 2SLS in context of heteroscedasticity or clustered data. This procedure combines the restricted efficient residual bootstrap (RE bootstrap) proposed by Davidson and MacKinnon (2008) for the 2SLS, with wild bootstrap, which allows for intra-cluster correlation and heteroscedasticity (Cameron et al., 2008). For completeness, we apply wild bootstrap also to the t statistic of the instrument in the first stage, as well as to the t statistic of the regressor of interest when estimating (2.2) by OLS.[39] The bootstrap procedures are explained in detail in Appendix 2.10.

Because of few clusters, also the F statistic of the first stage is overestimated. To adjust it, we exploit the fact that in case of one instrument the F statistic is the square of the t statistic of the instrument in the first stage: i.e. with G clusters, $F(1, G-1) = t^2(G-1)$. Therefore, the bootstrap F statistic is the critical value of the $F(1, G-1)$ distribution that corresponds to the bootstrap P-value of the t statistic of the instrument in the first stage.[40]

2.5 Results

Table 2.1 summarizes the results for the low educated from estimating (2.2) by OLS and 2SLS on alternative labor market outcomes, measured 6 years after graduation. As a matter of space, we report only the effects of interest, i.e. the effect of early non-employment β in the structural equation and the impact of the instrument in the first stage regression. The complete regressions are reported in Appendix 2.11. Odds and even columns show heteroscedastic-robust and cluster robust standard errors, respectively. The former takes into account that the dependent variables are dichotomous or censored at zero, while the latter allows for intra-cluster correlation induced by the fact that the instrument varies at the gp level. The fact that the 2SLS cluster robust standard errors are smaller than the 2SLS heteroscedastic-robust ones (columns 3 and 4 in Panel A) suggests that clustering is ineffective because of too few clusters. We ensure to make correct inference by bootstrapping the t statistic of the effects of interest and by reporting the corresponding P-value.

Panel B summarizes the results of the first stage. We report the original F statistic (10.51) as well as the bootstrap one (9.25), which accounts for the problem of few clusters. As expected, the former is overestimated. According to the Stock-Yogo critical values, the IV estimator of β

[39] In the first stage the instrument is grouped, hence we need to cluster. In contrast, when estimating (2.2) by OLS the regressor of interest varies at the individual level: hence, clustering is not a major issue. For completeness we provide both heteroscedastic-robust and cluster-robust standard errors.

[40] We are aware of only one study by Baltagi et al. (2013) on the performance of wild bootstrap applied to the F statistic in context of heteroscedastic - but not clustered - data. Bootstrapping directly the F test in our wild bootstrap procedure did not always yield the expected results (sometimes, the bootstrap P-value of the F statistic was smaller than the P-value of the original F). For this reason, we relied on the bootstrap P-value of the t statistic of the instrument in the first stage.

over-rejects the null, as it leads to a rejection rate close to 15% when the true rejection rate is 5% (Stock and Yogo, 2005).[41] Hence, because of this test size, the IV estimates should to be taken with caution. The reported estimate suggests that one percentage point (pp) increase in the unemployment rate at graduation increases early non-employment by 5 pp. For Flanders, the unemployment rate rises on average by 1.4 pp in the 1994-2010 period (and by 1.6 pp in the Great Recession in 2008). Thus, graduating in an average downturn increases the proportion of hours spent in non-employment early in the career by about 7% (1.4×5).

The upper part of Panel A refers to the employment indicators. The sign of the estimates suggests that early non-employment has a positive (negative) impact on the probability to be self-(salaried) employed, but the size of the effect is very small. In contrast to OLS, 2SLS are not significant: this may be a consequence of too small power of the test, because of the limited variation of the instrument. The null hypothesis of the exogeneity test is largely not rejected for all indicators, indicating that both estimators are consistent but the OLS is more efficient than the 2SLS one.[42] We therefore focus on OLS: for one pp increase in early non-employment, the probability to be salaried (overall) employed decrease by 0.17% (0.12%). These effects are statistically significant. Self-employment increases by 0.05%, but the impact is statistically insignificant.

More significant effects are shown in the bottom part of Panel A, which reports the unconditional effects of interest on continuous labor market outcomes. The null of the exogeneity test is rejected in all cases, meaning that the 2SLS estimator is consistent while the OLS one is not. A comparison between the estimates suggests that OLS underestimate scarring, which is in line with the hypothesis that the bias is caused by returns to search, liquidity constraints or measurement errors in the endogenous regressor, and consistent with what found in the literature. The 2SLS results indicate that one pp rise in early non-employment reduces earnings and hours worked by 10% and 7%, respectively (column 4). Both estimates are highly significant (at 1% level). Note that, for the cluster robust case (column 2 and 4) the P-values are computed according to the $t(G-1)$ distribution (with G being the number of clusters), to make a conservative inference.[43] However, the P-values of column 4 may be still underestimated due to the small number of clusters. We tackle this computing the bootstrap P-value for the t statistic of β. The latter is higher than the P-value from cluster robust standard errors, but still lower than 0.05. Hence, despite the small number of clusters, the impact of early non-employment on continuous outcomes is significant.

These estimates suggest that the low educated who, as a consequence of graduating in the adverse labor market conditions, found it difficult to get a stable position at the start of the career

[41] With one instrument, the critical value for maximal size test of 10% and 15% is 16.3 and 8.96.

[42] With clustered standard errors, the exogeneity test is defined as the difference of two Sargan-Hansen statistics: one for the equation with a smaller set of instruments where the suspect regressor is treated as endogenous, and one for the equation where the suspect regressor is treated as exogenous. Under the null that both sets of instruments are valid (i.e. the suspect regressor is exogenous), the statistic is distributed as $\chi^2(1)$. Note that this statistic is not corrected for the problem of few clusters. Hence, the P-value may be too small.

[43] In Stata this is automatically done in clustered OLS, but not in clustered 2SLS, where the P-values in the second stage are computed with the Normal distribution. In this case, the estimate remains significant at 1% level. Table 2.8 in Appendix 2.11 reports the clustered 2SLS with the significance level based on the Normal distribution.

(i.e. the compliers), are still significantly penalized in terms of hours worked and earnings, 6 years after graduation. The results on hours worked are not directly comparable but consistent with the existing literature, which reports persistent effects of early unemployment on subsequent unemployment: for UK new graduates aged 16-21, Gregg (2001) estimates that a 3-months increase in the unemployment duration before age 23 significantly increases the time out of work between age 28 and 33 by 2 months. Schmillen and Umkehrer (2013a) focus on new graduates from the German apprenticeship program and find that one additional day of unemployment during the first 8 years since graduation increases unemployment in the following 16 years by 0.96 days.

Table 2.8 in Appendix 2.11 show the entire OLS and 2SLS regressions. The individual controls show the expected signs: in the first stage, grade repetitions in secondary education is positively associated with early non-employment, while technical, vocational and apprenticeship programs are associated with a lower proportion of time spent in early non-employment, compared to general education. This suggests that the former programs ease the transition from school to work. An interesting result refers to mother education, which has a positive effect on early non-employment in the first stage, whereas a significant and negative (positive) impact on hours worked, earnings and salaried employment (self-employment) in OLS. This may capture the effect of unobserved liquidity constraints on the time worked, so that less constrained individuals (associated to higher mother's education) spend more time in early non-employment and work less hours in salaried employment (with consequent lower earnings). In the same spirit, low educated individuals with low educated mothers are also more likely to opt for salaried employment (with expected stable income under long-term contracts), while they are more likely to engage in (riskier) self-employment if their mothers are high educated. These effects become insignificant in 2SLS, to the extent that the endogeneity problem due to omitted liquidity constraints in (2.2) is tackled by the IV approach.

We repeat the analysis measuring the outcomes 8 years after graduation, to investigate the persistence of such scar. These results are reported in Table 2.2. The effects are qualitatively similar to the ones measured 2 years earlier: the estimates on discrete outcomes remain very small, while the adverse effects on continuous outcomes are still significant, but smaller than the ones measured at experience 6. These results may suggest that, for low educated youth, the scar originating from early non-employment on continuous outcomes persists still 8 years after graduation, i.e. it fades away slowly. However, the bootstrap F statistic in the first stage decreases to 6, warning against the problem of weak instrument. In this case, we know that 2SLS estimator is biased towards the OLS one. Therefore, we cannot discriminate to what extent 2SLS are able to identify the fact that the scar decreases over time, and to what extent they are simply biased towards the OLS: according to the F statistic, we may worry about the latter possibility. Thus, unfortunately these results are not very informative.

Table 2.1: Effect of Interest on Outcomes Measured 6 Years After Graduation for Low Educated.

Panel A: Effect of early non-employment in the structural equation:					
		OLS		2SLS	
	standard errors†	robust	cluster $g * p$	robust	cluster $g * p$
outcomes:		(1)	(2)	(3)	(4)
salaried empl.	coeff	-0.00169***	-0.00169***	-0.00256	-0.00256
	se	(0.00034)	(0.00041)	(0.00375)	(0.00290)
	P-val§		0.00019		0.38202
	Bootstrap P-val‡		0		0.45646
	Exogeneity test P-val§§				0.767
self-empl.	coeff	0.00054*	0.00054	0.00248	0.00248
	se	(0.00030)	(0.00041)	(0.00338)	(0.00258)
	P-val		0.19177		0.34175
	Bootstrap P-val		0.18619		0.37437
	Exogeneity test P-val				0.438
overall empl.	coeff	-0.00115***	-0.00115***	-0.00008	-0.00008
	se	(0.00021)	(0.00025)	(0.00207)	(0.00151)
	P-val		0.00005		0.95655
	Bootstrap P-val		0		0.96697
	Exogeneity test P-val				0.467
log earnings	coeff	-0.0269***	-0.0269***	-0.1002**	-0.1002***
	se	(0.0033)	(0.0040)	(0.0419)	(0.0291)
	P-val		2.51E-08		0.0013
	Bootstrap P-val		0		0.0060
	Exogeneity test P-val				0.00970
log hours worked	coeff	-0.0203***	-0.0203***	-0.0722**	-0.0722***
	se	(0.0024)	(0.0029)	(0.0307)	(0.0207)
	P-val		8.96E-09		0.0011
	Bootstrap P-val				0.0060
	Exogeneity test P-val				0.0113

Panel B: Effect of the instrument in the first stage : OLS			
outcome:	*standard errors:*	robust	cluster $g * p$
early non-empl.	coeff	5.4615***	5.4615***
	se	(1.7273)	(1.6848)
	P-val		0.00230
	Bootstrap P-val		0.00400
	F stat		10.51
	Bootstrap F stat††		9.25

*** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. Panel A reports results from estimating β in (2.2) on outcomes measured at potential experience 6. β is the effect of one pp increase in $y^0_{igpt_1}$, i.e. the % of hours spent in non-employment in the first 2.5 years after graduation relative to potential total hours if one would work full-time during the whole period. For clustered standard errors, we report the P-value and the wild bootstrap P-value. Column 1-2 (3-4) show OLS (2SLS). In 2SLS the provincial unemployment rate at graduation is used as instrument for $y^0_{igpt_1}$. Panel B shows the effect of the instrument on $y^0_{igpt_1}$ in the first stage and the corresponding F statistic.

† Robust indicates heteroscedastic-robust standard errors; clusters are defined by year g and province of residence at graduation p (G=44 clusters).

§ The P-value from clustered standard errors is computed using the $t(G-1)$ distribution, with G=44 (stars are reported accordingly).

‡ Bootstrap P-values are computed according to the wild bootstrap procedures explained in Appendix 2.10 for 999 repetitions.

†† Bootstrap F statistic is the F statistic corresponding to the bootstrap P-value of the t statistic of the instrument: we rely on the equivalence between F and t distribution: for $G = 44$, $t^2(G-1) = F(1, G-1)$.

§§ With clustered standard errors, this test is defined as the difference of two Sargan-Hansen statistics: one for the equation where y^0_{t1} is treated as endogenous and one for the equation where $y^0_{t_1}$ is treated as exogenous. Under the null that $y^0_{t_1}$ is exogenous, the statistic is distributed as $\chi^2(1)$.

Table 2.2: Effect of Interest on Outcomes Measured 8 Years After Graduation for Low Educated.

Panel A: Effect of early non-employment in the structural equation:					
		OLS		2SLS	
	standard errors†	robust	cluster $g*p$	robust	cluster $g*p$
outcomes:		(1)	(2)	(3)	(4)
salaried empl.	coeff	-0.00119***	-0.00119**	-0.00009	-0.00009
	se	(0.00035)	(0.00046)	(0.00430)	(0.00323)
	P-val§		0.01252		0.97670
	Bootstrap P-val†		0.01401		0.94494
	Exogeneity test P-val§§				0.739
self-empl.	coeff	0.00030	0.00030	-0.00087	-0.00087
	se	(0.00031)	(0.00043)	(0.00385)	(0.00354)
	P-val		0.48347		0.80640
	Bootstrap P-val		0.48448		0.82282
	Exogeneity test P-val				0.744
overall empl.	coeff	-0.00089***	-0.00089***	-0.00097	-0.00097
	se	(0.00019)	(0.00021)	(0.00220)	(0.00187)
	P-val		0.00013		0.60815
	Bootstrap P-val		0		0.68068
	Exogeneity test P-val				0.966
log earnings	coeff	-0.0230***	-0.0230***	-0.0674*	-0.0674**
	se	(0.0033)	(0.0042)	(0.0394)	(0.0311)
	P-val		2.34E-06		0.0360
	Bootstrap P-val		0		0.0841
	Exogeneity test P-val				0.162
log hours worked	coeff	-0.0172***	-0.0172***	-0.0487*	-0.0487**
	se	(0.0024)	(0.0032)	(0.0290)	(0.0224)
	P-val		2.53E-06		0.0355
	Bootstrap P-val		0		0.0861
	Exogeneity test P-val				0.171

Panel B: Effect of the instrument in the first stage : OLS			
outcome:	*standard errors:*	robust	cluster $g*p$
early non-empl.	coeff	4.9717***	4.9717***
	se	(1.6605)	(1.6060)
	P-val		0.00345
	Bootstrap P-val		0.01802
	F stat		9.584
	Bootstrap F stat††		6.05

*** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. Panel A reports results from estimating β in (2.2) on outcomes measured at potential experience 8. β is the effect of one pp increase in $y^0_{igpt_1}$, i.e. the % of hours spent in non-employment in the first 2.5 years after graduation relative to potential total hours if one would work full-time during the whole period. For clustered standard errors, we report the P-value and the wild bootstrap P-value. Column 1-2 (3-4) show OLS (2SLS). In 2SLS the provincial unemployment rate at graduation is used as instrument for $y^0_{igpt_1}$. Panel B shows the effect of the instrument on $y^0_{igpt_1}$ in the first stage and the corresponding F statistic.

† Robust indicates heteroscedastic-robust standard errors; clusters are defined by year g and province of residence at graduation p (G=44 clusters). § The P-value from clustered standard errors is computed using the $t(G-1)$ distribution, with G=44 (stars are reported accordingly).

‡ Bootstrap P-values are computed according to the wild bootstrap procedures explained in Appendix 2.10 for 999 repetitions.

†† Bootstrap F statistic is the F statistic corresponding to the bootstrap P-value of the t statistic of the instrument: we rely on the equivalence between F and t distribution: for $G = 44$, $t^2(G-1) = F(1, G-1)$.

§§ With clustered standard errors, this test is defined as the difference of two Sargan-Hansen statistics: one for the equation where y^0_{t1} is treated as endogenous and one for the equation where $y^0_{t_1}$ is treated as exogenous. Under the null that $y^0_{t_1}$ is exogenous, the statistic is distributed as $\chi^2(1)$.

2.5.1 Sensitivity Analysis for the Low Educated

As a first sensitivity analysis, we want to rule out that results in Table 2.1 are driven by multicollinearity, which may arise because in (2.2) we add many controls for the persistence of the unemployment rate (the current unemployment rate, the average unemployment rate since the end of the early period and the moment of evaluation as well as the minimum unemployment rate since graduation). Therefore, to rule this possibility we re-run the model including only the current unemployment rate (thereby excluding $\overline{UR}_p$ and $minUR_{pt}$): the results of this restricted specification are reported in Table 2.12 of Appendix 2.12. Note that the effect of interest should be interpreted as an average effect between the scarring of early non-employment due to adverse labor market conditions at graduation and the persistence of the unemployment rate in the period before the moment of evaluation.[44] The stability of the results ensures that multicollinearity is not driving the results.

Next, we assess the impact of measurement error in the endogenous regressor, which arises from the combination of survey and administrative data. In fact, early non-employment is measured precisely for individuals graduating in the 1998-2002 period - by means of administrative data - but it is imputed for those graduating in the period 1994-1997, based on the Sonar database (see Appendix 2.8 for details). This allows us to maximize the variation of the instrument considering the entire graduation period 1994-2002, at the cost of introducing some measurement error in the endogenous regressor. Therefore, in this second sensitivity analysis we re-run the analysis for low educated restricting the sample to graduation period 1998-2002, so that the endogenous regressor is measured uniquely by administrative data. Of course, clusters are drastically reduced from 44 to 24.[45] This is problematic not only because it exacerbates the problem of few clusters, but also because equation (2.2) contains too many parameters ($k = 30$) compared to the number of clusters, and as a consequence the rank condition in 2SLS is not satisfied.[46] Therefore, we need to reduce the parameters in equation (2.2).

We decide to exclude some of the non-significant individual controls and rather include in the specification all the aggregate regressors, which are very important to ensure the validity of the exclusion restriction.[47] In particular, we drop the following controls that are jointly not significant at the 5% level according to an F test in the first stage regression: dummy for living with single parent, dummy for not living with parents, number of household members aged 12-17, 18-29, 30-64, 65+ (we keep the number of household members aged 0-11 since it is significant); plus, we aggregate all educational tracks different from general education (technical, vocational, part-time education or apprenticeship) and include a dummy for general education instead.[48] Therefore, the new regression includes the following individual controls: father and mother education, repeated grades at age 17, dummy for general education at age 17, number of household members aged 0-

[44]This is because $\overline{UR}_p$ is significant in (2.2) while $minUR_{pt}$ is not: see Table 2.8 and 2.9 Appendix 2.11.

[45]In principle, 5 graduation years times 5 provinces, i.e. $G = 25$. However, $g2002p3$ is empty.

[46]The variance-covariance matrix of moment conditions has size (30×30) and rank=24 (Baum et al., 2003).

[47]Compared to the aggregate regressors individual controls play a minor role, as they alleviate the problem of omitted individual characteristics, which is anyway tackled by the IV approach.

[48]The reference is an aggregated category for technical, vocational, part-time education and apprenticeship.

11 when the individual is aged 17, birth cohort dummies. Table 2.11 shows that results are robust to this alternative specification, as the OLS estimate of the endogenous regressor is very stable in the full specification (column 1 and 4) and in the restricted specification (column 2 and 5); both specifications consider the graduation period 1994-2002.

Panel B of Table 2.10 reports the first stage regression for the 1998-2002 period. This table should be compared to Table 2.1. First, the bootstrap F statistic is 3.6 for the graduation period 1998-2002 compared to 9.2 for the period 1994-2002: therefore, the instrument becomes weak by restricting to the former period. As expected, the increased discrepancy between the original F and the bootstrap F statistic in Table 2.10 compared to Table 2.1 shows that the few-clusters bias worsens a lot by shifting from 44 to 24 clusters. Second, the direct effect of the instrument on early non-employment for the 1998-2002 period doubles compared to the period 1994-2002. An explanation is that the former period focuses on the dot-com recession, whose effects are mitigated in considering a larger span. Given the low F statistic, 2SLS are not reliable.

However, we can focus on the OLS results in Table 2.11 to shed lights on the importance of the measurement error in the endogenous regressor. In principle, this should bias the OLS estimate for the period 1994-2002 (Column 1 and 4) towards zero. At the same time, the OLS estimate in Column 3 and 6 should not be affected by measurement error "by construction", since the endogenous regressor is entirely measured by administrative data for the 1998-2002 period. We therefore compare the first row across columns (1 with 3 and 4 with 6): for each outcome, the estimate based on the graduation period 1994-2002 is slightly smaller than the corresponding estimate for the period 1998-2002. This is consistent with the presence of measurement error in the endogenous regressor for the graduation period 1994-2002. However, this difference is small (0.2 *pp*), which suggests that overall the OLS estimates are quite close in the 1998-2002 and 1994-2002 period: as a consequence, we conclude that measurement error in early non-employment is not a major issue in the main results.

Of course, another explanation could be that the combination of bias from various sources (ability, returns to search, liquidity constraints) may differ in the period 1998-2002 - when measurement error is absent - compared to the period 1994-2002 - when measurement error is present - and yet yield the same net effect: however, this requires that these (fixed) omitted factors affect individual labor market performances differently in the 1994-1997 and 1998-2002 period, which is peculiar. We think that this last story is more difficult to be argued.

2.5.2 Discussion on the validity of the IV approach

Under the assumption of homogenous effects, the 2SLS identifies the effect of interest under the Condition 1 and 2 illustrated in Section 2.4. This section discusses the limitations of this approach by assessing the validity of each assumption and the role played by each of them in the identification. Consistently with the notation used in Section 2.4, let Y be the outcome, Z the instrument, Y^0 the endogenous regressor. The IV estimator is $\beta_{IV} = cov(Y,Z)/cov(Y^0,Z)$. We start noting that it can be expressed as the ratio between the causal effect of Z on Y in the reduced form regression (RF) and the causal effect of Z on Y^0 in the first stage ($1^{st} stage$): $\beta_{IV} =$

$\frac{cov(Y,Z)}{cov(Y^0,Z)} = \frac{cov(Y,Z)/Var(Z)}{cov(Y^0,Z)/Var(Z)} = \frac{RF}{1^{st}stage}$. Condition 1 (exogeneity) is required to interpret the numerator as causal. By virtue of the exclusion restriction, this causal effect can only result from Z effecting Y^0 and Y^0 in turn affecting Y. Condition 2 allows us to quantify the effect of interest (the effect of Y^0 on Y), dividing $cov(Y,Z)$ by $cov(Y^0,Z)$, which yields the IV estimator. At the extreme in which Z has no explanatory power in the first stage, $cov(Y^0,Z) = 0$ and the ratio would not exist. If condition 1 is not satisfied, the reduced form yields a biased estimate of the effect of Z on Y and, as a consequence, the IV estimator provides a biased estimate of the effect of Y^0 on Y. Furthermore, the bias of the IV estimator is magnified in presence of weakly correlated instruments.

In Section 2.4 we extensively discuss the validity of Condition 1. We formally test the exogeneity of the timing of graduation, and argue that the place of graduation is also exogenous since we hardly observe any inter-provincial mobility in the data. However, commuting across provinces is a source of endogeneity that we cannot control for. Moreover, we identify a number of factors that may represent a source of violation of the exclusion restriction and set up the specification in (2.2) so that the latter is most likely satisfied, conditional on the covariates. In our case, Condition 2 is the most problematic, as the instrument is not strong enough. As already mentioned, this may magnify problems related to even small violations of Condition 1.

Differently from the IV estimator, the effect of Z on Y in the reduced form equation only requires the exogeneity of Z to be given a causal interpretation. For this reason, we complement the IV approach with the estimates of the reduced form equation, which are reported in Table 2.13 together with the first stage:[49] for continuous outcomes, these effects are unconditional on being salaried employed. The effect of Z on the employment indicators evaluated 6 years after graduation are not significant (Columns 4-6 of Table 2.13). Instead, the unconditional effects of Z on the continuous variables are negative and significant: one *pp* increase in the unemployment rate at graduation decreases earnings by 54% and hours worked by 39%, 6 years after graduation. These effects are very big. The reason is that they are unconditional average effects, which include also the effects for those who are not salaried employed at the moment of evaluation. Dividing these effects by the effect of Z on Y^0 in the first stage (Column 7) gives the 2SLS estimates of interest in Table 2.1.

So far the discussion has relied on the quite unrealistic assumption of homogenous effects. Allowing for heterogeneous effects, an IV estimator identifies an average causal effect for the sub-population that reacts to the instrument under two additional assumptions:[50]

- Stable Unit Treatment Value Assumption (SUTVA): no interference between units.
- Monotonicity: the effect of Z on Y^0 has the same sign for everybody.

In a simple framework with no covariates where both Z and Y^0 are discrete, the IV estimator identifies the local average treatment effect (LATE) for the compliers. In a more complicated framework as in our case, the IV estimator can be interpreted as a weighted average of local

[49] In the reduced form, $y^0_{igpt_1}$ in (2.2) is replaced by the unemployment rate at graduation; in the first stage, $y^0_{igpt_1}$ in (2.2) is replaced by the unemployment rate at graduation while at the same time y_{iT} is replaced by $y^0_{igpt_1}$.

[50] Condition 1 is split into random assignment and exclusion restriction. Condition 2 remains invariant.

average treatment effects (LATEs) (Angrist and Pischke, 2008).[51]

The SUTVA assumption requires that the outcomes Y are independent across individuals. This is satisfied if one can rule out crowding-out effects: these effects occur if there are less job openings than new graduates so that the latter compete for the same jobs and, as a consequence, only some of them get good positions while some others remain unemployed, get lower-paying jobs, or work less hours. This is a quite strong assumption to make in the context of labor markets. The monotonicity assumption requires that a higher unemployment rate at graduation Z prolongs (shortens) early non-employment for every low educated person. The direction of the effect does not matter to the extent that it has the same sign for all individuals.[52] Also this assumption boils down to crowding out effects in the labor market, as it rules out the possibility that, by graduating in a downturns, some new graduates may experience longer early non-employment while other luckier ones work since the start of the career. As for the case of homogenous effects, the IV is the ratio between the effect of Z on Y in the reduced form equation and the effect of Z on Y^0 in the first stage. In the program evaluation literature, the numerator is the Intention-To-Treat (ITT) effect. The SUTVA ensures that the numerator is not biased, which typically occurs in case of externalities (Heckman et al., 1999). Monotonicity (together with the exclusion restriction) ensures that the IV estimator identifies the causal effect of interest for the compliers - i.e. in this case those who experience high incidence of non-employment because graduated with high unemployment rates. If either the exclusion restriction or the monotonicity are violated, the IV estimates the LATE for the compliers plus a bias, which can be characterized depending on the assumption that is not satisfied. Clearly, allowing for heterogeneous effects comes at the cost of requiring more assumptions which, as already mentioned, may be restrictive in this specific case.

We conclude relating the analysis of this chapter with Chapter 1, since both of them rely on the same data (see Appendix 2.7 for details). In the current section we estimate unconditional ITT effects of the instrument on alternative outcomes which are evaluated at potential experience 6. Given that the dependent variables are fixed at a given point in time, we exploit the cross-section variation. By contrast, Chapter 1 exploits the entire panel structure of the data and jointly estimate distinct ITT effects on the same outcomes - one for each year of potential experience. The effects on continuous variables are conditional on salaried employment. Consequently, these results cannot be directly compared since the two studies use different frameworks. Namely, in Chapter 1, the panel structure allows to control for calendar time FE in the specification (which accounts for common shocks affecting the entire economy),[53] and in general contain much more informa-

[51] With covariates X, the IV is a weighted average of covariates-specific LATEs (one for each value of X), where more weight is attributed to covariate values where Z creates more variation in the fitted values. If Y^0 is continuous, the IV is a weighted average of LATEs, where the weights depend on how the compliers are distributed over the range of Y^0. With multiple instruments, the IV is a weighted average of LATEs for instrument-specific compliant subpopulations, where the weights are proportional to the relative strength of each instrument in the first stage. Similarly, if Z is continuous, the IV is a weighted average of instrument values-specific LATEs, and bigger weights are given to the instrument values that contribute the most in explaining Y^0 in the first stage: here, bigger weights are given to the clusters gp whose unemployment rate variation is mostly correlated with early non-employment.

[52] Monotonicity is important given the heterogeneity of the effect, since the LATE is an average causal effect.

[53] There are additional differences in the specification: Chapter 1 includes asymmetric effects for graduating in

tion, which yields more precise estimates. By excluding $minUR_{pt}$ and $\overline{UR}_p$ from (2.2) - which gives the most similar specification to the one estimated in Chapter 1 - we estimate a conditional ITT of -3.8% on earnings and -0.9% on hours worked (estimation output not reported).[54] The corresponding effects in Chapter 1 are -3.6% on earnings and -2.5% on hours worked (see Table 1.3 in Section 1.9). As expected, results from the panel estimation are more precise.[55] The point estimates of the effect on earnings are quite close across the two studies, as opposed to the ones referring to the effect on hours worked. However, the latter are not inconsistent, given the large confidence intervals of the conditional ITT resulting from our approach.

2.6 Conclusions

We consider a sample of low educated youth graduating in the period 1994-2002 in Flanders, the most prosperous of three Belgian regions. We study the impact of early non-employment on workers' subsequent career, measured 6 and 8 years after graduation. We deal with the endogeneity of unobserved individual characteristics with an IV approach, where the provincial unemployment rate at graduation is used as instrument for non-employment. The problem of few clusters is addressed by wild bootstrap methods.

Since in Belgium labor market institutions differ for blue and white collar workers, we focus on low educated new graduates. Chapter 1 of this dissertation provides evidence that Belgian low and high educated graduating in the period 1994-2004 are exposed differently to adverse labor market conditions because of these institutional differences: due to strict EPL for white collars, the high educated graduating in downturns are forced to downgrade and as a consequence are trapped in lower-paying jobs. The low educated instead tend to experience non-employment because of the flexible EPL for blue collars, and because wages are downwards-rigid due to the presence of minimum wages. We define the identification strategy in light of these results, so to ensure that the exclusion restriction is most likely satisfied: that is, we consider a measure of early non-employment as endogenous variable, since for the low educated the scar from graduating in downturns occurs through the loss of early work experience.[56]

In this study we have applied an IV approach to unveil the causality between early non-employment and subsequent labor market outcomes. Throughout the article we have discussed the assumptions required by the IV estimator, their validity and the role they play in the identification. Unfortunately, our IV estimators suffer from weak identification problem, which may magnify any small violation of the exclusion restriction.

We find that one *pp* increase in the time spent in non-employment in the first 2.5 years since

upturns/downturns, and interact the unemployment rate at graduation and the current one with experience splines. In current chapter, $minUR_{pt}$ and $\overline{UR}_p$ are included in (2.2).

[54]We focus on the continuous outcomes as they yield significant ITT effects, as opposed to the discrete ones.

[55]Standard errors of conditional ITT on earnings and hours worked are respectively 0.015 and 0.014 in Chapter 1, and 0.027 and 0.025 in this study.

[56]For high educated we use early wage as endogenous regressor as accepting lower-paying jobs is the main channel to explain the scars of graduating in downturns. However, the instrument is weak (see Appendix 2.14).

graduation decreases annual earnings and hours worked from salaried employment by 10% and 7% respectively, 6 years after graduation. Provided that our identification strategy is correct, these effects are causal. They may originate by the foregone human capital that would have been accumulated in case of early work-experience, or because early non-employment is interpreted as a signal of low quality. From a policy perspective, it is therefore very important that low educated workers acquire work experience at the start of their career. This may be impeded in a rigid labor market where workers reallocation is costly and as a consequence exiting unemployment is harder and may have long-lasting consequences. Early work experience instead may be enhanced by a *flexicurity* system in which workers are reallocated easily while unemployed workers are protected by an unemployment insurance system. In this context, the majority of unemployment is temporary - to the extent that the hiring cost are low. Long-term unemployed, who are provided for by the unemployment insurance, may be additionally supported in their job search by active labor market policies.

The Belgian low educated workers face a number of rigidities which restrain workers reallocation: the short-term work compensation program, by anchoring the employees to the employers, is an example. A second rigidity is represented by the very high Belgian minimum wages, which limit the absorbtion of low educated new graduates for whom minimum wages are binding. A third one is the asymmetry between the flexible EPL for blue collars and a rigid EPL for white collars, which characterized the Belgian labor market until 2013: in adverse labor market conditions, low educated new graduates risked to be marginalized, if they had to face the additional competition of higher educated new graduates, who in turn downgraded because of rigid EPL for white collar workers - which increased the hiring costs to fill white collar positions. Thus the low educated experienced serious difficulties to find stable jobs as a consequence of this asymmetry. Note that this controversial discrimination between blue and white collar workers has been removed since the beginning of 2014, as a single employment contract has been introduced, stipulating the same EPL for white and blue collar workers.

Acknowledgements

I wish to thank Bart Cockx, William Parienté, Matteo Picchio, Bruno Vanderlinden, Dieter Verhaest, and the participants to the seminars at Ghent University (internal seminar), and the participants to the Doctoral Workshop at the Université Catholique de Louvain for all their valuable comments. I am grateful to the Cross Roads Bank for Social Security and the “Steunpunt Studie- en Schoolloopbanen" for the availability of the data. I acknowledge financial support from the Special Research Fund (“BOF”) of Ghent University, as well as from the National Bank of Belgium (project code 05Y00110). The usual disclaimer applies.

2.7 Appendix A. Description of the Final Sample

In this study we consider almost the same sample as in Cockx and Ghirelli (2014). They consider 3514 individuals (comprising both low and high educated) while we consider 3586 individuals (see Table 2.3): i.e. we add 72 individuals (2% of the sample), including low educated graduating in 2002 and high educated graduating in 1995-96. For details in the construction of the control and outcome variables, see Section A.1, A.2 and A.3 of this dissertation.

Table 2.3: Dividing the Sample in Low and High Educated

education	low educated	high educated	Total
1	2		2
2	36		36
3	89		89
4	113		113
5	185		185
6	1,111		1,111
7	366	289	655
8		55	55
9		707	707
10		367	367
11		232	232
12		33	33
13		1	1
Total	1,902	1,684	3,586

Completed education refers to the number of years of education successfully attained from the beginning of secondary education, i.e. at age 12. Low educated are those who graduate with at most secondary education, which consists in 7 years of education in case of vocational track and 6 years for all other educational programs. High educated are those with higher than secondary education.

Table 2.4: Correspondence Between Low-High Educated and Blue-White Collar Workers

Function Undertaken 6 Years After Graduation†						
hline	Low educated			High educated		
	Freq.	Percent	Cum.	Freq.	Percent	Cum.
blue collar	1,193	62.72	62.72	184	10.93	10.93
white collar	390	20.5	83.23	1,193	70.84	81.77
functionary	68	3.58	86.8	105	6.24	88
missing	251	13.2	100	202	12	100
Total	1,902	100		1,684	100	
Prevalent Function Undertaken up to 6 Years After Graduation‡						
	Low educated			High educated		
	Freq.	Percent	Cum.	Freq.	Percent	Cum.
blue collar	1,346	70.77	70.77	235	13.95	13.95
white collar	401	21.08	91.85	1,337	79.39	93.35
functionary	53	2.79	94.64	36	2.14	95.49
missing	102	5.36	100	76	4.51	100
Total	1,902	100		1,684	100	

† It refers to the type of function undertaken at potential experience 6.
‡ It refers to the function that is undertaken more than 50% of the time from graduation up to potential experience 6. 70% (14%) of low (high) educated are prevalently employed as blue collars. Thus, there is clear correspondence between low (high) educated and blue (white) collars.

Table 2.5: Number of Individuals by Graduation Year and Province of Residence at Graduation

	Low educated					
grad_year	prov1	prov2	prov3	prov4	prov5	Total
1994	30	9	31	48	25	143
1995	47	22	44	48	48	209
1996	84	45	65	85	38	317
1997	78	41	65	67	36	287
1998	111	46	78	90	61	386
1999	99	42	47	64	47	299
2000	56	18	30	28	31	163
2001	26	8	11	17	18	80
2002	10	3	0	2	3	18
Total	541	234	371	449	307	1902

The analysis considers the graduation period 1994-2002 for the low educated. The combination g2002 & prov3 is excluded since empty. Provinces are in the following order from 1 to 5: Antwerp, Flemish Brabant, Western Flanders, Eastern Flanders, Limburg. Each combination of graduation year and province of residence at graduation represents a cluster *gp* in the main analysis.

Table 2.6: Descriptive Statistics of Individual Control Variables: Low Educated

Variable	Obs	Mean	Std. Dev.	Min	Max	label
birth cohort76	1902	0.330	0.470	0	1	1 if born in 1976
birth cohort78	1902	0.332	0.471	0	1	1 if born in 1978
birth cohort80	1902	0.338	0.473	0	1	1 if born in 1980
live in single-parent	1902	0.120	0.326	0	1	1 if live with single parent at age17(Dec)
not live with parents	1902	0.064	0.244	0	1	1 if not live with either parents at age17(Dec)
HH members aged 0-11	1902	0.248	0.625	0	7	nr of other HH members aged0-11 at age17(Dec)
HH members aged 12-17	1902	0.508	0.689	0	7	nr of other HH members aged12-17 at age17(Dec)
HH members aged 18-29	1902	0.521	0.731	0	8	nr of other HH members aged18-29 at age17(Dec)
HH members aged 30-64	1902	1.889	0.400	0	5	nr of other HH members aged30-64 at age17(Dec)
HH members aged 65+	1902	0.037	0.211	0	2	nr of other HH members aged65+ at age17(Dec)
father education	1902	4.586	3.201	0	13	father completed education since age12
mother education	1902	4.212	3.062	0	13	mother completed education since age12
years of delay in sec.edu	1902	0.828	0.840	-1	4	years of delay at age17(Aug)
general education	1902	0.110	0.313	0	1	1 if general edu at age17(Aug)
technical education	1902	0.379	0.485	0	1	1 if technical edu at age17(Aug)
vocational education	1902	0.410	0.492	0	1	1 if vocational edu at age17(Aug)
apprenticeship/PT voc.	1902	0.100	0.301	0	1	1 if apprenticeship/PT voc. edu at age17(Aug)

Descriptive statistics are computed on the sample used in the main analysis, in which the dependent variables are measured 6 years after graduation.

Table 2.7: Descriptive Statistics of Outcomes and Endogenous Regressor: Low Educated

Variable	Obs§	Mean	Std. Dev.	Min	Max	Label
						At potential experience 6
log earnings	1902	8.419	3.588	0	10.770	log annual gross earnings from salaried empl.
log hours	1902	6.202	2.669	0	7.725	log annual hours worked in salaried empl.
self-empl.	1902	0.122	0.327	0	1	1 if only pos. earnings from salaried (& not self-empl)
salaried empl.	1902	0.837	0.369	0	1	1 if registered as self-empl.
overall empl.	1902	0.959	0.198	0	1	1 if pos.earnings from salaried or registered as self-empl.
early non-empl.	1902	30.596	29.647	0	100	% hours not worked relative to FT salaried empl.
						At potential experience 8
log earnings	1894	8.615	3.461	0	10.943	log annual gross earnings from salaried empl.
log hours	1894	6.311	2.563	0	7.725	log annual hours worked in salaried empl.
self-empl.	1894	0.150	0.357	0	1	1 if only pos. earnings from salaried (& not self-empl)
salaried empl.	1894	0.822	0.383	0	1	1 if registered as self-empl.
overall empl.	1894	0.971	0.166	0	1	1 if pos.earnings from salaried or registered as self-empl.
early non-empl.	1894	30.553	29.634	0	100	% hours not worked relative to FT salaried empl.

Descriptive statistics is based on the sample studied in the main analysis, in which low educated graduate in the period 1994-2002. The employment indicators are related as follows: salaried+self=overall employment. For continuous outcomes we add value one before taking the log, so that non-salaried employed at the moment of evaluation are included with outcomes=0 after the logarithmic transformation.

§ Low educated are observed in both years of potential experience, except for 8 individuals who are observed in the labor market at potential experience 6 but not 2 years later.

2.8 Appendix B. Construction of the Endogenous Variable

We define potential experience as a variable counting each calendar year since graduation. Potential experience 0 corresponds to the year of graduation and runs from the month after graduation until December of that calendar year; therefore, it potentially lasts less than 12 months (6 months for a student graduating in June). Potential experience 1 runs from January until December of the subsequent calendar year, thereby lasting 12 months. Subsequent potential experience years are defined similarly.

The regressor of interest is the percentage of hours spent in non-employment at potential experience 0-2, relative to potential total hours if one would work full-time during the whole period. We express everything in hours because this is the smallest unit of measurement used in the administrative data (it is used to measure time worked in part-time employment). The reference period is computed considering the entire calendar year for potential experience 1 and 2, and the part of the calendar year following the month of graduation for potential experience 0. That is, for one who graduated in June, the time spent working at potential experience 0-2 is divided by the total working hours in 30 months of full-time salaried employment. As already mentioned in Section 2.3 of the main text, for simplicity we refer to this reference period as "the first 2.5 years since graduation", i.e. the average reference period since most of the sample graduates in June.

Define a as the total hours worked (including self-employment) in the first 2.5 years from graduation and b the potential total hours if one wold work full-time during the whole period; then the regressor of interest is computed as $100 * (b - a)/b$. Below we explain in detail how its components are constructed.

1. Construct a according to the following steps.

   I. a is mostly based on the total hours worked in salaried employment and the date of registration and cancellation from the self-employment register from the Data Warehouse. Since hours worked are not available for self-employment, we assume that the latter work as much as a full-time salaried worker: i.e., 5 days per week and 8 hours per day until 2002, and 5 days per week and 7.6 hours per day from 2003 onwards. This is due to the introduction of a new law in Belgium that changed the daily working hours from 8 to 7.6 from the first of January 2003. Whenever one combines self-employment and salaried-employment in the same quarter we make the same assumption, so that the hours worked do not exceed the bounds.

   II. The construction of a requires an additional adjustment due to the limited availability of the administrative data, which cover the period 1998-2010. Since the sample contains 3 birth cohorts (1976, 1978, 1980) and that compulsory education ends at age 18 in Belgium, these data can be used in the following cases (68.5% of the final sample): all individuals born in 1980, those born in 1978 graduating at least at age 20 and those born in 1976 graduating at least at age 22. Figure 2.1 summarizes the availability of the data. To retain in the analysis also students born in 1978 (1976) graduating at age

18-29 (18-21),[57] we exploit the monthly working status from the Sonar database and impute the values of a following the procedure used for self-employed workers. That is, for each month in which individuals are working according to Sonar we attribute the working hours of a full-time salaried worker: i.e. 8 (7.6) working hours per day until (strictly after) 2002 and 21.6 working days per month (assuming 65 working days in a quarter gives 21.6 working days per month: $21.6 \times 3 = 65$).

2. Construct b. Recall that it is defined as the potential total hours if one wold work full-time during the first 2.5 years since graduation. As for a, we consider a full-time working regime of 5 days per week and 8 (7.6) hours per day until (strictly after) 2002. This gives a total of 2080 annual working hours until 2002 (8 hours/day $\times$ 65 days/quarter $\times$ 4 quarter/year) and 1976 annual working hours from 2003 onwards (7.6 hours/day $\times$ 65 days/quarter $\times$ 4 quarter/year).

3. The regressor of interest is computed as $(b - a)/b * 100$ and hence ranges between $[0, 100]$. In some cases (10% of the final sample) this percentage is negative because of overtime work. Therefore, it is censored at 0.

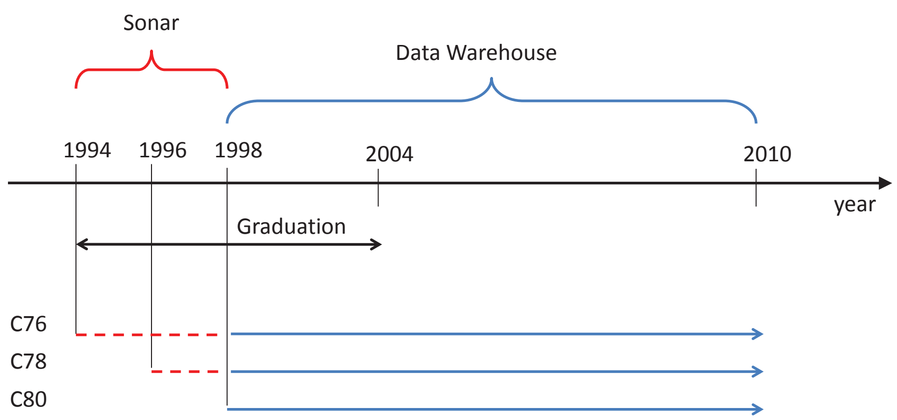


Figure 2.1: Availability of data for the construction of the main regressor. In 1998 birth cohorts 76, 78 and 80 are aged 22, 20 and 18, respectively. Birth cohort 76 (78) turns age 18 in 1994 (1996).

[57]These cases correspond to 31.5% of the final sample.

2.9 Appendix C. LFS - Provincial unemployment rate

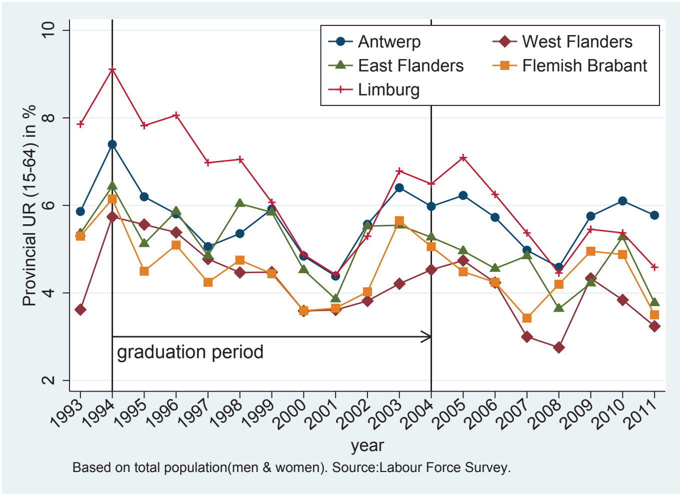


Figure 2.2: Provincial unemployment rates (15-64) for Flanders: graduation period 1994-2004, considering low and high educated together. For details, see Section A.1.5 of this dissertation.

2.10 Appendix D. Bootstrap Procedure

The basic idea of bootstrap testing is to compare the observed value of some test statistic with the empirical distribution of B bootstrap test statistics computed on as many pseudo-samples, where B is the number of bootstrap replications. We use a *Wild Restricted Efficient Residual Bootstrap* (WRE Bootstrap) proposed by Davidson and MacKinnon (2010). It is the wild version of the Restricted Efficient Residual bootstrap designed for 2SLS by Davidson and MacKinnon (2008). Few words on the terminology (which will become clearer below): *Residual* means that the objects to sample in generating the pseudo-samples are the residuals.[58] *Wild* refers to a procedure that creates pseudo-samples based on $residuals * 1$ with probability 0.5 and $residuals * (-1)$ with probability 0.5, with this assignment at the cluster level. This allows to preserve the intra-cluster correlation. *Efficient* means that the first stage of 2SLS is efficiently estimated in case of weak instruments. *Restricted* means that the null hypothesis of interest is imposed on the data generating process (DGP): this enhances efficiency in the procedure. Consider the following system of equation, which is a simplified version of the 2SLS model of interest:

$$y_{igp} = \beta y^0_{igp} + x'_i\delta + e_{igp} \tag{2.3}$$

$$y^0_{igp} = \pi Z_{gp} + x'_i\delta + u_{igp} \tag{2.4}$$

(2.3) is the structural equation where individual labor market outcomes y_{igp} are regressed on early non-employment y^0_{igp} and individual controls (for simplicity we omit the grouped covariates and

[58] Alternatively, one can sample pairs $[y\, X]$ of data.

the time subscripts in (1)), and (2.4) is the first stage regression where the endogenous explanatory variable is regressed on the grouped instrument Z_{gp} and all exogenous regressors x_i. The fact that the instrument is grouped requires cluster robust standard errors in 2SLS. We are interested in bootstrapping the t statistic of y^0_{igp}, i.e. $t(\hat{\beta}, \beta_0) = \frac{(\hat{\beta}-\beta_0)}{se(\hat{\beta})}$. Call $\hat{\tau}$ the observed value of this statistic. The bootstrap procedure will generate an empirical distribution of B bootstrap test statistics τ^*, with B being the number of repetitions, and where these statistics are generated using the bootstrap DGP which imposes the null hypothesis that is tested. In practice this is implemented as follows:

1. Estimate the system in (2.3)-(2.4) by 2SLS with cluster robust standard errors and obtain the statistic $\hat{\tau}$.

2. Estimate the restricted version of (2.3) by OLS imposing the null hypothesis $\beta = 0$ (with conventional standard errors). Predict the residuals $\tilde{e}_{igp}$ and the fitted values $\tilde{y}_{igp}$. (*Restricted*)

3. Estimate (2.4) including $\tilde{e}_{igp}$ as additional control, i.e.:$y^0_{igp} = \pi Z_{gp} + x'_i\delta + \gamma\tilde{e}_{igp} + residuals$. Compute the residuals $\tilde{u}_{igp} = residuals + \hat{\gamma}\tilde{e}_{igp}$. This allows the residuals of the first stage not to be too small in case of weak instrument (*Efficient*). Accordingly, compute the fitted values $\tilde{y}^0_{igp} = \hat{\pi}Z_{gp} + x'_i\hat{\delta}$.

4. At the cluster level, multiply the residuals $\tilde{u}_{igp}$ and $\tilde{e}_{igp}$ by a random variable ν^*, where $\nu^* = 1$ and $\nu^* = -1$ with probability $1/2$, respectively. Note that the same ν^* is applied to both residuals: this preserves the correlation across (2.3) and (2.4).[59] (*Wild*)

5. Construct $y^*_{igp} = \tilde{y}_{igp} + \nu^*\tilde{e}_{igp}$ and $y^{0*}_{igp} = \tilde{y}^0_{igp} + \nu^*\tilde{u}_{igp}$.

6. Estimate (2.3)-(2.4) by 2SLS where y_{igp} is replaced by y^*_{igp} and y^0_{igp} by y^{0*}_{igp}, with cluster robust standard errors and obtain the t statistics τ^*.

7. Repeat steps 4-6 B times, where B is the number of repetitions, so to get an empirical distribution of τ^*_j for $j = 1, ..., B$.

8. Calculate the bootstrap P-value as $p^*(\hat{\tau}) = 2min(\frac{1}{B}\sum_1^B \mathbf{1}[\tau^*_j < \hat{\tau}], \frac{1}{B}\sum_1^B \mathbf{1}[\tau^*_j > \hat{\tau}])$.

Below we describe the simpler procedure to compute wild bootstrap in the OLS case. In the main analysis, we apply this to the OLS estimations of the structural equation and to the first stage. Below we take the first stage as example: we are interested in estimating (2.4) by OLS and then in bootstrapping the t statistic of the instrument, $t(\hat{\pi}, \pi_0) = \frac{(\hat{\pi}-\pi_0)}{se(\hat{\pi})}$. The procedure is reported below:

1. Estimate (2.4) by OLS with cluster robust standard errors and obtain the statistics $\hat{\tau}$.

[59] Here we use the Rademacher weights, which have been shown to work well when the residuals are not too asymmetric. Other weights can be used.

2. Re-estimate (2.4) imposing the null hypothesis $\pi = 0$ (with conventional standard errors). Predict the residuals $\tilde{u}_{igp}$ and the fitted values $\tilde{y}^0_{igp}$. (*Restricted*)

3. At the cluster level, multiply the residuals $\tilde{u}_{igp}$ by a random variable ν^*, where $\nu^* = 1$ and $\nu^* = -1$ with probability $1/2$, respectively.

4. Construct $y^{0*}_{igp} = \tilde{y}^0_{igp} + \nu^* \tilde{u}_{igp}$.

5. Estimate (2.4) by OLS where y^0_{igp} is replaced by y^{0*}_{igp}, with cluster robust standard errors and obtain the t statistics τ^*.

6. Repeat steps 3-5 B times, where B is the number of repetitions, so to get an empirical distribution of τ^*_j for $j = 1, ..., B$.

7. Calculate the bootstrap P-value as $p^*(\hat{\tau}) = 2min(\frac{1}{B}\sum_1^B \mathbf{1}[\tau^*_j < \hat{\tau}], \frac{1}{B}\sum_1^B \mathbf{1}[\tau^*_j > \hat{\tau}])$.

2.11 Appendix E. Complete Results for the Low Educated

Table 2.8: Complete Estimations on Continuous Outcomes for the Low Educated

	second stage on continuous outcomes				first stage
Outcomes‡	log earnings		log hours worked		early non-empl.
	OLS	2SLS	OLS	2SLS	OLS
clustered standard errors:	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$
	(1)	(2)	(3)	(4)	(5)
UR_grad					5.4615***
					(1.6848)
early non-empl.	-0.0269***	-0.1002***	-0.0203***	-0.0722***	
	(0.0040)	(0.0291)	(0.0029)	(0.0207)	
UR_pe6	0.4822**	0.4172**	0.3340**	0.2879**	0.3779
	(0.1904)	(0.1919)	(0.1382)	(0.1391)	(1.6071)
lin_grad_year	-0.1588	0.0469	-0.1140	0.0317	3.5329
	(0.2945)	(0.2995)	(0.2133)	(0.2190)	(2.2946)
lin_grad_year|trend>3	0.9033**	0.6407	0.6680**	0.4821	-4.4281
	(0.4002)	(0.4242)	(0.2914)	(0.3108)	(2.8479)
lin_grad_year|trend>6	-0.4604*	-0.4394	-0.3337*	-0.3188	3.9169
	(0.2582)	(0.2852)	(0.1866)	(0.2035)	(3.1378)
d_province2	-1.3812*	-1.2603	-0.9473*	-0.8617	3.6574
	(0.7107)	(0.9544)	(0.5039)	(0.6678)	(8.6769)
d_province3	-2.1752***	-2.9225***	-1.5767***	-2.1058***	-11.7452
	(0.6066)	(0.9315)	(0.4343)	(0.6701)	(8.9621)
d_province4	0.0841	0.1825	0.0743	0.1439	0.7495
	(0.4801)	(0.4983)	(0.3503)	(0.3567)	(4.7080)
d_province5	0.8991	0.9211	0.6684	0.6840	-15.1893**
	(0.6099)	(0.6721)	(0.4494)	(0.4864)	(6.6042)
lin_calend_year_prov2	0.0498	0.0410	0.0349	0.0287	0.0774
	(0.1386)	(0.1413)	(0.1018)	(0.1024)	(0.9931)
lin_calend_year_prov3	0.0723	0.1638	0.0587	0.1235	2.2624
	(0.1136)	(0.1441)	(0.0810)	(0.1039)	(1.7325)
lin_calend_year_prov4	-0.0946	-0.1524	-0.0685	-0.1094	-1.0198
	(0.1267)	(0.1477)	(0.0927)	(0.1051)	(1.2618)
lin_calend_year_prov5	-0.0813	-0.0760	-0.0595	-0.0557	1.8419
	(0.1067)	(0.1194)	(0.0795)	(0.0868)	(1.1761)
avg_UR_pe3-6	-1.2836***	-1.4348**	-0.9654***	-1.0725***	-0.3037
	(0.4509)	(0.5863)	(0.3261)	(0.4136)	(4.4775)
min_UR_pe0-6	-0.5718	-0.4848	-0.3121	-0.2505	-6.4742
	(0.6219)	(0.6613)	(0.4388)	(0.4659)	(7.1097)
birth cohort76	0.7043	-0.3299	0.5424	-0.1897	-13.2641***
	(0.5632)	(0.7686)	(0.4125)	(0.5581)	(3.6802)
birth cohort78	0.4261	-0.0504	0.3190	-0.0183	-5.9156**
	(0.3677)	(0.4687)	(0.2690)	(0.3408)	(2.5815)
live in single-parent	0.3453	0.7897	0.2736	0.5882	6.1392
	(0.4632)	(0.5207)	(0.3466)	(0.3838)	(4.1354)
not live with parents	0.4272*	0.5203	0.3483*	0.4142	1.0913
	(0.2431)	(0.3501)	(0.1792)	(0.2529)	(2.5611)
HH members aged 0-11	-0.0178	0.0845	-0.0478	0.0246	1.3007
	(0.1123)	(0.1472)	(0.0882)	(0.1069)	(1.1570)
HH members aged 12-17	0.1680	0.1467	0.1262	0.1111	-0.2429
	(0.1177)	(0.1473)	(0.0884)	(0.1087)	(0.9030)
HH members aged 18-29	0.0112	0.2054	0.0074	0.1449	2.6587**

Continued on next page

Table 2.8 – continued from previous page

Outcomes‡	log earnings		log hours worked		early non-empl.
	OLS	2SLS	OLS	2SLS	OLS
	(1)	(2)	(3)	(4)	(5)
	(0.1164)	(0.1561)	(0.0869)	(0.1133)	(1.0291)
HH members aged 30-64	-0.0333	-0.0800	-0.0109	-0.0439	-0.4836
	(0.4065)	(0.4287)	(0.3021)	(0.3128)	(3.2574)
HH members aged 65+	-0.0680	0.0308	-0.0288	0.0412	1.2468
	(0.3639)	(0.4196)	(0.2683)	(0.3078)	(3.1042)
father education	0.0011	0.0294	0.0000	0.0200	0.3816
	(0.0255)	(0.0330)	(0.0188)	(0.0240)	(0.2531)
mother education	-0.1053***	-0.0485	-0.0766***	-0.0364	0.7800**
	(0.0356)	(0.0432)	(0.0270)	(0.0317)	(0.2925)
years of delay in sec.edu	-0.0545	0.3485*	-0.0424	0.2429	5.4123***
	(0.1112)	(0.2094)	(0.0808)	(0.1509)	(1.1119)
technical edu	0.4431	-0.4556	0.3576	-0.2787	-11.8663***
	(0.3274)	(0.4733)	(0.2473)	(0.3399)	(2.9533)
vocational edu	0.4572	-0.3519	0.3683*	-0.2045	-10.6291***
	(0.2772)	(0.4516)	(0.2105)	(0.3232)	(3.3428)
apprenticeship/PT voc	-0.1879	-1.1339*	-0.1077	-0.7774*	-12.5300**
	(0.4530)	(0.6234)	(0.3375)	(0.4495)	(4.7050)
Constant	15.0563***	18.1333***	10.7753***	12.9534***	24.5592
	(3.7528)	(5.2570)	(2.6939)	(3.7253)	(49.3992)
Observations	1,902	1,902	1,902	1,902	1,902
R-squared	0.0895	-0.2401	0.0895	-0.2090	0.1070
F stat of first step§					10.51
Exogeneity test P-val†		0.00970		0.0113	

*** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. Columns 1-4 report the results from estimating (2.2) by OLS (odds columns) and 2SLS (even columns). Column 5 reports OLS results from estimating the first stage. All estimations report cluster robust standard errors by graduation year g and province of residence at graduation p ($G = 44$). Column 5 reports the F statistic of the first stage and even columns report the exogeneity test for early non-employment ($y^0_{igpt_1}$).

‡ Continuous outcomes are measured at potential experience 6; early non-emp-loyment is measured in the first 2.5 years after gradaution. For continuous outcomes we add value one before taking the logarithm, so that those who are not salaried employed at the moment of evaluation are included with outcomes=0 after the logarithmic transformation.

§ This statistic is not corrected for the problem of few clusters. The corrected value resulting from the bootstrap procedure is 9.25 (see table 2.1).

† With clustered standard errors, the exogeneity test is defined as the difference of two Sargan-Hansen statistics: one for the equation where y^0_{t1} is treated as endogenous and one for the equation where $y^0_{igpt_1}$ is treated as exogenous. Under the null that $y^0_{igpt_1}$ is exogenous, the statistic is distributed as $\chi^2(1)$. This statistic is not corrected for the problem of few clusters.

Table 2.9: Complete Estimations on Discrete Outcomes for the Low Educated

	second stage					
Outcomes:‡	salaried empl.		self-empl.		overall empl.	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
clustered standard errors:	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$
	(1)	(2)	(3)	(4)	(5)	(6)
early non-empl.	-0.00169***	-0.00256	0.00054	0.00248	-0.00115***	-0.00008
	(0.00041)	(0.00290)	(0.00041)	(0.00258)	(0.00025)	(0.00151)
UR_pe6	0.00481	0.00404	0.01081	0.01253	0.01562	0.01657*
	(0.02385)	(0.02209)	(0.01947)	(0.01765)	(0.01020)	(0.00999)
lin_grad_year	-0.00253	-0.00009	-0.01367	-0.01911	-0.01621	-0.01920
	(0.03120)	(0.03028)	(0.02544)	(0.02343)	(0.01397)	(0.01340)
lin_grad_year|trend>3	0.02993	0.02680	0.00338	0.01032	0.03331*	0.03712**
	(0.04020)	(0.04043)	(0.03029)	(0.02903)	(0.01832)	(0.01813)
lin_grad_year|trend>6	-0.01720	-0.01695	0.01884	0.01829	0.00164	0.00134
	(0.02568)	(0.02477)	(0.02503)	(0.02252)	(0.01457)	(0.01375)
d_province2	-0.22288**	-0.22144**	0.17666*	0.17347**	-0.04621	-0.04797
	(0.09844)	(0.09548)	(0.09049)	(0.08719)	(0.04174)	(0.03996)
d_province3	-0.17150***	-0.18038***	0.13845**	0.15820**	-0.03304	-0.02218
	(0.06193)	(0.06407)	(0.06588)	(0.06577)	(0.03720)	(0.04285)
d_province4	-0.11409**	-0.11292**	0.10736**	0.10477**	-0.00673	-0.00816
	(0.04892)	(0.04910)	(0.04459)	(0.04794)	(0.02450)	(0.02611)
d_province5	0.04132	0.04158	-0.06171	-0.06229	-0.02039	-0.02071
	(0.06183)	(0.06032)	(0.04557)	(0.04483)	(0.03418)	(0.03449)
lin_calend_year_prov2	0.01976	0.01966	-0.01544	-0.01521	0.00432	0.00445
	(0.01597)	(0.01524)	(0.01253)	(0.01214)	(0.00565)	(0.00651)
lin_calend_year_prov3	-0.00496	-0.00387	0.00501	0.00259	0.00005	-0.00128
	(0.01207)	(0.01140)	(0.01565)	(0.01376)	(0.00716)	(0.00690)
lin_calend_year_prov4	0.01299	0.01231	-0.01369	-0.01217	-0.00070	0.00014
	(0.01118)	(0.01157)	(0.00964)	(0.01055)	(0.00642)	(0.00679)
lin_calend_year_prov5	-0.00608	-0.00602	0.01656	0.01642*	0.01048**	0.01041**
	(0.01074)	(0.01027)	(0.01012)	(0.00967)	(0.00471)	(0.00491)
avg_UR_pe3-6	-0.08309**	-0.08489**	0.03333	0.03733	-0.04976*	-0.04756*
	(0.04062)	(0.04005)	(0.03418)	(0.03230)	(0.02622)	(0.02516)
min_UR_pe0-6	-0.01193	-0.01089	0.00452	0.00222	-0.00741	-0.00867
	(0.06247)	(0.05970)	(0.05849)	(0.05672)	(0.02769)	(0.02854)
birth cohort76	0.00488	-0.00741	0.01471	0.04204	0.01959	0.03462
	(0.04397)	(0.06402)	(0.03588)	(0.05434)	(0.02535)	(0.03517)
birth cohort78	0.02309	0.01743	-0.00086	0.01173	0.02223	0.02916
	(0.03666)	(0.04407)	(0.03059)	(0.03645)	(0.01745)	(0.02152)
live in single-parent	-0.00309	0.00220	-0.03561	-0.04736	-0.03870	-0.04516
	(0.05525)	(0.05471)	(0.05233)	(0.05472)	(0.03217)	(0.03371)
not live with parents	0.02521	0.02631	-0.01502	-0.01748	0.01019	0.00883
	(0.02900)	(0.02995)	(0.02753)	(0.03031)	(0.01581)	(0.01592)
HH members aged 0-11	-0.00246	-0.00124	0.00596	0.00325	0.00350	0.00201
	(0.01217)	(0.01251)	(0.01277)	(0.01265)	(0.00574)	(0.00627)
HH members aged 12-17	0.02291*	0.02265*	-0.01252	-0.01196	0.01039	0.01070
	(0.01185)	(0.01173)	(0.01041)	(0.01069)	(0.00709)	(0.00683)
HH members aged 18-29	0.00009	0.00240	0.00374	-0.00139	0.00384	0.00101
	(0.01152)	(0.01355)	(0.01034)	(0.01150)	(0.00604)	(0.00688)
HH members aged 30-64	-0.01697	-0.01752	-0.02276	-0.02152	-0.03972	-0.03904
	(0.05212)	(0.05087)	(0.04874)	(0.04896)	(0.02847)	(0.02961)
HH members aged 65+	-0.02001	-0.01884	0.00730	0.00469	-0.01271	-0.01414
	(0.04165)	(0.04136)	(0.03573)	(0.03540)	(0.02182)	(0.02205)

Continued on next page

Table 2.9 – continued from previous page

	salaried empl.		self-empl.		overall empl.	
	(1)	(2)	(3)	(4)	(5)	(6)
father education	-0.00132	-0.00099	0.00045	-0.00030	-0.00087	-0.00128
	(0.00198)	(0.00226)	(0.00202)	(0.00213)	(0.00160)	(0.00170)
mother education	-0.00971***	-0.00904**	0.00937***	0.00787**	-0.00034	-0.00116
	(0.00341)	(0.00417)	(0.00294)	(0.00327)	(0.00181)	(0.00210)
years of delay in sec.edu	0.00440	0.00919	-0.02391**	-0.03456*	-0.01952***	-0.02537*
	(0.01275)	(0.02067)	(0.01070)	(0.01764)	(0.00702)	(0.01381)
technical edu	0.01038	-0.00031	0.02957	0.05333	0.03995***	0.05301**
	(0.03694)	(0.04459)	(0.03180)	(0.04181)	(0.01452)	(0.02678)
vocational edu	0.00818	-0.00144	0.03776	0.05914	0.04594***	0.05770**
	(0.03164)	(0.04096)	(0.02700)	(0.04065)	(0.01546)	(0.02569)
apprenticeship/PT voc	-0.06240	-0.07366	0.08533*	0.11033**	0.02292	0.03667
	(0.05336)	(0.06364)	(0.04311)	(0.05434)	(0.02842)	(0.03967)
Constant	1.41749***	1.45409***	-0.15519	-0.23651	1.26230***	1.21757***
	(0.35242)	(0.37325)	(0.32471)	(0.34533)	(0.20910)	(0.22952)
Observations	1,902	1,902	1,902	1,902	1,902	1,902
R-squared	0.04170	0.03730	0.03098	0.00332	0.05818	0.03540
Exogeneity test P-val†		0.767		0.438		0.467

*** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. Columns 1-6 report the results from estimating (2.2) by OLS (odds columns) and 2SLS (even columns). The first stage regression is reported in Table 2.8 (Column 5). All estimations report cluster robust standard errors by graduation year g and province of residence at graduation p ($G = 44$). Even columns report the exogeneity test for early non-employment ($y^0_{igpt_1}$).

‡ The discrete outcomes are measured at potential experience 6.

† With clustered standard errors, the exogeneity test is defined as the difference of two Sargan-Hansen statistics: one for the equation where y^0_{t1} is treated as endogenous and one for the equation where $y^0_{igpt_1}$ is treated as exogenous. Under the null that $y^0_{igpt_1}$ is exogenous, the statistic is distributed as $\chi^2(1)$. This statistic is not corrected for the problem of few clusters.

2.12 Appendix F. Sensitivity Analysis For the Low Educated

Table 2.10: Effect of Interest for the Low Educated: Graduation Period 1998-2002.

Panel A: Effect of early non-employment in the structural equation:					
		OLS		2SLS	
	standard errors†	robust	cluster $g*p$	robust	cluster $g*p$
Continuous outcomes:‡‡		(1)	(2)	(3)	(4)
log earnings	coeff	-0.0287***	-0.0287***	-0.1406**	-0.1406**
	se	(0.0045)	(0.0049)	(0.0666)	(0.0597)
	P-val§		6.09E-06		0.0273
	Bootstrap P-val‡		0		0.0821
	Exogeneity test P-val§§				0.0306
log hours worked	coeff	-0.0215***	-0.0215***	-0.1033**	-0.1033**
	se	(0.0034)	(0.0035)	(0.0492)	(0.0439)
	P-val		3.19E-06		0.0277
	Bootstrap P-val		0		0.0781
	Exogeneity test P-val				0.0318

Panel B: Effect of the instrument in the first stage : OLS			
outcome:	*standard errors:*	robust	cluster (g*p)
early non-empl.(% hours)	coeff	11.9484***	11.9484***
	se	(3.4994)	(3.4918)
	P-val	0.00233	
	Bootstrap P-val	0.07007	
	F stat	11.70921	
	Bootstrap F stat††	3.60923	

*** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. Panel A reports results from estimating β in (2.2) on continuous outcomes measured at potential experience 6. β is the effect of one pp increase in $y^0_{igpt_1}$, i.e. the % of hours spent in non-employment in the first 2.5 years after graduation relative to potential total hours if one would work full-time during the whole period. For clustered standard errors, we report the P-value and the wild bootstrap P-value. Column 1-2 (3-4) show OLS (2SLS). In 2SLS the provincial unemployment rate at graduation is used as instrument for $y^0_{igpt_1}$. Panel B shows the effect of the instrument on $y^0_{igpt_1}$ in the first stage and the corresponding F statistic.

‡‡ For continuous outcomes we add value one before taking the log, so that non-salaried employed at the moment of evaluation are included with outcomes=0 after the logarithmic transformation.

† Robust indicates heteroscedastic-robust standard errors; clusters are defined by year g and province of residence at graduation p (G=24 clusters). § The P-value from clustered standard errors is computed using the $t(G-1)$ distribution, with G=24 (stars are reported accordingly).

‡ Bootstrap P-values are computed according to the wild bootstrap procedures explained in Appendix 2.10 for 999 repetitions.

†† Bootstrap F statistic is the F statistic corresponding to the bootstrap P-value of the t statistic of the instrument: we rely on the equivalence between F and t distribution: $t^2(G-1) = F(1, G-1)$, with $G = 24$.

§§ With clustered standard errors, this test is defined as the difference of two Sargan-Hansen statistics: one for the equation where y^0_{t1} is treated as endogenous and one for the equation where $y^0_{igpt_1}$ is treated as exogenous. Under the null that $y^0_{igpt_1}$ is exogenous, the statistic is distributed as $\chi^2(1)$.

Table 2.11: Complete OLS Estimations for the Low Educated: Graduation Period 1994-2002 vs 1998-2002; Full vs Restricted Specification (excluding some individual controls).

Continuous outcomes‡‡:	log earnings			log hours worked		
	g94-02§		g98-02†	g94-02		g98-02
	full spec.§§	restricted spec.††		full spec.	restricted spec.	
	(1)	(2)	(3)	(4)	(5)	(6)
cluster	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$
early non-empl	-0.027***	-0.027***	-0.029***	-0.020***	-0.020***	-0.022***
	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)
UR_pe6	0.482**	0.617***	0.594	0.334**	0.432***	0.446
	(0.190)	(0.179)	(0.544)	(0.138)	(0.130)	(0.393)
lin_grad_year	-0.159	-0.114	0.812*	-0.114	-0.082	0.636*
	(0.294)	(0.284)	(0.460)	(0.213)	(0.205)	(0.333)
lin_grad_year|trend>3	0.903**	0.661*	-1.032	0.668**	0.492*	-0.797
	(0.400)	(0.360)	(0.701)	(0.291)	(0.260)	(0.503)
lin_grad_year|trend>6	-0.460*			-0.334*		
	(0.258)			(0.187)		
avg_UR_pe3-6	-1.284***	-0.826**	-1.879	-0.965***	-0.634**	-1.485
	(0.451)	(0.394)	(1.373)	(0.326)	(0.289)	(0.993)
min_UR_pe0-6	-0.572	-0.692	-3.139***	-0.312	-0.398	-2.203**
	(0.622)	(0.590)	(1.118)	(0.439)	(0.419)	(0.791)
d_province2	-1.381*	-0.913	-4.627***	-0.947*	-0.611	-3.214**
	(0.711)	(0.704)	(1.597)	(0.504)	(0.503)	(1.169)
d_province3	-2.175***	-1.756***	-4.127**	-1.577***	-1.275***	-3.026**
	(0.607)	(0.547)	(1.740)	(0.434)	(0.392)	(1.238)
d_province4	0.084	-0.109	-1.617	0.074	-0.069	-1.116
	(0.480)	(0.456)	(1.380)	(0.350)	(0.333)	(0.984)
d_province5	0.899	0.741	1.125	0.668	0.553	0.741
	(0.610)	(0.609)	(1.902)	(0.449)	(0.451)	(1.397)
lin_calend_year_prov2	0.050	0.075	0.126	0.035	0.055	0.064
	(0.139)	(0.144)	(0.307)	(0.102)	(0.106)	(0.228)
lin_calend_year_prov3	0.072	0.147	-0.135	0.059	0.114	-0.106
	(0.114)	(0.104)	(0.137)	(0.081)	(0.075)	(0.094)
lin_calend_year_prov4	-0.095	-0.007	-0.120	-0.069	-0.004	-0.102
	(0.127)	(0.109)	(0.304)	(0.093)	(0.080)	(0.219)
lin_calend_year_prov5	-0.081	-0.102	-0.078	-0.059	-0.075	-0.041
	(0.107)	(0.107)	(0.357)	(0.079)	(0.080)	(0.263)
birth cohort76	0.704	0.813	0.816	0.542	0.622	0.625
	(0.563)	(0.510)	(0.599)	(0.412)	(0.373)	(0.437)
birth cohort78	0.426	0.487	0.471	0.319	0.364	0.353
	(0.368)	(0.350)	(0.353)	(0.269)	(0.255)	(0.256)
HH members aged 0-11	-0.018	0.016	0.005	-0.048	-0.022	-0.028
	(0.112)	(0.116)	(0.136)	(0.088)	(0.091)	(0.103)
father education	0.001	-0.002	-0.017	0.000	-0.003	-0.014
	(0.026)	(0.025)	(0.041)	(0.019)	(0.019)	(0.031)
mother education	-0.105***	-0.100***	-0.103**	-0.077***	-0.073***	-0.074**
	(0.036)	(0.035)	(0.045)	(0.027)	(0.026)	(0.033)
years of delay in sec.edu	-0.055	-0.143	-0.190	-0.042	-0.106	-0.140
	(0.111)	(0.096)	(0.166)	(0.081)	(0.071)	(0.120)
general edu		-0.449	-0.161		-0.362	-0.145
		(0.299)	(0.353)		(0.227)	(0.266)
live in single-parent	0.345			0.274		
	(0.463)			(0.347)		

Continued on next page

Table 2.11 – continued from previous page

Continuous outcomes‡‡:	log earnings			log hours worked		
	g94-02§		g98-02†	g94-02		g98-02
	full spec.§§	restricted spec.††		full spec.	restricted spec.	
	(1)	(2)	(3)	(4)	(5)	(6)
not live with parents	0.427*			0.348*		
	(0.243)			(0.179)		
HH members aged 12-17	0.168			0.126		
	(0.118)			(0.088)		
HH members aged 18-29	0.011			0.007		
	(0.116)			(0.087)		
HH members aged 30-64	-0.033			-0.011		
	(0.406)			(0.302)		
HH members aged 65+	-0.068			-0.029		
	(0.364)			(0.268)		
technical edu	0.443			0.358		
	(0.327)			(0.247)		
vocational edu	0.457			0.368*		
	(0.277)			(0.210)		
apprenticeship/PT voc	-0.188			-0.108		
	(0.453)			(0.337)		
Constant	15.056***	12.907***	29.988***	10.775***	9.279***	22.077***
	(3.753)	(3.432)	(7.716)	(2.694)	(2.453)	(5.519)
Observations	1,902	1,902	946	1,902	1,902	946
R-squared	0.090	0.084	0.097	0.090	0.084	0.098

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors between parentheses. Columns 1 and 4 estimate (2.2) by OLS considering the graduation period 1994-2002: they are equivalent to the estimations reported in columns 1 and 3 of Table 2.8. Columns 2 and 5 estimate the restricted specification discussed for the sensitivity exercise in Section 2.5.1, based on the graduation period 1994-2002. Columns 3 and 6 estimate the same restricted specification on the graduation period 1998-2002, which is used in the second sensitivity analysis. Standard errors are clustered by graduation year g and province of living at graduation p.
§ graduation period 1994-2002 considered.
† graduation period 1998-2002 considered: for this reason, the third graduation year spline $lin\_grad\_year|trend > 6$ is omitted.
‡‡ For continuous outcomes we add value one before taking the log, so that non-salaried employed at the moment of evaluation are included with outcomes=0 after the logarithmic transformation.
§§ The full specification corresponds to equation (2.2).
†† The restricted specification is the one used in the second sensitivity analysis in Section 2.5.1.

Table 2.12: Effect of Interest for the Low Educated Excluding $\overline{UR}_p$ and $minUR_{pt}$ from the Specification.

Panel A: Effect of early non-employment in the structural equation:					
		OLS		2SLS	
	standard errors†	robust	cluster $g*p$	robust	cluster $g*p$
outcomes:		(1)	(2)	(3)	(4)
salaried employment	coeff	-0.00169***	-0.00169***	-0.00199	-0.00199
	se	(0.00034)	(0.00041)	(0.00387)	(0.00348)
	P-val§		0.00019		0.57020
	Bootstrap P-val‡		0.00000		0.62462
	Exogeneity test P-val§§			0.937	0.931
self-employment	coeff	0.00054*	0.00054	0.00219	0.00219
	se	(0.00030)	(0.00041)	(0.00340)	(0.00313)
	P-val		0.19253		0.48668
	Bootstrap P-val		0.19219		0.52853
	Exogeneity test P-val		0.619		0.587
overall employment	coeff	-0.00115***	-0.00115***	0.00020	0.00020
	se	(0.00021)	(0.00025)	(0.00229)	(0.00164)
	P-val		0.00005		0.90232
	Bootstrap P-val		0.00000		0.91291
	Exogeneity test P-val			0.540	0.386
log earnings	coeff	-0.0269***	-0.0269***	-0.0947**	-0.0947***
	se	(0.0033)	(0.0040)	(0.0447)	(0.0354)
	P-val		2.78E-08		0.0105
	Bootstrap P-val		0.0000		0.0300
	Exogeneity test P-val			0.0796	0.0361
log hours worked	coeff	-0.0203***	-0.0203***	-0.0666**	-0.0666***
	se	(0.0024)	(0.0029)	(0.0326)	(0.0251)
	P-val		1.00E-08		0.0111
	Bootstrap P-val		0.0000		0.0340
	Exogeneity test P-val			0.108	0.0483
Panel B: Effect of the instrument from the first stage (OLS)					
outcome‡‡*:*	*standard errors:*	robust	cluster (g*p)		
early non-empl.	coeff	5.0319***	5.0319***		
	se	(1.6519)	(1.7139)		
	P-val		0.0053		
	Bootstrap P-val		0.0120		
	F stat	9.279	8.620		
	Bootstrap F stat††		5.84		

Continued on next page

*** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. Panel A reports results from estimating β in (2.2) on outcomes measured at potential experience 6, excluding $\overline{UR}_p$ and $minUR_{pt}$. β is the effect of one pp increase in $y^0_{igpt_1}$, i.e. the % of hours spent in non-employment in the first 2.5 years after graduation relative to potential total hours if one would work full-time during the whole period. For clustered standard errors, we report the P-value and the wild bootstrap P-value. Column 1-2 (3-4) show OLS (2SLS). In 2SLS the provincial unemployment rate at graduation is used as instrument for $y^0_{igpt_1}$. Panel B reports the effect of the instrument on $y^0_{igpt_1}$ in the first stage and the F statistic. For continuous outcomes we add value one before taking the log, so that non-salaried employed at the moment of evaluation are included with outcomes=0 after the logarithmic transformation.

† Robust indicates heteroscedastic-robust standard errors; clusters are defined by year g and province of residence at graduation p (G=44 clusters).

§ The P-value from clustered standard errors is computed using the $t(G-1)$ distribution, with G=44 (stars are reported accordingly).

‡ Bootstrap P-values are computed according to the wild bootstrap procedures explained in Appendix 2.10 for 999 repetitions.

†† Bootstrap F statistic is the F statistic corresponding to the Bootstrap P-value of the t statistic of the instrument: we rely on the equivalence between F and t distribution: with $G = 44$, $t^2(G-1) = F(1, G-1)$.

§§ With clustered standard errors, this test is defined as the difference of two Sargan-Hansen statistics: one for the equation where y^0_{t1} is treated as endogenous and one for the equation where $y^0_{igpt_1}$ is treated as exogenous. Under the null that $y^0_{igpt_1}$ is exogenous, the statistic is distributed as $\chi^2(1)$.

2.13 Appendix G. Reduced Form Estimations for the Low Educated

Table 2.13: Complete Estimations of Reduced Form and First Stage for the Low Educated.

	Reduced Form (ITT effects)					1^{st} *stage*
Outcomes:§§	log earnings	log hours worked	salaried	self	overall empl.	early non-empl.
	(1)	(2)	(3)	(4)	(5)	(6)
clustered standard errors:	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$	$g*p$
UR_grad	-0.5475***	-0.3945***	-0.01400	0.01355	-0.00045	5.4615***
	(0.1662)	(0.1223)	(0.01675)	(0.01365)	(0.00851)	(1.6848)
P-val§	0.00198	0.00239	0.40799	0.32666	0.95777	0.00230
Bootstrap P-val‡	0.00801	0.00601	0.45646	0.36036	0.96496	0.00400
UR_pe6	0.3793*	0.2606*	0.00307	0.01347	0.01654	0.3779
	(0.1898)	(0.1390)	(0.02343)	(0.01867)	(0.01025)	(1.6071)
lin_grad_year	-0.3072	-0.2235	-0.00914	-0.01035	-0.01949	3.5329
	(0.3349)	(0.2428)	(0.03309)	(0.02628)	(0.01455)	(2.2946)
lin_grad_year|trend>3	1.0846**	0.8020**	0.03816	-0.00066	0.03749*	-4.4281
	(0.4259)	(0.3118)	(0.04150)	(0.02987)	(0.01878)	(2.8479)
lin_grad_year|trend>6	-0.8321***	-0.6017***	-0.02699	0.02800	0.00101	3.9169
	(0.2976)	(0.2193)	(0.02719)	(0.02393)	(0.01559)	(3.1378)
d_province2	-1.6270*	-1.1259*	-0.23081**	0.18254*	-0.04827	3.6574
	(0.8282)	(0.5902)	(0.10284)	(0.09202)	(0.04254)	(8.6769)
d_province3	-1.7451*	-1.2573*	-0.15028**	0.12907*	-0.02120	-11.7452
	(0.8831)	(0.6340)	(0.06874)	(0.06645)	(0.03691)	(8.9621)
d_province4	0.1073	0.0898	-0.11484**	0.10662**	-0.00822	0.7495
	(0.5737)	(0.4196)	(0.05039)	(0.04570)	(0.02701)	(4.7080)
d_province5	2.4438**	1.7812**	0.08051	-0.09996	-0.01945	-15.1893**
	(0.9123)	(0.6716)	(0.09174)	(0.07501)	(0.03869)	(6.6042)
lin_calend_year_prov2	0.0333	0.0231	0.01946	-0.01501	0.00444	0.0774
	(0.1463)	(0.1076)	(0.01658)	(0.01269)	(0.00669)	(0.9931)
lin_calend_year_prov3	-0.0630	-0.0400	-0.00967	0.00821	-0.00147	2.2624
	(0.1586)	(0.1134)	(0.01455)	(0.01646)	(0.00668)	(1.7325)
lin_calend_year_prov4	-0.0502	-0.0358	0.01492	-0.01469	0.00023	-1.0198
	(0.1322)	(0.0980)	(0.01172)	(0.00997)	(0.00685)	(1.2618)
lin_calend_year_prov5	-0.2607*	-0.1888*	-0.01074	0.02099	0.01025*	1.8419
	(0.1436)	(0.1064)	(0.01490)	(0.01350)	(0.00532)	(1.1761)
avg_UR_pe3-6	-1.4044***	-1.0505***	-0.08411**	0.03657	-0.04753*	-0.3037
	(0.4377)	(0.3218)	(0.04124)	(0.03380)	(0.02560)	(4.4775)
min_UR_pe0-6	0.1642	0.2171	0.00570	-0.01383	-0.00813	-6.4742
	(0.8643)	(0.6086)	(0.06138)	(0.05224)	(0.02860)	(7.1097)
birth cohort76	0.9998	0.7684*	0.02659	0.00914	0.03572	-13.2641***
	(0.6088)	(0.4473)	(0.04683)	(0.03551)	(0.02615)	(3.6802)
birth cohort78	0.5426	0.4090	0.03259	-0.00294	0.02965	-5.9156**
	(0.4035)	(0.2949)	(0.03871)	(0.03095)	(0.01872)	(2.5815)
live in single-parent	0.1743	0.1447	-0.01354	-0.03213	-0.04567	6.1392
	(0.4770)	(0.3578)	(0.05554)	(0.05226)	(0.03403)	(4.1354)
not live with parents	0.4109*	0.3354**	0.02352	-0.01477	0.00874	1.0913
	(0.2252)	(0.1661)	(0.02742)	(0.02702)	(0.01600)	(2.5611)
HH members aged 0-11	-0.0459	-0.0693	-0.00457	0.00648	0.00191	1.3007
	(0.1194)	(0.0950)	(0.01213)	(0.01265)	(0.00581)	(1.1570)
HH members aged 12-17	0.1711	0.1287	0.02328**	-0.01256	0.01072	-0.2429

Continued on next page

Table 2.13 – continued from previous page

	Reduced Form (ITT effects)					*1st stage*
Outcomes:§§	log earnings	log hours worked	salaried	self	overall empl.	early non-empl.
	(1)	(2)	(3)	(4)	(5)	(6)
	(0.1152)	(0.0866)	(0.01143)	(0.01027)	(0.00697)	(0.9030)
HH members aged 18-29	-0.0612	-0.0472	-0.00441	0.00521	0.00079	2.6587**
	(0.1209)	(0.0904)	(0.01253)	(0.01090)	(0.00622)	(1.0291)
HH members aged 30-64	-0.0316	-0.0090	-0.01628	-0.02272	-0.03900	-0.4836
	(0.4278)	(0.3192)	(0.05313)	(0.04852)	(0.03035)	(3.2574)
HH members aged 65+	-0.0942	-0.0489	-0.02203	0.00779	-0.01425	1.2468
	(0.3784)	(0.2784)	(0.04265)	(0.03599)	(0.02242)	(3.1042)
father education	-0.0089	-0.0075	-0.00196	0.00065	-0.00132	0.3816
	(0.0268)	(0.0198)	(0.00202)	(0.00203)	(0.00160)	(0.2531)
mother education	-0.1267***	-0.0927***	-0.01104***	0.00981***	-0.00123	0.7800**
	(0.0393)	(0.0297)	(0.00364)	(0.00299)	(0.00179)	(0.2925)
years of delay in sec.edu	-0.1941	-0.1481	-0.00468	-0.02114**	-0.02582***	5.4123***
	(0.1216)	(0.0888)	(0.01336)	(0.01037)	(0.00802)	(1.1119)
technical edu	0.7340**	0.5785**	0.03010	0.02389	0.05400***	-11.8663***
	(0.3559)	(0.2693)	(0.03912)	(0.03184)	(0.01708)	(2.9533)
vocational edu	0.7136**	0.5634**	0.02580	0.03278	0.05858***	-10.6291***
	(0.3132)	(0.2384)	(0.03306)	(0.02674)	(0.01792)	(3.3428)
apprenticeship/PT voc	0.1222	0.1277	-0.04154	0.07925*	0.03771	-12.5300**
	(0.4961)	(0.3700)	(0.05738)	(0.04296)	(0.03086)	(4.7050)
Constant	15.6713***	11.1794***	1.39113***	-0.17560	1.21553***	24.5592
	(4.5909)	(3.3431)	(0.38766)	(0.33394)	(0.22040)	(49.3992)
Observations	1,902	1,902	1,902	1,902	1,902	1,902
R-squared	0.0484	0.0469	0.02537	0.02905	0.03172	0.1070
F stat of first step						10.51
Bootstrap F stat†						9.25

*** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. Column 1-6 report the OLS result from estimating the reduced form equation of (2.2), i.e. where $y^0_{igpt_1}$ is replaced by the unemployment rate at graduation: the coefficients of the unemployment rate at graduation on the outcomes represent the Intention-To-Treat (ITT) effects. Column 7 reports the first stage regression, where y_{iT} is replaced by $y^0_{igpt_1}$ and $y^0_{igpt_1}$ is replaced by the unemployment rate at graduation. In all cases, for the coefficient of the unemployment rate at graduation we report the P-value and the wild bootstrap P-value.

§§ The outcomes of interest (Column 1-5) are measured at potential experience 6. The dependent variable in the first stage (Column 6) is measured in the first 2.5 years after graduation. For continuous outcomes (Column 1-2) we add value one before taking the log, so that non-salaried employed at the moment of evaluation are included with outcomes=0 after the logarithmic transformation.

§ The P-value from clustered standard errors is computed using the $t(G-1)$ distribution, with G=44 (stars are reported accordingly).

‡ Bootstrap P-values are computed according to the wild bootstrap procedures explained in Appendix 2.10 for 999 repetitions.

† Bootstrap F statistic is the F statistic corresponding to the Bootstrap P-value of the t statistic of the instrument: we rely on the equivalence between F and t distribution: $t^2(G-1) = F(1, G-1)$ with $G = 44$.

2.14 Appendix H. The Analysis for the High Educated

For the high educated we use the average hourly wages earned in the first 2.5 years since graduation as endogenous regressor, since the high educated are persistently damaged by adverse labor market conditions at graduation because they accept lower-paying jobs (see Chapter 1). Accordingly, early wages should be the relevant channel through which the instrument affects the outcomes of interest (exclusion restriction). In this case we have to restrict the sample to the graduation period 1998-2004, since we do not observe early wages for those graduating before 1998: this leaves us with a sample of 35 clusters (7 graduation years times 5 provinces). The size of the clusters is reported in Table 2.14.

As outcome of interest we use the log of wage at potential experience 6: therefore, the dependent variable is missing for those who were not salaried employed at the time of the evaluation, as the log-transformation is not defined (13% of the sample). The endogenous regressor instead is expressed in level; [60] similarly as for the dependent variable, it is missing for individuals who never earned a wage in the first 2.5 years since graduation (6.5% of the sample). Hence, the estimates should be interpreted as semi-elasticities, conditional on being salaried employed both 6 years after graduation and in the early period. Of course, conditional effects may be biased if the sub-sample of salaried employed is selected: we do not tackle this problem. Descriptive statistics of the individual controls and the endogenous variables are reported in Table 2.15. Note that, since high educated graduate in the period 1998-2004, the last graduation cohort is followed until potential experience 6. Later than that this sample gets smaller as the last graduation cohorts progressively drop out from the sample.

For the high educated, we study the long-term effect of early wages on subsequent wages: therefore, we expect a positive β and the persistence of early wages on subsequent labor market performances will be considered as evidence of scarring. In this case, we do not expect measurement error since we restrict the sample to the graduation period 1998-2004 and measure early wage only exploiting administrative data. Ability and returns to job search are positively correlated with both early wages and the outcomes of interest, thereby causing a positive bias. Liquidity constraints are instead negatively correlated with both early wages and the outcomes of interest, which also lead to a positive bias. Accordingly, in any case we expect OLS to overestimate β.

As for the low educated, also here we tackle the endogeneity problem of early wages with an IV approach where the unemployment rate at graduation is used as instrument for the endogenous regressor. In this case the exclusion restriction imposes that the long-term penalties of graduating in downturns are uniquely explained by the acceptance of lower-paying jobs early in the career. In this case, persistence would arise by the accumulation of human capital specific to lower-paying jobs or by the foregone human capital that one would have accumulated in a higher-paying job. Of course things can be a bit more blurry if we consider a wider definition of reservation wage which also includes non-pecuniary dimensions: then, high educated would be forced to accept lower quality jobs, such as temporary jobs or seasonal jobs, which may entail not only lower wages

[60] Transforming $y^0_{igpt_1}$ as $log(y^0_{igpt_1})$ yielded a lower F statistic. Hence, we decided to express $y^0_{igpt_1}$ in level.

but also unemployment spells upon termination of the contract. Under heterogenous effects, the monotonicity assumption requires that a higher unemployment rate at graduation makes every high educated person earning a higher (or lower) early wage.

In table 2.16, the bootstrap F statistic of the first stage is 2 which warns against the problem of weak instrument. For one *pp* increase in the unemployment rate at graduation, the average wage in the first 2.5 years since graduation decreases by 0.22 Euros. In case of weak instrument, the 2SLS is biased towards the OLS. Both OLS and 2SLS estimates show the expected sign: a higher early wage is associated with better labor market outcomes 6 years after graduation. The estimate is highly significant for OLS, which in principle could be due to endogeneity or causality. Unfortunately, in this case our IV approach is not effective to identify the latter.

Table 2.14: Number of Individuals by graduation year and province of residence at graduation.

	High educated					
grad_year	prov1	prov2	prov3	prov4	prov5	Total
1998	49	29	50	44	17	189
1999	55	54	61	63	37	270
2000	74	66	42	63	31	276
2001	61	39	44	48	33	225
2002	88	53	52	52	32	277
2003	53	35	24	32	26	170
2004	31	19	16	24	13	103
Total	457	319	327	361	220	1684

The analysis considers the graduation period 1998-2004 for the high educated. Provinces are in the following order from 1 to 5: Antwerp, Flemish Brabant, Western Flanders, Eastern Flanders, Limburg. Each combination of graduation year and province of residence at graduation represents a cluster *gp* in the main analysis.

Table 2.15: Descriptive Statistics for the High Educated

			Individual Control Variables:			
Variable	Obs	Mean	Std. Dev.	Min	Max	label
birth cohort76	1684	0.357	0.479	0	1	1 if born in 1976
birth cohort78	1684	0.328	0.470	0	1	1 if born in 1978
birth cohort80	1684	0.315	0.465	0	1	1 if born in 1980
live in single-parent	1684	0.078	0.268	0	1	1 if live with single parent at age17(Dec)
not live with parents	1684	0.027	0.161	0	1	1 if not live with either parents at age17(Dec)
HH members aged 0-11	1684	0.156	0.427	0	3	nr of other HH members aged0-11 at age17(Dec)
HH members aged 12-17	1684	0.573	0.683	0	4	nr of other HH members aged12-17 at age17(Dec)
HH members aged 18-29	1684	0.582	0.715	0	4	nr of other HH members aged18-29 at age17(Dec)
HH members aged 30-64	1684	1.929	0.298	1	4	nr of other HH members aged30-64 at age17(Dec)
HH members aged 65+	1684	0.030	0.188	0	2	nr of other HH members aged65+ at age17(Dec)
father education	1684	6.935	3.273	0	13	father completed education since age12
mother education	1684	6.322	2.947	0	13	mother completed education since age12
years of delay in sec.edu	1684	0.260	0.539	-1	3	years of delay at age17(Aug)
general education	1684	0.633	0.482	0	1	1 if general edu at age17(Aug)
technical education	1684	0.357	0.479	0	1	1 if technical edu at age17(Aug)
vocational education	1684	0.009	0.094	0	1	1 if vocational edu at age17(Aug)
apprenticeship/PT voc.	1684	0.001	0.024	0	1	1 if apprenticeship/PT voc. edu at age17(Aug)
			Endogenous Variables§:			
log hourly wage	1280	2.852	0.234	2.174	3.493	log hourly wage in salaried empl.
avg_early_wage	1280	13.566	2.560	8.014	31.375	average hourly wage in salaried empl.

Descriptive statistics are reported for the high educated graduating in the period 1998-2004.
§ "Log hourly wage" is measured at potential experience 6, whereas "avg_early_wage" is measured in the first 2.5 years since graduation. Both endogenous variables are conditional on being salaried employed at potential experience 6 and in the first 2.5 years after graduation. Hourly wages are computed as annual earnings divided by the annual hours worked in salaried employed. The outcome of interest ("Log hourly wage") is expressed in logarithm whereas the endogenous variable ("avg_early_wage") is expressed in level.

Table 2.16: Complete Estimations on Hourly Wages for the High Educated

outcomes†	second stage log hourly wage		first stage avg_early_wage
	OLS	2SLS	OLS
clustered se	$g * p$	$g * p$	$g * p$
	(1)	(2)	(3)
UR_grad			-0.2239*
			(0.1131)
avg_early_wage	0.0397***	0.0626	
	(0.0040)	(0.0430)	
P-val‡	1.60E-11	0.1547	
Bootstrap P-val§	0	0.3784	
UR_pe6	-0.0096	-0.0061	-0.1125
	(0.0106)	(0.0105)	(0.1264)
lin_grad_year	0.0262	0.0200	0.1559
	(0.0166)	(0.0216)	(0.1757)
lin_grad_year|trend>3	0.0054	-0.0025	0.5491**
	(0.0170)	(0.0220)	(0.2493)
d_province2	-0.0244	0.0135	-1.8109**
	(0.0556)	(0.0835)	(0.7469)
d_province3	0.1924**	0.1816**	0.3435
	(0.0890)	(0.0902)	(0.6809)
d_province4	0.0388	0.0944	-2.1026***
	(0.0782)	(0.1182)	(0.7124)
d_province5	0.0980	0.1026	0.0096
	(0.0982)	(0.0924)	(0.4239)
lin_calend_year_prov2	-0.0069	-0.0091	0.0963
	(0.0088)	(0.0089)	(0.0853)
lin_calend_year_prov3	-0.0387**	-0.0335*	-0.2369*
	(0.0162)	(0.0192)	(0.1293)
lin_calend_year_prov4	-0.0127	-0.0187	0.2210**
	(0.0112)	(0.0143)	(0.0926)
lin_calend_year_prov5	-0.0113	-0.0106	-0.0520
	(0.0129)	(0.0118)	(0.0559)
avg_UR_pe3-6	-0.0124	-0.0242	0.4844
	(0.0306)	(0.0357)	(0.3852)
min_UR_pe0-6	-0.0570	-0.0285	-1.1750**
	(0.0547)	(0.0793)	(0.4599)
birth cohort76	0.1144***	0.0790	1.5638***
	(0.0301)	(0.0724)	(0.3043)
birth cohort78	0.0527***	0.0313	0.9167***
	(0.0154)	(0.0409)	(0.1876)
live in single-parent	0.0091	0.0241	-0.6719
	(0.0527)	(0.0537)	(0.5381)
not live with parents	0.0083	0.0175	-0.4004
	(0.0417)	(0.0458)	(0.2521)
HH members aged 0-11	-0.0031	-0.0032	0.0091
	(0.0154)	(0.0151)	(0.1056)
HH members aged 12-17	0.0030	-0.0003	0.1442
	(0.0096)	(0.0110)	(0.1028)
HH members aged 18-29	0.0080	0.0081	0.0048
	(0.0083)	(0.0084)	(0.1017)
HH members aged 30-64	0.0014	0.0150	-0.6153
	(0.0397)	(0.0392)	(0.5187)

Continued on next page

Table 2.16 – continued from previous page

	second stage log hourly wage		first stage avg_early_wage
	OLS	2SLS	OLS
	(1)	(2)	(3)
HH members aged 65+	0.0213	0.0257	-0.2119
	(0.0288)	(0.0293)	(0.2499)
father education	0.0060**	0.0055*	0.0231
	(0.0024)	(0.0030)	(0.0234)
mother education	-0.0001	-0.0002	0.0014
	(0.0028)	(0.0028)	(0.0279)
years of delay in sec.edu	-0.0412***	-0.0241	-0.7543***
	(0.0133)	(0.0332)	(0.1078)
technical edu	-0.0320**	-0.0252*	-0.2940
	(0.0131)	(0.0145)	(0.1808)
apprenticeship/PT voc	-0.1069**	-0.1115**	0.2022
	(0.0511)	(0.0494)	(0.5282)
Constant	2.5052***	2.1346***	17.3247***
	(0.2562)	(0.7360)	(3.0029)
Observations	1,280	1,280	1,280
R-squared	0.2979	0.2433	0.1244
F stat of first step			3.920
Bootstrap F stat of first step§§			2.0824
Exogeneity test P-val††		0.614	

*** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. Columns 1 and 2 report the results from estimating (2.2) by OLS and 2SLS, respectively; the outcome is measured at potential experience 6. Column 3 reports the first stage regression, where the dependent variable is measured in the first 2.5 years since graduation. All estimations report cluster robust standard errors by graduation year g and province of residence at graduation p ($G = 35$). The sample is restricted to individuals who are salaried employed 6 years after graduation and in the first 2.5 years after graduation. Hence, the estimation reports conditional effects.

† Hourly wages are computed as annual earnings divided by the annual hours worked in salaried employed. The outcome of interest is expressed in logarithm whereas the endogenous variable is expressed in level.

‡ The P-value from clustered standard errors is computed using the $t(G-1)$ distribution, with G=35 (stars of the corresponding coefficient are reported accordingly).

§ Bootstrap P-values are computed according to the wild bootstrap procedures explained in Appendix 2.10 for 999 repetitions.

§§ Bootstrap F statistic is the F statistic corresponding to the Bootstrap P-value of the t statistic of the instrument: we rely on the equivalence between F and t distribution: $t^2(G-1) = F(1, G-1)$, with $G = 35$.

†† With clustered standard errors, the exogeneity test is defined as the difference of two Sargan-Hansen statistics: one for the equation where y^0_{t1} is treated as endogenous, and one for the equation where $y^0_{igpt_1}$ is treated as exogenous. Under the null that $y^0_{igpt_1}$ is exogenous, the statistic is distributed as $\chi^2(1)$. This statistic is not corrected for the problem of few clusters.

Bibliography

Amemiya, Y. (1985). Instrumental variable estimator for the nonlinear errors–in–variables model. *Journal of Econometrics 28*(3), 273–289.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Baltagi, B., C. Kao, and S. Na (2013). Testing for cross-sectional dependence in a panel factor model using the wild bootstrap F test. *Statistical Papers 54*(4), 1067–1094.

Baum, C. F., M. E. Schaffer, and S. Stillman (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal 3*(1), 1–31(31).

Beaudry, P. and J. DiNardo (1991). The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data. *Journal of Political Economy 99*(4), 665–688.

Bell, D. N. and D. G. Blanchflower (2010). Youth Unemployment: Déjà Vu? Iza discussion papers, Institute for the Study of Labor (IZA).

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics 90*(3), 414–427.

Clark, A. E. (2003). Unemployment as a Social Norm: Psychological Evidence from Panel Data. *Journal of Labor Economics 21*(2), 323–351.

Clark, A. E., Y. Georgellis, and P. Sanfey (2001). Scarring: The Psychological Impact of Past Unemployment. *Economica 68*(270), 221–241.

Cockx, B. and C. Ghirelli (2014). Scars of Recessions in a Rigid Labor Market. Mimeo, SHERPPA - Ghent University. Supplementary Online Appendix: `http://users.ugent.be/~bcockx/Ascars.pdf`.

Cockx, B. and M. Picchio (2013). Scarring effects of remaining unemployed for long-term unemployed school-leavers. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 176*(4), 951–980.

Davidson, R. and J. MacKinnon (2010). Wild Bootstrap Tests for IV Regression. *Journal of Business & Economic Statistics 28*(1), 128–144.

Davidson, R. and J. G. MacKinnon (2008). Bootstrap inference in a linear equation estimated by instrumental variables. *Econometrics Journal 11*(3), 443–477.

Devereux, P. J. and R. A. Hart (2007). The Spot Market Matters: Evidence on Implicit Contracts from Britain. *Scottish Journal of Political Economy 54*(5), 661–683.

Ellwood, D. T. (1982). Teenage Unemployment: Permanent Scars or Temporary Blemishes? In R. B. Freeman and e. David A. Wise (Eds.), *The Youth Labor Market Problem: Its Nature, Causes, and Consequences*, pp. 349–390. Chicago: University of Chicago Press.

Gangji, A. V. and R. Plasman (2007). The Matthew effect of unemployment: how does it affect wages in Belgium. DULBEA Working Papers 07-19.RS, ULB – Université Libre de Bruxelles.

Gartell, M. (2009). Unemployment and subsequent earnings for Swedish college graduates. A study of scarring effects. Arbetsrapport 2009:2, Institute for Futures Studies.

Gregg, P. (2001). The Impact of Youth Unemployment on Adult Unemployment in the NCDS. *The Economic Journal 111*(475), 626–653.

Gregg, P. and E. Tominey (2005). The wage scar from male youth unemployment. *Labour Economics 12*(4), 487–509.

Hagedorn, M. and I. Manovskii (2013). Job Selection and Wages over the Business Cycle. *American Economic Review 103*(2), 771–803.

Hausman, J. (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives 15*(4), 57–67.

Hausman, J. A., W. K. Newey, and J. L. Powell (1995). Nonlinear errors in variables Estimation of some Engel curves. *Journal of Econometrics 65*(1), 205–233.

Heckman, J. J., R. J. Lalonde, and J. A. Smith (1999). *The economics and econometrics of active labor market programs*, Volume 3 of *Handbook of Labor Economics*, Chapter 31, pp. 1865–2097. Elsevier.

Kilponen, J. and T. Santavirta (2010). New Evidence on Implicit Contracts from Linked Employer–Employee Data. *Scandinavian Journal of Economics 112*(4), 864–883.

Kletzer, L. G. and R. W. Fairlie (2003). The long-term costs of job displacement for young adult workers. *Industrial and Labor Relations Review 56*(4), 682–698.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013). Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment. *The Quarterly Journal of Economics*.

Lockwood, B. (1991). Information Externalities in the Labour Market and the Duration of Unemployment. *Review of Economic Studies 58*(4), 733–53.

MacKinnon, J. G. and M. D. Webb (2014). Wild Bootstrap Inference for Wildly Different Cluster Sizes. Working Papers 1314, Queen's University, Department of Economics.

Moulton, B. R. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units. *The Review of Economics and Statistics 72*(2), 334.

Mroz, T. A. and T. H. Savage (2006). The Long–Term Effects of Youth Unemployment. *The Journal of Human Resources 41*(2), 259–293.

Neumark, D. (2002). Youth Labor Markets in the United States: Shopping Around vs. Staying Put. *Review of Economics and Statistics 84*(3), 462–482.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012). The Short- and Long-Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Oreopoulos, P., T. v. Wachter, and A. Heisz (2008). The Short– and Long–Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates. Discussion Paper 3578, IZA, Bonn.

Pissarides, C. A. (1992). Loss of Skill During Unemployment and the Persistence of Employment Shocks. *The Quarterly Journal of Economics 107*(4), 1371–1391.

Schmillen, A. and M. Umkehrer (2013). The scars of youth: effects of early-career unemployment on future unemployment experience. IAB Discussion Paper 201306, Institute for Employment Research, Nürnberg, Germany.

Sonar (2003). Hoe maken de jongeren de overgang van school naar werk? Basisrapportering Cohorte 1978 (eerste golf). Technical report, Leuven: Steunpunt WAV.

Sonar (2004a). Hoe maken de jongeren de overgang van school naar werk? Basisrapportering Cohorte 1976 (tweede golf). Technical report, Leuven: Steunpunt WAV.

Sonar (2004b). Hoe maken de jongeren de overgang van school naar werk? Cohorte 1980 (eerste golf). Technisch rapport. Technical report, Leuven: Steunpunt WAV.

Stock, J. and M. Yogo (2005). Testing for Weak Instruments in Linear IV Regression. In D. W. Andrews (Ed.), *Identification and Inference for Econometric Models*, pp. 80–108. New York: Cambridge University Press.

Topel, R. (1990). Specific capital and unemployment: Measuring the costs and consequences of job loss. *Carnegie-Rochester Conference Series on Public Policy 33*, 181–214.

Topel, R. H. and M. P. Ward (1992). Job Mobility and the Careers of Young Men. *The Quarterly Journal of Economics 107*(2), 439–79.

Yedid-Levi, Y., N. Jaimovic, H. Siu, and M. Gervais (2014). What Should I Be When I Grow Up? Occupations and Employment over the Life Cycle and Business Cycle. Working Paper Series 20628, NBER.

3

Is it socially efficient to impose job search requirements on unemployed benefit claimants with hyperbolic preferences?

Published in the Journal of Public Economics (2014, Volume 113, Pages 80-95)

3.1 Introduction

Long-term unemployment is a major problem, in particular in a number of European labor markets (OECD, 2011). This pattern comes along with evidence that the unemployed have a very low search activity (Krueger and Mueller, 2010; Manning, 2011). It is well known that the provision of Unemployment Insurance (UI) raises moral hazard problems, i.e. the more generous UI, the lower the search incentives for the unemployed (e.g. Lalive et al., 2006). Many countries impose job search requirements on benefit recipients to cope with moral hazard in UI (OECD, 2007). To verify compliance, job search effort is monitored and, in the case of non-compliance, benefit recipients are sanctioned. However, as any policy addressing moral hazard, monitoring involves an insurance-efficiency trade-off (Boone et al., 2007; Boone and van Ours, 2006; Cockx et al., 2011). Restoring incentives comes at the cost of reducing the capacity of UI to adequately insure workers against the risk of unemployment. Job search monitoring is different from other policy instruments as it does not directly affect the unemployment benefit (UB) level. However, monitoring increases job search costs and decreases the average quality of prospective jobs, since rational, forwardlooking unemployed workers typically reduce their reservation wage in response to the higher job search requirements. Hence, the expected lifetime utility of the unemployed is negatively affected.

These results apply for individuals with standard exponential time preferences. These individuals discount the future at a constant rate and, hence, behave consistently over time. However, both laboratory experiments and empirical studies find evidence that procrastination in intertemporal choices is common (e.g. see Ainslie, 1992; Loewenstein and Thaler, 1989; Thaler and Shefrin, 1981; for a critical review see Frederick et al., 2002). That is, people seem to show self-control problems whenever they have to commit to a plan entailing present costs and future rewards (or vice versa). They may keep postponing the costly task over time and end up not achieving the future rewards, even if it was rationally optimal to reach them. This is evidence of *hyperbolic* discounting. Individuals exhibit a high degree of discounting in the short run and a relatively low degree of discounting in the long run. To cope with this limitation, a new branch of economics has been investigating intertemporal choices under the assumption of hyperbolic time preferences (e.g., Loewenstein and Prelec, 1992; O'Donoghue and Rabin, 1999).[1]

Recently, based on a longitudinal experiment on intertemporal effort choices, Augenblick et al. (2013) found limited evidence of a present bias in choices over monetary payments. By contrast, individuals procrastinate substantially in effort choices. Moreover, these individuals are more likely to choose a commitment device that forces them to complete more effort than they instantaneously desire, since they are aware of their present bias and take actions to limit their future behavior, i.e. they are *sophisticated* hyperbolic agents. This is consistent with earlier research by DellaVigna and Paserman (2005) - hereafter referred to as PDV - and Paserman (2008) who find evidence that hyperbolic preferences are particularly relevant to explain the patterns of job search behavior observed in the US. Job search effort typically entails immediate costs and delayed benefits. Consequently, individuals with hyperbolic preferences are always tempted to delay job search. Since unemployed workers engage too little in job search, PDV show that they are willing to pay a positive price for a commitment device that forces them to search more intensively if they are sophisticated hyperbolic agents. Job search monitoring could be such a commitment device. Based on simulations of an estimated structural job search model on US data, Paserman (2008) has indeed demonstrated that, if workers are impatient, monitoring job search can improve their long-run utility by lowering the expected duration of unemployment and raising the expected wages. In other words, to the extent that monitoring is relatively cheap to implement (Boone et al., 2007; Cockx et al., 2011), it can unambiguously lower government expenditures and increase social welfare without facing an insurance-efficiency trade-off. This contrasts with the conclusions for unemployed people with exponential time preferences.

Empirical evidence does not unambiguously support these positive conclusions with regards

[1]Researchers have studied the implications of this different behavioral assumption on various economic decisions. For instance, among others Laibson (1997) and Angeletos et al. (2001) examined saving-consumption decisions, while Carrillo and Mariotti (2000) focused on learning decisions and Fang and Silverman (2009) on labor supply and welfare participation. Others investigated specific consumption decisions: e.g. Mullainathan and Gruber (2005) focused on smoking, Fang and Wang (2010) on preventive health care, while DellaVigna and Malmendier (2006) studied contract choices and attendance to health clubs.

to job search monitoring. Neither does the job finding rate, nor the job quality always increase, and sometimes the unemployed rather exit to inactivity. For instance, Klepinger (1998), McVicar (2008) and Cockx and Dejemeppe (2012b) find that monitoring enhances the job finding rate. By contrast, Ashenfelter et al. (2005) find that tighter search requirements have insignificant effects on transitions to employment, and Klepinger et al. (2002) even find negative effects. In addition, Petrongolo (2009) reports negative impacts on the job quality (mainly earnings and employment duration) and, together with Manning (2009), she reports evidence that tighter search requirements lead to abandoning the UB claimant status.

In this paper, we show that these ambiguous findings on the effectiveness of job search monitoring need not be incompatible if the unemployed behave as agents with hyperbolic time preferences. This is because the decision to comply with the imposed job requirements does not depend on the long-run utility of these agents, but rather on the short-run utility of the current self for whom the benefits of enhanced search are shown to be smaller. Consequently, even if job search requirements are set at a sub-optimal, i.e. too low, level from the perspective of the future selves, unemployed procrastinators may nevertheless stop complying because the search requirements are too demanding from their perspective. Hence, it is shown that increasing job search requirements to a level that is optimal from the perspective of the future selves or from the perspective of society may after all still lead to a sub-optimal level of search effort and a long-run utility that is even lower than it was in the absence of job search requirements. Furthermore, we show that imperfections in the monitoring technology induced by caseworker discretion or by measurement error reinforce this problem.

The policy implication of this analysis is that job search monitoring may improve social welfare unambiguously only if the job search requirements are not set at too high a level. Moreover, it is shown that if, as a consequence of measurement error, benefit claimants always face a strictly positive sanction probability, job search monitoring may not be socially efficient. This means that other policies, such as job search assistance, may be more efficient socially than a system imposing job search requirements on hyperbolic unemployed benefit claimants. In the end, determining whether this is the case is an empirical matter.

The model extends the basic partial equilibrium job search model (Mortensen, 1986) in three directions. First, we introduce hyperbolic discounting as in PDV. We consider agents with sophisticated hyperbolic preferences. The case of agents with naive preferences is relegated to the Supplementary Appendix in Chapter B. Second, we include a perfect job search monitoring scheme in this model, in a very similar way to what Manning (2009) and Petrongolo (2009) do for individuals with exponential preferences. Finally, we allow for imperfections in the monitoring technology by allowing, first, caseworkers to have some discretion regarding whom they sanction and, second, search effort to be measured with error (see e.g. Boone et al., 2007; Cockx et al., 2011). We contribute to the literature on hyperbolic discounting by developing a graphical exposition of the impact of hyperbolic preferences on the choice of job search effort and the reservation wage. This graphical exposition contributes to a better intuitive understanding of the main results of this paper.

The rest of the paper is organized as follows. Section 3.2 describes the basic model. In this model, the monitoring technology is assumed to be perfect. We describe the assumptions and notations, the optimization problem of the sophisticated agent and present the first-order conditions of the solution. We devote a separate section to the graphical analysis of the solution. In Section 3.4, we discuss why raising the job search effort of benefit recipients with hyperbolic preferences can be socially efficient and how non-compliance affects this property. In Section 3.5, we generalize the model by incorporating an imperfect monitoring technology. First, we consider the case in which the caseworker has discretionary power as to whether a non-complying benefit claimant should be sanctioned or not. Second, we allow job search effort to be measured with error. In Section 3.6, we briefly contrast monitoring search with job search assistance. A final section concludes. All propositions are proved in the Appendix to this paper, while the case of a naive agent is treated, as mentioned, in the Supplementary Appendix (see Chapter B).

3.2 The Basic Model

3.2.1 Assumptions and Notations

We develop a partial equilibrium job search model under hyperbolic preferences in a stationary discrete-time setting. Infinitely-lived unemployed workers choose their reservation wage x and a scalar search-effort intensity σ to maximize their expected discounted lifetime utility. We denote $c(\sigma)$ the cost of effort and make the standard assumptions that $c(0) = 0$, $c'(\sigma) > 0$ and $c''(\sigma) > 0$. Unemployed workers are entitled to a flat unemployment benefit (UB) with no time limit. The total income while unemployed, $y_b > 0$, is equal to the UB plus any other external income (e.g. income from a partner). The payment of the UB is conditional on a search requirement $\bar{\sigma} > 0$. In our stylized benchmark representation, we assume that monitoring is perfect, meaning that job search effort is observed with perfect precision and that, if search effort falls below the requirement $\bar{\sigma}$, a sanction is imposed with probability one. This stylized representation of a monitoring scheme has also been adopted by other researchers (Manning, 2009; Petrongolo, 2009, e.g.). In Section 3.5, we study the consequences of measurement error in the monitoring technology. If the benefit claimant does not comply with the search requirement, the UB is withdrawn permanently. The exogenous income of sanctioned individuals is their external income plus, depending on the institutional context, income provided by charities and/or a (means-tested) assistance benefit. The exogenous income of sanctioned individuals is denoted y_z, with by assumption $y_z < y_b$. The job search of sanctioned individuals is no longer monitored and is therefore chosen freely. As is standard in the job search literature, we assume risk-neutral agents[2] and a separable instantaneous utility in income and search effort $y_u - c(\sigma)$ for $u \in \{b, z\}$.

In each period, job offers arrive with a probability $\lambda(\sigma) > 0$ increasing with job search effort. We assume that $\lambda(0) = 0$, $\lambda'(\sigma) > 0$ and $\lambda''(\sigma) \leq 0$. The net wage associated with a job

[2](Cockx et al., 2013) show that the setting and its predictions can be generalized to the case of risk-averse individuals.

offer is randomly drawn from an exogenous cumulative wage offer distribution $F(w)$ defined on $[\underline{w}, \bar{w}] \in \mathbb{R}^+$. The disutility of the effort in employment is normalized to zero, so that the instantaneous net utility in employment is the net wage w. Employed individuals are laid off with an exogenous probability $q \in (0, 1)$.

The timing within a period is as follows. First, search effort is chosen by the insured unemployed. Under perfect monitoring, the unemployed then know for sure whether they will be sanctioned or not. Next, a job offer may arrive. If an offer arrives and is accepted, the employment spell starts at the beginning of the next period and the job search effort for the current period is not monitored. If instead the offer is rejected or no offer is received, the monitoring takes place. If search effort is too low, the sanction applies instantaneously, meaning that the income is y_z in the current period.[3] The income y_b instead accrues to the unemployed whose search efforts are deemed sufficient and to the job seekers who have found a job.

In order to capture hyperbolic preferences, we need to distinguish between the lifetime utility of the unemployed current self (referred to by superscript c) and the lifetime utility of the unemployed and employed future selves *viewed from the perspective of the current self* (referred to by superscript f). The discount factor from the current period to the next one is $\beta\delta$, while the discount factor between two successive future periods is $\delta \leq 1$, with $0 < \beta\delta \leq \delta \leq 1$. In the case of an agent discounting at an exponential rate, $\beta = 1$. For a procrastinator, $\beta < 1$. We consider a homogenous population, but let β take any value in $[0, 1]$.[4] For any level of search effort σ, the Bellman equation defining the expected lifetime utility of an unemployed current self (or short-run utility) and the expected lifetime utility of the unemployed future selves (or long-run utility) U^f verifies:

$$W\left(\sigma, U^f \mid y_u, \beta\delta\right) \equiv y_u - c(\sigma) + \beta\delta\left\{\lambda(\sigma)E_F\left\{\max\left(V^f, U^f\right)\right\} + (1-\lambda(\sigma))U^f\right\} \quad (3.1)$$

The current net income, $y_u - c(\sigma)$, depends on whether the agent complies with the requirement $\bar{\sigma}$ ($y_u = y_b$) or not ($y_u = y_z$). With probability $\lambda(\sigma)$, a job offer is received. If it is accepted, at the beginning of the subsequent period a job spell starts yielding the long-run utility V^f. Otherwise, the unemployment spell continues. The second term on the right-hand side of (3.1) measures the discounted expected lifetime utility in the subsequent period. E_F designates the expectation over the wage offer distribution $F(.)$. This second term depends on whether the current self complied or not. If the current self did not comply, the continuation value in unemployment U^f is denoted Z^f. If the current self did comply, then she believes that she will comply in the next period as well, so that the continuation value U^f is denoted B^f. This is because the agents act in a stationary environment. Moreover, sophisticated agents are aware that their future selves are impatient and will set their search effort to the same level as the current self.

[3](Cockx et al., 2013) develop the analysis under the alternative assumption that the UB is only withdrawn from the period after detection. The assumption made here simplifies the exposition.

[4]In reality, people have different propensities to procrastinate and labor market policies can hardly discriminate between them. This makes their design complex.

3.2.2 The Optimization Problem

We first consider the optimization problem of the current self as a function of the continuation payoffs of the future selves. Subsequently, we state the optimization problems of the future selves. An unemployed worker maximizes her expected lifetime utility with respect to three choices in the following order: (i) the decision to comply or not with the search requirement $\bar{\sigma}$; (ii) the job search intensity σ to set; and (iii) to accept or not, if a job is offered. These choices involve very different intertemporal trade-offs. Both the decision to comply and the choice of the search intensity induce an instantaneous increase in search costs that should be balanced out by the expected future benefit stream to which one remains entitled by not being sanctioned. Accepting a job or continuing the search does not impose immediate costs or generate immediate benefits, but it affects the stream of benefits and costs during the future working life. This means that the first and the second choice involve a comparison between short- and long-term pay-offs, whereas the last one consists in trading off long-term utility streams only. Short-term impatience β matters for the first two decisions, but not for the last one. This difference has important implications for the optimal behavior of impatient agents.

The aforementioned choice problem can be formalized by the following optimization problem.

$$B^c(\bar{\sigma}) = \max_{\sigma \geq \bar{\sigma}} W\left(\sigma, B^f(\bar{\sigma}) \mid y_b, \beta\delta\right) \tag{3.2}$$

$$Z^c = \max_{\sigma} W\left(\sigma, Z^f \mid y_z, \beta\delta\right) \tag{3.3}$$

$$\Omega^c = \max\left\{B^c(\bar{\sigma}), Z^c\right\} \tag{3.4}$$

$B^c(\bar{\sigma})$ (resp., $B^f(\bar{\sigma})$) denotes the optimal short-run (resp., long-run) utility of an unemployed agent who complies with the job search requirement. In what follows, we will write B^i instead of $B^i(\bar{\sigma})$ ($i \in \{c, f\}$), except if we explicitly consider the case in which the search requirement is binding. Z^c (resp., Z^f) is the optimal short-run (resp., long-run) utility of a non-complying (sanctioned) unemployed agent.[5] The unemployed worker will comply or not, depending on what yields the highest short-run utility: $\Omega^c = \max\{B^c, Z^c\}$. Note that the latter decision clearly depends on the value of the short-term discount factor β.

The optimization problem of the current self is a function of the continuation payoffs of the future selves. Since the intertemporal values of the future selves are viewed from the perspective of the current self, they are all discounted by a factor δ instead of $\beta\delta$. Consequently, the intertemporal value in employment, V^f, verifies:

$$V^f = w + \delta\left[(1-q)V^f + qU^f\right] \tag{3.5}$$

from which it is clear that V^f varies with the wage but also with $U^f \in \{B^f, Z^f\}$. The expected lifetime utility of being employed is equal to the wage plus the discounted benefit of the continuation payoff in the next period. With probability $1-q$ the agent remains employed, whereas

[5]Notice that under perfect monitoring, a non-complying agent behaves in the same way as a a sanctioned individual. Under imperfect monitoring, this will no longer be the case.

with probability q she is laid off. In the case of a layoff, we assume that, if the individual was never sanctioned in the past, the eligibility to the UB is restored irrespectively of the length of the employment spell.[6] By contrast, we assume that a dismissed worker is not entitled to the UB if the individual did not comply with the search requirements and was sanctioned in the past. In other words, we ignore the "entitlement effect". We make this assumption, since it is not essential and simplifies the derivations. Note that these assumptions are implicit in the continuation payoff in the case of a layoff in (3.5), i.e. $U^f = B^f$ in the first case and $U^f = Z^f$ in the second one.

Turning to the intertemporal value of the unemployed future selves, sophisticated agents are aware that they will procrastinate in the future and set their search effort like their current selves. So,

$$U^f = W\left(\sigma_u, U^f \mid y_u, \delta\right) \tag{3.6}$$

where for $U^f = B^f$ and $y_u = y_b$ (resp., $U^f = Z^f$ and $y_u = y_z$) $\sigma_u = \sigma_b$ (resp., $\sigma_u = \sigma_z$) denotes the optimal search effort solving (3.2) (resp., (3.3)).

3.2.3 The First-Order Conditions of the Optimization Problem

The acceptance and search effort decisions are now characterized by first-order conditions. Next, we combine these conditions with the compliance decision to characterize the behavior of unemployed workers.

Using (3.5), we can write

$$V^f - U^f = \frac{w - (1-\delta)U^f}{1 - \delta(1-q)}, \qquad U^f \in \{B^f, Z^f\} \tag{3.7}$$

As the right-hand side is strictly increasing in w, the indifference condition $V^f - U^f = 0$ defines a unique reservation wage x_u ($u \in \{b, z\}$)

$$x_u = (1-\delta)U^f, \qquad U^f \in \{B^f, Z^f\} \tag{3.8}$$

Any job offer above (resp., below) x_u is accepted (resp., rejected). Using this reservation wage property, we can rewrite $W\left(\sigma, U^f \mid y_u, \beta\delta\right)$ as follows (see Appendix 3.8):

$$W\left(\sigma, U^f \mid y_u, \beta\delta\right) = y_u - c(\sigma) + \beta\delta\left\{\frac{\lambda(\sigma)}{1-\delta(1-q)} Q\left((1-\delta)U^f\right) + U^f\right\} \tag{3.9}$$

where

$$Q\left((1-\delta)U^f\right) \equiv \int_{(1-\delta)U^f}^{\bar{w}} \left[w - (1-\delta)U^f\right] dF(w) \tag{3.10}$$

with $Q'(x) = -\bar{F}(x) < 0$: if a job is found, the expected wage gain Q declines with the reservation wage. Inserting (3.9) with $\beta = 1$ in (3.6), the first-order condition (3.8) becomes:

$$y_u + \frac{\delta\lambda\left(\sigma_u\right)}{1-\delta(1-q)} Q\left(x_u\right) = c\left(\sigma_u\right) + x_u, \qquad u \in \{b, z\} \tag{3.11}$$

[6] This is a simplifying assumption, as in reality the entitlement to the UB usually depends on the past record of insurance contributions while employed. However, our assumption is not too restrictive: Paserman (2008) shows that the findings do not crucially depend on this assumption.

The left-hand side of (B.1) is equal to the benefit of continuing search for one more period rather than accepting a job offer at the reservation wage. This benefit is the instantaneous income in unemployment y_u plus the expected discounted wage gain in the case of continued search. This sum is equal to the cost of continued search on the right-hand side, i.e. the direct cost of job search plus the foregone income due to rejecting a job offer at the reservation wage.

Inserting (3.9) in (3.2) and (3.3), and differentiating with respect to σ, yields the first-order condition of job search effort:

$$\frac{\beta\delta\lambda'(\sigma_u)}{1-\delta(1-q)}Q(x_u)+\mu_u=c'(\sigma_u) \text{ and } \mu_u(\sigma_u-\bar{\sigma})=0, \qquad u\in\{b,z\} \tag{3.12}$$

where $\mu_b \geq 0$ is the Lagrange multiplier associated with the inequality constraint $\sigma \geq \bar{\sigma}$ in (3.2) and where $\mu_z = 0$, since the constraint $\sigma \geq \bar{\sigma}$ is absent in (3.3). Equation (B.2) states that the marginal benefit of search should equal its marginal cost, unless the constraint is binding. In the latter case, $\sigma_b = \bar{\sigma}$ and the marginal cost of search exceeds the marginal benefit. The agent would then like to decrease search effort, but cannot, since she would then violate the job search requirement.

3.3 A Graphical Representation of the Solution

In the previous subsection, we characterized the solution of the optimization problem by the first-order conditions of the reservation wage and the job search effort chosen by agents with hyperbolic time preferences. Here, we discuss the properties of this solution. We describe in particular how the differential discounting of the future affects the solution.

The graphical characterization introduced below fosters a more intuitive understanding of the results proven by PDV. In addition, it helps to explain how the solution changes as job search requirements are imposed through a monitoring scheme, a complication ignored in the analysis of PDV. More specifically, it clearly explains the point at which agents stop complying with the search requirement and illustrates how this point is affected by the time preferences of the agents. It will be shown that, as for the decision regarding job search intensity, the decision to comply or not involves a conflict between the current and future selves.

We define the implicit function $R(\sigma,x\mid y_u)=0$ (resp., $S(\sigma,x\mid\beta)=0$) to represent first-order condition (B.1) (resp., (B.2) when $\mu=0$) in a (σ,x)-space:

$$R(\sigma,x\mid y_u) \equiv y_u+\frac{\delta\lambda(\sigma)}{1-\delta(1-q)}Q(x)-c(\sigma)-x=0, \tag{3.13}$$

$$S(\sigma,x\mid\beta) \equiv \frac{\beta\delta\lambda'(\sigma)}{1-\delta(1-q)}Q(x)-c'(\sigma)=0 \tag{3.14}$$

We assume that this system of equations admits an interior solution $(\sigma_u,x_u)\in\mathbb{R}_0^+\times\mathbb{R}_0^+$.[7] The decision to comply or not to the job search requirements only shifts the first-order condition of the

[7]Theorem 1 of PDV shows that the solution is unique. Cockx et al. (2013) generalize the proof to the case of risk-averse agents.

reservation wage (3.13). If the unemployed worker complies, $U^f = B^f$ and $y_u = y_b$. If she does not comply, $U^f = Z^f$ and $y_u = y_z$. $S(\sigma, x \mid \beta)$ is not directly affected by the level of y_u. It is also easily seen that β only plays a direct role in the condition (3.14) characterizing search effort.

Let $x = r(\sigma \mid y_u)$ and $x = s(\sigma \mid \beta)$ denote the explicit relationship between the reservation wage and search effort described by the implicit equations $R(\sigma, x \mid y_u) = 0$ and $S(\sigma, x \mid \beta) = 0$. Proposition 1 shows that $s(\sigma \mid \beta)$ is always decreasing in σ. Since the job arrival rate displays decreasing returns and the job search effort increasing marginal costs, the *net* marginal return to job search effort is a decreasing function of σ for any given reservation wage x. Consequently, in order to compensate for this lower return, the reservation wage should be lowered, because this increases the expected wage gain if a job is found ($Q'(x) < 0$), and thereby ensures that the net marginal benefit of job search does not fall, i.e. that $S(\sigma, x \mid \beta) = 0$. The costlier the job search, the steeper the $x = s(\sigma \mid \beta)$ curve.

Proposition 1 also demonstrates that $r(\sigma \mid y_u)$ is increasing (decreasing) in σ for all σ smaller (greater) than the optimal job search effort σ_u^e of an exponential agent verifying the system (3.13)-(3.14) for $\beta = 1$. This property means that the reservation wage and, hence, by (3.8), the long-run utility are maximized if job search effort is set to the optimal level for an exponential agent. This makes sense, since only the long-run discount factor δ matters for setting the reservation wage. The $x = r(\sigma \mid y_u)$ relationship is therefore hump-shaped and reaches a maximum where it crosses the $x = s(\sigma \mid 1)$ curve.

Proposition 1. *The relationship $x = s(\sigma \mid \beta)$ implicitly defined by* (3.14) *is downward sloping. A rise in β shifts this relationship upwards in the (σ, x) space. The relationship $x = r(\sigma \mid y_u)$ implicitly defined by* (3.13) *reaches a maximum when it crosses the $x = s(\sigma \mid 1)$ curve, i.e. at the level of search optimally chosen by an exponential agent. A rise in y_u shifts the humped-shaped curve $x = r(\sigma \mid \beta)$ upwards.*

Proof. See Appendix 3.9. □

Before considering the sophisticated procrastinator, let us briefly consider the exponential agent. If the job search requirement $\bar{\sigma}$ is not binding, the interior solution (σ_b^e, x_b^e) can be found by solving the first-order conditions (3.13) and (3.14) in which $\beta = 1$ and $y_u = y_b$. Graphically, this corresponds to the intersection of the curves $r(\sigma \mid y_b)$ and $s(\sigma \mid 1)$ at point A in Figure 3.1. Note that, in line with Proposition 1, this intersection occurs at the maximum of $r(\sigma \mid y_b)$. Assume now that the job search requirement starts to be binding and the agent complies. Search effort equals $\bar{\sigma} > \sigma_b^e$ and the first-order condition with respect to search is no longer satisfied. So, the solution is now characterized by the pair $(\bar{\sigma}, r(\bar{\sigma} \mid y_b))$, i.e. one moves along the curve $r(\sigma \mid y_b)$ to the right of A. By (3.8), the reservation wage $r(\bar{\sigma} \mid y_b)$ is proportional to the intertemporal utility level. So, by moving to the right of A, the lifetime utility of the complier shrinks. The agent will stop complying as soon as the search requirement $\bar{\sigma}$ is set so high that her expected utility falls below the level she would obtain if she were sanctioned. Since the optimal solution for a sanctioned exponential agent (σ_z^e, x_z^e) is located at the intersection of the curves $r(\sigma \mid y_z)$ and $s(\sigma \mid 1)$, i.e. at point C in Figure 3.1, the maximum search requirement with which the exponential agent complies is denoted $\bar{\sigma}^e$ (point F) and is formally defined by the following equality:

$$r(\bar{\sigma}^e \mid y_b) = r(\sigma_z^e \mid y_z) \tag{3.15}$$

Beyond $\bar{\sigma}^e$, the exponential unemployed agent withdraws from the register of claimants preferring to devote a lower level of search effort σ_z^e.

Turning to the sophisticated procrastinator, let us first consider an agent for whom the job search requirement $\bar{\sigma}$ is not binding. The interior solution (σ_b, x_b) solves the first-order conditions (3.13) and (3.14) for $\beta < 1$ and $y_u = y_b$. Graphically, this corresponds to the intersection of the curves $r(\sigma \mid y_b)$ and $s(\sigma \mid \beta < 1)$ at point O in Figure 3.1. As expected and already shown by PDV, a hyperbolic agent searches less intensively than an exponential one. Since she is aware that she will also procrastinate in the future, her reservation wage is also lower than that of an exponential agent. Following a similar reasoning, we find that the sophisticated agent who does not comply chooses (σ_z, x_z) at point L in Figure 3.1.

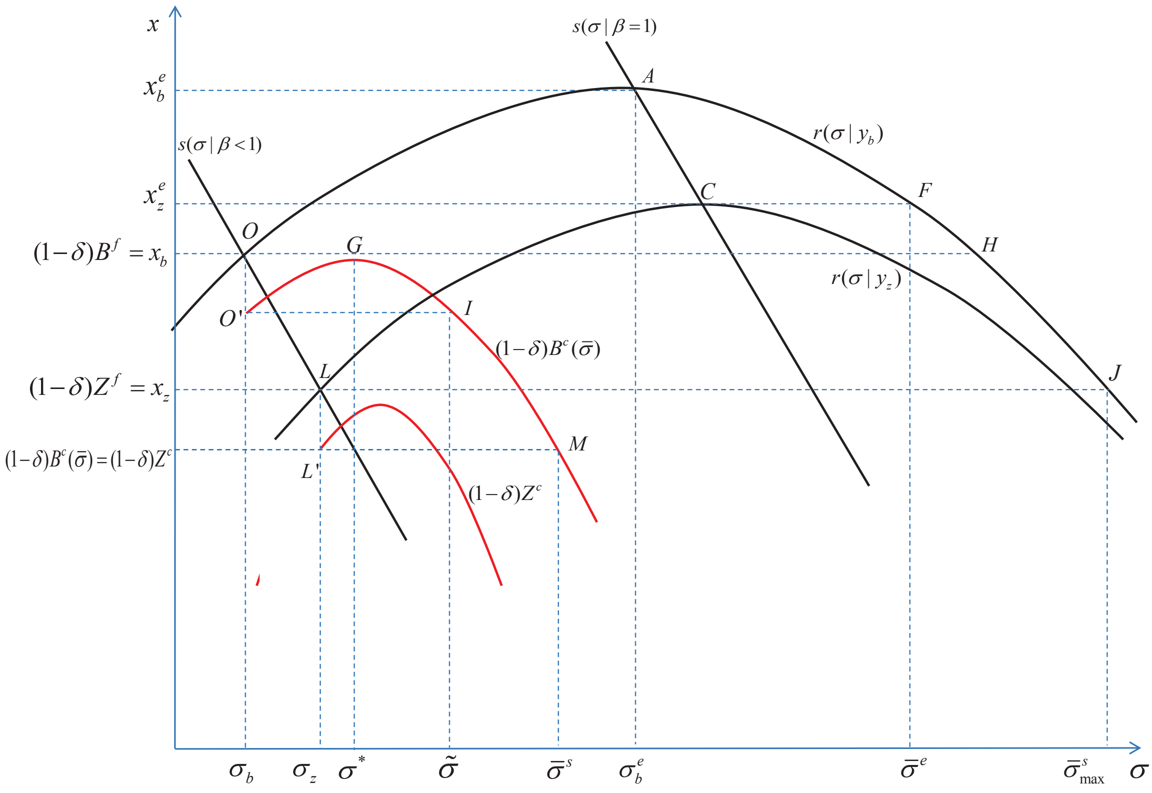


Figure 3.1: The Solution for the Exponential Agent and the Sophisticated Hyperbolic Agent in the Case of Perfect Monitoring. x = reservation wage; σ = search effort.

As the job search requirement $\bar{\sigma}$ is raised above σ_b, it starts to t be binding. A sophisticated agent knows that she has a self-control problem. So, her long-run utility (and hence her reservation wage) strictly increases with the requirement up to x_b^e (at point A) and strictly decreases afterwards. However, the decision to comply depends on short-run utility. From (3.1), (3.2) and (3.6), it should be clear that in Figure 3.1 the short-run utility $B^c(\bar{\sigma})$ lies below $B^f(\bar{\sigma})$ for any $\bar{\sigma}$. When the search requirement is binding ($\bar{\sigma} \geq \sigma_b$), it should also be understood that $B^c(\bar{\sigma})$ (resp., $B^f(\bar{\sigma})$) are equal to $W\left(\bar{\sigma}, B^f(\bar{\sigma}) \mid y_u, \beta\delta\right)$ (resp., $W\left(\bar{\sigma}, B^f(\bar{\sigma}) \mid y_u, \delta\right)$). Similarly to the long-run, the short-run utility, $B^c(\bar{\sigma})$, also initially increases with the requirement starting from σ_b, because it increases in the long-run utility and because σ_b is optimally chosen from the perspective of the current self, so that close to the right of σ_b the marginal cost of search is only slightly higher than the marginal benefit. However, as the search requirement is raised further and approaches the optimal level σ_b^e of the exponential agent, the net marginal cost of search increases at an increasing rate, while the expected utility of the future selves decreases at a decreasing rate. Proposition 2 below formally shows that the short-run utility attains a maximum strictly between σ_b and σ_b^e. The inverse U-shaped curve in Figure 3.1 with a maximum at G (whose abscissa is σ^*) represents the short-run utility.

The unemployed agent complies with the search requirement as long as her short-run utility does not fall below that of the non-complying current self. The maximum search requirement $\bar{\sigma}^s$ beyond which the unemployed stops complying is the abscissa of point M, defined by the equality $B^c(\bar{\sigma}^s) = Z^c$, i.e. by using (3.2), (3.8) and (3.9):

$$y_b - c(\bar{\sigma}^s) + \beta\delta\left\{\frac{\lambda(\bar{\sigma}^s)}{1-\delta(1-q)}Q\left[r(\bar{\sigma}^s \mid y_b)\right] + \frac{r(\bar{\sigma}^s \mid y_b)}{1-\delta}\right\} = Z^c, \tag{3.16}$$

where Z^c has been characterized by (3.3) and is obviously lower than Z^f in Figure 3.1. The level of $\bar{\sigma}^s$ above which a sophisticated agent stops complying depends on functional forms and parameter configurations. However, we can bracket the threshold $\bar{\sigma}^s$ (see the proof of Proposition 2). First, it can be checked that $B^c(\sigma_z) > Z^c$. For, $y_b > y_z$ and $r(\sigma_z \mid y_b) > r(\sigma_z \mid y_z)$. Therefore, we are sure that $\bar{\sigma}^s > \sigma_z$. Second, let $\tilde{\sigma}$ verify $B^c(\tilde{\sigma}) = B^c$, i.e. the equality between the short-run utility levels of the constrained and the unconstrained agent ($\tilde{\sigma}$ is the abscissa of point I in Figure 3.1; $\tilde{\sigma} > \sigma_b$). As the unemployed agent freely chooses search effort σ_b, because of the inverse U-shaped profile of $B^c(\bar{\sigma})$, it should be clear that any search effort requirement between σ_b and $\tilde{\sigma}$ cannot induce the unemployed agent to withdraw from the register.[8] Third, let $\bar{\sigma}^s_{max}$ verify $B^f\left(\bar{\sigma}^s_{max}\right) = Z^f$, i.e. it is the highest acceptable search requirement if the current self is as patient as the future ones ($\bar{\sigma}^s_{max}$ is the abscissa of point J in Figure 3.1; $\bar{\sigma}^s_{max} > \sigma_b$). In sum,

$$\bar{\sigma}^s \in (\max\left\{\sigma_z, \tilde{\sigma}\right\}, \bar{\sigma}^s_{max}), \tag{3.17}$$

which implies that a rise in the search requirement eventually leads to a drop in search effort when the benefit claimant stops complying and is sanctioned (since $\sigma_z < \bar{\sigma}^s$). The following proposition summarizes these results:

Proposition 2. *When the search requirement starts to be binding, the sophisticated hyperbolic unemployed agent first complies because by doing so she increases her short-run utility. Raising the*

[8] In Figure 1 $\tilde{\sigma} > \sigma_z$, but note that we cannot generally exclude that $\tilde{\sigma} < \sigma_z$.

search requirement further eventually leads to a decline in her short-run utility. The unemployed agent stops complying when the search requirement reaches a level $\bar{\sigma}^s$ which is above the maximum of two search intensities: free choice if she stops complying (σ_z), and the search intensity yielding the same short-run utility as in the absence of any search requirement ($\tilde{\sigma}$). The requirement $\bar{\sigma}^s$ must always be lower than the highest acceptable one if the current self is as patient as the future selves ($\bar{\sigma}^s_{max}$). Any further increase above $\bar{\sigma}^s$ induces search effort to discretely fall to the level σ_z.

Proof. See Appendix 3.10. □

3.4 Can Monitoring the Search Effort Be Socially Efficient?

In the previous section we have shown that, although job-seekers with time-consistent preferences always loose if binding job search requirements are imposed, the lifetime utility of both the current and future selves of unemployed benefit claimants with sophisticated hyperbolic preferences may increase together with the search requirement. For the current self, this occurs up to point G in Figure 3.1 and for the future selves up to point A. PDV were the first to demonstrate this. In addition, they have shown that the increase in lifetime utility can go hand in hand with a higher job finding rate. Indeed, they have shown that, under mild conditions on the wage offer distribution, the direct positive effect on the exit rate to employment of the enhanced job search effort induced by a stricter search requirement dominates the indirect negative effect of the higher reservation wage induced by this more intensive search.[9] This finding suggests that imposing job requirements on impatient unemployed job seekers may be Pareto improving, since it simultaneously raises the welfare of the unemployed and the employed, who need to contribute less to finance the UI if the unemployed find jobs more rapidly. Based on simulations, Paserman (2008) indeed shows that job search monitoring can simultaneously improve workers' long-run utility, reduce unemployment duration, and lower government expenditures.

Since the cost of monitoring is negligible relative to the cost of the UI (e.g. Boone et al., 2007; Cockx et al., 2011), this suggests that it is socially efficient to impose job search requirements on hyperbolic unemployed benefit claimants. This view can, however, be challenged. First, our analysis shows that the lifetime utility of the current self stops increasing at lower levels of search effort and, hence, at lower levels of savings on UB payments, than that of the future selves. In Figure 3.1, the Pareto-efficient range of search requirements from the perspective of the current self lies between σ^* and $\tilde{\sigma}$ (on the line segment between G and I), to the left of the corresponding frontier for the future selves, between the abscissas of points A and H. Point H is defined by the level of search requirement at which the long-run utility of a complying individual is equal to that of a benefit recipient on whom no requirement has been imposed. Moreover, in the Supplementary Appendix (Chapter B) we show that, identically to an agent with exponential preferences, the

[9]Cockx et al. (2013) have shown that this result applies for an even wider range of wage offer distributions than those considered by PDV and that it also remains valid if job seekers are risk averse.

current self of a *naive* procrastinator incurs a welfare loss as soon as the search requirement is set above the optimal freely chosen level. Based on a Pareto efficiency criterion that requires that the welfare of all period selves be raised, the social efficiency gains are therefore much reduced or even nonexistent if hyperbolic job seekers have naive instead of sophisticated time preferences.

However, referring to Akerlof (1991)'s view on procrastination, O'Donoghue and Rabin (1999) argue that the aforementioned welfare criterion is too strong for welfare analysis. This is because the preferences of the current self are biased. The current self faces a self-control problem and, if she is naive, a misperception problem in addition (Gruber and Köszegi, 2000, 2001). The preferences of the sophisticated future selves are, by contrast, not biased. The authors therefore claim that the preferences of the future sophisticated selves are more appropriate to base a welfare analysis on. This view is confirmed by Noor (2011), who argues that normative judgments cannot be based on revealed preferences but instead on what he calls *normative* preferences, that is, preferences which reflect the choices the individual thinks she *should* make. Consequently, according to this view, irrespectively of whether the hyperbolic job seeker is naive or sophisticated, Pareto efficiency should only be considered from the perspective of the future selves, so that this first criticism on the social efficiency of the imposition of search requirements is not relevant from a welfare analysis point of view.

Even if the current self is not relevant to base a *welfare* analysis on, the *behavior* of job seekers is, nevertheless, determined by the current self. Our analysis in the previous section demonstrates in particular that it is the current self who decides whether or not to comply with the imposed job search requirement. We have shown that imposing too high a search requirement may induce unemployed benefit claimants to stop complying before the Pareto-efficient frontier AH in Figure 3.1 is attained, and that in the case of non-compliance, job search effort is reduced and the utility decreases, because a sanction is imposed. In Proposition 2 we have, nevertheless, shown that the maximum search requirement with which the exponential agent complies $\bar{\sigma}^e$ is lower than $\bar{\sigma}^s_{max}$. Hence, benefit claimants do not necessarily stop complying before the socially efficient range is reached. This depends on preferences and the search technology, i.e. the cost of search and the job arrival function. This result analytically confirms the findings of a simulation that Paserman (2008) conducted, based on his estimation results of a structural job search model in which job seekers display hyperbolic preferences. Depending on the type of worker and the level of effort requirement, Paserman found that workers would comply and experience a gain in long-run utility, or they would opt out of the UI and experience a substantial loss in utility. In the next section we show that if imperfections in the monitoring technology are taken into account, the likelihood that monitoring can entail social efficiency is further reduced.

3.5 The Consequences of an Imperfect Monitoring Technology

Up to this moment, we assumed that the monitoring technology was perfect. This has the two following implications. First, if a benefit claimant does *not* comply with the search requirement, she is sanctioned for sure. Second, if she *does* comply, she is sanctioned with probability *zero*.

In this section, we relax these assumptions. First, we consider a case in which only the second assumption is maintained, while the non-complier is sanctioned with an exogenous probability. This first case could proxy a situation in which caseworkers have discretion in determining which non-complying agents are sanctioned. In this decision, caseworkers would typically take other criteria, such as social need, into account. Since such decisions may be very heterogeneous and lack systematics, the non-complying benefit claimant may perceive that the sanction is to a certain extent randomly imposed, independently of her behavior and characteristics. We label this "imperfection due to caseworker discretion". Second, we assume that search effort is measured with error. If the range of measurement errors is sufficiently broad, it turns out that the benefit claimant is always sanctioned with a certain probability, irrespectively of whether she complies or not with the requirement. This second part also considers the case of a more limited range of errors, which shares features with the two previous cases.

3.5.1 Imperfection Due to Caseworker Discretion

The monitoring technology can in this case, for any $\bar{\sigma} > 0$, formally be described as follows:[10]

$$p(\sigma) = \begin{cases} p_0 \in (0,1) & \text{for } \sigma < \bar{\sigma} \text{ and} \\ 0 & \text{for } \sigma \geq \bar{\sigma}. \end{cases} \tag{3.18}$$

If the agent complies with the search requirement ($\sigma \geq \bar{\sigma}$), the Bellman equation defining the short-run utility for any level of search effort σ and long-run utility B^f corresponds to the one considered for a perfect monitoring technology, i.e. to (3.1) for $y_u = y_b$ and $U^f = B^f$. However, in the case of non-compliance ($\sigma < \bar{\sigma}$), the Bellman equation becomes:

$$W_p\left(\sigma, EU_0^f \mid Ey_0, \beta\delta\right) \equiv Ey_0 - c(\sigma) + \beta\delta\left\{\lambda(\sigma)E_F\left\{\max\left(V^f, EU_0^f\right)\right\} + (1-\lambda(\sigma))\, EU_0^f\right\} \tag{3.19}$$

where $Ey_0 \equiv p_0 y_z + (1-p_0)y_b$, $EU_0^f \equiv p_0 Z^f + (1-p_0)N^f$ and N^f denotes the long-run utility in the case of non-compliance after the monitoring in the current period has taken place and no sanction has been imposed. Note that N^f replaces U^f in the definition (3.5) of V^f.

The optimization problem of the current self is therefore given by (3.2) and the following equations:

$$N^c = \max_{\sigma<\bar{\sigma}} W_p\left(\sigma, EU_0^f \mid Ey_0, \beta\delta\right) \tag{3.20}$$

$$\Omega_n^c = \max\left\{B^c(\bar{\sigma}), N^c\right\} \tag{3.21}$$

Following a similar reasoning as in the case of perfect monitoring, we prove in Appendix 3.11

[10] Abbring et al. (2005) have considered this type of monitoring technology to capture the fact that "any sanction policy needs to be backed up by explicit rules and individuals can appeal against sanctions". This does not justify, however, a constant sanction probability below the threshold $\bar{\sigma}$: the higher the actual search effort, the easier it is to prove compliance. This is why we believe that this monitoring technology may be better suited to describing the aforementioned caseworker discretion.

that:

$$W_p\left(\sigma, EU_0^f \mid Ey_0, \beta\delta\right) \equiv y_b - c(\sigma) + \beta\delta\left\{\frac{\lambda(\sigma)Q\left((1-\delta)N^f\right)}{1-\delta(1-q)} + N^f\right\}$$
$$- p_0\left\{y_b - y_z + \beta\delta\left[1 - h\left(\sigma, (1-\delta)N^f\right)\right]\left(N^f - Z^f\right)\right\} \quad (3.22)$$

where $h\left(\sigma, (1-\delta)N^f\right) \equiv \lambda(\sigma)\bar{F}\left((1-\delta)N^f\right)$, i.e. the probability of finding a job in the current period. Using this expression, similarly to the case of perfect monitoring, we can derive the implicit functions representing the first-order conditions of $\sigma < \bar{\sigma}$ and x for a *non-complying* agent:

$$R_n(\sigma, x \mid Ey_0, p_0) \equiv Ey_0 + \frac{\delta\lambda(\sigma)Q(x)}{1-\delta(1-q)} - x - c(\sigma) - p_0\frac{\delta}{(1-\delta)}\left[1 - h(\sigma, x)\right](x - x_z) = 0 \quad (3.23)$$

$$S_n(\sigma, x \mid \beta, p_0) \equiv \frac{\beta\delta\lambda'(\sigma)}{1-\delta(1-q)}Q(x) + p_0\frac{\beta\delta}{(1-\delta)}\lambda'(\sigma)\bar{F}(x)(x - x_z) - c'(\sigma) = 0 \quad (3.24)$$

Not complying with the job requirement is less costly than in the case of perfect monitoring, since the sanction is no longer imposed with certainty: $Ey_0 > y_z$. This increases the benefit of continued search. There is also an additional cost to continued search. If one does not find a job by the end of the period (which explains the weighting by $[1 - h(\sigma, x)]$), one risks the capital loss of the sanction in the next period $(x - x_z)/(1-\delta)$. This additional cost must, however, always be smaller than the additional benefit, since otherwise $x - x_z < 0$ for any given σ, and then the additional cost in (3.23) would become a benefit, which is a contradiction. So, we can deduce from (3.23) that $x_z < x < x_b$ for any given σ and $0 < p_0 < 1$. Consequently, since $x > x_z$, the non-complying job seeker is more selective in her job acceptance decision than in the perfect monitoring case.

On the other hand, the considered imperfection in the monitoring technology raises the job search incentive in the case of non-compliance. This is because the unemployed agent may avoid the (possible) sanction if she finds a job before the end of the period. This additional marginal benefit of search is equal to $p_0\frac{\beta\delta}{(1-\delta)}\lambda'(\sigma)\bar{F}(x)(x - x_z)$ in (3.24). Consequently, in Figure 3.2, the $x = s_n(\sigma \mid \beta, p_0)$ curve (i.e. the explicit relationship between the reservation wage and search effort described by the implicit equation $S_n(\sigma, x \mid \beta, p_0) = 0$) lies to the right of the corresponding $x = s(\sigma \mid \beta)$ curve under perfect monitoring.

Observe that the first-order conditions tend towards those of perfect monitoring if $p_0 \to 1$ or $p_0 \to 0$. If $p_0 \to 1$, $Ey_0 \to y_z$ and the additional terms drop out, since then $x - x_z \to 0$. If $p_0 \to 0$, $Ey_0 \to y_b$, and the additional terms also shrink to zero. So, in the limit, the system (3.23)-(3.24) provides the optimality conditions for the behavior of an unemployed agent on whom no job search requirement is imposed ($p_0 = 0$) or who is sanctioned for sure ($p_0 = 1$).

In Figure 3.2, we illustrate the behavior of an impatient UB claimant for a given $p_0 \in (0, 1)$. From what is mentioned above, it should be clear that for any σ we have the following ranking in the (σ, x) space: $r(\sigma \mid y_z) < r_n(\sigma \mid Ey_0, p_0) < r(\sigma \mid y_b)$ where $r_n(\sigma \mid Ey_0, p_0)$ is the explicit

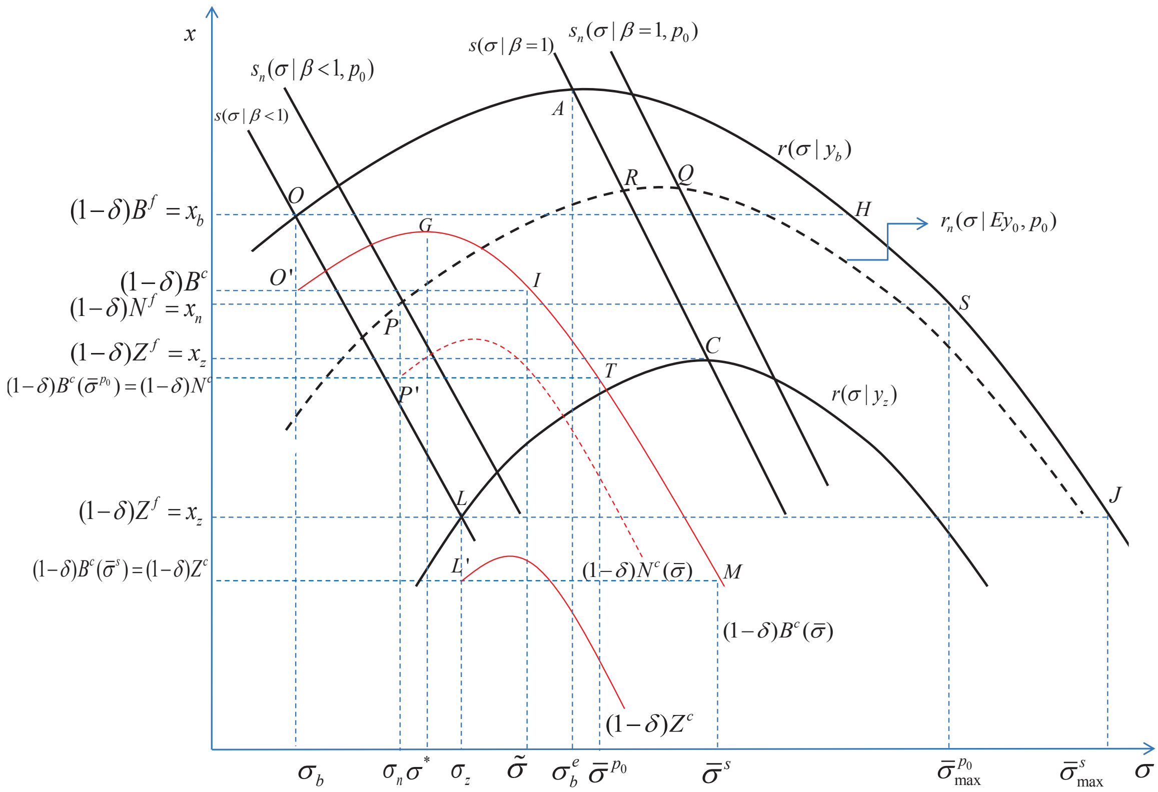


Figure 3.2: The Solution in the Case of Imperfection Due to Caseworker Discretion for p_0 such that the Curves $S_n(\sigma \mid \beta < 1, p_0)$ and $S_n(\sigma \mid \beta = 1, p_0)$ Are Downward-Sloping.

relationship between the reservation wage and search effort described by the implicit equations $R_n(\sigma, x \mid Ey_0, p_0) = 0$. The optimal solution in the case of perfect compliance is given at point O. The ordinate of point O', below point O, indicates the (monotonic transformation of) the short-run utility level at this optimum, i.e. $(1-\delta)B^c$. As $\bar{\sigma}$ is gradually raised above the optimal free level σ_b, the short-run utility first increases up to point G, falls thereafter until point T, at which it attains the short-run utility level for a non-complier, at the ordinate of P'. The search effort $\bar{\sigma}^{p_0}$ on the abscissa of point T and implicitly defined by $B^c(\bar{\sigma}^{p_0}) = N^c$ is the maximum level of the search requirement with which the benefit claimant complies. Let (σ_n, x_n), point P, denote the search effort and reservation wage combination solving (3.23)-(3.24). Whenever $\bar{\sigma}$ is set higher than $\bar{\sigma}^{p_0}$, job search effort falls discontinuously to σ_n and search effort stays at that level until the individual is sanctioned with probability p_0. Then the job search effort level jumps to σ_z. In Figure 3.2, $\sigma_z > \sigma_n$. However, as shown in the proof of Proposition 4, if p_0 is sufficiently close to one, it is possible that $\sigma_z \leq \sigma_n$. But observe that in the case of non-compliance, (σ_n, x_n) always lies to the right of the $x = s(\sigma \mid \beta < 1)$ curve (OL), because of the additional returns to job search present in the aforementioned first-order condition (3.24) if $0 < p_0 < 1$.

From Figure 3.2, it is clear that a UB claimant stops complying with the search requirement at a lower search effort level (at the abscissa of T) than in the case of perfect monitoring (at the abscissa of M): $\bar{\sigma}^{p_0} < \bar{\sigma}^s$. This is because the benefit claimant is not sanctioned with probability one if she does not comply, so that her short-run utility is strictly higher in this case: the (monotonic transformation of) the long-run utility $r_n(\sigma \mid Ey, p_0)$ in the imperfect monitoring case lies strictly above $r(\sigma \mid y_z)$ in the case of perfect monitoring. Similarly, we can demonstrate along the lines of the proof of Proposition 2 that the maximum search effort requirement above which an unemployed benefit claimant stops complying is strictly smaller than in the perfect monitoring case: $\bar{\sigma}^{p_0}_{max} < \bar{\sigma}^s_{max}$ (point S lies to the left of J), where $\bar{\sigma}^{p_0}_{max}$ verifies $B^f(\bar{\sigma}^{p_0}_{max}) = N^f$. In sum,

Proposition 3. *In the case of imperfect monitoring due to case worker discretion as described by* (3.18)*, a UB claimant stops complying with the search requirement at a lower search effort level than in the case of perfect monitoring. In addition, it is less likely that the Pareto-efficient frontier AH is attained.*

If agents have naive instead of sophisticated time preferences, it can even be shown that, if p_0 is not too large, some points of the Pareto-efficient frontier can never be reached.[11] We therefore conclude that the considered imperfection in the monitoring technology reduces the likelihood that social efficiency can be attained by imposing a job search requirement.

In Figure 3.2, we illustrated the behavior of the benefit claimant for a particular value of p_0. In Proposition 4 and Figure 3.3, we describe how the optimal search effort and reservation wage combination (σ_n, x_n) of a non-complying agent, the solution to (3.23)-(3.24), varies with p_0. We demonstrate that σ_n should increase in p_0 for p_0 close to zero, but also that it should decrease in p_0 for p_0 close to one. This reflects that the aforementioned additional benefit to job search in (3.24), induced by the fact that a job seeker may avoid a sanction if she finds a job before being monitored, is not only small if the sanction probability is low ($p_0 \to 0$), but also if the expected gain in the case of job acceptance, i.e. $x - x_z$, is small. This gain is small if $p_0 \to 1$, since $x - x_z = 0$ if $p_0 = 1$. By continuity, this means that the σ_n will be maximal at some value of p_0 strictly between zero and one, at point V in Figure 3.3. It is obvious that the optimal reservation wage lies somewhere between the $x = r(\sigma \mid y_b)$ and the $x = r(\sigma \mid y_z)$ curves. However, since $r(\sigma \mid y_b)$ is increasing in σ and σ_n is increasing in p_0 for p_0 close to zero, we cannot exclude that x_n is increasing in p_0 close to zero. But eventually x_n must decrease in p_0. In sum, the OL line in Figure 3.3 displays how the behavior of the non-complying agent qualitatively changes when p_0 varies between 0 and 1.

Proposition 4. *If, in the case of non-compliance, the monitoring technology is described by* (3.18)*, the optimal search effort σ_n strictly increases (decreases) in p_0 for $p_0 \to 0$ ($p_0 \to 1$), and the optimal reservation wage strictly decreases in p_0 for $p_0 \to 1$.*

Proof. See Appendix 3.12. □

[11] This occurs if the maximum of the $x = r_n(\sigma \mid Ey, p_0)$ curve, i.e. the level at which a naive agent stops complying, exceeds the OH-line.

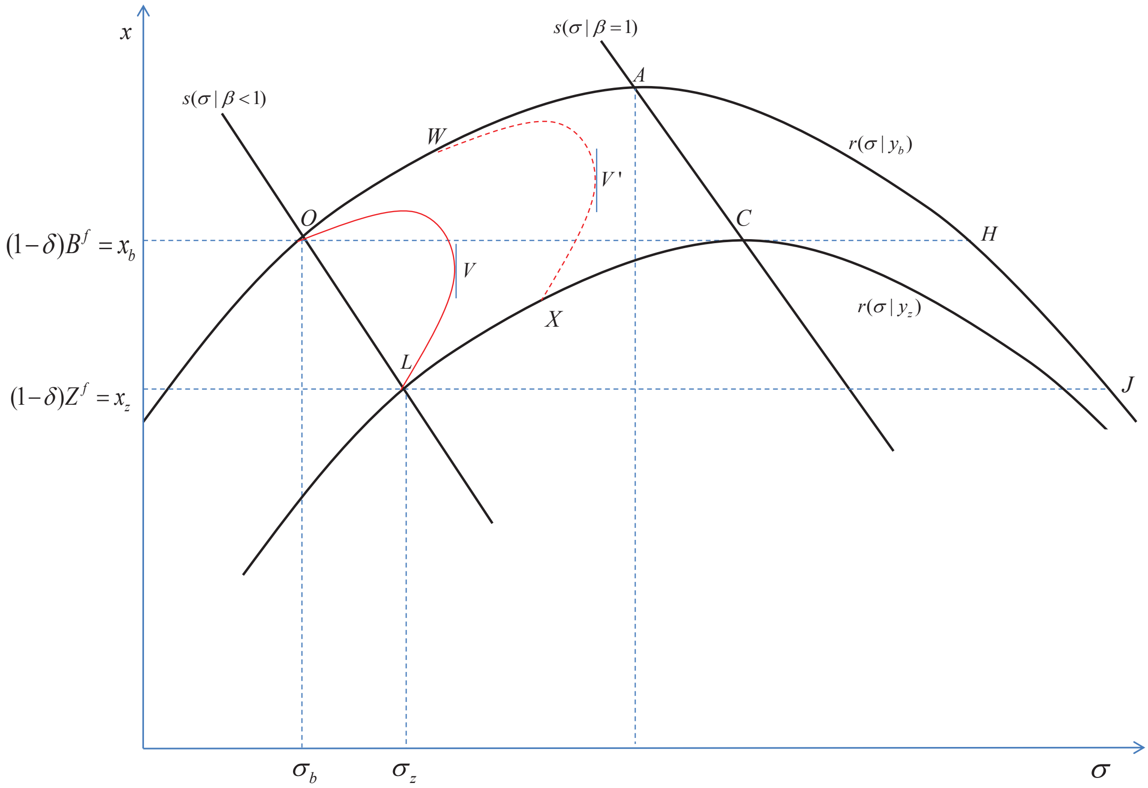


Figure 3.3: The Range of Solutions in the Case of Imperfect Monitoring as the Sanction Probability Raises from Zero to One: Non-Complying Agents in Sub-Section 3.5.1 and any UB Claimant in Sub-Section 3.5.2.

Contrary to what happens under perfect monitoring, the behavior of the non-complier differs from that of a sanctioned unemployed agent. We have seen that a non-complier has a higher reservation wage. As far as search effort is concerned, the non-complier searches less than a sanctioned individual if the search requirement is sufficiently close to σ_b and the sanction probability p_0 is sufficiently low. If the search requirement and p_0 are instead sufficiently high, then the search effort of a non-complier is above σ_z. As these qualitative conclusions also hold in the following sections, we will not repeat this comparison with the sanctioned job seeker.

3.5.2 Imperfection Due to Measurement Error

In this subsection, we allow for imperfect measurement of job search effort.[12] We follow the formalization along the lines of Boone et al. (2007), but assume a log-linear instead of a linear measurement error, since this is more natural for an error on a non-negative variable. If σ_0 denotes the search effort observed by the caseworker and ε the measurement error on the logarithm of

[12] We assume that the search requirement $\bar{\sigma}$ is clearly defined. Cockx et al. (2011) consider instead that it is a random variable from the perspective of the unemployed. We do not discuss this case explicitly here, since it can be shown to be very similar to the case of measurement error with unbounded support discussed below.

actual search effort, we can write:

$$\log \sigma_o = \log \sigma + \varepsilon \tag{3.25}$$

where $\varepsilon \in [-\underline{\varepsilon}, \overline{\varepsilon}]$ for $-\underline{\varepsilon} < 0 < \overline{\varepsilon}$, and $G(\varepsilon)$ denotes the cumulative distribution function of ε. Since the evaluation of job search effort now depends on the observed search effort level, which is no longer deterministically related to the *actual* effort level, the sanction is imposed with a probability $p(\bar{\sigma}/\sigma)$ which is directly related to the distribution of the measurement error:

$$p\left(\frac{\bar{\sigma}}{\sigma}\right) \equiv \text{Prob}\left(\sigma_o < \bar{\sigma}\right) = \text{Prob}\left(\log \sigma_o < \log \bar{\sigma}\right) = \text{Prob}\left[\varepsilon < \log\left(\frac{\bar{\sigma}}{\sigma}\right)\right] = G\left[\log\left(\frac{\bar{\sigma}}{\sigma}\right)\right], \tag{3.26}$$

This probability clearly decreases with σ:

$$\frac{\partial p\left(\frac{\bar{\sigma}}{\sigma}\right)}{\partial \sigma} = -\frac{g\left[\log\left(\frac{\bar{\sigma}}{\sigma}\right)\right]}{\sigma} < 0, \tag{3.27}$$

where $g(.)$ denotes the density function of ε. So, in contrast to the previously discussed imperfection, measurement error implies that the sanction probability strictly decreases with search effort. Nevertheless, if the measurement error has a bounded support, the sanction probability may still be constant for low or high *actual* levels of search effort σ:[13]

$$p\left(\frac{\bar{\sigma}}{\sigma}\right) \equiv \begin{cases} 1 & \text{for } \sigma \leq \bar{\sigma}\exp(-\overline{\varepsilon}) \\ \int_{-\underline{\varepsilon}}^{\log\left(\frac{\bar{\sigma}}{\sigma}\right)} g(\varepsilon)d\varepsilon & \text{for } \bar{\sigma}\exp(-\overline{\varepsilon}) < \sigma < \bar{\sigma}\exp(\underline{\varepsilon}) \\ 0 & \text{for } \sigma \geq \bar{\sigma}\exp(\underline{\varepsilon}) \end{cases} \tag{3.28}$$

Definition (3.28) shows that measurement error raises the level of the *actual* search requirement: $\bar{\sigma}\exp(\underline{\varepsilon}) > \bar{\sigma}$ for $\underline{\varepsilon} > 0$. Indeed, if the actual search intensity σ is below this level, the individual faces a strictly positive sanction probability. This means that an individual who actually complies ($\sigma = \bar{\sigma}$) still faces a strictly positive sanction probability: $p(1) = G(0) > 0$. Moreover, an individual who actually does not comply with the requirement ($\sigma < \bar{\sigma}$) is not sanctioned for sure as long as her search effort is greater than the lower bound of the *actual* search requirement, $\sigma > \bar{\sigma}\exp(-\overline{\varepsilon})$, where $\bar{\sigma}\exp(-\overline{\varepsilon}) < \bar{\sigma}$ for $\overline{\varepsilon} > 0$. Therefore if $\underline{\varepsilon} > 0$ or $\overline{\varepsilon} > 0$, i.e. as soon as there is some measurement error, a third regime emerges, alongside with full compliance and full non-compliance. If $\bar{\sigma}\exp(-\overline{\varepsilon}) < \sigma < \bar{\sigma}\exp(\underline{\varepsilon})$, we say that the unemployed agent *partially* complies with the job search requirement in the sense that she is sanctioned with an endogenous probability strictly between zero and one.

3.5.2.1 The Benchmark Case: Unbounded Support for the Measurement Error

We consider first the benchmark case in which the measurement error has an unbounded support. We assume throughout that the density of measurement error satisfies the following properties: $\forall \sigma, \bar{\sigma} \in (0, +\infty) : g\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right] > 0$, $\lim_{\varepsilon \to -\infty} g(\varepsilon) = \lim_{\varepsilon \to +\infty} g(\varepsilon) = 0$, $\lim_{\varepsilon \to -\infty} g'(\varepsilon) > 0$, and

[13] We use the fact that $\log(\bar{\sigma}/\sigma) \geq \overline{\varepsilon} \Leftrightarrow \sigma \leq \bar{\sigma}\exp(-\overline{\varepsilon})$ and that $\log(\bar{\sigma}/\sigma) \leq -\underline{\varepsilon} \Leftrightarrow \sigma \geq \bar{\sigma}\exp(\underline{\varepsilon})$.

$\lim_{\varepsilon \to +\infty} g'(\varepsilon) < 0$.[14] These properties are satisfied for a wide range of densities, among which the Normal one.

With these assumptions, it is clear that an unemployed worker can never fully comply or fully fail to comply. As soon as the search requirement $\bar{\sigma}$ is set above zero, the sanction probability rises above zero and will never attain one, irrespectively of how high the actual search intensity σ is set: $0 < p(\bar{\sigma}/\sigma) < 1$, where the limit values 0 and 1 are important to characterize the solution (see Proposition 5). In this benchmark case therefore only the partial compliance regime applies. Consequently, the Bellman equation defining the short-run utility $W_p\left(\sigma, EU^f \mid Ey, \beta\delta\right)$ takes the same form as the one defined by Equation (3.19) for the non-complier in the previous subsection. Only arguments EU_0^f and Ey_0 are replaced by EU^f and Ey, where $EU^f \equiv p\left(\frac{\bar{\sigma}}{\sigma}\right) Z^f + \left(1 - p\left(\frac{\bar{\sigma}}{\sigma}\right)\right) P^f$ and $Ey \equiv p\left(\frac{\bar{\sigma}}{\sigma}\right) y_z + \left(1 - p\left(\frac{\bar{\sigma}}{\sigma}\right)\right) y_b$. P^f denotes the expected lifetime utility of the partially complying future selves after monitoring in the current period has taken place. P^f also replaces U^f in the definition (3.5) of V^f.

The optimization problem of the current self therefore becomes:

$$P^c = \max_{\sigma} W_p\left(\sigma, EU^f \mid Ey, \beta\delta\right) \tag{3.29}$$

Let (σ_p, x_p) designate a solution to this problem. Observe that the decision to comply or not is no longer discrete, but incremental. The degree of compliance is measured by the sanction probability $p\left(\frac{\bar{\sigma}}{\sigma}\right)$. This probability is optimally set by trading off the benefits and costs of marginal increments of the decision variables σ and x. Consequently, since we assume that the first-order conditions define a global maximum, these decision variables are sufficient to characterize the solution. A consequence is also that the constraint $\sigma \geq \bar{\sigma}$ is no longer required in the optimization problem.

In Appendix 3.11, we prove that that:

$$\begin{aligned} W_p\left(\sigma, EU^f \mid Ey, \beta\delta\right) &\equiv y_b - c(\sigma) + \beta\delta\left\{\frac{\lambda(\sigma)Q\left((1-\delta)P^f\right)}{1-\delta(1-q)} + P^f\right\} \\ &\quad - p\left(\frac{\bar{\sigma}}{\sigma}\right)\left\{y_b - y_z + \beta\delta\left[1 - h\left(\sigma, (1-\delta)P^f\right)\right]\left(P^f - Z^f\right)\right\} \end{aligned} \tag{3.30}$$

Using this expression, the first-order conditions of x and σ for a partially complying agent are:

$$R_p(\sigma, x \mid Ey, \bar{\sigma}) \equiv Ey + \frac{\delta\lambda(\sigma)Q(x)}{1-\delta(1-q)} - x - c(\sigma) - p\left(\frac{\bar{\sigma}}{\sigma}\right)\frac{\delta}{(1-\delta)}\left[1 - h(\sigma, x)\right](x - x_z) = 0 \tag{3.31}$$

$$\begin{aligned} S_p(\sigma, x \mid \beta, \bar{\sigma}) \equiv &\frac{\beta\delta\lambda'(\sigma)}{1-\delta(1-q)}Q(x) - \frac{\partial p(\bar{\sigma}/\sigma)}{\partial\sigma}\left\{y_b - y_z + \frac{\beta\delta}{(1-\delta)}\left[1 - h(\sigma, x)\right](x - x_z)\right\} \\ &+ p\left(\frac{\bar{\sigma}}{\sigma}\right)\frac{\beta\delta}{(1-\delta)}\lambda'(\sigma)\bar{F}(x)(x - x_z) - c'(\sigma) = 0 \end{aligned} \tag{3.32}$$

These expressions differ from (3.23)-(3.24), defined in the previous subsection, in that $p\left(\frac{\bar{\sigma}}{\sigma}\right)$ replaces p_0, and that there is an additional marginal returns to job search:

[14] We also implicitly assume that the first-order conditions are not only necessary, but also sufficient conditions for a maximum. These conditions are made explicit in the proof of Proposition 5. They require the natural assumption in particular that $p(\bar{\sigma}/\sigma)$ is concave in σ, or at least not too convex in σ.

$-\frac{\partial p(\bar{\sigma}/\sigma)}{\partial \sigma}\left\{y_b - y_z + \frac{\beta\delta}{(1-\delta)}\left[1 - h(\sigma, x)\right]\ (x - x_z)\right\} > 0$. Indeed, the unemployed agent can reduce the sanction probability by marginally increasing her search effort. The term between braces in (B.13) is the sum of the instantaneous benefit sanction and the discounted capital loss as from the following period. The latter is only incurred if the individual did not find a job before being monitored. This explains the weighting by $[1 - h(\sigma, x)]$. Observe also that this capital loss decreases as $p(\bar{\sigma}/\sigma)$ tends to one, since then $x - x_z$ tends to zero. Note in addition that the sanction probability is now endogenous. Now p_0 is no longer the exogenous parameter, but $\bar{\sigma}$ is. We henceforth assume that any variation in $\bar{\sigma}$ is followed by a less than proportional adjustment in optimal search effort, so that a rise in $\bar{\sigma}$ has a net positive effect on $p(\bar{\sigma}/\sigma)$.

Given (3.27) and since in this benchmark case $\lim\limits_{\varepsilon\to-\infty} g(\varepsilon) = \lim\limits_{\varepsilon\to+\infty} g(\varepsilon) = 0$, it follows that, as in the previous subsection, the first-order conditions (B.12)-(B.13) converge to (3.13)-(3.14) for a perfect monitoring technology if $p(\bar{\sigma}/\sigma)$ converges to zero or to one. In addition, from Proposition 5 below, it is clear that the set of optimal combinations of search effort and reservation wage (σ_p, x_p) in this benchmark case evolves qualitatively in a very similar way to the search requirement $\bar{\sigma}$ (and, hence, with $p(\bar{\sigma}/\sigma)$) as (σ_n, x_n) with p_0 in the previous subsection (cf Proposition 4). This all means that Figure 3.3 which was introduced in the previous subsection also characterizes the solution in this benchmark case with measurement error, but the interpretation differs.

As in the case of perfect monitoring, point O characterizes the solution (σ_b, x_b) if no job search requirement is imposed. As soon as the requirement is raised above zero, the benefit claimant faces a strictly positive sanction probability even if she *actually* complies with the requirement: $\sigma_b > \bar{\sigma}$. As a consequence of the measurement error, the individual can indeed only comply partially. This is a major difference with Subsection 3.5.1, where in particular Figure 3.3 described the adjustment of the non-complying agent only. As explained below, the OL line in Figure 3.3 now shows how the behavior of the procrastinator qualitatively varies when $\bar{\sigma}$ rises, such that $p(\bar{\sigma}/\sigma)$ increases from zero to one. When $\bar{\sigma}$ is raised above zero, search effort is enhanced. The reservation wage as well, but at a lower rate than in the perfect monitoring case, because the unemployed cannot comply perfectly, so that the reservation wage (and, hence, long-term utility) is strictly lower than $r(\sigma \mid y_b)$. As the search requirement is further increased, the optimal solution shifts further to the right along the OV line segment. At some point, the cost of the higher search intensity and the higher sanction probability start to dominate the gain in long-run utility that enhanced search effort induces for an impatient individual, and the reservation wage starts to fall. At point V, search effort attains its maximum. Beyond that point, the additional marginal returns to job search reflected in the first-order conditions (B.12)-(B.13) start to decrease because (1) the sanction probability becomes so high that the expected capital gain, which is proportional to $(x - x_z)$, eventually falls so much (it tends to zero for $p(\bar{\sigma}/\sigma) \to 1$) that it dominates the gain from the decrease in the sanction probability induced by the enhanced search effort, and (2) the sanction probability eventually decreases in σ at a decreasing rate, so that the additional returns, which are proportional to $-\partial p(\bar{\sigma}/\sigma)/\partial\sigma$, eventually fall.[15] From point V, the solution then converges to L,

[15]This is because $\frac{\partial^2 p}{[\partial \sigma]^2} = \frac{g\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right] + g'\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right]}{\sigma^2} < 0$ if $\bar{\sigma} \to \infty$, since we assume that $\lim\limits_{\varepsilon\to+\infty} g(\varepsilon) = 0$ and

the free choice of a sanctioned individual, but never attains this point unless the unemployed agent is actually sanctioned.

Observe that in this benchmark case with measurement error, the unemployed job benefit claimant always faces a strictly positive sanction probability. Consequently, for any level of σ, the long-run utility is always strictly lower than in the case of perfect monitoring. This means that, irrespectively of preferences or job search technology, the Pareto-efficient frontier AH can no longer be attained. This demonstrates that the social benefit of imposing job search requirements on impatient unemployed benefit claimants is further reduced if search effort is measured with error.

> **Proposition 5.** *If the sanction probability is expressed by* (3.28) *and the support of measurement error is unbounded, then both the optimal search effort σ_p and reservation wage x_p strictly increase (resp., decrease) in $\bar{\sigma}$ for $\bar{\sigma} \to 0$ and, hence, $p(\bar{\sigma}/\sigma_p) \to 0$ (resp., $\bar{\sigma} \to +\infty$ and, hence, $p(\bar{\sigma}/\sigma_p) \to 1$). Furthermore, the Pareto-efficient frontier AH can no longer be attained.*

Proof. See Appendix 3.13. □

3.5.2.2 The Case of a Bounded Support for Measurement Error

We now turn to the case in which the support of measurement error is bounded. This is an interesting case, because its solution exhibits features of both the perfect monitoring case and the benchmark with measurement error, which are both limiting cases: perfect monitoring is attained if both $\underline{\varepsilon} \to 0$ and $\overline{\varepsilon} \to 0$, while in the benchmark both $\underline{\varepsilon} \to \infty$ and $\overline{\varepsilon} \to \infty$. We will demonstrate that in this case perfect compliance and non-compliance are again possibilities, alongside with partial compliance.

To ensure that the solution does not differ too much from the benchmark case, we assume in addition to the bounded support that $\forall \bar{\sigma} > 0,\ \forall \sigma \in [\bar{\sigma} \exp(-\overline{\varepsilon}), \bar{\sigma} \exp(\underline{\varepsilon})] : g\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right] > 0$, $g'(-\underline{\varepsilon}) \geq 0$, and $g'(\overline{\varepsilon}) \leq 0$. These assumptions are satisfied if the distribution of measurement error is for instance Uniform (in which case $g'(-\underline{\varepsilon}) = g'(\overline{\varepsilon}) = 0$) or truncated Normal (in which case $g'(-\underline{\varepsilon}) > 0$ and $g'(\overline{\varepsilon}) < 0$). With these assumptions, Proposition 5 applies, except for the fact that the reservation wage may be decreasing in $\bar{\sigma}$ for a sanction probability close to zero ($\sigma \to \bar{\sigma} \exp(\underline{\varepsilon})$) if $g(-\underline{\varepsilon})$ is sufficiently large.[16]

The assumptions that $g\left(-\underline{\varepsilon}\right) > 0$ and $g\left(\overline{\varepsilon}\right) > 0$ have consequences on the solution when the probability of being sanctioned tends to zero and to one. This follows from the first-order condition (B.13) where the additional marginal returns to job search effort proportional to $-\partial p(\bar{\sigma}/\sigma)/\partial\sigma = g\left[\log\left(\bar{\sigma}/\sigma\right)\right]/\sigma$ remain strictly positive in this case, even if the sanction probability tends to zero or to one. Consequently, when $\sigma \to \bar{\sigma} \exp(\underline{\varepsilon})$ so that $p(\bar{\sigma}/\sigma) \to 0$, the solution in the case of

$\lim_{\varepsilon \to +\infty} g'(\varepsilon) < 0$.

[16]Sufficiently large in a sense defined in the proof of Proposition 5, where the case of a bounded support of measurement error is explicitly discussed.

partial compliance (σ_p, x_p) no longer converges to the solution (σ_b, x_b). At the other extreme, when $\sigma \to \bar{\sigma}\exp(-\bar{\varepsilon})$ so that $p(\bar{\sigma}/\sigma) \to 1$, (σ_p, x_p) no longer converges to (σ_z, x_z). This means that in these limiting cases, search effort must be strictly greater than σ_b, respectively σ_z. Moreover, notice that the first-order condition for the reservation wage (B.12) is not affected in these limiting cases. In Figure 3.3, these results imply that the solution converges to point W to the right of O on the $x = r(\sigma \mid y_b)$ curve, as search effort converges to $\bar{\sigma}\exp(\underline{\varepsilon})$. Similarly, the solution converges to point X to the right of point L on the $x = r(\sigma \mid y_z)$ curve, as search effort converges to $\bar{\sigma}\exp(-\bar{\varepsilon})$.

Let us now describe how the solution evolves as the search requirement is increased. If the search requirement is raised above zero, the search effort σ_b is initially unaffected as long as σ_b is greater than the *actual* search requirement $\bar{\sigma}\exp(\underline{\varepsilon})$. As soon as the actual search requirement exceeds σ_b, the unemployed will choose between full and partial compliance depending which yields the highest short-term utility. Based on the reasoning in the previous paragraph, we know, however, that for a sanction probability close to zero in the case of partial compliance, the optimal search level σ_p must be be strictly higher than σ_b, so that the benefit claimant fully complies for a range of actual search requirements above σ_b up to the abscissa of point W in Figure 3.3. If the search effort is set marginally higher than the latter, the short-run utility in the case of partial compliance strictly exceeds the utility in the case of full compliance. Therefore, from point W, the benefit claimant starts complying partially. Subsequently, the optimal search effort σ_p qualitatively evolves with $\bar{\sigma}$ as in the benchmark. It continues to increase at first up to point V' after which it declines up to point X. At this point, the sanction probability attains one and the short-run utility of a partial complier equals that of a sanctioned individual. If then $\bar{\sigma}$ is further increased, the optimal search effort discontinuously falls to σ_z at point L, which is the optimal solution for a sanctioned individual.

Notice that as the support of measurement error is reduced, the abscissas of points W and X will move closer together, and V' will also lie less to the right of these points. In the limit, when both $\underline{\varepsilon} \to 0$ and $\bar{\varepsilon} \to 0$, the abscissas of these three points coincide and the solution corresponds to the one in case of perfect monitoring. This means that the abscissa of these points becomes $\bar{\sigma}^s$ in Figure 3.1. If the search requirement increases above this level, search effort falls to σ_z (at point L). If, by contrast, the support of measurement error is further widened, $g(-\underline{\varepsilon})$ and $g(\bar{\varepsilon})$ must eventually decrease, so that W tends to O and X tends to L. In the limit, when both $\underline{\varepsilon} \to \infty$ and $\bar{\varepsilon} \to \infty$, these points will coincide as they did in the benchmark in which measurement error had unbounded support.

From the aforementioned analysis it is clear that in the case of a measurement error with bounded support, we cannot in general exclude that one can attain the Pareto-efficient segment AH on the $x = r(\sigma \mid y_b)$ curve by imposing a search requirement on impatient benefit claimants. Indeed, we cannot exclude in general that W is located to the right of A. From the discussion it is clear that this depends in part on the magnitude of the density of the measurement error on the lower bound of the support. One may expect that the magnitude of this density is roughly inversely proportional to the width of the support of measurement error,[17] so that loosely speaking the more precisely search effort is measured, the likelier the social efficiency of search requirement. In the limiting case of an unbounded support, social efficiency cannot be attained. Since search effort is not directly observable, it is difficult to measure precisely, so that the latter hypothesis may not be so far from reality.

3.6 Monitoring Search versus Job Search Assistance

In this section, we briefly evaluate the social efficiency of imposing job search requirements on procrastinating unemployed benefit claimants relative to job search assistance (JSA). To keep the analysis simple, we here assume that the monitoring is perfect. JSA typically raises the returns to job search by increasing search efficiency (i.e. by increasing $\lambda'(\sigma)$ for any given level of σ) and/or by reducing the marginal cost of search (i.e. by reducing $c'(\sigma)$ for any given level of σ). This shifts the $x = s(\sigma \mid \beta)$ curve in Figure 3.1 to the right and thereby raises job search effort closer to the Pareto frontier AH. As such, the effects of JSA resemble those of imposing a search requirement. However, JSA also reduces the search cost level and/or increases the level of the job arrival rate. Consequently, the $x = r(\sigma \mid y_b)$ curve shifts at the same time upwards. It can be shown that both search effort and reservation wage increase.

This ignores that JSA is costly. A way to make these implementation costs apparent in our analysis is to assume that the unemployed are charged for these expenses. As a consequence (see the first-order condition), the current income y_b decreases, inducing the $x = r(\sigma \mid y_b)$ curve to shift downwards. Since the implementation costs are typically higher than the reduction in the search costs for the job seeker, this downward shift will typically dominate the aforementioned upward shift of the $x = r(\sigma \mid y_b)$ curve. Moreover, the marginal implementation cost per unit of search effort of a monitoring job search in the case of perfect monitoring technology are negligible (assumed zero in the analysis) relative to those of JSA,[18] so that at first sight monitoring job search seems socially more efficient than JSA.

Our analysis shows, however, that such a conclusion is deceptive. First, it does not take into account the fact that by imposing a job search requirement, a procrastinating benefit claimant may decide to stop complying at lower levels of search effort than an unemployed agent with exponential time preferences, and that in that case both welfare and search effort may eventually be much lower than they would have been had an alternative policy, such as JSA, been implemented. Second, this analysis ignores that job search effort is measured imprecisely. This is a hidden cost that induces the $x = r(\sigma \mid y_b)$ to shift downwards more than what may be expected based on the negligible directly observable implementation costs of a monitoring scheme.

A fair comparison of the two policies requires, however, that imperfections in JSA be also

[17] It is exactly inversely proportional in the case of a Uniform distribution.

[18] Notice that monitoring exhibits increasing returns to scale, since the cost of verifying search is close to fixed, so that the cost of increasing the search requirement decreases at an increasing rate. By contrast, the implementation costs of JSA typically exhibit decreasing returns.

taken into account. For JSA to work, the unemployed job seeker should be willing to devote time to this assistance, which could take the form of classroom training, counseling sessions, help to write resumes, etc. This therefore imposes, as job search effort, instantaneous costs and delayed rewards. A job seeker with hyperbolic preferences is therefore likely to underinvest in the offered JSA. That is why JSA may have to be imposed to make it work. Obviously, in this case, the aforementioned imperfections apply. Nevertheless, an important difference is that participation in JSA is much easier to verify than job search effort. Hence, measurement error is much smaller and this, as we have shown, increases the likelihood of perfect compliance. In any case, determining which policy is more efficient is clearly an empirical issue.

3.7 Conclusion

Given the size of the insured unemployed population and the financial pressure on public budgets, many governments have intensified the monitoring of unemployed benefit recipients' job search effort. If unemployed workers have exponential time preferences, this intensification does not unambiguously enhance social welfare, since, even if it raises the exit rate from unemployment and thereby reduces outlays on benefit payments, it hurts the unemployed, both by the higher effort level that this monitoring imposes and by the lower expected job quality, because reservation wages are reduced. However, if the unemployed procrastinate, DellaVigna and Paserman (2005) and Paserman (2008) have shown that imposing job search requirements may even be Pareto improving, at least if the unemployed are *sophisticated* procrastinators who realize that they exert too little search effort if they are not monitored. In this research, we have first confirmed this conclusion, based on a graphical analysis that aims at clarifying this result. In addition, we have argued that, even if the utility of the current self does not increase in the case when the unemployed are *naive* procrastinators, social welfare may unambiguously increase, if, as argued in the literature, the welfare of the unemployed is evaluated according to a social preference function that disregards the biases induced by self-control and misperception problems.

These conclusions with regards to the optimality of a monitoring scheme for hyperbolic discounting unemployed benefit recipients are, however, inconsistent with empirical studies that report the ambiguous effects of the monitoring of job search effort on the job finding rate and on job quality, and find that this monitoring may sometimes induce exits out of the labor force and, hence, lead to lower welfare levels for the unemployed. We have argued in this paper that this mixed evidence can nevertheless be compatible with the hypothesis that the unemployed procrastinate. This is because the aforementioned optimality results are derived under two strong and unrealistic assumptions. First, the unemployed can only exit towards employment. In reality, they can in particular withdraw from the claimant register and live from welfare benefits, or from the help of family members or charities. Taking this alternative into account, and assuming that job search effort is no longer monitored then, we have shown that social efficiency may no longer be attainable. Raising job search requirements to a level that is optimal from the perspective of the future selves or from the perspective of society may therefore eventually lead to a sub-optimal level of search effort and a long-run utility that is even lower than in the absence of job search requirements. A second oversimplifying assumption is that monitoring is perfect. Imperfections can be induced by caseworker discretion or measurement error. In any case, the utility of non-complying then increases, because the imperfection induces a strictly positive probability that the sanction is not imposed. Moreover, in the case of measurement error, the welfare of a complier is reduced, because she may still face a strictly positive sanction probability even if *actual* search effort satisfies the search requirement. Therefore, an imperfect monitoring technology raises the likelihood of non- or partial compliance, and, hence, lowers the likelihood that social efficiency is attained by imposing search requirements on procrastinating unemployed benefit claimants. If in particular whatever the search effort level, there still is a non-zero risk of being sanctioned, our analysis demonstrates that social efficiency cannot be achieved then. Consequently, even if the unemployed procrastinate, other policy instruments, such as job search assistance, could improve on monitoring schemes from a social welfare point of view. Whether this is indeed the case remains to be determined in future empirical research.

This analysis has been conducted in a stationary framework. In reality, unemployment benefits expire after a finite duration. Everything else equal, this reduces the intertemporal value of claiming benefits as the time limit draws closer, and hence the value of complying with search effort requirements. Consequently, this reduces the efficiency of imposing a job search requirement, as it not only decreases the welfare of the unemployed, but it also lowers the likelihood that the benefit claimant complies with the requirement, at least if it is set above the search effort level that an unemployed individual would freely choose (the latter being at most the one chosen by a sanctioned agent). This suggests that monitoring job search should rather be considered as a *substitute* for a UI scheme with a time limit than as a *complementary* policy. Such a comparison remains for further research. A different source of non-stationarity appears if the monitoring of search effort no longer takes place at the end of each time period, but at some future time which is known in advance, as in the analysis of Cockx et al. (2011) for unemployed agents with exponential time preferences. Looking at the behavior of procrastinators in such an environment is however outside the scope of this paper.

Acknowledgements

We wish to thank Mathias Hungerbühler, Dan Silverman (the editor), and two anonymous referees for their valuable comments and discussions. We are also very grateful for the comments received from the participants to the seminars at CORE (IUAP/PAI Public Economics Day), KULeuven (Day for Labour Economists), Ghent University (internal seminar), Louvain-la-Neuve (Doctoral Workshops) and the EALE annual conference in Bonn. Corinna Ghirelli acknowledges financial support from the Special Research Fund ("BOF") of Ghent University, as well as from the National Bank of Belgium (project code 05Y00110). The usual disclaimer applies.

3.8 Appendix A. Derivation of $W\left(\sigma, U_j^f \mid y_u, \beta\delta\right)$ in Equation (3.9)

Proof. Using the reservation wage property, the expectation term in (3.1) can be expanded to give:

$$W\left(\sigma, U^f \mid y_u, \beta\delta\right) = y_u - c(\sigma) + \beta\delta\left\{\lambda(\sigma)\left[\int_{\underline{w}}^{(1-\delta)U^f} U^f dF(w) + \int_{(1-\delta)U^f}^{\bar{w}} V^f dF(w)\right] + (1-\lambda(\sigma))U^f\right\}$$

Substituting V^f from (3.5), the previous equality becomes:

$$W\left(\sigma, U^f \mid y_u, \beta\delta\right) = y_u - c(\sigma) + \beta\delta\left\{\lambda(\sigma)\int_{(1-\delta)U^f}^{\bar{w}} \frac{w-(1-\delta)U^f}{1-\delta(1-q)} dF(w) + U^f\right\} \quad (3.33)$$

which is (3.9). □

3.9 Appendix B. Proof of Proposition 1

Proof. In (3.14), shift $c'(\sigma)$ to the right-hand-side. Then, take the log of both sides of this rearranged condition. Next, log differentiate this equality taking into account that $x = s(\sigma \mid \beta)$. This immediately proves that:

$$\forall\sigma \in \mathbb{R}^+ : \frac{\partial \log s(\sigma \mid \beta)}{\partial \log \sigma} = \left[\frac{\partial \log Q\left[s(\sigma \mid \beta)\right])}{\partial \log s(\sigma \mid \beta)}\right]^{-1}\left(\frac{d\log c'(\sigma)}{d\log\sigma} - \frac{d\log\lambda'(\sigma)}{d\log\sigma}\right) < 0.$$

Given that $x = r(\sigma \mid y_u)$, partially differentiating (3.13) with respect to σ, and using the definition of $S\left(\sigma, r(\sigma \mid y_u) \mid \beta\right)$ in (3.14) yields:

$$S\left(\sigma, r(\sigma \mid y_u) \mid 1\right) + \left[\frac{\delta\lambda(\sigma)}{1-\delta(1-q)}Q'\left[r(\sigma \mid y_u)\right] - 1\right]\frac{\partial r(\sigma \mid y_u)}{\partial\sigma} = 0 \quad (3.34)$$

From (3.10), $Q'\left[r(\sigma \mid y_u)\right] = -\bar{F}\left[r(\sigma \mid y_u)\right]$. Inserting this in (3.34) and rearranging gives:

$$\frac{\partial r(\sigma \mid y_u)}{\partial\sigma} = S\left[\sigma, r(\sigma \mid y_u) \mid 1\right]\frac{1-\delta(1-q)}{\left[1-\delta(1-q)+\delta h\left[\sigma, r(\sigma \mid y_u)\right]\right]}, \quad (3.35)$$

where $h\left[\sigma, r(\sigma \mid y_u)\right] \equiv \lambda(\sigma)\bar{F}\left[r(\sigma \mid y_u)\right]$ is the probability of leaving unemployment ($\bar{F}[.] \equiv 1 - F[.]$). Since the second term of the product is strictly positive, $\frac{\partial r(\sigma|y_u)}{\partial\sigma}$ has the same sign as $S\left(\sigma, r(\sigma \mid y_u) \mid 1\right)$. As this corresponds to the first-order condition of search effort for an exponential agent, we have that $S\left(\sigma_u^e, r(\sigma_u^e \mid y_u) \mid 1\right) = 0$ and is positive (negative) for $\sigma < \sigma_u^e$ ($\sigma > \sigma_u^e$). So,

$$\forall\sigma \lesseqgtr \sigma_u^e : \frac{\partial r(\sigma \mid y_u)}{\partial\sigma} \gtreqless 0. \quad (3.36)$$

Given that $x = s(\sigma \mid \beta)$, partially differentiating (3.14) with respect to β and x, and recalling that $Q'\left[r(\sigma \mid y_u)\right] < 0$ yields

$$\frac{\partial s(\sigma \mid \beta)}{\partial\beta} = -\frac{Q\left[r(\sigma \mid y_u)\right]}{\beta Q'\left[r(\sigma \mid y_u)\right]} > 0 \quad (3.37)$$

It is obvious that $s(\sigma \mid \beta)$ is independent of y_u. Given that $x = r(\sigma \mid y_u)$, partially differentiating (3.13) with respect to y_u, and following the same steps as earlier in this proof yields:

$$\frac{\partial r(\sigma \mid y_u)}{\partial y_u} = \frac{[1-\delta(1-q)]}{[1-\delta(1-q)+\delta h\left[\sigma, r(\sigma \mid y_u)\right]]} > 0 \tag{3.38}$$

□

3.10 Appendix C. Proof of Proposition 2

First we prove the inverse U-shaped profile of $B^c(\bar{\sigma})$:

$$\exists\, \sigma^* \in (\sigma_b, \sigma_b^e) \text{ such that } \forall \bar{\sigma} \lesseqqgtr \sigma^* \text{ and } \bar{\sigma} > \sigma_b : \frac{\partial B^c(\bar{\sigma})}{\partial \bar{\sigma}} \gtreqqless 0.$$

Knowing this, it is obvious that raising the search requirement $\bar{\sigma}$ somewhat above σ_b will improve the short-run utility of the claimant. Above $\bar{\sigma} = \sigma^*$, the short-run utility shrinks. Second, we show that the level of search at which the claimant stops complying, $\bar{\sigma}^s$, lies in the following open interval: $(\max\{\sigma_z, \tilde{\sigma}\}, \bar{\sigma}^s_{max})$. If the search requirement $\bar{\sigma}$ is set higher than $\bar{\sigma}^s$, the agent withdraws from insured unemployment and chooses the pair (σ_z, x_z).

Proof. 1. As $\partial B^c/\partial \bar{\sigma} = 0$ if the free choice σ_b is strictly greater than $\bar{\sigma}$, let us focus on the case where $\bar{\sigma} \geq \sigma_b$. Partially differentiating the left-hand side of (B.26) with respect to $\bar{\sigma}$ yields:

$$\frac{\partial B^c(\bar{\sigma})}{\partial \bar{\sigma}} = S\left(\bar{\sigma}, r(\bar{\sigma} \mid y_b) \mid \beta\right) + \beta\delta \frac{\left[(1-\delta)\left(1-h\left[\bar{\sigma}, r(\bar{\sigma} \mid y_b)\right] + \delta q\right)\right]}{[1-\delta(1-q)](1-\delta)} \frac{\partial r(\bar{\sigma} \mid y_b)}{\partial \bar{\sigma}}$$

First, consider $\bar{\sigma} = \sigma_b$. Then, $\partial B^c(\bar{\sigma})/\partial \bar{\sigma} > 0$, since $S\left(\sigma_b, r(\sigma_b \mid y_b) \mid \beta\right) = 0$ and, by (3.36), $\partial r(\bar{\sigma} \mid y_b)/\partial \bar{\sigma} > 0$ when $\bar{\sigma} < \sigma_b^e$. Now, consider $\bar{\sigma} = \sigma_b^e$. Then, $\partial B^c(\bar{\sigma})/\partial \bar{\sigma} < 0$, since $S\left(\sigma, r(\sigma \mid y_b) \mid \beta\right) < 0$ for all $\sigma > \sigma_b$ and, by (3.36), $\partial r(\bar{\sigma} \mid y_b)/\partial \bar{\sigma} = 0$ if $\bar{\sigma} = \sigma_b^e$. As $\partial B_s^c(\bar{\sigma})/\partial \bar{\sigma}$ is a continuous function, the following proposition must hold: $\exists\, \sigma^* \in (\sigma_b, \sigma_b^e)$ such that $\forall \bar{\sigma} \lesseqqgtr \sigma^*$ and $\bar{\sigma} > \sigma_b : \frac{\partial B^c(\bar{\sigma})}{\partial \bar{\sigma}} \gtreqqless 0$.

2. The second part of the proof is divided into four steps:

   (a) Using (3.2), (3.6), (3.8) and (3.9), and restricting the analysis to the cases where $\bar{\sigma} \geq \sigma_b$, we obtain

$$\forall \bar{\sigma} : B^f(\bar{\sigma}) - B^c(\bar{\sigma}) = \frac{\delta(1-\beta)}{[1-\delta(1-q)]} \left\{ \lambda(\bar{\sigma}) Q(r(\bar{\sigma}|y_b)) + \frac{[1-\delta(1-q)] r(\bar{\sigma}|y_b)}{(1-\delta)} \right\} > 0 \tag{3.39}$$

   Following a similar reasoning, we find

$$Z^f - Z^c = \frac{\delta(1-\beta)}{[1-\delta(1-q)]} \left\{ \lambda(\sigma_z) Q(x_z) + \frac{[1-\delta(1-q)] x_z}{(1-\delta)} \right\} > 0. \tag{3.40}$$

(b) To transform $Q(x_z)$ in the last expression, consider definition (3.10) and assume two values x_1, x_2 of the argument of Q such that $x_1 < x_2$. $Q(x_1)$ can be rewritten as follows:

$$\begin{aligned} Q(x_1) &= \int_{x_1}^{x_2} [w - x_1]\, dF(w) + Q(x_2) + \bar{F}(x_2)\,[x_2 - x_1] \\ &= Q(x_2) + \bar{F}(x_1)\,[x_2 - x_1] - \{x_2 - x_1 - E\,[w - x_1 | x_1 \leq w < x_2]\} \\ &\times \left[\bar{F}(x_1) - \bar{F}(x_2)\right] \end{aligned} \quad (3.41)$$

Recall that $\bar{\sigma}^s_{max}$ is defined by $B^f(\bar{\sigma}^s_{max}) = Z^f$. Since $\forall \bar{\sigma} \in [\sigma_z, \bar{\sigma}^s_{max}) : r(\bar{\sigma}|y_b) > x_z$, we can apply the above transformation and rewrite (3.40) as:

$$Z^f - Z^c = \frac{\delta(1-\beta)}{[1-\delta(1-q)]}\Big\{\lambda(\sigma_z)Q(r(\bar{\sigma}|y_b)) + h(\sigma_z, x_z)[r(\bar{\sigma}|y_b) - x_z] + x_z\frac{1-\delta+\delta q}{1-\delta}$$
$$-\lambda(\sigma_z)\{r(\bar{\sigma}|y_b) - x_z - E[w - x_z|x_z \leq w < r(\bar{\sigma}|y_b)]\}[\bar{F}(x_z) - \bar{F}(r(\bar{\sigma}|y_b))]\Big\} \quad (3.42)$$

(c) Subtracting (B.9) from (3.39) then yields for any $\bar{\sigma} \in [\sigma_z, \bar{\sigma}^s_{max})$:

$$B^f(\bar{\sigma}) - B^c(\bar{\sigma}) - Z^f + Z^c = \frac{\delta(1-\beta)}{[1-\delta(1-q)]}\Big\{Q[r(\bar{\sigma}|y_b)][\lambda(\bar{\sigma}) - \lambda(\sigma_z)]$$
$$+[r(\bar{\sigma}|y_b) - x_z] \times \frac{[(1-\delta)(1-h(\sigma_z, x_z)) + \delta q]}{(1-\delta)}$$
$$+\lambda(\sigma_z)\{r(\bar{\sigma}|y_b) - x_z - E[w - x_z|x_z \leq w < r(\bar{\sigma}|y_b)]\}[\bar{F}(x_z) - \bar{F}(r(\bar{\sigma}|y_b))]\Big\} > 0$$

Therefore,

$$\forall \bar{\sigma} \in [\sigma_z, \bar{\sigma}^s_{max}) : B^f(\bar{\sigma}) - Z^f > B^c(\bar{\sigma}) - Z^c \quad (3.43)$$

Since $B^f(\bar{\sigma}^s_{max}) = Z^f$, (3.43) implies that $B^c(\bar{\sigma}^s_{max}) < Z^c$. So, $\bar{\sigma}^s_{max}$ is an upper bound of the interval of search requirements that do not induce a withdrawal from the register of the insured unemployed. Because, $B^c(\sigma_z) > Z^c$ (see the discussion after (B.26)) and $B^c(\cdot)$ is a continuous function, it must be that $\sigma_z < \bar{\sigma}^s < \bar{\sigma}^s_{max}$.

(d) Let $\tilde{\sigma}$ make the constrained current self as well off as the unconstrained one, i.e. $B^c(\tilde{\sigma}) = B^c$. As $B^c > Z^c$, it follows that $B^c(\tilde{\sigma}) > Z^c$.

Therefore, $\bar{\sigma}^s > \max\{\sigma_z, \tilde{\sigma}\}$.

□

3.11 Appendix D. Derivation of $W_p\left(\sigma, EU^f \mid Ey, \beta\delta\right)$

Proof. In this proof p is generic and stands either for p_0 or for $p\left(\frac{\bar{\sigma}}{\sigma}\right)$. In the first case $U^f = N^f$ and in the second $U^f = P^f$. Using the reservation wage property, the expectation term in (3.1) can be expanded to give:

$$W_p\left(\sigma, EU^f \mid Ey, \beta\delta\right) = Ey - c(\sigma) + \beta\delta\left\{\lambda(\sigma)\left[\int_{\underline{w}}^{(1-\delta)U^f} EU^f dF(w) + \int_{(1-\delta)U^f}^{\bar{w}} V^f dF(w)\right]\right.$$
$$\left. +(1-\lambda(\sigma))EU^f\right\} \quad (3.44)$$

Substituting (3.5) in this equation yields:

$$W_p\left(\sigma, EU^f \mid Ey, \beta\delta\right) = Ey - c(\sigma) + \beta\delta\left\{\lambda(\sigma)\left[\int_{(1-\delta)U^f}^{\bar{w}}\left[\frac{w+q\delta U^f}{1-\delta(1-q)} - EU^f\right]dF(w)\right] + EU^f\right\} \tag{3.45}$$

Substituting $EU^f = U^f - p(U^f - Z^f)$ and $Ey = y_b - p(y_b - y_z)$ then results in:

$$\begin{aligned} W_p\left(\sigma, EU^f \mid Ey, \beta\delta\right) &= y_b - c(\sigma) + \beta\delta\left\{\frac{\lambda(\sigma)Q\left[(1-\delta)U^f\right]}{1-\delta(1-q)} + U^f\right\} \\ &- p\left\{y_b - y_z + \beta\delta\left[1 - h\left[\sigma, (1-\delta)U^f\right]\right](U^f - Z^f)\right\} \end{aligned} \tag{3.46}$$

where $h\left[w, (1-\delta)U^f\right] \equiv \lambda(\sigma)\bar{F}\left[(1-\delta)U^f\right]$. □

3.12 Appendix E. Proof of Proposition 4

Proof. This proposition is essentially a special case of Proposition 5 (next appendix) in which $p = p_0$, all derivatives of p with respect to σ and $\bar{\sigma}$ are set to zero, p_0 replaces $\bar{\sigma}$ as exogenous variable, $r_n(\sigma \mid Ey_0, p_0)$ replaces $r_p(\sigma \mid Ey, \bar{\sigma})$ and $s_n(\sigma \mid \beta, p_0)$ $s_p(\sigma \mid \beta, \bar{\sigma})$. This means that a_1 to a_4 and D have the same signs as in Proposition 5. However, a_5 and a_6 should be replaced respectively by:

$$-\frac{\partial R_n(\sigma \mid Ey_0, p_0)}{\partial p_0} = y_b - y_z + \frac{\delta(1-h)(x - x_z)}{1-\delta} > 0 \tag{3.47}$$

$$-\frac{\partial S_n(\sigma \mid \beta, p_0)}{\partial p_0} = -\frac{\beta\delta\lambda'\bar{F}(x)(x - x_z)}{1-\delta} \leq 0 \tag{3.48}$$

where the last derivative tends to zero if p_0 tends to one. This means that $-\frac{\partial R_n(\sigma|Ey_0,p_0)}{\partial p_0}$ has the same sign as a_5 in the proof of Proposition 5, but $-\frac{\partial S_n(\sigma|\beta,p_0)}{\partial p_0}$ has the same sign as a_6 only for p_0 sufficiently close to zero. In terms of the signs of the derivatives of the optimal choices σ_n and x_n with respect to p_0 the only consequence is that in this case $\frac{\partial x_n}{\partial p_0}$ is now ambiguous instead of strictly increasing for $p_0 \to 0$. □

3.13 Appendix F. Proof of Proposition 5

Proof. The proof does not only derive the partial derivatives of the optimal choices σ_p and x_p with respect to $\bar{\sigma}$, but also the partial derivatives of $r_p(\sigma \mid Ey, \bar{\sigma})$ and $s_p(\sigma \mid \beta, \bar{\sigma})$ with respect to σ and $\bar{\sigma}$. We provide this information, since, as in the perfect monitoring case, the optimal choices for any given $\bar{\sigma}$ are determined by the intersection between the $x = r_p(\sigma \mid Ey, \bar{\sigma})$ and $x = s_p(\sigma \mid \beta, \bar{\sigma})$ curves. These partial derivatives allow to determine the slope of these curves for different values of $\bar{\sigma}$ and in which direction these curves are shifted for marginal changes in $\bar{\sigma}$. It allows in particular to determine that the $x = s_p(\sigma \mid \beta, \bar{\sigma})$ curve has a negative slope for $\bar{\sigma} \to 0$ (and, hence, $p(\bar{\sigma}/\sigma) \to 0$) and shifts upwards if $\bar{\sigma}$ is marginally raised close to that point, and that it has, for $\bar{\sigma} \to +\infty$ (and, hence, $p(\bar{\sigma}/\sigma) \to 1$), a positive slope and is shifted downwards if it is marginally *lowered* close to that the point. For the purpose of not overloading the figures and the discussion in the main text, we neither displayed these curves in the graphical analysis, nor mentioned them in the main text.

The first-order conditions (B.12) and (B.13) define a system of two equations in two unknowns as a function of $\bar{\sigma} > 0$. Totally differentiating this system yields:

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} dx \\ d\sigma \end{bmatrix} = \begin{bmatrix} a_5 \\ a_6 \end{bmatrix} d\bar{\sigma} \tag{3.49}$$

where, without recalling the arguments of $\lambda, c, p, f, \bar{F}, Q, h$ and their derivatives,

$$\begin{aligned}
a_1 &= -\frac{\delta\lambda\bar{F}}{1-\delta(1-q)} - 1 - p\frac{\delta}{1-\delta}\left[\lambda f(x-x_z) + 1 - h\right] < 0 && (3.50)\\
a_2 &= \frac{\delta\lambda' Q}{1-\delta(1-q)} - c' - \frac{\partial p}{\partial\sigma}\left\{y_b - y_z + \frac{\delta(1-h)(x-x_z)}{1-\delta}\right\} + p\frac{\delta\lambda'\bar{F}(x-x_z)}{1-\delta} \\
&= S_p(\sigma, x \mid 1, \bar{\sigma}) && (3.51)\\
a_3 &= -\frac{\beta\delta\lambda'\bar{F}}{1-\delta(1-q)} - \frac{\partial p}{\partial\sigma}\frac{\beta\delta}{1-\delta}\left[\lambda f(x-x_z) + 1 - h\right] \\
&+ p\frac{\beta\delta}{1-\delta}\lambda'\left[\bar{F} - f(x-x_z)\right] && (3.52)\\
a_4 &= \frac{\beta\delta\lambda'' Q}{1-\delta(1-q)} - \frac{\partial^2 p}{[\partial\sigma]^2}\left\{y_b - y_z + \frac{\beta\delta(1-h)(x-x_z)}{1-\delta}\right\} + 2\frac{\partial p}{\partial\sigma}\frac{\beta\delta\lambda'\bar{F}(x-x_z)}{1-\delta} \\
&+ p\frac{\beta\delta\lambda''\bar{F}(x-x_z)}{1-\delta} - c'' && (3.53)\\
a_5 &= \frac{\partial p}{\partial\bar{\sigma}}\left\{y_b - y_z + \frac{\delta(1-h)(x-x_z)}{1-\delta}\right\} && (3.54)\\
a_6 &= \frac{\partial^2 p}{\partial\sigma\partial\bar{\sigma}}\left\{y_b - y_z + \frac{\beta\delta(1-h)(x-x_z)}{1-\delta}\right\} - \frac{\partial p}{\partial\bar{\sigma}}\frac{\beta\delta\lambda'\bar{F}(x-x_z)}{1-\delta} && (3.55)
\end{aligned}$$

Solving system (B.12) and (B.13) gives

$$\begin{bmatrix} \frac{\partial x}{\partial\bar{\sigma}} \\ \frac{\partial\sigma}{\partial\bar{\sigma}} \end{bmatrix} = \frac{1}{D}\begin{bmatrix} a_4 & -a_2 \\ -a_3 & a_1 \end{bmatrix}\begin{bmatrix} a_5 \\ a_6 \end{bmatrix} \tag{3.56}$$

where $D \equiv a_1 a_4 - a_2 a_3$.

In order to sign these partial derivatives, we make a couple of observations. First, using the definition of p in (3.28), we have that on the interior of its support:

$$\begin{aligned}
\frac{\partial p}{\partial\bar{\sigma}} &= \frac{g\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right]}{\bar{\sigma}} = -\frac{\partial p}{\partial\sigma}\frac{\sigma}{\bar{\sigma}} \geq 0 && (3.57)\\
\frac{\partial^2 p}{\partial\sigma^2} &= \frac{g\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right] + g'\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right]}{\sigma^2} && (3.58)\\
\frac{\partial^2 p}{\partial\sigma\partial\bar{\sigma}} &= -\frac{g'\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right]}{\sigma\bar{\sigma}} && (3.59)
\end{aligned}$$

By inserting (3.57) in (3.54), it follows that $a_5 \geq 0$. Also, by inserting (3.58) into (3.53), it can be seen that $a_4 < 0$ if $g' \left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right]$ is positive or not "too negative". We make this assumption throughout.

Second, $\forall\beta < 1 : a_2 > 0$ and if $\beta = 1$, then $a_2 = 0$ in the neighborhood of the solution (σ_p, x_p). This is because $a_2 = S_p(\sigma, x \mid 1, \bar{\sigma})$ and the FOC with respect to search effort equates this expression to zero if $\beta = 1$. Since the search effort of a procrastinating individual is always set too low compared to the optimal value of an exponential agent, the marginal benefit of search will always exceed the corresponding benefit for an exponential agent in the neighborhood of the solution. Hence, $S_p(\sigma, x \mid 1, \bar{\sigma}) > 0$ in this region. To formally prove this, observe that $R_p(\sigma, x \mid Ey, \bar{\sigma})$ does not depend on β and that the $S_p(\sigma, x \mid \beta, \bar{\sigma}) = 0$ curve monotonically shifts to the right with β in the (σ, x)-plane:

$$\left.\frac{\partial\sigma}{\partial\beta}\right|_{S_p(\sigma,x|\beta,\bar{\sigma})=0} = -\frac{\frac{\partial S_p}{\partial\beta}}{\frac{\partial S_p}{\partial\sigma}} > 0 \tag{3.60}$$

which is strictly positive because $\frac{\partial S_p}{\partial\beta} = \frac{\delta\lambda' Q}{1-\delta(1-q)} - \frac{\partial p}{\partial\sigma}\frac{\delta}{(1-\delta)}(1-h)(x-x_z) + p\frac{\delta}{1-\delta}\lambda'\bar{F}(x-x_z) > 0$ and $\frac{\partial S_p}{\partial\sigma} = a_4 < 0$. The solution of the procrastinating agent being at the intersection of the $R_p(\sigma, x \mid Ey, \bar{\sigma}) = 0$ and $S_p(\sigma, x \mid \beta < 1, \bar{\sigma}) = 0$ curve will therefore always be to the left of the $S_p(\sigma, x \mid 1, \bar{\sigma}) = 0$ curve, so that $S_p(\sigma, x \mid 1, \bar{\sigma}) = a_2 > 0$.

Third, we assume throughout that the first-order conditions define a unique maximum. Consequently, $a_1 < 0$, $a_4 < 0$ and $D > 0$. The first condition is automatically satisfied and the second conditional on the aforementioned assumption that $g' \left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right]$ should be positive or not too negative. Since $D = a_1 a_4 - a_2 a_3$, and $a_1 < 0$, $a_4 < 0$ and $a_2 > 0$, D is > 0 if a_3 is negative or not too positive.

Finally, we insert (3.27), (3.57) and (3.59) in (3.52) and (3.55) and rearrange to observe that a_3 and a_6 cannot in general be signed:

$$a_3 = -\frac{\beta\delta}{(1-\delta)}\left\{\lambda'\bar{F}\frac{[(1-p)(1-\delta) - p\delta q]}{[1-\delta(1-q)]} + p\lambda' f(x-x_z) - \frac{g}{\sigma}\left[\lambda f(x-x_z) + (1-h)\right]\right\} \tag{3.61}$$

$$a_6 = -\frac{g'}{\sigma\bar{\sigma}}\left\{y_b - y_z + \frac{\beta\delta(1-h)(x-x_z)}{1-\delta}\right\} - \frac{g}{\bar{\sigma}}\frac{\beta\delta\lambda'\bar{F}(x-x_z)}{1-\delta} \tag{3.62}$$

The sign is, however, unambiguous for the limiting cases in which p tends to zero or to one. In the following, we successively consider the environment where the support of the measurement error ε is unbounded and the one where it is bounded.

Case 1: $\mathbf{p \to 0}$

We will show that in this case $a_3 < 0$ and $a_6 < 0$. In the benchmark when the measurement error has an infinite support, $p \to 0$ only if $\bar{\sigma} \to 0$ and therefore $\log(\bar{\sigma}/\sigma) \to -\infty$. We have assumed that $\lim_{\varepsilon \to -\infty} g(\varepsilon) = 0$ and $\lim_{\varepsilon \to -\infty} g'(\varepsilon) > 0$. Inserting this in (3.61) and (3.62) yields

$$a_3 \to -\frac{\beta\delta\lambda'\bar{F}}{1-\delta(1-q)} < 0 \tag{3.63}$$

$$a_6 \to -\frac{g'}{\sigma\bar{\sigma}}\left\{y_b - y_z + \frac{\beta\delta(1-h)(x-x_z)}{1-\delta}\right\} < 0 \tag{3.64}$$

Now, consider the case in which the measurement error has a finite support. Then, the limiting case for which $p = 0$ is attained if $\log(\bar{\sigma}/\sigma) = -\underline{\varepsilon}$. In Subsection 3.5.2.2, we assume in this case $g(-\underline{\varepsilon}) > 0$ and $g'(-\underline{\varepsilon}) \geq 0$. Therefore, we obtain that $a_6 < 0$ by substitution in (3.62), but

$$a_3 = -\frac{\beta\delta\lambda'\bar{F}}{1-\delta(1-q)} + \frac{\beta\delta g}{(1-\delta)\sigma}\left[\lambda f(x-x_z) + (1-h)\right] \lesseqgtr 0 \tag{3.65}$$

which is negative if g is not "too large".

Case 2: $\mathbf{p \to 1}$

Since $x - x_z \to 0$ for $p \to 1$, (3.61)-(3.62) simplify to

$$a_3 \to \frac{\beta\delta}{(1-\delta)}\left\{\lambda'\bar{F}\frac{p\delta q}{[1-\delta(1-q)]} + \frac{g}{\sigma}(1-h)\right\} > 0 \tag{3.66}$$

$$a_6 \to -\frac{g'}{\sigma\bar{\sigma}}\{y_b - y_z\} \geq 0 \tag{3.67}$$

where the last inequality follows from the fact that $\lim_{\varepsilon \to +\infty} g'(\varepsilon) < 0$ in the benchmark when the measurement error has an infinite support, and $g'(\bar{\varepsilon}) = 0$ in the case of a finite support.

To sum up, we found that $a_1 < 0$, $a_2 > 0$, $a_4 < 0$, $a_5 \geq 0$, $D > 0$, while $a_3 < 0$ and $a_6 < 0$ if $p \to 0$ and $a_3 > 0$ and $a_6 \geq 0$ if $p \to 1$. By continuity, there exists at least one $\tilde{p} \in (0,1)$ at which $a_3 = 0$ and for the benchmark case at least one $\hat{p} \in (0,1)$ at which $a_6 = 0$. If the measurement error has finite support $\hat{p} = 1$.

Using these results and partially differentiating (B.12), we obtain:

$$\frac{\partial r_p(\sigma \mid Ey, \bar{\sigma})}{\partial\sigma} = -\frac{a_2}{a_1} > 0 \tag{3.68}$$

$$\frac{\partial r_p(\sigma \mid Ey, \bar{\sigma})}{\partial\bar{\sigma}} = \frac{a_5}{a_1} \leq 0 \tag{3.69}$$

where the latter is strictly negative only iff $g\left[\log\left(\frac{\bar{\sigma}}{\sigma}\right)\right] > 0$.

The following partial derivatives change signs at least once as p is raised from zero to one.

$$\frac{\partial s_p(\sigma \mid \beta, \bar{\sigma})}{\partial\sigma} = -\frac{a_4}{a_3} \tag{3.70}$$

$$\frac{\partial x_p}{\partial\bar{\sigma}} = \frac{a_4 a_5 - a_2 a_6}{D} \tag{3.71}$$

$$\frac{\partial\sigma_p}{\partial\bar{\sigma}} = \frac{a_1 a_6 - a_3 a_5}{D} \tag{3.72}$$

$\frac{\partial s_p(\sigma|\beta,\bar{\sigma})}{\partial \sigma}$ is strictly negative for p close to zero and increases to a strictly positive value for p close to one. Since $a_5 = 0$ for $g = 0$ in the benchmark case, $\frac{\partial x_p}{\partial \bar{\sigma}}$ is at first positive for p close to zero and negative as p approaches one. In the finite support case, the sign of this derivative is at first ambiguous close to $p = 0$, since then $a_5 > 0$. In both environments, $\frac{\partial \sigma_p}{\partial \bar{\sigma}}$ is strictly positive for p close to zero, but strictly negative for p close to one. Finally,

$$\frac{\partial s_p\left(\sigma \mid \beta, \bar{\sigma}\right)}{\partial \bar{\sigma}} = \frac{a_6}{a_3} \tag{3.73}$$

This partial derivative is strictly positive both when p is close to zero ($p < \min(\tilde{p}, \hat{p})$) and close to one ($p > \max(\tilde{p}, \hat{p})$). At $p = \tilde{p}$ the derivative becomes infinite and changes signs. At $p = \hat{p}$ the derivative becomes zero and changes a second time. In general, we do not know whether $\hat{p}$ is greater or smaller than $\tilde{p}$. In the case of a finite support, however, $\hat{p} = 1 > \tilde{p} > 0$ and the derivative changes only once for sure. In the other cases, the derivative changes signs at least twice. $\square$

Bibliography

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. *The American Economic Review 95*(1), 437–449.

Card, D. and T. Lemieux (2001). Dropout and Enrollment Trends in the Postwar Period: What Went Wrong in the 1970s? In J. Gruber (Ed.), *Risky Behavior among Youths: An Economic Analysis*, Chapter 9, pp. 439–482. University of Chicago Press.

Clark, D. (2011). Do Recessions Keep Students in School? The Impact of Youth Unemployment on Enrolment in Post-compulsory Education in England. *Economica 78*(311), 523–545.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics 10*(4), 417–451.

Frick, J. R. and M. M. Grabka (2003). Missing income data in the German SOEP: incidence, imputation and its impact on the income distribution. Discussion Paper 376, German Institute for Economic Research (DIW Berlin).

Jenkins, S. P. (1995). Easy Estimation Methods for Discrete-Time Duration Models. *Oxford Bulletin of Economics and Statistics 57*(1), 129–136.

Kalwij, A. and A. van Soest (2005). Item Non-Response and Alternative Imputation Procedures. In A. Börsch-Supan and H. Jürges (Eds.), *The Survey of Health, Ageing and Retirement in Europe. Methodology*, pp. 128–150. Mannheim Research Institute for the Economics of Aging.

Kiefer, N. M. (1988). Analysis of Grouped Duration Data. In N. U. Prabhu (Ed.), *Statistical Inference from Stochastic Processes*. American Mathematical Society, Providence.

Micklewright, J., M. Pearson, and S. Smith (1990). Unemployment and Early School Leaving. *The Economic Journal 100*(400), 163–169.

Moulaert, T. and J. Verly (2006). Le revenu minimum mensuel moyen garanti. Chronique internationale de l'IRES 103, Université Catholique de Louvain.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012). The Short– and Long–Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Petrongolo, B. and M. J. San Segundo (2002). Staying–on at school at 16: the impact of labor market conditions in Spain. *Economics of Education Review 21*(4), 353–365.

Särndal, C.-E. and S. Lundström (2005). *Imputation*, pp. 153–165. John Wiley and Sons, Ltd.

SONAR (2003). Hoe maken de jongeren de overgang van school naar werk? Basisrapportering Cohorte 1978 (eerste golf). Technical report, Leuven: Steunpunt WAV.

SONAR (2004a). Hoe maken de jongeren de overgang van school naar werk? Basisrapportering Cohorte 1976 (tweede golf). Technical report, Leuven: Steunpunt WAV.

SONAR (2004b). Hoe maken de jongeren de overgang van school naar werk? Cohorte 1980 (eerste golf). Technisch rapport. Technical report, Leuven: Steunpunt WAV.

Tumino, A. and M. Taylor (2013). The impact of local labour market conditions on school leaving decisions. Mimeo, ISER, University of Essex.

Wooldridge, J. M. (2006). Cluster-Sample Methods in Applied Econometrics: An Extended Analysis. Mimeo, Department of Economics, Michigan State University.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section And Panel Data (second ed.)*. The MIT Press, Cambridge, MA.

Overall Conclusions

This dissertation investigates the scarring effects of early unemployment on the labor market career of Flemish youth. This is relevant from a policy perspective, as if unemployment at the start of the career does not entail long-lasting impacts on the subsequent labor market career, workers are capable to offset by themselves any adverse effect on their career and hence no policy interventions should be advocated. By contrast, if early unemployment inflicts scars on workers' labor market careers, policies targeting the reduction of unemployment should be promoted. In this dissertation we find evidence on favour of the latter hypothesis.

In the first chapter we study the long-term effects of graduating in economic downturns on the labor market career of low and high educated youth. While the labor market conditions at graduation are found to have a persistent negative effect on earnings for both groups, the low educated experience higher unemployment, whereas the high educated downgrade towards lower-paying jobs. This results are explained by the labor market rigidities affecting differently the labor market of the white and blue collar workers. Moreover, the results of the second chapter corroborate the hypothesis of the first chapter by showing that the early experience of non-employment entails long-run effects on the labor market of low educated youth.

At a micro level, one can think of specific curative policies to reduce the impact of experiencing early non-employment, depending on the channel through which the scar materializes. For instance, the scar may be originated by the human capital depreciation occurring in the unemployment spell, by the foregone human capital that would have been accumulated in case of early work-experience, or because early non-employment is interpreted as a signal of low quality. If the first mechanism is the main driver of the scar, training schemes targeting young unemployed would be an appropriate cure. If instead the main cause of the scar is the loss of early work experience, policies that foster the integration of youth in the labor market should be advocated, such as wage subsidy programs. However, if the cause of the scar is rather the bad signals conveyed by the status of unemployment, it is not sure that entering in subsidized programs would improve the perception of the youth unemployed by the employers. More reflection on the actual channel should be needed to come up with a specific policy to target the Flemish youth.

From a more global perspective, however, the results of the first chapter suggests that the penalties of unemployment are not only limited to those who experience it. That is, already a higher risk of unemployment at graduation - without necessarily experiencing it - is sufficient to provoke a scar. These penalties are explained by the labor market rigidities of the Belgian labor market. This suggests that the aforementioned scars could be diminished by policy measures that target a reduction of unemployment. We argue that this could be addressed by a *flexicurity* system, where flexibility in the labor market goes hand in hand with a generous insurance system that provides for the unemployed.

On the one hand, flexibility should be enhanced to encourage employers to take the entrepreneurial risks and create more jobs. This is because labor market rigidities such as excessive employment protection may hamper productivity growth if they prevent workers reallocation or new hirings due to high expected future firing costs in future downswings. In this view, employment protection should be limited to promote job creation and these protections should increase with tenure. On the other hand, if workers bear a higher risk of unemployment because of the increased flexibility in the labor market, they should be protected by a more generous unemployment insurance system. To the extent that the inflow in employment is also enhanced in flexible labor markets, the experience of unemployment should be temporary for most of the workers. However, less employable individuals may be long-term unemployed. This group could be supported by active labor market programs as well as training schemes to improve their employability and help them to find a good match. Moreover, a generous unemployment insurance system entails a moral hazard problem that may result in insufficient job search by the unemployed. This inefficiency should be appropriately tackled in the context of flexicurity. A popular measure is to condition the provision of unemployment benefits to job search requirements and monitor the job search of the unemployed to verify compliance, eventually imposing a sanction in case of non-compliance. This is the topic of the third paper, in which we discuss the welfare implications of job search monitoring as a way to restore the incentives to search under unemployment insurance. Note that, from a flexicurity perspective, the moral hazard in the context of a generous unemployment insurance system may represent a source of rigidity since it reduces the turnover rate in the labor market, i.e. the unemployed are less likely to leave unemployment, which may also influence the likelihood that employers open up vacancies. Hence, measures that counteract reducing the efficiency-insurance trade-off of the unemployment insurance system play an important role in the design of flexicurity.

Thus, a flexicurity system entails three different dimensions (flexibility, security and the associated moral hazard), each of them crucial for the functioning of the entire system. As a consequence, its implementation requires a number of reforms affecting different aspects of the labor market: (i) eliminating the rigidities resulting from excessive jobs protection; (ii) shifting from job protection towards workers protection by means of a generous insurance system; (iii) addressing the moral hazard associated to unemployment insurance to restore the incentives to search for jobs. In what follows we discuss each of these dimensions in the Belgian context.

- *Flexibility:* The Belgian labor market is characterized by a number of rigidities that, according to the results of this dissertation, have caused substantial penalties for both low and high educated youth. First, in the period under investigation, Belgium was characterized by a flexible EPL for blue collar workers and a rigid EPL for white collar workers. This encouraged the high educated who graduated in period of low demand to downgrade and accept mismatched jobs. Since the beginning of 2014, however, a single employment contract has been introduced stipulating the same EPL for white and blue collar workers. Therefore, this controversial discrimination between these two types of workers has been finally removed. Second, the short-term work compensation program for blue collar workers represents an additional source of rigidity, as it ties the employers to the employees during downturns. This may have similar consequences to a strict EPL to the extent that it prevents workers reallocation or new hirings due to high expected future firing costs in future downswings. A last source of rigidity is represented by the very high level of minimum wages. According to our findings, a reduction of this level may facilitate the absorption of low educated new graduates, for which minimum wages are binding.

- *Security:* In Belgium a generous unemployment insurance system represents an important safety net for the unemployed. The Belgian unemployment insurance system is more generous in comparison to many other countries for two main reasons. First, there is no time limit on the payment of unemployment benefits. Second, the eligibility for the provision of unemployment benefits is granted not only to unemployed workers with sufficient employment records, but also to unemployed school-graduates with no employment records, conditional on a waiting period.

- *Tackling moral hazard:* Of course, the problem of moral hazard entailed by the unemployment insurance system should be appropriately tackled to restore the incentives to search for jobs. Since 2004 a job search monitoring scheme has been introduced in Belgium, targeting the long-term unemployed (at least 13 months unemployed). Despite the fact that this measure has been shown to effectively raise the job finding rate on this specific group of long-term unemployed (Cockx and Dejemeppe, 2012),[19] other dimensions such as the quality of subsequent employment have not yet been considered. In our third theoretical chapter we demonstrate that, from welfare perspective, job search monitoring reaches efficiency under limited conditions and that, allowing for (more realistic) imperfections in the monitoring technology, efficiency may not be achieved. Therefore, we suggest that other policies such as job search assistance may be socially more efficient. However, whether this is the case remains an open question, which should be answered empirically. We leave this for future work.

[19] Bart Cockx, Muriel Dejemeppe, “Monitoring job search effort: An evaluation based on a regression discontinuity design”, Labour Economics, Volume 19, Issue 5, October 2012, Pages 729–737.

A

Supplementary Appendix to "Scars of Recessions in a Rigid Labor Market"

A.1 Variables Construction

This analysis exploits data from different sources: (i) individual and family background variables are taken from the Sonar dataset, which is a survey-based dataset that follows a representative sample of three birth cohorts (born in 1976, 1978 or 1980) living in Flanders at age 23. The surveys were conducted when respondents were aged 23, 26 or 29 and register retrospectively and on monthly basis their most important activity - among which education - from the beginning of secondary education.[1] These observations are matched with the following administrative data of Belgian Social Insurance institutions centralized at the Cross Roads Bank of Social Security (CBSS, hereafter, Data Warehouse): (ii) yearly data on the province of residence, obtained by the Regional Register of Belgium; these data are available from the year in which individuals turn age 18 (i.e. year 1994, 1996 and 1998 for cohort born in 1976, 1978 and 1980, respectively) up to year 2010; (iii) quarterly data on individual labor market performances for the period 1998-2010, provided by the Regional Social Security office (RSZ), the Regional Social Security office of Provincial and Local Administration (RSZPPO) and the Regional Institute for the Social Security of the Self-employed (RSVZ). Finally, (iv) the URates series are provided by the Labor Force Survey (LFS). In what follows we describe the data and the construction of the variables for each of these datasets.

[1] For details, see SONAR (2003, 2004a,b).

A.1.1 Sonar Dataset: Educational Variables and the Moment of Graduation

The Sonar database is exploited to construct the following educational measures: (a) the moment of graduation, (b) the years of successfully attained education at age 17 (hereafter, "completed education at age 17"), (c) the total years of successfully attained education until graduation (hereafter, "completed education"). Based on the latter, we divide the sample into (d) low and high educated.

(a): The moment of graduation - The moment of graduation (month and year) is defined as the first exit from schooling after the completion of compulsory education (age 18 in Belgium). This variable is based on the monthly individual status provided by Sonar, which reports on a monthly basis whether the individual is studying, working, studying and working at the same time, or not working (and in case for which reason), from December of the year in which one turns 17 until at least the year in which one turns age 26.[2] According to our definition, individuals are considered out of education whenever they are not observed in regular education (i.e. if they are observed working, unemployed or not working). Note that those reporting as being in regular education and working at the same time are still considered as in education. In December of the year in which they turn age 17 all individuals in the sample are observed in education.[3] Whenever we cannot observe directly the transition from education to employment or unemployment, the exit from education is defined as not being in education for more than 4 subsequent months, independently of the destination (working, unemployment, not working, missing).

Imputation procedure for censored observations in Sonar - This procedure allows us to define the age of first exit from school for 96.5% of the sample. Among the observations for which the age of first exit is not defined, some individuals are censored at age 23 or 26 while still in ongoing education (3%): for these individuals, the age of first school exit is inferred (when possible) by exploiting the socio-economic position as reported at the end of each quarter for the period 1998-2010 by the Data Warehouse.[4] Intuitively, this is done assuming that, from the moment of censoring in the Sonar dataset, these observations remain in regular education until the moment in which they appear in the labor market according to the Data Warehouse. We cannot apply this procedure to individuals whose socio-economic position is never observed by the Data Warehouse: these individuals are dropped (0.11%). In some other cases, individuals are observed as workers - even before censoring - according to the Data Warehouse data. Among these cases we retain only individuals who are observed as employed in student jobs, while still in full-time education according to Sonar (2.75%). In contrast, we drop those who result employed in regular jobs while still in full-time education according to Sonar (0.13%), because of inconsistency between Sonar and Data Warehouse. Student jobs are defined according to the following criteria: (1) we exploit a

[2]That is, in each month, an individual can be observed in full-time education, education and working at the same time, employment, unemployment, or non-employment. For the 1978 wave, information is only collected until age 26. For the other waves, it is gathered until October or November of the year one turns 29, depending on the date of the survey.

[3]We drop few observations who quit education before the end of compulsory education.

[4]Participation to the labor market is defined as being employed or looking for a job; inactive status is not considered since could as well be in education.

variable provided by the Data Warehouse that identifies jobs that are “typically" undertaken while studying. (2) In case (1) is satisfied, we ensure that the time-worked in this potential student job does not exceed the maximum time a child is allowed to work, for his parents not to lose the family allowance that they receive for their children: this threshold amounts to 240 hours (i.e. roughly 30 days) per quarter except in the summer quarter, during which a child is allowed to work more. Hence, a potential student job is defined as student job if the individual has worked at most 30 days for each quarter except the summer.

For all other censored observations, the age of first exit from school is imputed by assuming that the individuals remain in education until the moment in which they are observed for the first time as active in the labor market, i.e. either as employed or as unemployed looking for a job. Accordingly, the moment of first exit from education is set just before the moment when they are first observed in the labor market. Note that the age of first exit from school is imputed differently depending on whether the individuals are first observed in the labor market as employed or unemployed looking for a job. The reason is that an individual is observed as unemployed when he first receives the unemployment benefits (UB), but in Belgium the receipt of these UB implies a waiting period that varies between 9 and 12 months from the registration to the unemployment agency, depending on the age; this means that, for the unemployed, the actual moment of entry in the labor market is at the moment of registration to the unemployment agency rather than at the moment of receiving the UB. Accordingly, the age of first school exit for the unemployed take into account this waiting period.

Check inconsistencies between Sonar and Data Warehouse - Recall that the year of graduation is defined mostly on information reported in Sonar, i.e. on survey-data. We verify the reliability of this information by exploiting the consistency with the (high-quality) administrative data by Data Warehouse: in particular, we check whether individuals are observed in the labor market according to the Data Warehouse in the years when they report to be still in full-time education in Sonar. Among the cases for which graduation is defined, 19.2% of the individuals are indeed observed in the labor market before their graduation year. Consistently with the aforementioned “student job” criterion, these observations are considered consistent if they are employed in student jobs. Inconsistent observations are dropped from the sample (4%).

(b):Completed education at age 17 - The Sonar dataset contains detailed information about the educational path of the individuals since secondary education: in each year, one can observe the type of education in which one is enrolled (e.g. full-time or part-time vocational), the type of program (general, vocational, artistic, technical, apprenticeship), the grade and the result at the end of each grade (successfully passed or not). This information is exploited to build completed education at age 17. This variable starts counting from the beginning of the first year of secondary education, and increases every time one passes a year, until the academic year in which one turns 17 (included). In the analysis, this variable is expressed so to give information about the educational progression of the individual at age 17 with respect to the theoretical years of schooling: 0 means that the individual is on time, positive numbers indicate the number of repeated grades, and -1 indicates that the individual has skipped one academic year. Note that this contains exactly the

same information as completed education at age 17.

(c): Completed education - This is constructed similarly to “completed education at age 17”. It counts the years of schooling successfully attained from the beginning of the secondary education. For individuals who are censored in Sonar database at 23 or 26 years old while still in ongoing education - and hence for which graduation is not observed in Sonar - completed education is corrected so that it is consistent with the imputed graduation, which is inferred using information from the Data Warehouse, as explained above. For these cases, it is assumed that, from the moment of censoring, all years in education were successfully passed until graduation. In each case, we ensure the consistency between completed education and the year of graduation, dropping few individuals for which the imputed completed education is unrealistically too high with respect to the year of graduation.

(d): Low and High Educated - The sample is divided into 2 educational groups based on “completed education”: low-educated are individuals who completed secondary education, that is who graduated at age 18, or at age 19 if enrolled in vocational program in secondary education (this is because secondary education ends after seven years for the vocational track and after 6 years for all other educational tracks.). High-educated are those who graduated later. In the main analysis these 2 groups are always studied separately.

A.1.2 Data Warehouse - Individual Labor Market Outcomes

Individual labor market outcomes are constructed by exploiting quarterly data from RSZ and RSZPPO database, as well as yearly data from RSVZ, for the period 1998-2010. RSZ and RSZPPO databases collect information on salary, earnings and time worked in dependent employment in the private and public sector, respectively; RSVZ database reports the registration in self-employment for the period 1998-2010 and yearly earnings from self-employment for the period 1998-2007. In the main analysis earnings from self-employment are not exploited as outcomes of interest since they are self-reported data, and hence, likely to be under-reported.[5] Based on these data, we construct the following yearly labor market outcomes: (a) log of earnings, (b) log of hours worked and (c) log of hourly wage in salaried employment;[6] (d) a working status as salaried employed (defined by strictly positive earnings and not being self-employed); (e) a self-employment status (if registered as such part of the year, irrespectively of being a salaried worker in the same year); (f) a working status comprising both salaried and self-employment. Descriptive statistics of these individual labor market outcomes is reported in Table A.2 and A.1 below, for low and high educated, respectively.

(a): Earnings - It is based on gross earnings, which for both RSZ and RSZPPO is defined as the sum of all remunerations that are subject to social contributions (including holiday allowances), excluding the allowances from contract termination: RSZ excludes premia from the definition,

[5]However, a worker is considered as self-employed if a given year he is registered as self-employed according to RSVZ, or if he is not registered as such but reports positive earnings from self-employment to RSVZ.

[6]The notion of hours worked accounts for part-time versus full-time work. Hourly wage is constructed by dividing yearly earnings by the total number of hours worked in a year. For details, see below.

whereas RSZPPO includes premia and the Christmas bonus.

Earnings from salaried employment are provided in classes of 100 Euros, so that the earnings class equal to 0 refers to earnings between 1 and 100 Euros. We transform yearly earnings multiplying by 100, and adding 50 - the midpoint of the interval of each class - in order not to underestimate real earnings. In addition to quarterly earnings by separated sources, the Warehouse Data provide us with the sum of all earnings in a year (adding up yearly earnings from RSZ, RSZPPO and RSVZ). Hence, for the cases in which the individuals are not working as self-employed, we use this sum as a measure of annual earnings from dependent employment, while for the years in which individuals are also self-employed, we compute by ourselves the sum of earnings from RSZ and RSZPPO - hence without considering the earnings from self-employment. Note that in this case, yearly earnings from RSZPPO and RSZ are first transformed from classes of 100 Euros in real earnings, and then summed up to obtain a yearly measure of earnings from dependent employment.

(b): *Hours worked* - The RSZ provides us with the number of working days in case of full-time work and the number of working hours in case of part-time work, while RSZPPO give us the number of working hours both for full-time and part-time work. To clean the data, we compute the total number of working days in case of full-time work - in RSZ and RSZPPO, separately - and drop the yearly observations whose values are above 312 working days per year, which corresponds to the maximum number of working days in case of full-time work in a 6 days per week regime. The equivalent number under a 5 days per week regime is 260. Then, for annual observations corresponding to full-time workers who work between 260 and 312 days per year, we assume a working regime of 6 days per week, while for the other cases we assume a working regime of 5 days per week. We decide to focus on hours worked in order to take into account also part-time work. Hence, for full-time work in RSZ, we convert working days in working hours, assuming a working regime of 8 (7.6) hours per day for the period 1998-2002 (2003-2010); this is because, in Belgium, the 7.6 hours per day regime for full-time work was introduced by law in January 2003 to replace the 8 hours per day regime. Accordingly, for full-time work in RSZ we transform the yearly working days in yearly working hours, by multiplying the former by 8 (7.6) until (strictly after) 2002. Then, we add up the yearly working hours for full-time and part-time work, and across sources (RSZ and RSZPPO) - in order to get a measure of the total number of hours worked in a year in dependent employment.

(c):Average Hourly Wage - Average hourly wage for dependent employment is obtained by dividing annual earnings by the annual number of hours worked, i.e. $(a)/(b)$.

Cleaning procedure for (a),(b) and (c) - For variables (a),(b) and (c), we check the presence of outliers according to the following procedure. Given the presence of minimum wages in Belgium, we detect bottom outliers by comparing the average hourly wage (c) with the corresponding hourly minimum wage.[7] The official monthly minimum wages are provided by the Conseil Regional du

[7]As a conservative measure - i.e. in order not to detect too many outliers - for this specific comparison we assume that earnings are at the top of each earnings-class instead of at the midpoint: i.e., multiply by 100 earnings in classes of 100 Euros and add 100 rather than 50 (as in the procedure explained in the text).

Travail of Belgium for each year. These numbers are adjusted by the Consumer Price Index (using the first quarter of 2011 as reference) and by age, as different percentages of the official minimum wage are applied for workers below age 21.[8] Then, for each minimum wage, year and age category, we construct the corresponding hourly minimum wages as follows: (i) compute the total annual earnings of a full-time worker paid at the minimum wage - by multiplying the monthly minimum wage times 12. (ii) Compute the total number of working hours in full-time work under the 5 and 6 days per week regimes, assuming the 8 (7.6) hours per day regime for the period 1998-2002 (2003-2010). This corresponds to 2496 (2080) annual working hours for the period 1998-2002 under the 6 (5) days per week regime, and 2371.2 (1976) annual working hours after 2003 under the 6 (5) days per week regime. (iii) Divide the total annual earnings calculated in (i) by the alternative results in (ii) to obtain the hourly minimum wages in full-time work under the 6 and 5 days per week regime for the period 1998-2002 and 2003-2010, respectively. (iv) For full-time workers that were observed in the 6 days per week regime, we apply the hourly minimum wage computed in (iii) under a 6 days per week regime; for all other workers we use the hourly minimum wages computed in (iii) under the 5 days per week regime. We detect as bottom outliers the annual observations in which the average hourly wage is below the corresponding hourly minimum wage: in this case, annual hours worked, annual earnings and average hourly wage are all replaced to missing.

Next, we check for the presence of outliers in the upper part of the distribution of annual earnings, average hourly wages and annual hours worked. As a general rule, we consider as top outliers the top percentile in the distribution of each of these variables. For average hourly wage and annual hours worked, we look at the distribution pooling all years together. For annual earnings, we identify as outliers the last percentile of the distribution of annual earnings by age, under the assumption that earnings do not differ systematically across years, but change over age. Similarly to the procedure used for bottom outliers, we replace to missing each of the variables (a), (b) and (c) whenever a top outlier in any of these three variables is detected. In total, according to this procedure 3% of the annual observations are detected as outliers. Finally, annual earnings, annual hours worked and average hourly wages are log-transformed.

(d): Salaried Employment Rate - The dummy for working in dependent employment is built upon the variable on annual earnings from salaried employment (a): the dummy is 1 when annual earnings are positive (i.e. if the worker has received a wage from salaried employment for at least 1 day in the year) or when earnings are missing because outliers. This is because outliers should be considered as employed individuals for whom one cannot calculate earnings or the time worked. Note that this outcome considers workers who in a given year worked (only) as salaried workers and never registered as self-employed.

(e): Self-employment Rate - Workers are defined self-employed if they are registered in the

[8] Workers aged 17, 18, 19 and 20 receive 76%, 82%, 88% and 94% of the official monthly minimum wage, respectively. Workers aged 21 or more receive the entire official monthly minimum wage (Moulaert and Verly, 2006). As a conservative choice, we consider the minimum wage of the previous age, i.e. we take the minimum age of those aged 17 in the year in which one turns 18.

self-employment register in a given year. Workers who combined in the same year both salaried and self-employment are considered self-employed (and not salaried employed).

(f): *Overall Employment Rate* - This dummy considers working in both salaried and self-employed: it is the sum of (d) and (e).

A.1.3 Data Warehouse - Firm Permanent Characteristics and Firm Mobility

The Data Warehouse gives us access to data on firm characteristics where the individuals are working in the observation period, as well as the firm identification number. We use the latter to come up with a measure of firm mobility. Among firm characteristics, quarterly data on the median daily wage paid out on June 30 in recruiting firms are exploited to build an indicator of permanent firm quality. Descriptive statistics of these data outcomes is reported in Table A.2 and A.1 below, for low and high educated, respectively.

Firm Mobility - This variable exploits the changes in the quarterly firm identification number. Transitions between self-employment and salaried employment are included in the definition of firm mobility, as we assign to self-employment a specific firm identifier. An individual is defined to change firm in year t if he is observed in a different firm in at least two quarters of the year t, or if the first firm in which he was employed in year t differs from the last firm in year $t-1$.

Permanent Firm Quality - In order to obtain this indicator of permanent firm quality we apply the following procedure, which is very close to Oreopoulos et al. (2012). (i) Quarterly nominal values of median daily wages paid by the firm are deflated by using two indexes of nominal wage trend from Belgostat (base year 1997) for white and blue collar workers, respectively; these are then converted in real terms by using the CPI (base year 2011). (ii) The data are log-transformed. (iii) Seasonal effects are taken out by regressing the quarterly time-varying data on dummies for quarters. (iv) The residuals from this regression are averaged by firm over the observed quarters, as many quarters as median daily wages are observed for each firm. Note that, due to this procedure, permanent characteristics are expressed in deviation from the average: for instance, a negative (positive) value of the median salary paid by the firm in a given quarter means that the firm paid less (more) compared to the average. (v) Since individuals may have changed firm within a calendar year, for each individual we average over the quarters of year t the permanent characteristics corresponding to the firms where an individual is observed working in t. This allows to get an annual indicator of the permanent characteristics of the average firm where an individual was employed.

A.1.4 Data Warehouse - Geographical Mobility

Geographical mobility is based on the province of residence, which for year t is measured in December of year $t-1$. The definition of the variable is as follows: an individual is defined to move in year t, if he lives in another province at the end of year t than where he lived at the end of year $t-1$. Note that, differently from all other outcomes which are observed for the period 1998-2010, residence is observed for the year in which individuals turn age 18 (which corresponds to year 1994, 1996 and 1998 for individuals born in 1976, 1978 and 1980, respectively) until year 2010. Therefore, using our definition we are able to define the province of residence from experience 1 onwards for low (high) educated, for all graduation cohorts - in particular for low-educated who graduates in the period 1994-1996, whose first year of experience occurs before 1998 (period at which the other outcomes are not observed). Note that this extra-availability of data on residence does not make a difference for high educated, as the first year of experience for the first graduation cohort occurs in 1998. Descriptive statistics of geographical mobility is reported in Table A.2 below.

A.1.5 LFS - Provincial URate

We use the 1994-2010 provincial unemployment rate series of the working population aged 15-64 (considering both men and women) based on the Labour Force Survey (LFS), since these series use the internationally accepted ILO definition of unemployment. In order to check the reliability we compared these series to the administrative ones provided by the National Employment Office of Belgium (RVA - Rijksdienst voor Arbeidsvoorziening) also available from 1994 onwards. In the latter series the unemployment rate is defined as the ratio of the number of unemployment benefit recipients searching for jobs to the number of individuals insured against unemployment. In general this results into higher unemployment rate figures than those of the LFS, but the evolution over time is overall very similar. Nevertheless, for the province Limburg, the two series displayed a very different pattern between 1994 and 1997. In those of the LFS the unemployment rates were increasing during these years while they were evolving downwards in the series of RVA. Since the unemployment rates in the other provinces were moving down in this period according to both data sources, we believe that a serious measurement error biases the LFS unemployment rate of Limburg severely during these years. Based on the RVA data, we therefore adjusted the LFS series of the unemployment rate of Limburg for the period 1994-1997. The details of the adjustment procedure can be obtained from the authors on request.

Table A.1: Descriptive Statistics of Individual Labor Market Outcomes, Permanent Firm Quality and Mobility: High Educated.

Variable	Missings	Obs	Mean	Std. Dev.	Min	Max	Label
Annual Discrete Individual Labor Market Outcomes							
salaried	0	14490	0.830	0.375	0	1	1 if earnings$>$0 from salaried empl. in year t(& not self-empl.)
self-empl	0	14490	0.142	0.349	0	1	1 if registered as self-empl. in t
overall empl.	0	14490	0.972	0.165	0	1	1 if either salaried or self-empl. in t
Annual Continuous Individual Labor Market Outcomes§							
earnings	1979	12511	29705.360	10424.370	150	65550	gross earnings if salaried empl. in t (level)
hours worked	1979	12511	1778.674	431.405	5	2256	hours worked (FT+PT) if salaried empl. in t (level)
hourly wage	1979	12511	16.708	4.490	6.667	33.093	average hourly wage if salaried empl. in t (level)
FT hours	1979	12511	1633.485	655.300	0	2256	Full-Time hours worked if salaried empl. in t (level)
PT hours	1979	12511	145.189	378.582	0	2252	Part-Time hours worked if salaried empl. in t (level)
log earnings	1979	12511	10.191	0.603	5.011	11.091	log gross earnings if salaried empl. in t (logarithm)
log hours worked	1979	12511	7.409	0.535	1.609	7.721	log hours worked if salaried empl. in t (logarithm)
log hourly wage	1979	12511	2.782	0.258	1.897	3.499	log average hourly wage if salaried empl. in t(logarithm)
Annual Permanent Firm Quality (raw values)							
median daily wage‡	1780	12710	134.2078	33.50615	36.143	394.07	avg over quarters of median daily wage paid in recruiting firm in t
Annual Mobility							
firm mobility‡‡	676	13814	0.217	0.412	0	1	1 if change firm compared to last observed quarter in $t-1$ or within t
geographical mobility††	34	14456	0.032	0.175	0	1	1 if change province of residence in year t compared to $t-1$

This table is based on the sample of high-educated graduating in the period 1997-2007, i.e. the sample studied in Chapter 1.

§ Outliers in any of the continuous outcomes are replaced to missings (for all countinuos outcomes) and still considered as salaried employed with not reliable info in terms of working hours, hourly wage and earnings. Continuous outcomes in level are related as follows: earnings=hours worked*hourly wage; FT hours+PT hours=hours worked. Log-transformed continuous variables are linearly related as follows: log earnings=log hours worked+log hourly wage.

‡ We do not report descriptive statistics based on the variable used in the main analysis, since it is based on residuals (see Section A.1.3 for details) and therefore its values are not informative of permanent firms' quality. In contrast, to get some descriptive statistics we construct this variable according to the following "naive" procedure, which maintains the original values of the data. (i) Quarterly nominal values of median daily wage are deflated by using two indexes of nominal wage trend from Belgostat (base year 1997), for white and blue collar workers respectively, and then converted in real terms by using the CPI (base year 2011). (ii) Quarterly values are averaged by firm over all quarters in the observation period to get a measure of permanent characteristics of each firm (in the main analysis we additionally take out seasonal trends in a regression-based procedure). (iii) Average over quarters to get permanent characteristic of the average firm when a worker works in more than one firm in t.

‡‡ Firm mobility is a dummy=1 in year t if he is observed in a different firm in at least two quarters of the t, or if the first firm in which he was employed in t differs from the last firm in $t-1$. Transitions between self-employment and salaried employment are included in the definition.

†† Geographical mobility is a dummy=1 in year t if an individual lives in another province at the end of t than where he lived at the end of $t-1$. Note that the missings in geographical mobility refer to missing residence in year 2011, which is used to look at geographical mobility for the last calendar year (2010). These observations were not dropped because year 2011 was not used in the main analysis.

Table A.2: Descriptive Statistics of Individual Labor Market Outcomes, Permanent Firm Quality and Mobility: Low Educated.

Variable	Missings	Obs	Mean	Std. Dev.	Min	Max	Label
Annual Discrete Individual Labor Market Outcomes							
salaried	0	20507	0.833	0.373	0	1	1 if earnings$>$0 from salaried empl. in year t(& not self-empl.)
self-empl.	0	20507	0.128	0.335	0	1	1 if registered as self-empl. in t
overall empl.	0	20507	0.961	0.194	0	1	1 if either salaried or self-empl. in t
Annual Continuous Individual Labor Market Outcomes§							
earnings	2993	17514	22667.740	8627.685	50	64550	gross earnings if salaried empl. in t (level)
hours worked	2993	17514	1619.597	472.970	5	2264	hours worked (PT+FT) if salaried empl. in t (level)
hourly wage	2993	17514	13.931	3.217	6.579	32.895	average hourly wage if salaried empl. in t (level)
FT hours	2993	17514	1474.114	644.758	0	2264	Full-Time hours worked if salaried empl. in t (level)
PT hours	2993	17514	145.483	359.528	0	2252	Part-Time hours worked if salaried empl. in t (level)
log earnings	2993	17514	9.891	0.686	3.912	11.075	log gross earnings if salaried empl. in t (logarithm)
log hours worked	2993	17514	7.281	0.647	1.609	7.725	log hours worked if salaried empl. in t (logarithm)
log hourly wage	2993	17514	2.610	0.213	1.884	3.493	log average hourly wage if salaried empl. in t (logarithm)
Annual Permanent Firm Quality (raw values)							
median daily wage‡	2898	17609	116.1369	34.23114	7.0392	727.32	avg over quarters of median daily wage paid out in recruiting firm in t
Annual Mobility							
firm mobility‡‡	1362	19145	0.246	0.431	0	1	1 if change firm compared to last observed quarter in $t-1$ or within t
geographical mobility††	35	21636	0.015	0.120	0	1	1 if change province of residence in year t compared to $t-1$

This table is based on the sample of low-educated graduating in the period 1994-2001, i.e. the sample studied in Chapter 1.

§ Outliers in any of the continuous outcomes are replaced to missings (for all countinuos outcomes) and still considered as salaried employed with not reliable info in terms of working hours, hourly wage and earnings. Continuous outcomes in level are related as follows: earnings=hours worked*hourly wage; FT hours+PT hours=hours worked. Log-transformed continuous variables are linearly related as follows: log earnings=log hours worked+log hourly wage.

‡ We do not report descriptive statistics based on the variable used in the main analysis, since it is based on residuals (see Section A.1.3 for details) and therefore its values are not informative of permanent firms' quality. In contrast, to get some descriptive statistics we construct this variable according to the following "naive" procedure, which maintains the original values of the data. (i) Quarterly nominal values of median daily wage are deflated by using two indexes of nominal wage trend from Belgostat (base year 1997), for white and blue collar workers respectively, and then converted in real terms by using the CPI (base year 2011). (ii) Quarterly values are averaged by firm over all quarters in the observation period to get a measure of permanent characteristics of each firm (in the main analysis we additionally take out seasonal trends in a regression-based procedure). (iii) Average over quarters to get permanent characteristic of the average firm when a worker works in more than one firm in t.

‡‡ Firm mobility is a dummy=1 in year t if he is observed in a different firm in at least two quarters of the t, or if the first firm in which he was employed in t differs from the last firm in $t-1$. Transitions between self-employment and salaried employment are included in the definition.

†† Geographical mobility is a dummy=1 in year t if an individual lives in another province at the end of t than where he lived at the end of $t-1$. Note that, differently from all other outcomes which are observed for the period 1998-2010, this variable is observed from the year in which individuals turn age 18 (1994, 1996, 1998 if born in 1976, 1978 and 1980, respectively) until 2010. Thus the residence is observed from experience 1 also for low educated who graduated in 1994, 1995, 1996, whose residence is observed in calendar year 1995-1997. Missings in geographical mobility refer to missing residence in 2011, which is used to look at geographical mobility for 2010. These observations were not dropped because year 2011 was not used in the main analysis.

Table A.3: Employment Status Decomposition.

		Low educated				High educated			
		freq.	%	freq.	%	freq.	%	freq.	%
not working		800	3.9			407	2.81		
overall employment		19,707	96.1			14,083	97.19		
	of which:								
	salaried employed			17073	83.25			12,030	83.02
	self-employed			2634	12.84			2,053	14.17
	total			19,707	96.1			14,083	97.19
total		20,507	100			14,490	100		

Not working refers to situations of unemployment, education, or being out of the labor market. Salaried employment and self-employment add up to overall employment.

Table A.4: Availability of Firms' Characteristics.

	Individual Continuous Labor Market Outcomes If Salaried					
	Low educated			High educated		
Firm's Characteristics	Missings	Obs	Total	Missings	Obs	Total
Missings	2,453	505§	2,958	1,582	235§	1,817
Obs	540†	17,009	17,549	397†	12,276	12,673
total	2,993	17,514	20,507	1,979	12,511	14,490

§ These observations are outliers in earnings. These individuals are still considered as salaried employed, but their corresponding earnings, hours worked and hourly wages are replaced to missing because the information is not reliable. For this reason firms' characteristics may be available also for outliers in earnings.
† For these observations firms' characteristics are missings even if the individuals are working as salaried employed.

A.2 Sample selection

The original Sonar sample contains 8958 individuals between men and women. From this sample we exclude the following observations, so to have a homogeneous sample: individuals who did not complete compulsory education, i.e. those who quit compulsory education by 31 December of the year one turns age 17 (0.17% of original Sonar sample);[9] individuals who are not Flemish, i.e. either do not have Belgian nationality or do not use to speak Dutch at the parental house (5.10%); individuals who attend special needs education (0.92%). Moreover, we focus on men (50.85%), since female labor supply is likely to differ from the male one due to mothering. We drop individuals with missing values in the following individual control variables, since this involves a small number of observations (2.19%): number of brothers, number of sisters, cohort, educational track at age 17 and completed education at age 17.

[9] In the remainder of this section the percentages of observations are computed relative to the size of the original Sonar sample, i.e. out of 8958 individuals.

Imputation procedure for parental education - Father and mother education are defined as completed school-leaving age minus 12, so that these variables give the number of years of schooling attained from the beginning of secondary education. Missing values are about 10.69% and 8.86% for father and mother respectively. In order to maximize the size of the sample used in the analysis, these missing values have been imputed, according to a regression based procedure, which adds a randomized residual to the predicted variables.[10] The imputation procedure exploits the individual controls listed above - i.e. controls with no missing - and mother and father education - which have missings. Intuitively, the imputation is performed such that, for observations with missing values in one parent's education but not in the education of his/her partner, the linear prediction for the missing values exploits also the information on his/her partner's education, in addition to the information provided by the controls with no missings. In contrast, for observations with missing value in both mother and father education, the linear predictions for each parental education exploits only information provided by the controls with no missings. An error term extracted from a logistic distribution is added to the linear predictions. Then, the probability of an outcome is computed, by transforming the corresponding predictions by the inverse logit function. Finally, to each missing we assign the outcome for which the predicted probability is maximum.

Imputation procedure for residence - Data on the province of residence are collected from the year in which individuals turn age 18 (which corresponds to year 1994, 1996 and 1998 for cohort born in year 1976, 1978 and 1980, respectively). Considering all the observation period, this information is missing in at least 1 year for 3.45%. In order to maximize the size of the sample used in the analysis, missings in residence are imputed if the following two conditions are met: (i) the socio-economic position of the individual is known in the year in which residence is missing (i.e. if he is employed, unemployed or inactive according to the socio-economic position variable provided from the Data Warehouse): this is an indication that he still resided in Belgium in that particular year; (ii) the residence does not change the year before and the year after missing(s). That is, missings in residence are not imputed if individuals have unknown socio-economic position in the year of missing residence (which means that they are abroad in that year), or if their residence changes before and after missing. Note that condition (i) is directly available from the socio-economic position variable, which attributes value 4 if the position in unknown (values 1, 2 or 3 are assigned to employed, unemployed or inactive status respectively). This procedure allows us to impute and hence retain 2.22% missing observations.

We further restrict to individuals residing in Flanders in the moment of graduation (1.1%) and drop individual graduating from age 25 onwards (2.15%). This selection leaves us with a sample of 3624 male individuals graduating at age 18-24. Last, to avoid complications in the estimation procedure, we restrict the sample to the following graduation periods: graduation year 1994-2001 for low-educated and graduation year 1997-2004 for high-educated. The final sample consists of 3514 male individuals. Descriptive statistics of the individual control variables for the final sample

[10]The addition of the error term aims at improving the regression-based imputation, which alone shows the following drawbacks: (i) distortion of variance, with respect to the trade-off between variance and bias; (ii) normative decision of the covariates for the predictions (Frick and Grabka, 2003; Kalwij and van Soest, 2005; Särndal and Lundström, 2005).

are reported in Table 1.1 of Appendix 1.8.

Finally, we consider 12 years of experience for the low educated and 10 years of experience for the high educated: low educated are followed longer since they graduate earlier. We make this selection based on the availability of the labor market outcomes, because we want to observe at least 4 graduation cohorts for a particular number of years of experience. The reason for that is that there needs to be some variation in the unemployment rate to identify the effect of the unemployment rate at graduation. Given that labor market outcomes are observed until year 2010 and we consider graduation year 1994-2001 (1997-2004) for low (high) educated, it follows that experience 12 (10) is observed for low (high) educated for the graduation cohorts 1994-1998 (1997-2000).

A.3 Description of the Final Sample

Table A.5: Final Sample by Graduation Year and Birth Cohort: Entire Sample.

Number of Individuals					Fraction of sample				
grad_year	c76	c78	c80	Total	grad_year	c76	c78	c80	Total
1994	143	0	0	143	1994	0.04	0.00	0.00	0.04
1995	209	0	0	209	1995	0.06	0.00	0.00	0.06
1996	168	149	0	317	1996	0.05	0.04	0.00	0.09
1997	154	240	0	394	1997	0.04	0.07	0.00	0.11
1998	185	203	187	575	1998	0.05	0.06	0.05	0.16
1999	191	137	242	570	1999	0.05	0.04	0.07	0.16
2000	115	163	163	441	2000	0.03	0.05	0.05	0.13
2001	0	153	154	307	2001	0.00	0.04	0.04	0.09
2002	0	138	142	280	2002	0.00	0.04	0.04	0.08
2003	0	0	172	172	2003	0.00	0.00	0.05	0.05
2004	0	0	106	106	2004	0.00	0.00	0.03	0.03
Total	1,165	1,183	1,166	3,514	Total	0.33	0.34	0.33	1.00

Table A.6: Dividing the Sample between Low and High Educated.

completed education§	low educated	high educated	Total
2	39	0	39
3	89	0	89
4	113	0	113
5	185	0	185
6	1,096	0	1,096
7	363	224	587
8	0	53	53
9	0	710	710
10	0	371	371
11	0	236	236
12	0	34	34
13	0	1	1
Total	1,885	1,629	3,514

§ Completed education refers to the number of years of education successfully attained from the beginning of secondary education, i.e. at age 12. Low educated are those who graduated with at most secondary education, which consists in 6 years of education (7 years in case of vocational track; 6 years for all other educational tracks). Therefore, low educated are those with at most 6 years of completed education if enrolled in general, technical, Part-Time vocational/apprenticeship program, and 7 years if enrolled in vocational education. High educated are those with higher completed education.

Table A.7: Final Sample: Number of Individuals by Graduation Year and Province of Residence at Graduation.

	Low educated						High educated					
grad_year	prov1	prov2	prov3	prov4	prov5	Total	prov1	prov2	prov3	prov4	prov5	Total
1994	30	9	31	48	25	143						
1995	47	22	44	48	48	209						
1996	84	45	65	85	38	317						
1997	78	41	65	67	36	287	29	14	24	19	21	107
1998	111	46	78	90	61	386	49	29	50	44	17	189
1999	99	42	47	64	47	299	56	54	61	63	37	271
2000	57	18	30	28	31	164	74	66	42	63	32	277
2001	26	8	11	17	18	80	62	39	45	48	33	227
2002							89	53	53	53	32	280
2003							54	35	24	33	26	172
2004							32	19	17	25	13	106
Total	532	231	371	447	304	1,885	445	309	316	348	211	1,629

The analysis considers the graduation period 1994-2001 and 1997-2004 for low and high educated, respectively. Provinces are in the following order from 1-5: Antwerp, Flemish Brabant, Western Flanders, Eastern Flanders, Limburg. Each combination of graduation year and province of residence at graduation represents a cluster *gp* in the main analysis. The number of individuals in each cluster corresponds to the cluster size.

Table A.8: Average Years of Completed Education by Graduation Year & Birth Cohort.

	Low educated				High educated			
grad_year	all cohorts	c76	c78	c80	all cohorts	c76	c78	c80
1994	5.02	5.02						
1995	5.91	5.91						
1996	5.61	6.03	5.14					
1997	5.84	6.04	5.77		8.09	8.41	7.00	
1998	5.43	5.95	6.04	4.80	8.78	9.05	7.12	
1999	6.10	6.15	6.16	6.08	9.13	9.58	8.55	7.00
2000	6.19	6.50	6.25	6.18	9.23	9.79	9.10	7.05
2001	6.19		5.80	6.21	9.40		9.86	8.53
2002					9.45		9.85	9.06
2003					9.76			9.76
2004					10.33			10.33
Total	5.75	5.76	5.75	5.74	9.28	9.28	9.23	9.31

The analysis considers the graduation period 1994-2001 and 1997-2004 for low and high educated, respectively. Completed education refers to the number of years of education successfully attained from the beginning of second education, i.e. at age 12.

Table A.9: Prevalent Function§ Undertaken in the Observation Period.

	Low educated			High educated		
	Freq.	Percent	Cum.	Freq.	Percent	Cum.
blue-collar	1,297	68.81	68.81	177	10.87	10.87
white-collar	432	22.92	91.72	1,306	80.17	91.04
fonctionnaire	78	4.14	95.86	81	4.97	96.01
missing	78	4.14	100	65	3.99	100
Total	1,885	100		1,629	100	

§ "Prevalent function" means the function that is undertaken more than 50% of the time in the observation period. 69% of the low educated are prevalently employed as blue collar workers, while for the high educated this figure is only 11%. Therefore, there is clear correspondence between low (high) educated and blue (white) collar workers.

A.4 Firm Mobility by Potential Experience.

Table A.10: Firm Mobility by Potential Experience.

	Low Educated					High Educated				
Pot. exp:	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
1	1,089	0.43	0.49	0	1	1,472	0.35	0.48	0	1
2	1,396	0.35	0.48	0	1	1,524	0.27	0.45	0	1
3	1,598	0.30	0.46	0	1	1,548	0.24	0.43	0	1
4	1,753	0.28	0.45	0	1	1,557	0.23	0.42	0	1
5	1,761	0.25	0.43	0	1	1,558	0.22	0.41	0	1
6	1,765	0.24	0.43	0	1	1,557	0.18	0.38	0	1
7	1,769	0.22	0.42	0	1	1,459	0.17	0.37	0	1
8	1,767	0.21	0.41	0	1	1,299	0.17	0.37	0	1
9	1,778	0.21	0.40	0	1	1,032	0.16	0.36	0	1
10	1,687	0.19	0.39	0	1	808	0.13	0.33	0	1
11	1,537	0.17	0.38	0	1					
12	1,245	0.17	0.37	0	1					
overall	19,145	0.25	0.43	0	1	13,814	0.22	0.41	0	1

The analysis considers the graduation period 1994-2001 and 1997-2004 for low and high educated, respectively.

Table A.11: Cumulative Firm Mobility by Experience (Proportion).

Low educated	Potential experience years												
cumulative firm mobility	1	2	3	4	5	6	7	8	9	10	11	12	Total
0	0.63	0.44	0.36	0.31	0.26	0.22	0.20	0.18	0.16	0.15	0.14	0.13	3.16
1	0.37	0.38	0.34	0.30	0.31	0.28	0.27	0.25	0.23	0.21	0.21	0.20	3.35
2	0.00	0.17	0.22	0.21	0.19	0.21	0.20	0.18	0.18	0.18	0.18	0.17	2.09
3	0.00	0.00	0.09	0.13	0.14	0.14	0.15	0.16	0.15	0.15	0.14	0.14	1.41
4	0.00	0.00	0.00	0.05	0.08	0.08	0.10	0.11	0.12	0.13	0.12	0.13	0.92
5	0.00	0.00	0.00	0.00	0.02	0.05	0.05	0.07	0.08	0.08	0.09	0.08	0.52
6	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.03	0.04	0.05	0.05	0.06	0.28
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.03	0.05	0.05	0.17
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.05
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.03
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	12.00
changed $>$ 2 out of 5 y§					0.25								
changed $>$ 3 out of 5 y					0.10								
changed $>$ 4 out of 5 y					0.02								
changed $>$ 2 out of 10 y										0.46			
changed $>$ 3 out of 10 y										0.31			
changed $>$ 4 out of 10 y										0.18			

High educated	Potential experience years										
cumulative firm mobility	1	2	3	4	5	6	7	8	9	10	Total
0	0.71	0.56	0.45	0.37	0.31	0.27	0.23	0.20	0.18	0.17	3.45
1	0.29	0.35	0.38	0.37	0.36	0.34	0.33	0.31	0.28	0.26	3.28
2	0.00	0.09	0.15	0.19	0.21	0.24	0.23	0.23	0.25	0.24	1.83
3	0.00	0.00	0.02	0.06	0.09	0.10	0.12	0.14	0.15	0.16	0.85
4	0.00	0.00	0.00	0.01	0.03	0.05	0.06	0.08	0.08	0.09	0.40
5	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.03	0.04	0.05	0.15
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.03
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	10.00
changed $>$ 2 out of 5 y§					0.12						
changed $>$ 3 out of 5 y					0.03						
changed $>$ 4 out of 5 y					0.00						
changed $>$ 2 out of 10 y										0.32	
changed $>$ 3 out of 10 y										0.16	
changed $>$ 4 out of 10 y										0.07	

The analysis considers the graduation period 1994-2001 and 1997-2004 for low and high educated, respectively. Moreover, the sample of low educated in this case is restricted to individuals who are observed continuously from experience 1 to 12. Note that in the sample of analysis not all low educated individuals are observed continuously throughout 12 years of experience because of availability of the data: labor market outcomes are provided for the period 1998-2010, while graduation period ranges 1994-2001 for low educated. This means that only individuals graduating from 1997 are observed from experience 1, which occurs in year 1998; in contrast, those for instance graduating in 1994 are observed from experience 4 (1994+4=1998). For high educated this is not a problem as the first graduation cohort graduates in 1997.

§ This statistic means "proportion who changed job in more than 2 out of the first 5 years after graduation": it is computed by summing up the cells in column 5 of the above matrix from row 3 until the last one. The other statistics are computed similarly.

A.5 Testing for the Endogeneity of the Graduation Cohort

A.5.1 Is the Timing of Graduation Endogenous?

According to economic theory the effect of economic conditions on the timing of graduation is ambiguous. On the one hand, a recession decreases the expected labor market income and, hence, the opportunity cost of education. On the other hand, it also reduces the expected returns to education and liquidity of the parents to finance education, so that early school leaving is enhanced. Existing empirical evidence is also mixed, but usually finds that unemployment raises the enrollment rate (see e.g. Card and Lemieux, 2001 and Clark, 2011). Micklewright et al. (1990) by contrast find that the regional unemployment rate tends to reduce the demand for schooling. Petrongolo and San Segundo (2002) and more recently Tumino and Taylor (2013a) report that the youth unemployment rate, as proxy for the opportunity cost, raises the probability of remaining in education, while the adult unemployment rate, as proxy for the returns, reduces this probability.

For our purpose it is important to rule out that the adult unemployment rate affects the timing of graduation. If this were the case, then it would affect the composition of the graduation cohort over the business cycle and any association between the unemployment rate and some labor market outcome could just reflect this variable composition rather than a causal effect. To test this we check whether the age of graduation is related to the provincial unemployment rate in that year. Since in Belgium education is compulsory until age 18, we can implement this test by estimating a discrete duration model in which we regress an indicator of graduating since age 17 on birth cohort dummies, individual characteristics x_i and the province of living measured at age 17, the elapsed duration in education since age 17, and the unemployment rate in each potential year of graduation (interacted with the elapsed duration), and by subsequently testing whether the coefficients of latter interactions are jointly significantly different from zero. We deal with selectivity induced by unobserved heterogeneity. The data are clustered in 15 clusters according to the birth year b (1976, 1978 or 1980) and the five provinces p of living at age 17. Problems of inference induced by the small number of clusters are solved in a similar two step approach as in the main analysis.[11]

We follow Kiefer (1988) and Jenkins (1995) to estimate the discrete duration model as a sequence of (yearly) binary choices from age 17 until age 24 ($a \in \{17, 18, ..., 24\}$).[12] In order to obtain correctly sized standard errors, we first regress the discrete-time hazard rate of graduating at a particular age on x_i and the group-age fixed effects $\mu^h_{bpa^*}$, where superscript h allows distinguishing these effects from the μ_{gpt} in the main analysis and $a^* \equiv a - 17$. In a second step the estimated $\hat{\mu}^h_{bpa^*}$ are then linearly regressed on the covariates that vary at the group-age level, one of which is the provincial unemployment rate.

In the first step, the conditional discrete-time hazard $h_{ibpa^*}(x_i, \epsilon^h)$ is assumed to take on the

[11] Notice that the size of all groups always satisfy the aforementioned rule of Cochran (1954), so that the asymptotic inference should work in this case.

[12] In order to maintain the same data as those that are used in the main analysis as well as to avoid problems of inference induced by too small cell sizes, we right censor duration at the end of the year in which individuals become 24.

complementary log-log specification:

$$h_{ibpa^*}(x_i,\epsilon^h) \equiv P(A^*_{ibp}=a^*|A^*_{ibp}\geq a^*;x_i,\epsilon^h) = 1-\exp\left[-\exp\left(\mu^h_{bpa^*}+x_i'\delta^h+\epsilon^h\right)\right] \tag{A.1}$$

where A^*_{ibp} is the random age (minus 17) at which individual i of birth cohort b and living at age 17 in province p graduates and ϵ^h is realization of a random individual unobserved heterogeneity term $\mathcal{E}^h$ that is independently distributed from x_i, b and p.

This model is estimated by maximum likelihood. To form the likelihood, note that the discrete survival rate at age a^* is simply $\prod_{s=1}^{a^*}\left(1-h_{ibps}(x_i,\epsilon^h)\right)$. Consequently, if c_i denotes an indicator that is equal to zero in case of right censoring, i.e. in case that individual i is still in education at the start of the calendar year in which he becomes 25 ($a^*=25-17=8$), and one otherwise. Then the log-likelihood function (from which the unobserved heterogeneity is integrated out) can be expressed as follows:

$$\log\mathcal{L}=\sum_{i=1}^{N}\log\int_{-\infty}^{\infty}\left[h_{ibpa^*}(x_i,\epsilon^h)\right]^{c_i}\prod_{s=1}^{a^*-1}\left(1-h_{ibps}(x_i,\epsilon^h)\right)dG(\epsilon^h) \tag{A.2}$$

where N denotes the total number of observations and $G(\epsilon^h)$ is the distribution of unobserved heterogeneity. We perform a sensitivity analysis in which (i) $\epsilon^h=0$, (ii) ϵ^h is Normally distributed with mean zero, or (iii) $\exp(\epsilon^h)$ is Gamma distributed with mean one.

In the second step, the following linear regression is estimated by FGLS according to the methods described in Section 2.4.4:

$$\hat{\mu}^h_{bpa^*}=\gamma^h_{a^*}+\beta^h_{a^*}u_{pt}+\eta^h_p+\lambda^h_b+v^h_{bpa^*} \tag{A.3}$$

where $t\equiv b+a$ is the year of potential graduation, $\gamma^h_{a^*}$ is an age specific fixed effect describing the evolution of the baseline hazard, η^h_p a provincial specific effect, λ^h_b a birth cohort fixed effect, and $v^h_{bpa^*}=e^h_{pba^*}+(\hat{\mu}^h_{bpa^*}-\mu^h_{bpa^*})$ is completely analogously defined as in Section 2.4.4. The parameters of interest are $\beta^h_{a^*}$. They measure the effect of the provincial unemployment rate u_{pt} on the hazard rate of graduating in that year. We test their joint significance.

Table A.12: Test for Exogenenous Timing of Graduation: Second Step FGLS.

	(1)	(2)	(3)
Unobserved heterogeneity (UH)	without UH	Gamma distributed	Normally distributed
Top panel: β^h_{a*} restricted to be equal for all ages			
β^h_{a*}	0.0239	0.0497	0.0438
	(0.0299)	(0.0347)	(0.0369)
Parameters in second step	14	14	14
Obs in second step	105	105	105
P-value of chi2 test	0.2983	0.7168	0.9251
Bottom panel: β^h_{a*} allowed to vary over ages			
β^h_{a*} at a*=1	-0.0039	-0.0021	0.0033
	(0.0488)	(0.0547)	(0.0640)
β^h_{a*} at a*=2	0.1172**	0.1401**	0.1513**
	(0.0504)	(0.0575)	(0.0642)
β^h_{a*} at a*=3	0.0319	0.0705	0.0683
	(0.0524)	(0.0559)	(0.0581)
β^h_{a*} at a*=4	0.0486	0.0647	0.0540
	(0.0613)	(0.0653)	(0.0671)
β^h_{a*} at a*=5	-0.0646	-0.0498	-0.0545
	(0.0619)	(0.0656)	(0.0675)
β^h_{a*} at a*=6	0.0452	0.0878	0.0794
	(0.0529)	(0.0592)	(0.0617)
β^h_{a*} at a*=7	-0.0908	-0.0377	-0.0514
	(0.0807)	(0.1048)	(0.1019)
Parameters in second step	20	20	20
Obs in second step	105	105	105
P-value of chi2 test	0.423	0.800	0.948
Test of joint significance of β^h_{a*} (P-val)†:	0.334	0.075	0.118
log variance of UH (first step)§	-	-0.0041	1.3160***
	-	(0.0827)	(0.2398)

Significance level: *** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. The table shows the effect of increasing the provincial unemployment u_{pt} by one *pp* on the hazard rate of graduating in that year. Since the event starts from age 18 and is right-censored after age 24, the baseline hazard $a^* = age - 17$ ranges from 1-7. The top panel reports the estimate of β^h_{a*} in (A.3); the bottom panel reports the estimates of β^h_{a*} obtained from estimating (A.3) in which the provincial unemployment rate u_{pt} is interacted with age specific FE of the baseline hazard. The estimates are obtained from a two-step FGLS described in detail in Section 2.4.4 and briefly below. First step: (A.1) is estimated by maximum likelihood, either neglecting the unobserved heterogeneity (UH) (column 1), or assuming that the UH is Gamma or Normally distributed (column 2 and 3). Second step: (A.3) is estimated by FGLS, where the inverse of the variance matrix of the $\hat{\mu}^h_{bpa*}$ in the first step is used as weight. If the χ^2 goodness-of-fit statistic rejects the model (P-value>0.05), standard errors clustered at the *bp* level are reported; otherwise conventional ones. These estimations are computed on the entire sample, i.e. without distinguishing between low and high educated.

† For the bottom panel it tests the null hypothesis that all β^h_{a*} are equal to zero.

§ The estimated log of the variance of the unobserved heterogeneity is obtained from the first step.

Table A.12 presents the outcome of the test regarding the exogeneity of the timing of graduation. We report for all three duration models (without, Gamma and Normally distributed unobserved heterogeneity) the parameters of interest in the second step FGLS regression of Equation (A.3), i.e. the coefficients γ^h_{a*} of the provincial unemployment rate u_{pt}, and the log of the variance of the distribution of unobserved heterogeneity (if applicable). In the top panel we report the results of the models in which we restrict all $\gamma^h_{a^*}$ to be equal over age, while the bottom panel the results of the unrestricted models are displayed.

According to the χ^2 goodness-of-fit statistics none of the six specifications can be rejected against the saturated model, so that the conventional standard errors are reported. The unrestricted model assuming Normally distributed unobserved heterogeneity provides the best fit to the data. In line with theory, the estimated coefficients in the models accounting for heterogeneity are in most cases larger in absolute value than in the models neglecting it. In the restricted models none of the parameters of interest are significantly different from zero. In the unrestricted models, a higher unemployment rate highly significantly accelerates school leaving at age 19 ($a^* = 2$). However, the coefficients at other ages are never significant, do not display any systematic pattern and can even have the opposite sign. Moreover, we do not have any clear explanation for this finding. We therefore argue that the significant result is obtained by chance. The fact that we cannot reject the hypothesis that all $\gamma^h_{a^*}$ are equal to zero (see the bottom line of bottom panel in Table A.12), is in line with this interpretation. We therefore conclude that the timing of graduation is exogenous to the business cycle. Further evidence for this conclusion is reported in Section A.9, where we show that our main findings are not sensitive to the inclusion of the completed number of years of education as a control variable.

A.5.2 Is the Province of Residence at Graduation Influenced by Local Labor Market Conditions?

Our identification strategy requires that youths (or their parents, if youth still live at their parents' house) do not move prior to graduation to provinces where the unemployment rate falls relatively to other provinces. For then the composition of recent graduates in provinces would be correlated with local labor market conditions, and it would no longer be possible to disentangle the effects of the latter from the former. To check whether this is a threat to identification, we measured the fraction of youth in our sample that has changed residence between the first year that our data inform about the place of living, i.e. on December 31 of the year in which the individual turns 17, and the moment at which the unemployment rate at graduation u_{gp}, our main regressor of interest, is measured, i.e. at the start of the year of graduation. Since only 0.44% of the individuals in the sample changed residence in that period, the issue can be safely ignored.

A.6 Complete Estimation Results for Low Educated

Table A.13: First Step Estimation: Low Educated.

Outcomes:	discrete		continuous			
	salaried	self-empl.	log wage	log hours	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
live in single-parent	0.0284	-0.0460	-0.0152	0.0481	-6.4318	34.1092
	(0.0516)	(0.0520)	(0.0177)	(0.0639)	(56.9915)	(24.6182)
not live with parents	0.0537**	-0.0436*	-0.0217	-0.0488	-54.1461	19.8639
	(0.0265)	(0.0251)	(0.0143)	(0.0460)	(41.9977)	(17.0054)
HH members aged 0-11	-0.0142	0.0136	0.0015	-0.0566**	-40.4398**	1.7354
	(0.0147)	(0.0143)	(0.0060)	(0.0241)	(18.8287)	(7.4735)
HH members aged 12-17	0.0147	-0.0138	-0.0043	0.0132	8.8881	1.0011
	(0.0111)	(0.0102)	(0.0058)	(0.0142)	(16.6629)	(7.6926)
HH members aged 18-29	-0.0008	0.0028	-0.0031	-0.0167	-16.9359	3.8838
	(0.0109)	(0.0102)	(0.0052)	(0.0130)	(14.8053)	(6.4400)
HH members aged 30-64	0.0037	-0.0114	0.0013	0.0561	46.8355	-3.4421
	(0.0457)	(0.0466)	(0.0137)	(0.0557)	(45.1878)	(17.7092)
HH members aged 65+	0.0132	-0.0106	-0.0041	0.0773**	64.2378*	-18.0643
	(0.0356)	(0.0347)	(0.0133)	(0.0306)	(37.0874)	(21.8596)
father education	-0.0027	0.0020	-0.0013	-0.0038	-6.4999*	3.7912**
	(0.0028)	(0.0026)	(0.0013)	(0.0032)	(3.6161)	(1.6774)
mother education	-0.0107***	0.0102***	-0.0018	-0.0092**	-11.1093***	3.6916*
	(0.0031)	(0.0030)	(0.0014)	(0.0040)	(4.0517)	(1.8991)
years of delay in sec.edu	0.0118	-0.0250**	-0.0116**	-0.0892***	-96.0414***	20.2979***
	(0.0110)	(0.0101)	(0.0053)	(0.0144)	(14.9899)	(7.1780)
technical edu†	0.0304	0.0189	0.0223*	0.1598***	137.5642***	-28.0554
	(0.0276)	(0.0262)	(0.0134)	(0.0402)	(41.6928)	(18.6938)
vocational edu†	0.0346	0.0210	-0.0179	0.0706*	-2.3482	-3.4916
	(0.0281)	(0.0267)	(0.0134)	(0.0401)	(41.5358)	(18.9772)
apprenticeship/PT voc†	-0.0460	0.0814**	-0.0187	0.1766***	89.3406	-20.8304
	(0.0430)	(0.0402)	(0.0186)	(0.0580)	(59.9006)	(27.6646)
birth cohort78††	0.0084	-0.0220	-0.0135	-0.0991***	-128.2615***	14.3591
	(0.0280)	(0.0273)	(0.0134)	(0.0321)	(35.1430)	(15.3584)
birth cohort80††	-0.0087	-0.0234	-0.0348**	-0.2619***	-305.1146***	55.8014**
	(0.0384)	(0.0370)	(0.0177)	(0.0486)	(51.1070)	(22.5020)
cluster-time FE§	*Not reported*					
Observations	32,466		34,164		34,164	
R-squared	0.7253		0.9928		0.8171	

Significance level: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors between parentheses. The table reports the results from estimating (3) in Section 2.4.4 on mentioned outcomes. This is the first step regression of the two-step FGLS approach, described in detail in the aforementioned Section. Since the outcomes satisfy adding-up constraints (salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked), this first step is estimated from an OLS SUR on the first two outcomes in the sum. For discrete outcomes in column 1-2, (3) is estimated by Linear Probability Model. Standard errors are clustered at the individual level. The individual control variables, measured when the individuals are aged 17, are reported in order: 1 if live with single parent; 1 if not live with either parents; number of other household members in different age classes; parent's education; years of delay in secondary education; choice of educational track in secondary education; birth cohort dummies. These variables are expressed in deviation from the mean. Cluster-time fixed effects $\hat{\mu}_{gpt}$ are not reported.

† Choice of educational track in secondary education: reference is general education.

†† Birth dummies: reference is born in 1976.

§ Cluster-time FE identify observations referring to individuals graduating in year g, living in province p at graduation and whose outcome is measured in calendar year y. The FE included in the regression on discrete outcomes differ from the ones included in the regression on continuous outcomes, because different selection rules are applied for discrete and continuous outcomes, and hence different groups are retained (see Table 1.2 in Appendix 1.8 and Section 2.4.4).

Table A.14: Second Step OLS Estimation: Low Educated.

Outcomes:	discrete		continuous			
	salaried	self-empl.	log wage	log hours§	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
URate_grad	-0.0115	0.0328	-0.0080	-0.0119	-24.9982	27.2606
	(0.0239)	(0.0221)	(0.0177)	(0.0414)	(42.3072)	(19.7239)
URate_grad*lin_exp	0.0128**	-0.0160***	0.0000	-0.0093	-10.0671	-2.8690
	(0.0057)	(0.0049)	(0.0052)	(0.0181)	(12.9797)	(9.2303)
URate_grad*lin_exp|exp>3	-0.0158**	0.0131**	-0.0016	0.0096	16.1812	-0.8707
	(0.0071)	(0.0061)	(0.0055)	(0.0206)	(15.0691)	(11.9548)
URate_grad*lin_exp|exp>6	0.0006	0.0051	0.0003		-9.3893	6.8536
	(0.0056)	(0.0057)	(0.0021)		(8.5323)	(5.7855)
URate_grad*lin_exp|exp>9	0.0061	-0.0052	0.0010	0.0008	1.3590	1.8411
	(0.0058)	(0.0040)	(0.0039)	(0.0080)	(10.9054)	(5.7275)
URate_grad*lin_exp*upturn				0.0021		
				(0.0114)		
URate_grad*lin_exp|exp>3*upturn				0.0012		
				(0.0168)		
URate_grad*lin_exp|exp>6*upturn						
URate_grad*lin_exp|exp>9*upturn				0.0071		
				(0.0108)		
d_exp1	0.7512***	0.1165**	2.3176***	6.9075***	958.9185***	298.9458***
	(0.0457)	(0.0457)	(0.0356)	(0.1224)	(126.9019)	(94.2751)
d_exp2	0.7461***	0.1226***	2.3619***	7.0090***	1,123.5383***	228.2590***
	(0.0438)	(0.0427)	(0.0312)	(0.1034)	(111.9305)	(80.9229)
d_exp3	0.7618***	0.1226***	2.4055***	7.0345***	1,217.5205***	164.6059***
	(0.0378)	(0.0381)	(0.0262)	(0.0861)	(88.0276)	(58.8217)
d_exp4	0.7607***	0.1386***	2.4389***	7.0675***	1,283.7424***	137.2792***
	(0.0352)	(0.0373)	(0.0230)	(0.0673)	(73.2986)	(42.1305)
d_exp5	0.7707***	0.1459***	2.4769***	7.0772***	1,307.9451***	126.0374***
	(0.0347)	(0.0367)	(0.0199)	(0.0623)	(65.3467)	(30.5050)
d_exp6	0.7710***	0.1474***	2.5213***	7.1275***	1,403.3019***	83.4630***
	(0.0317)	(0.0337)	(0.0191)	(0.0684)	(72.0378)	(29.6944)
d_exp7	0.7754***	0.1597***	2.5574***	7.1523***	1,447.5870***	66.9784
	(0.0291)	(0.0303)	(0.0212)	(0.0751)	(83.4648)	(40.3594)
d_exp8	0.7781***	0.1707***	2.6028***	7.1260***	1,444.8401***	64.4567
	(0.0306)	(0.0290)	(0.0249)	(0.0971)	(105.4114)	(55.5343)
d_exp9	0.7633***	0.1883***	2.6408***	7.1943***	1,508.9936***	40.3436
	(0.0323)	(0.0292)	(0.0292)	(0.1112)	(126.6040)	(74.3594)
d_exp10	0.7637***	0.1918***	2.6761***	7.1655***	1,495.2992***	39.0506
	(0.0342)	(0.0279)	(0.0341)	(0.1275)	(155.2295)	(94.9612)
d_exp11	0.7620***	0.2083***	2.7130***	7.1683***	1,529.4159***	11.1327
	(0.0373)	(0.0292)	(0.0401)	(0.1555)	(182.0271)	(115.4204)
d_exp12	0.7492***	0.2204***	2.7425***	7.1461***	1,544.8297***	-12.8784
	(0.0437)	(0.0318)	(0.0467)	(0.1729)	(207.2307)	(133.9152)
current_URate*lin_exp			-0.0024	-0.0127	-13.1490	4.6741
			(0.0026)	(0.0079)	(10.2617)	(6.0726)
current_URate*lin_exp|exp>3			0.0021	0.0195	18.4495	-5.8028
			(0.0047)	(0.0140)	(17.0719)	(9.9964)
current_URate*lin_exp|exp>6			0.0005	0.0056	1.0209	2.1206

Continued on next page

Table A.14 – continued from previous page

	salaried	self-empl.	log wage	log hours§	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
			(0.0040)	(0.0117)	(13.8847)	(6.9309)
current_URate*lin_exp|exp>9			-0.0010	-0.0291**	-4.2898	-10.2813
			(0.0046)	(0.0127)	(15.8712)	(8.0814)
lin_grad_year	0.0403***	-0.0214	0.0279***	0.0472	65.4507**	-8.6015
	(0.0137)	(0.0138)	(0.0073)	(0.0320)	(29.6637)	(21.6567)
lin_grad_year|trend>3	-0.0240	0.0215	-0.0069	0.0140	15.5502	-11.5497
	(0.0179)	(0.0169)	(0.0054)	(0.0249)	(25.3218)	(14.1348)
lin_grad_year|trend>6	-0.0132	0.0144	-0.0155	0.0307	6.5676	11.5773
	(0.0285)	(0.0283)	(0.0126)	(0.0221)	(32.5098)	(20.9785)
d_y2000			-0.0125	-0.0172	-38.8673	26.1713
			(0.0104)	(0.0543)	(44.5270)	(28.7443)
d_y2001			-0.0182	-0.0247	-74.2369	49.9492
			(0.0164)	(0.0820)	(75.5566)	(47.9611)
d_y2002			-0.0126	-0.0565	-102.6201	49.9582
			(0.0219)	(0.0946)	(92.2693)	(64.9251)
d_y2003			0.0200	-0.0836	-185.5176	68.4335
			(0.0287)	(0.1172)	(116.2514)	(85.2820)
d_y2004			0.0018	-0.0544	-187.7900	76.6086
			(0.0353)	(0.1384)	(145.0028)	(105.8795)
d_y2005			-0.0177	-0.0954	-231.4900	91.1613
			(0.0406)	(0.1591)	(171.9990)	(126.0796)
d_y2006			-0.0366	-0.0984	-263.5384	116.7552
			(0.0475)	(0.1884)	(200.9022)	(145.7559)
d_y2007			-0.0440	-0.0899	-257.7645	127.6621
			(0.0550)	(0.2165)	(229.4389)	(163.9872)
d_y2008			-0.0730	-0.0695	-272.8240	147.2443
			(0.0625)	(0.2441)	(262.2820)	(186.8023)
d_y2009			-0.0610	-0.1343	-359.7139	160.6956
			(0.0687)	(0.2624)	(286.1192)	(206.2585)
d_y2010			-0.0785	-0.0978	-348.4011	181.8591
			(0.0745)	(0.2868)	(315.8287)	(229.1536)
lin_calend_year|trend>3	-0.0132*	-0.0011				
	(0.0067)	(0.0047)				
lin_calend_year|trend>6	0.0027	0.0051				
	(0.0052)	(0.0044)				
lin_calend_year|trend>9	0.0023	-0.0064				
	(0.0041)	(0.0047)				
d_province2	-0.0467*	0.0323	-0.0192	-0.0614	-202.5254**	122.4041***
	(0.0254)	(0.0225)	(0.0220)	(0.0772)	(98.6645)	(39.8929)
d_province3	-0.0599*	0.0782***	0.0169	0.0557	15.7333	30.6915
	(0.0302)	(0.0276)	(0.0201)	(0.0576)	(73.9502)	(34.2322)
d_province4	-0.0299*	0.0355**	0.0258	0.0107	-49.9896	33.2321*
	(0.0169)	(0.0155)	(0.0169)	(0.0361)	(41.1422)	(17.4238)
d_province5	-0.0578***	0.0616***	0.0242	0.1082	48.6183	15.0094
	(0.0206)	(0.0217)	(0.0210)	(0.0734)	(88.1180)	(39.9302)
lin_calend_year_province2			-0.0081***	0.0086	14.2801	-6.1889
			(0.0027)	(0.0088)	(13.7433)	(5.7350)
lin_calend_year_province3			-0.0092***	-0.0036	-2.7219	-3.3067
			(0.0029)	(0.0065)	(8.7463)	(3.7209)
lin_calend_year_province4			-0.0027	0.0017	4.1164	-2.1860
			(0.0027)	(0.0045)	(5.8421)	(3.3370)
lin_calend_year_province5			-0.0030	-0.0121	-8.2976	-2.0106

Continued on next page

Table A.14 – continued from previous page

	salaried	self-empl.	log wage	log hours§	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
			(0.0027)	(0.0081)	(8.9590)	(3.5839)
Observations	564		756		756	
Amemiya-Nold test†	-0.0015	-0.0013	-0.0004	-0.0020	-1826.0490	-526.2726
R-squared	0.9941		0.9998		0.9944	

Significance level: *** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. The table reports the results from estimating (4) in Section 2.4.4 by OLS on mentioned outcomes. This is the second step regression of the two-step approach described in detail in the aforementioned Section, where this second step is estimated by OLS. Since the outcomes satisfy adding-up constraints (salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked), this second step is estimated from a OLS SUR on the first two outcomes in the sum. Effects on the third outcome (the sum) are then obtained from the adding-up constraints. Standard errors are clustered at graduation time-province (gp) level. The dependent variables of the second step are the cluster-time FE $\hat{\mu}_{gpt}$ estimated in each corresponding first step regression, given by (3) in Section 2.4.4. The covariates of the second step are described below (in order): $URate\_grad$ is the unemployment rate at graduation u_{gp}. $URate\_grad * lin\_exp$ is an interaction between u_{gp} and a linear trend in experience. $URate\_grad * lin\_exp|exp > 3$ is u_{gp} interacted with a linear trend in experience that starts from exp>3. The subsequent 2 covariates are defined similarly. Variables $URate\_grad * lin\_exp$-$URate\_grad * lin\_exp|exp > 9$ specify the linear spline in potential experience $f_g(e)$ that multiplies the provincial unemployment rate at graduation u_{gp}. The slope of the spline changes at experience 3,6 and 9 for low educated, as explained in (2) in Section 5.1 of the main text. Variables $URate\_grad * lin\_exp * upturn$-$URate\_grad * lin\_exp|exp > 9 * upturn$ are triple interactions of u_{gp} with $f_{gu}(e)$ and a dummy=1 if graduation occurs in an upturn. They allows u_{gp} to have a different effect over experience in upturn/downturn. d_exp1-d_exp12 are experience FE. Since they define constant-terms for each experience level, the constant is omitted in the specification. $current\_URate * lin\_exp$-$current\_URate * lin\_exp|exp > 9$ are interactions between the current unemployment rate u_{tp} with $f_t(e)$. Variables lin_grad_year-$lin\_grad\_year|trend > 6$ is the spline for the graduation year $f_0(g)$. d_y2000-d_y2010 are calendar year FE for the period 1998-2010. Identification requires dropping both year 1998 and 1999: the first is the reference FE, while the second has to be dropped because of the following accounting identity: $calend\_year = exp + grad\_year$. In this choice of dropping a second calendar year we followed Oreopoulos et al. (2012). Calendar year is alternatively specified with a spline, by means of variables $lin\_calend\_year|trend > 3$-$lin\_calend\_year|trend > 9$, where the first term of the spline is omitted because of the aforementioned accounting identity. $d\_province2$-$d\_province5$ are province FE (province 1 is the reference). $lin\_calend\_year * d\_province2$-$lin\_calend\_year * d\_province5$ are province-specific linear time trends. Depending on the outcome, we impose restrictions which cannot be jointly rejected at the 5% level, as follows: the effect of u_{gp} is restricted to be symmetric in upturn and downturn (for salaried and self-employment, log hourly wage, full-time and part-time hours worked); the effect of u_{tp} is set to be zero (for salaried and self-employment); calendar year FE are specified with a spline (for salaried and self-employment); provincial time trends are set to be zero (for salaried and self-employment).

§ For log hours worked, it is also imposed $\beta_{g2} = 0$, i.e. the slope of the linear spline remains fixed after 6 years of experience. This restriction cannot be rejected.

† The Amemiya-Nold test is an estimate of the variance of the cluster-time errors e_gpt, described in (5) in Section 2.4.4. A negative statistics is evidence that the unobserved cluster-time shocks are indeed zero. Accordingly, one can implement the FGLS procedure as in Wooldridge (2006, 2010).

Table A.15: Second Step FGLS Estimation: Low Educated.

Outcomes:	discrete		continuous			
	salaried	self-empl.	log wage	log hours§	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
URate_grad	-0.0238	0.0501**	0.0100	-0.0525*	-164.8746***	56.8591***
	(0.0220)	(0.0236)	(0.0154)	(0.0275)	(39.6889)	(16.4111)
URate_grad*lin_exp	0.0063	-0.0144*	-0.0055	0.0087	41.1335***	-20.5795***
	(0.0063)	(0.0075)	(0.0037)	(0.0092)	(8.4565)	(5.3600)
URate_grad*lin_exp|exp>3	-0.0040	0.0087	0.0040	-0.0085	-30.4640***	18.6047***
	(0.0079)	(0.0089)	(0.0037)	(0.0118)	(8.9870)	(6.7193)

Continued on next page

Table A.15 – continued from previous page

	salaried	self-empl.	log wage	log hours§	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
URate_grad*lin_exp|exp>6	-0.0026	0.0074*	0.0042**		-24.5012***	10.5285**
	(0.0047)	(0.0042)	(0.0018)		(6.8968)	(3.9159)
URate_grad*lin_exp|exp>9	0.0017	-0.0057*	-0.0058**	-0.0030	19.8884***	-8.1435*
	(0.0034)	(0.0033)	(0.0022)	(0.0079)	(6.9654)	(4.0476)
URate_grad*lin_exp*upturn				-0.0108		
				(0.0075)		
URate_grad*lin_exp|exp>3*upturn				0.0183		
				(0.0117)		
URate_grad*lin_exp|exp>6*upturn						
URate_grad*lin_exp|exp>9*upturn				0.0013		
				(0.0115)		
d_exp1	0.8569***	0.0796	2.2692***	7.1194***	1,225.9931***	182.0557***
	(0.0646)	(0.0637)	(0.0378)	(0.1094)	(126.8483)	(58.2625)
d_exp2	0.8516***	0.0793	2.3175***	7.2186***	1,415.8790***	91.2670
	(0.0639)	(0.0617)	(0.0359)	(0.0974)	(114.1311)	(54.6505)
d_exp3	0.8457***	0.0847	2.3647***	7.2388***	1,465.8398***	73.3560
	(0.0623)	(0.0600)	(0.0350)	(0.0838)	(105.5893)	(50.3020)
d_exp4	0.8543***	0.0972	2.4002***	7.2710***	1,500.8617***	66.5529
	(0.0613)	(0.0599)	(0.0347)	(0.0741)	(99.0005)	(47.0799)
d_exp5	0.8612***	0.1073*	2.4316***	7.2705***	1,525.3920***	81.6274*
	(0.0609)	(0.0595)	(0.0343)	(0.0647)	(91.6504)	(44.6073)
d_exp6	0.8524***	0.1058*	2.4691***	7.3052***	1,532.2297***	94.6461**
	(0.0602)	(0.0594)	(0.0348)	(0.0572)	(86.7440)	(40.3467)
d_exp7	0.8530***	0.1226**	2.5001***	7.3216***	1,543.8087***	95.9874**
	(0.0601)	(0.0594)	(0.0346)	(0.0525)	(84.4805)	(38.9072)
d_exp8	0.8633***	0.1165*	2.5330***	7.3568***	1,520.6944***	109.3179***
	(0.0605)	(0.0602)	(0.0354)	(0.0570)	(87.0377)	(36.0375)
d_exp9	0.8576***	0.1263**	2.5667***	7.3706***	1,528.2080***	141.4640***
	(0.0614)	(0.0612)	(0.0369)	(0.0624)	(89.3094)	(34.8296)
d_exp10	0.8578***	0.1340**	2.5935***	7.3918***	1,530.9900***	152.9140***
	(0.0632)	(0.0628)	(0.0394)	(0.0699)	(100.0108)	(37.5158)
d_exp11	0.8569***	0.1404**	2.6227***	7.4047***	1,534.6193***	148.0725***
	(0.0652)	(0.0652)	(0.0425)	(0.0844)	(112.5650)	(39.2476)
d_exp12	0.8533***	0.1350**	2.6424***	7.4290***	1,524.9435***	148.2635***
	(0.0683)	(0.0677)	(0.0461)	(0.0995)	(128.1002)	(43.1278)
current_URate*lin_exp			-0.0008	-0.0073**	-2.7920	4.4547*
			(0.0011)	(0.0035)	(3.9967)	(2.5871)
current_URate*lin_exp|exp>3			-0.0011	0.0070	4.0244	-7.0951
			(0.0021)	(0.0065)	(8.6607)	(5.0969)
current_URate*lin_exp|exp>6			0.0025	-0.0032	8.6961	3.2832
			(0.0024)	(0.0058)	(9.5344)	(3.8670)
current_URate*lin_exp|exp>9			0.0020	0.0117*	-9.0042	-5.1401
			(0.0029)	(0.0067)	(7.4738)	(3.7716)
lin_grad_year	0.0173	-0.0100	0.0396***	0.0264	19.0216	16.0035
	(0.0201)	(0.0197)	(0.0102)	(0.0258)	(35.5687)	(15.2251)
lin_grad_year|trend>3	-0.0012	0.0030	-0.0235*	0.0277*	21.4500	-10.2731
	(0.0221)	(0.0218)	(0.0124)	(0.0164)	(37.5160)	(14.2848)
lin_grad_year|trend>6	-0.0152	0.0060	0.0087	-0.0004	44.8529	-23.3993**
	(0.0283)	(0.0270)	(0.0157)	(0.0200)	(37.5199)	(11.1892)
d_y2000			-0.0089	-0.0207	-21.8402	0.0906
			(0.0066)	(0.0233)	(23.2243)	(10.8590)
d_y2001			-0.0108	-0.0419	-5.2062	-20.1063

Continued on next page

Table A.15 – continued from previous page

	salaried	self-empl.	log wage	log hours§	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
			(0.0116)	(0.0399)	(38.5634)	(17.5764)
d_y2002			-0.0049	-0.0355	10.6677	-53.7633***
			(0.0151)	(0.0489)	(54.0165)	(19.3777)
d_y2003			0.0292	-0.0821	-69.3960	-74.0548***
			(0.0199)	(0.0697)	(70.6908)	(25.9086)
d_y2004			0.0161	-0.0719	-29.2518	-92.8033***
			(0.0242)	(0.0818)	(85.3414)	(30.5642)
d_y2005			-0.0009	-0.0818	-56.0160	-114.5512***
			(0.0289)	(0.0964)	(104.9215)	(36.2211)
d_y2006			-0.0117	-0.1070	-16.5626	-126.2454***
			(0.0340)	(0.1119)	(121.1634)	(41.2826)
d_y2007			-0.0170	-0.1288	4.0424	-132.1354***
			(0.0391)	(0.1307)	(140.5737)	(47.5882)
d_y2008			-0.0315	-0.1495	3.5323	-121.5956**
			(0.0453)	(0.1484)	(159.4959)	(54.3311)
d_y2009			-0.0187	-0.1670	-76.4815	-149.4521**
			(0.0505)	(0.1632)	(173.8986)	(58.7294)
d_y2010			-0.0378	-0.1970	-54.4050	-155.5810**
			(0.0565)	(0.1761)	(192.4969)	(64.9975)
lin_calend_year|trend>3	-0.0057	-0.0005				
	(0.0063)	(0.0059)				
lin_calend_year|trend>6	-0.0064	0.0073*				
	(0.0048)	(0.0043)				
lin_calend_year|trend>9	0.0027	-0.0025				
	(0.0045)	(0.0039)				
d_province2	-0.0448	0.0339	-0.0169	-0.0166	-33.2013	5.9581
	(0.0276)	(0.0262)	(0.0217)	(0.0526)	(92.8610)	(39.2545)
d_province3	-0.0710***	0.0677***	0.0423**	0.0635	156.8683**	-63.1090**
	(0.0228)	(0.0223)	(0.0206)	(0.0504)	(68.5089)	(30.5171)
d_province4	-0.0391**	0.0319**	0.0434***	0.0267	29.4486	-24.5258
	(0.0167)	(0.0155)	(0.0148)	(0.0412)	(56.6118)	(24.7125)
d_province5	-0.0382	0.0407	0.0201	0.1331**	198.9715**	-26.1553
	(0.0272)	(0.0262)	(0.0181)	(0.0598)	(76.9946)	(32.3431)
lin_calend_year_province2			-0.0052**	0.0035	14.3407*	-1.9295
			(0.0022)	(0.0050)	(7.3382)	(3.5427)
lin_calend_year_province3			-0.0058***	-0.0115**	-14.5633**	6.0920*
			(0.0020)	(0.0052)	(6.0903)	(3.4753)
lin_calend_year_province4			0.0004	-0.0030	-5.0574	3.8559
			(0.0022)	(0.0046)	(4.8954)	(2.8497)
lin_calend_year_province5			-0.0028	-0.0121**	-11.7671*	-0.4693
			(0.0022)	(0.0058)	(6.7259)	(2.7682)
R-squared	0.9986		0.9999		0.9953	
WSSR (2nd step)	331		1289		1519	
Obs (2nd step)	375		756		754	
Parameters (2nd step)	54		88		86	
Test joint signif. all imposed restr.(P-val)	0.286		0.155		0.268	
P-value of chi2 test	0.341		0.000		0.000	
cluster (at which level)	no		g*p		g*p	
Imposed Restrictions:						
effect URate at grad. symmetric up/downturn	yes	yes	yes	no	yes	yes
effect Current Urate over exp=0	yes	yes	no	no	no	no
spline for calendar year FE	yes	yes	no	no	no	no

Continued on next page

Table A.15 – continued from previous page

	salaried	self-empl.	log wage	log hours§	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
effect prov-time trends=0	yes	yes	no	no	no	no

Significance level: *** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. The table reports the results from estimating (4) in Section 2.4.4 by FGLS on mentioned outcomes. This is the second step regression of the two-step FGLS approach described in detail in the aforementioned Section. Since the outcomes satisfy adding-up constraints (salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked), this second step is estimated from a FGLS SUR on the first two outcomes in the sum. Effects on the third outcome (the sum) are then obtained from the adding-up constraints. The dependent variables of the second step are the cluster-time FE $\hat{\mu}_{gpt}$ estimated in each corresponding first step regression, given by (3) in Section 2.4.4. A description of the covariates reported in the Table is in the footnote of Table A.14 above. In this FGLS SUR, the data are weighted by the inverse of the cluster-robust variance matrix of $\hat{\mu}_{gpt}$ estimated in the first step. For discrete outcomes, the Moore-Penrose generalized inverse of this matrix is used as weight, to take into account the perfect serial correlation induced by the fact that, for specific clusters, the outcomes do not vary over time (see Section 2.4.4 for details). Depending on the outcome, we impose restrictions which cannot be jointly rejected at the 5% level: these restrictions are listed in the bottom panel of the table. If the χ^2 goodness-of-fit statistic rejects the model (P-value>0.05), standard errors clustered at the *gp* level are reported; otherwise conventional ones. For completeness we report also the weighted sum of squared residuals (WSSR), the number of estimated parameters and the number of observations, which are used to compute the χ^2 test.

§ For log hours worked the following additional restriction (not mentioned in the table) is also imposed: $\beta_{g2} = 0$, i.e. the slope of the linear spline remains fixed after 6 years of experience. This restriction cannot be rejected.

A.7 Complete Estimation Results for High Educated

Table A.16: First Step Estimation: High Educated.

Outcomes:	discrete		continuous			
	salaried	self-empl.	log wage	log hours	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
live in single-parent	-0.0142	0.0185	-0.0108	-0.0948	-58.7515	-23.5552
	(0.1065)	(0.1089)	(0.0494)	(0.0769)	(97.1894)	(38.6768)
not live with parents	0.0055	-0.0545	-0.0066	-0.0186	12.8753	-13.5174
	(0.0458)	(0.0384)	(0.0281)	(0.0462)	(66.2764)	(31.5662)
HH members aged 0-11	-0.0093	0.0050	-0.0024	-0.0281	-59.1713*	24.8043
	(0.0202)	(0.0194)	(0.0117)	(0.0220)	(30.7132)	(16.6696)
HH members aged 12-17	0.0071	-0.0014	0.0117	-0.0212	-34.8318*	14.4708
	(0.0149)	(0.0144)	(0.0080)	(0.0134)	(20.0060)	(10.9179)
HH members aged 18-29	0.0003	0.0014	0.0030	-0.0343**	-31.4121*	5.0918
	(0.0132)	(0.0129)	(0.0075)	(0.0140)	(18.9413)	(9.4834)
HH members aged 30-64	0.0055	0.0163	-0.0120	-0.0171	32.9106	-53.7095**
	(0.0990)	(0.1021)	(0.0467)	(0.0626)	(80.2774)	(26.7216)
HH members aged 65+	0.0113	-0.0076	-0.0168	0.0506*	70.0408	-28.3348
	(0.0456)	(0.0421)	(0.0215)	(0.0271)	(62.1322)	(35.4918)
father education	-0.0048	0.0034	0.0048***	-0.0010	4.2756	-4.3901*
	(0.0034)	(0.0033)	(0.0018)	(0.0028)	(4.3456)	(2.3938)
mother education	-0.0043	0.0023	-0.0009	-0.0113***	-16.0538***	6.2931**
	(0.0039)	(0.0038)	(0.0019)	(0.0033)	(4.8368)	(2.6061)
years of delay in sec.edu	0.0099	-0.0070	-0.0703***	-0.0562***	-83.9288***	20.8350
	(0.0168)	(0.0165)	(0.0097)	(0.0181)	(23.2658)	(12.6673)
technical edu†	0.0130	-0.0057	-0.0276**	0.0158	5.8643	2.7490
	(0.0217)	(0.0211)	(0.0117)	(0.0208)	(30.7576)	(16.8961)
vocational edu†	0.0398	-0.0062	-0.0990**	0.0463	-61.2107	90.4418
	(0.0846)	(0.0822)	(0.0417)	(0.0741)	(146.4928)	(84.9388)
apprenticeship/PT voc†	-0.6675***	0.7161***	-0.1879***	0.0344	-323.0605***	265.5299***
	(0.0551)	(0.0537)	(0.0341)	(0.0450)	(73.6523)	(41.9571)
birth cohort78††	0.0193	-0.0259	-0.0873***	-0.0242	-10.6798	-15.6678
	(0.0274)	(0.0259)	(0.0149)	(0.0283)	(37.6116)	(20.8565)
birth cohort80††	0.0735**	-0.0810**	-0.1795***	-0.0664*	-100.0433*	14.4197
	(0.0359)	(0.0337)	(0.0213)	(0.0395)	(53.7439)	(28.6471)
cluster-time FE§	*Not reported*					
Observations	22,886		24,314		24,314	
R-squared	0.7108		0.9951		0.8345	

Significance level: *** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. The table reports the results from estimating (3) in Section 2.4.4 on mentioned outcomes. This is the first step regression of the two-step FGLS approach, described in detail in this Section. Since the outcomes satisfy adding-up constraints (salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked), this first step is estimated from an OLS SUR on the first two outcomes in the sum. For discrete outcomes in column 1-2, (3) is estimated by Linear Probability Model. Standard errors are clustered at the individual level. The individual control variables, measured when the individuals are aged 17, are reported in order: 1 if live with single parent; 1 if not live with either parents; number of other household members in different age classes; parent's education; years of delay in secondary education; choice of educational track in secondary education; birth cohort dummies. These variables are expressed in deviation from the mean. Cluster-time fixed effects $\hat{\mu}_{gpt}$ are not reported.

† Choice of educational track in secondary education: reference is general education.

†† Birth dummies: reference is born in 1976.

§ Cluster-time FE identify observations referring to individuals graduating in year g, living in province p at graduation and whose outcome is measured in calendar year y. Note that the FE included in the regression on discrete outcomes differ from the ones included in the regression on continuous outcomes, because different selection rules are applied for discrete and continuous outcomes, and hence different groups are retained (see Table 1.2 in Appendix 1.8 and Section 2.4.4).

Table A.17: Second Step OLS Estimation: High Educated.

Outcomes:	discrete		continuous			
	salaried	self-empl.	log wage	log hours	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
URate_grad	-0.0427**	0.0355*	-0.0088	-0.0835	-91.4880*	32.3351
	(0.0169)	(0.0201)	(0.0098)	(0.0526)	(45.3291)	(19.2652)
URate_grad*lin_exp	0.0169***	-0.0145**	-0.0027	0.0311*	30.9963**	-5.5913
	(0.0047)	(0.0068)	(0.0026)	(0.0162)	(13.3616)	(6.8823)
URate_grad*lin_exp|exp>3	-0.0209***	0.0199***	0.0039	-0.0217	-25.9922*	-0.2021
	(0.0060)	(0.0072)	(0.0040)	(0.0184)	(14.6282)	(9.9671)
URate_grad*lin_exp|exp>6	0.0058	-0.0084	0.0017	-0.0224**	4.8011	2.7938
	(0.0061)	(0.0053)	(0.0031)	(0.0099)	(20.9446)	(9.5876)
URate_grad*lin_exp*upturn				0.0129	5.6285	
				(0.0110)	(11.3489)	
URate_grad*lin_exp|exp>3*upturn				-0.0351*	-10.2750	
				(0.0205)	(23.0995)	
URate_grad*lin_exp|exp>6*upturn				0.0356**	-9.1289	
				(0.0135)	(27.8938)	
d_exp1	0.9689***	0.0333	2.3866***	7.4621***	1,501.2335***	230.6942***
	(0.0430)	(0.0461)	(0.0203)	(0.1105)	(120.6049)	(46.0044)
d_exp2	0.9292***	0.0619	2.4581***	7.4502***	1,646.6508***	136.5209***
	(0.0442)	(0.0480)	(0.0245)	(0.0717)	(90.1631)	(34.4444)
d_exp3	0.9159***	0.0690	2.5161***	7.3154***	1,638.2554***	79.8997*
	(0.0435)	(0.0485)	(0.0275)	(0.0690)	(76.6823)	(43.6743)
d_exp4	0.8922***	0.0885	2.5699***	7.1631***	1,621.2628***	25.6041
	(0.0483)	(0.0524)	(0.0339)	(0.1158)	(98.2383)	(70.0681)
d_exp5	0.8657***	0.1114**	2.6220***	7.0046***	1,567.8224***	-17.4379
	(0.0497)	(0.0529)	(0.0431)	(0.1781)	(142.7774)	(102.0089)
d_exp6	0.8294***	0.1342**	2.6595***	6.8335***	1,516.2816***	-54.8387
	(0.0527)	(0.0552)	(0.0517)	(0.2477)	(193.0891)	(135.4781)
d_exp7	0.8024***	0.1582***	2.6984***	6.6729***	1,442.1611***	-74.7287
	(0.0557)	(0.0573)	(0.0604)	(0.3079)	(240.9555)	(169.6860)
d_exp8	0.7659***	0.1857***	2.7469***	6.5074***	1,391.0482***	-116.9754
	(0.0569)	(0.0598)	(0.0681)	(0.3783)	(296.0228)	(202.4608)
d_exp9	0.7368***	0.2091***	2.7816***	6.3768***	1,348.6060***	-142.4168
	(0.0603)	(0.0632)	(0.0767)	(0.4463)	(352.6231)	(236.8329)
d_exp10	0.7055***	0.2376***	2.8009***	6.1864***	1,253.7807***	-169.8802
	(0.0671)	(0.0678)	(0.0865)	(0.5135)	(402.3159)	(271.3453)
current_URate*lin_exp					-0.9215	
					(6.7461)	
current_URate*lin_exp|exp>3					5.4568	
					(16.1967)	
current_URate*lin_exp|exp>6					-5.0503	
					(13.4528)	
lin_grad_year	-0.0246	0.0090	0.0514***	-0.1554**	-69.1863	-16.0394
	(0.0162)	(0.0174)	(0.0068)	(0.0655)	(55.3554)	(29.6588)
lin_grad_year|trend>3	-0.0051	0.0193	-0.0038	0.0002	22.9224	-17.7204
	(0.0209)	(0.0222)	(0.0124)	(0.0144)	(30.7012)	(22.1972)
lin_grad_year|trend>6	-0.0332*	0.0270	-0.0052	0.0206	42.8455	-21.7279
	(0.0195)	(0.0176)	(0.0118)	(0.0214)	(31.9825)	(18.6639)
d_y2000			-0.0243*	0.2275**	154.9824*	7.2023
			(0.0140)	(0.1076)	(81.1079)	(48.7754)
d_y2001			-0.0399*	0.3992**	231.8595	26.4881
			(0.0211)	(0.1788)	(139.5020)	(78.2480)

Continued on next page

Table A.17 – continued from previous page

	salaried	self-empl.	log wage	log hours	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
d_y2002			-0.0425	0.5822**	290.4235	88.4649
			(0.0290)	(0.2439)	(183.7660)	(110.8750)
d_y2003			0.0145	0.6609**	124.6172	187.8584
			(0.0360)	(0.3013)	(235.8363)	(144.9946)
d_y2004			-0.0048	0.8437**	208.4257	213.7200
			(0.0449)	(0.3674)	(288.0389)	(174.9526)
d_y2005			-0.0203	1.0296**	293.3327	239.4540
			(0.0556)	(0.4369)	(343.7920)	(208.1228)
d_y2006			-0.0172	1.2162**	368.0552	275.0646
			(0.0649)	(0.5033)	(395.8863)	(241.0492)
d_y2007			-0.0164	1.3988**	466.0036	308.0828
			(0.0750)	(0.5723)	(447.9135)	(275.3444)
d_y2008			-0.0346	1.5682**	558.6072	326.8093
			(0.0847)	(0.6428)	(501.5399)	(309.9391)
d_y2009			-0.0124	1.7273**	573.8484	379.8259
			(0.0951)	(0.7092)	(558.5416)	(343.9877)
d_y2010			-0.0414	1.8996**	658.0239	400.9781
			(0.1041)	(0.7763)	(612.3930)	(377.9615)
lin_calend_year|trend>3	0.0215**	-0.0131				
	(0.0104)	(0.0097)				
lin_calend_year|trend>6	-0.0001	0.0050				
	(0.0068)	(0.0063)				
lin_calend_year|trend>9	-0.0015	-0.0057				
	(0.0065)	(0.0059)				
d_province2	0.0117	0.0014	-0.0566*	-0.0049	74.7160	-28.7352
	(0.0176)	(0.0169)	(0.0309)	(0.1011)	(73.5696)	(28.2519)
d_province3	-0.0345	0.0458	-0.0194	0.0624	13.5300	43.8943
	(0.0274)	(0.0288)	(0.0336)	(0.0597)	(83.0897)	(45.6237)
d_province4	-0.0306*	0.0248	-0.0288	0.0980*	35.6973	57.3384
	(0.0169)	(0.0151)	(0.0218)	(0.0530)	(81.6450)	(41.3521)
d_province5	-0.0196	0.0208	0.0297	-0.0290	-32.0965	7.4747
	(0.0141)	(0.0138)	(0.0231)	(0.0694)	(101.3922)	(45.8290)
lin_calend_year_province2			0.0042	0.0029	-3.2237	1.7653
			(0.0027)	(0.0102)	(8.8182)	(3.8674)
lin_calend_year_province3			-0.0040	-0.0082	-3.8914	-2.3396
			(0.0036)	(0.0064)	(11.0287)	(5.0573)
lin_calend_year_province4			0.0032	-0.0121**	-2.8661	-7.6270*
			(0.0028)	(0.0057)	(8.5770)	(3.8364)
lin_calend_year_province5			-0.0061**	0.0016	1.7957	-1.0461
			(0.0026)	(0.0071)	(11.7184)	(5.4416)
Observations	482		646		646	
Amemiya-Nold test[†]	-0.00243	-0.00211	-0.00035	-0.00231	-5238.837	-1361.398
R-squared	0.9951		0.9999		0.9958	

Significance level: *** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. The table reports the results from estimating (4) in Section 2.4.4 by OLS on mentioned outcomes. This is the second step regression of the two-step approach described in detail in Section 2.4.4, where this second step is estimated by OLS. Since the outcomes satisfy adding-up constraints (salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked), this second step is estimated from a OLS SUR on the first two outcomes in the sum. Effects on the third outcome (the sum) are then obtained from the adding-up constraints. Standard errors are clustered at graduation time-province (gp) level. The dependent variables of the second step are the cluster-time FE $\hat{\mu}_{gpt}$ estimated in each corresponding first step regression, given by (3) in Section 2.4.4. The covariates of the second step are described below (in order): $URate\_grad$ is the unemployment rate at graduation u_{gp}. $URate\_grad * lin\_exp$ is an interaction between u_{gp} and a linear trend in experience. $URate\_grad * lin\_exp|exp > 3$ is u_{gp} interacted with a linear trend in experience that starts from exp>3. The subsequent covariate is defined similarly. Variables $URate\_grad * lin\_exp$-$URate\_grad * lin\_exp|exp > 6$ specify the linear spline in potential experience $f_g(e)$ that multiplies the provincial unemployment rate at graduation u_{gp}. The slope of the spline changes at experience 3 and 6 for high educated, as explained in (2) in Section 5.1 of the main text. Variables $URate\_grad * lin\_exp * upturn$-$URate\_grad * lin\_exp|exp > 6 * upturn$ are triple interactions of u_{gp} with $f_{gu}(e)$ and a dummy=1 if graduation occurs in an upturn. They allows u_{gp} to have a different effect over experience in upturn/downturn. d_exp1-d_exp10 are experience FE. Since they define constant-terms for each experience level, the constant is omitted in the specification. $current\_URate * lin\_exp$-$current\_URate * lin\_exp|exp > 6$ are interactions between the current unemployment rate u_{tp} with $f_t(e)$. Variables lin_grad_year-$lin\_grad\_year|trend > 6$ is the spline for the graduation year $f_0(g)$. d_y2000-d_y2010 are calendar year FE for the period 1998-2010. Identification requires dropping both year 1998 and 1999: the first is the reference FE, while the second has to be dropped because of the following accounting identity: $calend\_year = exp + grad\_year$. In this choice of dropping a second calendar year we followed Oreopoulos et al. (2012). Calendar year is alternatively specified with a spline, by means of variables $lin\_calend\_year|trend > 3$-$lin\_calend\_year|trend > 9$, where the first term of the spline is omitted because of the aforementioned accounting identity. $d\_province2$-$d\_province5$ are province FE (province 1 is the reference). $lin\_calend\_year * d\_province2$-$lin\_calend\_year * d\_province5$ are province-specific linear time trends. Depending on the outcome, we impose restrictions which cannot be jointly rejected at the 5% level, as follows: the effect of u_{gp} is restricted to be symmetric in upturn and downturn (for salaried and self-employment, log hourly wage, part-time hours worked); the effect of u_{tp} is set to be zero (for salaried and self-employment, log hourly wage, log hours worked, part-time hours worked); calendar year FE are specified with a spline (for salaried and self-employment); provincial time trends are set to be zero (for salaried and self-employment).

† The Amemiya-Nold test is an estimate of the variance of the cluster-time errors e_gpt, described in (5) in Section 2.4.4. A negative statistics is evidence that the unobserved cluster-time shocks are indeed zero. Accordingly, one can implement the FGLS procedure as in Wooldridge (2006, 2010).

Table A.18: Second Step FGLS Estimation: High Educated.

Outcomes:	discrete		continuous			
	salaried	self-empl.	log wage	log hours	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
URate_grad	-0.0258	0.0087	-0.0184	-0.0565***	-132.7637***	31.0621**
	(0.0183)	(0.0165)	(0.0114)	(0.0207)	(28.4491)	(13.0104)
URate_grad*lin_exp	0.0096*	-0.0076*	-0.0046*	0.0214***	50.7776***	-7.7613*
	(0.0052)	(0.0045)	(0.0024)	(0.0067)	(8.1339)	(4.3468)
URate_grad*lin_exp|exp>3	-0.0089	0.0104*	0.0042	-0.0248***	-63.7346***	10.0000**
	(0.0061)	(0.0055)	(0.0031)	(0.0071)	(8.7159)	(4.9163)
URate_grad*lin_exp|exp>6	-0.0041	-0.0001	-0.0023	0.0046	20.5970**	-2.1085
	(0.0039)	(0.0038)	(0.0021)	(0.0066)	(8.5333)	(1.2834)
URate_grad*lin_exp*upturn				-0.0090**	-13.0287**	
				(0.0044)	(4.8606)	
URate_grad*lin_exp|exp>3*upturn				0.0186***	21.0758**	
				(0.0067)	(8.0703)	
URate_grad*lin_exp|exp>6*upturn				-0.0136**	-12.9939*	
				(0.0066)	(6.7360)	
d_exp1	0.9697***	0.0308	2.3992***	7.4988***	1,709.5018***	141.1729***
	(0.0568)	(0.0557)	(0.0334)	(0.0572)	(101.9240)	(31.5225)
d_exp2	0.9408***	0.0574	2.4429***	7.5401***	1,905.9920***	64.8359***
	(0.0553)	(0.0543)	(0.0333)	(0.0388)	(74.3620)	(21.5103)
d_exp3	0.9331***	0.0648	2.4848***	7.5086***	1,937.7451***	43.0083**
	(0.0540)	(0.0533)	(0.0334)	(0.0326)	(69.4179)	(20.7862)
d_exp4	0.9258***	0.0763	2.5181***	7.4682***	1,955.6993***	35.3342
	(0.0538)	(0.0532)	(0.0353)	(0.0502)	(95.3896)	(31.1004)
d_exp5	0.9079***	0.0945*	2.5499***	7.4050***	1,938.0647***	24.3331
	(0.0543)	(0.0535)	(0.0384)	(0.0764)	(136.8705)	(46.1787)
d_exp6	0.8982***	0.1083*	2.5708***	7.3546***	1,941.3592***	7.7160
	(0.0562)	(0.0554)	(0.0432)	(0.1034)	(183.4455)	(63.8553)

Continued on next page

Table A.18 – continued from previous page

	salaried	self-empl.	log wage	log hours	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
d_exp7	0.8877***	0.1167**	2.5894***	7.3062***	1,937.0128***	3.6514
	(0.0589)	(0.0579)	(0.0479)	(0.1339)	(230.4968)	(80.2243)
d_exp8	0.8613***	0.1369**	2.6090***	7.2465***	1,907.5481***	0.4558
	(0.0627)	(0.0613)	(0.0539)	(0.1630)	(281.2384)	(98.1468)
d_exp9	0.8377***	0.1579**	2.6192***	7.1994***	1,891.2816***	2.2098
	(0.0671)	(0.0653)	(0.0598)	(0.1911)	(330.2644)	(115.9887)
d_exp10	0.8150***	0.1707**	2.6230***	7.1421***	1,852.2196***	6.1605
	(0.0729)	(0.0709)	(0.0673)	(0.2196)	(381.8000)	(133.9872)
current_URate*lin_exp					-7.9607***	
					(2.4082)	
current_URate*lin_exp|exp>3					19.0362***	
					(4.8666)	
current_URate*lin_exp|exp>6					-13.7122***	
					(4.8368)	
lin_grad_year	-0.0169	0.0116	0.0353***	-0.0628**	-22.8865	-2.8056
	(0.0194)	(0.0191)	(0.0103)	(0.0306)	(53.1461)	(19.2831)
lin_grad_year|trend>3	-0.0042	0.0060	-0.0109	0.0256**	-6.0510	14.6038
	(0.0221)	(0.0221)	(0.0133)	(0.0122)	(25.7994)	(11.0858)
lin_grad_year|trend>6	-0.0362	0.0392	0.0104	0.0366**	80.7248**	-51.1515***
	(0.0284)	(0.0284)	(0.0109)	(0.0143)	(35.2520)	(15.2233)
d_y2000			0.0165*	0.0610	-26.6306	-9.5551
			(0.0093)	(0.0403)	(63.6992)	(26.7663)
d_y2001			0.0347*	0.1076	-33.6986	1.0926
			(0.0180)	(0.0690)	(118.5847)	(43.9962)
d_y2002			0.0544**	0.1799*	-6.4181	12.6021
			(0.0245)	(0.0998)	(168.1877)	(60.9035)
d_y2003			0.1253***	0.1695	-91.7889	33.3459
			(0.0322)	(0.1310)	(220.8156)	(79.5331)
d_y2004			0.1289***	0.2448	-70.6336	40.0388
			(0.0382)	(0.1595)	(271.2868)	(96.7843)
d_y2005			0.1344***	0.3112	-78.1185	43.9727
			(0.0458)	(0.1897)	(320.7529)	(115.4804)
d_y2006			0.1587***	0.3620	-60.0532	47.7650
			(0.0534)	(0.2216)	(372.7299)	(133.7550)
d_y2007			0.1873***	0.4257	-13.9987	46.8146
			(0.0607)	(0.2525)	(425.3406)	(152.0748)
d_y2008			0.1917***	0.4870*	14.9304	49.7238
			(0.0688)	(0.2821)	(476.3703)	(170.3990)
d_y2009			0.2328***	0.5344*	12.3347	61.1454
			(0.0772)	(0.3120)	(529.1195)	(188.3118)
d_y2010			0.2285**	0.5976*	40.1475	69.7336
			(0.0843)	(0.3412)	(581.2636)	(207.3524)
lin_calend_year|trend>3	0.0096	-0.0079				
	(0.0093)	(0.0081)				
lin_calend_year|trend>6	-0.0006	0.0044				
	(0.0050)	(0.0044)				
lin_calend_year|trend>9	0.0032	-0.0068*				
	(0.0041)	(0.0040)				
d_province2	0.0302	-0.0306	-0.0594*	0.0563	116.7313**	-36.1011**
	(0.0260)	(0.0260)	(0.0345)	(0.0390)	(51.7205)	(15.2201)
d_province3	-0.0221	0.0208	-0.0293	0.0256	-32.9023	-10.5819
	(0.0280)	(0.0280)	(0.0217)	(0.0351)	(60.0183)	(16.2143)
d_province4	0.0016	-0.0023	-0.0053	0.0567	157.2709***	-39.6587**

Continued on next page

Table A.18 – continued from previous page

	salaried	self-empl.	log wage	log hours	FT hours	PT hours
	(1)	(2)	(3)	(4)	(5)	(6)
	(0.0206)	(0.0205)	(0.0208)	(0.0369)	(56.3724)	(15.3425)
d_province5	-0.0184	0.0186	0.0081	0.0642*	83.7151	-14.2460
	(0.0279)	(0.0278)	(0.0237)	(0.0343)	(56.5859)	(17.6747)
lin_calend_year_province2			0.0009	-0.0055	-11.1725*	4.3444**
			(0.0035)	(0.0037)	(5.5786)	(2.0338)
lin_calend_year_province3			-0.0072***	-0.0024	7.9390	-1.8570
			(0.0025)	(0.0033)	(5.9834)	(1.8436)
lin_calend_year_province4			0.0013	-0.0045	-10.7012*	0.8637
			(0.0030)	(0.0037)	(5.6221)	(1.6140)
lin_calend_year_province5			-0.0034	-0.0042	0.5409	-3.3208*
			(0.0028)	(0.0035)	(5.3448)	(1.7501)
R-squared	1.0000		0.9999		0.9988	
WSSR (2nd step)	271.55829		1084.3865		1058.984	
Obs (2nd step)	310		646		646	
Parameters (2nd step)	48		75		78	
Test joint signif. all imposed restr.(P-val)	0.494		0.309		0.390	
P-value of chi2 test	0.3293804		2.89E-34		4.48E-32	
cluster (at which level)	no	no	g*p	g*p	g*p	g*p
Imposed Restrictions:						
effect URate at grad. symmetric up/downturn	yes	yes	yes	no	no	yes
effect Current Urate over exp=0	yes	yes	yes	yes	no	yes
spline for calendar year FE	yes	yes	no	no	no	no
effect prov-time trends=0	yes	yes	no	no	no	no

Significance level: *** p<0.01, ** p<0.05, * p<0.1. Standard errors between parentheses. The table reports the results from estimating (4) in Section 2.4.4 by FGLS on mentioned outcomes. This is the second step regression of the two-step FGLS approach described in detail in Section 2.4.4. Since the outcomes satisfy adding-up constraints (salaried employment+self-employment=overall employment; log hourly wage+log hours worked=log earnings; FT hours worked+PT hours worked=total hours worked), this second step is estimated from a FGLS SUR on the first two outcomes in the sum. Effects on the third outcome (the sum) are then obtained from the adding-up constraints. The dependent variables of the second step are the cluster-time FE $\hat{\mu}_{gpt}$ estimated in each corresponding first step regression, given by (3) in Section 5.2 of the main text. A description of the covariates reported in the Table is in the footnote of Table A.17 above. In this FGLS SUR, the data are weighted by the inverse of the cluster-robust variance matrix of $\hat{\mu}_{gpt}$ estimated in the first step. For discrete outcomes, the Moore-Penrose generalized inverse of this matrix is used as weight, to take into account the perfect serial correlation induced by the fact that, for specific clusters, the outcomes do not vary over time (see Section 5.2 of the main text for details). Depending on the outcome, we impose restrictions which cannot be jointly rejected at the 5% level: these restrictions are listed in the bottom panel of the table. If the χ^2 goodness-of-fit statistic rejects the model (P-value>0.05), standard errors clustered at the *gp* level are reported; otherwise conventional ones. For completeness we report also the weighted sum of squared residuals (WSSR), the number of estimated parameters and the number of observations, which are used to compute the χ^2 test.

A.8 The Effects of the Control Variables in the First Step

This section discusses briefly the estimated effects of the control variables in the first step estimation, which are reported in Table A.13 and A.16 for low and high educated, respectively.

- *Household composition*: For both low and high educated, a higher number of young household members (aged less than 11 when the individual was aged 17) is associated with lower hours worked in full-time employment, on average. These individuals may allocate more time in household activities or leisure because of particular preferences or time constraints due to the presence of younger members in the household.

- *Parental education*: (i) For low educated, mother and father education are both negatively associated with hours worked (and full-time hours worked) and positively related with part-time hours worked. For high educated the same relationship holds for mother education. This may capture the effect of unobserved liquidity constraints on time worked, so that less constrained individuals (associated to higher parental education) work more (less) hours in part-time (full-time) jobs, and overall less hours in salaried employment. In the same spirit, low educated individuals with low educated mothers are also more likely to opt for salaried employment (with expected stable income under long-term contracts), while they are more likely to engage in (riskier) self-employment if their mothers are high educated. (ii) In addition, for high educated, father education is positively associated with hourly wages. This is a standard result: father education may indicate family connections; or, it can be a proxy for (not observed) father earnings or family wealth, thereby reflecting better initial conditions or particular values shared in the family (i.e. ambitions). For low educated, this relation is not significant. This suggests that family background is more important predictor of individual wages for high than for low educated.[13]

- *Delay in secondary education at age 17*: For low educated, the number grades repeated at age 17 has a negative effect on self-employment, wages, hours worked and full-time hours worked, while a positive effect on part-time hours worked. For high educated, it has a negative impact on wages, hours worked and full-time hours worked. The signs of the effects are expected: students that already repeated a grade at age 17 are likely to be negatively selected from the population of students and therefore have, on average, worse labor market performances.

- *Educational track compared to general education*: For low educated, choosing a technical track at age 17 is associated with higher wages and more hours worked, especially in full-time jobs, compared to general education. Therefore, for this group technical education is the most profitable track. For high educated instead general education is associated with better labor market performances than the other options. This is also a standard results, since general education is the typical choice for those who want to access tertiary education.

- *Birth cohort 1978 and 1980 compared to 1976*: For both high and low educated, cohort 1978 and 1980 are associated with lower wages than cohort 1976. This is explained by the sample selection, in particular by the fact that labor market outcomes are observed from 1998 onwards, while for cohort 1976 and 1978 graduation can occur earlier, i.e. as from 1994 and 1996 respectively (at age 18, i.e. the end of compulsory education). Consequently, cohort 1976 (1978) is under-represented at potential experience 1-4 (1-2) compared to cohort 1980, since these years of experience are not observed for those born in 1976 (1978) who gradu-

[13] Parents with higher education levels have children with higher education level (Black et al., 2005). As a matter of fact, the correlation between completed education and father education is larger for high (0.2054) than for low educated (0.0768). The same correlation for mother education is similar - still positive, but slightly smaller: 0.1808 and 0.0665 for high and low educated, respectively.

ated before age 23 (21).[14] The tabulations below show that labor market outcomes of cohort 1976 and 1978 are under (over)-represented for first (last) years of experience. Given that wages increase with tenure and seniority, this may explain the negative correlation between wage and birth cohort.

In what calendar year is experience 1 observed?										
	all				low educated			high educated		
year	c1976	c1978	c1980	Total	c1976	c1978	c1980	c1976	c1978	c1980
1998	154	240	0	394	71	216	0	83	24	0
1999	185	203	187	575	22	177	187	163	26	0
2000	191	137	242	570	13	61	225	178	76	17
2001	115	163	163	441	2	20	142	113	143	21
2002	0	153	154	307	0	5	75	0	148	79
2003	0	138	142	280				0	138	142
2004	0	0	172	172				0	0	172
2005	0	0	106	106				0	0	106
Total	645	1,034	1,166	2,845	108	479	629	537	555	537

In what calendar year is experience 10 observed?										
	all				low educated			high educated		
year	c1976	c1978	c1980	Total	c1976	c1978	c1980	c1976	c1978	c1980
2004	143	0	0	143	143	0	0	83	24	0
2005	209	0	0	209	209	0	0	159	26	0
2006	168	148	0	316	168	148	0	173	76	17
2007	154	238	0	392	71	214	0	111	142	20
2008	181	198	186	565	22	172	186			
2009	185	137	241	563	12	61	224			
2010	113	161	159	433	2	19	139			
Total	1,153	882	586	2,621	627	614	549	526	268	37

In addition, for low educated cohort 1978 and 1980 are associated with less hours worked (and less hours worked in full-time jobs) than cohort 1976; cohort 1980 is also associated with more hours worked in part-time jobs than cohort 1976. For high educated instead, cohort 1980 has lower (higher) probability to be self- (salaried) employed than cohort 1976. We do not have an explanation for these correlations.

[14]For someone born in 1976 (1978) who graduated at age 23 (21), graduation occurs in 1997 and the first year of potential experience is observed, as it occurs in 1998.

A.9 Tables of the Sensitivity Analysis

Table A.21: Individual Discrete Labor Market Outcomes: Low Educated.

		sensitivity tests:				
	baseline	education[†]	national[§]	probit[‡]	2nd-step OLS	one-step
Imposed Restrictions:[§§]						
Effect URate at grad. symmetric up/downturn	yes	yes	-	yes	yes	yes
Effect Current Urate over exp=0	yes	yes	yes	yes	yes	yes
Spline for calendar year FE	yes	yes	-	yes	yes	yes
Effect prov-time trends=0	yes	yes	yes	yes	yes	yes
Level of clustering[††]	no	no	no	no	$g*p$	$g*p$
Salaried employment						
potential exp	(1)	(2)	(3)	(4)	(5)	(6)
1	-0.017	-0.016	0.004	0.000	0.001	-0.006
	(0.018)	(0.018)	(0.020)	(0.023)	(0.020)	(0.021)
2	-0.011	-0.008	-0.001	0.009	0.014	0.005
	(0.015)	(0.015)	(0.020)	(0.021)	(0.017)	(0.017)
3	-0.005	0.000	-0.007	0.017	0.027	0.017
	(0.015)	(0.015)	(0.020)	(0.018)	(0.015)	(0.015)
4	-0.002	0.003	-0.011	0.018	0.024	0.014
	(0.015)	(0.014)	(0.019)	(0.017)	(0.014)	(0.014)
5	0.000	0.006	-0.014	0.019	0.021	0.011
	(0.014)	(0.014)	(0.020)	(0.017)	(0.014)	(0.014)
6	0.002	0.009	-0.018	0.018	0.018	0.009
	(0.015)	(0.015)	(0.021)	(0.017)	(0.015)	(0.015)
7	0.002	0.007	-0.010	0.018	0.015	0.007
	(0.014)	(0.014)	(0.020)	(0.016)	(0.013)	(0.013)
8	0.002	0.005	-0.003	0.016	0.013	0.005
	(0.014)	(0.014)	(0.021)	(0.016)	(0.012)	(0.013)
9	0.002	0.003	0.005	0.013	0.010	0.003
	(0.015)	(0.015)	(0.022)	(0.014)	(0.012)	(0.013)
10	0.003	0.005	0.013	0.014	0.014	0.007
	(0.015)	(0.014)	(0.022)	(0.013)	(0.010)	(0.011)
11	0.004	0.007	0.020	0.019	0.018	0.011
	(0.015)	(0.015)	(0.024)	(0.016)	(0.009)	(0.010)
12	0.006	0.009	0.027	0.019	0.021	0.015
	(0.015)	(0.015)	(0.028)	(0.015)	(0.010)	(0.011)
Self-employment						
1	0.036	0.048	0.015	0.021	0.017	0.025
	(0.018)	(0.018)	(0.016)	(0.031)	(0.019)	(0.019)
2	0.021	0.024	0.022	0.006	0.001	0.008
	(0.015)	(0.014)	(0.016)	(0.026)	(0.016)	(0.016)
3	0.007	0.001	0.028	-0.008	-0.015	-0.010
	(0.015)	(0.015)	(0.017)	(0.022)	(0.015)	(0.014)
4	0.001	-0.004	0.024	-0.016	-0.018	-0.011
	(0.014)	(0.014)	(0.017)	(0.021)	(0.014)	(0.014)
5	-0.005	-0.009	0.020	-0.024	-0.021	-0.013
	(0.014)	(0.014)	(0.017)	(0.022)	(0.015)	(0.014)
6	-0.010	-0.013	0.015	-0.029	-0.024	-0.015
	(0.014)	(0.014)	(0.018)	(0.021)	(0.016)	(0.016)
7	-0.009	-0.011	0.012	-0.026	-0.021	-0.014
	(0.014)	(0.013)	(0.018)	(0.020)	(0.014)	(0.014)
8	-0.007	-0.009	0.008	-0.021	-0.019	-0.013

Continued on next page

Table A.21 – continued from previous page

Potential experience	baseline (1)	sensitivity tests: education† (2)	national§ (3)	probit‡ (4)	2nd-step OLS (5)	one-step (6)
	(0.014)	(0.013)	(0.018)	(0.020)	(0.012)	(0.013)
9	-0.005	-0.006	0.004	-0.014	-0.017	-0.012
	(0.014)	(0.014)	(0.019)	(0.017)	(0.011)	(0.012)
10	-0.009	-0.011	-0.002	-0.017	-0.020	-0.014
	(0.014)	(0.014)	(0.020)	(0.017)	(0.009)	(0.011)
11	-0.013	-0.015	-0.008	-0.024	-0.023	-0.017
	(0.014)	(0.014)	(0.022)	(0.021)	(0.009)	(0.010)
12	-0.017	-0.019	-0.014	-0.026	-0.026	-0.020
	(0.014)	(0.014)	(0.024)	(0.019)	(0.010)	(0.011)
Overall-employment						
1	0.018	0.031	0.019	0.021	0.018	0.018
	(0.015)	(0.015)	(0.014)	(0.023)	(0.008)	(0.006)
2	0.010	0.017	0.020	0.016	0.015	0.013
	(0.009)	(0.009)	(0.012)	(0.015)	(0.006)	(0.005)
3	0.002	0.002	0.021	0.009	0.012	0.007
	(0.009)	(0.009)	(0.012)	(0.012)	(0.007)	(0.007)
4	-0.001	0.000	0.013	0.002	0.006	0.003
	(0.007)	(0.007)	(0.011)	(0.010)	(0.006)	(0.006)
5	-0.005	-0.003	0.005	-0.005	0.000	-0.002
	(0.006)	(0.006)	(0.010)	(0.010)	(0.006)	(0.005)
6	-0.008	-0.005	-0.003	-0.011	-0.006	-0.006
	(0.008)	(0.007)	(0.011)	(0.011)	(0.007)	(0.006)
7	-0.007	-0.004	0.001	-0.008	-0.006	-0.007
	(0.006)	(0.006)	(0.010)	(0.009)	(0.006)	(0.006)
8	-0.005	-0.004	0.005	-0.004	-0.006	-0.008
	(0.005)	(0.005)	(0.010)	(0.008)	(0.006)	(0.006)
9	-0.004	-0.003	0.009	-0.001	-0.007	-0.009
	(0.006)	(0.006)	(0.011)	(0.008)	(0.007)	(0.007)
10	-0.006	-0.006	0.011	-0.003	-0.006	-0.007
	(0.005)	(0.005)	(0.011)	(0.007)	(0.005)	(0.005)
11	-0.009	-0.008	0.012	-0.005	-0.005	-0.006
	(0.006)	(0.005)	(0.013)	(0.009)	(0.004)	(0.004)
12	-0.011	-0.010	0.013	-0.007	-0.005	-0.005
	(0.007)	(0.007)	(0.016)	(0.009)	(0.004)	(0.004)

Standard errors between parentheses. Column 1 reports the point estimates of the benchmark model reported in Section 2.5. Columns 2-6 report the following sensitivity tests (in order): including completed education FE as individual control variables in the first step of the two-step approach; national model; estimating the first step by Probit rather than by a Linear Probability Model; estimating the second step by OLS rather than FGLS; estimating the model in a one-step rather than two-step approach. Results in Column 1-4 are obtained from the second step FGLS of the two-step approach procedure explained in the text; Column 5 is obtained from the second step OLS, i.e. neglecting the measurement error introduced by the two-step procedure. Column 6 is estimated in one step. For all columns, salaried and self-employment are estimated jointly by means of a SUR, and the effects of interest on the overall employment rate are obtained as a linear combination of the effects on the first two outcomes, exploiting the adding-up constraint: overall employment=salaried employment+self-employment.

§ In the national model cells need not be dropped, since cells are always sufficiently large (see selection rule in Table 1.2 in Appendix 1.8). In contrast, all other sensitivity analyses (Column 2,4-6), which are all based on the provincial model, rely on the same sub-sample retained for the benchmark and consider the same specification used in the benchmark; in the one-step approach (column 6), this specification is augmented by the individual control variables used in the first step.

† Completed education is measured as the number of years of education successfully attained from the beginning of secondary education, i.e. from age 12. Therefore, repeated grades are counted only as soon as that grade is successfully passed.

‡ The table reports the partial effects on the probability of employment for each year of potential experience, where the other aggregate regressors are evaluated at their sample mean.

§§ The restrictions listed at the top of the table are imposed on both salaried and self-employment.
†† Clustered (conventional) standard errors are reported if the model is (not) rejected on the basis of the χ^2 goodness-of-fit test at the 5% level. In contrast to all provincial models, in the national model the variance of unobserved cluster-time shocks, as calculated by Equation (5) in the main text is strictly positive for salaried employment (see Section 2.4.4 for details). We therefore add the estimated variance of the cluster-time shocks to the diagonal of the variance matrix of the measurement error, and use the inverse of the resulting matrix as weight in the FGLS transformation.

Table A.22: Individual Continuous Labor Market Outcomes: Low Educated.

		Sensitivity tests			
	baseline	education†	national§	2nd-step OLS	one-step
	log hourly wage				
Imposed Restrictions:					
Effect URate at grad. symmetric up/downturn	yes	yes	yes	yes	yes
Effect Current Urate over exp=0	no	no	-	no	no
Spline for calendar year FE	no	no	no	no	no
Effect prov-time trends=0	no	no	-	no	no
Level of clustering††	$g*p$	$g*p$	g	$g*p$	$g*p$
Potential experience	(1)	(2)	(3)	(4)	(5)
1	0.005	0.004	-0.008	-0.008	-0.004
	(0.013)	(0.013)	(0.017)	(0.014)	(0.010)
2	-0.001	-0.001	-0.005	-0.008	-0.007
	(0.010)	(0.011)	(0.015)	(0.011)	(0.008)
3	-0.006	-0.006	-0.002	-0.008	-0.010
	(0.009)	(0.010)	(0.015)	(0.010)	(0.008)
4	-0.008	-0.007	0.000	-0.009	-0.012
	(0.008)	(0.009)	(0.015)	(0.009)	(0.007)
5	-0.009	-0.008	0.003	-0.011	-0.014
	(0.008)	(0.008)	(0.015)	(0.008)	(0.006)
6	-0.011	-0.010	0.005	-0.013	-0.016
	(0.007)	(0.008)	(0.016)	(0.007)	(0.006)
7	-0.008	-0.008	0.004	-0.014	-0.015
	(0.007)	(0.008)	(0.016)	(0.007)	(0.005)
8	-0.006	-0.005	0.003	-0.015	-0.014
	(0.007)	(0.008)	(0.017)	(0.006)	(0.006)
9	-0.003	-0.003	0.002	-0.017	-0.014
	(0.008)	(0.009)	(0.017)	(0.007)	(0.006)
10	-0.006	-0.006	0.003	-0.017	-0.016
	(0.008)	(0.009)	(0.018)	(0.006)	(0.005)
11	-0.009	-0.009	0.004	-0.017	-0.018
	(0.009)	(0.009)	(0.019)	(0.007)	(0.005)
12	-0.012	-0.011	0.006	-0.018	-0.021
	(0.010)	(0.010)	(0.022)	(0.008)	(0.006)
	log hours worked				
Imposed Restrictions:§§					
Effect URate at grad. symmetric up/downturn	no	no	yes	no	no
Effect Current Urate over exp=0	no	no	-	no	no
Spline for calendar year FE	no	no	no	no	no
Effect prov-time trends=0	no	no	-	no	no
Level of clustering††	$g*p$	$g*p$	g	$g*p$	$g*p$
1	-0.044	-0.044	-0.067	-0.021	-0.024
	(0.022)	(0.022)	(0.049)	(0.029)	(0.025)
2	-0.035	-0.036	-0.045	-0.031	-0.033
	(0.020)	(0.019)	(0.043)	(0.024)	(0.022)

Continued on next page

Table A.22 – continued from previous page

Potential experience	baseline (1)	sensitivity tests: education† (2)	national§ (3)	2nd-step OLS (4)	one-step (5)
3	-0.026	-0.028	-0.023	-0.040	-0.041
	(0.022)	(0.021)	(0.044)	(0.032)	(0.029)
4	-0.026	-0.027	-0.021	-0.040	-0.038
	(0.019)	(0.018)	(0.040)	(0.029)	(0.026)
5	-0.026	-0.026	-0.019	-0.039	-0.035
	(0.016)	(0.015)	(0.037)	(0.026)	(0.023)
6	-0.025	-0.025	-0.017	-0.039	-0.031
	(0.014)	(0.013)	(0.035)	(0.024)	(0.021)
7	-0.025	-0.024	-0.014	-0.039	-0.028
	(0.012)	(0.012)	(0.034)	(0.022)	(0.019)
8	-0.025	-0.023	-0.012	-0.039	-0.025
	(0.012)	(0.012)	(0.035)	(0.022)	(0.019)
9	-0.025	-0.021	-0.010	-0.038	-0.022
	(0.013)	(0.013)	(0.036)	(0.022)	(0.020)
10	-0.027	-0.023	-0.007	-0.037	-0.023
	(0.010)	(0.011)	(0.039)	(0.021)	(0.020)
11	-0.030	-0.026	-0.004	-0.036	-0.023
	(0.010)	(0.010)	(0.049)	(0.022)	(0.021)
12	-0.033	-0.028	-0.001	-0.035	-0.024
	(0.012)	(0.012)	(0.063)	(0.024)	(0.023)
log earnings					
1	-0.039	-0.040	-0.075	-0.029	-0.028
	(0.026)	(0.026)	(0.052)	(0.034)	(0.029)
2	-0.036	-0.037	-0.050	-0.038	-0.040
	(0.023)	(0.021)	(0.046)	(0.028)	(0.025)
3	-0.033	-0.034	-0.025	-0.048	-0.051
	(0.022)	(0.021)	(0.048)	(0.032)	(0.030)
4	-0.034	-0.034	-0.021	-0.049	-0.050
	(0.019)	(0.019)	(0.045)	(0.029)	(0.027)
5	-0.035	-0.034	-0.016	-0.050	-0.048
	(0.017)	(0.017)	(0.043)	(0.027)	(0.024)
6	-0.036	-0.035	-0.011	-0.052	-0.047
	(0.015)	(0.015)	(0.041)	(0.025)	(0.022)
7	-0.033	-0.031	-0.010	-0.053	-0.043
	(0.014)	(0.015)	(0.040)	(0.023)	(0.021)
8	-0.031	-0.028	-0.009	-0.054	-0.039
	(0.014)	(0.016)	(0.040)	(0.023)	(0.021)
9	-0.028	-0.024	-0.008	-0.055	-0.035
	(0.016)	(0.017)	(0.042)	(0.024)	(0.022)
10	-0.033	-0.029	-0.004	-0.054	-0.039
	(0.014)	(0.016)	(0.044)	(0.022)	(0.021)
11	-0.039	-0.034	0.000	-0.054	-0.042
	(0.014)	(0.016)	(0.052)	(0.023)	(0.022)
12	-0.045	-0.039	0.004	-0.053	-0.045
	(0.017)	(0.018)	(0.065)	(0.024)	(0.024)

Standard errors between parentheses. Column 1 reports the point estimates of the benchmark model reported in Section 2.5. Columns 2-5 report the following sensitivity tests (in order): including completed education FE as individual control variables in the first step of the two-step approach; national model; estimating the second step by OLS rather than FGLS; estimating the model in a one-step rather than two-step approach. Results in Column 1-3 are obtained from the second step FGLS of the two-step approach procedure explained in the text; Column 4 is obtained from the second step OLS, i.e. neglecting the measurement error introduced by the two-step procedure. Column 5 is estimated in one step. For all columns, log hourly wage and log hours worked are estimated jointly

by means of a SUR, and the effects of interest on the log earnings are obtained as a linear combination of the effects on the first two outcomes, exploiting the adding-up constraint: log earnings=log hourly wage+log hours worked.
§ In the national model cells need not be dropped, since cells are always sufficiently large (see selection rule in Table 1.2 in Appendix 1.8). In contrast, all other sensitivity analyses (Column 2,4-5), which are all based on the provincial model, rely on the same sub-sample retained for the benchmark and consider the same specification used in the benchmark; in the one-step approach (column 5), this specification is augmented by the individual control variables used in the first step.
† Completed education is measured as the number of years of education successfully attained from the beginning of secondary education, i.e. from age 12. Therefore, repeated grades are counted only as soon as that grade is successfully passed.
§§ For log hours worked the following additional restriction (not mentioned in the table) is also imposed: $\beta_{g2} = 0$, i.e. the slope of the linear spline remains fixed after 6 years of experience. This restriction cannot be rejected.
†† Clustered (conventional) standard errors are reported if the model is (not) rejected on the basis of the χ^2 test at the 5% level.

Table A.23: Individual Discrete Labor Market Outcomes: High Educated.

		sensitivity tests:				
	baseline	education†	national§	probit‡	2nd-step OLS	one-step
Imposed Restrictions:§§						
Effect URate at grad. symmetric up/downturn	yes	yes	yes	yes	yes	yes
Effect Current Urate over exp=0	yes	yes	-	yes	yes	yes
Spline for calendar year FE	yes	yes	yes	yes	yes	yes
Effect prov-time trends=0	yes	yes	-	yes	yes	yes
Level of clustering††	no	no	no	no	$g*p$	$g*p$
Salaried employment						
Potential experience	(1)	(2)	(3)	(4)	(5)	(6)
1	-0.016	-0.021	-0.012	-0.011	-0.026	-0.021
	(0.015)	(0.015)	(0.013)	(0.023)	(0.014)	(0.013)
2	-0.006	-0.012	-0.010	0.001	-0.009	-0.007
	(0.013	(0.013	(0.011	(0.022	(0.011	(0.011)
3	0.003	-0.004	-0.009	0.012	0.008	0.008
	(0.014)	(0.013)	(0.013)	(0.019)	(0.010)	(0.011)
4	0.004	-0.003	-0.003	0.008	0.004	0.004
	(0.013)	(0.013)	(0.012)	(0.017)	(0.009)	(0.011)
5	0.005	-0.002	0.003	0.005	0.000	-0.001
	(0.013)	(0.013)	(0.012)	(0.019)	(0.009)	(0.012)
6	0.005	-0.002	0.009	0.002	-0.004	-0.005
	(0.014)	(0.013)	(0.013)	(0.022)	(0.011)	(0.014)
7	0.002	-0.004	0.000	0.000	-0.002	-0.004
	(0.013)	(0.013)	(0.012)	(0.016)	(0.009)	(0.013)
8	-0.001	-0.005	-0.009	-0.001	0.000	-0.003
	(0.014)	(0.013)	(0.012)	(0.017)	(0.010)	(0.013)
9	-0.005	-0.007	-0.018	-0.002	0.002	-0.003
	(0.014)	(0.013)	(0.014)	(0.016)	(0.012)	(0.015)
10	-0.008	-0.009	-0.027	-0.003	0.004	-0.002
	(0.015)	(0.014)	(0.016)	(0.015)	(0.014)	(0.017)
self-employment						
1	0.001	0.007	0.007	-0.013	0.021	0.017
	(0.014)	(0.014)	(0.009)	(0.026)	(0.015)	(0.015)
2	-0.007	-0.001	0.005	-0.020	0.006	0.007
	(0.013)	(0.013)	(0.009)	(0.023)	(0.011)	(0.012)
3	-0.014	-0.008	0.003	-0.025	-0.008	-0.003
	(0.014)	(0.013)	(0.010)	(0.021)	(0.011)	(0.011)
4	-0.011	-0.005	0.002	-0.018	-0.003	0.002
	(0.013)	(0.013)	(0.010)	(0.019)	(0.010)	(0.011)
5	-0.009	-0.002	0.002	-0.014	0.003	0.006

Continued on next page

Table A.23 – continued from previous page

		sensitivity tests:				
	baseline	education†	national§	probit‡	2nd-step OLS	one-step
Potential experience	(1)	(2)	(3)	(4)	(5)	(6)
	(0.013)	(0.013)	(0.010)	(0.021)	(0.010)	(0.012)
6	-0.006	0.001	0.001	-0.010	0.008	0.011
	(0.014)	(0.013)	(0.011)	(0.023)	(0.010)	(0.013)
7	-0.003	0.002	0.008	-0.006	0.005	0.010
	(0.013)	(0.013)	(0.011)	(0.018)	(0.009)	(0.012)
8	0.000	0.003	0.016	-0.004	0.002	0.008
	(0.014)	(0.013)	(0.011)	(0.019)	(0.009)	(0.012)
9	0.002	0.004	0.023	-0.002	-0.001	0.006
	(0.014)	(0.013)	(0.012)	(0.018)	(0.011)	(0.013)
10	0.005	0.005	0.031	0.001	-0.004	0.004
	(0.015)	(0.015)	(0.013)	(0.017)	(0.013)	(0.015)
	overall employment					
1	-0.015	-0.014	-0.005	-0.025	-0.005	-0.004
	(0.006)	(0.006)	(0.009)	(0.011)	(0.008)	(0.008)
2	-0.013	-0.013	-0.005	-0.020	-0.002	0.000
	(0.004)	(0.004)	(0.007)	(0.006)	(0.006)	(0.005)
3	-0.011	-0.012	-0.005	-0.013	0.000	0.004
	(0.004)	(0.004)	(0.008)	(0.006)	(0.006)	(0.006)
4	-0.007	-0.008	0.000	-0.010	0.002	0.005
	(0.003)	(0.003)	(0.006)	(0.004)	(0.005)	(0.005)
5	-0.004	-0.004	0.005	-0.009	0.003	0.006
	(0.002)	(0.002)	(0.006)	(0.003)	(0.004)	(0.004)
6	0.000	-0.001	0.010	-0.008	0.004	0.007
	(0.003)	(0.003)	(0.008)	(0.005)	(0.004)	(0.004)
7	-0.001	-0.002	0.008	-0.006	0.003	0.006
	(0.002)	(0.002)	(0.006)	(0.003)	(0.004)	(0.004)
8	-0.002	-0.002	0.007	-0.005	0.002	0.004
	(0.002)	(0.002)	(0.006)	(0.003)	(0.004)	(0.004)
9	-0.002	-0.003	0.006	-0.004	0.001	0.003
	(0.003)	(0.003)	(0.007)	(0.004)	(0.004)	(0.004)
10	-0.003	-0.004	0.004	-0.003	0.000	0.002
	(0.005)	(0.005)	(0.010)	(0.005)	(0.004)	(0.004)

Standard errors between parentheses. Column 1 reports the point estimates of the benchmark model reported in Section 2.14. Columns 2-6 report the following sensitivity tests (in order): including completed education FE as individual control variables in the first step of the two-step approach; national model; estimating the first step by Probit rather than by a Linear Probability Model; estimating the second step by OLS rather than FGLS; estimating the model in a one-step rather than two-step approach. Results in Column 1-4 are obtained from the second step FGLS of the two-step approach procedure explained in the text; Column 5 is obtained from the second step OLS, i.e. neglecting the measurement error introduced by the two-step procedure. Column 6 is estimated in one step. For all columns, salaried and self-employment are estimated jointly by means of a SUR, and the effects of interest on the overall employment rate are obtained as a linear combination of the effects on the first two outcomes, exploiting the adding-up constraint: overall employment=salaried employment+self-employment. § In the national model cells need not be dropped, since cells are always sufficiently large (see selection rule in Table 1.2 in Appendix 1.8). In contrast, all other sensitivity analyses (Column 2,4-6), which are all based on the provincial model, rely on the same sub-sample retained for the benchmark and consider the same specification used in the benchmark; in the one-step approach (column 6), this specification is augmented by the individual control variables used in the first step. † Completed education is measured as the number of years of education successfully attained from the beginning of secondary education, i.e. from age 12. Therefore, repeated grades are counted only as soon as that grade is successfully passed. ‡ The table reports the partial effects on the probability of employment for each year of potential experience, where the other aggregate regressors are evaluated at their sample mean. §§ The restrictions listed at the top of the table are imposed on both salaried and self-employment. †† Clustered (conventional) standard errors are reported if the model is (not) rejected on the basis of the χ^2 goodness-of-fit test at the 5% level. In contrast to all provincial models, in the national model the variance of unobserved cluster-time shocks, as calculated by Equation (5) in the main text is strictly positive for salaried employment (see Section 2.4.4 for details). We therefore add the estimated variance of the cluster-time shocks to the diagonal of the variance matrix of the measurement error, and use the inverse of the resulting matrix as weight in the FGLS transformation.

Table A.24: Individual Continuous Labor Market Outcomes: High educated.

		Sensitivity tests			
	baseline	education†	national§	2nd-step OLS	one-step
	log hourly wage				
Imposed Restrictions:					
Effect URate at grad. symmetric up/downturn	yes	yes	yes	yes	yes
Effect Current Urate over exp=0	yes	yes	-	yes	yes
Spline for calendar year FE	no	no	no	no	no
Effect prov-time trends=0	no	no	-	no	no
Level of clustering††	$g*p$	$g*p$	g	$g*p$	$g*p$
Potential experience	(1)	(2)	(3)	(4)	(5)
1	-0.023	-0.019	-0.013	-0.011	-0.009
	(0.010)	(0.009)	(0.011)	(0.009)	(0.008)
2	-0.028	-0.023	-0.022	-0.014	-0.013
	(0.009)	(0.009)	(0.010)	(0.008)	(0.007)
3	-0.032	-0.027	-0.031	-0.017	-0.017
	(0.009)	(0.009)	(0.010)	(0.008)	(0.007)
4	-0.033	-0.027	-0.023	-0.016	-0.016
	(0.008)	(0.009)	(0.014)	(0.007)	(0.007)
5	-0.033	-0.026	-0.015	-0.014	-0.014
	(0.008)	(0.009)	(0.018)	(0.008)	(0.007)
6	-0.033	-0.026	-0.007	-0.013	-0.012
	(0.009)	(0.009)	(0.022)	(0.008)	(0.008)
7	-0.036	-0.028	-0.008	-0.010	-0.010
	(0.009)	(0.010)	(0.020)	(0.009)	(0.009)
8	-0.039	-0.029	-0.010	-0.007	-0.007
	(0.010)	(0.010)	(0.018)	(0.010)	(0.010)
9	-0.042	-0.031	-0.012	-0.004	-0.004
	(0.010)	(0.011)	(0.017)	(0.012)	(0.012)
10	-0.044	-0.032	-0.013	-0.001	-0.001
	(0.011)	(0.012)	(0.015)	(0.014)	(0.014)
	log hours worked				
Imposed Restrictions:					
Effect URate at grad. symmetric up/downturn	no	no	yes	no	no
Effect Current Urate over exp=0	yes	yes	-	yes	yes
Spline for calendar year FE	no	no	no	no	no
Effect prov-time trends=0	no	no	-	no	no
Level of clustering††	$g*p$	$g*p$	g	$g*p$	$g*p$
1	-0.035	-0.025	-0.120	-0.052	-0.055
	(0.015)	(0.014)	(0.035)	(0.039)	(0.035)
2	-0.014	-0.007	-0.049	-0.021	-0.026
	(0.011)	(0.010)	(0.027)	(0.028)	(0.024)
3	0.008	0.012	0.023	0.010	0.004
	(0.009)	(0.010)	(0.027)	(0.024)	(0.020)
4	0.004	0.009	0.019	0.019	0.014
	(0.009)	(0.009)	(0.019)	(0.018)	(0.015)
5	0.001	0.005	0.014	0.028	0.025
	(0.009)	(0.009)	(0.013)	(0.015)	(0.012)
6	-0.003	0.002	0.010	0.038	0.035
	(0.011)	(0.011)	(0.010)	(0.014)	(0.013)

Continued on next page

Table A.24 – continued from previous page

Potential experience	baseline (1)	sensitivity tests: education† (2)	national§ (3)	2nd-step OLS (4)	one-step (5)
7	-0.002	0.003	-0.005	0.025	0.020
	(0.009)	(0.009)	(0.009)	(0.013)	(0.012)
8	-0.001	0.005	-0.020	0.012	0.004
	(0.008)	(0.008)	(0.011)	(0.014)	(0.013)
9	0.001	0.006	-0.035	-0.001	-0.011
	(0.009)	(0.009)	(0.015)	(0.016)	(0.015)
10	0.002	0.007	-0.050	-0.014	-0.027
	(0.011)	(0.012)	(0.019)	(0.020)	(0.017)
log earnings					
1	-0.058	-0.044	-0.133	-0.064	-0.065
	(0.019)	(0.018)	(0.034)	(0.038)	(0.036)
2	-0.041	-0.029	-0.071	-0.035	-0.039
	(0.015)	(0.014)	(0.027)	(0.028)	(0.025)
3	-0.025	-0.015	-0.008	-0.007	-0.013
	(0.013)	(0.014)	(0.026)	(0.024)	(0.021)
4	-0.029	-0.018	-0.004	0.003	-0.001
	(0.012)	(0.013)	(0.021)	(0.020)	(0.017)
5	-0.032	-0.021	-0.001	0.014	0.011
	(0.013)	(0.013)	(0.021)	(0.017)	(0.016)
6	-0.036	-0.024	0.003	0.025	0.023
	(0.014)	(0.014)	(0.026)	(0.018)	(0.017)
7	-0.038	-0.024	-0.013	0.014	0.010
	(0.012)	(0.013)	(0.025)	(0.018)	(0.017)
8	-0.040	-0.025	-0.030	0.004	-0.003
	(0.011)	(0.012)	(0.024)	(0.019)	(0.018)
9	-0.041	-0.025	-0.047	-0.006	-0.015
	(0.012)	(0.012)	(0.024)	(0.022)	(0.021)
10	-0.043	-0.025	-0.064	-0.016	-0.028
	(0.014)	(0.013)	(0.025)	(0.026)	(0.024)

Standard errors between parentheses. Column 1 reports the point estimates of the benchmark model reported in Section 2.14. Columns 2-5 report the following sensitivity tests (in order): including completed education FE as individual control variables in the first step of the two-step approach; national model; estimating the second step by OLS rather than FGLS; estimating the model in a one-step rather than two-step approach. Results in Column 1-3 are obtained from the second step FGLS of the two-step approach procedure explained in the text; Column 4 is obtained from the second step OLS, i.e. neglecting the measurement error introduced by the two-step procedure. Column 5 is estimated in one step. For all columns, log hourly wage and log hours worked are estimated jointly by means of a SUR, and the effects of interest on the log earnings are obtained as a linear combination of the effects on the first two outcomes, exploiting the adding-up constraint: log earnings=log hourly wage+log hours worked.

§ In the national model cells need not be dropped, since cells are always sufficiently large (see selection rule in Table 1.2 in Appendix 1.8). In contrast, all other sensitivity analyses (Column 2,4-5), which are all based on the provincial model, rely on the same sub-sample retained for the benchmark and consider the same specification used in the benchmark; in the one-step approach (column 5), this specification is augmented by the individual control variables used in the first step.

† Completed education is measured as the number of years of education successfully attained from the beginning of secondary education, i.e. from age 12. Therefore, repeated grades are counted only as soon as that grade is successfully passed.

†† Clustered (conventional) standard errors are reported if the model is (not) rejected on the basis of the χ^2 test at the 5% level.

Bibliography

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. *The American Economic Review 95*(1), 437–449.

Card, D. and T. Lemieux (2001). Dropout and Enrollment Trends in the Postwar Period: What Went Wrong in the 1970s? In J. Gruber (Ed.), *Risky Behavior among Youths: An Economic Analysis*, Chapter 9, pp. 439–482. University of Chicago Press.

Clark, D. (2011). Do Recessions Keep Students in School? The Impact of Youth Unemployment on Enrolment in Post-compulsory Education in England. *Economica 78*(311), 523–545.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics 10*(4), 417–451.

Frick, J. R. and M. M. Grabka (2003). Missing income data in the German SOEP: incidence, imputation and its impact on the income distribution. Discussion Paper 376, German Institute for Economic Research (DIW Berlin).

Jenkins, S. P. (1995). Easy Estimation Methods for Discrete-Time Duration Models. *Oxford Bulletin of Economics and Statistics 57*(1), 129–136.

Kalwij, A. and A. van Soest (2005). Item Non-Response and Alternative Imputation Procedures. In A. Börsch-Supan and H. Jürges (Eds.), *The Survey of Health, Ageing and Retirement in Europe. Methodology*, pp. 128–150. Mannheim Research Institute for the Economics of Aging.

Kiefer, N. M. (1988). Analysis of Grouped Duration Data. In N. U. Prabhu (Ed.), *Statistical Inference from Stochastic Processes*. American Mathematical Society, Providence.

Micklewright, J., M. Pearson, and S. Smith (1990). Unemployment and Early School Leaving. *The Economic Journal 100*(400), 163–169.

Moulaert, T. and J. Verly (2006). Le revenu minimum mensuel moyen garanti. Chronique internationale de l'IRES 103, Université Catholique de Louvain.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012). The Short– and Long–Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Petrongolo, B. and M. J. San Segundo (2002). Staying–on at school at 16: the impact of labor market conditions in Spain. *Economics of Education Review 21*(4), 353–365.

Särndal, C.-E. and S. Lundström (2005). *Imputation*, pp. 153–165. John Wiley and Sons, Ltd.

SONAR (2003). Hoe maken de jongeren de overgang van school naar werk? Basisrapportering Cohorte 1978 (eerste golf). Technical report, Leuven: Steunpunt WAV.

SONAR (2004a). Hoe maken de jongeren de overgang van school naar werk? Basisrapportering Cohorte 1976 (tweede golf). Technical report, Leuven: Steunpunt WAV.

SONAR (2004b). Hoe maken de jongeren de overgang van school naar werk? Cohorte 1980 (eerste golf). Technisch rapport. Technical report, Leuven: Steunpunt WAV.

Tumino, A. and M. Taylor (2013). The impact of local labour market conditions on school leaving decisions. Mimeo, ISER, University of Essex.

Wooldridge, J. M. (2006). Cluster-Sample Methods in Applied Econometrics: An Extended Analysis. Mimeo, Department of Economics, Michigan State University.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section And Panel Data (second ed.)*. The MIT Press, Cambridge, MA.

B

Supplementary Appendix to "Is it Socially Efficient to Impose Job Search Requirements on Unemployed Benefit Claimants with Hyperbolic Preferences?" The Solution and Discussion for the Naive Agent

In the main text we have only considered the case in which agents have sophisticated time preferences. In this Appendix we analyze the behavior of agents who have naive time preferences. The literature contrasts two types of hyperbolic agents, a sophisticated and a naive one. They differ in the perception of how their respective future selves will behave. A sophisticated agent correctly realizes that her future selves will act exactly as the current self (discounting by $\beta\delta$), while a naive agent wrongly believes that her future selves will behave as an exponential agent (discounting by δ). Using the terminology of (Gruber and Köszegi, 2000, 2001) they both have a *self-control* problem, but only the naive agent has a *misperception* problem.

B.1 Perfect Monitoring

The optimization problem of the current self of a *naive* or a *sophisticated* agent coincides and can be formally represented by (2)-(4) in the main text. By contrast, the optimization problem of the future selves differs between the *naive* and *sophisticated* agent. While the sophisticated agent knows that she will continue to set her search effort in the future to the current level, a naive agent believes that she will set her future search effort to the same level as an exponential discounter would do. This means that the first-order condition for search effort, (12) in the main text, is not affected, but in the first-order condition of the reservation wage (11) the optimal search effort of the exponential agent σ_u^e replaces σ_u of the hyperbolic agent:

$$y_u + \frac{\delta\lambda\left(\sigma_u^e\right)}{1-\delta(1-q)} Q\left(x_u\right) = c\left(\sigma_u^e\right) + x_u, \qquad u \in \{b, z\} \tag{B.1}$$

where σ_u^e is determined by the first-order condition of search effort for the exponential agent, which is equivalent to (12) in the main text, but where $\beta = 1$:

$$\frac{\delta\lambda'\left(\sigma_u^e\right)}{1-\delta(1-q)} Q\left(x_u\right) + \mu_u^e = c'\left(\sigma_u^e\right) \text{ and } \mu_u^e\left(\sigma_u^e - \bar{\sigma}\right) = 0, \qquad u \in \{b, z\} \tag{B.2}$$

where $\mu_b^e \geq 0$ is the Lagrange multiplier associated with the inequality constraint $\sigma_b^e \geq \bar{\sigma}$ and where $\mu_z^e = 0$, since job search is not constrained in the case of a sanction. Since search effort of an exponential agent is higher than that of a hyperbolic agent, this constraint is less likely to bind than the one in (12).

So, the solution for the naive agent is characterized by three first-order conditions instead of two: (B.1), (B.2) and (12) jointly determine the optimal solution $(\sigma_u, x_u, \sigma_u^e)$ for $u \in \{b, z\}$. Notice that we use the same notations as in the main text, but σ_u and x_u now designate the optimal solution for a naive agent instead of a sophisticated one. This convention is maintained throughout this Appendix. Observe that the naive agent sets her reservation wage at the level that the exponential agent sets it: $x_u = x_u^e$.

These first-order conditions can be represented by the following functions defined in the main text: $x = r(\sigma \mid b)$, $x = s(\sigma \mid 1)$, and $x = s(\sigma \mid \beta < 1)$. This allows to represent the solution of the naive agent graphically. The reservation wage of the unconstrained unemployed benefit claimant x_b is determined, as for an exponential agent, by the intersection between $x = r(\sigma \mid b)$, $x = s(\sigma \mid 1)$ at point A in Figure B.1 below. The level of search effort is then set at point point B, the intersection of the horizontal line through x_b and $x = s(\sigma \mid \beta < 1)$. Similarly, the reservation wage and search effort of the sanctioned or non-complying agent are determined respectively by points C and D.

Let us now consider the behavior of the agent as the search requirement $\bar{\sigma}$ is raised. As long as $\bar{\sigma} < \sigma_b$, the agent does not change her behavior characterized by point B. If $\bar{\sigma}$ is set at a higher level than σ_b, she will at first comply and increase search effort accordingly. However, since a naive agent believes that she will act as an exponential agent in the future, she will not adjust her reservation wage as long as the requirement is set below the optimal search level of an exponential agent σ_b^e at point A. So, the optimal (σ_b, x_b) pair first follows the straight line BA. The reservation wage (and hence lifetime utility of the future selves) is lowered along the line passing through AHF only if $\bar{\sigma} > \sigma_b^e$.

The decision to comply or not depends on the lifetime utility of the current self. This utility starts decreasing as soon as $\bar{\sigma}$ is set at a higher level than σ_b to the right of point B, because, as formally demonstrated in the proof of Proposition 6, the instantaneous cost of search increases

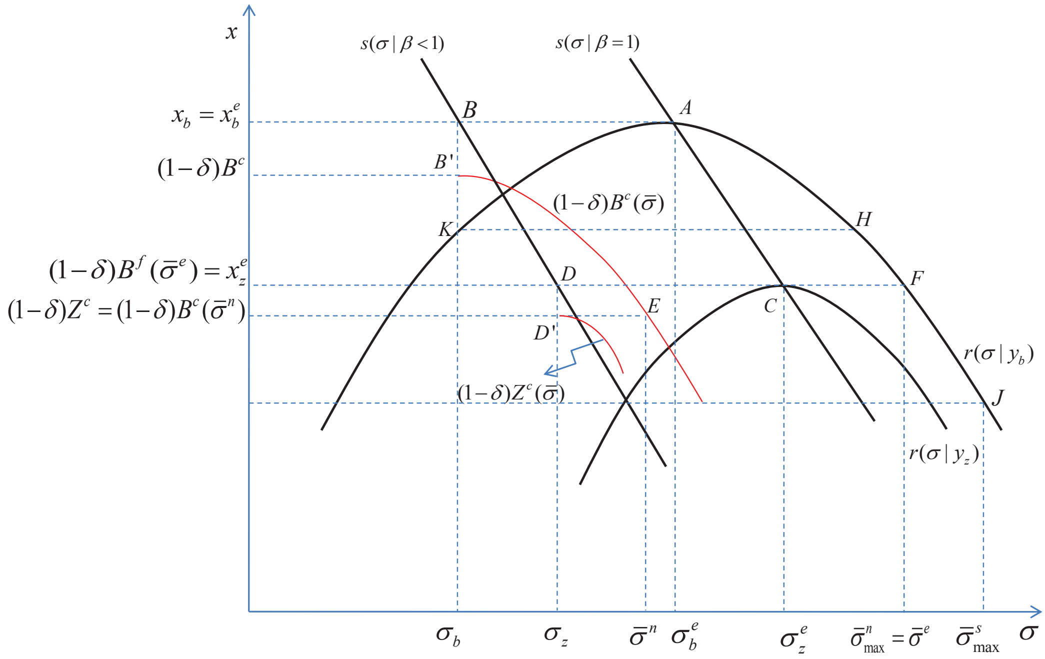


Figure B.1: The Solution for the Naive Agent in Case of Perfect Monitoring. x = reservation wage; σ = realized search effort.

while the reservation wage remains unaffected. The naive agent will stop complying as soon as the search requirement is raised above the effort level $\bar{\sigma}^n$ at which the current self is indifferent between complying or being sanctioned. The maximum search requirement $\bar{\sigma}^n$ verifies the following equation:

$$B^c(\bar{\sigma}^n) = Z^c \tag{B.3}$$

In Figure B.1, this corresponds to the search intensity attained at point E.

Without being more precise about preferences and the search technology of the agent, we can say little about the exact level of $\bar{\sigma}^n$. Nevertheless, in Proposition 6 it is demonstrated that this level can be bracketed: $\bar{\sigma}^n \in (\sigma_z, \bar{\sigma}^n_{max})$, where $\bar{\sigma}^n_{max}$ is equal to the maximum search requirement $\bar{\sigma}^e$ before an exponential agent stops complying. In Figure B.1, this corresponds to the search intensity attained at point F, where the reservation wage of the complier $r(\bar{\sigma} \mid y_b)$ is equal to that of a sanctioned exponential agent x^e_z. This means that the imposition of a search requirement leads to non-compliance at lower effort levels for naive hyperbolic agents than for exponential agents, as considered by Manning (2009) and Petrongolo (2009), since $\bar{\sigma}^n \in (\sigma_z, \bar{\sigma}^n_{max} = \bar{\sigma}^e) < \bar{\sigma}^e$, and than sophisticated hyperbolic agents, as considered in the main text. In the latter case the maximum search requirement that can be attained before an agent stops complying is equal to

σ^s_{max} located on the abscissa of point J.

Proposition 6.

(i) The lifetime utility of the naive current self is unaffected by the search requirement for $\bar{\sigma}$ lower than her optimal free choice σ_b and is strictly decreasing in $\bar{\sigma}$ if it is higher.

(ii) The maximum search requirement $\bar{\sigma}$ at which a naive agent stops complying is not lower than the optimal search effort σ_z of a sanctioned naive hyperbolic agent and is strictly lower than the search effort $\bar{\sigma}^e$ at which an exponential agent stops complying.

Proof. See Appendix B.3. □

B.2 The Consequences of an Imperfect Monitoring Technology

We will not develop the complete analysis for a naive agent in the case of an imperfect monitoring technology, since the analysis is very similar to the case of a sophisticated agent. The main difference is that the long-run utility of the naive agent is constantly at the level of an agent with exponential time preferences because of her misperception problem. This means that the long-run utility of a naive agent is either constant or decreasing in σ, and never strictly increasing. Consequently, if in the benchmark case of an unbounded support of the measurement error the search requirement and, hence, the sanction probability is raised above zero, the long-run utility immediately strictly decreases, while for a sophisticated agent it initially increases. Proposition 5 in the main text differs therefore in this respect and we will provide a version of it, as well as its proof, in Proposition 7 below.

Proposition 7.

Assume that the sanction probability is expressed by (28) in the main text and that the support of measurement error is unbounded. Then, the optimal search effort for a naive agent σ_p strictly increases (resp., decreases) in $\bar{\sigma}$ for $\bar{\sigma} \to 0$ and, hence, $p(\bar{\sigma}/\sigma_p) \to 0$ (resp., $\bar{\sigma} \to +\infty$ and, hence, $p(\bar{\sigma}/\sigma_p) \to 1$). The reservation wage x_p is always decreasing in $\bar{\sigma}$. The Pareto frontier cannot be reached.

Proof. See Appendix B.4. □

B.3 Proof of Proposition 6

(i) *Proof.* Using (1), (2), (8) and (9) in the main text, and noting that the reservation wage of the naive agent is only affected by $\bar{\sigma}$ if $\bar{\sigma} > \sigma_b^e$, we can write:

$$B^c(\bar{\sigma}) = \max_{\sigma \geq \bar{\sigma}} y_b - c(\sigma) + \beta\delta \left\{ \frac{\lambda(\sigma)}{1-\delta(1-q)} Q\left[x_b^e(\bar{\sigma})\right] + \frac{x_b^e(\bar{\sigma})}{1-\delta} \right\} \tag{B.4}$$

where $x_b^e(\bar{\sigma}) \equiv 1_{\{\bar{\sigma} \leq \sigma_b^e\}} x_b^e + 1_{\{\bar{\sigma} > \sigma_b^e\}} r(\bar{\sigma} \mid y_b)$ and $1_{\{A\}} = 1$ if A is true and $1_{\{A\}} = 0$ otherwise. The search effort solving the maximization problem (B.4) is equal to σ_b if the latter is strictly higher than $\bar{\sigma}$ (with $x_b^e(\bar{\sigma}) = x_b^e$) and $\bar{\sigma}$ otherwise.

Partially differentiating $B^c(\bar{\sigma})$ in (B.4) with respect to $\bar{\sigma}$ is equal to zero if $\sigma_b > \bar{\sigma}$. Otherwise, using that $Q'\left[r(\sigma \mid y_u)\right] = -\bar{F}\left[r(\sigma \mid y_u)\right] r(\sigma \mid y_u)$:

$$\frac{\partial B^c(\bar{\sigma})}{\partial \bar{\sigma}} = S\left(\bar{\sigma}, x_b^e(\bar{\sigma}) \mid \beta\right) + 1_{\{\bar{\sigma} > \sigma_b^e\}} \beta\delta \frac{\left[(1-\delta)\left(1 - h\left[\bar{\sigma}, r(\bar{\sigma} \mid y_b)\right] + \delta q\right)\right]}{[1-\delta(1-q)](1-\delta)} \frac{\partial r(\bar{\sigma} \mid y_b)}{\partial \bar{\sigma}}$$

Since $S\left(\bar{\sigma}, x_b^e(\bar{\sigma}) \mid \beta\right) \leqq 0$ iff $\bar{\sigma} \geqq \sigma_b$ and since, by Proposition 1 in the main text, $\partial r(\bar{\sigma} \mid y_b)/\partial\bar{\sigma} < 0$ iff $\bar{\sigma} > \sigma_b^e$, $\partial B^c(\bar{\sigma})/\partial\bar{\sigma} \leqq 0$ iff $\bar{\sigma} \geqq \sigma_b$. In sum, $\partial B^c(\bar{\sigma})/\partial\bar{\sigma} = 0$ iff $\bar{\sigma} \leq \sigma_b$ and $\partial B^c(\bar{\sigma})/\partial\bar{\sigma} < 0$ iff $\bar{\sigma} > \sigma_b$. □

(ii) *Proof.* We prove (ii) by showing that for any $\bar{\sigma} \geq \sigma_b$ the difference between the expected lifetime utility of a complying and a non-complying current self must always be strictly greater for an agent discounting the future at an exponential rate ($\beta = 1$) than for an agent discounting the future at a hyperbolic rate ($\beta < 1$). Consequently the hyperbolic agent will always stop complying (when $B^c(\bar{\sigma}^n) = \tilde{Z}^c$) at a lower level of search effort than that of an exponential agent.

(1) Using (2), (8) and (9) of the main text, and restricting the analysis to the cases where $\bar{\sigma} \geq \sigma_b$, we obtain:

$$\forall \bar{\sigma} < \sigma_b^e : B^f(\bar{\sigma}) - B^c(\bar{\sigma}) = \delta(1-\beta) \left\{ \frac{\lambda(\bar{\sigma})}{[1-\delta(1-q)]} Q(x_b^e) + \frac{x_b^e}{(1-\delta)} \right\} \tag{B.5}$$

and

$$\forall \bar{\sigma} \geq \sigma_b^e : B^f(\bar{\sigma}) - B^c(\bar{\sigma}) = \delta(1-\beta) \left\{ \frac{\lambda(\bar{\sigma})}{[1-\delta(1-q)]} Q(r(\bar{\sigma}|y_b)) + \frac{r(\bar{\sigma}|y_b)}{(1-\delta)} \right\} \tag{B.6}$$

Similarly, using (3), (8) and (9) in the main text, we find

$$Z^f - Z^c = \delta(1-\beta) \left\{ \frac{\lambda(\sigma_z)}{[1-\delta(1-q)]} Q(x_z^e) + \frac{x_z^e}{(1-\delta)} \right\} \tag{B.7}$$

(2) Consider Definition (10) in the main text and assume that $x_1 < x_2$. Then we can rewrite $Q(x_1)$ as follows:

$$\begin{aligned} Q(x_1) &= \int_{x_1}^{x_2} (w - x_1)\, dF(w) + Q(x_2) + \bar{F}(x_2)\,(x_2 - x_1) \\ &= Q(x_2) + \bar{F}(x_1)\,(x_2 - x_1) - \left\{(x_2 - x_1) - E\left(w - x_1 | x_1 \leq w < x_2\right)\right\} \\ &\times \left[\bar{F}(x_1) - \bar{F}(x_2)\right] \end{aligned} \tag{B.8}$$

(3) Since $x_z^e < x_b^e$, we can use (10) in the main text and (B.8) to rewrite (B.7) as follows:

$$\begin{aligned} Z^f - Z^c &= \frac{\delta(1-\beta)}{[1-\delta(1-q)]}\Big\{\lambda(\sigma_z)Q(x_b^e) + h(\sigma_z, x_z^e)(x_b^e - x_z^e) + x_z^e\frac{1-\delta+\delta q}{1-\delta} \\ &- \lambda(\sigma_z)(x_b^e - x_z^e) + \lambda(\sigma_z)E(w - x_z^e | x_z^e \leq w < x_b^e) \\ &\times [\bar{F}(x_z^e) - \bar{F}(x_b^e)]\Big\} \end{aligned} \tag{B.9}$$

Subtracting (B.9) from (B.5) then yields for $\bar{\sigma} \in [\sigma_z, \sigma_b]$:

$$\begin{aligned} &B^f(\bar{\sigma}) - B^c(\bar{\sigma}) - Z^f + Z^c = \frac{\delta(1-\beta)}{[1-\delta(1-q)]}\Big\{Q(x_b^e)[\lambda(\bar{\sigma}) - \lambda(\sigma_z)] + (x_b^e - x_z^e) \\ &\times\frac{[(1-\delta)(1-h(\sigma_z, x_z^e)) + \delta q]}{(1-\delta)} + \lambda(\sigma_z)\left\{(x_b^e - x_z^e) - E(w - x_z^e)|x_z^e \leq w < x_b^e)\right\} \\ &\times\left[\bar{F}(x_z^e) - \bar{F}(x_b^e)\right]\Big\} > 0 \end{aligned} \tag{B.10}$$

Since $\forall\bar{\sigma} \in [\sigma_b^e, \bar{\sigma}_{max}^n] : r(\bar{\sigma}|y_b) > x_z^e$, we can derive using (B.6) a similar expression as (B.10) in which x_b^e is replaced by $r(\bar{\sigma}|y_b)$. Consequently,

$$\forall\bar{\sigma} \in [\sigma_z, \bar{\sigma}_{max}^n] : B^f(\bar{\sigma}) - Z^f > B^c(\bar{\sigma}) - Z^c \tag{B.11}$$

Since $B^f(\bar{\sigma}_{max}^n) = Z^f$, (B.11) implies that $B^c(\bar{\sigma}_{max}^n) < Z^c$. Because $B^c(\sigma_z) > Z^c$ and $B^c(\cdot)$ is a continuous function, it must be that $\sigma_z < \bar{\sigma}^n < \bar{\sigma}_{max}^n$. Finally, since the reservation wage of a naive agent (and hence lifetime utility of the future selves Z^f) is equal to that of an exponential agent ($x_z = x_z^e$), $B^f(\bar{\sigma}_{max}^n) = Z^f$ implies that $\bar{\sigma}_{max}^n$ is equal to the maximum search requirement $\bar{\sigma}^e$ at which an exponential agent stops complying.

□

B.4 Proof of Proposition 7

Proof. Following arguments that are similar as those for the sophisticated agent in the case of measurement error, but adjusted, as in Section 1 of this Internet Appendix, the first-order conditions define a system of three equations in three unknowns: x, σ and σ^e. σ denotes the search effort of the naive agent, while σ^e refers to the effort of an exponential agent. Notice that because of the misperception problem the reservation wage of the naive and the exponential agent are equal: $x = x^e$. The system is thus

$$Ey + \frac{\delta\lambda(\sigma^e)Q(x)}{1-\delta(1-q)} - x - c(\sigma^e) - p\left(\frac{\bar{\sigma}}{\sigma^e}\right)\frac{\delta}{(1-\delta)}\left[1-h\left(\sigma^e, x\right)\right](x - x_z) = 0 \tag{B.12}$$

$$\frac{\beta\delta\lambda'(\sigma)}{1-\delta(1-q)}Q(x) - \frac{\partial p(\bar{\sigma}/\sigma)}{\partial\sigma}\left\{y_b - y_z + \frac{\beta\delta}{(1-\delta)}\left[1-h\left(\sigma, x\right)\right](x - x_z)\right\}$$
$$+p\left(\frac{\bar{\sigma}}{\sigma}\right)\frac{\beta\delta}{(1-\delta)}\lambda'(\sigma)\bar{F}(x)(x - x_z) - c'(\sigma) = 0 \tag{B.13}$$

$$\frac{\delta\lambda'(\sigma^e)}{1-\delta(1-q)}Q(x) - \frac{\partial p(\bar{\sigma}/\sigma^e)}{\partial\sigma^e}\left\{y_b - y_z + \frac{\delta}{(1-\delta)}\left[1-h\left(\sigma^e, x\right)\right](x - x_z)\right\}$$
$$+p\left(\frac{\bar{\sigma}}{\sigma^e}\right)\frac{\delta}{(1-\delta)}\lambda'(\sigma^e)\bar{F}(x)(x - x_z) - c'(\sigma^e) = 0 \tag{B.14}$$

The first equation is the first-order condition of the reservation wage, Equation (31) in the main text, in which σ is replaced by σ^e. The second is the first-order condition of search for the naive agent, corresponding to (32) in the main text, but in which σ and x refer to the behavior of a naive agent instead of a sophisticated one. Finally, the third equation is the first-order condition of search effort for the exponential agent. This corresponds to equation (32) in the main text for which β is set to one and σ is replaced by σ^e.

Totally differentiating this system yields:

$$\begin{bmatrix} A_1 & A_2 & A_3 \\ A_4 & A_5 & A_6 \\ A_7 & A_8 & A_9 \end{bmatrix}\begin{bmatrix} dx \\ d\sigma \\ d\sigma^e \end{bmatrix} = \begin{bmatrix} A_{10} \\ A_{11} \\ A_{12} \end{bmatrix} d\bar{\sigma} \tag{B.15}$$

where, without recalling the arguments of $\lambda, c, p, f, \bar{F}, Q, h$ and their derivatives and without writing a subscript e to denote that a function is evaluated at σ^e

$$A_1 = -\frac{\delta\lambda_e\bar{F}}{1-\delta(1-q)} - 1 - p_e\frac{\delta}{1-\delta}\left[\lambda_e f(x - x_z) + 1 - h_e\right] < 0 \tag{B.16}$$

$$A_2 = 0 \tag{B.17}$$

$$\begin{aligned}
A_3 &= \frac{\delta\lambda_e' Q}{1-\delta(1-q)} - c_e' - \frac{\partial p_e}{\partial \sigma^e}\left\{y_b - y_z + \frac{\delta(1-h)(x-x_z)}{1-\delta}\right\} + p_e\frac{\delta\lambda_e'\bar{F}(x-x_z)}{1-\delta} \\
&= S_p(\sigma^e, x \mid 1, \bar{\sigma}) && \text{(B.18)} \\
A_4 &= -\frac{\beta\delta\lambda'\bar{F}}{1-\delta(1-q)} - \frac{\partial p}{\partial \sigma}\frac{\beta\delta}{1-\delta}\left[\lambda f(x-x_z) + 1 - h\right] \\
&+ p\frac{\beta\delta}{1-\delta}\lambda'\left[\bar{F} - f(x-x_z)\right] && \text{(B.19)} \\
A_5 &= \frac{\beta\delta\lambda'' Q}{1-\delta(1-q)} - \frac{\partial^2 p}{[\partial\sigma]^2}\left\{y_b - y_z + \frac{\beta\delta(1-h)(x-x_z)}{1-\delta}\right\} + 2\frac{\partial p}{\partial\sigma}\frac{\beta\delta\lambda'\bar{F}(x-x_z)}{1-\delta} \\
&+ p\frac{\beta\delta\lambda''\bar{F}(x-x_z)}{1-\delta} - c'' && \text{(B.20)} \\
A_6 &= 0 && \text{(B.21)} \\
A_7 &= -\frac{\delta\lambda_e'\bar{F}}{1-\delta(1-q)} - \frac{\partial p_e}{\partial\sigma^e}\frac{\delta\left[\lambda_e f(x-x_z) + 1 - h_e\right]}{1-\delta} + p_e\frac{\delta\lambda_e'\left[\bar{F} - f(x-x_z)\right]}{1-\delta} && \text{(B.22)} \\
A_8 &= 0 && \text{(B.23)} \\
A_9 &= \frac{\delta\lambda_e'' Q}{1-\delta(1-q)} - \frac{\partial^2 p_e}{[\partial\sigma^e]^2}\left\{y_b - y_z + \frac{\delta(1-h_e)(x-x_z)}{1-\delta}\right\} + 2\frac{\partial p_e}{\partial\sigma^e}\frac{\delta\lambda_e'\bar{F}(x-x_z)}{1-\delta} \\
&+ p_e\frac{\delta\lambda_e''\bar{F}(x-x_z)}{1-\delta} - c_e'' && \text{(B.24)} \\
A_{10} &= \frac{\partial p_e}{\partial\bar{\sigma}}\left\{y_b - y_z + \frac{\delta(1-h_e)(x-x_z)}{1-\delta}\right\} && \text{(B.25)} \\
A_{11} &= \frac{\partial^2 p}{\partial\sigma\partial\bar{\sigma}}\left\{y_b - y_z + \frac{\beta\delta(1-h)(x-x_z)}{1-\delta}\right\} - \frac{\partial p}{\partial\bar{\sigma}}\frac{\beta\delta\lambda'\bar{F}(x-x_z)}{1-\delta} && \text{(B.26)} \\
A_{12} &= \frac{\partial^2 p_e}{\partial\sigma^e\partial\bar{\sigma}}\left\{y_b - y_z + \frac{\delta(1-h_e)(x-x_z)}{1-\delta}\right\} - \frac{\partial p_e}{\partial\bar{\sigma}}\frac{\delta\lambda_e'\bar{F}(x-x_z)}{1-\delta} && \text{(B.27)}
\end{aligned}$$

Solving system (B.12)-(B.14) yields

$$\begin{bmatrix} \frac{\partial x}{\partial\bar{\sigma}} \\ \frac{\partial\sigma}{\partial\bar{\sigma}} \\ \frac{\partial\sigma^e}{\partial\bar{\sigma}} \end{bmatrix} = \frac{1}{D_n}\begin{bmatrix} A_5A_9 & 0 & -A_3A_5 \\ -A_4A_9 & (A_1A_9 - A_3A_7) & A_3A_4 \\ -A_5A_7 & 0 & A_1A_5 \end{bmatrix}\begin{bmatrix} A_{10} \\ A_{11} \\ A_{12} \end{bmatrix} \tag{B.28}$$

where $D_n = A_5(A_1A_9 - A_3A_7)$.

In order to sign these partial derivatives, we use a couple of results of the proof of Proposition 5 in the Appendix of the main text. First, by substituting (A.25) into (B.25), we obtain that $A_{10} \geq 0$. Second, substituting (A.26) into (B.20) and (B.24) yields $A_5 < 0$ and $A_9 < 0$ if $g'\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right]$ is positive or not too negative, an assumption we make throughout. Third, since $A_3 = S_p(\sigma^e, x \mid 1, \bar{\sigma})$, we have that $A_3 = 0$ if we evaluate it at the optimal choice (σ_b^e, x_b), since for a naive agent the reservation wage and search effort of the future selves are set at the optimum of an exponential agent ($x_b = x_b^e$). Observe that as a result, at the optimum, $D_n = A_1A_5A_9 < 0$.

Finally, we insert (A.25)-(A.27) in (B.19) and (B.26)[1] to observe that these have ambiguous signs:

$$A_4 = -\frac{\beta\delta}{(1-\delta)}\left\{\lambda'\bar{F}\frac{[(1-p)(1-\delta)-p\delta q]}{[1-\delta(1-q)]} + p\lambda' f(x-x_z) - \frac{g}{\sigma}\left[\lambda f(x-x_z)+(1-h)\right]\right\} \tag{B.29}$$

$$A_{11} = -\frac{g'}{\sigma\bar{\sigma}}\left\{y_b - y_z + \frac{\beta\delta(1-h)(x-x_z)}{1-\delta}\right\} - \frac{g}{\bar{\sigma}}\frac{\beta\delta\lambda'\bar{F}(x-x_z)}{1-\delta} \tag{B.30}$$

However, as in the proof of Proposition 5, we now show that these partial derivatives can be unambiguously signed in the limiting cases where p tends to zero or to one. We only consider the benchmark case in which the support of the measurement error is unbounded. This means that the following properties are satisfied: $\forall \sigma, \bar{\sigma} \in (0,+\infty) : g\left[log\left(\frac{\bar{\sigma}}{\sigma}\right)\right] > 0$, $\lim_{\varepsilon\to-\infty} g(\varepsilon) = \lim_{\varepsilon\to+\infty} g(\varepsilon) = 0$, $\lim_{\varepsilon\to-\infty} g'(\varepsilon) > 0$, and $\lim_{\varepsilon\to+\infty} g'(\varepsilon) < 0$.

Case 1: $\mathbf{p} \to \mathbf{0}$
Substituting $\lim_{\varepsilon\to-\infty} g(\varepsilon) = 0$ and $\lim_{\varepsilon\to-\infty} g'(\varepsilon) > 0$ in (B.29) and (B.30) yields

$$A_4 \to -\frac{\beta\delta\lambda'\bar{F}}{1-\delta(1-q)} < 0 \tag{B.31}$$

$$A_{11} \to -\frac{g'}{\sigma\bar{\sigma}}\left\{y_b - y_z + \frac{\beta\delta(1-h)(x-x_z)}{1-\delta}\right\} < 0 \tag{B.32}$$

Case 2: $\mathbf{p} \to \mathbf{1}$
Using $\lim_{\varepsilon\to+\infty} g'(\varepsilon) < 0$ and $x - x_z \to 0$ for $p \to 1$ allows to simplify equations (B.29) and (B.30) as follows:

$$A_4 \to \frac{\beta\delta}{(1-\delta)}\left\{\lambda'\bar{F}\frac{p\delta q}{[1-\delta(1-q)]} + \frac{g}{\sigma}(1-h)\right\} > 0 \tag{B.33}$$

$$A_{11} \to -\frac{g'}{\sigma\bar{\sigma}}\left\{y_b - y_z\right\} > 0 \tag{B.34}$$

where the last inequality follows from the fact that $\lim_{\varepsilon\to+\infty} g'(\varepsilon) < 0$ in the benchmark when the measurement error has an infinite support, and $g'(\bar{\varepsilon}) = 0$ in the case of a finite support.

To sum up, we found that $A_1 < 0$, $A_2 = 0$, $A_3 = 0$, $A_5 < 0$, $A_6 = 0$, $A_8 = 0$, $A_9 < 0$, $A_{10} \geq 0$ and $D_n < 0$, while $A_4 < 0$ and $A_{11} < 0$ if $p \to 0$ and $A_4 > 0$ and $A_{11} > 0$ if $p \to 1$. By continuity, there exists at least one $\tilde{p}_n \in (0,1)$ at which $A_4 = 0$ and at least one $\hat{p}_n \in (0,1)$ at which $A_{11} = 0$.

Inserting these results in (B.28) and evaluating these partial derivatives at the optimal solution $(\sigma_p, x_p, \sigma_p^e)$ yields:

$$\frac{\partial x_p}{\partial \bar{\sigma}} = \frac{A_5 A_9 A_{10}}{D_n} = \frac{A_{10}}{A_1} \tag{B.35}$$

$$\frac{\partial \sigma_p}{\partial \bar{\sigma}} = \frac{A_9(A_1 A_{11} - A_4 A_{10})}{D_n} = \frac{(A_1 A_{11} - A_4 A_{10})}{A_1 A_5} \tag{B.36}$$

[1]We do not consider A_7 and A_{12}, because we are not interested in the behavior of the future selves of a naive agent.

Consider first the case where $p \to 0$. Observe that since $\lim_{\varepsilon \to -\infty} g(\varepsilon) = 0$, $A_{10} \to 0$, so that in the limit $\frac{\partial x_p}{\partial \bar{\sigma}}$ tends to zero. However, for any finite ε, $g(\varepsilon) > 0$, so that $A_{10} > 0$, and, hence, $\frac{\partial x_p}{\partial \bar{\sigma}} < 0$. From (B.36) it is clear that $\frac{\partial \sigma_p}{\partial \bar{\sigma}} > 0$ in the case where $p \to 0$. Inserting the values of the partial derivatives for the case that $p \to 1$ yields again that $\frac{\partial x_p}{\partial \bar{\sigma}} < 0$ close to $p = 1$. Since $A_{10} \to 0$ if $p \to 1$, while A_{11} remains strictly negative, we obtain that $\frac{\partial \sigma_p}{\partial \bar{\sigma}} < 0$ sufficiently close to $p = 1$. □

Bibliography

Gruber, J. and B. Köszegi (2000). Is Addiction "Rational"? Theory and Evidence. Working Paper 7507, NBER.

Gruber, J. and B. Köszegi (2001). Is Addiction “Rational”? Theory and Evidence. *The Quarterly Journal of Economics 116*(4), 1261–1303.

Manning, A. (2009). You can’t always get what you want: The impact of the UK jobseeker’s allowance. *Labour Economics 16*(3), 239–250.

Petrongolo, B. (2009). The long-term effects of job search requirements: Evidence from the UK JSA reform. *Journal of Public Economics 93*(11–12), 1234–1253.