Essays on Unemployment and Active Labour Market Policies

By Jonas Staghøj

A dissertation submitted to
the Faculty of Social Sciences, University of Aarhus
in partial fulfilment of the requirements of
the PhD degree in
Economics and Management
# Table of Contents

Preface .......................................................... v

Summary .......................................................... vii

Dansk resume (Danish Summary) ............................... xi

Chapter 1  
A Danish Profiling System .................................... 1

Chapter 2  
Choosing the Best Training Programme: Is there a Case for Statistical Treatment Rules? .............................................. 31

Chapter 3  
Using Statistical Treatment Rules for Assignment of Participants in Labor Market Programs .................................................... 81

Chapter 4  
Analyzing the Ins and Outs of Unemployment using Danish Spell Data ................................................................. 139
Preface

This thesis was written in the period from September 2005 to September 2009 while I was a PhD student at the School of Economics and Management, Aarhus University and during my visit at University of Michigan. I am grateful for the financial support from the School of Economics and Management that has allowed me to participate in numerous courses, workshops and conferences in Denmark, Germany, Netherland, Switzerland, Finland and USA. Travelling around the world to present your work is indeed a very nice feature of the PhD programme.

A number of people have contributed to this thesis. First, I would like to thank my main advisor, Michael Svarer. He taught the class in Labour Economics that led me into the PhD study and has always been ready to provide comments and constructive suggestions. I also thank my secondary advisor, Michael Rosholm. Chapters one and two of my thesis was written in close cooperation with both of my advisors, and I have always felt that the straightforward way in which we have discussed the progress along the way, has made it easy to be a PhD student.

From September 2008 to March 2009 I had the great pleasure to visit Professor Jeffrey A. Smith at the Department of Economics at University of Michigan. This visit was no doubt the most inspiring experience during my PhD study. To meet and listen to so many enthusiastic economists made a deep impression on me. I want to thank Jeffrey A. Smith for very valuable discussions and for really taking an interest in making me feel comfortable during the stay; whether it be the weekly suggestions of interesting seminars to attend or invitations to dinner parties, concerts etc.

I would also like to thank my fellow PhD students for providing a friendly working environment and for numerous trips to the coffee machine and the football table. Not least Rune Vejlin and my office mate Torben Sørensen, who both started at the School of Economics and Management together with me almost eight years ago. Finally, I thank my family for love, support and encouragement throughout the years.

Jonas Staghøj

Aarhus, August 2009
Updated preface

The pre-defence was held on October 26, 2009. I am grateful to the members of the assessment committee, Mette Ejrnæs, Mike Elsby and Marianne Simonsen (chair), for their careful reading of the thesis and their many useful and constructive comments and suggestions. Some of the suggestions have been incorporated in the present version of the thesis while others remain for future work on the chapters.

Jonas Staghøj

Aarhus, November 2009
Summary

This thesis consists of four self-contained chapters about unemployment and active labour market policies. Unemployment has always been a key concept for economists for a number of reasons. It is important for the government’s budget since unemployed workers are unproductive and expensive when unemployment benefits are paid. And for individuals, unemployment can have substantial effects on the standards of living, especially for those individuals experiencing long-term unemployment. These consequences of unemployment naturally lead to policies designed to decrease unemployment. And in Denmark, an important part of the policies has the form of Active Labour Market Programmes (ALMPs). In this thesis I analyze the effects of these programmes and suggest specific roads for improvements. The starting point for the analysis is that previous evaluations of ALMPs in Denmark and internationally generally find that the effects of these programmes are modest and sometimes even negative (Heckman, LaLonde & Smith, 1999 and Kluve, 2006). One reason for this may be that ALMPs are not used in an optimal way; i.e. if the effects of ALMPs are heterogeneous across the population, it may be that the unemployed are not allocated to the programmes in such a way that the programme effect is maximized.

The first chapter is written together with my advisors, Michael Svarer and Michael Rosholm, and it is called “A Danish Profiling System”. In the paper we develop a method to identify unemployed at risk of long-term unemployment. The reason for this is that not all workers entering unemployment need help to find a new job. A major part of unemployment in Denmark is actually made up of workers who are only unemployed for a very short period before they move on to a new job, and it would inefficient to assign expensive active labour market programmes to these individuals. But other unemployed workers are not able or willing to find a job and they stay unemployed for longer periods. It seems appropriate to target programmes towards these individuals. A potential problem is that it may not be optimal to assign unemployed workers to a programme, when they have actually experienced a long period of unemployment. It may be better immediately to identify those at risk of long-term unemployment and assign them to a programme before they experience the potentially negative effects of being long-term unemployed. This approach is suggested in OECD (1998).
Models that seek to identify those in risk of long-term unemployment have been labelled *profiling* models in the literature and in this paper we use a duration model to estimate the probability of becoming long-term unemployed based on observable characteristics of the unemployed worker. Caseworkers can then be informed about this probability and use the information in their decision of whether or not to assign an individual to a programme. Our results show that the model performs reasonably well in terms of discriminating between short- and long-term unemployed. The predictions are however far from perfect, which is as expected when trying to predict a complex outcome based on relatively sparse information.

The second chapter is also written together with Michael Svarer and Michael Rosholm, and is called “Choosing the Best Training Programme: Is there a Case for Statistical Treatment Rules?” The paper is a rather natural extension of the first chapter, where we asked who should be assigned to a programme, since we now ask which program should be assigned to a particular unemployed worker? The latter question is aimed directly at improving the *effectiveness* of the programmes, whereas the question asked in chapter one, might be seen as an attempt to improve *equity* by assigning programmes to those most in need. Apart from distributional concerns across the pool of unemployed workers, the assignment of programmes is most interesting when effects are heterogeneous, so in this paper we seek to exploit the information about heterogeneous treatment effects over some observed individual characteristics. The model presented is a duration model that uses the timing-of-events framework to identify causal effects. We compare different assignment rules, and the results suggest that a significant reduction in the average duration of unemployment may be the result if a statistical treatment rule is introduced. The results corroborate the finding in Lechner & Smith (2007), where the current assignments based on caseworkers’ decisions are found to be no better than a random assignment of programmes.

The third chapter is written together with Jeff Smith and is called “Using Statistical Treatment Rules for Assignment of Participants in Labor Market Programs”. *Profiling* and *targeting* models analyzed in the previous chapters can more generally be seen as examples of Statistical Treatment Rules (STRs), and in this paper we survey the literature about the use of STRs for the assignment of participants in labor market programs. We first discuss the
theoretical advantages and disadvantages of STRs compared to alternative assignment mechanisms. Then we discuss a number of econometric issues regarding the estimation and evaluation of STRs that have not previously been emphasized in the literature. And finally, we analyse the empirical results available in the literature. We note, that the available evidence does often not allow for actual evaluations of the STRs, mainly because most studies lack information about the performance of alternative assignment mechanisms. The results do, however, suggest that profiling models can be used to increase equity but that targeting models are more useful if the goal is to increase efficiency. The positive results about targeting models are mostly based on simulation studies, so more evidence on the actual use of targeting models is needed in order to determine whether targeting models are empirically as relevant as theoretical and simulation studies suggest.

The final chapter is not about active labour market policies, but may still provide relevant policy implications. The paper is called “Analyzing the Ins and Outs of Unemployment using Danish Spell data”. In the paper I analyze the dynamics of unemployment using register data with weekly information for the entire population in Denmark followed over the period 1985-2003. This data is used to analyze whether fluctuations in the unemployment rate is mainly a consequence of fluctuations in the inflow to or the outflow from unemployment. Spurred by the new findings in Shimer (2007), a number of papers have sought to address this classic question and the main contribution in my paper is to use a new and more informative type of data for the analysis. I use various decomposition methods suggested in the recent literature and consistently find that the outflow rate is the most important flow accounting for fluctuations in the unemployment rate. But it does not account for all the variation and contrary to recent results in the literature, I find that it is important to allow for flows in and out of the labour force in order to get a more complete characterization of unemployment fluctuations in Denmark. Furthermore, I find that outflow fluctuations are not due to compositional changes in the pool of unemployed. I also find that the use of less informative survey data, often used in the previous literature, implies that the importance of the inflow rate is overstated in the Danish context. Finally, I find that the inflow rate is relatively more important for younger workers and for male workers.
References


Dansk resume (Danish Summary)


Det første kapitel er skrevet sammen med mine vejledere, Michael Svarer og Michael Rosholm, og det har titlen ”A Danish Profiling System” (et dansk profilerings system). I dette papir udvikler vi en metode til at identificere arbejdsløse, der er i risiko for at blive langtidsarbejdsløse. Baggrunden for dette er, at det ikke er alle arbejdere, der bliver arbejdsløse, som har brug for hjælp til at finde et nyt job. En stor del af den samlede danske arbejdsløshed udgøres faktisk af arbejdere som kun er arbejdsløse i en meget kort periode, inden de finder videre til et nyt job, og det vil derfor være ineffektivt at lade disse personer deltage i omkostningsfulde aktiveringsprogrammer. Men andre arbejdsløse forbliver arbejdsløse i længere tid, fordi de enten ikke kan, eller vil, finde et nyt arbejde. Det synes naturligt at fokuse den aktive arbejdsmarkedspolitik på denne gruppe af arbejdsløse. Et muligt problem er dog, at det ikke nødvendigvis er optimalt først at lade arbejdsløse deltage i
aktivering, når de har været arbejdsløse i længere tid. Det er sandsynligvis bedre med det samme at identificere de arbejdsløse, der er i risiko for arbejdsløshed, og så lade dem deltage i aktivering inden de oplever de potentielle negative effekter af at være arbejdsløse i længere tid. Denne tilgang anbefales bl.a. i OECD (1998). Modeller der forsøger at identificere arbejdsløse, der er i risiko for at blive langtidsarbejdsløse, kaldes i litteraturen for Profiling modeller og i dette papir bruger vi en varighedsmodel til at estimere sandsynligheden for at blive langtidsarbejdsløs baseret på observerbare karakteristika for den enkelte person. Sagsbehandlerene kan så blive informeret om denne sandsynlighed og bruge denne information, når de skal vælge om en arbejdsløs skal aktiveres eller ej. Vores resultater viser, at modellen er forholdsvis god til at diskriminere mellem kort- og langtidsarbejdsløse. Men forudsigelser er lang fra perfekte, hvilket dog er forventeligt, når vi forsøger at forudsige et meget komplekst udfald baseret på relativt sparsomme informationer.

Kapitel 2 er ligeledes skrevet sammen med mine vejledere, Michael Svarer og Michael Rosholm, og det har titlen ”Choosing the Best Training Programme: Is there a Case for Statistical Treatment Rules?” (Om at vælge det bedste aktiveringsprogram: kan det hjælpe at bruge statistisk baserede udvælgelsesregler?). Papiret er en naturlig udvidelse af det første kapitel, hvor vi spurgte hvem der skulle deltage i aktivering, da vi i dette papir spørger hvilken type aktivering den enkelte arbejdsløse skal deltager i? Sidstnævnte spørgsmål er rettet mod at forbedre effektiviteten af aktivering, mens spørgsmålet der blev stillet i kapitel 1 kan ses som et forsøg på at forbedre ligheden, ved at lade de dårligst stillede deltage i aktivering. Hvis vi ser bort fra fordelingsmæssige hensyn over gruppen af arbejdsløse, så er det mest interessant at kigge på hvordan arbejdsløse bliver udvalgt til aktivering, når vi faktisk observerer heterogene aktiveringseffekter. I dette papir forsøger vi at udnytte en sådan information om heterogene effekter. Modellen vi anvender, er en varighedsmodel, som bruger ”timing-of-events” metoden til at identificere kausale effekter. Vi sammenligner forskellige metoder til at udvælge arbejdsløse til aktivering og resultaterne tyder på, at det er muligt at opnå en betydelig afkortelse af den gennemsnitlige arbejdsløshedszvarighed, hvis man introducerer en statistisk metode til at udvælge arbejdsløse til aktivering. Disse resultater underbygger dermed tidligere resultater i Lechner & Smith (2007), hvor de finder, at den
nuværende sagsbehandler-baserede udvælgelse til aktivering ikke er bedre end en tilfældig udvælgelse.

Kapitel 3 er skrevet sammen med Jeff Smith og har titlen “Using Statistical Treatment Rules for Assignment of Participants in Labor Market Programs” (Om brug af statistisk baserede behandlingsregler til at udvælge deltagere til aktivering). Profiling og targeting modeller, analyseret i de forrige kapitler, kan mere generelt ses som eksempler på statistisk baserede behandlingsregler og i dette papir analyserer vi litteraturen om brug af sådanne regler til at udvælge deltage i aktivering. Vi diskuterer først de teoretiske fordele og ulemper ved statistisk baserede behandlingsregler sammenlignet med alternative metoder til at udvælge deltagere. Dernæst diskuterer vi nogle økonometrisk problemstillinger omkring estimation og evaluering af statistisk baserede behandlingsregler, som den hidtidige litteratur ikke har lagt vægt på. Og endelig analyserer vi de empiriske resultater i litteraturen. Vi gør opmærksom på, at de tilgængelige studier sjældent giver mulighed for en faktisk evaluering af de statistisk baserede behandlingsregler, først og fremmest fordi der ofte mangler information om hvor godt alternative udvælgelsesmetoder klarer sig. Resultaterne tyder dog på, at profiling modeller kan bruges til at øge lighed, mens det er bedre at bruge targeting modeller, hvis målet er øget effektivitet. De positive resultater for targeting modeller er dog baseret på simuleringsstudier, så der er brug for mere viden om faktisk anvendte targeting modeller, hvis vi vil være sikre på at disse modeller også i praksis kan fungere så godt som både teoretiske studier og simuleringsstudier tyder på.

har forsøgt at besvare dette klassiske spørgsmål, og hovedbidraget fra mit papir er at vise resultater, der er baseret på en ny og mere informativ type af data. Jeg bruger forskellige dekompositionsmetoder, der er blevet foreslået i den seneste litteratur, og finder konsekvent, at raten hvormed arbejdsløse strømmer ud af arbejdsløshed er den vigtigste, når man vil forklare svingninger i arbejdsløshed. Men variation i denne rate forklarer ikke alle svingningerne i arbejdsløshed og i modsætning til den hidtidige litteratur, så finder jeg at det er vigtigt tage højde for strømme ind og ud af arbejdsstyrken, hvis man vil have en mere komplet beskrivelse af svingningerne i arbejdsløshed i Danmark. Jeg finder desuden, at svingningerne i raten hvormed arbejdsløse forlader arbejdsløshed, ikke kan forklares ud fra kompositoriske forandringer i gruppen af arbejdsløse. Derudover finder jeg, at hvis man bruger en type af aggregerede data, som ofte bruges i den eksisterende litteratur, så tildeler man for stor betydning til raten hvormed arbejdsløse strømmer ind i arbejdsløshed, i forhold til resultater baseret på de mere informative resultater, der er baseret direkte på danske register data. Og endeligt finder jeg, at raten hvormed arbejdsløse strømmer ind i arbejdsløshed, er relativt mere betydnende for unge arbejdere og for mandlige arbejdere.
Referencer


CHAPTER 1

________________________________________________________

A Danish Profiling System

________________________________________________________
A Danish Profiling System*

Michael Rosholm† Jonas Staghøj‡ and Michael Svarer§


March 2006

Abstract

This paper describes the statistical model used for profiling new unemployed workers in Denmark. When a worker - during his or her first six months in unemployment - enters the employment office for the first time, this model predicts whether or not he or she will be unemployed for more than six months from that date. The case workers’ assessment of how to treat the person is partially based upon this prediction.

Keywords: Unemployment duration, profiling

JEL-Codes: J64, J68

---

*We are grateful to Annette B. Andersen for reading the manuscript and to Klaus Langager from the Danish National Labour Market Authority for access to the data used in the paper. Michael Svarer thanks the Danish National Research Foundation for support through its grant to CAM. Michael Rosholm thanks the Danish Ministry of Social affairs for financial support through the Danish Graduate School for Integration, Production and Welfare.

†Department of Economics, Aarhus School of Business, Aarhus University, Denmark, E-mail: rom@asb.dk
‡School of Economics and Management, Aarhus University, Denmark, E-mail: jstaghoej@econ.au.dk
§School of Economics and Management, Aarhus University, Denmark, E-mail: msvarer@econ.au.dk
1 Introduction

“Across OECD countries, millions of unemployed have been out of work for more than a year. And others are at risk of becoming so. One possible way to combat the drift into long-term unemployment is to offer more assistance to job losers before they reach the stage of long-term unemployment. But it would be very costly to offer in-depth help to all of the job losers. This has led some countries to develop methods to both identify jobseekers at risk of becoming long-term unemployed and refer them to suitable labour market programmes, usually known as profiling. But is it possible to accurately identify such jobseekers?” (OECD, 1998).

In this paper we present the first statistical component of such a profiling model, which as of December 1, 2004, has become an integrated part of the Danish national labour market policy after extensive experimentation with the statistical model and with the way the information is presented to caseworkers. When discussing how to develop and implement a profiling system, a first requirement is to define which goals the system should be designed to fulfill, because this obviously influences the way the system should be designed. Two natural goals are equity and efficiency and hence it would be interesting to investigate whether these goals are correlated or in fact conflicting goals, but we will not go into a deep discussion of this issue in the present paper, although we will propose some arguments indicating that both goals may be fulfilled in the present project. After properly defining the goals the next task is to choose a profiling variable based upon which the group of unemployed is divided into different categories. A measure of the probability of becoming long-term unemployed is chosen and the main focus of this paper is on the estimation of this probability and on the predictive power of the probability when using it in the profiling system.1

The main purpose of the paper is to verify whether the profiling system is capable of identifying unemployed workers who are at risk of ending up in long-term unemployment (LTU, henceforth). The profiling model consists of a statistical model (presented here) to be used as an initial screening device for identifying potentially long-term unemployed workers, combined with in-depth interviews by caseworkers with those asserted to have a high risk of LTU. The intention is to extend the profiling model by a statistical model and additional

1See Black et al. (2000) for a discussion of how to design and evaluate profiling systems.
interviews designed to identify the 'best' strategy and optimal timing for helping a given unemployed person at risk of LTU in order to reduce the risk of individual LTU\(^2\).

The statistical component of the profiling system consists of a duration model for the time spent in unemployment. The model is estimated on 120 subgroups, stratified according to age, gender, benefit eligibility, and region of residence. The data used for estimation is the entire inflow into unemployment in Denmark during the period January 1999 - June 2003. Based on the estimated models, it is possible to calculate the probability that a worker attending a meeting with a caseworker at the employment agency will still be unemployed six months from that date, conditional on the elapsed duration of unemployment. A set of threshold values are then calculated in order to maximize the number of correct predictions of the model, and the caseworkers are presented with information about whether the calculated probability is far above, far below, or close to the threshold value.

In several countries attempts have been made to specify worker profiling models. Fröligh \textit{et al.} (2003) state that profiling models are currently used or being tested in Australia, Finland, France, Germany, Ireland, South Korea, New Zealand, Sweden, and the United States. The predictive power of the various models is mixed and to some extent discouraging. Nevertheless, worker profiling is now used in certain states in the U.S. and in South Korea and is, as mentioned above, used on a large scale in Denmark.

Our model is readily comparable to the New Zealand worker profiling model (Watson \textit{et al.}, 1997). The New Zealand model was in active use for some time, but has since been abandoned, allegedly due to a new government that wanted to shift attention from active policies towards incentive-based policies (benefit cuts). Compared to the New Zealand Worker Profiling model, the Danish model provides a substantial improvement in predictive power.

The paper is organized in the following way: Section 2 offers a brief overview of Danish labour market institutions and a recently implemented labour market reform of which the profiling model is one component. Section 3 presents the data used in the estimation process. Section 4 contains a description of the statistical model, and Section 5 shows selective results and evaluates the predictive power of the model. Finally, Section 6 discusses policy issues \footnote{See Fröligh \textit{et al.} (2003) for a discussion of statistically assisted programme selection.}
and offers a few conclusions.

2 The Institutional Framework and The Labour Market Reform

Denmark has a two-tiered system for unemployed workers. Most workers in Denmark - around 80% - are members of an unemployment insurance fund. These individuals have, upon the fulfillment of a few conditions, the right to receive unemployment insurance (UI) benefits, which correspond to 90% of the previous wage with an upper limit of approximately 1,800 Euro per month. UI benefit payments are heavily subsidized by the state, which finances around 80% of total payments. This system is administered by the Central Labour Market Authority (Arbejdsmarkedsstyrelen), which is a unit operating under the Ministry of Employment.

Unemployed workers without UI benefit eligibility may instead receive social assistance (SA) benefits. While non-insured workers make up only around 20% of the workforce, they make up a larger fraction of the unemployed, as the group typically consists of workers with a low attachment to the labour market. Hence, they are more often unemployed, and on average they are unemployed for longer periods. Social assistance benefits are means tested, but typically the amount is below the UI benefit level. Social assistance is administrated by the municipal authorities. There are 279 municipalities in Denmark. Needless to say, they are all subject to the same rules and regulations, but the administration differs considerably between municipalities, and recent research has shown that the differences in efficiency between municipalities in bringing SA recipients back to work cannot be explained solely by individual and municipality-specific variables. In other words, the causes of the differences in efficiency of the local labour market policy are unknown.

Up until 2002, the rules and regulations regarding contacts with caseworkers, participation in labour market programs etc. differed between the two systems. With the labour market reform of 2002, of which the profiling system is one component, the aims are to eventually have identical rules and regulations in the two systems, and in fact to merge

---

3 Arendt et al. (2004).
the system in the sense that the two-tiered system should become one system. There will still be UI benefits and SA benefits, but the rules regarding meetings, job search etc. will eventually be the same. The goal of the system of reforms is to reduce the emerging public finance problem triggered by an ageing population by increasing the labour force by some 90,000 individuals by 2010. The development of a common profiling model for assessing the employability of unemployed workers marks a step towards a single-tier system.

2.1 Profiling

The profiling model is designed to increase equity and efficiency in the labour market.

If equity is defined in terms of some utility measure, the profiling model will help achieving the goal of increasing equity if it is facilitates helping the unemployed back into work and if the unemployed are in fact among the people in the society who are worst off measured on the utility scale. If instead equity is defined as equal opportunities, the profiling model should be particularly able to increase equity, as it introduces an effective way to provide equal treatment of unemployed workers across different municipalities.

The profiling model will achieve the goal of increasing efficiency if it allocates the labour market programs to the unemployed workers who have the largest (expected) effects of participating in the programs. However, it is not clear that these workers are the same as those who become long-term unemployed, and hence efficiency in this sense also requires a model for targeting active policies. However, efficiency in terms of identifying those at risk of LTU and leaving alone those who are perfectly able to find jobs themselves (thus avoiding deadweight losses) is an important aim of the profiling model; each year a very large number of workers experience a short period of unemployment, and this leads to at least two important reasons for identifying as quickly as possible those who are at risk of LTU. First, early identification of individuals at risk of LTU allows preventive policies to be implemented during the early stages of unemployment. Second, as mentioned above, early identification is necessary in order to avoid treating persons who are perfectly able to find jobs on their own.

The profiling model consists of several components. First, there is a ‘job barometer’,
which is a graphical representation of the predictions based upon the statistical profiling model presented in this paper. This is used by the caseworker to assess employability before the first meeting with a newly unemployed person. Next, there is a public assistance record, which gives the caseworker an overview of the person’s previous periods on public assistance.\(^5\) Third, there is a dialogue guide for the caseworkers’ communications with clients designed to identify strengths and weaknesses in relation to the labour market. Finally, the unemployed person has to prepare some personal information before the first meeting. This should make it easier across employment offices to treat similar persons alike, and eventually to conduct labour market policies as efficiently as possible.

The aim of the profiling system is to assess the employability of newly unemployed workers. This will be done by eventually placing each individual in one of five categories, ranging from fully employable to (at present) fully unemployable. The statistical profiling tool basically calculates a probability that an individual with certain characteristics - including the labour market history for the past five years - will still be unemployed in six months time, conditional on the elapsed duration of unemployment, which, at the date of the meeting, can be anything from 4 to 30 weeks. This information is presented to the caseworker in the job-barometer, which is shown in Figure 1.

\[\text{INSERT FIGURE 1 HERE}\]

The area to the left indicates 'high risk of LTU’, the intermediate area indicates 'medium risk of LTU’, and the area to the right indicates 'low risk of LTU’. Which area is highlighted depends on the way the individual probability deviates from a population mean. The empirical foundation for these probabilities is described in the following sections.

### 3 The Empirical Model

When faced with the challenge of constructing the statistical tool of a worker profiling model, one must choose an appropriate econometric/statistical model. We first note that the problem of early identification of potentially long-term unemployed workers is related to

\(^5\)This information is also used in the statistical profiling model.
similar economic problems in the finance literature.\textsuperscript{6} Early detection of financially distressed firms with high risk of bankruptcy, or ways to correctly classify profitable investments has obviously received considerable attention for many years. This has led to the estimation of profiling models by many different techniques including multivariate discriminant analysis, probit and logit models, and more sophisticated nonlinear models such as neural networks models. However, the overall pattern appears to be that the differences between statistical models in terms of predictive accuracy is relatively unimportant.\textsuperscript{7}

In the U.S., the original Worker Profiling and Reemployment Services (WPRS) model applied a discrete choice model, where the dependent variable was UI benefit exhaustion. Recently, Black \textit{et al.} (2003) have criticized the WPRS model on that choice. Their main concern is that by using a dichotomous dependent variable all data variation among individuals who do not exhaust their UI benefits is ignored. Instead they suggest in the Kentucky Profiling Model (henceforth KPM), that a continuous dependent variable is employed. Specifically, they suggest as dependent variable the ratio of benefits drawn to benefit entitlement (i.e., fraction of benefits claimed). In an earlier paper (using the same data and the same dependent variable) several different statistical models are compared (Berger \textit{et al.}, 2000). Among ordinary least squares (OLS), Cox proportional hazard models, and tobit models, the difference in predictive power between the models is very low, and they suggest, for simplicity, using the OLS model.

The dependent variable of interest in the Danish context is really an indicator of whether a given individual, conditional on the elapsed unemployment duration, is still unemployed after an additional 6 months. Hence, for each value of the elapsed duration of unemployment, we could create a dichotomous variable taking the value 1 if the individual ‘survives’ 26 additional weeks in unemployment, and 0 otherwise. We could then estimate a probit or logit model for each value of the elapsed duration of unemployment and use the estimated parameters of those models for predictive purposes. However, such a strategy is also vulnerable to Black \textit{et al.}’s (2003) criticism, since we do not fully exploit all information in

\textsuperscript{6}Prediction and profiling problems are also seen in various other kinds of literatures, like criminology (Auerhahn, 1999), insurance (Yeo \textit{et al.}, 2001), marketing (Shaw \textit{et al.}, 2001) and medicine (Khan \textit{et al.}, 2001).

\textsuperscript{7}See Altman (1968) for an early discussion and O’Leary (1998) for a more recent discussion of different models.
the data to reduce the uncertainty of the parameter estimates. We have therefore chosen, instead, to estimate the duration of unemployment and subsequently use the parameters estimated in the duration model to calculate the probability of 'survival' for 26 additional weeks, conditional on the elapsed duration. Hence, our dependent variable of interest is the duration of unemployment, and the econometric/statistical models to be employed are duration models.

Let the continuous stochastic variable $T$, $T \in (0, \infty)$ denote unemployment duration. The hazard rate, which denotes the probability that an individual with observed characteristics $x$ finds a job in the interval $t + dt$ given that the individual is still unemployed at time $t$, is then given by

$$
\begin{align*}
    h(t|x_t) &= \lim_{dt \to 0} \frac{\mathbb{P}(t < T \leq t + dt|T > t, x_t)}{dt} \\
    &= \frac{f(t|\{x_s\}_{0}^{t})}{S(t|\{x_s\}_{0}^{t})},
\end{align*}
$$

where $f(t|\{x_s\}_{0}^{t})$ is the density function, $S(t|\{x_s\}_{0}^{t})$ is the survivor function, and $\{x_s\}_{0}^{t}$ denotes the entire path of the explanatory variables from the start of the unemployment spell until time $t$. The survivor function denotes the probability that an individual is unemployed more than $t$ weeks. The association between the hazard function and the survivor function can also be expressed as

$$
S(t|\{x_s\}_{0}^{t}) = \exp(- \int_{0}^{t} h(s|x_s)ds).
$$

The objective of the profiling model is to calculate the probability of remaining in unemployment for more than 26 additional weeks conditional on the elapsed unemployment duration being between $4 - 30$ weeks. Suppressing the dependency on $x$, this conditional probability can be written as

$$
\begin{align*}
    \Pr(T > \tau + 26|T > \tau) &= \frac{S(\tau + 26)}{S(\tau)} \\
    &= \frac{\exp(- \int_{0}^{\tau + 26} h(s)ds)}{\exp(- \int_{0}^{\tau} h(s)ds)} \\
    &= \exp(- \int_{\tau}^{\tau + 26} h(s)ds),
\end{align*}
$$

where $\tau$ denotes the elapsed duration of the unemployment spell. In practice, as mentioned
above, \(4 < \tau \leq 30\) for individuals in the UI system, since the first interview conducted by the Public Employment Service (PES, henceforth) takes place after 1 month of unemployment. Thus, in the estimations we consider a population that has survived 4 weeks of unemployment. In order to calculate (3) as accurately as possible, we restrict attention to the first 52 weeks of the unemployment spell, that is \(T \in [4, 56]\). Consequently, all unemployment spells longer than 56 weeks are censored at a duration of 56 weeks. For individuals in the SA system, the first interview may take place from the first day of entry, hence for this system, we will have \(0 < \tau \leq 26\), and accordingly we can censor all durations in this system at 52 weeks.

The hazard function is specified as a proportional hazard model. That is, the hazard is the product of the baseline hazard, which captures the time dependence, and a function of observed time-varying characteristics, \(x_t\)

\[
h(t|x_t) = \lambda(t) \cdot \varphi(x_t),
\]

where \(\lambda(t)\) is the baseline hazard, and \(\varphi(x_t)\) is the scaling function specified as \(\exp(x_t\beta)\). The baseline hazard is specified as a piecewise constant baseline hazard with splitting times \(\tau_0 = 4, \tau_1 = 5, \tau_2 = 6, \ldots, \tau_{52} = 56\) for individuals in the UI system, that is, there is a separate baseline component for each week. The baseline is defined similarly for the models for the SA system, with \(\tau_0 = 0, \tau_1 = 1, \tau_2 = 2, \ldots, \tau_{52} = 52\). The value of the baseline hazard in the \(k^{th}\) interval is denoted \(\lambda_k\).

In the scaling function \(\exp(x_t\beta)\), the explanatory variables are allowed to be time-varying, as noted above. Let \(d\) denote the censoring indicator, which takes the value 1 if the observation is shorter than 56 weeks and uncensored, and zero otherwise.

Let \(\theta\) denote the parameters of the model. To obtain estimates of the parameters, we perform maximum likelihood estimation based on the following (conditional) log-likelihood function (see Lancaster, 1990, for details on duration models).

\[
\log l(\theta) = \sum_{i=1}^{N} \left[ d_i \ln(h(t_i|x_{i,t})) - \int_{4}^{t_i} h(s_i|x_{i,s})ds \right],
\]

where \(N\) denotes sample size. This log-likelihood function is for the models for the UI system.
For the SA system, it looks similar, except the lower bound for the integral is 0 instead of 4.

Based on the estimated parameters, the probability that an individual who has been unemployed for $\tau$ weeks will experience 26 additional weeks of unemployment is easily calculated as

$$\hat{\Pr}(T > \tau + 26|T > \tau, x) = \exp \left( - \exp(x\beta) \sum_{k=\tau}^{\tau+26} \hat{\lambda}_k \right)$$

assuming that the $x$ does not change.\(^8\)

### 3.1 Unobserved Heterogeneity

Duration models would typically also include a component intended to capture unobserved heterogeneity. In duration models, it is well known that the baseline hazard is biased towards negative duration dependence if neglected unobserved heterogeneity is present. Moreover, the remaining parameter estimates will be biased too since the model is non-linear.

However, for the present model, the objective is not consistent estimation, but predictive ability. Neglect of unobserved heterogeneity implies that the baseline hazard and the other model parameters will be affected by unobserved heterogeneity. So, for example, if we know that an individual has survived some weeks in unemployment we also know that his or her unobserved characteristics are not that favourable. However, these characteristics are not observed, but their effect is reflected in the baseline hazard. So, for predictive purposes, this seems the best way to exploit all information. If we were to include unobserved heterogeneity, we would be forced to evaluate everyone at the mean (or some other arbitrarily chosen point) of the unobserved variable, and then knowing that the person survived half a year in unemployment is not allowed to influence the evaluation of the hazard.\(^9\) For this reason, the model does not correct for unobserved heterogeneity.

We should also note that the issue of unobserved heterogeneity could influence the degree of success for the profiling model. When evaluating the profiling model and comparing the model to the existing system, where allocation of unemployed into labour market programs

---

\(^8\)Since the purpose of the model is to predict whether an individual survives an additional 26 weeks in unemployment, we cannot use time-varying variables in the predictions; the path of the $x$’s is not known in advance. Hence, we make the simplifying assumption that the current value will prevail.

\(^9\)Of course, one could also calculate the distribution of unobservables conditional on the elapsed duration of unemployment, from that infer the mean of the unobserved variable given the elapsed duration and use that number for the predictions. Our approach is a shortcut.
is, to a large extent, subjectively chosen by the caseworker, we know that the caseworker can observe more detailed information about the individual. Some of the reasons for the unobserved heterogeneity are really data issues and we will discuss these in the section 'explanatory variables', while others may be more interesting; motivation, for example, is not observed by the econometrician but may be partially observed by the caseworker. If it is very important to observe motivation in order to allocate the unemployed into the optimal labour market program then the caseworker may be able to do a better job than the statistical model. Some experiments from Switzerland indicate, however, that caseworkers have difficulties in predicting treatment effects and so it seems there is indeed room for additional improvements.10

4 Data

The analysis here uses data from administrative registers from the Danish Labour Market Authority. This is the same data that the employment offices use and therefore the same information on which predictions have to be made. The advantage of the data set is that it is updated with a very short time lag. The disadvantage is that it basically only contains labour market data. Ideally, we would have liked to use more information by merging to other administrative registers, but since the aim of the analyses is to maximize predictive power based on the available information, we use only what is readily available. The register we use is called DREAM (Danish Register for Evaluation Of Marginalisation), and it is basically an event history file, that includes weekly information on each individual’s receipt of public transfer incomes, unemployment registrations, and participation in active labour market programs. Based on this information, a weekly event history is constructed, where the individual each week either occupies one of a number of public transfer states or is not receiving public transfers. When an individual is not registered as receiving public transfers, the person can either be employed or be outside the labour force without receiving transfer income. In the Danish welfare state, the latter is very unlikely; hence the assumption that not receiving public transfers in a given week corresponds to employment is innocuous. From DREAM, we sample the inflow to unemployment in both the UI and the SA system in the

10See Lechner & Smith (2005).
period January 1999 to June 2003. All exits from unemployment to states other than (what we assume to be) employment are treated as independently right censored observations.

For persons in the UI system, we exclude all unemployment durations shorter than four weeks, because the first meeting will never take place during the first four weeks.\footnote{This truncation from the left also implies that we eliminate a substantial amount of temporary layoff spells. Temporary layoff is very common in the Danish labour market since employers only pay UI benefits for the first two days of an unemployment period. Approximately 40\% of all unemployment spells in Denmark are temporary layoffs. They are, however, typically quite short and therefore only constitute around 16\% of total unemployment (Jensen & Svarer, 2003). Moreover, approximately 90\% of them are 4 weeks or shorter.} Moreover, all unemployment durations longer than 56 weeks in the UI system and longer than 52 weeks in the SA system are censored at these durations, because that is all the information we are going to use in the estimation process. As the profiling model is further developed, it will eventually be extended such that it can make predictions for persons with an elapsed unemployment duration larger than 26 (or 30) weeks, but to directly extend the current model implies an assumption that the effect of covariates does not change over unemployment duration. Several studies have shown that this assumption is not realistic for longer durations, hence the intention is to estimate new models for elapsed durations above 52 (or 56) weeks, thus essentially allowing for time-varying parameters of the models.

4.1 Sample selection and subsampling

Denmark is divided into 14 counties, plus a ‘region’ consisting of Copenhagen and Frederiksberg municipalities, each with different labour markets and different local labour market conditions. We follow that division in our estimations below and split the data by region of residence. This leads to 15 sub samples.

Moreover, as mentioned above, there are two parallel labour market systems, one for workers insured against unemployment (the UI system), and hence eligible for UI benefits, and one for the non-insured (the SA system). Hence, in each region, data is split according to the labour market system to which each worker belongs.

In each system, different rules apply to different age groups. In the UI system, the Youth Unemployment program applies to workers aged below 25. The data is therefore split into two groups: those under 25, and 25 and older. For workers in the SA system, more active policies are pursued for those aged below 30 than for those above. For workers in the SA
system, the data is split into those under 30, and those who are 30 or older. The sample is truncated from below to persons aged 16 or older. In addition, due to a mandatory retirement age of 65, all samples are restricted to those aged 64 or below.

Finally, previous investigations show that male and female workers have very different behaviour in unemployment, so the data is also divided by gender. In total, we thus end up with 15 (regions) \times 2 (systems) \times 2 (age groups) \times 2 (gender) data sets, that is, 120 sub samples of the inflow into unemployment during the period January 1999 - June 2003. The duration model specified above is estimated separately for each of these 120 sub samples.

The dependent variable of the study is the duration of unemployment. In the UI system, the dependent variable is the duration of unemployment given that it is at least four weeks. After these sub sample definitions and the reduction in the samples, we end up with a total of almost 2 million unemployment spells that are used in the estimations.

### 4.2 Explanatory Variables

Since the purpose of this exercise is to make sound predictions, we use the 'kitchen sink' approach to determine which explanatory variables to include in the model. However, since the data are obtained directly from the Danish Labour Market Authorities, we only have access to the variables that are available in their databases. The implication is that the information that is usually available when working with Statistics Denmark’s register based data is not generally available to us. For example, measures like education, previous wage, and working experience are not in this data set.\textsuperscript{12} The information available is the following:

**Age:** The individual’s age is known, and it is used to construct a set of dummies for age group. In the samples of young individuals (aged below 25 or 30), there is a dummy for each age from 16 – 29 or 16 – 24 (with 29 or 24 being the reference category), and for the samples of 'older' individuals, we construct dummies for belonging to 5-year age intervals.

**Year:** We have included a set of indicators measuring the year in which the unemployment spell begins. As we are only looking at short spells, censoring all spells at 56 weeks, it is not important to take into account time-varying calendar time effects during an unemployment spell.

\textsuperscript{12}The intention is to increase the information available in the register, so that the caseworker also has this information and so that we can base predictions on it. It is not yet available, however.
**Municipality:** We have a set of indicators - a different set for each county - for the municipality of residence of the unemployed person.

**Local unemployment rate:** The municipal unemployment rate is included to allow for cyclical effects and thereby improve the predictive power of the model when a new year is entered without the model being updated. This variable is identified because it can vary over time and between municipalities. Hence, it is not perfectly correlated with a linear combination of annual dummies and municipal dummies.

**Unmarried:** This measures whether an individual is unmarried and does not co-habit either.

**Sick:** Indicates that an individual is currently reported sick (receiving sick pay) while unemployed. It is thus a time-varying variable.

**Immigrant:** We have four indicators for whether the individual is first or second generation immigrant from more or less developed countries. The reference category is native Danes.

**UI-fund:** We have a set of indicators for unemployment insurance fund membership. There are several UI-funds in Denmark, and membership is often categorized according to education/skills and/or by industry. We have 36 different UI funds, and we have an indicator for each. These funds may be seen as broad proxies for the missing information concerning education and skills. This set of variables is (naturally) only used in the samples of workers insured against unemployment.

**Maternity Leave and Holiday Pay:** We know whether an individual has been on maternity leave and whether an individual has received holiday pay while unemployed; in employment individuals accumulate rights to holiday payments. If the individual, due to unemployment in the previous year, has not accumulated a sufficient amount of money it is possible to receive money from the state for holiday. Individuals currently employed will count as unemployed in the period they receive vacation pay, therefore we take that into account in the model. These variables are thus time-varying.

**Active Labour Market Policies:** We have a set of time-varying variables indicating whether the individual is currently in a labour market program, and whether the individual has completed a labour market program during the past 26 weeks. This information is
naturally time-varying.

**Labour market history:** The most important information, however, for our predictive purposes, is the history on past labour market performance. We know, for each of the five years preceding the current unemployment spell, the fraction of the year spent on some kind of income transfer (UI, SA, temporary leave schemes including parental leave, or other public transfer schemes). We have the same information for sickness periods as well. So we have constructed 10 variables, five for public transfers and five for sickness, measuring the fraction of each of the past five years spent in either sickness or on a transfer scheme. Moreover, we use the number of unemployment spells the individual has had over the same period. For the ‘young samples’, we only use the information for the past two years. If the information is missing (because the individual was ‘too’ young), these variables are set at zero.

## 5 Empirical findings

This section contains a short description of the estimated parameters in the duration model. The main focus of the paper, however, is on the predictive power of the model, so this will be brief, and as already noted the estimated parameters may be biased due to unobserved heterogeneity. The entire set of estimation results is available on request. To get an idea of the results, Table 1 presents the effects of various explanatory variables for insured unemployed men above 25 from Aarhus county, i.e., these are the results for one out of 120 subgroups. Note that we only present a subset of the coefficients, as the UI fund indicators (36), the municipality indicators (279 in total), and indicators for participation in and completion of active labour market programs are not shown.

### 5.1 Baseline hazard function

In the empirical model we have modelled the baseline hazard as a piecewise constant function. This very flexible specification is attractive when the baseline hazard exhibits non-monotone behaviour. Watson et al. (1997) impose a Weibull distribution on the baseline hazard.

In Figure 2 we show the baseline hazard for insured men over 25 years old in Aarhus county.
We find that the baseline hazard generally exhibits negative duration dependence. This holds for all 120 sub samples. However, it may to some extent just reflect neglected unobserved heterogeneity, see the discussion in section 3.1. The peaks in the baseline every 4 – 5 weeks probably reflect that most jobs begin and end at the start of a month. The hazard rates are generally much lower for persons on social assistance than for persons receiving unemployment insurance benefits.

5.2 Effects of explanatory variables

The effects of some of the explanatory variables differ across the different subgroups, but some consistent patterns arise. The Danish economy has, like most of the western world, experienced an economic downturn after the IT-bubble burst and the September 11 terror attacks in NY. This is captured by the year dummies. They show that compared to the reference year 1999 the hazard rate out of unemployment has been lower in subsequent years.
Table 1: Hazard model for men, insured, above 25 from Aarhus county

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999 (reference year)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>-0.033</td>
<td>0.016</td>
</tr>
<tr>
<td>2001</td>
<td>-0.036</td>
<td>0.018</td>
</tr>
<tr>
<td>2002</td>
<td>-0.218</td>
<td>0.018</td>
</tr>
<tr>
<td>2003</td>
<td>-0.231</td>
<td>0.029</td>
</tr>
<tr>
<td>Age 26-29 (reference age group)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 30-34</td>
<td>-0.025</td>
<td>0.017</td>
</tr>
<tr>
<td>Age 35-39</td>
<td>-0.097</td>
<td>0.018</td>
</tr>
<tr>
<td>Age 40-44</td>
<td>-0.132</td>
<td>0.019</td>
</tr>
<tr>
<td>Age 45-49</td>
<td>-0.217</td>
<td>0.020</td>
</tr>
<tr>
<td>Age 50-54</td>
<td>-0.346</td>
<td>0.020</td>
</tr>
<tr>
<td>Age 55-59</td>
<td>-0.624</td>
<td>0.022</td>
</tr>
<tr>
<td>Age 60-64</td>
<td>-0.759</td>
<td>0.037</td>
</tr>
<tr>
<td>Temporarily on Holiday Pay</td>
<td>1.292</td>
<td>0.042</td>
</tr>
<tr>
<td>Temporarily on Paternity leave</td>
<td>0.369</td>
<td>0.114</td>
</tr>
<tr>
<td>Temporarily on Sickness benefits</td>
<td>-0.424</td>
<td>0.033</td>
</tr>
<tr>
<td>Single</td>
<td>-0.171</td>
<td>0.011</td>
</tr>
<tr>
<td>1. generation immigrant from developed country</td>
<td>-0.176</td>
<td>0.029</td>
</tr>
<tr>
<td>1. generation immigrant from less developed country</td>
<td>-0.356</td>
<td>0.030</td>
</tr>
<tr>
<td>2. generation immigrant from developed country</td>
<td>-0.095</td>
<td>0.109</td>
</tr>
<tr>
<td>2. generation immigrant from less developed country</td>
<td>-0.401</td>
<td>0.134</td>
</tr>
<tr>
<td>Sickness benefit rate 1 year ago</td>
<td>-0.289</td>
<td>0.104</td>
</tr>
<tr>
<td>Sickness benefit rate 2 years ago</td>
<td>0.022</td>
<td>0.225</td>
</tr>
<tr>
<td>Sickness benefit rate 3 years ago</td>
<td>-0.013</td>
<td>0.353</td>
</tr>
<tr>
<td>Sickness benefit rate 4 years ago</td>
<td>-0.140</td>
<td>0.359</td>
</tr>
<tr>
<td>Sickness benefit rate 5 years ago</td>
<td>-0.197</td>
<td>0.241</td>
</tr>
<tr>
<td>Public transfers rate 1 year ago</td>
<td>-0.088</td>
<td>0.043</td>
</tr>
<tr>
<td>Public transfers rate 2 years ago</td>
<td>-0.159</td>
<td>0.092</td>
</tr>
<tr>
<td>Public transfers rate 3 years ago</td>
<td>0.019</td>
<td>0.137</td>
</tr>
<tr>
<td>Public transfers rate 4 years ago</td>
<td>0.017</td>
<td>0.168</td>
</tr>
<tr>
<td>Public transfers rate 5 years ago</td>
<td>-0.086</td>
<td>0.106</td>
</tr>
<tr>
<td>Number of unemployment spells last year</td>
<td>0.054</td>
<td>0.010</td>
</tr>
<tr>
<td>Number of unemployment spells two years ago</td>
<td>0.087</td>
<td>0.006</td>
</tr>
<tr>
<td>Local unemployment rate</td>
<td>-0.503</td>
<td>0.135</td>
</tr>
<tr>
<td>36 UI Fund Membership Indicators</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>26 Municipality of Residence Indicators</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>ALMP Participation and Completion Indicators</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Note: Bold figures indicate that the parameter is different from 0 at 5% level. In the regression we also corrected for municipality effects, for UI-fund membership, and for participation in ALMPs.

As it is witnessed in several studies of unemployment duration, the hazard rate out of unemployment decreases with age. This is also the case in our models. Not surprisingly, men who are out of work due to holiday or paternity leave leave unemployment faster. In addition single men are less likely to leave unemployment compared to their married or cohabiting counterparts. This result is consistent with previous investigations of unemployment duration. Being a first or second generation immigrant from less developed countries is associated with lower hazard rates and therefore longer unemployment durations. For immigrants from developed countries, the same pattern emerges, but it is less clear and the coefficients are smaller and more often insignificant. Second, the larger the fraction of time in the past five years spent on transfer incomes, the lower the probability of leaving unem-

---

13 For some reason, such individuals are characterized as unemployed while they are on these transfers.
ployment. The same results hold for sickness periods. However, when we look at the number of spells, we find the more unemployment spells an individual has had, e.g. during the past two years, the higher is the hazard rate out of unemployment. This coefficient, however, must be interpreted given the level of the variables reflecting the fraction of time spent on transfer schemes. That is, individuals with many short spells of unemployment in the past are also likely to have a short current spell of unemployment. Finally, we see that the local unemployment rate influences the hazard rate out of unemployment. This is consistent with e.g. Svarer et al. (2004). They find that the mobility among unemployed is very low in Denmark. As a consequence, people tend to be unemployed longer if they stay in the region with a high unemployment rate.

5.3 Assessment of predictive power

The primary purpose of this exercise is to construct a tool that can guide caseworkers in their work. The prime success criterion is of course that they can trust the outcome of the statistical model. Consequently, we are interested in identifying the group of newly registered unemployed that has the highest probability of experiencing more than 26 weeks of unemployment. We will denote that group 'potentially long-term unemployed' (PLTU) whereas their counterparts are the potentially short-term unemployed (PSTU). Define as the cut-off value the number which \( \Pr(T > \tau + 26 | T > \tau, x_t) \) shall exceed in order for an individual to be identified as PLTU. We can subsequently calculate the number of correct predictions; that is, the number of actually short-term unemployed (those with unemployment spells shorter than 26 weeks) who are also predicted to be short-term unemployed, plus the number of actually long-term unemployed who are also predicted to be long-term unemployed. This number can be related to the number of incorrect predictions. The choice of cut-off value will clearly have an effect on the outcome of this comparison. We have chosen to determine the cut-off value (separately for each of the 120 sub groups) such that the following sum is maximized\(^{14}\):

\[
\text{Number of short-term unemployed predicted to be short-term unemployed} + \\
\text{Number of long-term unemployed predicted to be long-term unemployed}
\]

\(^{14}\)This objective functions implies that the two groups are weighted by their relative sizes. By simply changing the weights it is possible to put more emphasis on one group if desired.
Note that when making an assessment of the correct predictions, we are forced to leave out all spells that are right censored at a duration shorter than 30 weeks in total. Moreover, when making predictions, we use only the value of explanatory variables at the beginning of the spell. That implies that we could make even better predictions if time-varying variables were taken into account in the process of prediction. Table 1 contains the aggregate numbers of correct and incorrect predictions for the entire country.

Table 2: The distribution of predictions and actual outcomes

<table>
<thead>
<tr>
<th>Groups</th>
<th>Fraction of correct predictions</th>
<th>STU PSTU</th>
<th>STU PLTU</th>
<th>LTU PSTU</th>
<th>LTU PLTU</th>
<th>No. of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women, ≤24, ins.</td>
<td>0.68</td>
<td>29,879</td>
<td>1,822</td>
<td>13,983</td>
<td>3,283</td>
<td>48,967</td>
</tr>
<tr>
<td>Women, ≤29, not ins.</td>
<td>0.66</td>
<td>71,485</td>
<td>18,182</td>
<td>34,466</td>
<td>29,709</td>
<td>153,842</td>
</tr>
<tr>
<td>Women, ≥25, ins.</td>
<td>0.61</td>
<td>174,840</td>
<td>70,063</td>
<td>110,001</td>
<td>104,331</td>
<td>459,235</td>
</tr>
<tr>
<td>Women, ≥30, not ins.</td>
<td>0.68</td>
<td>16,371</td>
<td>18,845</td>
<td>7,849</td>
<td>39,458</td>
<td>82,523</td>
</tr>
<tr>
<td>Men, ≤24, ins.</td>
<td>0.80</td>
<td>46,616</td>
<td>452</td>
<td>11,188</td>
<td>580</td>
<td>58,836</td>
</tr>
<tr>
<td>Men, ≤29, not ins.</td>
<td>0.68</td>
<td>102,444</td>
<td>12,030</td>
<td>43,316</td>
<td>16,004</td>
<td>173,794</td>
</tr>
<tr>
<td>Men, ≥25, ins.</td>
<td>0.71</td>
<td>271,420</td>
<td>25,947</td>
<td>104,353</td>
<td>36,474</td>
<td>438,194</td>
</tr>
<tr>
<td>Men, ≥30, not ins.</td>
<td>0.64</td>
<td>39,558</td>
<td>23,950</td>
<td>21,114</td>
<td>41,114</td>
<td>125,736</td>
</tr>
<tr>
<td>Total</td>
<td>0.66</td>
<td>752,613</td>
<td>171,291</td>
<td>346,270</td>
<td>270,953</td>
<td>1,541,127</td>
</tr>
<tr>
<td>Percentage</td>
<td>0.49</td>
<td>0.11</td>
<td>0.22</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: PSTU: Predicted to be short-term unemployed. PLTU: Predicted to be long-term unemployed.

The fraction of correct predictions is 0.66. Compared to the New Zealand profiling model that also employs a duration model, we gain a significant improvement in predictive power. In their model they are able to make 59% correct predictions. Since these fractions are dependent on the populations the models are trying to categorize, in particular the relative size of the groups, it is probably more interesting to consider the fraction of correct predictions in each group. 82% of the STU are predicted correctly while 55% of the LTU are predicted correctly, and the corresponding fractions are both equal to 59% in the New Zealand model. The key to the improved predictions is obviously the sub sampling and the large number of variables, especially the information on past labour market history, which
greatly improves the predictive power of the model.\textsuperscript{15} Furthermore a different objective function is used in their study, which in effect amounts to assigning equal weights to the two categories.

Looking across subgroups, it is revealed that the predictive power is higher for men than for women and for younger workers compared to their older counterparts. The former result is in line with previous research on modelling of individual unemployment.

We also performed out-of-sample predictions. In practice we randomly divided all the samples in halves. We estimated the model on the first half of the data and applied the parameters’ estimates to the second half for predictions. The predictions were almost identical to their full-sample counterparts.

Even though we obtain a reasonable level of correct predictions, there are still a substantial number of individuals who – if the model’s predictions were taken at face value – would be put into the wrong category, but that is exactly the reason why the statistical model is only a part of the profiling system. As discussed above, it is an input the caseworker can use to extract useful information regarding the potential risks of LTU facing an unemployed worker. In future versions, the profiling model will include information about unemployed workers gathered by the caseworkers. This information will give an impression of how motivated the individual is in terms of regaining employment, how employable the person is etc. When this information becomes available we expect to have a profiling model that is even better at predicting the LTU risk.

As a result of the estimation procedure with right censored observations it is not very informative to compute expected values and standard deviations. Instead we turn to a graphical depiction of some of issues concerning the precision of the predictions. In Figure 3 we look at the sub sample consisting of unemployed males in Aarhus who are insured and more than 25 years old, and see that the classification into the two categories is not nearly perfect. A considerable number of the PLTU find a job before 26 weeks, though the majority is unemployed in a longer period and 33\% actually more than a year. Similarly some of the PSTU are unemployed in more than 26 weeks but still, the majority get a job earlier, see Figure 4.

\textsuperscript{15}The variables used in the New Zealand model are: age, prior unemployment, gender, ethnicity, qualifications, urban location and regional location.
Another way to illustrate the prediction ability is to look at the fractions that are predicted correctly distributed over the actually experienced unemployment duration. Figure 5 shows this for the particular sub sample.

It should be noted that the reason for the low fractions of PLTU is that in this particular sub sample the group of STU is much bigger than the group of LTU and hence, predicting STU correctly receives a relatively high weight in the objective function, which is to maximize the number of correct predictions. However, abstracting from the low level, it is clearly seen that the fraction of correctly predicted long-term unemployed is increasing in the duration, so that we are more likely to correctly classify the "really long-term" unemployed, i.e. those who experiences more than a year of unemployment. This should then be compared to the cost of making incorrect predictions for different types of unemployed. If it is most costly to classify "really short-term" unemployed as PLTU and "really long-term" unemployed as PSTU compared to making mistakes in the interval in between, then we would like to have the highest fractions of correct predictions at very short and very long durations. But to the extent that the profiling system should work as an additional tool for the caseworker, it might actually be more valuable to be able to distinguish those in between if the caseworkers themselves are capable of correctly identifying the extremes. Another way to put this is that it could be a partial degree of explanation rather than an absolute that we want to maximize. In the discussion of this issue, it is also important to remember that we are trying to make a rather crude categorization of the unemployed into two groups and the results should hence be interpreted with this in mind.

6 Concluding remarks

In this paper we describe a statistical model used for profiling of newly unemployed workers in Denmark. When a worker - during his or her first six months in unemployment - enters
the employment office for the first time, this model predicts whether or not he or she will be
unemployed for more than six months from the current date. The caseworker’s assessment of
how to treat the person is partially based upon this prediction. The model – which performs
relatively well in terms of predicting actual unemployment – is the first step in the process
of developing statistical procedures to assist caseworkers in Denmark in their effort to bring
unemployed individuals back into employment. Future amendments to the model include
additional information based on caseworker’s assessment of the unemployed individuals in
order to make dramatic improvements of the model’s predictive ability. Moreover, assessment
of the effects of participation in various active labour market programs, that is, a targeting
system, seems to be the natural next generation in the world of profiling models.
References


Figure 1: The Job Barometer

Figure 2: Baseline hazard for Male UI Fund members, over 25, and resident in Aarhus County.
Figure 3: Histogram for actual durations for those who are PLTU.

Figure 4: Histogram for actual duration for those who are PSTU.
Figure 5: Fractions of PLTU and PSTU against actual duration
Choosing the Best Training Programme: Is there a Case for Statistical Treatment Rules?
Choosing the Best Training Programme: Is there a Case for Statistical Treatment Rules?*

Jonas Staghøj†, Michael Svarer‡ and Michael Rosholm§

Forthcoming in *Oxford Bulletin of Economics and Statistics*

August 2009

**Abstract**

When treatment effects of active labour market programmes are heterogeneous in an observable way across the population, the allocation of the unemployed into different programmes becomes particularly important. In this paper, we present a statistical model that can be used to allocate unemployed into different active labour market programmes. The model presented is a duration model that uses the timing-of-events framework to identify causal effects. We compare different assignment rules, and the results suggest that a significant reduction in the average duration of unemployment may result if a statistical treatment rule is introduced.

**Keywords**: Profiling, Targeting, Statistical Treatment Rules, Heterogeneous Effects

**JEL Codes**: J64, J68

---

* We thank two anonymous referees, seminar participants at University of Michigan, W. E. Upjohn Institute for Employment Research, Institut für Arbeitsmarkt und Berufsforschung, COST meeting in St. Gallen, NCoE Summer Institute in Helsinki and BA Conference on Evaluation of ALMPs for Operational Purposes in Nuremberg for very constructive comments and Birgitte Højklint for carefully reading the manuscript.

† School of Economics and Management, Aarhus University, E-mail: jstaghoej@econ.au.dk
‡ School of Economics and Management, Aarhus University, E-mail: msvarer@econ.au.dk
§ Department of Economics, Aarhau School of Business, Aarhus University, E-mail: rom@asb.dk
1 Introduction

In recent decades, active labour market programmes (ALMPs) have become an important component of labour market policies in many countries. In Denmark, ALMPs constitute an essential element in the so-called Flexicurity system of the labour market. The Flexicurity system consists of flexible employment relations in terms of low hiring and firing costs, a generous income replacement scheme in case of job loss, and an active labour market policy. ALMPs are essential in two ways. First, they ensure that workers who become unemployed can obtain the qualifications necessary for re-entering employment. Second, and perhaps more importantly, the ALMPs test the availability of unemployed workers to the labour market and provide an extra incentive to conduct active job search.

When ALMPs are used on a large scale and the costs are far from negligible on the government budget (costs of ALMPs constitute 1.5-2% of GDP in Denmark), it is crucial that they are used effectively. However, evaluation studies generally find that the effects of these programmes are modest and sometimes even negative.¹ One reason may be that ALMPs are not used in an optimal way; i.e. if the effects of ALMPs are heterogeneous across the population, it may be that the unemployed are not allocated to the programmes in such a way that the programme effect is maximized.

In this paper, we develop a statistical treatment rule to assign unemployed workers into the most effective programmes. The rule is based on the estimation of heterogeneous effects of ALMPs. We estimate programme effects in a duration model framework and use the timing-of-events method for identifying causal effects of programme participation (Abbring and van den Berg, 2003). We compare different assignments of unemployed to ALMPs, and the results show that the choice of assignment rule has a considerable impact on the expected

¹ See e.g. Heckman et al. (1999) and Kluve (2006) for reviews of the evaluations of ALMPs.
average duration of unemployment. The current assignment of unemployed workers to different programmes often involves a large degree of discretion to the caseworker. Sometimes deterministic eligibility rules state that programmes should be offered to particular groups of unemployed, but the choice of a particular programme is usually made by the caseworker (and the unemployed). We argue that this may not be optimal and propose a statistical treatment rule which exploits information on past programme effects to assign the unemployed to programmes.

The literature on the use of sophisticated statistical treatment rules to assign unemployed to ALMPs is still developing. The main contribution of this paper is to show the potential of such a rule based on detailed Danish labour market data. Another contribution is to use an alternative econometric strategy to identify causal programme effects since similar models developed and tested in e.g. Switzerland (SAPS) and Germany (TrEffEr)2 are based on matching estimators, and we use the timing-of-events method, as mentioned above.

The paper is organized as follows. In section 2, we first provide a framework for analyzing statistical treatment rules and review the existing literature. We then discuss the issue of caseworkers' discretion versus statistical treatment rules to set the stage for the rest of the paper. In section 3, we motivate and describe the econometric model, and section 4 contains a description of the data. Section 5 presents estimation results, discusses different ways of presenting the potential outcomes to caseworkers, and assesses the potential gains from implementing a statistical treatment rule. Finally, section 6 contains the conclusion and some considerations for future work.

---

2 See Fröhlich (2008), Fröhlich et al. (2003) or Behncke et al. (2006, 2007) for more details about the SAPS (Statistically Assisted Programme Selection) model and Stephan et al. (2006) for information about TrEffEr (Treatment Effects and Prediction).
2 Statistical Treatment Rules – concepts and considerations

The potential scope for statistical treatment rules is very broad and includes applications in finance, tax fraud detection, medicine, insurance, criminology, marketing and data mining. In this paper, we analyze the assignment of unemployed workers to ALMPs using a statistical treatment rule. We refer to Black et al. (2001) for a more detailed analysis of the key issues for a statistical treatment rule to be successful in this context.

The starting point for any statistical treatment rule is the goal it is supposed to accomplish. Historically, the first type of statistical treatment rules for assigning unemployed to ALMPs focused on identifying the unemployed at risk of long-term unemployment. This type of statistical treatment rule is sometimes called a profiling model, and it is mainly intended for improving equity by allocating help towards those most in need. If the unemployed at risk of long-term unemployment are also those who gain the most from participating in ALMPs, then profiling also improves efficiency. A newer type of statistical treatment rule, called a targeting model, focuses directly on improving efficiency. These models assign unemployed to programmes based on estimated programme effects.

2.1 Statistical treatment rules vs. caseworker discretion

As mentioned in the introduction, current assignment methods typically consist of a combination of some deterministic screening mechanism (e.g. target groups for ALMPs) and a large degree of discretionary power to caseworkers. In this section, we discuss the potential contribution of a statistical treatment rule.

---

3 For applications of statistical treatment rules in other literature areas, see e.g. Auerhahn (1999) for an application in criminology and Gottfredson and Moriarty (2006) for a recent overview, Yeo et al. (2001) for an application about insurance, Shaw et al. (2001) for an application in marketing, and Khan et al. (2001) or Murphy (2005) for applications in medicine.

4 This is based on the assumption that ALMP have positive effects.
The main contribution of a statistical treatment rule is the possibility to include literally hundreds of thousands of observations in performing inference on previously treated individuals, and subsequently use this information in an attempt to predict the future of those currently unemployed. Even experienced caseworkers will only meet a limited number of unemployed. It may therefore be impossible for caseworkers to infer an effect based on this small sample of unemployed workers. If, say, participation in a programme results in a 10% increase in the exit rate from unemployment and the baseline exit rate is 3% per week, then the post-treatment exit rate would be 3.3%. Such a difference would be impossible for a caseworker to detect without access to statistical analyses based on larger amounts of data. Moreover, the caseworker may not be able to follow the unemployed over a sufficiently long time period to observe the outcome after a programme is completed.\(^5\)

An important question to address is whether a statistical treatment rule should be seen as an alternative to caseworkers or as a tool offered to caseworkers. If implemented as an alternative to caseworkers, the system would ensure equal treatment of similar unemployed workers, which may be another important criterion\(^6\), and it would of course offer considerable scope for cost savings since no caseworker salaries would have to be paid. However, in this paper we argue that using statistically obtained treatment effects to equip caseworkers with additional information is a better idea. Obviously, caseworkers perform many tasks in addition to assigning the unemployed to ALMPs. Such tasks include assisting and advising in job search, motivating and monitoring the unemployed, etc. A statistical treatment rule would not render caseworkers redundant, but rather give them the possibility to focus more resources

---

\(^5\) In Denmark, caseworkers are not given automatic feedback on what happens after a worker is no longer observed to show up at the unemployment office. Hence, the caseworker does not necessarily know whether the worker found a job, left the labour force, moved to another municipality, etc. Moreover, an unemployed worker is rarely assigned to a single caseworker during a longer spell of unemployment. For these reasons, caseworkers do not always observe the duration of unemployment.

\(^6\) We do not attempt to say anything about fairness in this paper since the focus is on efficiency. There is a related economic literature on racial profiling where fairness issues are discussed more thoroughly, see Persico (2002), Persico and Todd (2005) or Harcourt (2004).
on other tasks. Since the institutions in charge of assignment are often public (or publicly funded), such an implementation may also serve as an attempt to compensate for the lack of a natural pricing-mechanism since the model advises the caseworkers on which programmes are the most valuable for different types of unemployed.

In addition, the caseworker may observe some information about the specific individual which the statistical model cannot take into account. Motivation and ability are typical examples. These unobserved variables are important to take into account, and if they are sufficiently important for estimating the true relationships between treatments and outcomes, then caseworkers may actually perform better than the statistical model. Lechner and Smith (2007), however, provide evidence suggesting that caseworkers are not very good at predicting effects, or interpreting the results differently, they are not seeking to maximize the expected outcomes. They show that caseworkers do no better than a random assignment mechanism. Similar results are found in Bell and Orr (2002), where caseworkers are also shown to have difficulties in identifying those with the largest gains.\(^7\)

When considering a combination of a statistical treatment rule and caseworker discretion, implementation is crucial. A good statistical treatment rule provides caseworkers with new and relevant information on potential outcomes with different treatments. A good caseworker identifies unobservable characteristics and uses these to modify the recommendations of the statistical treatment rule. Thus, the caseworkers could sort the unemployed into groups in terms of unobservables; for example by identifying the group of individuals who have problems beyond unemployment (e.g. psychiatric, health or substance

\(^7\) The finding that experts can be outperformed by a statistical method is by no means a rare finding. Grove and Meehl (1996) cites a meta study of 136 studies from a wide range of literatures and find that statistical methods outperformed experts' predictions in 64 cases, 64 showed no significant difference and only 8 favoured the informal expert approach.
abuse problems). The statistical predictions would probably not be very accurate for a person with an alcoholic abuse or a similar severe problem not observed in the data, but for a relatively 'mainstream unemployed', the statistical model may provide relevant information on potential outcomes to the caseworkers. In the Statistically Assisted Programme Selection model implemented in Switzerland and discussed in the following section, caseworkers are actually able to make subjective employability assessments and include these as a variable in the statistical model.

2.2 Existing evidence

Variations of the type of statistical treatment rule called profiling (assignment to programmes based on need) have been used in several countries. A general result is that unemployed at risk of long-term unemployment can indeed be identified with reasonable precision, but that these individuals do not experience systematically larger gains from participation in ALMPs than other unemployed workers.

The most prominent study on a targeting model, where the assignment to programmes is based on efficiency, is a randomized experiment conducted in Switzerland and described in Behncke et al. (2007). A statistical treatment rule called the Statistically Assisted Programme Selection (SAPS) model is estimated based on register data with detailed employment histories for the past 10 years as well as subjective assessments made by the caseworkers about the perceived employability about the particular unemployed worker. The model is estimated with matching techniques and hence relies on the conditional independence

---

8 As suggested by a referee, using our model we could actually provide three sets of predictions for an unemployed worker with observed characteristics, $X$, and good, average or bad unobserved characteristics. The caseworker could then chose which set of predictions that seems most appropriate based on a subjective judgement of characteristics unobservable for the statistical model.

9 These countries include Australia, Canada, Denmark, Finland, Germany, Netherlands, Sweden, Switzerland, UK and US.

10 For a more thorough review of the performance of these models, see OECD (1998) or Smith and Stagøj (2009).
assumption, which seems reasonable given the informative data base. This identification strategy is the main difference to our model, and we will discuss it in more detail in the econometrics section. The outcome variable in the SAPS model is the number of months in stable employment in the next 12 months. ALMPs are categorized into 6 or 7 types of programmes, depending on the region. When a caseworker meets with an unemployed worker, the SAPS model provides her with estimates of potential outcomes with different types of treatment. SAPS has been tested by random assignment within several employment offices. In each office, half the caseworkers were randomly selected to get access to the information from SAPS, and the other half was used as controls. Results from this randomized experiment show that caseworkers largely ignored the information generated by SAPS, and hence the experiment did not really reveal whether targeting could improve outcomes. Hence, for future experiments, it will be important to design incentive schemes in order to assure that caseworkers obtain and exploit the information provided by the targeting model.

In Germany, an ambitious project called Treatment Effect and Prediction (TrEffeR) is currently being tested.\(^{11}\) Like the SAPS model, this model is estimated with matching techniques. The introduction of TrEffeR is still in the experimental stage, and no results have been published yet. However, Arnkil \textit{et al.} (2008) state that initial pilot experiments show promising results.

In Canada and the US, two targeting models were developed but never implemented nor evaluated - the Service and Outcome Measurement System (SOMS) in Canada and the Frontline Decision Support System (FDSS) in the US. Both systems were intended to provide individual specific predictions of future labour market outcomes depending on the participation in different ALMPs. FDSS was pilot-tested in Georgia but discontinued before

an actual evaluation could take place. The reasons appear to be a combination of a sharp rise in unemployment and large administrative changes which both coincided with the introduction of FDSS.\textsuperscript{12} The SOMS project in Canada was also abandoned and Colpits (2002) states two main reasons for the failure to implement the system. First, some caseworkers opposed the system because they either did not trust the results or because they were afraid that the system would take over their jobs. The latter reason may have been reinforced by the timing of the introduction because it happened during a period with major reductions in the number of caseworkers. Secondly, there were concerns about data security because of the merging of detailed individual information into the database used to construct the statistical treatment rule.

The previous findings hence show that the processes of introduction and implementation of a targeting system are very important. It is also evident that we are still to see the first study documenting the performance of a fully functioning targeting system.

3 Choice of econometric model

The targeting model developed in this paper is a first attempt towards the development of a more complete model. In a duration model framework, we would ultimately like to model event histories, including all unemployment and employment spells, and allow for participation in sequences of programmes etc. However, in this paper we focus exclusively on unemployment spells, and we include only the first programme an individual participates in during an unemployment spell.\textsuperscript{13}

We estimate a model intended to maximize the effectiveness of programmes in terms of their ability to reduce unemployment duration. Hence, we first predict the future outcome for

\textsuperscript{12} See Behncke et al. (2007) who cite personal communications with Chris O'Leary.
\textsuperscript{13} Estimating effects of sequences of programmes would be a relevant but demanding extension. It is relevant for the way programmes are assigned in the real world, and it is demanding because a proper modelling of selection into later programmes should be allowed to depend on outcomes realized after the first programme.
each individual, conditional on participation in each of the possible programmes, and we then design a method for detecting the (set of) best programme(s). For each individual, we define the potential outcomes\(^{14}\) as

$$Y_0, Y_1, \ldots, Y_R$$

where \(\{0, 1, \ldots, R\}\) is the set of programmes available, denoting 0 as the no programme outcome. The fundamental evaluation problem is that for each individual, we observe at most one of the potential outcomes. We need additional identifying assumptions to estimate all other potential outcomes, and since the assignment of unemployed into programmes is not random, we have to distinguish causal programme effects from selection effects.

One possible identifying assumption used for estimating, e.g. the Swiss SAPS or German TrEffeR model, is the Conditional Independence Assumption (CIA), which can be stated as

$$Y_0, Y_1, \ldots, Y_R \perp D \mid X \quad \forall x \in \chi$$

where \(D \in \{0, 1, \ldots, R\}\) indicates the programme an individual is assigned to, and \(\chi\) is the relevant set of characteristics. With this assumption, we can estimate counterfactual outcomes conditional on \(X\).\(^{15}\) If the assignment of programmes is random, this assumption is clearly fulfilled. Proper use of it will typically require a rich data set containing detailed data on individual characteristics as well as market specific information. For the present analysis, we also have access to quite detailed data, so it might have been reasonable to assume that the CIA is fulfilled and proceed by constructing a matching estimator for duration outcomes.

We have chosen, instead, to estimate a parametric duration model within the timing-of-events framework which allows for selection on unobservables as well as observed

\(^{14}\) See Rubin (1974).

\(^{15}\) Frölich (2008) gives a detailed description of the estimation of effects underlying the SAPS model. It is explicitly described how they allow one set of conditioning variables, \(W\), needed for CIA to be fulfilled and a smaller subset of variables, \(X\), for future predictions - \(X\) being the variables observable to caseworkers. Predictions are then made by integrating out the conditional distribution of \(W\mid X\).
variables. The reasons for this choice are the following: First, if some unobserved variables exist which influence the selection process as well as the potential outcomes, approaches based on the CIA will result in biased estimates. Second, and more importantly, in the Danish flexicurity model, where programme participation becomes mandatory after some time, the construction of a matched comparison group for those participating in a programme after say 12 months of unemployment becomes suspiciously close to selection on the dependent variable. Those who do not participate in a programme in the current unemployment spell, but have more than 12 months of unemployment, are likely to i) have found employment a short time after the 12 months (otherwise they would have ended up in a programme), or ii) have unfavourable unobserved characteristics (rendering them unfit for programme participation). Hence, a model adequate for describing the dynamic process of selection into programmes (e.g. competing risks duration models for entry into different types of programmes) as well as dynamic sample selection (through explicit modelling of the exit rate from unemployment) is preferred in a situation with mandatory programme participation after some time in open unemployment. As described in the following section, we therefore use the timing-of-events model developed by Abbring and van den Berg (2003) to estimate the potential outcomes. This model allows for unobserved heterogeneity and corrects automatically for dynamic selection bias. The downside of this method is that, compared to the matching method, additional functional form assumptions are needed.

Finally, we assume that the Stable Unit Treatment Value Assumption (SUTVA) is fulfilled, which means that the potential outcomes for each individual do not depend on the treatment of other individuals. This implies that we ignore possible general equilibrium effects and restrict ourselves to a partial equilibrium analysis. Of course, this assumption is

---

16 A potential drawback for this approach is that the timing-of-events framework leaves us with less flexibility in terms of considering alternative outcome measures. With a matching estimator we could more easily change the outcome variable to be e.g. wages or productivity.

17 Rubin (1980).
particularly questionable in the Danish case of large scale mandatory programmes, but this
issue is left for future research.

3.1 The timing-of-events duration model

The stochastic variable of interest is the duration of unemployment, \( T_u \in (0, \infty) \). To model the
selection process into programmes, we define another stochastic variable, \( T_p \in (0, \infty) \) as the
time until programme participation. If \( T_p < T_u \), the individual participates in an ALMP, and if
\( T_p > T_u \), then the individual has not participated in an ALMP before leaving unemployment.
\( T_p \) is the minimum of four latent durations into one of the four types of programmes, \( T_{p1}, T_{p2},
T_{p3}, T_{p4} \). As we only evaluate each individual's first ALMP during an unemployment spell,
unemployment spells are right-censored at the time of the start of a second programme.

We model the hazard functions as Mixed Proportional Hazards (MPH), which means
that the hazards are modelled as the product of a baseline hazard, \( \lambda(t) \), depending on time, and
a scaling function, \( \phi(x_t, v) \), depending on the (possibly time-varying) observed characteristics,
\( x_t \in X \), and on an unobserved component, \( v \in V \), that is,

\[
\theta(t \mid x, v) = \lambda(t) \cdot \phi(x_t, v)
\]

With multi-spell data, the model is non-parametrically identified without making the
MPH assumption. Still, this functional form can be made fairly flexible. As shown in
Richardson and van den Berg (2008), multi-spell data also allows us to avoid the usual
exogeneity assumption requiring \( X \perp V \).

In the data, we observe when an individual starts and ends an ALMP, and using this
information, we construct two sets of time-varying indicator variables; one set of indicators
for participating in a programme of type \( j \in \{1,2,3,4\} \), \( d_{jt}^{1} \), and another set of indicators for

\[\text{18 The types of programmes are described in the data section.}\]
having completed programme \( j \), \( d_{j,t}\).\(^{19}\) We chose the functional form \( \phi(x_t,v) = \exp(x_t\beta + v_u) \) for the scaling function and write the hazard function out of unemployment as

\[
\theta_u[t \mid x_t, d_{j,t}^1, d_{j,t}^2, v_u] = \lambda_u(t) \exp \left[ x_t\beta_u + \sum_{j=1}^4 (d_{j,t}^1\gamma_j^1 + d_{j,t}^2\gamma_j^2) + v_u \right]
\]

(1)

As illustrated in Figure 1, the parameter \( \gamma_j^1 \) captures the effect on the hazard while participating in programme \( j \), and \( \gamma_j^2 \) captures the effect of having finished programme \( j \). We

---

\(^{19}\) \( d_{j,t}^2 \) is equal to 1 from the point in time where the programme ends and until the end of the current unemployment spell. If an individual is later observed in a new unemployment spell, then \( d_{j,t}^2 \) is set to 0. This is done because we find it most reasonable to assume that programme effects have faded away after an individual has left to employment and then re-entered into unemployment. As suggested by a referee, it might in principle be possible to test this assumption empirically if we observe comparable individuals with multiple unemployment spells. We have not pursued this any further. See Richardson and van den Berg (2008) for a more detailed discussion of the time-profile of programme effects.
would typically expect $\gamma^1_j$ to be negative while participating (a locking-in effect) since less time is available for searching for a job. A positive $\gamma^2_j$ (the post programme effect) indicates that the hazard rate increases after completion of the programme.

It is not obvious that a locking-in effect is a bad thing. If programmes are designed to provide valuable skills for the unemployed, it may be optimal for the unemployed to complete the programme. If this is the case, the post programme effects should be large.

We ignore any effects of ALMPs before programmes start, i.e. threat effects (see e.g., Black, Smith, Berger and Noel, 2003, and Rosholm and Svarer, 2008). These ex-ante effects are not necessarily so important when it comes to assigning individuals to specific programmes. Nevertheless, the no training option will be affected by the threat effect, see Rosholm and Svarer (2008). This potential bias is ignored in the following.

We specify the baseline hazards as piecewise constant within each of 12 distinct intervals.

The hazard into programmes is the sum of all four programme hazards

$$\theta_p(t \mid x, v_p) = \sum_{j=1}^{4} \theta_{p_j}(t \mid x, v_{p_j})$$  \hspace{1cm} (2)$$

The distributions of the unobserved variables are modelled as discrete with two mass-points, and we normalize one of the mass-points for each variable to 0 such that $V_u \in (v_u^1 = 0, v_u^2)$ and $V_{p_j} \in (v_{p_j}^1 = 0, v_{p_j}^2)$ for $j = 1, 2, 3, 4$. This way of introducing unobserved heterogeneity is based on Heckman and Singer (1984). The unobserved components in the four different programme hazards are assumed to be perfectly correlated such that $\text{cor}(V_{p_i}, V_{p_j}) \in (-1,1)$ for $i \neq j$. This assumption is relatively easy to relax, but the perfect
correlation restriction simplifies the estimation process. On the other hand, to allow for unrestricted correlation between $V_u$ and $V_p$ is important because this is the channel through which we allow for selection on unobservables.

Van den Berg (2001) writes that "... a consensus has emerged that multi-spell data allow for reliable inference that is robust with respect to the specification of the unobserved heterogeneity distribution."

Defining a non-censoring indicator, $C_i$, we can construct the likelihood function for individual $i$ with $K$ unemployment spells as

$$L_i(v_u, v_p) = \prod_{k=1}^{K} \theta_p \left[ x_{1k} | x_{1k}, v_p \right] \theta_u \left[ x_{uk} | x_{uk}, d_{1k}, d_{2k}, v_u \right] \cdot \exp \left[ - \int_{0}^{t_{i}^{st}} \theta_p \left[ s \mid x_s, v_p \right] ds - \int_{0}^{t_{i}^{st}} \theta_u \left[ t \mid x_t, d_{1k}, d_{2k}, v_u \right] dt \right]$$

(3)

Summing over the support of the discrete distributions for the unobservables, and summing over the sample of individuals, we can construct the complete likelihood function as

$$\prod_{j=1}^{N} \sum_{j=1}^{4} P_j \cdot L_j(v_u, v_p)$$

(4)

where $P_j$ is the associated probabilities of the mass-points. The parameters to be estimated are

$$\Psi = \{ \beta_u, \lambda_u \}, (\beta_j, \lambda_j), (v_u, v_p, P_j) \}_{j=1,2,3,4}.$$

3.2 Estimation of heterogeneous treatment effects

We estimate heterogeneous effects by allowing the effects to depend on the observable characteristics, and we assume that all heterogeneity is captured in this way. This also

---

20 In practice, it is difficult to use a completely unrestricted correlation structure. Several of the mass points collapse to the same values, and several of the associated probabilities converge to zero. Still, in an actual policy implementation, a more thorough estimation process would be recommendable.

21 Gaure et al. (2007) show in a Monte Carlo analysis that the method proposed by Heckman and Singer (1984) is quite precise and robust.

22 An interesting extension would be to use the results from Richardson and van den Berg (2008) where they show that the model is still identified even if the effect indicators are interacted with the estimated unobserved heterogeneity components.
loosens the assumptions about proportional effects in the sense that the size of the multiplicative effect is now allowed to depend directly on time-varying observed variables. Conditional on observables, the effects are assumed homogeneous, and hence we do not have to distinguish between the average treatment effect on the treated (ATET) and the average treatment effect (ATE) as long as we condition on the observable characteristics.\textsuperscript{23} This is useful since we want to predict all potential outcomes for every unemployed individual and not just for the subgroups actually participating in the programmes.\textsuperscript{24}

We thus include interaction terms between a subset of the characteristics, $x_t^s$, and the programme indicators, $d_{jt}^1$ and $d_{jt}^2$. Apart from a larger set of parameters, the estimation procedure is as before, and the hazard function out of unemployment can be written as

$$
\theta_u \left[ y_u \mid x_t, d_{jt}^1, d_{jt}^2, v_u \right] = \lambda_u(t) \exp \left[ x_t \beta_u + \sum_{j=1}^{4} \left( (1 \cdot x_t^s) \cdot d_{jt}^1 \cdot \gamma_{j}^1 + (1 \cdot x_t^s) \cdot d_{jt}^2 \cdot \gamma_{j}^2 \right) + v_u \right] \tag{5}
$$

where, if $x_t^s$ is a $1 \times K$ vector of characteristics, $(1 \cdot x_t^s) \cdot d_{jt}^1$ is a $1 \times (K+1)$ vector, and $\gamma_{j}^1$ is now a $(K+1) \times 1$ parameter vector, and similarly for $\gamma_{j}^2$. At first, we included all characteristics in the interactions, but this led to quite large standard errors, so instead we ended up using only the subset of characteristics which were most often significant in the interaction terms.

### 3.3 Identification

The exact identifying assumptions of the timing-of-events model are stated in Abbring and van den Berg (2003), and the two main assumptions are (i) \textit{proportional hazards}\textsuperscript{25} and (ii) \textit{no}...
anticipation. The proportionality assumption is needed in order to identify the unobserved heterogeneity terms. To see how this works, note that the distribution of unobservables will change over time because those unemployed with higher values of the unobserved variables will leave unemployment at a faster rate. Next, when the unobserved heterogeneity enters the hazard rate as a multiplicative term, it has a larger absolute effect on those individuals with a higher value of the systematic part of the scaling function, \( \exp(x, \beta) \). Combining these observations, this means that the distribution of unobservables among the individuals with high \( x\beta \) will change more rapidly over time than among those individuals with low \( x\beta \). However, this introduces an apparent nonproportionality since the time profile supposed to be captured by the baseline hazard \( \lambda(t) \) now looks different for different kinds of \( x\beta \) individuals. That is, we observe an apparent interaction between the scaling function, \( \exp(x, \beta) \), and the baseline hazard \( \lambda(t) \), and since we have assumed the hazard to be proportional in these terms, the only way we can capture the observed nonproportionality is by introducing unobserved variables, and this is how the distribution of these is identified.

The no anticipation assumption states that individuals are not allowed to know in advance the precise time at which they will start in a programme. If they had this information, they would be able to adjust their behaviour even before programme start, and that would invalidate our identification strategy. Individuals are, however, allowed to know the distribution of time until a programme starts. We argue that the no anticipation assumption is reasonable in this analysis since programme participation is typically planned only a few weeks before programme start.

The intuition behind the identifying strategy in the timing-of-events approach is to use exogenous variation in the time until the unemployed are assigned into programmes. This identification provided that we observe a sufficient amount of variation in covariates over time and across observations. He states that the required variation over time is minimal.
strategy is well-suited for an evaluation of ALMPs in Denmark because we observe a lot of variation in the time until individuals start in a programme. Some unemployed are assigned into a programme very early in their unemployment spell, and if there is exogenous variation in the timing of the assignments, we can use similar unemployed, not yet assigned to a programme, as the relevant counterfactuals. We model the selection processes into programmes conditional on observed and unobserved variables, and argue that some of the remaining variation is exogenous. This exogenous variation may emerge for several reasons. First, supply constraints (programme availability) at the local unemployment office influence the duration of time before an unemployed can be sent into a programme. Second, assignment into ALMPs is often planned at meetings between the caseworker and the unemployed, and these meetings occur with some exogenous variation for several practical reasons.

A potential problem with the timing-of-events approach in this context could be that unemployed in Denmark are required to participate in some kind of ALMP after a certain period of unemployment. After one year of open unemployment, the unemployed enter the so-called active period in which they must participate in a programme for 75% of the time. But apparently this is not enforced too strictly by the labour market authorities, so even for longer unemployment durations, we still have some individuals which can be used as counterfactuals. In this case, where the dynamic selection process is explicitly modelled and allowed to depend on unobservables, this is not a problem.

4 Data

4.1 Institutional settings in the Danish labour market

Denmark has a two-tiered system for unemployed workers. Most workers in Denmark - around 80% - are members of an unemployment insurance (UI) fund. These individuals have - upon the fulfilment of a few conditions - the right to receive UI benefits corresponding to
90% of the previous wage with an upper limit of approximately 1800 Euro per month. UI benefit payments are heavily subsidized by the state, which finances around 80% of total payments.

Unemployed workers without UI benefit eligibility may instead receive social assistance (SA) benefits. While non-insured workers only make up around 20% of the workforce, they make up a much larger fraction of the unemployed as the group typically consists of workers with a low attachment to the labour market. Hence, they are more often unemployed, and on average they are unemployed for longer periods. Social assistance benefits are means tested, but the amount is typically below the UI benefit level. Social assistance is administrated by the municipal authorities.

Unemployed workers receiving any of the two types of benefits are required to search for a job and to be fully available for work. If they do not fulfil these requirements, they run the risk of being sanctioned, so in principle they cannot turn down reasonable job offers or programme assignments made by the authorities. In the observation period which we consider, i.e. 1998-2003, those who stay unemployed for more than a year enter what is called the 'active period', during which they have to participate in an ALMP for 75% of the remaining time in unemployment. Before that period, they may also participate in programmes, and quite many do so.

4.2 The sample and the variables

The data we use is a 10% sample of the adult population in Denmark followed in the period 1998-2003, and it is constructed by merging information from several Danish administrative registers maintained by Statistics Denmark. We can follow each individual on a weekly basis and observe transitions between different labour market states. This individual labour market
history is then combined with socioeconomic variables in order to get a detailed data set suited for the analysis.

We define the population of interest as UI benefit recipients, men aged 25-55, who enter unemployment in the period from 1998 to 2003. This sub sample is chosen because there are special rules related to individuals below 25 as well as those above 55. During this time window, there were no major reforms in the regulation of the labour market.\textsuperscript{26} We distinguish temporary unemployment from ordinary unemployment by defining temporary unemployment as unemployment spells where the individual returns to a job within the same firm. This is done because former analyses have shown that temporary unemployment spells are of a different nature than ordinary unemployment.\textsuperscript{27} We only include ordinary unemployment spells in our analysis.

In the following section, we provide a brief description of the data used in the analysis. The dependent variable is unemployment duration measured in weeks.

The explanatory variables are the following: We include YEAR dummies to capture major business cycle and other calendar time effects. To capture local labour market effects, we include the LOCAL UNEMPLOYMENT RATE at the county level. A dummy for each COUNTY is also used.\textsuperscript{28}

**Individual background variables:** We use a dummy for having CHILDREN as well as a dummy for having CHILDREN LESS THAN 7 YEARS OLD and for being MARRIED. Dummies for different AGE groups are included: 25-29, 30-39, 40-49 (reference), and above 50. A dummy for being member of a UNION is included, and dummies for membership of different UI FUNDS are included. The UI funds are categorized into 9 different groups:

\textsuperscript{26} We have access to data from the period 1988-2003 and some of this information is used to construct labour market history variables.

\textsuperscript{27} See Jensen and Svarer (2003).

\textsuperscript{28} In the time period considered, Denmark consisted of 14 counties, and we included a dummy for the municipality of Copenhagen as well.
Building, Production, Technology, Trade, Service, Academic, Others and Self-employed. The reference group is Metal. We also add a dummy for being an immigrant from a NON-OECD country. Years of working EXPERIENCE in the labour market is included along with the hourly WAGE in the last observed job. Also, from this wage we construct the UI REPLACEMENT RATE. Considering education, we include dummies for NO FURTHER EDUCATION, FURTHER EDUCATION, and the reference group is VOCATIONAL EDUCATION. We have constructed a variable with the number of WEEKS REMAINING BEFORE ACTIVE PERIOD$^{29}$ to account for the fact that after the period of open unemployment, the unemployed enters the so-called active period and is more likely to be assigned to an ALMP. Former studies have shown an increase in the hazard rate out of unemployment in the weeks preceding this transition into a new period (Geerdsen, 2006).

**ALMP variables:**

i. PRIVATE JOB TRAINING: The individual is employed in a temporary job in the private sector where the employer receives a subsidy. The duration of these programmes is typically 6-9 months.

ii. PUBLIC JOB TRAINING: The individual is employed in a temporary job in the public sector, and the duration of these programmes is typically 6-12 months.

iii. CLASSROOM TRAINING: The individual participates in some kind of classroom training which includes short courses as well as ordinary education. It typically lasts only a few months.

---

$^{29}$ In 1998 the maximum number of weeks before entering the active period was 104. In the period until 2003 this number was gradually reduced to 65 weeks. The individual's number of weeks remaining before the active period is constructed as the above mentioned number less the accumulated number of weeks in unemployment since the individual was last observed to be in ordinary employment. A very short employment spell may not reset the accumulated time spent in unemployment, so the possible accumulation of time in unemployment over several unemployment spells, combined with the discretionary changes in the rules over time, implies that the variable "weeks remaining before active period" is not perfectly correlated with the duration of unemployment.
iv. OTHER TRAINING: A somewhat mixed group of all programmes which do not fall into any of the other categories. Typical programmes in this group include job search assistance, programmes of competence detection, individualized job training etc.

These quite detailed variables concerning the history of the individuals and the state of local labour markets should allow us to identify the selection process into programmes and important heterogeneity in the impacts of the programmes. However, we only have access to very broad measures of the different types of ALMPs. Although there are trade-offs when deciding how finely to define different groups, the four groups we have are almost certainly too broad. For the moment, however, it is all we have. The residual category other training may not in itself be very helpful for caseworkers, but given that a caseworker will receive information on all four types of programmes, it may still be possible to extract some useful information on the relative performance of the programmes.

Variables like motivation and ability are unobserved to the econometrician, but it may eventually become possible to improve the quality of the data by including information supplied by the caseworkers. This is done in the SAPS project in Switzerland, where the statistical model incorporates several variables with caseworkers' ratings of different skills of the unemployed. Unfortunately, this information is not yet available in the Danish data.

4.3 Descriptive data analysis

In this section, we present some descriptive statistics to get a first impression of the data. As seen in Table 1, the sample of men in the considered age group consists of 29,221 individuals with a total of 58,673 unemployment spells beginning during the observation period. 8,578 are observed to be assigned to a programme, of which classroom training is by far the largest programme.
TABLE 1

Description of the sample

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Mean duration of unemployment (weeks)</th>
<th>Mean duration before ALMP (weeks)</th>
<th>Mean duration of ALMP† (weeks)</th>
<th>Censored (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men aged 25-55</td>
<td>29,221</td>
<td>58,673</td>
<td>20.3</td>
<td>26.7</td>
<td>26.7</td>
</tr>
<tr>
<td>Unemployment Spells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participates in ALMP</td>
<td>8,578</td>
<td>50.6</td>
<td>24.8</td>
<td>18.1</td>
<td>58.0</td>
</tr>
<tr>
<td>Private Job Training</td>
<td>834</td>
<td>46.8</td>
<td>21.1</td>
<td>22.1</td>
<td>31.4</td>
</tr>
<tr>
<td>Public Job Training</td>
<td>777</td>
<td>67.4</td>
<td>27.8</td>
<td>35.4</td>
<td>62.0</td>
</tr>
<tr>
<td>Classroom Training</td>
<td>5,790</td>
<td>50.9</td>
<td>26.5</td>
<td>15.8</td>
<td>59.6</td>
</tr>
<tr>
<td>Other Training</td>
<td>1,177</td>
<td>40.7</td>
<td>17.3</td>
<td>15.2</td>
<td>66.4</td>
</tr>
</tbody>
</table>

Note: † Observed duration of the programmes.

The large amount of right-censored spells means that one should be careful when interpreting the raw mean durations. Still, it is clear that unemployment spells which involve an ALMP are much longer than the typical unemployment spell. We can also see that the programmes differ both with respect to the length of the time period until programmes start and with respect to their observed duration.

Apart from the descriptive statistics shown in Table 1, participants in the four programmes also differ on other dimensions. Particularly, those assigned to public job training are generally older, less educated, had lower wages in their last job, and have experienced a longer period of unemployment before programme participation than participants in other programmes. This fits quite well with the perception that public job training is a 'last resort'.

As stated when describing the econometric model, it is a key identifying assumption that we observe some (exogenous) variation in the time until being assigned to a programme. Taking a closer look at the hazard rates into different programmes in Figure 2, we see that

---

30 More descriptive statistics are available on request from the authors.
there is indeed a lot of variation in these durations. The entry rates are relatively stable for about a year and then increase as the unemployed enter the active period.

![Graph showing Kaplan-Meier hazard rates from unemployment to ALMPs]

Figure 2. Kaplan-Meier hazard rates from unemployment to ALMPs

5 Results

In this section, we first present the results from the timing of events model. We then consider different ways to exploit this information and evaluate the outcomes under the different treatment rules.

To save space, we only present the most relevant parameters; remaining results are available on request. Table 2 contains estimates of the coefficients to the interactions between the programme participation dummies (for classroom training), $d_{3,t}^1$ and $d_{3,t}^2$, and the most important observed characteristics. First, it should be noticed that it is not straightforward to interpret the coefficients in the tables since the model is non-linear. We therefore calculate expected durations in the next subsection. However, it is obvious from the tables that the estimated effects are certainly heterogeneous. For example, it appears that programmes are
more effective when the local unemployment rate is low, and that immigrants from Non-OECD countries have more favourable effects than the reference group (natives and immigrants from OECD countries). Second, the estimated standard errors are quite large for some parameters. This points to a requirement for large amounts of data (e.g. 100 % rather than 10 % sample) when estimating models with heterogeneous effects.31

TABLE 2

Programme effects for classroom training

<table>
<thead>
<tr>
<th>Coefficients in the hazard out of unemployment</th>
<th>Post programme effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Locking-in effects</strong></td>
<td><strong>Post programme effects</strong></td>
</tr>
<tr>
<td>Constant</td>
<td>0.470***</td>
</tr>
<tr>
<td></td>
<td>(0.160)</td>
</tr>
<tr>
<td>Age over 50</td>
<td>-0.273***</td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
</tr>
<tr>
<td>Non-OECD</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>(0.166)</td>
</tr>
<tr>
<td>Labour Market Experience</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
</tr>
<tr>
<td>Wage in Previous Job</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
</tr>
<tr>
<td>Elementary or High School</td>
<td>-0.095</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
</tr>
<tr>
<td>Further Education</td>
<td>-0.114</td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
</tr>
<tr>
<td>Local Unemployment Rate</td>
<td>-0.055***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
</tr>
</tbody>
</table>

Note: The table shows a subset of the coefficients from the hazard rate out of unemployment. The first column shows the coefficients to interactions with the participation dummy, and the second column shows the coefficients to interactions with the dummy for having completed a programme. *, ** and *** denote significance at the 10%, 5% and 1% significance levels. Standard errors in parenthesis.

5.1 Calculating counterfactual outcomes

We now construct expected unemployment durations conditional on participation in the different programmes. That is, we construct all the potential outcomes for each individual. The expected duration of unemployment is a natural way of presenting the results, and

31 Danish administrative registers cover the entire population; hence, had these been available to us, we could have estimated the interaction parameters more precisely.
facilitates a more clear interpretation compared to just analyzing the coefficients on all the interaction terms.

When constructing counterfactual outcomes, it is useful to look at the issue from the perspective of the case worker: At a given point in time, a caseworker may assign the unemployed to a programme, or she may not, in which case the unemployed may be assigned to a programme some time in the future. Hence, in computing counterfactual outcomes, it would seem that the relevant \textit{no training} potential outcome is not \textit{no training in this unemployment spell}, but rather \textit{no training this period}. This implicitly means that the caseworker would compare the outcome of assigning the person to a programme now versus not doing so now but perhaps later. Thus, the \textit{no training this period} counterfactual would require a calculation of the potential outcomes of assigning the individual to a programme at a later date, possibly incorporating some regulatory constraints; e.g. if a person has to participate before some elapsed unemployment duration. This would be intellectually straightforward but computationally burdensome to do. Since the programme effect is not allowed to depend on elapsed unemployment duration, the optimal treatment does not vary over time, this issue is not immediately relevant in the present model.\footnote{The design of the no training outcome has consequences when it comes to presenting the results to the case workers, which is discussed in a later Section.} Hence, at present we restrict ourselves to comparing training now to \textit{no training during this unemployment spell}. This may also be a more clear-cut counterfactual, and hence easier to understand when it comes to presenting the results to caseworkers, see the next section.

We need to make a few additional assumptions. The durations of \textit{private job training} and \textit{public job training} are set to 26 weeks, while the durations of \textit{classroom training} and \textit{other training} are set to 16 weeks.\footnote{In a more elaborate model, the programme durations could easily be an additional choice parameter, but for now it is taken as given. These durations are chosen to reflect realistic durations of the programmes. The average durations shown in Table 1 are observed and not planned durations of the programmes, which are unfortunately}
different assumptions, each of which provides us with useful information on the effectiveness of the system. First, we consider what happens if we set the starting time of the programme to $t=0$; that is, we assume that the unemployed enter a programme immediately after entering unemployment. This assumption makes it possible to do the analysis for all individuals entering unemployment and indicates the effects of using the statistical treatment rule at this point in time.\footnote{Unfortunately, this choice of starting time does not allow us to compare the statistical model with the current system of caseworker assignment because caseworkers choose \textit{no training} for nearly all unemployed during the first week. For this reason, we also consider setting the starting time to the point in time where the unemployed in our sample are in fact assigned to a programme. This allows us to compare caseworker assignment with the suggested assignments from the statistical treatment rule, but only for those unemployed who enter a programme during the observation period. In the following section, we present the results with the starting time set to $t=0$.}

By combining the parameters of the model with the observed characteristics and the assumptions on the duration and timing of programmes, we can calculate the expected duration of unemployment conditional on participation in each of the programmes

$$E[T_u \mid \{x_j, d_{j,s}, d_{j,s}^2, v_s\}] = \sum_{m=1}^{2} \sum_{n=1}^{2} \Pr(v_m^n, v_p) \sum_{k=1}^{\infty} \exp \left[ - \sum_{s=1}^{k} \theta_s(s \mid x_j, d_{j,s}, d_{j,s}^2, v_m^n, v_p) \right]$$

(6)

To illustrate the importance of allowing for a richer set of explanatory variables and to control for selection effects, we present results from three different specifications. In the 'Simple' model we only use three dummy variables (age 25-39, Native Dane, Low local

not available in the data. A strong locking-in effect for a specific type of programme will increase the observed duration. In particular, this is the case for public job training.\footnote{We have repeated the analysis at various points in time, $t=3,6,9,12$ months of unemployment. The findings corroborate the results presented in the rest of the paper and are available on request.}
unemployment rate). This specification is included to analyze whether a subgroup analysis would actually suffice to identify heterogeneous effects. The specifications 'No Unobs.' and 'Unobs.' both use all variables described in the data section, but only the latter corrects for selection based on unobserved heterogeneity. A summary of the results is shown in Table 3.

TABLE 3

Programme effects. Mean expected unemployment durations under different specifications

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Simple</th>
<th>No Unobs.</th>
<th>Unobs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Training</td>
<td>31.4</td>
<td>37.5</td>
<td>35.1</td>
</tr>
<tr>
<td>Private Job Training</td>
<td>39.2</td>
<td>33.3</td>
<td>39.7</td>
</tr>
<tr>
<td>Public Job Training</td>
<td>84.0</td>
<td>59.3</td>
<td>59.4</td>
</tr>
<tr>
<td>Classroom Training</td>
<td>44.9</td>
<td>40.0</td>
<td>43.3</td>
</tr>
<tr>
<td>Other Training</td>
<td>62.2</td>
<td>49.9</td>
<td>48.3</td>
</tr>
</tbody>
</table>

Note: The ‘Simple’ specification only contains three variables (age 25-39, Native Dane and Low local unemployment rate). The ‘No Unobs.’ specification contains all variables but does not correct for unobserved heterogeneity. The ‘Unobs.’ specification contains all variables and allows for selection based on unobserved heterogeneity. Results show the simulated durations in unemployment averaged over the 58,863 spells starting in the sample.

The results from the 'Simple' specification differ a lot from the full specifications, and we believe this is a result of inadequate modelling of the selection process. The difference between 'No Unobs.' and 'Unobs.' is less pronounced, but especially private job training is evaluated more positively under the 'No Unobs.' specification. Private job training is typically the type of programme where we would be most likely to find 'cream skimming'; i.e. caseworkers assigning the 'better' unemployed to this type of programme. Our results show that this does not only take place based on observed characteristics, but also based on unobserved characteristics. In the full model (Unobs.), the average causal effect of the private job training on unemployment duration is positive. In the remaining analysis, we will consider the 'Unobs.' specification as our best model of the truth and use results under this specification to evaluate different assignment rules. We will only use the other model.
specifications to consider what would happen if these specifications were used to construct the statistical treatment rules.

Notice that - in line with the previous literature on short-term effects of ALMPs on individual job finding rates as summarized in e.g. Heckman et al. (1999) and Kluve (2006) - participation in ALMPs generally prolongs unemployment periods.\(^{35}\) Still, private job training is the best programme on average and public job training the worst. Moreover, there is variation in the estimated effects over the individuals in the sample. In Figure 3, we show histograms for the distributions of estimated effects. A negative effect in the histogram implies that unemployment duration would be shortened as a result of participating in the programme, when compared to no training. It is seen that there are negative effects for all programmes, and so it would be desirable to target the programmes to these individuals and avoid assigning individuals to programmes where the expected unemployment duration is much longer.

\[^{35}\text{These results are also consistent with similar Danish studies, see e.g. Rosholm and Svarer (2008).}\]
Figure 3. Histograms for the distribution of the estimated programme effects over the 58,673 starting spells in the sample. The dark parts highlight spells with negative programme effects.

### 5.2 Presenting results to caseworkers

When a caseworker meets with an unemployed client, they agree on a plan for getting the client back into work as soon as possible. At such a meeting, it might be helpful for the caseworker to know something about the effectiveness of the available programmes for the particular client. This information is exactly what the targeting model provides. The question is how to present the results and the related uncertainty in the most comprehensible way? Frölich (2008) describes a method called Multiple Comparisons with the Best (MCB), which is used to present results to caseworkers in the SAPS project in Switzerland.\footnote{Horrace and Schmidt (2000) contains a description of this statistical method when used in economic contexts.} For a given significance level, this method calculates, for each individual, a set of programmes, $\hat{S}$, which contains the best programme with the chosen probability. This set of best programmes may
contain only one programme, in which case this programme is significantly better than all the other possible programmes, or it may contain more than one programme, which would allow us to say that the best programme is one of the programmes in the set, but not to point at a particular programme. MCB can also provide a set of programmes which is significantly worse than the best programme, \( \hat{W} \).

This way of presentation seems appropriate when the statistical treatment rule is intended as an information tool to caseworkers who still have discretionary power. The caseworker may choose a programme from the set of best programmes, or she may take additional information into account - e.g. supply constraints and unobserved characteristics - and overrule the suggestions made by the programme.

MCB is also appropriate from a statistical point of view; the testing is automatically carried out using the joint distribution of the impact estimates. This is not always done. Often, \( t \)-statistics are calculated to test whether a particular programme effect is significantly larger than 0, or larger than another programme effect. A problem with this test procedure may arise if it is repeated several times because the probability of making a type 1 error, that is, to reject a true hypothesis, would not be equal to our chosen significance level. If we just test a sufficient amount of parameters, some of them will probably be considered significant, even if they are not. We could then adjust the significance level for the individual \( t \)-tests to take into account that we are going to make a certain number of tests. However, using the MCB method, this is automatically taken into account. This latter point may not be very important when we have only four types of programmes, but if we could divide the programmes into perhaps 50 categories, then it might be important for the results. Finally, if we want to test all programmes against the best programme, we need to include the additional uncertainty stemming from the fact that we do not \textit{ex ante} know the best programme and, hence, which programme to use for comparison.
Note that, when using MCB, it would be tricky to have *no training this period* as the *no training* potential outcome. The reason is that, even at the optimal time of participation in, say, a *private job training* programme, we would still compare *private job training now* to *private job training a month from now*, and with the present size of the standard errors, *no training* would always be in the set of best programmes. However, in a full scale implementation of a targeting model, it would be important to address the issue of optimal timing of programme participation.

### 5.3 Multiple comparison with the best – results

In Table 4, we present the MCB results for a typical individual and include a traditional $t$-test analysis for comparison. The standard errors are found using parametric bootstrap with 500 parameter draws from the multivariate normal parameter vector. The first two columns in Table 4 show that *public job training, classroom training* and *other training* all have effects that are significantly different from zero for this individual, and that the 2.6 weeks increase in unemployment for *private job training* is not significant. The last two columns in Table 4 show the results from a MCB analysis for the same typical individual. Here, *no training* and *private job training* constitute the set of best programmes, while none of the remaining programmes are in the set of worst programmes.
### Table 4

**Example of output for a typical individual**

<table>
<thead>
<tr>
<th>Programme</th>
<th>Effect</th>
<th>Standard error</th>
<th>Duration</th>
<th>In set of best</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Training</td>
<td>0</td>
<td>0</td>
<td>38.8</td>
<td>Yes</td>
</tr>
<tr>
<td>Private Job Training</td>
<td>2.6</td>
<td>5.86</td>
<td>41.5</td>
<td>Yes</td>
</tr>
<tr>
<td>Public Job Training</td>
<td>34.7**</td>
<td>15.68</td>
<td>73.6</td>
<td>No</td>
</tr>
<tr>
<td>Classroom Training</td>
<td>5.7***</td>
<td>1.76</td>
<td>44.6</td>
<td>No</td>
</tr>
<tr>
<td>Other Training</td>
<td>19.3**</td>
<td>7.62</td>
<td>58.1</td>
<td>No</td>
</tr>
</tbody>
</table>

*Note:*, ** and *** denote significance at the 10%, 5% and 1% significance levels. The set of best programmes in MCB analysis is found using a 95% confidence interval. The ‘typical’ individual becomes unemployed in 1998, has no children, is 30-39 years old, is married, is a native Dane, has 12 years of labour market experience, deflated wage in last job is 100 Dkr., has a vocational education, is a member of a union in the category called ‘production’, has a UI compensation rate at 90%, has 50 weeks remaining before active period, the local unemployment rate is 7%, and he lives in Copenhagen.

Summary statistics for the MCB analysis for all the individuals in the sample (available on request) show that while we can seldom point to a single programme as being the best, we can quite often distinguish the two or three best programmes. Not surprisingly, the set of best programmes is most often constituted by no training and private job training. Whether one prefers the MCB way of presenting the results over the usual $t$-statistic way is ultimately a matter of taste, but we argue that the MCB method provide a constructive way to inform caseworkers about uncertainty, and it often gives them a set of programmes to choose from. Also, it seems that the value of the MCB method is increasing in the number of programmes, such that this method is suited for allowing a more detailed categorization of programmes.

### 5.4 Comparing assignment rules

Having calculated expected durations for each individual, we turn to the selection of an assignment rule. We first consider the scenario where individuals are assigned to programmes...
at \( t=0 \), and we consider the following assignment rules: i) Assign *no training* to everyone and ii) Assign the *best*, *worst* or a *random programme* to everyone.\(^{37}\)

These assignment rules ignore constraints on the availability of different types of programmes, which may be somewhat unrealistic, although the composition of programmes has in fact changed quite dramatically over time. Therefore, we also consider some assignment rules where we have imposed the restriction that the fraction of unemployed assigned to each programme should be as actually observed.\(^{38}\) The restrictions impose an upper limit on the number of unemployed that can participate in each programme, but allow as many as desired to be assigned to *no training*. Such restricted assignment rules will not be optimal because of the imposed constraints, but they may serve as a lower limit of the effects of imposing the assignment rules.

When considering restricted assignment rules, the sorting of the unemployed becomes important because some programmes may eventually be closed for further assignments. To make a fair comparison to the current system, we have chosen to sort the sample of unemployed randomly. This should mimic the current system where caseworkers do not have information about all unemployed who will enter the unemployment office during the relevant period. Hence, we only allow the statistical treatment rule to use the same information as caseworkers have.

We apply the different assignment rules to each of the three different model specifications to investigate if the complexity of the preferred model pays off in terms of improved performance. We would expect assignment rules based on the 'Unobs.' model to outperform the other specifications. A potential problem when comparing assignment rules is related to overfitting since some of the assignment rules (i.e. choosing the best programme)

\(^{37}\) This section follows the procedures suggested in Lechner and Smith (2007), which contains a comparable analysis.

\(^{38}\) The restrictions of the fractions in each programme are calculated from the actual assignment as seen in Table 1.
have an 'option like' element such that the relative performance of these assignment rules are increasing in the variance of the estimates. To avoid this overfitting bias, we use the following procedure. First, the resulting allocations under each assignment rule are calculated. Then, we again use a parametric bootstrap with 500 parameter draws from the multivariate normal distribution to simulate 500 sets of potential outcomes for each individual. And finally, the performance of each assignment rule is assessed by averaging over the 500 simulated outcome sets. Using this procedure, the results will show if an assignment rule chooses effective programmes, or if an assignment rule just chooses programmes that seem effective based on a particular draw of a noisy estimated programme effect.39

In Table 5, the average unemployment duration for each assignment rule is shown.

---

39 This 'trick' actually turns out not to be of major importance. Similarly, as suggested by a referee, we have tried to estimate the underlying models based on a smaller sample to get higher variance and see how this affects the 'assign to best programme' rule. We would expect this assignment rule to perform better if variance is increased, but our results show little change. These additional results are available on request.
### TABLE 5
Comparing assignment rules used at t=0. The outcome variable is weeks of unemployment

<table>
<thead>
<tr>
<th>Assignment rule</th>
<th>Simple Average duration</th>
<th>No Unobs. Average duration</th>
<th>Unobs. Average duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Training</td>
<td>35.2</td>
<td>35.2</td>
<td>35.2</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.30)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>Best</td>
<td>34.8</td>
<td>33.5</td>
<td>31.7</td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td>(0.73)</td>
<td>(0.46)</td>
</tr>
<tr>
<td>Worst</td>
<td>60.7</td>
<td>67.1</td>
<td>67.8</td>
</tr>
<tr>
<td></td>
<td>(1.89)</td>
<td>(2.08)</td>
<td>(2.05)</td>
</tr>
<tr>
<td>Random</td>
<td>45.7</td>
<td>45.7</td>
<td>45.7</td>
</tr>
<tr>
<td></td>
<td>(0.67)</td>
<td>(0.67)</td>
<td>(0.67)</td>
</tr>
</tbody>
</table>

**Restricting the maximum number of individuals in each of the 4 programmes**

<table>
<thead>
<tr>
<th>Assignment rule</th>
<th>Simple Average duration</th>
<th>No Unobs. Average duration</th>
<th>Unobs. Average duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best, restricted</td>
<td>35.0</td>
<td>35.4</td>
<td>34.4</td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td>(0.33)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>Worst, restricted</td>
<td>38.8</td>
<td>39.6</td>
<td>39.7</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.46)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>Random, restricted</td>
<td>38.2</td>
<td>38.2</td>
<td>38.2</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.35)</td>
<td>(0.35)</td>
</tr>
</tbody>
</table>

**Interesting differences**

<table>
<thead>
<tr>
<th></th>
<th>Simple Average duration</th>
<th>No Unobs. Average duration</th>
<th>Unobs. Average duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impose restrictions</td>
<td>7.5***</td>
<td>7.5***</td>
<td>7.5***</td>
</tr>
<tr>
<td>(Random – Random, restr.)</td>
<td>(0.54)</td>
<td>(0.54)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>Introduce Targeting, restricted</td>
<td>3.2***</td>
<td>2.8***</td>
<td>3.8***</td>
</tr>
<tr>
<td>(Random, restr. – Best, restr.)</td>
<td>(0.23)</td>
<td>(0.24)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Introduce Targeting, unrestricted</td>
<td>3.3***</td>
<td>4.7***</td>
<td>6.5***</td>
</tr>
<tr>
<td>(Random, restr. – Best)</td>
<td>(0.23)</td>
<td>(0.70)</td>
<td>(0.46)</td>
</tr>
</tbody>
</table>

*Note:* *, ** and *** denote significance at the 10%, 5% and 1% significance levels. Standard errors are constructed using a parametric bootstrap method. See the text for details. Programmes are assumed to start immediately after entering unemployment. ‘Restricted’ means that the maximum number of individuals in each programme is fixed and the number assigned to no training is not restricted. Individuals are ordered randomly before assigning to programmes.

The upper part of Table 5 shows that the full model (Unobs.) is better in terms of assigning unemployed to the best programme. In addition, being assigned to the best programme is significantly better than assigning everyone to no training. In the middle part of Table 5, we restrict the allocation to correspond to the actual number of slots available. It is seen that the simple specification performs almost as good as the more advanced model.
In Table 6, we show the allocation of individuals in the different programmes. Assigning everyone to the best programme in the simple specification implies that the vast majority are assigned to no training. This is not the case for the comparable assignment rule under the 'Unobs.' specification, where much more unemployed are actually assigned to programmes. So while it might first appear as if the 'Simple' subgroup analysis seems sufficient, this model does not really exploit the heterogeneous effects in the same way as under our preferred 'Unobs' specification.

In the lower part of Table 5, we first show the difference between restricted and unrestricted assignment rules. It is seen that imposing the restrictions has a large impact on the average unemployment duration. More interestingly, we next try to demonstrate the potential gain from a targeting model, which shows that average unemployment duration can be lowered significantly by choosing the best programme instead of a random programme. This effect is due to the targeting model if random assignment is a good approximation to the current system. The 'Unobs.' specification is seen to provide the most effective targeting model and offers a decrease in average duration of unemployment of 3.8 weeks. If it is possible to introduce the targeting model without the restrictions, then the implied decrease in unemployment duration is 6.5 weeks. Hence, depending on whether the number of open slots in the programmes are binding, an estimate of the overall decrease in the unemployment duration when introducing a targeting model would be in the interval 3.8 - 6.5 weeks.

---

40 We shall elaborate on this issue below.
TABLE 6
Resulting assignments using different assignment rules

<table>
<thead>
<tr>
<th>Assignment rule</th>
<th>No Training</th>
<th>Private</th>
<th>Public</th>
<th>Classroom</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Training</td>
<td>58,673</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>When the ‘Simple’ specification is used to estimate the targeting model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best</td>
<td>54,547</td>
<td>1,158</td>
<td>2,698</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Worst</td>
<td>0</td>
<td>0</td>
<td>36,290</td>
<td>0</td>
<td>22,383</td>
</tr>
<tr>
<td>Random</td>
<td>11,752</td>
<td>11,870</td>
<td>11,612</td>
<td>11,709</td>
<td>11,730</td>
</tr>
<tr>
<td>Best, restricted</td>
<td>55,044</td>
<td>1,675</td>
<td>1,561</td>
<td>393</td>
<td>0</td>
</tr>
<tr>
<td>Worst, restricted</td>
<td>41,448</td>
<td>1,675</td>
<td>1,561</td>
<td>11,626</td>
<td>2,363</td>
</tr>
<tr>
<td>Random, restricted</td>
<td>41,448</td>
<td>1,675</td>
<td>1,561</td>
<td>11,626</td>
<td>2,363</td>
</tr>
<tr>
<td>When the ‘No Unobs.’ specification is used to estimate the targeting model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best</td>
<td>17,214</td>
<td>30,916</td>
<td>6,481</td>
<td>1,657</td>
<td>2,405</td>
</tr>
<tr>
<td>Worst</td>
<td>1,593</td>
<td>4,804</td>
<td>31,382</td>
<td>1,292</td>
<td>19,602</td>
</tr>
<tr>
<td>Random</td>
<td>11,752</td>
<td>11,870</td>
<td>11,612</td>
<td>11,709</td>
<td>11,730</td>
</tr>
<tr>
<td>Best, restricted</td>
<td>43,307</td>
<td>1,675</td>
<td>1,561</td>
<td>9,767</td>
<td>2,363</td>
</tr>
<tr>
<td>Worst, restricted</td>
<td>41,448</td>
<td>1,675</td>
<td>1,561</td>
<td>11,626</td>
<td>2,363</td>
</tr>
<tr>
<td>Random, restricted</td>
<td>41,448</td>
<td>1,675</td>
<td>1,561</td>
<td>11,626</td>
<td>2,363</td>
</tr>
<tr>
<td>When the ‘Unobs.’ specification is used to estimate the targeting model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best</td>
<td>32,159</td>
<td>13,090</td>
<td>6,714</td>
<td>196</td>
<td>6,514</td>
</tr>
<tr>
<td>Worst</td>
<td>356</td>
<td>10,645</td>
<td>28,850</td>
<td>4,345</td>
<td>14,477</td>
</tr>
<tr>
<td>Random</td>
<td>11,752</td>
<td>11,870</td>
<td>11,612</td>
<td>11,709</td>
<td>11,730</td>
</tr>
<tr>
<td>Best, restricted</td>
<td>51,493</td>
<td>1,675</td>
<td>1,561</td>
<td>1,581</td>
<td>2,363</td>
</tr>
<tr>
<td>Worst, restricted</td>
<td>41,448</td>
<td>1,675</td>
<td>1,561</td>
<td>11,626</td>
<td>2,363</td>
</tr>
<tr>
<td>Random, restricted</td>
<td>41,448</td>
<td>1,675</td>
<td>1,561</td>
<td>11,626</td>
<td>2,363</td>
</tr>
</tbody>
</table>

Note: Number of individuals assigned to the programmes under different models and assignment rules. ‘Restricted’ means that the maximum number of slots in each programme is fixed and the number assigned to no training is not fixed. Individuals are ordered randomly before assigning to programmes. The model specifications ‘Simple’, ‘No Unobs.’ and ‘Unobs.’ are described in the text.

Assigning everyone to no training is almost as good as assigning everyone to the best programme when the restrictions apply. However, due to neglect of general equilibrium effects in the estimated model, we would not stretch this result too far. The elimination of all ALMPs may very well affect both unemployed and employed workers in ways we have not taken into account. It has, for example, been argued that the presence of mandatory ALMPs acts as a tax on leisure and serves as a threat effect for all unemployed even before they are
assigned to the programmes. Furthermore, as seen in Table 6, *no training* is the best programme for more than half of the unemployed in the sample. However, in contrast to the results on mean programme effects, this actually suggests that many unemployed may be better off when assigned to a programme. Notice that the restrictions imposed are binding for most assignment rules. A noticeable deviation from this observation is that the assignment to the best programme (both restricted and not restricted) only assigns very few unemployed to *classroom training* programmes.

In the interpretation of the results above, we argued that a reasonable approximation of the current allocation rule is random assignments with restrictions. To provide some evidence of this assumption, we now consider the situations where caseworkers actually assigned an unemployed into a programme. Results from this exercise are shown in Table 7.\textsuperscript{41}

\textsuperscript{41} When calculating remaining unemployment durations from a time $t>0$, we encounter a problem of which distribution of the unobserved heterogeneity to use. We only have an estimate for this distribution at $t=0$, and we know that the conditional distribution will change over $t$. In lack of a better method, we have used the distribution at $t=0$ for later $t>0$ as well. This simplification biases the estimated remaining duration, but because we know that the distribution we use will be too 'favourable', we know the direction of the bias, namely that we will underestimate the remaining duration. But more importantly, the simplification has no influence on the relative ranking of the programmes because the unobserved component enters the hazard multiplicatively. If we were to use the correct distributions at $t>0$, the estimated differences between programmes would be larger, and the differences between assignment rules would similarly increase. Our estimates are hence conservative estimates of the differences.
### TABLE 7

*Comparing actual choice by caseworkers to Targeting model*

<table>
<thead>
<tr>
<th>Caseworkers’ Choice</th>
<th>No Training an Option (%)</th>
<th>No Training not an Option (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Programme</td>
<td>15.6</td>
<td>13.5</td>
</tr>
<tr>
<td>Second Best Programme</td>
<td>13.9</td>
<td>43.4</td>
</tr>
<tr>
<td>Third Best Programme</td>
<td>15.2</td>
<td>27.4</td>
</tr>
<tr>
<td>Fourth Best Programme</td>
<td>19.7</td>
<td>15.7</td>
</tr>
<tr>
<td>Fifth Best Programme</td>
<td>35.5</td>
<td>-</td>
</tr>
<tr>
<td>In the Set of Best Programmes</td>
<td>71.5</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assignment Rules</th>
<th>No Training an Option (mean duration)</th>
<th>No Training not an Option (mean duration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caseworkers’ Choice</td>
<td>61.6 (2.05)</td>
<td>61.6 (2.05)</td>
</tr>
<tr>
<td>Best, restricted</td>
<td>42.4 (1.46)</td>
<td>52.6 (1.64)</td>
</tr>
<tr>
<td>Random, restricted</td>
<td>55.6 (1.73)</td>
<td>59.2 (1.88)</td>
</tr>
</tbody>
</table>

*Note:* The actual choice by caseworkers are compared to the outcomes under the targeting model estimated using the ‘Unobs.’ specification. Standard errors are in parenthesis.

It shows that caseworkers chose the worst programme most of the time (35.5%). On the other hand, they quite often choose a programme from the MCB set of best programmes (71.5%), and this is comparable to a random assignment which would result in 69.2%. The average difference in the remaining duration of unemployment under the actually chosen programme and the best programme is 19.2 [std. err. 1.74] weeks. Compared to random assignment, it is seen that caseworkers perform a little worse. However, since caseworkers are sometimes forced (by the rules) to assign unemployed into a programme, we have also analyzed whether they then chose the best available programme. That is, we repeat the analysis shown in the first column in Table 7, but remove the option of assigning unemployed to *no training*. These results are shown in the last column of Table 7, and the change is seen to explain part of the difference between caseworkers' choice and the best choice. The difference between caseworker assignment and assignment to the best programme is now 9.0
... weeks, and caseworkers perform about as well as random assignment in this comparison.

The reality about the available choice set for caseworkers is probably a combination, implying that they sometimes have the possibility of choosing the option *no training*, while at other times they are effectively forced to choose a programme. Hence, the estimated effect of introducing the statistical treatment rule at the point in time where caseworkers have actually assigned unemployed to a programme is a reduction in the average remaining unemployment duration in the interval 9.0 - 19.2 weeks.

Note that when caseworkers are actually observed to assign unemployed to a programme, they may have faced a number of earlier choice situations where they have chosen *no training*, which is most often the best option. Nevertheless, when finally assigned to a programme, many of the unemployed are not assigned to the programme which is estimated to be most effective. Some reasons for the poor performance of caseworkers in this respect may be that they are not aware of what is the best programme, or that they have different objectives when assigning the unemployed to programmes; e.g. that they are making the assignment based on $Y_0$, rather than $Y_j - Y_0$, that is, out of equity rather than effectiveness considerations. As discussed previously they could also be more interested in other outcome variables than the duration in unemployment, i.e. the risk of repeated unemployment or the job quality match as measured by the wage.

6 Conclusion

The results shown in this paper clearly indicate the potential benefits from introducing statistical treatment rules to assist caseworkers in assigning unemployed workers to ALMPs. The estimated difference between the current system of caseworker assignment and the statistical treatment rule is in the range of a 9.0 - 19.2 weeks (14.6% - 31.2%) decrease in the
average remaining duration in unemployment. Effects in the range 3.8 - 6.5 weeks (11.0% - 17.0%) were found in an analysis of introducing the statistical model at the very first week of unemployment. These large impacts are obtained using a fairly small sample (for the study of heterogeneous effects, a much larger sample would be preferred), a few selected explanatory variables in the interaction, and a crude grouping of programmes. This suggests to us that a statistical rule may be worth serious consideration by policy makers.

There are many possible extensions of the current analysis. Among these are the introduction of finer categorization of programmes and more explanatory variables for the heterogeneous programme effects combined with estimation on the full population of unemployed workers. In addition, it would be natural to explore these potential gains further and eventually to conduct a randomized pilot study of the implementation of such a system.
References


CHAPTER 3

Using Statistical Treatment Rules for Assignment of Participants in Labor Market Programs
Using Statistical Treatment Rules for Assignment of Participants in Labor Market Programs

Jeffrey Smith* and Jonas Staghøj†

November 25, 2009

Abstract

Statistical Treatment Rules have in recent years been tested and implemented as a method to assign individuals into specific labor market programs in several countries. This paper discusses the basic problem facing the policymaker and the available solutions. We describe two different types of Statistical Treatment Rules called profiling and targeting and survey the available literature on the topic. Main results in the literature indicate that it is possible to identify the risk of long-term unemployment which is used to determine assignment in profiling models, but that this risk is not clearly related to the size of program impacts. Targeting models, on the other hand, assign individuals to programs based on estimated impacts and are hence directly aimed at improving efficiency. Simulation results suggest that targeting models can improve the assignment compared to the currently used methods, but more clear evidence from actual experiments is needed.

Keywords: Profiling, Targeting, Statistical Treatment Rules, Unemployment, Labor Market Programs

JEL Codes: J68, H00

*Department of Economics, University of Michigan. E-mail: econjeff@umich.edu
†Department of Economics, Aarhus School of Business, Aarhus University. E-mail: jonass@asb.dk
1 Introduction

There are at least two reasons to consider how to assign individuals to active labor market programs (ALMPs) in an optimal way. First, there is the question of finding the optimal program for a particular individual. And secondly, when one or more programs cannot be provided to everyone in the target population, the decision maker has to prioritize between the individuals. The two choices are of course also related. When prioritizing programs between individuals in the population, it would be natural to take into account what the impact of the program is for the individual and also what the best alternative program is. Different methods can be used to solve these problems and this paper surveys the literature about a particular type of methods called Statistical Treatment Rules (STRs). A STR provides a mapping from observed individual characteristics to specific labor market programs. We adopt the naming convention introduced in Frölich et al. (2003) to distinguish two different types of STRs called profiling and targeting. Profiling aims at providing programs to those unemployed who are most in need of being helped, whereas targeting assigns programs to those individuals expected to gain the most from participation.

Similar decision problems arise in many other areas and the explicit use of statistics in the solution method is by no means new. Psychologist Paul E. Meehl actually published a book called *Clinical Versus Statistical Prediction* as early as in 1954 and asked the question: Who is more likely to make better predictions - the clinician who knows his patients through close and personal contact, or the statistician who has only actuarial data at his disposal, but knows how to arrange them in a mathematical formula that will provide a predictive answer? The answer, which initiated a long lasting debate was, that statistical formulas almost always performed at least as good or better than clinicians. Although controversial, this conclusion has been confirmed in the majority of subsequent studies which is summarized in a meta-analysis conducted in the year 2000 by Grove et al.. They analyzed the results from 136 studies from a diverse list of research areas in (among others) medicine, psychology, crime, finance and labor markets. Results showed that 64 studies favored the statistical method, 64 showed approximately equivalent accuracy, and 8 favored the clinician. The general conclusion hence seems to be clear and suggests to us that the possibility of using

---

1 Taylor (1956).
STRs should be considered seriously in many situations. In this paper we analyze whether it seems to be appropriate to use STRs in the particular context of assigning unemployed workers to ALMPs. The availability of increasingly more detailed data only adds to the potential benefits of using statistical methods.

We first describe the basic problem facing the decision maker and define the relevant alternatives to STRs which are broad deterministic rules, random allocation, caseworker discretion and self-selection by the unemployed. We provide a brief summary of the theoretical advantages and disadvantages of the competing methods. The problem of assigning programs to individuals is complicated by the inherent uncertainty about future labor market outcomes, and we have therefore included a section to discuss the econometrics of the problem. This section considers the identification of the relevant parameters needed to construct different types of STRs. A randomized experiment would usually provide the relevant parameters, but as this is often not available, we also discuss the non-experimental estimators that have been applied. We describe how the estimation of the targeting method requires more detailed data compared to the more simple profiling method and discuss the type of heterogeneity that can be identified and used in the construction of different STRs.

STRs have been implemented or tested in a number of countries and we review the available results and seek to identify important factors which seem to affect the effectiveness of these methods. We discuss results from Australia, Canada, Denmark, Finland, Germany, Sweden, Switzerland, the UK and the US. Some of the consistent results are that it is possible to identify the risk of long-term unemployment which is used to determine assignment in profiling models, but that this risk is not clearly related to the size of program impacts. Targeting models on the other hand assign individuals to programs based on estimated impacts and are hence directly aimed at improving efficiency. Simulation results suggest that targeting models can improve the assignment compared to the currently used methods, but more clear evidence from actual experiments is needed.
2 The assignment of programs

2.1 Potential outcomes

Consider a situation where the unemployed can be assigned to one of \( R \) different labor market programs which are mutually exclusive.\(^2\) These programs could include job search assistance, subsidized job training, classroom training etc. and also the option of not being assigned to a program.\(^3\) The types of programs can hence be quite heterogeneous as they are supposed to help the unemployed workers with different aspects which are assumed to constitute a barrier to their reentrance into the labor market. Some ‘carrot’ programs are designed to increase skills while other ‘stick’ programs are designed mainly to act as a leisure tax in order to avoid moral hazard problems arising from insufficient search-incentives while receiving unemployment insurance. It hence seems natural to assume that impacts of participating in these programs can be heterogeneous with the impact depending on how well the program is suited to solve the individual’s specific problems. This idea of heterogeneous treatment impacts is the main reason why we want to analyze whether the present assignment into programs could be improved by using a STR. We now describe the basic problem facing the decision-maker.

Denote the option of no training by 0 and let \( T_j \in \{0, 1, \ldots, R\} \) be a dummy variable indicating assignment to a particular program for individual \( j \), where \( j \in \{1, 2, \ldots, J\} \) is an index for the individuals in the relevant population. Corresponding to each of the possible options in \( \{0, 1, \ldots, R\} \), we consider the potential outcomes to be the resulting labor market outcome conditional on assignment to a particular program and denote these as

\[
Y^0_j, Y^1_j, \ldots, Y^R_j
\]

The labor market outcomes are assumed to measure the relevant outcome of interest

\(^2\)This section draws on the work in Berger et al. (2001) as well as Frölich (2008).

\(^3\)In this paper we take the existence of ALMPs as given. Barnow (2009) offers a discussion of when it may be reasonable to offer public training programs instead of relying on markets to provide training. Individuals may for instance encounter capital constraints preventing them from appropriate training programs, and more generally, the individually chosen amount of training may not be equal to the social optimum when individuals have imperfect information, face externalities, have other goals or employ a different discount rate compared to a social planner.
for the decision maker. Examples of labor market outcomes could be earnings, duration of unemployment, probability of being employed at a particular point in time or more comprehensive measures like the present value of future earnings net of the costs associated with participating in a program.\footnote{Even more comprehensive measures could also include effects on health as analyzed in Huber et al. (2009a).}

In the absence of binding constraints on the number of slots available in the programs, the decision maker wants to assign each individual to the best program, that is, the program which maximizes the potential outcome\footnote{Manski has a series of papers on this topic that give a more detailed analysis of the decision problem. See e.g. Manski (2001, 2004, 2009) or his book from 2005.}

\[
T_j^* = \arg\max_{T_j \in \{0,1,...,R\}} Y^{T_j}_j 
\]  

(1)

None of the potential outcomes are, however, observed ex ante and hence the assignment has to be based on a prediction of the future. This prediction is at the heart of the discussion of different assignment mechanisms as this is where STRs are proposed to be able to improve on the existing methods. In STRs, the prediction of future labor market outcomes is based on historical data about the outcomes for previously unemployed workers. But even in the historical data we are only able to observe one of the potential outcomes for each individual as they are defined to be mutually exclusive. If an individual has actually participated in program $T_j = 1$, we will be able to observe $Y^1_j$, but not all the other potential outcomes which are hence counterfactual outcomes. To construct estimated counterfactual outcomes, we have to employ further identifying assumptions. We discuss identification and estimation issues in more detail in Section 3. The other related dimension in which STRs may be able to improve the allocation, which is the prioritizing between individuals, is discussed in Section 2.3 where different social welfare functions are introduced.

### 2.2 Alternative assignment mechanisms

An evaluation of STRs requires a comparison to other possible assignment mechanisms. In this paper we compare profiling (PRO) and targeting (TAR) to the following alternatives: deterministic rules (DET), random assignment (RAN), caseworker discretion (CASE) and
self-selection by the unemployed (SELF). Observed assignment mechanisms are often a combination of several of the mechanisms, but we start out by considering each mechanism in isolation.

All assignment mechanisms are basically rules providing a mapping for each individual into the set of available programs. The mapping may or may not depend on the observed characteristics of the individual. We loosely define a deterministic rule as a rule stating some broad criteria to be fulfilled in order to be able to participate in a program. Examples include eligibility rules which are based on a few observed characteristics like age, gender and labor market status. We denote this mapping by $\Phi^{DET}$:

$$\Phi^{DET} : X_j \rightarrow T_j^{DET}$$

The success of a rule like this should be based on the resulting outcomes $Y_j^{DET}$, for $j \in \{1, 2, ..., J\}$. The main advantages of deterministic rules are that they are easy to implement and operate and they may be able to take care of the prioritization between individuals when some programs only admit a limited number of participants. It may also be seen as an advantage that all individuals inside the defined groups are treated in the same manner. A deterministic rule may, however, not be particularly effective when impacts are heterogeneous. An example of such ineffectiveness could be if an unemployed worker is assigned to a program based on some broad characteristics even though this individual easily would be able to find a job on her own. Similarly, individuals who would actually have gained from participating in a program may be excluded from participation by the broad rules. Situations like this may occur when program impacts are heterogeneous and this naturally leads us to consider STRs, which should be able to be more customized towards individuals’ needs. STRs are in principle also deterministic rules, but we distinguish them out as another type of assignment mechanism because they include an additional step in the mapping into the set of programs. The additional step is to define an intermediate variable,

---

6Simple 'characteristics screens' also belong to this type of assignment mechanism unless it is specifically based on statistical knowledge in which case we would consider it to be a STR.

7Manski (2005, 2009) analyzes alternatives to this consequentialist approach including deontological welfare economics where actions may have intrinsic value, apart from their consequences. As an example, strict enforcement of a principle of 'equal treatment of equals' may require the use of deontological welfare analysis.
$Q_j^{STR}$, that summarizes the relevant information from $X_j$. In practice, $Q_j^{STR}$ is not observed and the first part of a STR is therefore to estimate the intermediate variable based on the observed variables. We denote this estimated intermediate variable as, $\hat{Q}_j^{STR}$, and the second part of a STR is to provide a mapping from $\hat{Q}_j^{STR}$ to the set of programs.

$$\Phi_{STR} : X_j \rightarrow \hat{Q}_j^{STR} \rightarrow T_j^{STR}$$

Again, the success of this mechanism should be based on the resulting outcomes, $Y_j^{TSTR}$, for $j \in \{1, 2, ..., J\}$. We consider two different STRs which differ in the choice of the intermediate variable, $Q_j^{STR}$. The first method is denoted profiling (PRO), and in this method the intermediate variable is set equal to the estimated value of the potential outcome if the individual is not assigned to a program. So $\hat{Q}_j^{PRO} = E(Y_j^0|X_j)$, and this measures the expectation of the outcome if the individual is left to herself. Individuals can then be ordered according to $\hat{Q}_j^{PRO}$ and the most intensive programs can be assigned to those unemployed with the lowest values. The use of profiling can be based on both equity and efficiency concerns depending on the relationship between $\hat{Q}_j^{PRO}$ and labor market outcomes. First, the decision-maker may want to spend the most resources on helping specific groups because they are assumed to experience the lowest welfare and because of an aversion to inequality across individuals. This argument for increasing equity presumes that the money is spent on programs that are actually beneficial for the unemployed. The profiling approach could also be based on efficiency concerns if it is generally the case that individuals with a low value of the $\hat{Q}_j^{PRO}$ have the largest impacts of participating in the programs. One argument supporting this relationship is that by focusing on the unemployed who are most likely to stay unemployed for a longer period, we should be able to decrease the deadweight loss caused by assigning costly programs to individuals who would find a job rather quickly even without participating in a program. Alternatively, policymakers could just wait and only allocate long-term unemployed into programs, but as it is often suggested that the skills of unemployed decrease while staying unemployed, then it would be more efficient to identify these individuals earlier on and provide the relevant help in the beginning of the unemployment spell.\textsuperscript{8} It has also been argued that this focus on individuals in risk of long-term unem-

\textsuperscript{8}Jackman and Layard (1991) argue that this could happen as a result of declining skills, loss of motivation,
ployment is effective because the inclusion of these marginalized individuals in the ordinary labor market would to some extent amount to an increase in the real labor supply, and hence decrease potential wage pressure. General equilibrium effects like this are hard to estimate and are not considered in the present framework, but should be kept in mind in situations where there are reasons to believe that they may be of particular importance.

The other STR we analyze in this paper is called targeting (TAR). This method aims directly at maximizing the effectiveness of the programs, by using information about all potential outcomes to estimate the intermediate variable which then allows a selection of the program with the highest expected outcome. Formally, 

\[ \hat{Q}^{TAR}_j = \max \{ E(Y^1_j - Y^0_j | X_j), ..., E(Y^R_j - Y^0_j | X_j) \} \].

In terms of efficiency, the targeting approach goes directly at the target without the additional assumption about a negative correlation between the intermediate variable and the impacts needed when using profiling. If impacts are heterogeneous in a more complex way than just depending monotonically on \( Y^0_j \), then targeting should be more effective than profiling. The additional effectiveness from using the targeting method compared to profiling is of course conditional on being able to estimate all the potential outcomes correctly and we will discuss this issue further in Section 3.2. Apart from the efficiency gains, targeting might also be seen as a way to ensure that individuals are treated equally, because all unemployed will be treated according to an objective rule. This would typically not be the case for the next assignment rule we will consider, which is caseworker discretion. The equality in treatment when using targeting might, however, be a little controversial as it at the same time implies that individuals with different characteristics may be treated in very different ways if their impacts are estimated to be very different. But this is basically a political question of how to define the groups within which individuals are supposed to be treated in the same way.

Caseworker discretion is often at least partly the method used today when assigning the unemployed to programs. Some eligibility rules may restrict the group of unemployed workers who can be assigned to programs, but then it is often up to caseworkers and perhaps

---

9 Fraser (1999). See also Johnson (1979) where treatment effects are analyzed in a model with divided labor markets and where the wage depends on the unemployment rate within a particular labor market.

10 A number of studies have shown that rather large differences may appear across caseworkers with regards to the program assignments made. See e.g. Bell and Orr (2002).
the individual unemployed to choose a particular program. Caseworker discretion can also be seen as a mapping from observed characteristics into programs. But caseworkers may include some additional information, $C_j$, which is only observable to caseworkers. The individual’s degree of motivation, some judgement of ability, general job readiness or updated information about the local labor market might for example be partly observed by caseworkers, but not included in the observed variables used in the econometric model. It may also contain information about the caseworker if for instance a particular caseworker is especially good at guiding unemployed through a particular type of program. We define this mapping as

$$\Phi^{CW} : X_j, C_j \rightarrow T_j^{CW}$$

The possibility of including the additional variables of what we could call local information is of course the main advantage for this method and caseworkers can react very flexibly to individuals’ needs. The main disadvantage of using caseworker discretion is that caseworkers may not be particularly good at realizing the relationship between the observed variables and the outcomes. They will presumably form their expectations about the potential outcomes based on the experience with previously observed unemployed. But even experienced caseworkers will only have observed a limited number of unemployed workers compared to the much larger number that could be included in a statistical model. The relationship between the observed characteristics and potential outcomes may hence be less precisely estimated by caseworkers who, on the other hand, can include some local information.\footnote{Meehl (1954) provides some interesting thoughts about how clinicians perform tasks that contain some implicit statistical modelling. Bishop and Trout (2002) state that unjustified overconfidence is one of the most robust findings in contemporary psychology, and so this may also be a problem for caseworkers when thinking about the quality of their judgements. See also Ægisdóttir et al. (2006) p. 371.} Caseworkers may also implicitly or explicitly, tend to think about unemployed in terms of different stereotypes. This may be particularly true if caseworkers face strict time and resource constraints, such that a detailed individual assessment is not possible. A related disadvantage which has been mentioned in the literature is that the individuals’ assignment to programs will depend too much of the expertise of the particular caseworker. Another risk when using caseworker discretion is the possibility of what has been called cream-skimming, which means that unemployed are assigned to a program based on their
probability of a successful outcome when participating. This method may not be efficient if those unemployed who are most likely to succeed after participating in a program would most likely have succeeded anyway.\textsuperscript{12} Cream-skimming may be a logical outcome in some situations where the incentive schemes for caseworkers encourage this kind of behavior.\textsuperscript{13}

A natural alternative to caseworker discretion would be to allow unemployed workers to self-select into programs, since individuals have access to local information as well.

\[
\Phi^{\text{SELF}} : X_j, C_j \rightarrow T_j^{\text{SELF}}
\]

Compared to caseworkers, we expect individuals to have even more information about personal skills and motivation, but they may be less informed about the situation on the local labor market. One way to allow unemployed workers to self-select into programs would be to use vouchers. The unemployed workers can then use vouchers to buy training from a list of training providers, and in principle, this arrangement may both be cheap and effective and at the same time imply increased competition between training providers. Empirically, the case for vouchers is less optimistic, and important problems seem to be informational restrictions for individuals as well as for the government outsourcing to training providers.\textsuperscript{14}

The individuals may not have adequate information about the gains they will experience from different training programs or perhaps more importantly, they may not chose programs that are in accordance with the social goals. The latter hypothesis is supported by results showing that voucher programs in general are quite popular among participants, even though it does not appear to be especially effective.\textsuperscript{15} Informational restrictions may also complicate the outsourcing to training providers since this may imply less control over what is actually provided. Therefore, it may in general be difficult for the social planner to design effective solutions to the principal-agents problems arising between the society on the one hand and

\begin{itemize}
\item \textsuperscript{12} In a situation where individuals can either participate in a program or not, cream-skimming can only be efficient if \(\text{cor}(Y^1, Y^1 - Y^0)\) is large whereas the profiling approach previously discussed only can be efficient if \(\text{cor}(Y^0, Y^1 - Y^0)\) is large.
\item \textsuperscript{13} See Heckman et al. (2002) or Barnow and Smith (2008) for analyses of performance standards. See also Heckman and Heinrich (2005).
\item \textsuperscript{14} McConnell et al. (2006) show that vouchers are more effective when combined with counseling. The counseling provides information about individuals’ own ability, training quality and employment prospects in the occupation and counseling constrains choices to reflect social goals rather than private goals.
\item \textsuperscript{15} Barnow (2009).
\end{itemize}
the combination of individual self-selection and training providers on the other hand.

The last assignment mechanism we consider briefly in this paper is random assignment.

\[ \Phi^{\text{RAN}} : X_j \rightarrow T^{\text{RAN}}_j \]

The mapping could depend on \( X_j \) if sample is stratified before randomization. This method may be used when a limited number of program slots have to be assigned to a population where program impacts are more or less homogeneous. Then we may see random assignment as the most fair approach to solve the problem of prioritizing between individuals, and furthermore this kind of assignment may allow for unbiased evaluations of the program impacts, which is a very valuable by-product. Random assignment is also easy to implement and operate, but with heterogeneous impacts it will in general be ineffective. This mechanism would neither be able to provide training for those most in need or those who are most likely to benefit from the programs.

### 2.3 How to evaluate the assignment mechanisms?

In the previous section we assumed that the labor market outcomes described by the potential outcomes could summarize all relevant information to the decision-maker. The definition of this potential outcome is hence important, but not always immediately obvious. Among other things, it has to be decided who the decision-maker is and what should be included in the outcome variable. If the decision-maker is thought to be a benevolent social planner, then it would probably make sense to define the welfare function in terms of the individuals’ utility which would then also reflect such things as the value of leisure time for the unemployed. If the decision-maker instead is the government, then a cost-savings approach without any measure of the value of leisure time may be appropriate. The key point is that the outcome variable is defined such that it is in accordance with the social agenda. In some cases it may be hard to define a single variable, but then we could instead use a summary index weighting different goals, such that the approach can in fact be made quite general.\(^\text{16}\) The definition of the relevant outcome variable needs to be done paying attention to the available data,

\[^{16}\text{This possibility is explicitly described in Frölich (2001).}\]
in order to get a variable which it is actually possible to measure with some precision and which at the same time is highly correlated with the goals of the programs.

Assuming that a reasonable outcome variable has been defined, the success of an assignment mechanism can be based on the resulting outcomes, \( Y_{jT_k} \), for \( j \in \{1, 2, ..., J\} \) and \( k \in \{DET, PRO, TAR, CASE, SELF, RAN\} \). With the introduction of uncertainty and perhaps also constraints on the feasible assignments because of a limited number of available slots in each program, we can no longer apply the simple solution in (1). In general we could instead imagine the decision-maker to evaluate outcomes based on a social welfare function

\[
W = \left( \sum_{j=1}^{J} Y_{jT_k} \cdot \tau(\cdot) \right) - Cost^k
\]

(2)

where we have included a cost term to reflect the fact that different assignment mechanisms cost different amounts to operate.\(^{17}\) \( \tau(\cdot) \) is a weighting function which allows the outcomes for different individuals to be weighted differently. A standard case often implicitly assumed in economic analyses is a utilitarian welfare function where the weights are equal for all individuals such that the total welfare is just the sum of individuals’ outcomes. This kind of welfare function is sometimes justified by an assumption of transferability of welfare between individuals. But in general we should allow the decision-maker to use a non-trivial weighting function. The profiling approach could for example be justified if the weighting function is decreasing in \( Y^0 \).

The decision-maker may also want to include some measure of uncertainty about future outcomes directly in the welfare function.\(^{18}\) This is not relevant ex post when the actual outcomes can be observed, but when evaluating different assignment mechanisms ex ante, it may be reasonable to take uncertainty into account. An example could be if a lower estimated outcome is preferred to a higher estimated outcome with a higher variance. This could be

---

\(^{17}\)One extension analyzed in Manski (2009) is to allow for multiple periods, thereby introducing learning motives for a decision-maker that has to assign programs based on partially identified outcomes. We leave such dynamic considerations implicit in our analysis.

\(^{18}\)A slightly different approach is taken in Manski (2004) where he makes the point that we should explicitly take into account that the variable \( W \) is maximized under uncertainty. He considers different approaches apart from the usual maximization of \( W \). A more conservative decision-maker may want to use a maximin rule to assure that the worst case is as good as possible. Or perhaps an intermediate minimax-regret rule which seeks to minimize the dependency on the uncertainty by minimizing the distance to what would have been the optimal decision in a situation with no uncertainty. See also Dehejia (2005) for a Bayesian approach.
incorporated in the social welfare function with a weighting function that is decreasing in \( \text{var}(Y_{T_j}^{T_k}) \) or by transforming the potential outcomes into certainty equivalent outcomes.

A more general point to take away from this subsection is that the evaluation of an assignment mechanism and the programs themselves are two different aspects although they are also related. We always observe the combined results from a specific assignment mechanism and some specific programs, but this may cover different stories about the effectiveness of the individual components. An effective assignment of poor programs or an ineffective assignment of good programs could show the same combined results. The traditional evaluation literature about ALMPs has focused on the performance of the programs, and it is somewhat surprising that there is still very little theoretical or empirical literature on assignment mechanism choice across programs. The choice of assignment mechanism, including STRs, could potentially help to cast light on the reasons why it is generally found that ALMPs have limited impacts on average.\(^{19}\)

3 Econometrics and data requirements

We build the discussion of some of the econometric details around three main points. First, we explain the important distinction between the estimation of statistical associations and the estimation of causal relationships in the context of STRs. Then we discuss the identification requirements when we are interested in causal relationships. And finally, we consider different types of heterogeneity and discuss when it is possible to estimate a credible and interesting STR.

3.1 Predictive or causal relationships?

To use a STR, we have to predict future labor market outcomes. The identification requirements vary for the different types of STRs. Profiling models are in general easier to estimate and require less detailed data compared to targeting models. A targeting model always requires the use of causal impact estimates and we discuss the identification of these impacts in the next section. In this section, we concentrate on profiling models and discuss

\(^{19}\)For general reviews of the effectiveness of ALMPs, see Heckman et al. (1999), Kluve (2006) and Card et al. (2009).
when it is sufficient to estimate predictive relationships. The literature about STRs has not always been particularly clear about how to interpret results from a profiling model, so we elaborate a little on this point in the following.

A profiling rule is based on $Y^0_j$, so we need information about this potential outcome for all individuals in the relevant population. At the time where the assignment to programs is carried out, $Y^0_j$ is unknown and is hence considered as a stochastic variable, which outcome is determined in the future. So we have to predict the value using the available information, which in this case is the observed characteristics, $X_j$. Let $Y_j$ be the observed outcome, equal to $Y^t_j$ if $T_j = t$, and consider a situation where the available data contains information on $Y^0_j$ for all $j$. Then it is fairly straightforward to estimate the historical relationship between $Y^0_j$ and $X_j$. This allows the construction of a profiling model that can be used to predict future values of $Y^0_j$ conditional of some $X_j$, given stability of the relationship over time. In a regression model, stability means that the coefficients of the model should not change over time, whereas the distributions of $X_j$ and $Y^0_j$ are allowed to change. When the relationship is stable, we do not necessarily need the profiling model to be causal since it is used for predictive purposes. A coefficient may in this case reflect a causal relationship or just reflect that the particular variable is acting as an approximation for variables not included in the model.

Given stability, a predictive model may be sufficient for profiling, but the difference between a predictive and a causal model has implications for the way results should be interpreted. In a predictive model we should be particularly clear that the relationship is estimated conditional on the current institutional environment. To illustrate this, consider as an example a situation where we want to estimate the probability of recidivism for a convicted criminal who are considered to be offered a release on parole. A regression of $Y$ on $X$ would not take into account that the current system already tries to identify those individuals who are most likely to commit a new crime and then orders these individuals to stay in prison. If detention in prison is an effective way to prevent recidivism, then we will not be able to interpret the estimated associations between $X$ and $Y$ as a causal relationship, because of the current selection of individuals who are offered release on parole. In this situation, a

---

low explanatory power of a predictive model based on observed characteristics may either reflect a genuine lack of relationship between $Y$ on $X$ or it may reflect the fact that the current prediction method is already using the information conveyed by the characteristic variables in an optimal way. To allow for a more complete understanding of the mechanisms involved, we would have to model the current selection directly, so as to ‘clear’ the data from the effects of the current system. This would require detailed knowledge about the current assignment mechanism.

As the above discussion indicates, the assumption about a stable relationship over time is crucial for the validity of the model. In a complex world where many variables affect the labor market, this assumption may be hard to maintain when important factors like institutional settings or the business cycle situation are changing over time. The profiling model may therefore need to be updated to reflect the current situation more accurately. But this raises another problem since the profiling model has now been in use and has thus contaminated the new data used to reestimate the model. If the programs assigned by the profiling mechanism are relatively ineffective, then it may be possible to ignore the problem and estimate a profiling model using the contaminated data. But this basically requires that those previously predicted to have a low value of $Y_j$ still have a low value even after being assigned to a program under the profiling mechanism. To circumvent this problem about contaminating future data, it may be reasonable to design the profiling model such that some fraction of the unemployed is left out from the profiling model. This will allow future updates of the model to be estimated on the subsample.

Until now, we have implicitly discussed how to use ‘reduced form’ models, and as the discussion shows, it is important to know what assumptions are invoked when the model is used to extrapolate to new periods or to new institutional settings. An alternative approach may be to estimate a more complete structural model where the selection issues arising from the decisions made by the involved economic agents are modelled directly. A valid structural model will in principle allow the policymaker to compare all different assignment mechanisms including some mechanisms which have not yet been put in place. In the example mentioned above, a structural model could be used to estimate the effect introducing a profiling model to make parole decisions instead of the currently used method. A structural approach is
demanding in the sense that it requires us to know the entire structure of the problem at hand, but on the other hand we are implicitly assuming much of same structure when we for instance use a reduced form model to simulate resulting outcomes under different assignment mechanisms. Whether the direct structural approach is feasible ultimately depends on the specific context and the available data.

The choice between predictive or causal models will depend on the context. One argument for using causal models could be that causal models, where more care is taken to estimate a well specified model, are less prone to changes in the economic environment, such that the stability assumption is more plausible. A pure predictive model, where the estimated coefficients partly reflects correlation with variables not included, will generally require the distribution of unobservables to be constant over time. This requirement may not be needed in a causal model where the error term can be assumed to be independent of the observed variables.

### 3.2 Identification of different STRs

To estimate a STR based on causal relationships, we need some exogenous variation in the assignment to programs. The exogenous variation is used to get unbiased estimates of counterfactual outcomes. For all individuals we would like to compare the actual outcome to what would have happened if the individual had participated in another program or no program at all. The counterfactual outcomes can only be estimated if we have data on comparable individuals who were actually assigned to the other programs, and this is exactly what the exogenous variation in assignments can assure. Different sources of exogenous variation can basically be divided into two groups: The Conditional Independence Assumption (CIA)\(^{21}\) and Instrumental Variables (IV)\(^{22}\).

The CIA approach relies on the availability of very detailed data. If we have access to all variables which jointly influence the assignment to programs and the outcome of interest, then it may actually be possible to estimate the relevant counterfactual outcomes. In the

\(^{21}\)This assumption is also called ‘selection on observables’ (Heckman and Robb, 1985), or ‘unconfoundedness’ (Rosenbaum and Rubin, 1983).

\(^{22}\)The term exclusion restrictions is also used to describe the identification approach used with this method.
context of a profiling model, identification requires the weak CIA which can be written as

$$Y_0^j \perp T_j|X_j$$  \hspace{1cm} (3)

This means that the potential outcome when no program is provided is independent of assignment to programs conditional on $X_j$. When this assumption is combined with support conditions which ensure that all individuals in principle could have been assigned to 'no program', we can estimate the needed counterfactual outcomes based on $X_j$.\textsuperscript{23} There are different ways to exploit this identification assumption including regression techniques and potentially more flexible estimation methods like matching or inverse probability weighted estimation.\textsuperscript{24} The CIA approach can also be used for estimation of a targeting model in which case we need the strong CIA,

$$Y_0^j, Y_1^j, ..., Y_T^j \perp T_j|X_j$$  \hspace{1cm} (4)

The support conditions then have be to strengthened accordingly to assure that the relevant counterfactuals can be constructed for all individuals.\textsuperscript{25} The CIA estimation approach is as already stressed a data-demanding strategy because it will not work unless we have access to all variables jointly influencing $T_j$ and $Y_j$. But with the increasing availability of high quality register data, we may argue that this is actually fulfilled in some situations. Frölich (2008) discusses the use of this assumption in much more detail and among other things he shows how to deal with a situation where only a subset of the variables is available for predicting future outcomes. He proposes an estimation strategy where a very detailed data set is constructed for the estimation of a causal model, but where it is not feasible to use all the detailed information when using the model for predictions afterwards. This situation

\textsuperscript{23}The formal support condition is

$$0 < \Pr(T_j = 0|X_j) < 1 \text{ for all } X_j \in \chi \text{ and } j \in \{1, ..., J\}$$

where $\chi$ is the relevant set of possible characteristics in the population.


\textsuperscript{25}Formally, we need

$$0 < \Pr(T_j = t|X_j) < 1 \text{ for all } X_j \in \chi \text{ and } j \in \{1, ..., J\} \text{ and } t \in \{0, 1, ..., R\}$$

where $\chi$ is the relevant set of possible characteristics in the population.
could arise for many reasons including cost considerations, data confidentiality reasons, or time delays in data availability. The solution is basically to integrate out for the variables that cannot be used for predictions conditional on the observed variables, but see Frölichs paper for much more details. A related problem may arise if (4) is not fulfilled, but we instead have that

\[ Y_j^0, Y_j^1, \ldots, Y_j^R \perp T_j|X_j, C_j \]  

such that the CIA is only fulfilled when we condition on \( X_j \) as well as \( C_j \) which is the local information observed by caseworkers. In a situation like this, we would estimate biased program effects by using (4) whereas caseworkers may be able to estimate the correct impacts by using their additional information, \( C_j \). We have to keep this possibility in mind when STRs are compared to caseworkers in a ‘simulation’ type of study, and this is discussed in further details in the results section.

As an alternative to CIA-based methods, we may use the IV approach. This method requires access to an instrument, \( Z_j \), satisfying the following assumptions\(^{26}\)

\[ \text{cor}(Z_j, Y_j|X_j) = 0; \]

\[ \text{cor}(Z_j, T_j|X_j) \neq 0. \]

The ideal IV situation is to have data from a well conducted randomized experiment where the instrument is constructed directly by the researcher, and this is often seen as the gold standard of evaluations. But as this kind of data is not always available, the evaluation literature has searched for other valid instruments. Sometimes it is possible to exploit certain institutional changes over time or specific program design features to isolate some exogenous variation. This is often called natural experiments to indicate that it is intended to work as an approximation for a controlled randomized experiment. It has also been argued that the time until programs start are assigned with some degree of exogenous variation which can be exploited using an approach called timing-of-events.\(^{27}\) This method is applicable when the outcome of interest is to be modeled in a duration framework.

\(^{26}\)See for instance Heckman et al. (1999) for a much more thorough analysis of IV estimators.
\(^{27}\)This method is developed by Abbring and van den Berg (2003).
3.3 Types of heterogeneity

The idea of using a STR is to predict program impacts based on observable characteristics, but how detailed heterogeneity should we estimate? We have already argued against the use of common impacts, but have not specifically considered the possibility of heterogeneous program impacts. STRs are, however, most important when impacts are indeed heterogeneous so in this section we discuss the types of heterogeneity it is possible to include in a STR.²⁸ We distinguish between three types of heterogeneity ranging from subgroup heterogeneity over individual heterogeneity based on observables to individual heterogeneity based on variables not observable to the econometrician.

If program impacts are heterogeneous across subgroups of the population but (approximately) homogeneous within these subgroups, then it would make sense to estimate impacts for the various major types of subgroups like age, gender and education, and this is indeed often conducted in many evaluation analyses. This simple type of heterogeneity can then be used to target programs to the subgroups with the largest impacts. But it is often only done in one dimension in a time, and this approach may not be sophisticated enough to actually capture more complex heterogeneous program impacts. Profiling can be seen as a model where the heterogeneity is modelled in one dimension based on the assumption that this dimension can be used to distinguish individuals with different impacts.

But as documented in Rudolph and Konle-Seidl (2005), many countries have started to move away from the assumption that program impacts are homogeneous within broadly defined groups. The subgroup approach can then be stretched to allow for a finer definition of the subgroups, such as defining subgroups along several dimensions at a time, and we denote this as individual heterogeneity based on observables. This allows impacts to depend on all variables observed for the individuals, such that an older person with low education and a relatively stable employment history can have an impact that differs from the mean effect among older individuals etc. The crucial point in the estimation of this type of heterogeneous impacts is that the heterogeneity is only based on observables. First, this allows us to estimate the impacts consistently, and secondly this allows us to use the estimated parameters

²⁸STRs may seem irrelevant when impacts are not heterogenous. Notice however that STRs may still help to prioritize between individuals in a situation where the number of program participants is fixed.
to predict impacts for individuals with specific values of the observed variables. The basic requirement needed when using the estimated impacts to construct a STR is that we do not only estimate average treatment effects on the treated (ATET), but can actually generalize to average treatment effects (ATE) by integrating over the distribution of $X$ in the relevant population.\footnote{See Heckman et al. (1999) for more details on these concepts.} In a context with one program compared to not participating in a program we need

$$\Delta_{ATET|X} = E(Y^1 - Y^0 | X, T = 1) = E(Y^1 - Y^0 | X, T = 0) = E(Y^1 - Y^0 | X) = \Delta_{ATE|X}$$

This equality between ATET and ATE is assured by the strong version of the CIA and the appropriate support conditions, and it allows us to predict the counterfactual outcomes. Within the context of IV estimation, an instrument allowing the identification of a local average treatment effect (LATE) is not sufficient for the purpose of constructing a STR. Without going into details about LATE estimates, this is basically an estimate of a program impact for a particular subgroup, namely those among the treated who are only treated because of the value of the instrument. That is, the individuals who are moved from the control group not participating in the program to the participation group by the instrument. See Imbens and Angrist (1994) or Heckman et al. (1999) for much more details about this widely used estimator. In this context we just mention the LATE estimator to stress that we actually need a quite robust source of exogenous variation to estimate the parameters needed to use STRs. Very specific instruments relating to the impact from a specific context which cannot be generalized to a broader group are not very helpful for the construction of STRs. But if all heterogeneity can be modeled in terms of observable characteristics, then the distinction between LATE and ATE estimates becomes irrelevant.

The third type of heterogeneity is allowed to be based on all individual characteristics, including variables not observed by the econometrician. If individuals are indeed heterogeneous in this unobservable way, then it may be impossible to estimate consistent impacts if individuals have selected voluntarily into the different program types based on personal knowledge about their specific impacts. So in this case, the exogenous variation in the as-
ignment to programs which is used for the estimation of unbiased impacts, may be hard to
isolate. A randomized experiment will still provide the necessary exogenous variation given
that it has been conducted in a proper way\textsuperscript{30}, but other sources of exogenous variation may
not always be convincing. If individuals either do not know anything about the part of their
impacts which is based on variables not observable to the econometrician, or if they do not
select into programs based on this information, then we may not face any problems even
under this specific type of heterogeneity. Knowledge about which type of heterogeneity it is
reasonable to assume in a specific context, depends on the institutions and in particular on
detailed knowledge about the selection into programs.\textsuperscript{31}

Apart from the above discussion of which kind of heterogeneity is actually present in
the situation, we may also face estimation problems caused by finite sample sizes when
estimating a STR. Individual interaction impacts may be hard to estimate with a sufficient
degree of precision since specific interaction terms often will be estimated based on a small
number of individuals. This should, however, be reflected in the estimated standard errors,
and hence we should be able to take this into account when constructing a STR based on the
estimated impacts. If the sample size problems become too large, then we may sometimes
be better off by only estimating the more simple subgroup impacts instead of the detailed
individual impacts based on observables.

To summarize, STRs are constructed using individual heterogeneity based on observable
characteristics and the crucial point for estimation is that we can capture all relevant het-
erogeneity by using interactions with observables. This allows us to identify the impacts and
extrapolate to predict potential outcomes for future individuals.

\textsuperscript{30}See Heckman and Smith (1995), Heckman et al. (1999) and Duflo et al. (2008).
\textsuperscript{31}See e.g. Heckman and Smith (2004) for an analysis of selection into social programs. Heckman, Urzua
and Vytlacil (2006) examine the properties of IV in models where program impacts are heterogeneous and
where individuals select into programs with at least partial knowledge of their idiosyncratic impact. Djebbari
and Smith (2008) contain an analysis of the Mexican conditional cash transfer program called PROGRESA
and show how to distinguish the different types of heterogeneity. In the particular example, they find
strong evidence for subgroup heterogeneity and modest evidence for heterogeneous impacts conditional on
observables.
3.4 Estimating the max of set

A STR that tries to identify the best program from a set of programs encounters a couple of potential problems arising from uncertainty. The first problem is a risk of biased results if we do not use out-of-sample predictions. The second problem is how to reflect the uncertainty of the model when doing inference.

A targeting model estimates the max of a set of discrete choices. And due to the order statistic nature of this estimator, the expected value of the max is not equal to the max of the observed estimates. To illustrate the importance of this result, consider an example where individuals can be assigned to one of \( n \) programs, all of which having a true impact of zero. The impacts are, however, estimated with some uncertainty, so suppose for simplicity that they are independently normally distributed with zero mean and equal variance, \( \sigma \). The graphs shown in Figure 1 show some results based on a simulation of impacts for 5,000 individuals.

The vertical axis in both graphs measures the simulated maximum impact as a function of the number of programs, \( n \), and the variance of the impacts, \( \sigma \), respectively. Even though the true maximum of the impacts is zero, it is clearly seen that the estimated maximum impact is approximately linearly increasing in \( \sigma \) and also increasing in \( n \). The graphs in the left part of Figure 1 show that each additional program essentially works as an option that is only exercised by the targeting model choosing the maximum if the realization of the stochastic variable is sufficiently positive. The relationship is concave such that the bias of an additional program increases most dramatically when the number of programs is small.

The example presented above is a stylized representation of a situation where targeting may be used, but it illustrates that we should be careful not to misuse spurious sample variation. It also indicates the likely result of comparing two hypothetical targeting systems with for instance 5 and 10 programs, respectively. Simulations for the targeting system with the higher number of programs will be more biased both because of the additional number of programs and also because of increased sample variation if the finer categorization also leads to higher variance of the impact estimates.
So we should obviously try to avoid the bias arising from spurious sample variation. Unfortunately, we cannot completely remove the bias by using the usual out-of-sample techniques since they address only part of the problem. We could for instance estimate the targeting model based on a sub-sample and reserve the remaining sub-sample for the evaluation of the model.\textsuperscript{32} The sub-samples may be chosen randomly or they may divide the sample in two periods based on calendar time. The latter choice of sub-samples would mimic the actual use of STRs where a model is estimated on historical data and used to estimate future outcomes. Such estimation procedures decrease the risk of overfitting and lead to more robust models that perform better in out-of-sample predictions, which is really where the models are supposed to perform well.

If the STR is to be evaluated in an actual experiment, then the above mentioned out-of-sample estimation procedures would suffice. But an additional problem arises when the performance of the model is based on simulated outcomes. Suppose for example that we use a robust model to simulate outcomes for an evaluation sample that has been left out for the estimation of the STR. For each individual, the STR will then use the simulated outcomes to identify the best program. But if we do not take uncertainty into account in this step as well, then the STR will be allowed to exploit spurious sample variation. A solution to this problem could be to use the following procedure. Simulate one set of outcomes and let the STR chose the best programs. Fix the chosen program for each individual and evaluate the performance of the STR over a number of new simulations of outcomes. If the STR exploits spurious variation in the first simulation, then this will show when evaluating the performance over a number of new simulations. If the simulations account for all uncertainty of the model, then this procedure should allow for unbiased estimates of the performance of STRs.\textsuperscript{33}

To sum up, we need to be careful to avoid biased estimates of the performance of STRs, especially when the model evaluation is based on simulations. The reason is that we need

\textsuperscript{32}If small sample problems are likely to be of importance, it may be necessary to use more sophisticated methods where a process of estimation and evaluation on different sub-samples is repeated. See for instance the approach used in Plesca and Smith (2005) or Payne and Payne (2000) where cross-validation and shrinkage correction methods are used when estimating a profiling model.

\textsuperscript{33}An alternative to the suggested approach could be explicitly to construct the STR to assign programs based on statistically significant impact estimates. This could at least decrease the bias, but would probably not be efficient in the sense that it does not exploit all information in the model.
out-of-sample techniques to estimate the model underlying the STR, and we need out-of-sample techniques to evaluate the performance of the model.

Uncertainty also has consequences for the interpretation of results. Statistically significant impacts can be identified from estimated impacts and standard errors as usual. In practice, it may, however, be worthwhile to think about how to present results to caseworkers (or individuals) in situations where the STR is supposed to work as an information system for those who assign individuals into programs. Frölich (2008) shows how to use a statistical method called Multiple Comparison with the Best (MCB) as a method to present results to caseworkers. The objective for the MCB method is to identify the best program within a set of programs, with a given significance level. If no program is significantly better than the rest of the programs, then the method will instead identify a set of programs which contains the best program. The MCB method can simultaneously be used to identify another set of programs which are significantly worse than the best program. This method may therefore provide a pedagogical way to present and interpret uncertainty of programs to caseworkers and serve as an alternative to the usual interpretations based on a t-statistics for each program. Apart from being intuitively attractive, the MCB method also solves a couple of potential statistical issues regarding t-statistics. First, if we test a large number of programs, then we should take this into account and adjust the significance level accordingly. Otherwise, we might judge some programs to be significant just because of chance when testing too many parameters. The MCB method is based on estimates for the joint distribution of impacts and therefore takes the number of programs into account. And secondly, if we are indeed interested in identifying the best program, then we should test a particular program against all other programs since we do not ex ante know which program is the best. The MCB also adjusts for this additional uncertainty about which program to test against.

3.5 A good STR

In the beginning of this paper, we mentioned two dimensions in which a STR could be valuable: choosing the best program for the individual unemployed and prioritizing the

---

34 This method was actually used in the Swiss targeting model called Statistically Assisted Program Selection (SAPS) which is described in the literature review. See Behncke et al. (2009) for this practical example and Horrace and Schmidt (2000) for a description of MCB with applications in economics.
assignment of scarce programs between unemployed. The variation needed to help solving these two problems differs. To solve the problem for the individual, we would like to have a STR with a high predictive accuracy such that the program selection can be done with less uncertainty. According to this dimension, predictive accuracy is a valuable feature of a STR, because fewer mistakes will be made. A profiling model offering a predictive accuracy of 80% may hence be preferred to caseworker assignment if they accomplish a smaller number than this. If we take costs into account, then a profiling model may even be preferred at much lower levels of accuracy, if it is cheaper than caseworkers.

A high accuracy does, however, not necessarily help prioritizing between individuals, because this basically just needs the correct ranking of the individuals. In this case a STR with a predictive accuracy of 80% may not necessarily be preferred to an alternative method with a smaller number. The relevant criterion is instead whether the model can rank the individuals according to how important it is to assign them to a program, as measured by how much this would increase the social welfare function described in Section 2.3. And the ability to rank individuals may not always increase with an increase in the absolute accuracy. This point should be taken into account when considering empirical results where predictive accuracy often is the measure of success.

The use of predictive accuracy as the relevant measure may also be problematic for other reasons. First, this measure will depend on the relative size of the subgroups in the population which is in focus. If for example we seek to identify long-term unemployed, then it is trivial to achieve a higher degree of accuracy in a population with only 10% long-term unemployed compared to a population with 30% long-term unemployed. More informative measures can be defined conditional on different subgroups, such as the probability of predicting long-term unemployment among those who are actually long-term unemployed. Secondly, if we have no information about the accuracy of alternative methods, then it is hard to evaluate whether a given degree of precision is acceptable. The STR may be inadequate, or the prediction problem may just be inherently difficult. But in the latter case, we

---

35 Properly cleaned from the effects of the current assignment mechanism.
36 See also Plesca and Smith (2005) for a discussion of this point.
37 Simply predicting everyone in the first case to be short-term unemployed gives a correct prediction for 90% of the individuals whereas a similar strategy in the second case only achieves correct predictions for 70% of the individuals.
should still use the STR if it is the best available method even though it may be imprecise.

The social welfare functions discussed in Section 2.3 are based on outcomes. So how do we get the relevant outcomes that are needed to compare different assignment methods? Basically, by the same methods used to identify program impacts, namely by exploiting exogenous variation. The ideal case would again be to have data from a randomized experiment where the unemployed are randomly selected to be assigned by the different assignment methods we want to compare. And in the absence of a randomized experiment, we would have to argue that some exogenous variation could be exploited to identify the relevant counterfactual outcomes. Apart from randomized experiments, the only method which has been used is to simulate what would have happened under an alternative assignment method.\textsuperscript{38} The simulation method is obviously not as robust as a randomized experiment, but may nevertheless provide some indications of the size and direction of the results, which could then perhaps open up for a randomized pilot experiment.

The idea of this section is to stress that data availability is important for the design of STRs, and as the last section showed we would ideally need to have access to two randomized experiments to implement and estimate the effects of a STR, indeed a quite hard requirement. A predictive profiling model may only need one randomized experiment if the estimation of the STR itself can be based on administrative data. We have not discussed the use of different functional form assumptions and estimation techniques, because these appear to be of less importance compared to the quality of the data.\textsuperscript{39} Note also, that the use of specific individual variables in a model should be legal, in the sense that they are not prohibited by anti-discrimination policies. In the US and Canada this has been a real problem since some variables like age, race, ethnic group, sex, color, national origin, disability, religion, political affiliation and citizenship have been excluded from the variables used to estimate STRs.\textsuperscript{40} Pope and Sydnor (2009) discuss this problem and show how socially unacceptable predictors are likely to be correlated with other variables included in models. This implies

\textsuperscript{38}If we take a structural model very seriously, then we may not need further evidence from randomized experiments, since the model basically reflects the truth. We prefer to label the structural extrapolations as simulation studies explicitly to distinguish them from actual experiments.

\textsuperscript{39}See Black et al. (2003a) for more detailed recommendations about estimation issues in the US profiling model called Worker Profiling and Reemployment Services system (WPRS). Sullivan et al. (2007) provide an updated analysis.

\textsuperscript{40}See Wandner and Messenger (1999). There is a related literature about ‘racial profiling’. See e.g. Persico (2002) where equity and efficiency goals are analyzed in the context of policing.
that we cannot simply exclude specific variables since a STR will still implicitly discriminate based on the socially unacceptable variables. An inefficient solution could then be also to exclude all variables that are correlated with the socially unacceptable variables from the model. Pope and Sydnor (2009), however, suggest that we could instead include all variables when estimating the model, to clean the data from their effects. And then we may set the coefficients on the socially unacceptable variables to zero when making predictions. They show that this approach is efficient since it makes use of all available information which is not based on socially unacceptable variables.

4 Literature review - results and experiences

In this section we review the existing literature about the use of STRs to allocate unemployed into ALMPs. We first consider the profiling type of models and then move on to targeting models before we try to summarize factors which seem to be correlated with a good performing STR. A general issue with the reviewed results is, that it is hard to find clear-cut studies giving a complete evaluation of a STR compared to the alternatives. Results from a number of studies can, however, be put together to at least give us some indication of the value of using STRs. Earlier reviews of the literature can be found in OECD (1998), Ramboll PLS (2001), Hasluck (2004) and Bimrose et al. (2007).

4.1 Results with profiling

The profiling type of models is the simpler type of STRs and it is hence also the model most countries started out using. The US, the UK, Australia, Canada, Denmark, Finland, Germany, the Netherlands, South Korea and Sweden have all either experimented with or actually used models which can be seen as profiling models. Evaluation of profiling models requires knowledge about the goals of the system and knowledge about the performance of alternative methods. Unfortunately, the available studies all have insufficient information about at least one of these issues.

Consider first the Worker Profiling and Reemployment Services (WPRS) system which was introduced in the US in 1993. This system is probably the most thoroughly investigated
profiling model. In WPRS, the profiling variable is a measure of the risk of long-term unemployment. A national statistical model was developed as a baseline for the individual states which were required by a new law to use some sort of systematic sorting of the unemployed before referring them to reemployment programs. Some states used the nationally developed model directly, while others developed their own models. Most states have chosen to rank the pool of unemployed according to the probability of benefit exhaustion, and reemployment programs are then provided to individuals according to this list starting with those in greatest risk of benefit exhaustion. The system is still in use, which may be seen as an indirect indicator of the system’s value. A more direct evaluation of the system runs into the problem that it is not entirely clear whether the WPRS is supposed to fulfill equity and/or efficiency goals. The available evidence focuses mostly on the accuracy of the model which is reasonable if the main goal is equity. These evaluations may, however, not be especially informative about the efficiency of profiling. Sullivan et al. (2007) provide a recent overview of the history and previous evaluations of WPRS. An early evaluation concluded that 'The models clearly identified claimants who were most likely to exhaust their benefits' (Hawkins et al., 1996, p.III-10) but a later study resulted in a more pessimistic interpretation: 'However, the targeting power of the model is modest... Exhaustion seems to be very difficult to predict accurately with available demographic and labor market data.' (Olsen et al., 2002, p. 53). The apparent differences between these conclusions arise because of inadequate information about the performance of alternative allocation mechanisms. This leaves more room for a subjective interpretation of whether a certain accuracy is actually good or bad.

Sullivan et al. (2007) show how to construct a metric that specifically compares the accuracy of the profiling model in WPRS to random predictions. They consider the perfor-

---

41 Sullivan et al. (2007) explain that 50 out of the 53 implemented models (50 states plus the District of Columbia, Puerto Rico and the Virgin Islands) use benefit exhaustion as the profiling variable. Alternative variables include benefit duration and proportion of total benefits claimed. Most states define exhaustion as claiming 100% of the available benefits, but some states use a 90% threshold level.

42 The Unemployment Compensation Amendments of 1993 which mandated that states identify unemployed workers likely to exhaust benefits and refer them to reemployment services, were at least partly based on an experiment conducted in New Jersey (New Jersey Unemployment Insurance Reemployment Demonstration Project). Results from this experiment indicated that reemployment services could be successfully targeted towards a group with somewhat greater reemployment problems and the results also showed some evidence that treatment impacts were higher for this group. This may have lead policymakers to believe that profiling could fulfill both equity and efficiency goals. See Corson and Haimson (1996) for more details about the experiment.

43 More evaluations of WPRS are available in Wandner and Messenger (1999), Dickinson et al. (1997) and Dickinson et al. (2002).
mance of 28 profiling models for which they have data, and the profiling model used in New York is found to be the most accurate when comparing to random allocation. In New York 40.4% of the unemployed workers are expected to exhaust their benefits. If the profiling model was used to identify this fraction of the unemployed based on their profiling scores, then the resulting exhaustion rate among the profiled individuals would instead be 55.5%. While not nearly a perfect profiling model, the model is clearly better than random prediction. The average impact over all 28 profiling models considered in the evaluation is 5.7 percentage points. These comparisons to random allocation show the performance against one alternative that is easy to understand. The obvious problem is that some researchers or politicians may see this test as being too easy to pass and therefore not especially relevant. It would therefore be highly relevant to gain more empirical evidence about the performance of the alternatives, especially the performance of caseworker assignment since this is the most widespread alternative.

Black et al. (2003b) exploit a design feature of the implementation of WPRS in Kentucky to get some exogenous variation that can be used to evaluate the system. In Kentucky, individuals were placed in 20 groups based on their expected unemployment duration. Individuals predicted to claim 95 to 100 percent of their UI benefits were placed in group 20, those predicted to claim 90 to 95 percent were placed in group 19, etc. Programs were then referred to the groups in descending order and in some weeks the number of available slots in reemployment programs implied that not all individuals within a group could be referred to a program. The prioritization within the group was then solved by randomization. This provides the variation used in the paper and reemployment services are actually found to be relatively effective.

To investigate whether the task of profiling adds value, Black et al. (2003b) estimate

---

44Sullivan et al. (2007) revise and update the profiling models for some states when the required data is accessible to them. This leads to an average difference between random allocation and profiling models of 8.5 percentage points.

45The average duration of UI benefit is reduced by 2.2 weeks, the mean UI benefits claimed received is reduced by $143 and earnings in the first year following the start of a program increase by more than $1,050. These results are not interpreted as average impacts for all unemployed, but rather as marginal impacts for particular subgroups around the points where the randomization took place. But a subsequent paper by Black et al. (2007) actually shows that a generalization of these results to a broader group of unemployed is probably not too far from the truth. Considering the dynamics of the program impacts, Black et al. (2003b) show that most of the impacts arise from an increased exit from unemployment immediately after unemployed workers receive a notice that they have been selected to participate in a program.
the program impacts for different values of the profiling variable. This exercise shows that there is no clear relationship between the profiling variable and the size of the impact, and profiling is therefore not found to be efficient in this study. It seems likely that the results from WPRS would have been similar even without the initial task of profiling the unemployed before referring to programs. This result is consistent with the results described in Peck and Scott (2005), where a small randomized experiment is conducted in Arizona. An instrument called the Case Management Screening Guide (CMSG) is used to categorize the unemployed into three groups based on scores calculated from a survey filled out by the unemployed. The conclusion from the randomized study is that the instrument is capable of sorting unemployed according to their employability, but the screening has no impact on employment outcomes.

Eberts (2002) describes another small experiment conducted in Kalamazoo, Michigan, and since this experiment was actually designed to evaluate alternative assignment mechanisms, we will discuss the results in more details. In the experiment, unemployed workers could be assigned to one of three training providers. A profiling model was estimated to predict the probability to find a job and hold it for 90 days. Caseworkers were asked to suggest the most appropriate relationship between the profiling variable and the three types of programs. Caseworkers suggested to categorize the unemployed workers into three groups of roughly equal size according to the estimated probabilities. Unemployed workers with low $Y^0$ were assigned to the most intensive program in terms of hours of participation and unemployed workers with high values of $Y^0$ were assigned to a program where the philosophy of delivering services was more self-directed and self-paced. The middle group of unemployed workers were assigned to the third program. To see whether this assignment based on $Y^0$ was effective, unemployed workers within the three programs were then randomly assigned to a treatment group and a control group. The treatment group would participate in the program they were assigned to and the control group would randomly be assigned to one of the three programs. This design was specifically chosen to allow for a comparison of the profiling model to random assignment. Results from this experiment show first

---

46 These providers delivered somewhat similar services but ‘... providers differed in their style and philosophies in delivering services and in the number of hours in which participants were engaged in specific activities’ (Eberts, 2002, p. 24). The average number of hours for the three programs were 7.3, 11.2 and 16.0.

that the profiling model was quite accurate in identifying unemployed workers with different probabilities to find and hold a job. And secondly, the results show that the assignment based on the profiling model was actually also relatively efficient, such that different types of unemployed workers defined by their $Y^0$ values were actually assigned to the most effective program. This result was found by considering alternative combinations of the three groups of unemployed workers defined by the profiling model and the three types of programs. The mapping suggested by caseworkers resulted in a job retention rate that was 27% higher than the average of the alternative combinations. These results can of course not say whether even more efficient assignments could be made if a targeting model was used, but they do show that a profiling model may help to enhance efficiency in situations where it is possible to make a reasonable mapping from the profiling variable to the available programs. In situations with a larger number of more diverse programs, it may be more questionable whether the mapping into programs can be made efficiently based on a single dimension, like the probability to find and hold a job.

Australia has quite similar experiences with profiling models. In Australia, profiling models have been used since 1994 to screen the unemployed according to their estimated risk of long-term unemployment.\textsuperscript{48} The most recent profiling model called Job Seeker Classification Instrument (JSCI) has been in use since 1998. This model is partly based on an estimation of a logit model for the probability of staying unemployed for 12 months conditional on a list of individual characteristics, and partly based on a subjective judgement. Only statistically significant parameters from the logit model are used, and these are converted into points for each characteristic. Additional characteristics are then added as a result of advice from a Classification Working Group and from major stakeholders.\textsuperscript{49} The resulting profiling model is therefore not a pure STR. The exclusion of statistically insignificant variables and inclusion of other variables not included in the model seem inefficient from an econometric point of view. These efficiency losses should be compensated for by increased transparency and

\textsuperscript{48}The early intervention strategy implemented in 1994 was initially based on two instruments, the Job Seeker Screening Instrument (JSI) and the Client Classification Level (CCL). JSI constituted the profiling part of this mechanism, as this was based on estimated probabilities of long-term unemployment. See OECD (1998) and DEWRSB (1998) for more details.

\textsuperscript{49}The Classification Working Group include caseworkers and occupational psychologists. The major stakeholders include representatives from case management organisations and the employment service industry.
increased goodwill from the caseworkers and managers that are going to use to system. The original JSCI developed in 1998 contained 18 variables. Since then, the model has been updated in 2001, 2003 and 2008 to reflect changing conditions in the business cycle as well as the institutional environment. These updates may be appropriate to the extent that the changing environment has a large impact on the identification of long-term unemployed workers. But as the updated models do not appear to have taken into account that the new data have been contaminated by the use of previous profiling models, it is not necessarily clear that the updated models will be more accurate than the original. This again points to the possibility of leaving aside a subsample of individuals which can be used for future estimations of the profiling model and to design the system in such a way that it is possible to evaluate the accuracy of different models. The available evidence about the accuracy of JSCI is similar to the US experience in the sense that it is generally believed to be relatively accurate, but that it is hard to find numbers for the accuracy especially if we want to compare to alternative methods. DEEWR (2009) contains the most recent evaluation and is based on submissions from Australian Government employment services providers, employment services peak industry bodies and a range of community organizations and is hence more a representation of opinions than an actual measure of accuracy. This evaluation of profiling is quite positive. More solid evidence about accuracy is found in Lipp (2005) where it is quite clearly demonstrated that the profiling score from JSCI is positively correlated with the proportion of unemployed workers who enter long-term unemployment. This appears to be one of the studies that most clearly demonstrate the possibility of using a profiling model to accomplish equity goals, but it does not show whether the efficiency goal is obtained as well. And it does not show the accuracy of alternative allocation mechanisms.

50 The transparency of the model has the possible disadvantage that to some degree it is possible to manipulate the JSCI score. Unemployed workers can learn the points assigned to different variables directly from an official webpage. The cut-off score for the profiling model has been changed a couple of times during the years and a spike in the distribution of the JSCI score is consistently seen around the cut-off score (Lipp, 2005). For some reason, this problem, however, seems to have disappeared since the introduction of the latest JSCI version in 2003.

51 See O’Connell et al. (2009) and DEEWR (2009).

52 This relationship is found even without taking into account that individuals with high profiling scores from JSCI will receive more assistance in order to find a job. If this assistance has a positive effect on the job finding rate, then the actual correlation between the profiling score and the risk of long-term unemployment should be even higher than what is found in Lipp (2005).

53 Lipp (2005) also contains an interesting discussion of a recent experiment where information about the motivation of the unemployed is collected. The use of this additional measure in the model may improve
The final evidence on profiling models introduced on a full scale is from Denmark, where a profiling model was implemented in 2004.\textsuperscript{54} The model estimated the individuals' probability of staying unemployed for at least 26 weeks after entering unemployment.\textsuperscript{55} This information was then given to caseworkers in the form of a 'Job barometer' dividing individuals into three categories representing different degrees of risk. Caseworkers could use the information or not. The system was never really accepted and used by caseworkers and was abandoned before a statistical evaluation of the performance was made.

Apart from the evidence from actual implementations and experiments, a number of studies have analyzed the potential benefits from profiling models in other countries. A number of studies present their results in the form of a 2 by 2 matrix with model predictions in one dimension and actual outcomes in the other dimension. We present some of these results here, but note that it is generally hard to compare for instance the overall percentage of correct predictions across the studies since they depend on a number of context-specific issues, including the actual fraction of long-term unemployed in the sample and the method used to choose a cut-off value when predicting long-term unemployment based on probabilities. The cut-off level should ideally depend on the costs associated with the different types of errors.

Consider first the Danish profiling model described above. The actual use of the model was not evaluated, but Rosholm et al. (2006) contain some numbers about the predictive power. In a population with 40\% long-term unemployed (LTU), the cut-off is chosen such that 29\% of the unemployed workers are predicted to become LTU. The overall fraction of correct predictions with the profiling model is 66.4\%. This compares to a random selection of 29\% of the unemployed workers which would result in an overall fraction of correct predictions of 54.2\%. Among the actual LTU, the profiling model correctly identifies 43.9\% whereas random selection would correctly identify 29\%.

Similar results are found in Wong et al. (1999, 2002) where the possibility to predict LTU is analyzed using data from Canada. In this study, LTU is defined as staying unemployed for more than 52 weeks, and a profiling model is estimated using a probit model. In the predictive performance of the JSCI.

\textsuperscript{54}See Rosholm et al. (2006).

\textsuperscript{55}The underlying econometric model was a duration model for the duration in unemployment. This model was then used to estimate probabilities of long-term unemployment.
sample, 21% of the unemployed workers are LTU. The cut-off used to convert estimated
probabilities into predictions is set such that only 5.8% of the unemployed are predicted to
become LTU. The overall fraction of correct predictions for the profiling model is 79.3% and
randomly predicting 5.8% of the sample to be LTU would result in 75.6%. Among the actual
LTU, the profiling model correctly identifies 15.0% and random prediction would correctly
identify 5.8%.

Payne and Payne (2000) use data from the UK to analyze the predictive power of a
profiling model based on a logit model for the probability of staying unemployed for more
than 12 months. The sample contains 25% LTU. They show results for different cut-off
values, and here we show the results for a cut-off level that predicts 9% of the sample as LTU.
The overall fraction of correct predictions for this profiling model is 75% which compares to
70.6% using random selection. Among actual LTU, the profiling model correctly identifies
19.2% and a random prediction identifies 9.0%. The study is done on a fairly small sample,
but Bryson and Kasparova (2003) reach similar conclusions based on another set of data from
the UK. They use a methodology similar to the one used in Lipp (2005), which is to analyze
the actual probability of LTU for different values of the profiling variable. Again, a clear
correlation is found, such that individuals with higher profiling scores have a higher risk of
LTU. They also consider the accuracy of profiling models for different groups of unemployed
workers. Interesting, they find that profiling models are more accurate for groups thought
to have a lower contact with the labor market. Specifically, they find profiling models to be
more accurate for sick and disabled unemployed workers. To the extent that the employment
prospects for these unemployed workers are dominated by other factors not observed in the
data, then it is somewhat surprising that the statistical profiling models are more accurate for
this group. One reason for this may be that there is more variation in employment outcomes
for this group, since this would usually allow for more accurate statistical predictions.

Finally, Germany, Sweden and Finland are all in the process of testing actual profiling
models. In Germany, a profiling experiment has been conducted as a part of the Treatment
Effect and Prediction Project (TrEffRe). A profiling instrument was used in combination

\[56\text{‘Treffer’ is German for hitting the target. The project includes profiling as well as targeting models and the latter is discussed in the next section. For more information about the profiling part, see Rudolph (2005) and Rudolph and Müntnich (2001).}\]
with caseworker discretion to categorize newly unemployed workers according to their risk of long-term unemployment. Hence, it is not possible to evaluate the profiling instrument in isolation, but Rudolph (2005) presents some results showing that both caseworkers and the statistical instrument are relatively successful in identifying the long-term unemployed. The results also indicate that the statistical model outperforms caseworkers. Finland and Sweden are both reported to be in the process of testing pilot profiling models. In Finland, a profiling model was implemented based on Moisala et al. (2006). The model predicted the duration in unemployment based on information available when the unemployed first register. This was used to present an index of risk of long-term unemployment to caseworkers. No formal evaluations of the model are available, but caseworkers are reported not to pay much attention to predictions from the model, since they find it to add no value. An experiment was also conducted in Sweden. Evaluation results are not yet available, but a follow-up survey showed that caseworkers generally found the system to be helpful although it was not considered a major break through. The profiling model is meant to be implemented nationwide as soon as it is practically feasible. Bennmarker et al. (2007) present the underlying model and have some results on the potential accuracy. They find the profiling model to be relatively successful in predicting who will stay unemployed for at least 6 months. They also show that a categorization based on subgroups is less precise compared to the profiling model.

4.2 Results with targeting

The development of targeting models is more demanding, but the US, Canada, Germany and Switzerland have all experimented with this type of STR. Results from these experiments are, however, only available for Switzerland, since the targeting models used in Canada and the US were abandoned before evaluations were possible and the German targeting experiment is still ongoing. An implication of this is, that we have to rely more on evidence from simulation studies where the potential performance of targeting models have been evaluated based on estimations and extrapolations of economic models. This type of studies have been carried

---

57 Personal communication with Roope Uusitalo and the Ministry of Employment and the Economy. See also Viinikainen (2007).
58 Personal communication with Anders Forslund.
out using data from the US, Denmark, Germany and Switzerland.  

Consider first the targeting model called Statistical Assisted Programme Selection (SAPS) which was tested in a pilot experiment in Switzerland. Behncke et al. (2009) describe this experiment which seems to provide the ideal way of evaluating the importance of the assignment mechanism in isolation. The randomization was carried out in such a way that a random sample of caseworkers gained access to the predictions from the targeting model. The model provided predictions of the number of months in stable employment in the next 12 months, conditional on participation in different programs. Caseworkers could then use this additional information when assigning unemployed workers to programs. Unfortunately, the experiment did not provide an answer to the question about the performance of the targeting model, because caseworkers did not use the information provided by the targeting model. When allowed to use the additional information or not, caseworkers simply choose not to act according to the model predictions, and so the experimental estimates were predestined to be substantively insignificant. Although this was a disappointing result from an ambitious attempt to conduct an actual experiment, this study does provide some important lessons. Most importantly, the implementation of a STR is found to be extremely important. To get caseworkers to use the information provided by the targeting model, caseworkers should either be given less discretion or some sort of incentive schemes should be designed to ensure that the STR is actually used. Hopefully, if the system shows the ability to improve the impacts of the programs, then caseworkers will be less resistant.

A study called Frontline Decision Support System (FDSS) was supposed to provide a similar experimental evaluation of a targeting model providing caseworkers with information about estimated program impacts. The experiment took place in selected sites in Georgia, US, and it is described in Eberts et al. (2002). It was, however, abandoned before the effects of the system could be estimated. W. E. Upjohn Institute for Employment Research helped develop the targeting model, and according to their webpage the system was abandoned partly because of 'bad timing' and partly because of changes in priorities by a new senior executive.

---

59 More generally, all evaluations finding heterogeneous program impact estimates implicitly make the case for targeting as a more efficient assignment mechanism, provided that heterogeneity can be predicted from observable characteristics.

60 An employment spell was characterized as stable if the duration was at least 3 months.

61 www.upjohninst.org/fdss/index.html

62 The system was introduced just before the business cycle turned downwards. The downturn resulted in
management in the Georgia Department of Labor.

A targeting model was also estimated in Canada and this model was supposed to provide a 'user-friendly vehicle for letting scientific research inform the management and practice of employment service delivery'\textsuperscript{63}. The system was called the Service and Outcome Measurement System (SOMS) and was designed to provide caseworkers with detailed information about individual unemployed workers, including estimates of the most effective employment programs. SOMS was developed and tested and was actually ready for a full-scale introduction. Unfortunately, the ambitious project was abandoned before it was really introduced. Colpits (2002) states two main reasons for the failure to implement the system. First, some caseworkers opposed the system because they either did not trust the results or because they were afraid that the system would take over their jobs. The latter reason may have been reinforced by the timing of the introduction because it happened during a period with major reductions in the number of caseworkers employed. Secondly, there were concerns about data security because of the merging of detailed individual information into the database used to construct the STR. Many of the lessons learned in the process of developing and implementing the SOMS are important for other countries who want to design a targeting system.

Finally, in Germany, a targeting model was developed as a part of the TrEffeR project already discussed in the review of profiling models.\textsuperscript{64} The first part of the TrEffeR project aims at establishing a large database with detailed individual information which is now used to provide semi-annual estimates of the effectiveness of all available labor market instruments for each of the 178 regional labor market agencies. The second part of the TrEffeR project uses this information to develop a targeting tool (Produkteffekte auf Kunden - PeaK). The impacts underlying this targeting tool are estimated by matching techniques relying on the CIA. This identification strategy may be reasonable given access to very detailed data that contain most variables jointly determining program participation and outcomes. A detailed description about the matching method used is not yet available. A potential problem with the method is, however, that the control group appears to be constructed based on an increased pressure on the caseworkers such that they did not have the necessary resources to operate the new system.


\textsuperscript{64}Information about the targeting part of the TrEffeR project can be found in Stephan et al. (2006).
individuals that will never participate in a program. This comparison may be relevant in the sense that results are easy to interpret, but as discussed in a number of papers this type of comparison group may differ because of selection. Unemployed workers who do not participate in a program for a long period are more likely to be disadvantaged and hence not judged by caseworkers to be fit for participation in a program, or they may have found a job starting in the future and were excluded from program participation for this reason. If the control group is selected in ways not accounted for in the matching method, then results may be biased. Notice however, that the targeting model still may be able to estimate the ranking of the different programs if the bias is the same for all programs. The introduction of the targeting tool is still in the experimental stage, and no actual results have been published yet. Arnkil et al. (2007) do, however, state that the initial pilot experiment seems to have been a success. Simulation studies based on the initial pilot experiment indicate that program impacts are larger if assignments are made according to the targeting model.

The above mentioned experimenting with targeting models was initiated because several simulation studies had shown the potential benefits from targeting models. To start out, Plesca and Smith (2005) estimate the performance of different types of assignment mechanisms including profiling and targeting models. They use experimental impact estimates from the National JTPA Study which allow for an unbiased estimation of program impacts, but the performance of the different STRs still has to be based on a simulation of what would have happened if different STRs had been followed. The performance of STRs are found to be somewhat mixed. For adult males and females, a targeting model based on estimated impacts seems to be able to improve average program impacts. This is, however, not true for young people, where caseworker assignment and random assignment are actually found to perform better than targeting. An important point made in the paper is to design the estimation procedure to avoid overfitting problems. If this is not done in these simulation type evaluations, the results will be biased towards estimating too large effects. The performance of profiling models is found to be less efficient than targeting in most circumstances,

---

66 Personal communication with Gesine Stephan and Torben Schewe.
67 The National Job Training Partnership Act Study was supposed to measure the benefits and costs of selected employment and training programs for economically disadvantaged adults and out-of-school youths. The experiment was based on random assignment for around 21,000 individuals during the years 1987-1989. Bloom et al. (1997) describe the key findings from the experiment.
and the general conclusion is that targeting models may imply increased program impacts compared to the present assignment mechanism, which is caseworker assignment.

A similar simulation exercise was conducted using Swiss data in Lechner and Smith (2007) and the results in this paper are more positive with regards to the performance of targeting. In this study, the program impacts are estimated using a matching estimator, and the results hence rely on the CIA. The authors argue that this assumption is plausible given the detailed administrative data which include a wealth of individual information, including assessments made by the caseworkers about the individual’s employability. They also stress that the simulation approach has some important limitations, and note that the results should not be seen as definitive results, but rather as suggestions about the value of a targeting model. The results are, however, quite clear. Caseworkers are found to obtain roughly the same post-program employment rate as random assignment of the unemployed to programs (or no program). The use of a targeting model is found to be able to increase the post-program employment rate significantly. Note that there is no correction for in-sample bias in this study. This means that the performance of the targeting model will be somewhat overstated as discussed in Section 3.4, but given the large difference between caseworker assignment and the targeting model, this in-sample bias is unlikely to explain all of the difference.

Two studies have considered the potential benefits of a better targeting of programs in Germany. Wunsch and Lechner (2008) estimate program impacts for 7 different programs during the period 2000-2002. They have access to a large and detailed administrative data set and argue that the variables included in the analysis justify the use of matching estimators relying on the CIA. Programs are generally found to be rather ineffective due to large locking-in effects and small or non-existing impacts on outcomes in the long run. Program impacts are, however, found to be heterogeneous over subgroups defined by observed characteristics. More specifically, program impacts are found to be most negative for unemployed workers with good employment prospects, because the locking-in effect is especially harmful for this group and because there is no long-run impacts to make up for this initial decline in outcomes. Unemployed workers with less favorable employment prospects have smaller locking-in effects and for some programs they actually experience positive impacts in the long run. This heterogeneity leads the authors to simulate program impacts under alterna-
ative assignment mechanisms. They identify the best programs conditional on different values of an estimated employability rating, $Y^0$, and compare simulated outcomes under the alternative assignments to the current assignment.\footnote{Note that this is actually targeting based on (detailed) subgroups. A more detailed estimation of heterogeneous impacts conditional on observed variables could be used to construct a more precise targeting model.} If all unemployed workers are assigned to the best programs, then the employment rate after 2.5 years is estimated to increase by 8% from 39%. This assignment is, however, much more expensive than the current assignment since many of those currently receiving no training would be assigned to one of the more expensive programs. Alternatively, by only changing the assignments for those unemployed who actually participate in programs under the current assignment, it is shown that the employment rate can be increased by 2% and at the same time the average costs can be decreased considerably.\footnote{The reduction in costs is estimated to be around EUR 1,000 per person which for the population considered would imply costs savings of around EUR 2.2 billion.} A later study by Huber et al. (2009) shows similar results for welfare recipients participating in training programs. The current assignment mechanism in this case results in an employment rate of 14%, random assignment of programs results in an employment rate of 15% and optimal assignment of programs (keeping the share of participants in each program constant) results in an employment rate of 23%. These two German studies are not aimed specifically at an evaluation of the assignment mechanism, but they do provide good examples of how to exploit the information obtained from evaluation studies where substantial impact heterogeneity is found. This also means that the results are to be seen as rough estimates for the potential of a targeting, and the studies do for instance not take in-sample bias into account.

Finally, a Danish study described in Staghøj et al. (2009) explicitly seek to evaluate whether a proposed targeting model can decrease the average duration of unemployment. The identification of program impacts in this study relies on exogenous variation in the timing of the assignment to programs using the timing-of-events approach.\footnote{This method is developed by Abbring and van den Berg (2003).} The authors argue that this method is particularly appropriate in this context where ongoing programs are evaluated in a setting where dynamic selection into programs needs to be modelled. Based on the estimated program impacts, different assignment mechanisms are considered, and the results suggest that a rather large decrease in the average duration of unemployment may be
the result if the targeting model is introduced. This study explicitly tries to eliminate the bias resulting from in-sample predictions as discussed in details in Section 3.4. A parametric bootstrap method is used to make sure that the suggested assignment mechanisms do not exploit spurious variation. If the assignments of the best programs are mostly based on spurious variation, then this assignment will show a poor performance when evaluated over the parametric bootstrap simulations. The bias arising from in-sample predictions is not found to be particularly important in this evaluation.

4.3 Summary of results

Two main results emerge from reading the literature about STRs. First, while it seems possible to use a profiling model to identify unemployed in risk of long-term unemployment, this approach does not seem to improve the effectiveness of programs significantly. The studies that have looked at the relationship between the value of the profiling variable and program impacts have not found a clear pattern. The negative correlation needed between \( Y^0 \) and the program impacts is only found in some situations. If efficiency is the main goal of the programs, then this result naturally opens up for the use of targeting models which aim much more specifically at improving efficiency. It simply answers the more appropriate question: which program is the most effective for this particular person? The natural reason being that program effectiveness may very well depend on other things than those factors influencing the risk of long-term unemployment. An important question with regards to targeting is whether a sufficient degree of accuracy is obtainable? There is no doubt that the estimation of a targeting model is a demanding task, and hence it can only be done with access to first-class data. As this is more often available, the case for estimating targeting models will only improve in the future. Studies considering the performance of targeting models have not yet established that targeting models will in fact imply a considerable improvement compared to the current methods used. But the growing list of simulation

\[ \text{Different introductions of the targeting system are considered in the paper. If the targeting model is used immediately after workers enter unemployment, then the targeting model is simulated to result in a decrease in the average duration of unemployment in the range of 3.8 – 6.5 weeks. If the targeting model is instead used at the time where unemployed workers are actually observed to be assigned to a program under the current assignment mechanism, then the average remaining duration of unemployment for these individuals is decreased by 9.0 – 19.2 weeks as a results of a better assignment of the programs.} \]
studies suggests that targeting models are capable of improving effectiveness of programs. And on the same time, exactly because the task of estimating program impacts is so hard, this also seems likely to be true for caseworkers, who are not generally found to be especially good at this particular task. The current and future experimentation with targeting models will increase our knowledge about the empirical relevance of this type of STR both in terms of the possible benefits and, if the experiment is well conducted, also in terms of the costs associated with different assignment mechanisms. This should allow for a more informed cost benefit analysis of the different types of assignment mechanisms.

The second main point coming out of reading the literature is, that implementation of a STR really needs to be considered seriously. There are examples where the model has become accepted, like in the US and in Australia\textsuperscript{72}, but also a list of examples, where the STRs never were really accepted and used after the implementation like in Canada, Denmark and Switzerland. A main reason for the rejection of the STR seems to be resistance from caseworkers. Hence, it is most important to specify in which way the STR and caseworkers are thought to interact and then to assure that incentive schemes are aligned to accomplish this desired use of the STR. In most cases (with the WPRS model used in the US as the main exemption), it has been the idea to use the STR as a pure information tool\textsuperscript{73} to caseworkers who then retain the discretion to choose the allocation of programs. Three advantages of this 'soft' introduction of a STR can be presented. First, it may seem less threatening to caseworkers in the sense that they should not be afraid that the computer will suddenly take over their job. Secondly and related to the first advantage, it may be seen as an indirect acceptance of the fact that caseworkers may indeed observe some local information which sometimes implies that suggestions from the STR are inappropriate and should be overruled. And finally, if the STR performs well, then it should be in the interest of caseworkers to actually use the available information to help their clients to get into work more quickly.

A main disadvantage of a soft implementation of STRs is that this may seriously weaken the cost-benefit potential of a STR where substantial cost savings may be possible if fewer

\textsuperscript{72}Although not without problems along the way. See Productivity Commission (2002).

\textsuperscript{73}A targeting model may not only provide relevant information to caseworkers. If some unemployed workers have unrealistic expectations about their job opportunities, then the targeting model may help the unemployed to make more informed choices. And targeting models may also be used as a relevant information tool for administrators who want to analyze the performance of local employment agencies or to target resources to areas where impacts are estimated to be largest.
Caseworkers are needed to make assignments to programs. But this is not the only reason why a soft implementation may not be optimal. The results from the experiment with SAPS and other studies suggest that the soft approach may not always be sufficient. If results suggest that the STR represents an improvement compared to caseworkers, then we would not want caseworkers just to override the STR suggestions. One reason for the comparatively worse performance by caseworkers may be that they are overconfident about their own ability to predict the program impacts, and this may then also lead them to reject the targeting model’s predictions.\footnote{Caseworkers may also deviate from the targeting model’s suggestions because they have other goals than the social planner. Reality may require that a number of different goals are weighted together to make the final decision on program assignment. A targeting model could then be seen as an information tool that should allow caseworkers to make more informed choices. The targeting model informs caseworkers about the costs of deviating from the best program.} It would in this case be necessary either to make the use of the targeting model mandatory or at least to design some incentive schemes that would ensure that the STR suggestions are followed in general. This may still be done with the possibility for caseworkers to overrule the STR, but it should then only be done on less frequent occasions.

The ‘broken leg’ example discussed by Meehl (1954) is exactly about the situations where a statistical rule should be overruled by clinicians. The example is about predicting whether an economics professor will go to the movie theater on a given Tuesday night. The statistical model may then use historical observations to predict that this will happen with a 90% probability. But if the clinician knows that the professor fell and broke his leg in the morning, then the clinician would do right to intervene and set the probability to 0% instead. Meehl gives a detailed discussion of the problem of these improbable events which nevertheless happen every day. But the improbable events are by definition rare, and he argues that clinicians tend to see them too often. So, an observation that caseworkers overrule the STR in say 50% of the cases should not be explained by the fact that all these individuals represent ‘broken leg’ examples. One way to monitor whether caseworkers intervene in the right situations is simply to record the deviations from the STR suggestions and evaluate the resulting outcomes. In this way, it should be possible over time to design the system in a way that implies an appropriate amount of caseworker discretion. The conclusion from this subsection is therefore that it may not always be optimal to use the ‘easy’ solution of
just saying that we should use both an STR and caseworker discretion, because this may not give the optimal balance between the two methods. To strike the optimal balance, it may be necessary to introduce different versions of the system in a randomized experiment which is in fact one of the recommendations coming out of the SAPS experiment.

Finally, it is obviously important to develop and maintain the STR along the way. A good STR should be designed such that continuous testing and improvement of the model is possible. It may be optimal to sacrifice some short-term efficiency by introducing random assignment for subgroups of individuals in order to obtain better estimates and hence more precise targeting in future periods. A systematic reestimation of program impacts may also be required because the relationship between the variables in the model may change over time. And it should allow us to learn more about the stability of the estimated parameters, which can then be used to determine an optimal frequency of updating the model. The implementation of WPRS in Kentucky and the experiments with profiling in Kalamazoo, Michigan provide some ideas of how to design a system that facilitates evaluation of both the programs and the assignment mechanisms.

5 Conclusion and discussion

In this paper we have discussed the use of profiling and targeting models to assign unemployed workers to ALMPs and reviewed the results in the literature. The theoretical foundation suggests that STRs may be used to optimize the targeting of programs whenever impacts are heterogeneous in an observable way. The empirical evidence has, however, not yet demonstrated that these theoretical improvements are possible to achieve in practice. Most evidence is available about profiling models and it generally seems possible to identify those in risk of long-term unemployment with reasonable accuracy. This implies that profiling models may be used to further equity goals if the group of long-term unemployed workers coincides with a group that are defined as most in need for help by a social planner, given that programs are beneficial for the participants. But if the main goal is efficiency rather than equity, then profiling models do not seem to be appropriately directed at the right question. Targeting models would then be more appropriate, exactly because they seek to provide an answer to the right question: which program is most effective for the particular unemployed
worker? Targeting models are, however, also more difficult to estimate and demand a lot in terms of data. More actual evaluation studies are needed in order to determine whether targeting models are empirically as relevant as a list of simulation studies suggests. But as the simulation studies all seem to point in the same direction, it seems worthwhile to go on with the experimentation with targeting models.

The reason that STRs are considered relevant is basically that caseworkers face some hard prediction problems when trying to help their clients. So an alternative to the use of STRs could be to help caseworkers perform these tasks more effectively in other ways. Specific training or improved feedback mechanisms could possibly improve their performance. For instance, statistical training may enable caseworkers to be better at identifying the relationship between the characteristics of the individual and the impacts of different programs.

An interesting extension of currently available STRs would be to consider dynamic editions of the models. When is it optimal to assign unemployed workers to a program? And is it possible to update the model dynamically to allow for an inclusion of intermediate results as observed variables? Such extensions would obviously just raise the already high data demands even further but in medicine, some studies of this sort has been conducted, and it may be possible to learn something from this literature about adaptive treatment effects.  

---

75 See Murphy (2005) for a pointer to this literature.
References


129


[78] Rudolph, Helmut and Michael Münt niche (2001). "Profiling' zur Vermeidung von Langzeitarbeitslosigkeit. Erste Ergebnisse aus einem Modellprojekt (The Use of Pro-
filing to Avoid Long-Term Unemployment. First Project Results)”, *Mitteilungen zur Arbeitmarkt- und Berufsforschung*, 34 (4), 530-553.


Figure 1: Simulated bias. Average values of the max of individuals’ program impacts. The averages are taken over simulations for 5,000 individuals. Individuals’ program impacts are drawn as independent normal variables with zero means. ’n’ denotes the number of programs and ’sigma’ denotes the variance of the impacts.
CHAPTER 4

Analyzing the Ins and Outs of Unemployment using Danish Spell Data
Analyzing the Ins and Outs of Unemployment using Danish Spell Data*

Jonas Staghøj†

November 25, 2009

Abstract

In this paper I analyze the dynamics of unemployment using register data with weekly information for the entire working population in Denmark followed over the period 1985-2003. I analyze whether fluctuations in the unemployment rate is mainly due to variations in the inflow to or the outflow from unemployment. Using various decomposition methods I consistently find that the outflow rate is the most important flow. The outflow rate accounts for at least half of the cyclical fluctuations in unemployment, with the remaining fluctuations accounted for by the inflow rate and flows in and out of the labour force. I find that outflow fluctuations are not due to compositional changes in the pool of unemployed. I also find that the use of less informative survey data implies that the importance of the inflow rate is overstated. Finally, I find that the inflow rate is relatively more important for younger workers and for male workers.

Keywords: Unemployment, Gross Worker Flows, Job Finding Rate, Separation Rate

JEL Codes: J63, J64, E24

*I am grateful for comments from Michael Svarer, Rune Vejlin, and Jesper Bagger and from participants at the DGPE 2008 Annual Meeting and LMDG’s Ebeltoft meeting.

†School of Economics and Management, Aarhus University, E-mail: jstaghoej@econ.au.dk
1 Introduction

In this paper I analyze the dynamics of unemployment using register data with weekly information about labour market states for the entire working population in Denmark followed over the period 1985-2003. \footnote{Data for the period 2004-2006 should be available shortly.} Two essential channels affect the unemployment rate, namely fluctuations in the inflow rate from employment to unemployment and fluctuations in the outflow rate from unemployment to employment. The conventional wisdom, developed on the basis of Darby, Haltiwanger & Plant (1986), has been that an increase in the unemployment rate was primarily a result of an increased inflow to unemployment. Shimer (2007), however, challenges this view and shows that the inflow rate is almost acyclical in U.S. data while the outflow rate is highly procyclical. This 'New View' of labour market flows has spurred a renewed interest in figuring out the relative importance of the two channels, in order to find out whether the focus should be on the ins or the outs if we want to understand the dynamics of unemployment.

The main contribution of this paper is to show results from a new type of data. The need for additional information is apparent from the differing conclusions in a number of papers that have followed Shimer (2007). Various modifications and alternative decomposition methods have been proposed by Elsby, Michaels & Solon (2009), Elsby, Hobijn & Sahin (2009), Fujita & Ramey (2006, 2007), and Petrongolo & Pissarides (2008). Most papers use data from the Current Population Survey (CPS) but even using the same data, not all papers agree with Shimer’s conclusions.

Five results emerge from the present analysis of detailed Danish register data. First, using various decomposition methods, the outflow rate is consistently found to be the most important factor accounting for fluctuations in the unemployment rate. The strength of this relationship does, however, vary over the decomposition methods. Some methodological assumptions lead to a very strong relationship where there is essentially no role left for the inflow rate, whereas other assumptions leave a role for the inflow rate even though the outflow rate is still more important. Secondly, a more complete and accurate description of
the dynamics of the Danish labour market should include transitions in and out of the labour force. This contrasts the findings in Shimer (2007) where a simple two state description (employment and unemployment) is found to be sufficient. The third result is that changes in the observed outflow rate cannot be contributed to compositional changes in the pool of unemployed. The fourth result is that the inflow rate is relatively more important for some subgroups, especially for younger workers and male workers. And finally, I find that the use of less informative survey data would imply that the importance of the inflow rate is overstated when compared to results based on register data.

The results in this paper are to be seen as descriptive. In the next section I will briefly discuss the underlying economic decisions that determine the various transition rates but without setting up a more elaborate model. So while it is sometimes tempting to interpret relationships as being causal, this is not necessarily the case. The motivation for a purely descriptive study is that the development of reliable stylized facts adds to our knowledge about the labour market and is a prerequisite for model building. It may also provide valuable information about where the marginal contributions of further research will be bigger.

If inflows are most important, then we should concentrate the analysis on the decisions underlying layoffs and quits. If, on the other hand, the outflows turn out to be more important, then the focus should be primarily on understanding the changes in the job finding probability. And finally, descriptive results may still have policy implications in terms of how to address problems arising from large unemployment fluctuations. An example could be to let the use of active labour market policies depend on the business cycle, which would make sense if the outflow plays a central role, but not necessarily so if the inflow is the most important factor behind unemployment fluctuations.

The paper is organized as follows. In section 2 I briefly discuss the flow approach to labour market dynamics and review the empirical evidence about the importance of the different flows. In section 3 I show how steady state approximations can be used to construct a number of decomposition methods. Section 4 describes the data used in this paper and includes a comparison to U.S. data. Section 5 shows the main results and section 6 concludes.
2 Labour Market Dynamics

2.1 The Flow Approach

The flow approach to the modelling of unemployment dynamics is based on the assumption that there are costly frictions in the process of matching unemployed workers and open vacancies.\(^2\) This means that workers cannot immediately move between all labour market states as they prefer. Unemployed workers have to search for a job and are matched with an appropriate open vacancy in a process that contains some generic uncertainty. Employers, similarly, have to search for appropriate workers, and this has implications for the dynamics of the labour market. Cyclical unemployment is a result of workers and employers changing behavior as a reaction to changing market conditions.

Employers continuously try to adjust the number of employed workers to an optimal level determined by the demand for the goods they sell and the prices of the input factors. This adjustment is carried out by hiring and firing workers and the costs associated with these two channels may vary over time, job types, sectors and countries. If for instance an employer needs to decrease the number of employees during a boom, this may be possible to accomplish by relying on the natural flow of workers to other employers and to other labour market states, by simply not hiring new employees. In a recession the employer may instead have to use the firing channel. The preferred adjustment channel may also vary over sectors, where sectors with a high turnover level and low firm specific human capital accumulation will be able to rely more on the hiring channel. Changes in the size of different sectors over time may hence also affect the dynamics of unemployment within an economy.

Workers may also contribute to fluctuations over time, if their behavior changes when faced with different labour market situations. Workers’ search behavior is basically determined by the relative value associated with the different labour market states. If market conditions change and unemployed workers experience a change in the probability of getting a job or a change in the terms of available jobs, they react by changing their search behavior. The relative value associated with different states may also vary due to institutional changes both in the long run and over the business cycle. An example of the latter could occur if the

\(^2\)See Mortensen & Pissarides (1994, 1999) for much more details.
stigmatizing effect of being unemployed is smaller during a recession where more individuals are unemployed. This would result in a reduced search effort and hence to a lower outflow rate.

It is not the purpose of this paper to analyze to what extent flows between labour market states are a result of employers’ or workers’ decisions, but rather to characterize the pattern of flows over time and the implications for the unemployment rate. These results can then guide us to the most important mechanisms to delve into and model more explicitly.

2.2 Existing Evidence

No clear consensus has yet emerged from the literature on the relative importance of the different types of flows for the fluctuations in the unemployment rate. The conventional wisdom holds that a recession is typically characterized by a sharp increase in the inflow to unemployment because more workers are fired. Changes in the outflow rate are less important since the outflow rate is primarily determined by the characteristics of the unemployed workers and not affected very much by the business cycle. The temporary increase in the inflow may however have long-lasting effects on the observed outflow rate because of heterogeneity. This will be the case if workers who are fired are typically less skilled workers, with a lower outflow rate. These workers will then be overrepresented in the unemployment pool, and dynamic selection out of unemployment may further postpone the return to normal conditions. The conventional view has some intuitive attractiveness and is consistent with the picture often drawn in the popular press about the situation of the labour market, where mass layoffs play a central role. It is also supported by a number of papers including Darby, Haltiwanger & Plant (1985, 1986), Blanchard & Diamond (1990), Davis, Haltiwanger & Schuh (1996), Davis & Haltiwanger (1995, 1999) and more recently by Fujita & Ramey (2006, 2007).

The 'New View' on the other hand is that the outflow rate from unemployment seems much more important for fluctuations in unemployment. Recent papers supporting this theory include Shimer (2007) and Hall (2005a, 2005b, 2007). But this view is in fact not entirely new, since Pissarides (1986) showed very similar results. The story is, basically, that while it is true that more workers are fired during the initial phase of a recession, this does not
really affect the unemployment rate very much. The reason being that even during normal periods, many workers flow into unemployment, and furthermore there is also a tendency for different types of flows out of employment to counteract each other. While more workers are fired during a recession, less workers quit and move freely to unemployment. The basic result in the papers supporting this view shows how the fluctuations in the unemployment rate can be matched quite accurately without any contribution from changes in the inflow rate. Shimer (2007) shows that the outflow rate accounts for 95% of the fluctuations in the unemployment rate in the U.S. during the period 1987-2007. He also shows that it is not possible to attribute this variation to heterogeneity in the outflow rate.

One way to grasp the support for the different theories is to conclude that both reflect some part of the picture. And there is a number of papers finding that both flows seem if not equally important, then at least sufficiently important not to neglect any of them. Davis, Faberman & Haltiwanger (2006) conclude that outflow changes are most important during mild recessions, while inflow changes are most important during more severe recessions. Elsby, Michaels & Solon (2009) agree with Shimer that the outflow rate accounts for most of the variation in the unemployment rate, but they also find that there is still a contribution from the inflow rate. Using the same data as Shimer (2007) but refining some of the analysis, they conclude that the outflow rate accounts for two thirds of the unemployment fluctuations, with the remaining fluctuations are accounted for by the inflow rate. Yashiv (2007) provides a survey of the literature on U.S. labour market dynamics and concludes that both flow rates are important. He also concludes that different adjustment methods and different methods to handle transitions in and out of the labour force contribute to differing results even when the same data are used in different papers. My results corroborate the importance of these issues and I discuss the importance of methodological choices in more details in Section 5.

Finally, a couple of papers have considered the empirical results from other countries. Petrongolo & Pissarides (2008) use data from UK, France and Spain and find differing results. For UK, they find that the outflow rate accounts for 67% of the unemployment fluctuations over the period 1967-2007 which is basically the same as Elsby, Michaels & Solon (2009) find using U.S. data. In the more recent period 1993-2007 this number is up to 75%, and they argue that this may be due to a period without dramatic changes in the unemployment.
rate, since this makes it more easy for employers to rely on changes in the job creation rate, which drives the outflow rate. Interestingly, they also find that the flows in and out of the labour force contribute to unemployment fluctuations in UK. Using data from the more strictly regulated French labour market, they find that the outflow rate accounts for 80% over the period 1991-2007. They argue that this is not surprising given that a high degree of employment protection raises the cost for employers of using the firing channel, and this seems especially important during periods with falling unemployment. For Spain they find that inflows and outflows seem equally important during the period 1987-2006. They explain that Spain was initially characterized by high employment protection which then resulted in a widespread use of fixed-term contract with a maximum duration of three years to avoid the strict firing rules.

Elsby, Hobijn & Sahin (2009) extend the method developed by Shimer (2007) in order to use annual data from the OECD countries to give a broader picture across countries. They find that the outflow rate typically accounts for around 80% of unemployment fluctuations in Anglo-Saxon countries whereas for Continental European and Nordic countries both flow rates seem equally important. I perform a similar analysis using Danish data in Section 5.5 and find that the results based on survey data are biased in the Danish context, with too much importance assigned to the inflow rate.

3 Decomposing Unemployment Fluctuations

The analysis of labor market flows is in principle very simple. Workers switch between different states in the labor market and the number of unemployed workers increases if the number of entries into unemployment is bigger than the number of exits. Hence, much of the discussion about this topic arises from the task of connecting the simple theory to the available data. Shimer’s main results are found in a model with only two states, unemployment ($U$) and employment ($E$). He later performs some robustness analysis to allow for transitions in and out of the labour force ($N$) as well, but find that this does not change the results substantially. The decomposition of fluctuations in the unemployment rate is based on

\footnote{They use data for Norway and Sweden but not for Denmark. Presumably because of a problematic structural break between 1991 and 1992 in the Danish survey data.}

\footnote{Shimer denote this state as Inactivity.}
on steady state approximations, and in the following I show the relevant approximations in
the two and three state models.

3.1 Steady State Approximations

Shimer (2007) argues that you should clearly define what time period you want to measure
and then adjust the available data if it is observed with a different time period. Failing to
take time-aggregation in observed data into account may result in biased findings, because
some short term transitions will not be measured.\(^5\) He proposes a way to adjust the observed
monthly transitions to get a measure of continuous transition rates. But as Elsby, Michaels
& Solon (2009) argue, it is not obvious that you should adjust the series 'all the way' to
continuous time. They argue that an adjustment from monthly to weekly transition rates
is probably more in agreement with the actual labor market definitions of unemployment
used to construct the data. I agree with this and use the weekly data in this paper without
adjusting for time-aggregation, not because of data limitations, but because I find a weekly
transition rate to be more in accordance with the real world decisions leading to transitions
between labour market states.

I assume worker homogeneity in the sense that all workers face the same transition
probabilities.\(^6\) The number of unemployed in period \(t+1\) is then given by the simple
accounting identity

\[
U_{t+1} = U_t + S_t - F_t
\]

(1)

where \(S_t\) is the number of inflows to unemployment and \(F_t\) is the number of outflows.

Now consider the situation with only two states, \(E_t\) and \(U_t\), and rewrite in terms of transition
rates to get

\[
U_{t+1} - U_t = s_t E_t - f_t U_t
\]

where \(s_t = \frac{S_t}{U_t}\) is the inflow rate from employment and \(f_t = \frac{F_t}{U_t}\) is the outflow rate.

\(^5\)Transitions into unemployment are more likely to be missing in the data, because the outflow rate is
much larger than the inflow rate. Time-aggregation is therefore most important for the inflow rate. And the
bias is likely to be time-varying since more transitions into unemployment will be missing in the data when
the outflow rate is high.

\(^6\)I relax this assumption to allow for heterogeneous outflow rates in Section 5.4.
Defining the conventional unemployment rate as \( u_t = \frac{U_t}{L_t} \), and assuming a constant labor force, \( L_t = E_t + U_t \), I get an expression for the change in the unemployment rate,

\[
\Delta u_t = (1 - u_t) s_t - f_t u_t \tag{2}
\]

And in steady state this leads to the main approximation used by Shimer

\[
u^{ss} = \frac{s}{s + f} \tag{3}\]

The two state approximation of unemployment is hence seen to depend only on the inflow and outflow rates and this equation is the starting point for many of the empirical decompositions of unemployment fluctuations described in the following section.

When I introduce out of the labour force as a third state, \( N_t \), the notation becomes a little more cumbersome. Denote a flow from state \( A \) to \( B \) as \( AB \) and the corresponding transition rate as \( \lambda^{AB} \). This transition rate is defined as the number of \( AB \) transitions divided by the number of workers in \( A \). In steady state, equality between the flows to and from employment and unemployment implies that

\[
(\lambda^{EU} + \lambda^{EN})E = \lambda^{UE}U + \lambda^{NE}N \]

\[
(\lambda^{UE} + \lambda^{UN})U = \lambda^{EU}E + \lambda^{NU}N \]

Shimer shows that these equations can be manipulated to give the following expressions

\[
E = k(\lambda^{UN}\lambda^{NE} + \lambda^{NU}\lambda^{UE} + \lambda^{NE}\lambda^{UE})
\]

\[
U = k(\lambda^{EN}\lambda^{NU} + \lambda^{NE}\lambda^{EU} + \lambda^{NU}\lambda^{EU})
\]

\[
N = k(\lambda^{EU}\lambda^{UN} + \lambda^{UE}\lambda^{EN} + \lambda^{UN}\lambda^{EN})
\]

where \( k \) is a constant set such that \( E, U \) and \( N \) sum to the number of individuals in the population. Using these expressions, I can now again approximate the steady state.
The simple approximation in (3) and the extended approximation in (4) are the main equations used to decompose the contribution of different transition rates to the fluctuation of the unemployment rate. In the following sections I describe the decomposition methods, which I will use in this paper.

### 3.2 Shimer’s Decomposition Method

Shimer (2007) suggests a new way to construct measures of the inflow and outflow rates, which has been used in the subsequent papers in the literature. The measures of the flow rates are constructed from publicly available data series from the Bureau of Labor Statistics. These series are measured at a monthly frequency so Shimer shows how to take time aggregation into account to get a continuous outflow rate that is assumed constant within a month. He takes quarterly averages of the monthly series to be less exposed to measurement errors. To judge the relative importance of the inflow and outflow rates, Shimer constructs two hypothetical unemployment series. One series where the outflow rate is allowed to fluctuate while the inflow rate is fixed at the mean value, ̅s, and a similar series where the inflow rate is allowed to fluctuate and the outflow rate is fixed at the mean, ̅f. I refer to these series as the contribution from the outflow rate, uto, and the contribution from the inflow rate, uti

\[
uto = \frac{\bar{s}}{\bar{s} + \bar{f}} \quad \text{uti} = \frac{s_t}{s_t + f_t}
\]

He then detrend the two contribution series and the actual unemployment rate using an HP filter with the smoothing parameter set to 100,000. And finally, he considers the covariance of each of the two contribution series and the unemployment rate divided by the variance of the unemployment rate.

\[
\beta^s = \frac{\text{cov}(ut, ut_i)}{\text{var}(ut)} \quad \beta^f = \frac{\text{cov}(ut, ut_o)}{\text{var}(ut)}
\] (5)
These measures are interpreted as the fraction of unemployment fluctuations which can be accounted for by the considered contribution series. He shows that this decomposition is fairly accurate such that $\beta^s + \beta^f \simeq 1$.

To measure the flows between states when allowing for flows in and out of the labour force, Shimer exploits that the Current Population Survey is conducted as a rotating panel where some individuals are in the sample in consecutive months. It is therefore possible to construct measures for the six flow rates between the three labour market states. He then calculates contribution series for each type of flow rate by fixing all other rates in (4) and then again analyzes how well the contribution series can account for the actual fluctuations in the unemployment rate.

### 3.3 Fujita & Ramey’s Decomposition Method

Fujita & Ramey (2007) make some additional adjustment of the gross flows based on CPS data, and then exploit an idea put forth in Elsby, Michaels & Solon (2009) to decompose the variation in unemployment more precisely. The idea is, basically, to take into account that the covariance between the two contribution series may be relevant such that $\beta^s + \beta^f \neq 1$.

The trend component of unemployment is defined from the steady state approximation in (3) as

$$u^\text{trend}_t = \frac{s^\text{trend}_t}{s^\text{trend}_t + f^\text{trend}_t}$$

where $s^\text{trend}_t$ and $f^\text{trend}_t$ are found using an HP filter with the smoothing parameter set to the standard value of 1,600.\(^7\) Then they log-linearize the unemployment rate around the trend values to get

$$\ln\left(\frac{u_t}{u^\text{trend}_t}\right) = (1 - u^\text{trend}_t) \ln\left(\frac{s_t}{s^\text{trend}_t}\right) - (1 - u^\text{trend}_t) \ln\left(\frac{f_t}{f^\text{trend}_t}\right) + \varepsilon_t$$

If the log-linearization is instead made around lagged values the decomposition becomes

$$\Delta \ln(u_t) = (1 - u^\text{trend}_t)^\Delta \ln(s_t) - (1 - u^\text{trend}_t)^\Delta \ln(f_t) + \varepsilon_t$$

\(^7\)This number was suggested as a reasonable smoothing parameter in Hodrick & Prescott (1997) when using quarterly data.
which is the decomposition suggested in Elsby, Michaels & Solon (2009). Fujita & Ramey (2007) show how to calculate \( \beta \) values equivalent to the concept used in finance to describe the covariation between each of the two terms on the RHS of equation (7) with the LHS of equation (7) as a fraction of the total variation of the LHS.

\[
\beta^s = \frac{\text{cov}(\ln\left(\frac{u_t}{u_t^{\text{trend}}}\right), (1 - u_t^{\text{trend}}) \ln\left(\frac{s_t}{u_t^{\text{trend}}}\right))}{\ln\left(\frac{u_t}{u_t^{\text{trend}}}\right)}
\]

\[
\beta^f = \frac{\text{cov}(\ln\left(\frac{u_t}{u_t^{\text{trend}}}\right), -(1 - u_t^{\text{trend}}) \ln\left(\frac{f_t}{u_t^{\text{trend}}}\right))}{\ln\left(\frac{u_t}{u_t^{\text{trend}}}\right)}
\]

A similar decomposition applies to equation (8). These beta values should provide a complete decomposition of the variation in cyclical unemployment given that the steady state approximation is appropriate.

### 3.4 Elsby, Hobijn & Sahin’s Decomposition Method

Elsby, Hobijn & Sahin (2009) use data for 14 different OECD countries and note that the steady state approximation used on U.S. data is not always appropriate for data in other countries. In fact, they show that the U.S. is an outliers in terms of the level of the flow rates with no other countries having nearly as high flow rates in and out of unemployment as in the U.S.. This is potentially important for the decomposition methods used, since they are all based on steady state approximations. In countries with lower flow rates, shocks changing the steady state unemployment will be more persistent, whereas the high flow rates in the U.S. ensure an almost instantaneous adjustment to the new steady state level, when measured on a monthly basis. Therefore, Elsby, Hobijn & Sahin (2009) propose a new log-linearization around steady state values, where the change in unemployment is allowed to depend on lagged values of the unemployment rate. The starting equation is the development of the unemployment rate over time in equation (2). This can be solved one period forward to get

\[
u_t = (1 - \lambda_t) \frac{s_t}{s_t + f_t} + \lambda_t u_{t-1}
\]

Where \( \lambda_t = e^{-(s_t+f_t)} \). When \( \lambda_t \) is close to zero, the usual steady state approximation
is appropriate. Alternatively, they propose a log-linearization of (10) around \( s_t = s_{t-1}, \)
\( f_t = f_{t-1} \) and \( u_{t-1} = \frac{s_{t-1}}{s_{t-1} + f_{t-1}} \) and show in their appendix how this can be used to get

\[
\Delta \ln u_t \simeq \lambda_{t-1} \left[ (1 - \frac{s_{t-1}}{s_{t-1} + f_{t-1}})(\Delta \ln s_t - \Delta \ln f_t) + \frac{1 - \lambda_{t-2}}{\lambda_{t-2}} \Delta \ln u_{t-1} \right]
\]  

(11)

This approximation is then used to calculate contribution series where the flow rates are allowed to vary one at a time. Since the change in unemployment now depends on the unemployment rate in the previous period, they calculate the contribution series recursively, starting from the first period in the sample window. These cumulative contribution series are defined as

\[
C_{ft} = \lambda_{t-1} \left[ -(1 - \frac{s_{t-1}}{s_{t-1} + f_{t-1}})\Delta \ln f_t + \frac{1 - \lambda_{t-2}}{\lambda_{t-2}} C_{f_{t-1}} \right]
\]  

(12)

\[
C_{st} = \lambda_{t-1} \left[ (1 - \frac{s_{t-1}}{s_{t-1} + f_{t-1}})\Delta \ln s_t + \frac{1 - \lambda_{t-2}}{\lambda_{t-2}} C_{s_{t-1}} \right]
\]  

(13)

\[
C_{0t} = \lambda_{t-1} \left[ \frac{1 - \lambda_{t-2}}{\lambda_{t-2}} C_{0_{t-1}} \right]
\]  

(14)

where \( C_{ft} \) and \( C_{st} \) denote the contributions from the outflow and the inflow rates and \( C_{0t} \) denotes the contribution from the initial deviation from steady state. The recursive construction of the series is started by setting \( C_{f0} = C_{s0} = 0 \) and \( C_{00} = \Delta \ln u_{0}^{ss} \).

### 3.5 Decompositions used in this Paper

I use all of the above proposed decomposition methods to see whether methodological differences have important implications for the results. In particular, I should be able to analyze the difference between two and three states decompositions, the difference between approximations around steady state and non steady state as proposed by Elsby, Hobijn & Sahin (2009), the difference between using the standard HP smoothing parameter for quarterly data and the higher value suggested by Shimer, and the importance of taking the covariance of the contribution series into account.
4 Data

4.1 Introduction to the Danish Labour Market

To put the analysis in context I provide a brief overview of the Danish labour market. The historical fluctuations in the unemployment rate in Denmark is illustrated in Figure 1 together with the unemployment rate in the U.S. As seen, Denmark has experienced many of the same ups and downs as the U.S., but with significant changes in the level over time. In particular, the unemployment rate in Denmark started out very low during the 1960’s but then seemed to jump permanently to a new level during the first oil crisis in 1973 and again after the second oil crisis around 1979-1980. The period analyzed in this paper (1985-2003) is characterized by a relatively high unemployment rate compared to the U.S. and covers the large increase in unemployment until 1994, the long recovery thereafter and most of the increase starting in 2002.

In a European context, the Danish labour market can be characterized as being very flexible, with few restrictions on employers’ possibilities to fire workers. This is combined with generous unemployment insurance schemes, both in terms of the coverage of workers, the amount of benefit received and not least the maximum duration where unemployment benefit can be collected.

4.2 Danish Register Data

The key data series used in this paper are time series for the number of individuals in unemployment \(U\), employment \(E\) and out of the labour force \(N\) and time series for the six transition rates between these states. These series are constructed from a huge data set with weekly spell histories for the population in Denmark followed over the years 1985-2003.\(^8\) This spell data set is constructed by merging several Danish register data sets, using an anonymized personal identifier. First, individuals are assigned to one of sixteen mutually exclusive labour market states in each week, using information from the different data sets. Then the sixteen states are combined to the three most important labour market states, \(E\), \(U\) and \(N\). And finally, I simply count the weekly number of individuals in a certain state

\(^8\)The preliminary results presented in this draft is based on a 10% random sample from the population.
and the number of individuals switching state. The remainder of this Section provides a more detailed description of the construction of the basis data set.

The overall quality of the data sets used is high, but sometimes the different data sets provide contradictory information. This is for instance the case if an individual appears to be in multiple states in the same week. The problem is resolved by using the most reliable data source. An implication of this is that some states are more precisely defined than others.

The most important information is obtained from a data set called CRAM (Det Centrale Register for Arbejdsmarkedsstatisitik/Central Register for Labour Market Statistics) that has weekly information about actual payouts of unemployment insurance. The official unemployment statistics published by Statistics Denmark are based on this data set, and I use it to identify individuals in state $U$. The reliability of the information from this data source is very high, and this state is therefore never overwritten with information from other data sources.

Another data set called SHS (Sammenhængende Socialstatistik/Consecutive Social Statistics) contains monthly information about individuals in a number of different support schemes, including sick pay, maternity/paternity leave, social welfare, rehabilitation, incapacity benefits, early retirement and ordinary retirement. Most of the individuals in these schemes are defined as being in state $N$, but individuals in some temporary schemes like maternity/paternity leave or sick pay are assigned to either $E$ or $U$ depending on the state they were in before they entered the temporary state.

Similarly, data from AMFORA (Arbejdsmarkedspolitiske Foranstaltninger/Active Labour Market Policies) is used to identify individuals participating in various types of active labour market programmes. Individuals in these schemes are either assigned to $U$ or $E$, depending on whether any salary is actually paid in the programme. This is done in an attempt to follow the guidelines from the International Labour Organization also used by Statistics Denmark.

From the above mentioned data sources I have collected information on all individuals receiving any kind of transfer and assigned some to state $U$ and other to state $N$. The next task is to identify individuals in employment. This is done using a data set called CON (Centrale Oplysningsseddelregister/Central Tax Information Sheet Register) which contains
annual information on all employers an individual has worked for during a year and the wage that was paid. The data should in principle contain annual information about the exact start and end dates of all employment spells, but this information is missing for many individuals. For most individuals the reason is that the employment spell has covered the entire year, in which case it is not a problem because the true start and end dates will then be available in the years where the employment spell starts and ends. So the general solution is to set missing start and end weeks to the first and last week of the year, respectively. But for some individuals either the start date or the end date is truly missing, and this turns out to be a problem for $EN$ and $NE$ transitions that are observed in the first week of a year. In the appendix I show how this is solved. I simply assume that the start and end weeks are missing at random, and then simulate a new start or end week during the interval where the transition must have happened.

4.3 Descriptive Statistics

From the initial data analysis I end up with weekly series of the number of individuals in the three states, $E$, $U$ and $N$, and the number of individuals who change state before the next week. I adjust the series to a monthly frequency, to make the analysis comparable to international results. For flow values this is done by summing flows over the weeks in a month and for stock values I take the value in the first week of the month.\(^9\) I adjust the flow values for the fact that the months defined in the data can consist of either 4 or 5 weeks by adjusting to an average duration of 4.33 weeks. The purpose of this paper is to analyze business cycle fluctuations so I perform a seasonal adjustment using the X-12-ARIMA procedure which is typically also used to adjust the CPS data series.

Figure 2 shows quarterly averages for the monthly unemployment rates along with the official unemployment rate from Statistics Denmark. The two series for the unemployment rate are constructed at least in part from the same basis data sets, and it is therefore reassuring to see that the final series are very similar. There is a level difference between the series, but the cyclical behavior appears to be similar, and the correlation between the

\(^9\)This is actually in the middle of normal calendar months, since a special calendar time for the payouts of unemployment benefits is used in this paper for practical reasons. It is, in general, shifted two weeks backwards compared to normal calendar time.
series is as high as 0.997. The unemployment rate is defined as \( u = \frac{U}{U+E} \) and it is clear from the underlying series for the number of workers in \( U \) and \( E \) that the level difference is a result of a lower \( E \) in the denominator in my data. When compared to official series for the population, the series in my data has too few workers in \( E \) and too many in \( N \). A minor part of this difference is due to the missing values problem for \( E \) spells handled in the appendix.\(^{10}\) The general conclusion is however that the constructed series matches the official series quite well in spite of the considerable number of data choices which have been made when constructing the final series from the basis data sets.

Table 1 shows descriptive statistics for the monthly series. The unemployment rate varies in the interval 5.3 – 14.2% with a considerable standard deviation compared to the mean. Note also that flows between \( U \) and \( N \) are actually the highest, which may indicate that these are important unless they are more or less constant over the time frequency considered in this paper. In the last two columns I have calculated the relative size of the different flow types and included similar numbers for the U.S. for comparison purposes. It is seen that the U.S. have relatively larger flows between \( E \) and \( N \) and relatively smaller flows between \( U \) and \( N \), and this may help to explain why I find the inclusion of \( N \) to be more important in Danish data.

The key data series in Shimer (2007) are the inflow and outflow probabilities. I show these in Figures 3 and 4 as quarterly averages of the monthly series in order to avoid excess volatility arising from noise. A first thing to note is that the level for the outflow probability is about ten times as high as the level for the inflow probability, mainly a consequence of the differences in the denominators. From a comparison with similar series from the U.S. it is seen that the level of the Shimer’s series is about three times as high for both the inflow and outflow probabilities. So while the Danish labour market is often characterized as being highly mobile, and hence more comparable to the U.S. than for instance continental European countries, these results indicate that there is still a substantial difference between

\(^{10}\)Other potential reasons for differences include different choices regarding unemployed on vacation (\( U \) in my series but \( N \) in the official series) and the fact that the official series use a constant labour force over the year, whereas I use the actual labour force in each week. The official unemployment series also takes into account that some workers are part time unemployed. This is not feasible in my analysis, since I have to assign each individual to one state in every week. I choose to assign part time unemployed to \( U \) if the worker is unemployed for more than half of the week and to \( E \) otherwise. Sensitivity analysis, however, shows that this has no visible influence on the aggregate unemployment rate.
the Danish and the U.S. labour market dynamics. This difference is also found in Elsby, Hobijn & Sahin (2009), where the U.S. is a clear outlier compared to all the OECD countries considered.

There is a downward sloping trend in the inflow probability regardless of the choice of smoothing parameter in the HP filter. This trend seems very similar to Shimer’s series over the same period. The trends in the outflow probability show more cyclical variation, and it is seen that the HP trend with parameter set to $\frac{1}{600}$ removes much of the variation.

4.4 Comparison with CPS data

Two different types of data from the CPS have been used to analyze the ins and outs of unemployment. The most simple type of data is the publicly available series on aggregate data, and the more complex data type measures gross flows constructed from microeconomic data on the individuals’ labour market status in consecutive surveys in the CPS.

Shimer (2007) shows how to use the simple aggregate data. He shows how estimates for the inflow and outflow rates can be obtained from series for the number of individuals in employment and unemployment and the number of short term unemployed. The upside of this analysis is that it uses publicly available and frequently updated data, and that the data contains reliable measures of well-defined economic variables. The downside is that the analysis relies on a number of simplifying assumptions. The method assumes no flows in and out of the labour force, homogeneous transition rates across individuals and constant flow rates within months. Shimer shows that the analysis is robust to these assumptions in his data, but they may be too restrictive in other settings. In Section 5.5 I show that these assumptions are not valid in the Danish case.

The more detailed gross flows data allows some of the assumptions to be relaxed, but as many authors have noted there are problematic issues with this data. Blanchard & Diamond (1990, p. 89) for instance write that 'Except from brief episodes, however, those tabulations have not been published because the Bureau of Labor Statistics (BLS) perceives them to be of poor quality.’ And Elsby, Michaels & Solon (2009, p. 87) sum up the most important problems when they write 'In addition, the gross flow data are subject to a number of drawbacks, including the systematic exclusion of individuals who change residence and the

158
many spurious transitions generated by misclassification of labor force status in either of the months used in the longitudinal match’.

To overcome some of the problems with the gross-flow data, Abowd & Zellner (1985) proposed several corrections of the flows. I refer to their paper and to Poterba & Summers (1986) for much more discussion of the need for data corrections and to Frazis, Robinson, Evans & Duff (2005), Ilg (2005) and Boon, Carson, Faberman & Ilg (2008) for more recent discussions of how the BLS has now again started to publish adjusted gross-flow data. Almost all papers about ins and outs of unemployment use the Abowd & Zellner adjusted data, which may implicitly speak about the appropriateness of their method. But the adjustments are far from negligible so it may potentially affect the results. In particular, the adjustments assure that the gross flow data is made to be consistent in the sense that the stocks in the following month can be obtained using the stocks and flows in the present month. This is the way the data should be, when correctly measured. But if this consistency is only assured by brute force adjustments, it may not necessarily improve the data for all purposes. Some of the differences between the U.S. and Denmark found in this paper, may partly be explained by the fact that the Danish data is not adjusted in this way.

Measurement issues are of course also present in the Danish data, and it is generally hard to judge the extend of these problems, but it seems reasonable to believe that the problems are less pronounced in Danish data. For instance, the problems with missing individuals between different months are not present in the Danish data where all individuals are observed in every week. Classification issues may, however, be present to the extent that individuals are assigned to a wrong labour market state. This is not likely to happen for $U$ spells, since this information is based on actual payments measured at a weekly frequency. But as already noted, classification errors happen more often when distinguishing the exact timing of $E$ and $N$ spells. The problem should, however, not result in spurious transitions that raise the flow rates, since it is really the timing of the transition that may sometimes be missing, and not a question of whether the transition actually takes place or not.

Blanchard & Diamond (1990, p. 94-95) point out that sample size may be problematic in the CPS data despite of the relatively large number of households in the CPS. The reason is that transitions are relatively rare events, when considered at a monthly basis, since most
individuals stay in the same state from one month to the next. They write that the expected number of monthly transitions averaged over the six transition types is around 750. This potential problem should be less evident using the Danish data, where the mean number of monthly transitions is 2.948 (using the 10% population sample). On the other hand, the Danish sample spans a much shorter period of time compared to the U.S. data. The ability to include historical data allows for the inclusion of a larger number of recessions and more generally, for more fluctuations in the unemployment rate. But the use of a long sample period may also imply that the sample period covers some structural breaks in the data, and Shimer (2007) for instance bases his main conclusion on the more recent history. So the lack of historical data series may not be as problematic as the lack of very recent data. It is obviously a nice feature of the U.S. data that they are so frequently and fast updated. If I am in a later version of this paper able to include 2004-2006 data, this will decrease this shortcoming of the Danish data.

5 Results

5.1 Approximations

All the decomposition methods are based on approximations that allow us to describe the unemployment rate as a function of the flow rates. Figure 5 shows the simple two state approximation from equation (3) and the three state approximation from equation (4). Both approximations capture the cyclical variation in the unemployment rate, with the three state approximation showing smoother fluctuations similar to the actual fluctuations in the unemployment rate. The correlation with the unemployment rate is 0.987 for the two state approximation and 0.983 for the three state approximation. These correlations appear very high, but a little caution is appropriate since correlation rates do not say much about the size of the fluctuations in two series.

There is a level difference in the approximations, and it is seen that the three state approximation is closer to the actual unemployment rate, than the two state approximation. So it seems to be necessary to include flows in and out of the labour force to be able to explain the level of unemployment in Denmark. The level of unemployment is not just determined
by the frictions between $E$ and $U$ but also by the additional frictions for flows going through $N$.\footnote{This should not be too surprising since it is well known that many individuals flow directly between $N$ and $E$. And it is also known that it is sometimes hard to distinguish the states $U$ and $N$ since some of the individuals in $N$ actually have a behavior very similar to the registered unemployed workers (see Blanchard & Diamond (1990), p.87). These patterns are found in Denmark as well as in the U.S., and it is therefore not entirely clear why transitions in and out of the labour force are not found to be important in U.S. data. Shimer (2007) and Elsby, Hobijn & Sahin (2009) for instance all claim that flows in and out of the labour market are not essential for their analysis. The rough comparison of the relative size of the different types of flows shown in Table 1 may indicate the reason for the difference. $EN$ and $NE$ flows are relatively larger in the U.S. while $UN$ and $NU$ flows are smaller. Pissarides & Petrongolo (2008) show results for U.K. that are very similar to the Danish results, where transitions in and out of the labour force are important.}

The approximations in Figure 5 are constructed using flow rates in the previous period. It is also possible to approximate the unemployment rate recursively from an initial value and flow rates. This is the dynamic approximation described in equation (10) and I show the resulting series in Figure 6 together with the extended dynamic approximation based on equation (11) where the unemployment rate is allowed to deviate from steady state values in a certain way. Both approximations are too low compared to the actual unemployment rate. But given that the approximations shown in Figure 6 only use information about flows between $E$ and $U$ it seems to be reasonable approximations. The correlation rate with the unemployment rate is 0.977 for the steady state approximation and 0.962 for the non steady state approximation.

The non steady state approximation is essentially a moving average of the present and previous steady state values for the unemployment rate and this is also seen from the fact that the approximation is lower than the actual rate when the actual rate is increasing and higher when the actual rate is decreasing. But even though the extended approximation may help to get a better approximation, it is still not perfect. The reason is that it only allows for certain deviations from steady state, namely those that arise because of a slow transition from one steady state to a new steady state. But apparently, slow transition to a new steady state is not the main reason for the inexactness of the approximations in the Danish data. This is consistent with the relatively high transition rates in Denmark.

Similar approximations based on U.S. data are much more precise. I believe there are three reasons for this difference. The first is that the inclusion of the state $N$ seems more important when using Danish data, since the three state approximation is more accurate. The second possible explanation is the lack of adjustment of the gross flows in Danish
data. Flows are measured directly as they occur and apart from seasonal adjustment and an adjustment around new year described in the appendix, I do not adjust the flows. In particular, I have not adjusted the flows to assure consistency between the flows and stocks. U.S. data is adjusted to assure this consistency. The third, and probably most important, explanation is that the U.S. results are constructed from survey data where the flows are not directly observed. The outflow is first inferred from the number of short term unemployed in the following period and the change in unemployment, and then the inflow is set to assure consistency between stocks and flows. It is therefore almost by construction the case that the steady state approximation based on these flows will match the actual unemployment. In Section 5.5 I show the results when using similar methods on Danish data.

5.2 Decompositions

Decomposition of cyclical fluctuations requires the series to be detrended. Figure 7 shows the unemployment rate together with two HP trends. The cyclical unemployment series implied by the different trends can be seen in Figure 8. The series with the label ’1. diff.’ is the result of using lagged values as the trend. The graphs show that the more flexible HP 1,600 trend removes too much variation, and this is the reason that I will, in the remainder of the paper, focus on results based on the HP 100,000 trend which leave more cyclical variation to be explained. I will, however, show some results using the flexible HP 1,600 trend to judge the robustness of this choice.

Figures 9 and 10 show the contribution series, where either the inflow rate or the outflow rate is allowed to vary. The contribution series for the outflow rate follows the actual unemployment rate quite accurately during the entire period except in the period around 1992-1994 where it is not able to explain the peak in the unemployment rate. This three year period seems to be the only period where the inflow contributes significantly to the fluctuations in the unemployment rate. Inflow may also contribute towards the end, but it is hard to say whether this is due to noise, since there is also a period from 1985-1989 where the inflow contribution actually goes in the opposite direction than the actual unemployment rate. So based on these graphs, the outs seem to be the winner.

The corresponding graphs for the inflow and outflow contributions using Elsby, Hobijn
Sahin’s method show basically the same results, and are available on request. In Table 2, I summarize the results from all decomposition methods that are based on 2 state approximations. The coefficients using Shimer’s method show $\beta_s$ and $\beta_f$ as defined in the equations in (5). These measures are interpreted as the degree of unemployment fluctuations that can be accounted for by the particular series. Fujita & Ramey’s method show $\beta_s$ and $\beta_f$ defined in equation (9) and Elsby, Hobijn & Sahin’s method show similar beta values based on the series defined in equations (12)-(14). The interpretation for these measures are basically the same as Shimer’s measures, but the latter measures take into account that the contribution series may covary.

Columns 1 – 3 show the results for the full sample period, with different detrending methods. The 1. difference detrending method is included since it has been used in the literature, but this is not my preferred method, because it does not seem to isolate the cyclical variation that is the subject of this paper. Results for using this method is only shown for the Fujita & Ramey method. The difference between using the HP $1,600$ and my preferred HP $100,000$ is not substantial, but it is generally the case that the approximation is less accurate when using the more flexible HP $1,600$ trend.

None of the two state approximations are able to decompose all the variation in the unemployment rate which would have required a coefficient of 1. But all approximations are able to account for a substantial part of unemployment fluctuations, and the decomposition between inflow and outflow shows that the outflow rate is most important. Setting aside the 1. diff. method, the inflow contribution accounts for $4 - 15\%$ of the variation in the unemployment rate according to the different methods, whereas the corresponding numbers for the outflow contribution are $44 - 87\%$.

There is not much difference in the decompositions for the early and late subperiods, which are generally characterized by increasing and decreasing unemployment, respectively. The only difference appears to be that the inflow rate was particularly irrelevant in the early period which is consistent with the 'false' increase in the inflow contribution series seen in Figure 9. In the more recent subperiod, the inflow rate seems slightly more important but still accounts for considerably less than the outflow contribution. 

\footnote{See figure 8.}
Now consider the more accurate decompositions based on three state approximations. I show the contributions from the six different flow rates in Figure 11, where one flow rate is allowed to vary at a time. Most contribution series are not able to generate fluctuations that match the actual fluctuations in the unemployment rate. The only exception is the contribution from $UE$ transitions which are actually able to generate sizable fluctuations and follow the unemployment rate over most of the sample period.

Table 3 summarizes the precision of the contribution series shown in Figure 11. First of all, it is seen that the approximations based on three states are able to explain a larger part of the fluctuations in unemployment. Secondly, it is seen that the most important flow rate is the $UE$ flow rate. Changes in this flow rate accounts for $42\% - 52\%$ of the fluctuations in the unemployment rate over the full sample period. But it is also interesting to see that some of the other series seem to contribute as well even though this may be hard to see from the graphs presented above. $UN$ and $NU$ flows are both seen to account for more than $10\%$ of the fluctuations in the unemployment rate. In the most recent period with a long lasting fall in unemployment, the $UN$ flow seems even more important. The general conclusion based on the results presented in Table 3 is, that even though the $UE$ flow accounts for around half of the fluctuations, this means that changes in the remaining flow rates also accounts for around half of unemployment fluctuations. All flows are hence needed in order to carefully characterize the variation in unemployment.

### 5.3 Subgroup Analysis

The spell data can be merged with other data sources to provide detailed individual characteristics which allows for an analysis of subgroups. A number of interesting results are found when analyzing subgroups defined by age, gender, ethnicity, marital status, education and sector. In Table 4, I show results from two and three state decompositions using Shimer’s method with the HP parameter set to 100,000.\(^\text{13}\)

First, it is seen that the outflow rate is more important than the inflow rate for all subgroup. The differences over subgroups is therefore whether the inflow rate seems completely irrelevant or has some contribution to unemployment fluctuations. The inflow rate are for

\(^{13}\text{Results using the other decomposition methods are available upon request.}\)
example seen to be relatively more important for the younger workers and for male workers, where the inflow contributions are 0.18 and 0.17, respectively. The contribution is even bigger when the two dimensions are combined, since young male workers have a contribution from the inflow rate of 0.24. One interpretation for the differences over age groups is that younger workers are less firmly attached to their employers and may also be more exposed to firings. Older workers may have more stable employment spells, and when they are hit by unemployment, it is the outflow rate that is important for their reentry into a job. The gender difference is most naturally interpreted as reflecting different types of jobs. To consider this possibility, I have divided workers into ten different sectors and it is seen that the sectors 'Industry', 'Utility & Construction', 'Hotels, Restaurants & Catering', 'Education', and 'Other' are sectors where the inflow is not irrelevant. Other male dominated sectors like 'Agriculture & Fishery' and 'Transportation' actually shows the opposite pattern than the gender results would indicate, since these sectors show little importance for the inflow. Hence, there may also be gender variation in the type of job within sectors, where male workers may for instance have more risky jobs.

The inflow rate also seems to be relatively more important for immigrants and unmarried workers. Stability of employment spells may again be a part of the explanation, but for immigrants, attachment to the labour force may also be important, since the three state decomposition shows relatively larger contributions from fluctuations in $EN$ and $UN$ flow rates for this subgroup.

Results with respect to education is somewhat surprising, since I expected the inflow rate to be relatively more important for unskilled workers. But I actually find that unskilled workers are largely unaffected by fluctuations in the inflow rate whereas the group of skilled workers (workers with vocational training) are more affected by the inflow rate. Note, however, that the decompositions concerns cyclical fluctuations in the unemployment rate, and

\[\text{Note that the sector 'Other' includes individuals for which no sector is observed. One reason that no sector is observed is that the individual is never observed to work during the observation period. So part of this group consists of individuals with a low attachment to the labour market.}\]
not the average level of unemployment. Unskilled workers actually have the highest inflow rate, and the lowest outflow rate, but the cyclical fluctuations in the inflow rate are relatively smaller than the cyclical fluctuations in the outflow rate. This means that inflow fluctuations are able to account for a smaller fraction of cyclical unemployment fluctuations even though the level of the inflow rate is important for the average level of unemployment for this group.

To draw policy implications from this decomposition analysis we should be aware of the distinction between how the flows affect the average level and the how they affect the cyclical fluctuations. If cyclical unemployment fluctuations are the main policy concern, then the above analysis may for instance indicate that the focus should be on policies concerned with the outflow rate, if unskilled workers are a target group. If policy makers are instead concerned with the average level of unemployment within a target group, then it is not possible to use the decompositions presented here to say that the inflow rate is irrelevant.

5.4 Heterogeneous Job Finding Rates

Darby, Haltiwanger & Plant (1985, 1986) suggest that observed changes in the outflow rate may not reflect real changes in the outflow rate, but instead be a consequence of changes in the composition of unemployed workers. When the outflow rate is for instance observed to fall, this may be due to the fact that less able unemployed workers now constitute a larger part of the unemployed workers. To analyze whether this 'heterogeneity hypothesis' is important, I follow the approach suggested in Shimer (2007).\textsuperscript{15} He shows that the overall outflow rate can be written as an average over outflow rates for subgroups of the population

\[ F_j = \frac{\sum_j u_{t,j} F_{t,j}}{\sum_j u_{t,j}} \]

where \( j \in \{1, ..., J\} \) is an index of \( J \) subgroups. The observed outflow rate is hence a function of the outflow rates in the subgroups and the relative size of the subgroups. I fix the relative size of the subgroups at the mean values over time, \( \overline{u}_j = \frac{1}{T} \sum_t u_{t,j} \), and allow the outflow rates, \( F_{t,j} \), to vary, to get a contribution from real changes in the outflow rates.

\textsuperscript{15}See also Baker (1992) for more about the 'heterogeneity hypothesis'.
Likewise, the outflow rate is fixed at the mean values over time, $F_j = \frac{1}{T} \sum_t F_{t,j}$, and the
number of unemployed in the subgroups is allowed to vary, to get a contribution from the
compositional changes in the relative size of the different subgroups.

If the heterogeneity hypothesis is to be supported by the data, we have to observe differ-
ces in the outflow rate across subgroups. As in the previous section I define subgroups
based on age, gender, ethnicity, marital status, education and sector. Consider, as an ex-
ample, the outflow rate for immigrants shown in Figure 12. As expected, there is a level
difference between the outflow rate for native Danes and immigrants. If there are cyclical
changes in the relative size of the subgroup of unemployed immigrants, then this could lead
to changes in the overall outflow rate, even without any real changes in the subgroup spe-
cific outflow rates. The problem is however, that such cyclical changes in the fraction of
unemployed who belong to a particular subgroup are hard to find in the data. Figure 13
shows the changes over time in the fraction of unemployed who are immigrants. There is
an increase over time but no distinctive cyclical patterns. Similar or less varying graphs
are found for the other subgroups I consider. It should therefore not be too surprising that
none of results presented in Figure 14 shows much relevance for the compositional effect.

The contribution series where the subgroup specific outflow rate is allowed to vary follows
the overall outflow rate very closely for all the subgroup divisions considered (the series are
graphically indistinguishable for some subgroups). The contribution series where only the
relative sizes of the subgroups are allowed to vary are on the contrary not able to generate
much variation in overall outflow rate.

The findings presented in this section are found using a limited number of subgroups
and for one dimension of heterogeneity at a time. But since none of the subgroup divisions
are able to generate any relevant compositional effects, I do not find it reasonable to be-
lieve that the compositional effect would play a major role even if I included more detailed
heterogeneity. This result is consistent with the findings in Gaure & Røed (2003), where
a heterogeneity-adjusted outflow rate is estimated on Norwegian data. They find that the
adjustments have a modest impact on the time-series behaviour of the outflow rate. Using
U.S. data, Dynarski & Shefrin (1990) similarly find that unemployment durations remain
procyclical (and hence the outflow rate remains countercyclical) after conditioning on a wide
range of controls. van den Berg & van der Klaauw (2001) analyze French data and find the effect on the outflow rate from compositional changes in the inflow to be small.

5.5 Comparison to Survey Data

Most of the previous literature has used survey data to construct the flow rate series whereas I have used register data in this paper. To analyze potential consequences of using different types of data, I use the high frequency Danish data to construct pseudo-survey data. Shimer (2007) uses monthly survey data for the number of unemployed and the number of short-term unemployed, defined as unemployed with an unemployment duration between 0 and 1 month. Elsby, Hobijn & Sahin (2009) use quarterly OECD survey data for the number of unemployed workers and annual data for the number of unemployed workers with durations in the intervals, $0 – 1$, $1 – 3$, $3 – 6$, $6 – 12$ and more $12$ months. The OECD data is also available for Denmark, but there appears to be a structural break between 1991 and 1992, where only the period after 1992 is comparable to the register data used in this paper. This leaves a rather short overlap of different types of data so I will not show results for the OECD data for Denmark, but instead show results for the pseudo-survey data I have constructed.\footnote{More results using the OECD survey data are available on request.}

I construct two different survey data sets.\footnote{I have done the analysis using a monthly survey data set as well, but this does not change the results.} A quarterly survey and an OECD-like survey, where the number of unemployed are observed on a quarterly basis and the distribution of durations are observed on an annual basis. The latter data set is constructed to compare results to Elsby et al. (2009), so I also assume that I do not know the exact timing of the annual survey, as in their paper. They describe in the appendix how to handle this problem, and I follow their approach for comparability.\footnote{The solution is to assume that the annual survey is conducted at the same time as one of the quarterly surveys in the year. Then take average values over the four different possibilities for the timing of the annual survey within the year.}

Elsby et al. (2009) show how to extend the method proposed in Shimer (2007) to calculate inflow and outflow rates from the type of survey data described above. The change in unemployment from month $t$ to month $t + 1$ can be written as

$$u_{t+1} - u_t = u_t^{<1} - F_t u_t$$
where $u_{t+1}^{<1}$ denotes the number of unemployed with a duration shorter than a month at time $t+1$. The outflow, since time $t$, is the number of unemployed in month $t$ times the outflow probability. Rewrite this equation to get an expression for the outflow probability, $F_t$

$$F_t = 1 - \frac{u_{t+1} - u_{t+1}^{<1}}{u_t}$$

which implies that the corresponding outflow rate can be found from

$$f_t^{<1} = -\ln(1 - F_t)$$

Shimer (2007) calculates the outflow rate from these equations and calculates the inflow rate as the solution for $s_t$ in equation (10). Elsby et al. (2009) note that similar equations holds for the number of unemployed with other unemployment durations, such that

$$f_t^{<d} = -\frac{\ln(1 - F_t^{<d})}{d}$$

where $f_t^{<d}$ denotes the outflow rate calculated from the number of unemployment with a duration less than $d$ months. From the OECD-like survey data, I can therefore calculate different measures of the outflow rate, $f_t^{<1}$, $f_t^{<3}$, $f_t^{<6}$ and $f_t^{<12}$. If the outflow rate is constant over the duration of unemployment, then these measure will coincide. Empirically, we would however expect the outflow rate to be decreasing in the duration of unemployment, either because of true duration dependency or dynamic sorting where the better unemployed leave unemployment at a faster rate. When this is the case, we should find that $f_t^{<1} > f_t^{<3} > f_t^{<6} > f_t^{<12}$, since measures based on higher durations will overweight the long-term unemployed. It is therefore preferable to use $f_t^{<1}$, but this measure can be quite noisy, since it is only based on a small sample of the unemployed. It will be more noisy for countries with a lower turnover where only a small fraction of the unemployed, at any given time, have a duration of less than a month. Using the constructed survey data for Denmark, I find that $f_t^{<1} \simeq f_t^{<3} > f_t^{<6} > f_t^{<12}$ and therefore I use $f_t^{<3}$ as a less noisy measure of the outflow rate.\(^{19}\)
The estimated inflow and outflow rates are then used to decompose unemployment fluctuations based on the approximation in equation (3). Interestingly, the steady state approximation is now almost perfect for the annual survey data, and also very precise for the quarterly survey data as seen in Figure 15. And it also appears that the contributions from the inflow rate is much closer to the actual unemployment rate, compared to when using the register data. In Table 5 I sum up the decomposition results based on survey data, using Shimer’s decomposition method and an HP trend with the smoothing parameter set to 100,000. The inflow rate is seen to contribute substantially and is in general able to account for around half of the fluctuations in unemployment. The more noisy measure based on $f_t^{<1}$ shows less importance for the inflow rate and the more biased measure based on $f_t^{<12}$ shows even more importance for the inflow rate. Note also, that there is no relevant differences between the quarterly survey data, and the OECD-like annual survey data.

The overall result is, that the inflow seems much more important when using survey data, compared to using register data. My explanation for this difference is that the inflow rate constructed from survey data differs from the inflow rate from register data. The survey inflow rate is actually deduced as a residual to make stocks and flows consistent. The outflow is inferred from the data, and the inflow is then set to fit the future stock of unemployed. For this reason, it is not surprising that the steady state approximation is very good, since it is among other things based on the inflow rate, which is set to fit the actual unemployment rate. In the register data, the inflow rate (and the outflow rate) is instead constructed from observed flows between $E$ and $U$. One reason for the discrepancy between the inflow rates between the to types of data, is that the survey methods rely heavily on the assumption that workers only flow between $E$ and $U$. When the inflow rate is deduced, it picks up all the fluctuations not accounted for by the outflow rate, and since outflow rate accounts for around half of the variation, the same is found for inflow rate. In reality however, the contribution assigned to the inflow rate is the sum of all other types of flows, including flows in and out of the labour force.

In the Danish case, the difference between the type of data used turns out to be substantial. This does not necessarily mean that the same difference would be found in U.S. data. More importantly, the failure to reject the hypothesis is not enough to justify the hypothesis in the Danish case, where there are substantial evidence in the literature for duration dependence.
When Shimer (2007) finds that the outflow rate accounts for around 95% of the unemployment fluctuations during the last couple of decades, then it does not really matter whether the remaining 5% is a consequence of fluctuations in the inflow rate or fluctuations in the other flow rates. But for other OECD countries where flows in and out of the labour force are more relevant, then the results based on survey data should probably be interpreted with caution. An inflow contribution which also captures the contribution from flows that goes through $N$ may be interesting in its own right, but should not be interpreted as the separation rate from $E$ to $U$.

6 Conclusion

In this paper I have used register data to analyze the ins and outs of unemployment in Denmark. I decompose the fluctuations of unemployment and find that the outflow rate is much more important than the inflow rate. A more detailed decomposition shows that flows in and out of the labour market are important for an accurate characterization the dynamics of the Danish labour market. I also find that the business cycle changes in the outflow rate cannot be explained by changes in the composition of unemployed workers. The results therefore provide further evidence for the ’New View’ of unemployment dynamics described by Hall (2005a, 2005b, 2007) and Shimer (2007).

I use the Danish register data to construct pseudo-survey data to analyze the importance of the type of data. Interestingly, I find that this has a large impact on the results. When I apply the methods widely used in the literature to the constructed survey data, I find that the contribution from the inflow rate is overstated. I argue that the reason is that the inflow rate is essentially constructed to pick up the part of the variation which is not accounted for by the outflow rate. In the analysis based on register data, where the inflow is observed directly, it is found to be less important for unemployment fluctuations.

Finally, I analyze the importance of some methodological assumptions that vary across different decomposition methods suggested in the recent literature. With respect to the detrending method, I find that this has some influence on the results, but it does not change the main conclusions. I also argue that I find the HP filter with a high smoothing parameter to be most appropriate since this preserves more of the interesting cyclical variation in the
series. A question left out in this paper is the long term fluctuations in the key variables.\textsuperscript{20} It is for instance quite noticeable that the inflow rate to unemployment has decreased steadily over the entire period considered in this paper. These long term changes are most likely determined by demographics and institutional changes, but in this paper I have kept the focus on variations at business cycle frequencies.

Another methodological assumption considered is the importance of the more precise decomposition methods suggested by Elsby, Michaels & Solon (2009) and Fujita & Ramey (2007) that takes the covariance of the inflow and outflow contributions into account. In the Danish data, this does not alter results markedly.

Appendix

Solving the problems for $EN$ and $NE$ transitions around new year

When an $EN$ transition is observed in the first week of a year it can either be a true transition that actually takes place in this week or a spurious transition. A spurious transition in this case means that I know that the transition has taken place, but I do not observe the exact timing of the transition. The problem in the first week of a year arise if the annual data on employment spells has a missing start or end week. I may for instance observe that a worker is in $E$ towards the end of the previous year and that the same worker is not working at all in the following year. If the information about the employment spell has a missing end week, then the transition is placed in the first week of the new year where I know for sure that the worker is no longer in $E$. But it is clear from the data, that there are too many $EN$ and $NE$ transitions in the first week of a new year\footnote{In monthly data the mean number of $EN$ transitions in the first month of a year is 19.780 and 1.252 in other months. For $NE$ transitions the corresponding means are 9.565 and 1.605.}, so I propose the following adjustment which basically assumes that the missing end week of the $E$ spell in the previous year in fact should be placed somewhere in between the last week where the worker was observed in some other state and the last week of the year, as illustrated in Figure 16. The solution is then to simulate a new transition week assuming a uniform distribution over the feasible period.

A similar adjustment is made for $NE$ transitions in the first week of a new year, where I know that the true transition takes place somewhere between the first week of the year and the first week in which I observe the worker to be in another state. Since some transitions truly take place in the first week of a year, I leave an average number of $EN$ and $NE$ transitions in this week and simulate new transition weeks for the remaining.

The consequences of the adjustments are seen in Figures 17 and 18 where I graph the unadjusted and adjusted numbers of $EN$ and $NE$ transitions during the year 1986, which illustrate the typical pattern observed over all the years.

$EN$ transitions are moved from the first month of the year to the months leading up to the year, whereas $NE$ transitions are moved from the first month of the year to the following couple of months. A drawback of this adjustment method is that it implies a minor decrease
in the number of workers in $E$ and a corresponding increase in workers in $N$. This actually
goes in the wrong direction in terms of matching the series for $E$ and $N$ to the official series,
but as shown in Figures 19 and 20 the difference is quite small.
References


<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>Denmark</th>
<th>U.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>209,693</td>
<td>6,209</td>
<td>196,634</td>
<td>218,380</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>21,211</td>
<td>6,005</td>
<td>12,199</td>
<td>32,617</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>114,179</td>
<td>4,443</td>
<td>105,840</td>
<td>125,678</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU</td>
<td>2,663</td>
<td>489</td>
<td>1,306</td>
<td>4,409</td>
<td>15%</td>
<td>13%</td>
</tr>
<tr>
<td>EN</td>
<td>2,788</td>
<td>297</td>
<td>1,891</td>
<td>4,730</td>
<td>16%</td>
<td>25%</td>
</tr>
<tr>
<td>UE</td>
<td>3,126</td>
<td>610</td>
<td>1,908</td>
<td>5,211</td>
<td>18%</td>
<td>14%</td>
</tr>
<tr>
<td>UN</td>
<td>3,225</td>
<td>619</td>
<td>1,980</td>
<td>5,405</td>
<td>18%</td>
<td>11%</td>
</tr>
<tr>
<td>NE</td>
<td>2,249</td>
<td>310</td>
<td>1,643</td>
<td>3,881</td>
<td>13%</td>
<td>24%</td>
</tr>
<tr>
<td>NU</td>
<td>3,638</td>
<td>795</td>
<td>1,923</td>
<td>6,240</td>
<td>21%</td>
<td>13%</td>
</tr>
</tbody>
</table>

| Unemployment rate | 0.092 | 0.026 | 0.053 | 0.142 |

Table 1: Monthly series for the number of workers in E, U and N and the number of transitions between these states. 'EU' refers to a transition from E to U, and the other transitions are defined similarly. The series are seasonally adjusted using X-12-ARIMA. The sample period is 1985M2-2003M12 and covers a 10% random sample from the population. 'Fraction of flows' denotes the number of flows divided by the sum of the six flow types. U.S. numbers are calculated using the means for monthly CPS flows presented in Boon, Carson, Faberman & Ilg (2008) p. 10.
Table 2: Decomposition of unemployment rate fluctuations based on two state approximations. See the text for details about the decomposition methods. The series are seasonally adjusted using X-12-ARIMA. '1.diff.' means that the series is detrended around values in the previous period. 'HP 1,600' and 'HP 100,000' means that the series is detrended using an HP filter with the smoothing parameter set to either 1,600 or 100,000.

<table>
<thead>
<tr>
<th></th>
<th>Full period</th>
<th>Rising unemployment</th>
<th>Falling unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. diff. HP 1,600 HP 100,000</td>
<td>1. diff. HP 1,600 HP 100,000</td>
<td>1. diff. HP 1,600 HP 100,000</td>
</tr>
<tr>
<td><strong>Shimer’s method</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full approximation</td>
<td>0.70 0.78</td>
<td>0.71 0.77</td>
<td>0.69 0.78</td>
</tr>
<tr>
<td>Contribution from inflow</td>
<td>0.09 0.08</td>
<td>0.05 0.03</td>
<td>0.12 0.14</td>
</tr>
<tr>
<td>Contribution from outflow</td>
<td>0.54 0.66</td>
<td>0.51 0.68</td>
<td>0.58 0.63</td>
</tr>
<tr>
<td><strong>Fujita &amp; Ramey’s method</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full approximation</td>
<td>0.76 0.71</td>
<td>0.62 0.71</td>
<td>0.85 0.77</td>
</tr>
<tr>
<td>Contribution from inflow</td>
<td>0.31 0.08</td>
<td>0.35 0.02</td>
<td>0.34 0.15</td>
</tr>
<tr>
<td>Contribution from outflow</td>
<td>0.46 0.63</td>
<td>0.31 0.69</td>
<td>0.51 0.63</td>
</tr>
<tr>
<td><strong>Elsby, Hobijn &amp; Sahin’s Method</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full approximation</td>
<td>0.68 0.85</td>
<td>0.59 0.77</td>
<td>0.58 0.71</td>
</tr>
<tr>
<td>Contribution from inflow</td>
<td>0.08 0.04</td>
<td>0.10 -0.03</td>
<td>0.14 0.15</td>
</tr>
<tr>
<td>Contribution from outflow</td>
<td>0.60 0.87</td>
<td>0.58 0.87</td>
<td>0.44 0.56</td>
</tr>
</tbody>
</table>

Table 3: Decomposition of unemployment rate fluctuations based on 3 state approximations. See the text for details about the decomposition method. The series are seasonally adjusted using X-12-ARIMA. ‘HP 1,600’ and ‘HP 100,000’ means that the series is detrended using an HP filter with the smoothing parameter set to either 1,600 or 100,000.

<table>
<thead>
<tr>
<th></th>
<th>Full period</th>
<th>Rising unemployment</th>
<th>Falling unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HP 1,600 HP 100,000</td>
<td>HP 1,600 HP 100,000</td>
<td>HP 1,600 HP 100,000</td>
</tr>
<tr>
<td><strong>Shimer’s method</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full approximation</td>
<td>0.93 1.04</td>
<td>1.02 1.13</td>
<td>0.84 0.93</td>
</tr>
<tr>
<td>Contribution from EU flows</td>
<td>0.06 0.06</td>
<td>0.04 0.02</td>
<td>0.09 0.10</td>
</tr>
<tr>
<td>Contribution from EN flows</td>
<td>0.02 0.07</td>
<td>0.04 0.07</td>
<td>0.01 0.06</td>
</tr>
<tr>
<td>Contribution from UE flows</td>
<td>0.42 0.52</td>
<td>0.42 0.54</td>
<td>0.43 0.49</td>
</tr>
<tr>
<td>Contribution from UN flows</td>
<td>0.12 0.11</td>
<td>0.06 0.09</td>
<td>0.17 0.14</td>
</tr>
<tr>
<td>Contribution from NU flows</td>
<td>0.05 0.06</td>
<td>0.06 0.09</td>
<td>0.05 0.03</td>
</tr>
<tr>
<td>Contribution from NE flows</td>
<td>0.12 0.13</td>
<td>0.17 0.16</td>
<td>0.06 0.09</td>
</tr>
</tbody>
</table>
### Table 4: Decomposition of unemployment rate fluctuations for subgroups. The two state decomposition shows the contributions from the inflow rate and the outflow rate. The three state decomposition shows the contributions from all six flow rates between the three states. All results are based on Shimer’s decomposition method with the HP parameter set to 100,000. The series are seasonally adjusted using X-12-ARIMA. See the text for further details.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Outflow</th>
<th>Inflow</th>
<th>Full approx.</th>
<th>EU</th>
<th>EN</th>
<th>UE</th>
<th>UN</th>
<th>NE</th>
<th>NU</th>
<th>Full approx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.66</td>
<td>0.08</td>
<td>0.78</td>
<td>0.06</td>
<td>0.07</td>
<td>0.52</td>
<td>0.11</td>
<td>0.06</td>
<td>0.13</td>
<td>1.04</td>
</tr>
<tr>
<td>Age 15-30</td>
<td>0.53</td>
<td>0.18</td>
<td>0.74</td>
<td>0.11</td>
<td>0.06</td>
<td>0.37</td>
<td>0.11</td>
<td>0.08</td>
<td>0.13</td>
<td>0.96</td>
</tr>
<tr>
<td>Age 31-50</td>
<td>0.72</td>
<td>0.03</td>
<td>0.77</td>
<td>0.02</td>
<td>0.06</td>
<td>0.61</td>
<td>0.09</td>
<td>0.11</td>
<td>0.09</td>
<td>1.07</td>
</tr>
<tr>
<td>Age 51-70</td>
<td>0.94</td>
<td>-0.04</td>
<td>0.90</td>
<td>-0.03</td>
<td>0.11</td>
<td>1.00</td>
<td>0.08</td>
<td>0.09</td>
<td>0.11</td>
<td>1.50</td>
</tr>
<tr>
<td>Female</td>
<td>0.81</td>
<td></td>
<td>0.81</td>
<td>0.00</td>
<td>0.08</td>
<td>0.62</td>
<td>0.12</td>
<td>0.07</td>
<td>0.10</td>
<td>1.08</td>
</tr>
<tr>
<td>Male</td>
<td>0.53</td>
<td>0.17</td>
<td>0.75</td>
<td>0.18</td>
<td>0.02</td>
<td>0.39</td>
<td>0.10</td>
<td>0.03</td>
<td>0.21</td>
<td>0.99</td>
</tr>
<tr>
<td>Female, Age 15-30</td>
<td>0.63</td>
<td>0.10</td>
<td>0.75</td>
<td>0.06</td>
<td>0.07</td>
<td>0.43</td>
<td>0.13</td>
<td>0.09</td>
<td>0.12</td>
<td>0.99</td>
</tr>
<tr>
<td>Female, Age 31-50</td>
<td>0.89</td>
<td>-0.07</td>
<td>0.81</td>
<td>-0.06</td>
<td>0.07</td>
<td>0.74</td>
<td>0.11</td>
<td>0.13</td>
<td>0.07</td>
<td>1.15</td>
</tr>
<tr>
<td>Female, Age 51-70</td>
<td>1.03</td>
<td>-0.03</td>
<td>0.99</td>
<td>-0.03</td>
<td>0.11</td>
<td>1.03</td>
<td>0.07</td>
<td>0.09</td>
<td>0.08</td>
<td>1.50</td>
</tr>
<tr>
<td>Male, Age 15-30</td>
<td>0.45</td>
<td>0.24</td>
<td>0.74</td>
<td>0.16</td>
<td>0.05</td>
<td>0.32</td>
<td>0.10</td>
<td>0.07</td>
<td>0.13</td>
<td>0.93</td>
</tr>
<tr>
<td>Male, Age 31-50</td>
<td>0.56</td>
<td>0.14</td>
<td>0.76</td>
<td>0.11</td>
<td>0.05</td>
<td>0.49</td>
<td>0.08</td>
<td>0.08</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Male, Age 51-70</td>
<td>0.82</td>
<td>-0.01</td>
<td>0.81</td>
<td>-0.05</td>
<td>0.10</td>
<td>1.04</td>
<td>0.11</td>
<td>0.22</td>
<td>0.11</td>
<td>1.64</td>
</tr>
<tr>
<td>Native Danes</td>
<td>0.66</td>
<td>0.08</td>
<td>0.78</td>
<td>0.06</td>
<td>0.06</td>
<td>0.53</td>
<td>0.10</td>
<td>0.07</td>
<td>0.12</td>
<td>1.04</td>
</tr>
<tr>
<td>Immigrants</td>
<td>0.67</td>
<td>0.15</td>
<td>0.84</td>
<td>0.08</td>
<td>0.18</td>
<td>0.36</td>
<td>0.19</td>
<td>0.02</td>
<td>0.11</td>
<td>1.05</td>
</tr>
<tr>
<td>Married</td>
<td>0.77</td>
<td>0.01</td>
<td>0.81</td>
<td>0.01</td>
<td>0.10</td>
<td>0.69</td>
<td>0.09</td>
<td>0.07</td>
<td>0.12</td>
<td>1.18</td>
</tr>
<tr>
<td>Not Married</td>
<td>0.60</td>
<td>0.12</td>
<td>0.76</td>
<td>0.06</td>
<td>0.04</td>
<td>0.43</td>
<td>0.13</td>
<td>0.08</td>
<td>0.12</td>
<td>0.97</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.73</td>
<td>0.04</td>
<td>0.80</td>
<td>0.02</td>
<td>0.06</td>
<td>0.49</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
<td>1.01</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.59</td>
<td>0.15</td>
<td>0.78</td>
<td>0.12</td>
<td>0.08</td>
<td>0.55</td>
<td>0.06</td>
<td>0.09</td>
<td>0.09</td>
<td>1.09</td>
</tr>
<tr>
<td>Professionals</td>
<td>0.58</td>
<td>0.12</td>
<td>0.72</td>
<td>0.09</td>
<td>0.11</td>
<td>0.54</td>
<td>0.07</td>
<td>0.11</td>
<td>0.08</td>
<td>1.08</td>
</tr>
<tr>
<td>Agriculture &amp; Fishery</td>
<td>0.77</td>
<td>0.03</td>
<td>0.84</td>
<td>-0.10</td>
<td>0.06</td>
<td>0.80</td>
<td>0.16</td>
<td>0.10</td>
<td>0.04</td>
<td>1.20</td>
</tr>
<tr>
<td>Industry</td>
<td>0.54</td>
<td>0.17</td>
<td>0.79</td>
<td>0.12</td>
<td>0.02</td>
<td>0.45</td>
<td>0.09</td>
<td>0.14</td>
<td>0.10</td>
<td>1.04</td>
</tr>
<tr>
<td>Utility &amp; Construction</td>
<td>0.50</td>
<td>0.27</td>
<td>0.87</td>
<td>0.42</td>
<td>0.02</td>
<td>0.30</td>
<td>0.01</td>
<td>0.03</td>
<td>0.17</td>
<td>1.02</td>
</tr>
<tr>
<td>Trade</td>
<td>0.62</td>
<td>0.12</td>
<td>0.78</td>
<td>0.08</td>
<td>0.06</td>
<td>0.45</td>
<td>0.11</td>
<td>0.09</td>
<td>0.12</td>
<td>1.03</td>
</tr>
<tr>
<td>Hotels, Restaurants &amp; Catering</td>
<td>0.57</td>
<td>0.24</td>
<td>0.79</td>
<td>0.15</td>
<td>0.08</td>
<td>0.37</td>
<td>0.15</td>
<td>0.06</td>
<td>0.19</td>
<td>1.02</td>
</tr>
<tr>
<td>Transportation</td>
<td>0.73</td>
<td>0.02</td>
<td>0.80</td>
<td>0.01</td>
<td>0.04</td>
<td>0.60</td>
<td>0.12</td>
<td>0.16</td>
<td>0.03</td>
<td>1.10</td>
</tr>
<tr>
<td>Finance, Insurance &amp; Property Service</td>
<td>0.65</td>
<td>0.09</td>
<td>0.76</td>
<td>0.06</td>
<td>0.01</td>
<td>0.49</td>
<td>0.13</td>
<td>0.11</td>
<td>0.10</td>
<td>0.99</td>
</tr>
<tr>
<td>Public Adm. &amp; Health Care</td>
<td>0.74</td>
<td>0.01</td>
<td>0.76</td>
<td>0.01</td>
<td>0.11</td>
<td>0.58</td>
<td>0.11</td>
<td>0.08</td>
<td>0.08</td>
<td>1.07</td>
</tr>
<tr>
<td>Education</td>
<td>0.54</td>
<td>0.31</td>
<td>0.85</td>
<td>0.25</td>
<td>0.04</td>
<td>0.47</td>
<td>0.08</td>
<td>0.01</td>
<td>0.11</td>
<td>0.97</td>
</tr>
</tbody>
</table>
| Other          | 0.25    | 0.54   | 0.78         | 0.14| 0.30| 0.07| 0.23| 0.31| 0.11| 1.10         

### Table 5: Decomposition of unemployment rate fluctuations using different types of constructed survey data. The two state decomposition shows the contribution from the inflow rate and the outflow rate. The OECD-like survey data covers the period 1986-2004 and the quarterly survey data covers the period 1985Q2-2003Q4. The quarterly series are de-seasoned using the X-12-Arima procedure. All series are detrended using a HP filter with the smoothing parameter set to 100,000. ’f1’ means that the outflow rate, f, is calculated using the number of unemployed with a duration of less than 1 month of unemployment and similarly for f3, f6 and f12. See the text for further details.

<table>
<thead>
<tr>
<th>OECD-like Survey Data (annual)</th>
<th>Quarterly Survey Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f1</td>
</tr>
<tr>
<td><strong>Shimer’s method</strong></td>
<td></td>
</tr>
<tr>
<td>Full approximation</td>
<td>1.00</td>
</tr>
<tr>
<td>Contribution from Inflow</td>
<td>0.32</td>
</tr>
<tr>
<td>Contribution from Outflow</td>
<td>0.62</td>
</tr>
</tbody>
</table>
Figure 1: Time series for the unemployment rates in the U.S. and Denmark over the period 1966-2007. The shaded area shows the period analyzed in this paper. The Danish unemployment rate is from Statistics Denmark and the U.S. unemployment rate is from the Bureau of Labor Statistics.

Figure 2: Quarterly averages of the monthly unemployment rate calculated using the spell data presented in this paper and the official unemployment rate from Statistics Denmark over the period 1985Q2-2009Q1.
Figure 3: Quarterly averages of the monthly inflow probability. HP trends are shown with the smoothing parameter set to 1,600 and 100,000. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.

Figure 4: Quarterly averages of the monthly outflow probability. HP trends are shown with the smoothing parameter set to 1,600 and 100,000. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.
Figure 5: Two and three state approximations based on equations (3) and (4). Quarterly averages of monthly data. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.

Figure 6: Elsby, Hobijn & Sahin (2009) steady state and non steady state approximations based on equations (10) and (11). Quarterly averages of monthly data. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.
Figure 7: Unemployment rate and HP trends with the smoothing parameter set to 1,600 or 100,000. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.

Figure 8: Cyclic unnatural unemployment rate after detrending. The series are seasonally adjusted using X-12-ARIMA. '1.diff.' means that the series is detrended around values in the previous period. 'HP 1,600' and 'HP 100,000' means that the series is detrended using an HP filter with the smoothing parameter set to either 1,600 or 100,000. Sample period is 1985Q2-2003Q4.
Figure 9: Actual unemployment and the contribution from the inflow rate. See the text for details. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.

Figure 10: Actual unemployment and the contribution from the outflow rate. See the text for details. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.
Figure 11: Actual unemployment and contributions from six different flow types. See the text for details. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.
Figure 12: Outflow rates for Immigrants and Native Danes. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.

Figure 13: Number of unemployed immigrants as a fraction of total unemployment. Sample period is 1985Q2-2003Q4.
Figure 14: Heterogeneous outflow rates. In the 'Outflow Rate based on Real changes' series, the subgroup specific outflow rate is allowed to vary and the relative sizes of the subgroups are fixed to the mean. In the 'Outflow Rate based on Compositional changes' series, the relative sizes of the number of unemployed in the subgroups are allowed to vary and the outflow rates are fixed to the mean within subgroups. The series are seasonally adjusted using X-12-ARIMA. Sample period is 1985Q2-2003Q4.
Figure 15: Decomposition of unemployment fluctuations using survey data. The two state steady state approximation is based on equation (3). The contribution from the Inflow rate is found by fixing the outflow rate to the mean value over time, and the contribution from the Outflow rate is found by fixing the inflow rate at the mean. The three graphs to the left show annual data for the period 1986-2004 and the three graphs to the right show quarterly data for the period 1985Q2-2003Q4. The quarterly series are deseasoned using the X-12-Arima procedure.
Figure 16: Illustration of the adjustment of EN transition in the first week of the year. The end week of the E spell is missing so I simulate an end week in the interval where I know that the transition must take place.

Figure 17: Monthly number of EN transitions in the year 1986, before and after adjustment for EN transitions in the first week of a year.
Figure 18: Monthly number of NE transitions in the year 1986, before and after adjustment for NE transitions in the first week of a year.

Figure 19: Monthly number of workers in E in the year 1986, before and after adjustment for EN and NE transitions in the first week of a year.
Figure 20: Monthly number of workers in N in the year 1986, before and after adjustment for EN and NE transitions in the first week of a year.
PhD Theses:

1999-6 Peter Raahauge, Dynamic Programming in Computational Economics.
1999-7 Torben Dall Schmidt, Social Insurance, Incentives and Economic Integration.
1999 Jørgen Vig Pedersen, An Asset-Based Explanation of Strategic Advantage.
1999 Birgitte Holt Andersen, Structural Analysis of the Earth Observation Industry.
2000-1 Jakob Roland Munch, Economic Integration and Industrial Location in Unionized Countries.
2001-3 Niels C. Beier, Propagation of Nominal Shocks in Open Economies.
2001-4 Mette Verner, Causes and Consequences of Interruptions in the Labour Market.
<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001-6</td>
<td>Søren Vester Sørensen</td>
<td>Three Essays on the Propagation of Monetary Shocks in Open Economies.</td>
</tr>
<tr>
<td>2001-8</td>
<td>Peter Ejler Storgaard</td>
<td>Price Rigidity in Closed and Open Economies: Causes and Effects.</td>
</tr>
<tr>
<td>2002-1</td>
<td>Mette Rose Skaksen</td>
<td>Multinational Enterprises: Interactions with the Labor Market.</td>
</tr>
<tr>
<td>2002-3</td>
<td>Boriss Siliverstovs</td>
<td>Multicointegration, Nonlinearity, and Forecasting.</td>
</tr>
<tr>
<td>2002-4</td>
<td>Søren Tang Sørensen</td>
<td>Aspects of Sequential Auctions and Industrial Agglomeration.</td>
</tr>
<tr>
<td>2002-5</td>
<td>Peter Myhre Lildholdt</td>
<td>Essays on Seasonality, Long Memory, and Volatility.</td>
</tr>
<tr>
<td>2002-6</td>
<td>Sean Hove</td>
<td>Three Essays on Mobility and Income Distribution Dynamics.</td>
</tr>
<tr>
<td>2002</td>
<td>Hanne Kargaard Thomsen</td>
<td>The Learning organization from a management point of view - Theoretical perspectives and empirical findings in four Danish service organizations.</td>
</tr>
<tr>
<td>2003-2</td>
<td>Morten Ørregaard Nielsen</td>
<td>Multivariate Fractional Integration and Cointegration.</td>
</tr>
<tr>
<td>2004-1</td>
<td>Lars Stentoft</td>
<td>Least Squares Monte-Carlo and GARCH Methods for American Options.</td>
</tr>
<tr>
<td>2004-2</td>
<td>Brian Krogh Graversen</td>
<td>Employment Effects of Active Labour Market Programmes: Do the Programmes Help Welfare Benefit Recipients to Find Jobs?</td>
</tr>
<tr>
<td>2004-3</td>
<td>Dmitri Koulikov</td>
<td>Long Memory Models for Volatility and High Frequency Financial Data Econometrics.</td>
</tr>
</tbody>
</table>
2005-1 Julia Chiriaeva, Credibility of Fixed Exchange Rate Arrangements.
2005-5 Ott Toomet, Marginalisation and Discouragement: Regional Aspects and the Impact of Benefits.
2005 Hesham Morten Gabr, Strategic Groups: The Ghosts of Yesterday when it comes to Understanding Firm Performance within Industries?
2006-2 Allan Sørensen, Economic Integration, Ageing and Labour Market Outcomes
2006-3 Philipp Festerling, Essays on Competition Policy
2006-4 Carina Sponholtz, Essays on Empirical Corporate Finance
2006-5 Claus Thrane-Jensen, Capital Forms and the Entrepreneur – A contingency approach on new venture creation
2006-6 Thomas Busch, Econometric Modeling of Volatility and Price Behavior in Asset and Derivative Markets
2007-1 Jesper Bagger, Essays on Earnings Dynamics and Job Mobility
2007-2 Niels Stender, Essays on Marketing Engineering
2007-3 Mads Peter Pilkjær Harmsen, Three Essays in Behavioral and Experimental Economics
2007-4 Juanna Schrøter Joensen, Determinants and Consequences of Human Capital Investments
2007-5 Peter Tind Larsen, Essays on Capital Structure and Credit Risk
<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-1</td>
<td>Toke Lilhauge Hjortshøj, Essays on Empirical Corporate Finance – Managerial Incentives, Information Disclosure, and Bond Covenants</td>
<td>Toke Lilhauge Hjortshøj</td>
</tr>
<tr>
<td>2008-3</td>
<td>David Glavind Skovmand, Libor Market Models - Theory and Applications</td>
<td>David Glavind Skovmand</td>
</tr>
<tr>
<td>2008-4</td>
<td>Martin Seneca, Aspects of Household Heterogeneity in New Keynesian Economics</td>
<td>Martin Seneca</td>
</tr>
<tr>
<td>2008-5</td>
<td>Agne Lauzadyte, Active Labour Market Policies and Labour Market Transitions in Denmark: an Analysis of Event History Data</td>
<td>Agne Lauzadyte</td>
</tr>
<tr>
<td>2009-1</td>
<td>Christian Dahl Winther, Strategic timing of product introduction under heterogeneous demand</td>
<td>Christian Dahl Winther</td>
</tr>
<tr>
<td>2009-2</td>
<td>Martin Møller Andreasen, DSGE Models and Term Structure Models with Macroeconomic Variables</td>
<td>Martin Møller Andreasen</td>
</tr>
<tr>
<td>2009-3</td>
<td>Frank Steen Nielsen, On the estimation of fractionally integrated processes</td>
<td>Frank Steen Nielsen</td>
</tr>
<tr>
<td>2009-4</td>
<td>Maria Knoth Humlum, Essays on Human Capital Accumulation and Educational Choices</td>
<td>Maria Knoth Humlum</td>
</tr>
<tr>
<td>2009-5</td>
<td>Yu Wang, Economic Analysis of Open Source Software</td>
<td>Yu Wang</td>
</tr>
<tr>
<td>2009-6</td>
<td>Benjamin W. Blunck, Creating and Appropriating Value from Mergers and Acquisitions – A Merger Wave Perspective</td>
<td>Benjamin W. Blunck</td>
</tr>
<tr>
<td>2009-7</td>
<td>Rune Mølgaard, Essays on Dynamic Asset Allocation and Electricity Derivatives</td>
<td>Rune Mølgaard</td>
</tr>
<tr>
<td>2009-8</td>
<td>Jens Iversen, Financial Integration and Real Exchange Rates: Essays in International Macroeconomics</td>
<td>Jens Iversen</td>
</tr>
<tr>
<td>2009-9</td>
<td>Torben Sørensen, Essays on Human Capital and the Labour Market</td>
<td>Torben Sørensen</td>
</tr>
<tr>
<td>2009-10</td>
<td>Jonas Staghøj, Essays on Unemployment and Active Labour Market Policies</td>
<td>Jonas Staghøj</td>
</tr>
</tbody>
</table>