Discussion Papers No. 675, January 2012
Statistics Norway, Research Department

Tom Kornstad, Ragnar Nymoen and Terje Skjerpen

Macroeconomic shocks and the probability of being employed

Abstract:
Macroeconomic theories take polar views on the importance of choice versus chance. At the micro level, it seems realistic to assume that both dimensions play a role for individual employment outcomes, although it might be difficult to separate these two effects. Nevertheless the choice and chance dimension are seldom treated symmetrically in models that use micro data. We estimate a logistic model of the probability of being employed among married or cohabitating women that are in the labor force. Besides variables that measure individual characteristics (choice), we allow a full set of indicator variables for observation periods that represent potential effects of aggregate shocks (chance) on job probabilities. To reduce the number of redundant indicator variables as far as possible and in a systematic way, an automatic model selection is used, and we assess the economic interpretation of the statistically significant indicator variables with reference to a theoretical framework that allows for friction in the Norwegian labor market. In addition, we also estimate models that use the aggregate female and male unemployment rates as 'sufficient' variables for the chance element in individual employment outcomes. Data are for Norway and span the period 1988q2-2008q4.

Keywords: Job probability; Automatic model selection; Random utility modeling

JEL classification: C21, J21, J64

Acknowledgements: An earlier version of this paper has been presented at a seminar at the Department of Economics, University of Oslo 14 October 2011, and at a workshop arranged by the Institute for new economic thinking at the Oxford Martin School, 29 October 2011. We would like to thank Arvid Raknerud for valuable comments

Address: Tom Kornstad, Statistics Norway, Research Department. Email: tko@ssb.no

Ragnar Nymoen, University of Oslo, Economics Department and Statistics Norway, Research Department. Email: ragnar.nymoen@econ.uio.no

Terje Skjerpen (Corresponding author), Statistics Norway, Research Department. Email: tes@ssb.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:
Statistics Norway
Telephone: +47 62 88 55 00
E-mail: Salg-abonnement@ssb.no

ISSN 0809-733X
Print: Statistics Norway

Sammendrag

Ulike makroøkonomiske teorier knyttet til arbeidsmarkedet har forskjellig syn på betydningen av valg kontra tilfeldigheter. Når det gjelder hvorvidt individer i arbeidsstyrken er i arbeid, er det rimelig å anta at både valg og tilfeldigheter har betydning. På tross av dette er det sjelden at disse to elementene blir behandlet på en symmetrisk måte i empiriske modeller basert på mikrodata. I denne artikkelen estimeres ulike logit-modeller for sannsynligheten for å være i arbeid gitt at en er i arbeidsstyrken. Modellene estimeres på norske data for gifte/samboende kvinner ved hjelp av kvartalsvise tidsserier av tverrsnittsdata for perioden 1988k2-2008k4. Ved siden av personspesifikke variabler som har betydning for den enkelte kvinnes valg, tillater vi også et komplett sett med indikatorvariabler som representerer potensielle effekter av aggregerte sjokk på sannsynligheten for å være i arbeid. For å finne fram til best mulig modell gjør vi bruk av et softwareprogram for automatisk modellseleksjon. Den økonomiske fortolkningen av de statistisk signifikante indikatorvariabler tar utgangspunkt i et teoretisk rammeverk som tillater friksjoner i det norske arbeidsmarkedet. Som et alternativ til bruken av det komplette settet med indikatorvariabler nevnt ovenfor, presenteres også noen beregninger hvor ulike mål på aggregert arbeidsledighet er med i informasjonssettet. Disse variablenes synes å kunne fange opp all nesten all signifikant kalendervariasjon når det gjelder gifte og samboende kvinners sysselsettingstilbøyelighet.

1. Introduction

In a situation with real wage flexibility and no frictions in the labor market, individuals' probabilities for work and unemployment may be expected to be unaffected by macroeconomic shocks that are common to a large number of workers. However, it is realistic to assume that real world labor markets are characterized by many frictions, and the relevant question is therefore whether individuals are able to adapt in ways that more or less offset the effects of aggregate shocks on their work prospects. This question is relevant for policy designs. If the probabilities for unemployment and work for a large number of workers are affected by aggregate shocks and fluctuations (frictions), the role for countercyclical macroeconomic policies is stronger than if friction effects are empirically irrelevant.

In macroeconomic theory, the standard real business cycle (RBC) model and the search theoretical model represent polar views on the issue about labor market frictions and about the importance of chance versus choice, see Krusell et al. (2010). In the frictionless models in the tradition of Kydland and Prescott (1982), changes in employment are explained by individual choice. In a macroeconomic search model of the Mortensen and Pissarides (1994) type, the emphasis is on chance rather than on choice, in the sense that changes in employment reflect changes in the probability of receiving a job offer.

In an econometric model of the probability of being employed it is unattractive to impose the dichotomy between chance and choice *a priori*, since it seems realistic to assume that both dimensions can play a role for individual employment outcomes. Nevertheless, in the literature on micro-econometric modeling of labor marked behavior, the custom is to concentrate on the choice aspect as captured by measured individual characteristics. That said, Dagsvik *et al.* (2010) report results where their model besides choice variables contains year-dummies that are intended to capture effects stemming from the business cycle. In this paper, we treat the choice and chance dimension symmetrically in the unrestricted model formulation, and we test econometrically, for married and cohabitating women in the work force, the hypothesis that the probability of being employed depends on the business cycle.

Our test is based on the assumption that if chance matters, fortunate and unfortunate episodes will be linked to fluctuations at the aggregate level of the economy. The data set is a sample of independent cross-sections for married and cohabitating women in the Norwegian labor force covering the period from 1988q2 to 2008q4.[1] The reason for focusing on married and cohabiting women is that empirical

[1] Realistically, the result can to some degree be sample dependent. We therefore also make use of a second data set that has been sampled in the same manner as the first one.

analyses typically find that these women are more responsive to policy changes than their male partners. One important reason is that women often take more responsibility for family and children, and thus they have stronger preferences for home work. Since their male partner is participating in the labor market in most cases, household incomes do not drop to zero even if the female is not working. Employing data for persons in the workforce may be interpreted as a strong test of the importance of frictions, since such persons have a strong tie to the labor market from the outset. In our sample there are 82 potential periods in which macroeconomic shocks might occur. Since the number of observations is large, 50 487, we might in principle estimate a general model that includes a dummy variable for each potential break together with the variables that measure individual attributes (education length and the number of children in different age groups for example). However, all periods in the sample are not likely to be equally important when it comes to friction. The methodological task is therefore to find the significant impulse period dummies objectively, and to retain in the final model only those dummies that represent significant frictional effects of macroeconomic shocks. We use the computer based automatic model selection algorithm Autometrics (see Doornik, 2009) as our main tool in the testing of the hypothesis that aggregate shocks (as represented by dummies) have no effect on the individual probability of being employed.

As a background, it is interesting to note that although Norway is often regarded as an “oil-driven” economy that is characterized by even growth, our sample contains periods where there have been large changes in job-creation and in job-destruction. At the start of our sample, in 1988, employment growth was still positive, following the credit led boom that started in 2003. But during 1988 the housing market collapsed. Real house prices fell by 40 percent from the first quarter of 1988 to the first quarter of 1993. There was a major banking crisis, and the first years of the 1990s were marked by financial consolidation among households and by low growth. During this period there was a sharp rise in the aggregate unemployment rate, and unemployment spells became longer. Employment growth also became weak and negative during the first five years of the new millennium, but then a period with unprecedented high employment growth started in 2005. A significant part of the increase in employment was made up of temporary as well as more permanent immigration of workers from East Europe, for instance Poland and the Baltic States. Our sample ends at the start of the international financial crisis, and a drop in the growth rate is visible at that point.

The rest of the paper is organized as follows: In section 2 we give our model and state our hypotheses. A description of the data set is given in section 3. Section 4 is devoted to automatic model selection.

Our empirical results are reported in sections 5–8. Section 9 concludes. An Appendix contains summary statistics for our main data set.

2. Logit model with variables representing frictions

As noted above, it is of interest to investigate whether the probability of being unemployed depends on macroeconomic fluctuations or intermittent shocks that are exogenous to the individual, but common to all employed and job-seeking married and cohabitating women. With reference to a theoretical model that includes the separation probability and the employment opportunity arrival rate, one way to introduce aggregate shocks is to allow both of the two rates to be non-constant as a result of macro-economic events. In the following we refer to such variations as *frictions*, cf. Krusell *et al.* (2010).

We investigate the friction hypothesis econometrically within the framework of a standard logit model. Assume the agent is searching for employment. When receiving a particular job offer, the agent compares the utility of the arriving job offer and the expected utility of continued search. In this comparison the female uses her perceptions about the job arrival rate and job separation rate. These rates depend on the skills of the agent (education and work experience), the functioning of the labor market including exogenous shocks and business cycles. In addition comparisons of utilities are influenced by the agent's non-labor income and the number of children in different age groups in the family.

Let q^* be the difference between the utility of the arriving job offer and the expected utility of continued search. In what follows we will assume that this difference can be modelled as

$$q^* = X_1\delta_1 + X_2\delta_2 + \varepsilon, \tag{1}$$

where X_1 includes years of schooling, experience, experience squared, number of children in three age groups,[2] a binary variable for urbanity and the logarithm of real non-labor income and with δ_1 as the corresponding vector of coefficients (including an intercept). Moreover, X_2 is a (row-)vector consisting of variables that capture joint fluctuations in the employment opportunity arrival rate and the job separation rate (at present, we have no ambition to identifying separate effects of the two friction parameters), and δ_2 denotes the associated parameter vector. ε denotes a random error term that is included to capture the effects of variables that are latent to the researcher, but known by the agent.

[2] The three age groups we consider are 0-3 years, 4-6 years and above 6 years.

While q^* is a latent variable that cannot be observed in our data, what we observe is whether the female is employed (job offer is accepted) or unemployed (job offer is not accepted). Therefore, what we observe is the indicator variable

$$\begin{aligned} q &= 1 \quad \text{if } q^* > 0 \\ q &= 0 \quad \text{otherwise.} \end{aligned} \tag{2}$$

Assuming ε is logistically distributed, the probability of being employed given that the female is in the labor force (q), is given by the well known logit model,

$$q = \frac{1}{1+\exp\left(-X_1\delta_1 - X_2\delta_2\right)}, \tag{3}$$

cf. Dagsvik, Kornstad and Skjerpen (2010) for a similar specification. Alternatively, the model can be written with $1-q$ on the left-hand side, which is the probability of being unemployed for agents with characteristics X_1 who experience the macroeconomic shock represented by a significant element in X_2. In both interpretations of the model, the null hypothesis we test is H_0: $\delta_2 = 0$.

In the following we make use of two alternative model specifications for estimating the effects of aggregate shocks. First, we include impulse dummies for each period of observation. In practice this leads to a model with 91 parameters which is not in itself a problem since we have more than 50K observations in each of the two data sets that we use. In principle, in the case of no friction, we should be unable to reject the hypothesis H_0: $\delta_2 = 0$ at the usual level of significance.

In the case of rejection of the null hypothesis, it is of interest to take the analysis one step further and investigate whether there is a sub-set of impulse dummies which explains the rejection. In fact this is theoretically reasonable, since it seems unlikely that the seekers of (new) jobs are at all times equally affected by friction. We therefore perform a model reduction by sifting out the significant dummy variables from the insignificant ones. In order to do the general-to-specific (GETS) modeling in an objective way (that can be replicated) we make use of the automatic modeling feature of PcGive, see Doornik and Hendry (2009). This approach is discussed in section 4 below, and the results are reported in sections 5 and 6.

The second approach is more direct, and is based on representing the fluctuations with the aid of one, or a few, observed macro variables which are correlated with the changes in both the employment opportunity arrival rate and the separation rate. In this paper we have used aggregate unemployment rates for women and men, see section 7.

In the case of rejection of the null hypothesis of no friction, it is of interest to assess the numerical significance of the retained dummies. We are interested in their estimated impact on the employment probability of an individual married or cohabitating woman with given characteristic, but also the aggregated implications in the form of expected increase in unemployment. Clearly, even if the changes in the individual probabilities are quite tiny, the expected change in the number of unemployed persons might be significant and thus more interesting from a practical point of view.

We report results for estimation on two samples of independent cross-sections for married and cohabitating women in Norway, covering the period 1988q2 to 2008q4. In the next section, we give a description of the main data set, whereas the other one is regarded more as a supplementary data set.

3. The data sets

The data set was obtained by merging the Labor Force Survey (LFS) 1988–2008 with three different register data sets—the Tax Register for personal tax payers, the Tax Return Register and the National Education database—with additional information about incomes, family composition and education.[3] While the Tax Return Register is our primary source of information about incomes and family formation covering the years 1993−2008, we have also included data from the Tax Register for the years 1988−1992 in order to include a period of more fluctuations in the unemployment rate.

The classification in the LFS is based on answers to a broad range of questions. Persons are asked about their attachment to the labor market during a particular week. For a person to be defined as unemployed, she must not be employed in the survey week, she must have been seeking work actively during the preceding four weeks, and she must wish to return to work within the next two weeks.

Unfortunately, the Tax Register for personal taxpayers does not include very detailed information about different types of incomes. We have chosen to use a measure of non-labor income that includes salaries of the husband as well as stipulated labor incomes for self-employed husbands. The nominal non-labor income variable is deflated by the official Norwegian consumer price index, with 1998 as the base year.

[3] This is possible owing to a system with unique personal identification numbers for every Norwegian citizen.

Education is measured in years of achieved level of schooling and experience is defined as age minus schooling minus age at school start. An area is defined as densely populated if at least 200 persons live in the area and the distance from one house to another normally is less than 50 meter. The age distribution of the children is considered by measuring the number of children aged 0−3 years, 4−6 years and 7−18 years, respectively.

Only married or cohabiting females ranging in age between 25 and 60 years are included in the sample. The motivation for the age restriction is that education is an important activity for women younger that 25 years, and that for those older than 60 years, early retirement is rather frequent.[4] From this sample we have also excluded self-employed women and women without non-labor income or with very high non-labor income (more than one million NOK in real terms).

In the estimations we apply two different samples in order to illustrate some aspects of sample variability effects and robustness. The actual samples we use are a subset of the LFS and consist of independent cross-sections for all quarters from 1988q2 to 2008q4. Both samples are selected such that each woman is observed in one quarter only, i.e., there is no dependency over time among records due to repeated observations of a particular woman. Choice of quarter is randomly determined.

The first sample (Sample I) obtained by this procedure includes 50,487 females to be used in the estimation of the logistic regression. Out of these 1,202 females are unemployed. The Appendix provides additional information about this sample.

Since Sample I was randomly drawn from a much larger sample where each female can be observed several times, it is possible to draw a second sample (Sample II) with the same females as in Sample I, but where the females might be observed in a different period.[5] The reason we do this is to investigate how robust the obtained empirical results are using the first data set. Sample II shows no noteworthy differences in the characteristics of the persons from Sample I. Mainly due to transitions over time from labor market participation to being out of the labor market, and vice versa, the number of observations in the two samples are not identical, but the difference is rather small (about 200 observations).

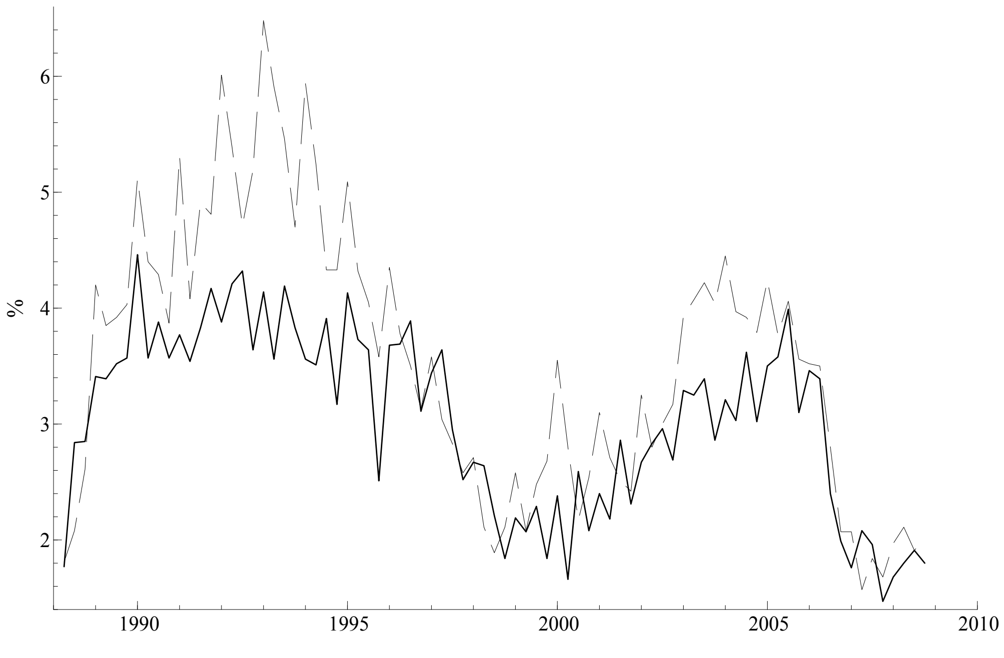
---

[4] Norway has an early retirement program for workers. It was introduced for the first time in 1988, originally only for 66 years old workers working in firms that were participating in the program. Today the program covers most workers aged 62−66 years.

[5] Since one of the aims of the current paper is to conduct model selection using Autometrics and since this routine does not handle a panel data design, we retain only one observation for each observational unit and do not utilize the panel data dimension in our analysis.

In addition to the individual specific variables described above, in some of the estimations we also introduce the aggregate unemployment rate for females and/or males. These variables are constant across all individuals observed in the same time period in the sample, but vary across quarters.[6] Figure 2 displays the unemployment rates over our sample period. In 1989 and 1990 the two unemployment rates rose from very low levels. This was the time of major credit and banking crisis in Norway, collapse in the housing market (real prices fell by 42 percent from 1988q1 to 1993q1), financial consolidation among households, and also a decline in the most important export markets. It is interesting to note though that after the initial rise, the female unemployment percentage levelled off long before the male unemployment percentage. On the other hand, it also took a longer time of economic recovery before the female rate fell decisively in 1996. Both rates increased gradually from 2000 to 2005, and then there was a very marked decline in both rates in the years before the financial crisis in 2008. This was a period of almost unprecedented employment growth in Norway.

Figure 1. Aggregate unemployment rates by gender, percent. The solid line shows the female rate, and the dashed line shows the male rate



[6] Data on the unemployment rates are provided by the macro economic research group, Statistics Norway. They only cover persons aged 25-59 years, but in addition to married/cohabitating persons they also include observations on single persons.

4. Automatic model selection

Logit estimation of (3) on our data set involves 91 parameters, since we have 8 variables with individual characteristics and 82 dummy variables (1988q2 is the reference quarter) and a constant term in the equation. In line with conventional terminology we refer to this model as the general unrestricted model (GUM). Since the number of observations is large, there is no degree-of-freedom problem involved in estimating such a model specification. For example, if the specified person specific explanatory variables are of importance for the probability of being employed, their parameters will almost certainly be estimated with some precision in large data set like ours. However, for the effects of the time period dummies one will typically obtain mixed results. Some, but not all 82 indicator variables may be significant when conventional t-tests are used for example. This motivates that we search for simplifications of the GUM with the aim of obtaining a final model that retains only the period dummies which represent significant effect of frictions on individual job probabilities.

Procedures for model specification are an old theme which involves several issues in econometrics. A classic dichotomy is *specific-to-general* versus *general-to-specific* methodologies, see Granger (1990). In our context, *specific-to-general* would entail that we start by a model that does not include the period dummies but only the individual explanatory variables and a constant term. If we primarily are interested in the effects of individual characteristics, and if these variables are more or less orthogonal to the (omitted) period dummies, this specific model is sufficient for the purpose. However, if we are interested in testing the potential effects of aggregated variations and the effects of friction outlined above, it is not clear that adding variables successively will give reliable results about the time period dummies. In such a situation it seems more sensible to start from the GUM mentioned above and to follow a general-to-specific (GETS) procedure.

GETS modeling strategies have been advocated and debated over several decades, as e.g. the different positions in Hendry *et al.* (1990) show. One advantage of GETS compared to specific–to-general modeling is that it lets itself to computer automatization. Good algorithms for GETS modelling have been shown to be able to retrieve the true model with great regularity if it is situated within the GUM, see Hoover and Perez (1999) and Hendry and Krolzig (1999, 2005).

In this paper we refer to the latest version of the computer program version of GETS which is dubbed Autometrics, see Doornik (2009) and Doornik and Hendry (2009). The main control parameter in Autometrics, set by the user, is the target *size* that determines the significance level below which a variable cannot be deleted from the model. We denote the target size by the symbol α, because of the obvious connotation to the significance level of a (one-off) statistical test.

Perhaps the most common argument against GETS modeling is that repeated testing results in loss of control of the size of the test, so that the probability of not-rejecting variables with true coefficient equal to zero may be considerably larger than the intended significance level, see Lovell (1983). However, and somewhat surprisingly, research shows that this is not an inherent problem for GETS modeling, but also that the search algorithm matters a lot, see Hendry (2000) and Doornik (2009). In order to fix ideas we might consider an idealized situation with n regressors that are orthogonal. This situation is particularly relevant in our case, because the individual variables in X_1 are almost uncorrelated with dummies in X_2. In such a situation a natural GETS algorithm would be to estimate the general unrestricted model (GUM), order the squared t-values in decreasing order and use a critical cut-off value based on the size parameter α. The cut-off value distinguishes between the m retained and the (n-m) excluded variables. In this case only one decision is needed, and "repeated testing" leading to loss of control of the size of the test, never occurs. "Goodness of fit" or other model selection criteria are never used.

In the above example, it is perhaps not surprising that a sensible GETS algorithm makes the actual retention rate of insignificant variables close to the chosen size or significance level α. In Autometrics the use of path searches gives impression of repeated testing, like in a test-tree. Therefore confusion might arise between selecting among 2^n models, and choosing m retained variables from the n variables of the GUM. Selecting between models becomes an impossible task for even quite low values of n, while the Autometrics algorithm for choosing variables has been shown to deliver a gauge close to the user set significance level α in a range of Monte-Carlo simulations that capture realistic degrees of colinearity, departures from normality assumptions and even non-stationary data, see e.g. Doornik (2009). Of course, sampling variation matters in practice, and we can retain irrelevant variables or miss relevant variables near the selection margin. Sometimes the selection margin is quite broad and Autometrics will then deliver more than one 'terminal model' before choosing one 'final model' based on a wider set of criteria, e.g. the Schwarz information criterion. This is only a default option though. In such cases it is usually worth assessing the full set of terminal models, and look for interpretational differences between the models that appear to be more or less identical from a choice of variables point of view. Instead of choosing among the set of terminal models, one may alternatively combine the information sets from the different terminal models. As we shall see below, this point becomes relevant for our application.

In sum, the average retention rate of irrelevant variables can in practical use be controlled by the Autometrics α-value. In order to control for selection bias (estimated coefficients are biased from zero for retained variables), a conservative strategy is usually advised. In practice this implies choosing α to for example 0.01 or 0.025 rather than the conventional significance level of 0.05.

The "potency" of a GETS algorithm is measured (in Monte Carlo studies) by the average retention rate of relevant variables. Of course, one would want this to be high and close to the theoretical power of a one-off test. Simulation studies show that potency is close to the theoretical power if the reduced model (terminal model) encompasses the GUM. Conversely, reduction without encompassing loses both gauge and potency.

5. Empirical results – Sample I

Table 1 gives an overview of estimation results using Sample I. In the GUM, the probability of being employed given that the female is in the work force is modelled as a function of the following individual variables: Education length, work experience, and its square, a dummy for living in a residential area, the number of children in three age groups and the log of real non-labor income. In addition the GUM contains a constant term and the full set of 82 period dummies for time periods, the reference quarter being 1988q2. This gives a model with 91 parameters, of which 8 represent the effects of changes in the individual explanatory variables. We refer to estimations based on this information set as *MODEL Class NU* (no unemployment rate).

The results when all the variables in the GUM are forced to be included in the estimation are shown in the first column of Table 1. We do not show the estimated dummy coefficients for the GUM in Table 1 because of the large number of period dummies. Instead, the last part of the table includes the Chi-square statistics of the joint hypothesis that all the period dummies in the GUM have zero coefficients. As the test statistic shows, this hypothesis is rejected at a very low significance level. Figure 2 shows a graph of the t-values of the full set of period dummies in the GUM. The graph shows that for the years between 1991 and 1999 there is a majority of negative t-values, and that some of these are significant. This period was marked by the consequences of the banking crisis in Norway, a 40 percent fall in the real housing prices, and a general recession caused both by financial consolidation among Norwegian households, and in contraction of export markets. Thus our estimation results confirm other findings that many women (like men) experienced layoffs, and a drop in the job offer rate, during these years. The years before and immediately after the new millennium were much better in terms of labor market performance, which seems to be reflected in the sequence of t-values in the graph. The period until

2005 was, however, marked by low employment growth, but then employment increased by almost 250 000 persons between 2005 and the onset of the financial crisis in 2008 (this is unusually high for Norway). As commented on above, this exceptional large increase in employment would not have been possible without immigration of female workers.

Overall, the graph seems to document that there are some effects of chance in individual employment probabilities, and that negative shocks (lay-offs) may be more detectible than positive ones. Of course, there is no reason to expect complete symmetry in the effects, and at present, the question is whether we can use automatic model selection to narrow down the set of dummies and interpret the results in the light of what we know about the Norwegian business cycle history.

Table 1. Estimation of the probability of being employed for married and cohabitating women in the work force. Sample I . *Model Class NU*

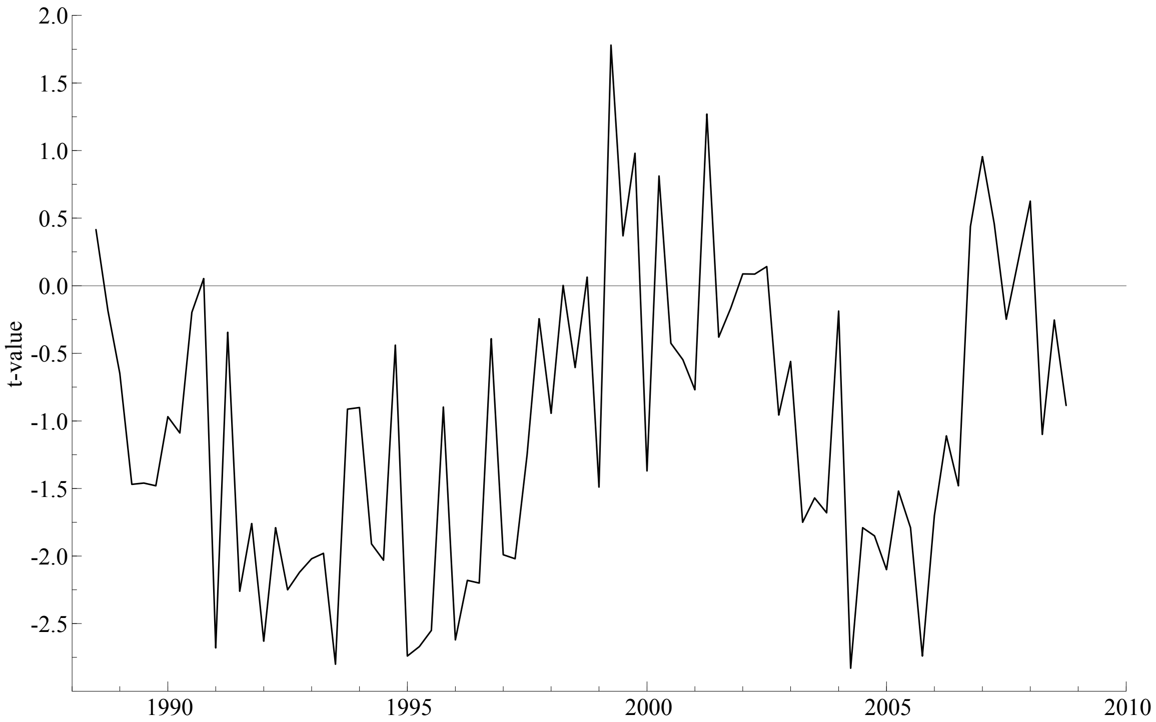
Explanatory Variable	Model specification			
	GUM		One-stage selection	
	Estimate	t-value	Estimate	t-value
Schooling	0.283	17.5	0.288	18.1
Experience	0.072	4.24	0.071	4.20
Experience^2/100[a]	−0.047	−1.32	−0.043	−1.22
Urban dummy	0.065	0.972	0.06	0.90
#children 0–3	−0.196	−3.07	−0.198	−3.11
#children 4–6	−0.141	−2.15	−0.135	−2.07
#children 7–18	−0.078	−2.96	−0.078	−2.07
Non-labor income[b]	0.149	4.25	0.157	4.54
Constant	−2.07	−3.81	−2.63	−5.62
1991q1	This model has a full set of break Dummies		−0.429	−2.32
1993q3			−0.454	−2.56
1995q1			−0.441	−2.46
1995q2			−0.432	−2.26
1999q2			2.26	2.26
2004q2			−0.606	-2.4
2005q2				
Chi-sq for full set of breaks[c]	123.422 [0.0021]			
Chi-sq for retained set of breaks[c]			31.665 [0.000]	
No. of param.	91		15	
No. of breaks	82		6	
Log-likelihood	−5,311.141		−5,364.468	
Target size, α			Small (0.01)	

[a] For better readability of the table, the estimated coefficient of the squared experience variable has been multiplied by 100.
[c] The variable is log transformed.
[c] Significance probability in square brackets.

We next use Autometrics to select models that retain only a sub-set of statistically significant dummies. By using a target size of 1 percent, we expect to keep perhaps 1 dummy too many, under the null hypothesis of no significance. It is reasonable to think that identification of breaks is sample dependent, and therefore we also estimate the model on a second data set below (Section 6).

The columns named *One-stage selection* shows the results when the individual explanatory variables are forced to be included as regressors, and Autometrics chooses only among the set of dummies for observation quarter (from one GUM only). The result is that only 6 calendar dummies are kept as significant and their estimated coefficients and t-values are shown in columns 4 and 5 of the table. Two of the retained dummies are from the early 1990s, two are from 1995, one is from 1999 and the last one is from 2004.

Figure 2. t-values of the coefficients of the dummies for observation period in the GUM in Table 1



The dummies for 1991q1, 1993q3, and also 1995q1 and 1995q2, correspond well with the recession mentioned above. The positive coefficient for the 1999q2 dummy is associated with the period of marked fall in female unemployment between 1996 and 2000, and the selection of the 2004q2-dummy comes from the period of prolonged growth in unemployment at the start of the millennium. The test of the null hypothesis that the parameters of the five dummies are jointly zero, gives rejection at arbitrary low levels of significance, as the *Chi-squared* test shows.

To assess the effects of the identified macroeconomic shocks it is informative to measure how the different shocks influence the job probability. In Table 2 we consider the implications of the significant negative breaks for female population employment.[7] To explain the table let us pick one of the quarters, for instance 1991q1. The number 0.0150 corresponds to the mean increase in predicted job probability when one compares a counterfactual situation without the break with the actual situation in which one has identified and estimated the impact of the break in 1991q1. The mean is taken over the total number of persons in the sample observed in this quarter. Although the increase in the predicted job probability is rather small, this does not mean that the aggregate increase in numbers of unemployed is trivial. Using information on the corresponding population of married and cohabiting females in the work force in the same quarter one may deduce that the absence of the shock yields an increase in the population employment corresponding to approximately 6,600 persons, which is a non-negligible increase in employment.[8] Similar type of calculations is carried out for the other quarters that are reported in Table 2.

Table 2. Marginal effects of different negative shocks on the mean job probability of the women in the sample and aggregate implications for employment. *Model Class NU One-stage* selection reported in Table 1

Period (t)	Increase in mean probability in a counterfactual situation with no break[a]	Increase in number of employed at the population level
1991q1	0.0150	6,600
1993q3	0.0144	6,800
1995q1	0.0131	6,500
1995q2	0.0121	6,100
2004q2	0.0142	7,900

[a] The mean is taken over the individuals in the sample in the specific quarter.

6. Empirical results - Sample II

To illustrate some aspects of sample variability and robustness, we repeat the estimations undertaken in the previous section using a different sample. Recall that Sample II shows no noteworthy differences in the characteristics of the persons from Sample I, but due to the sample selection procedure the number of observations in the various quarters might differ.

[7] Recall that the population we are considering is married and cohabitating women aged 25-60 years.

[8] This type of calculation is based on the simplifying assumption that the sample is drawn randomly from the population. More accurate calculations may be carried out taking account to the sample design used in conjunction with LFS.

Figure 3 shows the t-values for the period dummies in the GUM according to the two data sets. Although there are differences between the t-values for a given quarter, the two graphs show the same general qualitative evolution of friction over time. According to the graphs, there was (negative) friction early in the data period, and between 2003 and 2005. After 2005 there was a relatively marked improvement in labor market conditions and a lessening of frictional effects on job probabilities. The lower panel of the figure reduces the effect of sample variability by showing the centered moving average of the two sequences of t-values.

Figure 3. t-values of the coefficients of the friction dummies. The upper panel shows the t-values in the two GUMs for Sample I (solid line) and Sample II (dashed line). The lower panel shows the centered moving averages of the two sequences of t-values: Sample I (solid line) and Sample II (dashed line)

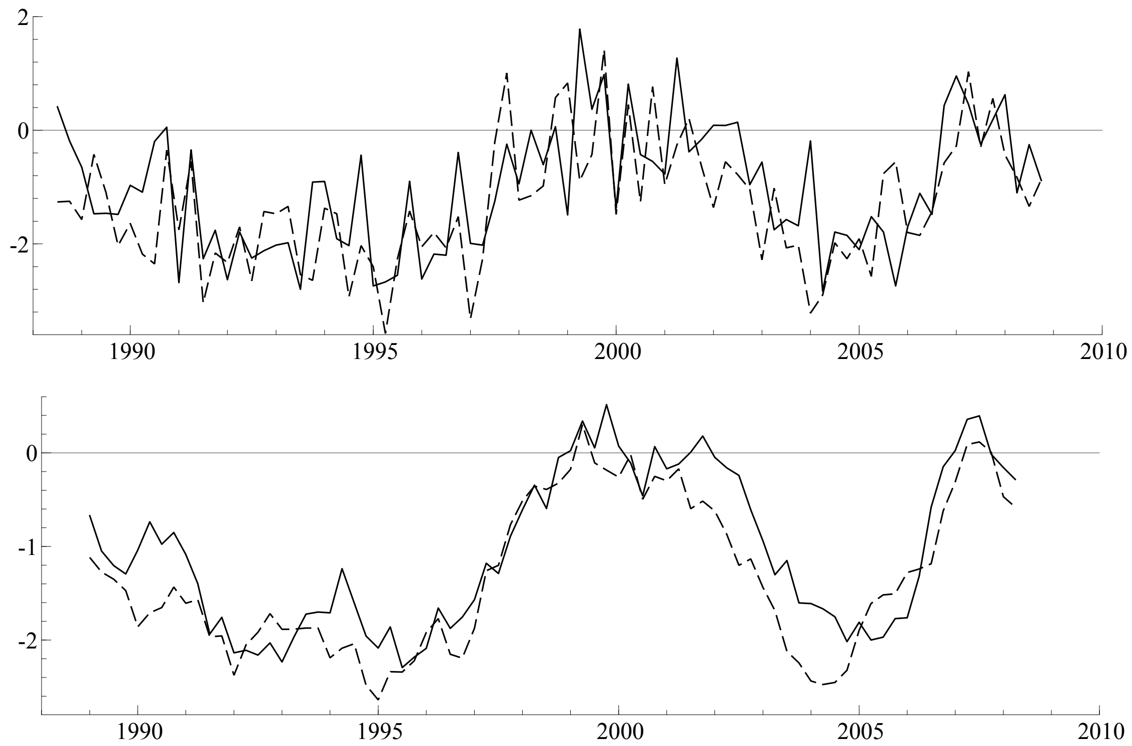


Table 3 makes it possible to compare the models obtained on the two datasets, and also the results from using two different model selection algorithms. The first column with estimates and t-values is the same as *Model Class NU One-stage* selection in Table 1, i.e., for Sample I. The next two columns show results for the first sample, but for a second selection procedure which can be associated with a t*wo-stage selection* procedure based on two different GUMs. The last four columns in Table 3 show the results from the same two selection procedures, but for Sample II.

Table 3. Final models of the probability of being employed for married and cohabitating women in the work force according to two different samples and two selection methods.[a] *Model* Class *NU*

Explanatory vars.	Sample I				Sample II			
	One-stage select.		Two-stage select.[b]		One-stage select.		Two stage select.[c]	
	Estimate	t-value	Estimate	t-value	Estimate	t-value	Estimate	t-value
Individual var.								
Schooling	0.288	18.5	0.289	18.6	0.274	18.0	0.271	17.9
Experience	0.071	4.20	0.072	4.25	0.086	5.14	0.085	5.14
$\text{Experience}^2/100$	–0.043	–1.22	–0.045	–1.26	–0.083	–2.41	–0.085	–2.44
Urban dummy	0.06	0.90	0.060	0.89	0.017	0.25	0.020	0.29
#children 0-3	–0.198	–3.11	–0.195	–3.06	–0.163	–2.53	–0.168	–2.62
#children 4-6	–0.135	–2.07	–0.131	–2.02	–0.219	–3.44	–0.226	–3.54
#children 7-18	–0.078	–2.07	–0.077	–2.06	–0.086	–2.30	–0.088	–2.34
Non-labor inc.[d]	0.157	4.54	0.156	4.51	0.115	3.10	0.111	2.97
Calendar var.								
Constant	–2.63	–5.62	–2.618	–5.59	–2.046	–4.19	–1.971	–4.02
1988q3			0.572	1.78				
1991q1	–0.429	–2.32	–0.458	–2.48				
1991q3							–0.435	–2.35
1992q1			–0.461	–2.34				
1993q3	–0.454	–2.56	–0.483	–2.72				
1994q3							–0.394	–2.17
1995q1	–0.441	–2.46	–0.470	–2.62				
1995q2	–0.432	–2.30	–0.462	–2.46	–0.612	–3.54	–0.598	–3.46
1995q3			–0.421	–2.24				
1996q1			–0.458	–2.33				
1997q1					–0.595	–2.85	–0.581	–2.78
1997q4							1.160	–2.30
1998q4							0.883	1.95
1999q1							1.057	2.10
1999q2	2.26	2.26	2.233	2.23				
1999q4					1.650	2.32	1.664	2.34
2000q2							0.816	1.81
2000q4							1.018	2.02
2004q1					–0.668	–2.58	–0.652	–2.52
2004q2	–0.606	–2.41	–0.636	–2.53				
2005q4			–0.636	–2.38				
No. of obs.	50,487		50,487		50,283		50,283	
No. of paramet.	15		20		13		20	
No. of breaks	6		11		4		11	
Log-likelihood	–5,364.468		–5,353.399		–5,417.333		–5,397.875	

[a]The target size, α, in conjunction with *One-stage selection* and the initial sequence of *Two-stage selection* is 0.01. The target size in conjunction with the second sequence of *Two-stage selection* is 0.025. [b]The GUM used in the final sequence includes the following right hand side variables: All individual variables, constant term, 1988q3, 1988q4, 1990q3, 1990q4, 1991q1, 1991q2, 1992q1, 1993q3, 1994q4, 1995q1, 1995q2, 1995q3, 1996q1, 1998q2, 1998q4, 1999q2, 1999q3, 2002q1, 2002q2, 2002q3, 2004q2, 2005q4 and 2006q4. The GUM used in the initial sequence yields 20 terminal models. [c]The GUM used in the second sequence includes the following right hand side variables: All individual variables, constant term, 1989q2, 1990q4, 1991q3, 1994q3, 1995q2, 1997q1, 1997q4, 1998q4, 1999q1, 1999q4, 2000q2, 2000q4, 2001q3 and 2004q1. The GUM used in the initial sequence yields 6 terminal models. [d]The variable is log transformed.

The difference between the two selection methods is the following. In *One-stage* selection we obtain a single final specification by employing Schwarz information criterion after having ended up with potentially several terminal models. In Sample I there are in fact 6 terminal models, meaning that the selection margin is broad. In this case it is worth considering other decision rules than the Schwarz criterion. Hence as an alternative, we also use a two-stage selection procedure where we form a union model of the six terminal models and then let Autometrics select friction dummies from this second-stage GUM, using a slightly more liberal significance level with target size, α, equal to 0.025. As we see from the table this yields more period dummies in the final model for both data sets. For Sample I the number of period dummies increases from 6 to 11, whereas it increases from 4 to 11 when one applies Sample II.

The results related to the effects of the individual variables are rather similar for the two data sets. A difference is that the squared experience term enters significantly for Sample II and in addition its estimated coefficient is larger in magnitude. With respect to the one-stage selection procedure the retained period dummies using the different data sets come from the same periods, but they are fewer using Sample II. Specifically, the final models using Sample II do not incorporate any quarters from 1991 and 1993 as periods of significant friction. These are the main differences between the selected variables in the two samples in conjunction with one-stage selection.

If one turns to the two-stage selection procedure there are 11 period dummies using either of the data sets. There is one common period dummy, 2005Q2. Otherwise the results differ somewhat. For instance, the results for Sample I show there are period dummies for 1991q1 and 2004q2, whereas according to Sample II there are period dummies for 1991q3 and 2004q1. However, there are also examples that significant period dummies are found in a time interval using one of the data sets, but not the other one. For instance according to the results for Sample II there are four significant period dummies in the years 1997–1998, but none according to Sample I. This difference may to some extent be attributed to sample variability.

7. Models with aggregate unemployment rate(s)

In this section we consider model selection after having added aggregated unemployment rate(s) to the information set. Our primary goal is to demonstrate that almost no period dummies are necessary when we include the unemployment rate(s) in the information set. We consider three different ways of incorporating information on unemployment: *Model Class FU* (Table 5), *Model Class MU* (Table 6) and *Model Class BU* (Table 7), respectively, correspond to (i) a specification with the female aggregate unemployment rate, (ii) a specification with the male aggregate unemployment rate and (iii) a specification with both the female and male unemployment rates. These cases are interacted with the two selection algorithms, cf. the one-stage and the two-stage selection procedures used in the previous sections. Recall that One-stage is based on deriving a single specification using a GUM with all the period dummies and the unemployment rate(s) included, while Two-stage is based on two sequences, where a "new" GUM is generated from the union of retained variables corresponding to the terminal models from the initial sequence. In both cases all the individual explanatory variables are forced to be included in the final model. A strong conclusion is that one unemployment rate is retained in all the final models.[9]

Table 4 below contains estimation results when all the period dummies have been omitted a priori. These specifications involve no model selection. Within both samples the parameter estimates are almost identical across model specifications. The parameter estimate associated with the unemployment rate is larger when we use the unemployment rate of females than the unemployment rate of males, but from Figure 1 we notice that for most periods the unemployment rate among men is higher than the one for females.

[9] One may argue that using the female unemployment rate as a regressor in the logit models will produce an endogeneity problem. However, one should recall that the female unemployment rate in addition to married and cohabitating women also covers single women. A reason for also including the male unemployment rate in the analysis is that the same type of objection can not be raised against this variable.

Table 4. Logit estimates of the probability of being employed for different measures of the unemployment rate.[a] No automatic model selection

Explanatory variable	Sample I			Sample II		
	Model Class			Model Class		
	FU	*MU*	*BU*	*FU*	*MU*	*BU*
Schooling	0.271	0.274	0.271	0.257	0.262	0.257
	(17.4)	(17.6)	(17.4)	(16.8)	(17.2)	(16.8)
Experience	0.070	0.070	0.070	0.084	0.085	0.084
	(4.13)	(4.13)	(4.13)	(5.05)	(5.11)	(5.04)
Experience^2/100	–0.046	–0.045	–0.047	–0.086	–0.086	–0.086
	(–1.32)	(–1.29)	(–1.32)	(–2.49)	(–2.50)	(–2.48)
Urban dummy	0.070	0.067	0.070	0.026	0.027	0.025
	(1.05)	(1.01)	(1.05)	(0.38)	(0.40)	(0.38)
#children 0–3	–0.210	–0.207	–0.210	–0.178	–0.171	–0.178
	(–3.30)	(–3.24)	(–3.29)	(–2.76)	(–2.66)	(–2.76)
#children 4–6	–0.152	–0.151	–0.153	–0.238	–0.235	–0.238
	(–2.33)	(–2.32)	(–2.35)	(–3.73)	(–3.69)	(–3.72)
#children 7–18	–0.085	–0.083	–0.085	–0.091	–0.089	–0.091
	(–2.27)	(–2.20)	(–2.26)	(–2.42)	(–2.38)	(–2.42)
Non-labor income[b]	0.141	0.144	0.140	0.092	0.097	0.093
	(3.99)	(4.08)	(3.96)	(2.41)	(2.56)	(2.42)
Female unemp. rate	–0.318		–0.241	–0.297		–0.333
	(–7.35)		(–3.28)	(–6.99)		(–4.54)
Male unemp. Rate		–0.172	–0.056		–0.138	0.026
		(–6.89)	(–1.29)		(–5.54)	(0.59)
Constant	–1.195	–1.637	–1.209	–0.568	–1.153	–0.561
	(–2.31)	(–3.28)	(–2.34)	(–1.05)	(–2.21)	(–1.04)
No. of parameters	10	10	11	10	10	11
Log-likelihood	–5,354.409	–5,359.011	–5,353.579	–5,408.105	–5,418.345	–5,407.931

[a] t-values in parentheses.
[b]The variable is log transformed.

In Table 5 we consider models with the female unemployment rate included in the information set (*Model Class FU*).

Using the single-sequence version of the model selection algorithm, we end up with one period dummy applying Sample I and no period dummies applying Sample II. For the corresponding estimations without the unemployment rate in the information set (Table 3) we ended up with 6 dummies applying sample I and 4 dummies applying Sample II. Applying Sample I the only retained period dummy in Table 5 (1999q2) is also among the 6 retained period dummies in the corresponding specification in Table 3.

Table 5. Logit estimates of the probability of being employed using the female unemployment rate (*Model Class FU*) as a regressor[a]

Explanatory variable	Sample I		Sample II	
	One-stage select.[b]	Two-stage select.[c]	One-stage select.[b]	Two-stage select.[d]
Schooling	0.271	0.273	0.257	0.259
	(17.4)	(17.5)	(16.8)	(16.9)
Experience	0.070	0.069	0.084	0.085
	(4.14)	(4.11)	(5.05)	(5.08)
Experience^2/100	–0.047	–0.045	–0.086	–0.087
	(–1.32)	(–1.29)	(–2.49)	(–2.50)
Urban dummy	0.071	0.070	0.026	0.020
	(1.06)	(1.04)	(0.38)	(0.30)
#children 0–3	–0.211	–0.213	–0.178	–0.175
	(–3.32)	(–3.34)	(–2.76)	(–2.71)
#children 4–6	–0.150	–0.149	–0.238	–0.236
	(–2.30)	(–2.28)	(–3.73)	(–3.69)
#children 7–18	–0.085	–0.084	–0.091	–0.092
	(–2.26)	(–2.23)	(–2.42)	(–2.44)
Non-labor income[e]	0.142	0.143	0.092	0.096
	(4.00)	(4.05)	(2.41)	(2.51)
Female unempl. Rate	–0.304	–0.311	–0.297	–0.265
	(–7.00)	(–7.11)	(–6.99)	(–6.08)
Constant	–1.255	–1.258	–0.568	–1.258
	(–2.43)	(–2.44)	(–1.05)	(–2.44)
1995q2				–0.489
				(–2.82)
1997q1				–0.547
				(–2.62)
1997q4				0.950
				(1.88)
1999q2	1.938	1.915		
	(1.93)	(1.91)		
1999q4				1.277
				(1.79)
2004q1				–0.668
				(–2.58)
2004q2		–0.649		
		(–2.58)		
2005q4		–0.623		
		(–2.33)		
No. of obs.	50,487	50,487	50,283	50,283
No. of parameters	11	13	10	15
No. of breaks	1	3	0	5
Log-likelihood	–5,350.526	–5,345.530	–5,408.105	–5,394.110

[a] t-values in parentheses.
[b] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in GUM: 1988q4, 1989q1,…,2008q4, female unemployment rate.
[c] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in GUM: 1991q1, 1992q1, 1993q3, 1995q1, 1995q2, 1995q3, 1996q1, 1999q2, 1999q3, 2004q2, 2005q4 and female unemployment rate.
[d] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in final sequence GUM: 1991q3, 1995q2, 1997q1, 1997q4, 1999q4, 2004q1 and female unemployment rate.
[e] The variable is log transformed.

Using the double-sequence version of the algorithm — still with the female unemployment rate included in the information set — the final model specifications contain some more period dummies. From the results in Table 5 we see that for Sample I there are 3 period dummies, while for Sample II 5 period dummies are included. The female unemployment rate enters with the correct sign in all specifications and is strongly significant.[10]

It might be argued that there is some overlap in the samples used in the estimation of the aggregate female unemployment rate and the sample used in the estimation of our model since both samples are based on data from the LFS. To consider this potential problem we apply the unemployment rate for males (*Model Class MU*) instead of females in the estimations since these two time series are highly correlated. The effect of the male unemployment rate is also negative and significant in Table 6, but the parameter estimates are not as large as the corresponding estimates of female unemployment reported in Table 5. Using the male unemployment rate and the one-stage selection procedure we end up with 1 period dummy for both samples. Recall that the corresponding estimation results in Table 3, in which the unemployment rate was omitted, included several period dummies. For the double-sequence selection algorithm, and with the male unemployment rate in the information set, the number of period dummies is 5 with Sample I and 2 with Sample II (Table 6). Here the GUM used in the final sequence is almost equal to the GUM used in the initial sequence. Only two period dummies are dropped. The estimated slope parameters attached to the person-specific covariates are rather equal regardless of whether we use the female (Table 5) or the male (Table 6) unemployment rate.

[10] Note that in conjunction with tables 5 and 6, we have omitted the period dummies for 1988q2 and 1988q3 in order to achieve identification which is necessary for the selection algorithm to be operative, while we in Table 7, of the same reason, have omitted the period dummies for 1988q2, 1988q3 and 1988q4.

Table 6. Logit estimates of the probability of being employed using the male unemployment rate (*Model Class MU*) as a regressor[a]

Explanatory variable	Sample I		Sample II	
	One-stage select.[b]	Two-stage select.[c]	One-stage select.[b]	Two-stage select.[d]
Schooling	0.274	0.276	0.262	0.262
	(17.6)	(17.7)	(17.2)	(17.2)
Experience	0.070	0.070	0.086	0.085
	(4.14)	(4.14)	(5.13)	(5.12)
Experience^2/100	–0.045	–0.045	–0.087	–0.087
	(–1.30)	(–1.28)	(–2.52)	(–2.51)
Urban dummy	0.068	0.068	0.024	0.022
	(1.01)	(1.01)	(0.36)	(0.33)
#children 0–3	–0.208	–0.209	–0.168	–0.167
	(–3.26)	(–3.28)	(–2.61)	(–2.60)
#children 4–6	–0.150	–0.148	–0.235	–0.232
	(–2.29)	(–2.27)	(–3.69)	(–3.64)
#children 7–18	–0.082	–0.082	–0.090	–0.091
	(–2.19)	(–2.17)	(–2.40)	(–2.42)
Non-labor income[e]	0.144	0.145	0.097	0.097
	(4.08)	(4.11)	(2.56)	(2.57)
Male unempl. rate	–0.164	–0.162	–0.141	–0.138
	(–6.55)	(–6.38)	(–5.64)	(–5.64)
Constant	–1.680	–1.709	–1.136	–1.133
	(–3.37)	(–3.43)	(–2.18)	(–2.17)
1995q2				–0.559
				(–3.23)
1995q3		–0.350		
		(–1.87)		
1997q1			–0.630	–0.643
			(–3.02)	(–3.08)
1999q2	2.003	1.992		
	(2.00)	(1.98)		
2001q2		1.062		
		(1.82)		
2004q2		–0.566		
		(–2.25)		
2005q4		–0.630		
		(–2.36)		
No. of obs.	50,487	50,487	50,283	50,283
No. of parameters	11	15	11	12
No. of breaks	1	5	1	2
Log-likelihood	–5,354.738	–5,346.271	–5,414.513	–5,410.013

[a] t-values in parentheses.
[b] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in GUM: 1988q4, 1989q1,…, 2008q4, male unemployment rate.
[c] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in GUM: 1991q1, 1991q3, 1992q1, 1992q3, 1993q3, 1995q1, 1995q2, 1995q3, 1996q1, 1999q2, 1999q3, 1999q4, 2000q2, 2001q2, 2004q2, 2005q4, 2006q4 and male unemployment rate.
[d] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in second second sequence GUM: all period dummies except 1998q4 and 2000q2, male unemployment rate.
[e] The variable is log transformed.

When we, cf. Table 7, include both unemployment rates, only the female unemployment rate is retained. This comes as no surprise since the two unemployment rates series are highly correlated, as is evident from Figure 1.[11] Using the single-sequence selection algorithm, none of the period dummies are retained in the final specification. This is the case for both samples. Using the double-sequence selection algorithm, we retain 2 period dummies in conjunction with Sample I and 4 with Sample II.

Table 7. Logit estimates of the probability of being employed using the female and male unemployment rates (*Model Class BU*) as regressors[a]

Explanatory variable	Sample I		Sample II	
	One-stage select.[b]	Two-stage select.[c]	One-stage select.[b]	Two-stage select.[d]
Schooling	0.271	0.272	0.257	0.260
	(17.4)	(17.5)	(16.8)	(16.9)
Expererience	0.070	0.069	0.084	0.084
	(4.13)	(4.11)	(5.05)	(5.05)
Exper^2/100	−0.046	−0.046	−0.086	−0.086
	(−1.32)	(−1.30)	(−2.49)	(−2.47)
Urban dummy	0.070	0.069	0.026	0.020
	(1.05)	(1.03)	(0.38)	(0.30)
#children 0−3	−0.210	−0.213	−0.178	−0.177
	(−3.30)	(−3.34)	(−2.76)	(−2.76)
#children 4−6	−0.152	−0.150	−0.238	−0.235
	(−2.33)	(−2.31)	(−3.73)	(−3.68)
#children 7−18	−0.085	−0.085	−0.091	−0.091
	(−2.27)	(−2.25)	(−2.42)	(−2.43)
Non-labor income[e]	0.141	0.142	0.092	0.095
	(3.99)	(4.02)	(2.41)	(2.51)
Female unempl. Rate	−0.318	−0.308	−0.297	−0.292
	(−7.35)	(−7.07)	(−6.99)	(−6.73)
Constant	−1.195	−1.245	−0.568	−0.626
	(−2.31)	(−2.41)	(−1.05)	(−1.16)
1995q2				−0.496
				(−2.86)
1997q1				−0.561
				(−2.69)
1999q2		1.925		
		(1.92)		
2004q1				−0.689
				(−2.66)
2004q2		−0.641		−0.638
		(−2.55)		(−2.39)
No. of obs.	50,487	50,487	50,283	50,283
No. of parameters	10	12	10	14
No. of breaks	0	2	0	4
Log-likelihood	−5,354.409	−5,347.814	−5,408.105	−5,396.578

[a] t-values in parentheses.
[b] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in GUM: 1988q4, 1989q1,…,2008q4, female unemployment rate, male unemployment rate.
[c] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in second sequence GUM: 1991q1, 1992q1, 1993q3, 1995q1, 1995q2, 1995q3, 19961q1, 1999q2, 2004q2, female unemployment rate, male unemployment rate.
[d] All individual explanatory variables and constant term forced to be included in the final model. Other exogenous variables in GUM: 1991q3, 1994q3, 1995q2, 1997q1, 1997q2, 1999q4, 2004q1, 2004q2, 2004q4, 2005q2, female unemployment rate, male unemployment rate.
[e] The variable is log transformed.

[11] The empirical correlation coefficient is 0.86.

One of the period dummies is common for both samples, namely the dummy for 2004q2.

The magnitudes of the estimated parameters attached to the individual-specific variables are rather equal to those reported in Table 5 and Table 6. The parameter estimate of the female unemployment rate is somewhat larger for Sample I than for Sample II. The estimated effect of Schooling is somewhat larger for Sample I as compared to Sample II.

8. The relative importance of chance and choice variables at the micro and macro level

One might hypothesize that the choice variables have only limited explanatory power when they are included in a model with controls for business cycle effects. To shed light on this question we consider the hypothesis that the effects of the choice variables are zero and confront it with the alternative that the effects of the choice variables are different from zero. We carry out such a test for all the four model classes, and consider both the models obtained using the *One-stage selection* and models based on the *Two-stage selection* methods. In Table B1 we report the log-likelihood value and the value of the AIC under the null and under the alternative. As seen from the table omitting the choice variables yields a significant drop in the log-likelihood value and a substantial increase in AIC. Using Chi-squared distributed LR-tests with 8 degrees of freedom the null hypothesis is clearly rejected in all cases, i.e., when one combines the four model classes with the two selection algorithms. Thus, undoubtedly the choice variables play a crucial role as explanatory variables for the probabilities of being employed or unemployed at the micro level.

It is also relevant to consider the importance of choice and chance variables from a macro perspective. To do so, we assess how the different models are capable of explaining the (within sample) unemployment rate. In Appendix B we explain how this comparison is carried out. Given a specific combination of model class and selection method we estimate three models. The first model (Model 1), which is a reference model, contains only a constant term. The second model (Model 2) takes as the point of departure the model specification obtained by model selection, but all the eight choice variables are omitted and the model is reestimated. The third model (Model 3) is the model obtained by model selection, cf. Table 3 and tables 5-7. In each of the 83 time periods we predict the number of unemployed workers relative to the working force in the sample and measure the deviation relative to the observed (within sample) unemployment rate. We obtain an overall measure of deviation by calculating the square root of the mean of the squared deviations over all periods, cf. Eq. (B2). The results are reported in Table B2. In all the cases there is a drop in the deviation measure when one compares Model 2 to Model 1. With respect to the NU model class there is a further drop in the deviation measure when one goes from Model 2 to Model 3. Within the other three classes the difference between Model 2 and Model 3 is rather modest. For model classes FU and BU, in which the female macro unemployment rate is used as an regressor in both Model 2 and Model 3, one does not get a better explanation of the (within sample) unemployment rate by adding the choice variables. For the model class MU, in which the macro male unemployment rate is used as an regressor in both Model 2 and Model 3, there is a slight drop in the deviation measure when one goes from model 2 to 3. Thus the choice variables seem to be very important at the micro level, but less so if the focus is on the macro level.

9. Conclusions

In this paper we have estimated models, using Norwegian time series of cross-sections data over a twenty year period, for individual job probabilities that include both choice and chance factors. The choice factors were represented by eight individual variables motivated by microeconomic theory and previous empirical evidence, while the chance factors were represented by 82 calendar period dummies, as well as by macro unemployment rates. We used automatic model selection to estimate parsimonious models which retained all the individual variables but only significant dummies, which we suggest can be interpreted as periods when chance, or friction elements, impinged significantly on individual employment probabilities. We applied this modelling to two different samples, and as may be expected, the results are subject to sampling variability. The quarters that are found to represent friction are not exactly the same in the two samples. However in terms of sequences of “good and bad times” the results are the same. In both samples the effects on individual employment probabilities are small, which confirms the insight that most individuals are able to hold on to a job through a macroeconomic downturns. Nevertheless, the aggregate number of people who become unemployed in those periods may be non-trivial, as our calculations suggest.

We also considered models that use the aggregate female and male unemployment rates as ‘sufficient’ variables for the chance element in individual employment outcomes. The results show that this is more or less the case, and the evidence from the two models is mutually supporting the interpretation that chance effects can play a role in empirical models of individual employment probabilities.

References

Dagsvik, J.K., Kornstad, T. and T. Skjerpen (2010): Labor Force Participation and the Discouraged Worker Effect. Discussion Paper No. 642. Statistics Norway.

Doornik, J.A. (2009): Autometrics. Chapter 4 in Castle, J. and N. Shephard (Eds.): *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. New York: Oxford University Press.

Doornik, J.A. and D.F. Hendry (2009): *Empirical Econometric Modelling. PcGive 13. Volume 1*, Timberlake Consultants.

Granger, C.W.J. (Ed.) (1990): *Modelling Economic Series: Readings in Econometric Methodology*. Oxford: Oxford University Press.

Hendry, D.F. (2000): Epilogue: the Success of General-to-Specific Model Selection, Ch 20 in D.F. Hendry: *Econometrics. Alchemy or Science? (New Edition)*. Oxford: Oxford University Press.

Hendry D.F and H.-M. Krolzig (1999): Improving on Data Mining Reconsidered by K.D. Hoover and S.J. Perez, *Econometrics Journal*, 2, 202–219.

Hendry, D. F. and H.-M. Krolzig (2005): The Properties of Automatic GETS Modelling. *Economic Journal*, 115, C32–C61.

Hendry, D.F., E. Leamer and D.J. Poirier (1990): A Conversation on Econometric Methodology, *Econometric Theory*, 6, 171−261.

Hoover, K.D. and S.J. Perez (1999): Data Mining Reconsidered: Encompassing and the General-To-Specific Approach to Specification Search. *Econometrics Journal*, 2, 167−191.

Krusell, P., Mukoyama, T., Rogerson, R. and A. Şahin (2010): Aggregate Labor Market Outcomes: The Roles of choice and Chance. *Quantitative Economics*, 1, 97−127.

Kydland, F. and E. Prescott (1982), Time to Build and Aggregate Fluctuations, *Econometrica*, 50, 1345−1370.

Lovell, M.C. (1983): Data mining. *Review of Economic Statistics*, 65, 1−12.

Mortensen, D. and C. Pissarides (1994): Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies*, 61, 397–415.

Appendix A. Summary statistics for Sample I

Table A1. The number of women in the workforce each quarter and the number belonging to the two labor market states

Quarter	No. of women	No. of employed women	No. of unemploy. women	Unemployment rate
1988q2	793	778	15	1.9
1988q3	623	613	10	1.6
1988q4	593	581	12	2.0
1989q1	662	646	16	2.4
1989q2	644	623	21	3.3
1989q3	694	672	22	3.2
1989q4	705	683	22	3.1
1990q1	698	681	17	2.4
1990q2	740	721	19	2.6
1990q3	752	738	14	1.9
1990q4	748	734	14	1.9
1991q1	792	760	32	4.0
1991q2	721	707	14	1.9
1991q3	785	756	29	3.7
1991q4	767	743	24	3.1
1992q1	725	697	28	3.9
1992q2	768	744	24	3.1
1992q3	756	730	26	3.4
1992q4	736	711	25	3.4
1993q1	851	823	28	3.3
1993q2	818	791	27	3.3
1993q3	852	817	35	4.1
1993q4	846	825	21	2.5
1994q1	858	837	21	2.4
1994q2	826	800	26	3.1
1994q3	824	796	28	3.4
1994q4	894	876	18	2.0
1995q1	891	857	34	3.8
1995q2	854	823	31	3.6
1995q3	868	837	31	3.6
1995q4	874	855	19	2.2
1996q1	788	760	28	3.6
1996q2	761	737	24	3.2
1996q3	715	691	24	3.4
1996q4	679	666	13	1.9
1997q1	620	601	19	3.1
1997q2	534	517	17	3.2
1997q3	497	484	13	2.6
1997q4	514	505	9	1.8
1998q1	476	465	11	2.3
1998q2	463	456	7	1.5
1998q3	452	443	9	2.0
1998q4	520	512	8	1.5
1999q1	469	457	12	2.6
1999q2	444	443	1	0.2
1999q3	483	477	6	1.2
1999q4	470	466	4	0.9
2000q1	468	457	11	2.4
2000q2	466	462	4	0.9
2000q3	465	457	8	1.7
2000q4	503	494	9	1.8
2001q1	490	481	9	1.8
2001q2	469	466	3	0.6
2001q3	510	502	8	1.6
2001q4	491	484	7	1.4

Table A1. Continued

Quarter	No. of women	No. of employed women	No. of unemploy. women	Unemployment rate
2002q1	500	494	6	1.2
2002q2	470	464	6	1.3
2002q3	505	499	6	1.2
2002q4	496	486	10	2.0
2003q1	435	428	7	1.6
2003q2	452	441	11	2.4
2003q3	437	427	10	2.3
2003q4	479	467	12	2.5
2004q1	457	451	6	1.3
2004q2	517	500	17	3.3
2004q3	456	445	11	2.4
2004q4	458	447	11	2.4
2005q1	459	447	12	2.6
2005q2	459	449	10	2.2
2005q3	478	467	11	2.3
2005q4	463	448	15	3.2
2006q1	452	442	10	2.2
2006q2	436	428	8	1.8
2006q3	441	432	9	2.0
2006q4	442	438	4	0.9
2007q1	305	304	1	0.3
2007q2	309	307	2	0.6
2007q3	377	373	4	1.1
2007q4	374	371	3	0.8
2008q1	545	542	3	0.6
2008q2	598	590	8	1.3
2008q3	839	831	8	1.0
2008q4	1,143	1,129	14	1.2
Total	50,487	49,285	1,202	2.4

Table A2. Summary statistics of explanatory variables

Year	Statistic	Real non-labor income[a]	Length of schooling	No. of children aged 0–3 years	No. of children aged 4–6 years	No. of children aged 7–18 years	Dummy for densely populated area
1988	Mean	183,461.21	11.12	0.19	0.18	0.77	0.77
	Std. dev.	77,255.38	2.52	0.45	0.41	0.90	0.42
	Min	182.38	6	0	0	0	0
	Max	494,463.72	20	3	2	4	1
	# obs.	2,009	2,009	2,009	2,009	2,009	2,009
1989	Mean	180,285.02	11.21	0.22	0.18	0.76	0.77
	Std. dev.	75,639.33	2.51	0.48	0.41	0.91	0.42
	Min	432.87	6	0	0	0	0
	Max	516,409.00	20	3	2	5	1
	# obs.	2,705	2,705	2,705	2,705	2,705	2,705
1990	Mean	184,393.46	11.45	0.20	0.18	0.80	0.77
	Std. dev.	76,902.08	2.70	0.46	0.42	0.91	0.42
	Min	250.22	6	0	0	0	0
	Max	556,812.00	20	3	3	5	1
	# obs.	2,938	2,938	2,938	2,938	2,938	2,938
1991	Mean	185,801.53	11.49	0.23	0.18	0.77	0.76
	Std. dev.	84,711.61	2.61	0.49	0.41	0.92	0.43
	Min	322.37	6	0	0	0	0
	Max	586,645.44	20	3	3	5	1
	# obs.	3,065	3,065	3,065	3,065	3,065	3,065
1992	Mean	191,292.31	11.53	0.21	0.18	0.81	0.76
	Std. dev.	88,864.42	2.61	0.47	0.42	0.93	0.43
	Min	156.89	6	0	0	0	0
	Max	585,173.88	20	3	2	5	1
	# obs.	2,985	2,985	2,985	2,985	2,985	2,985
1993	Mean	187,917.42	11.68	0.28	0.18	0.78	0.76
	Std. dev.	88,179.46	2.58	0.53	0.41	0.93	0.43
	Min	77.10	6	0	0	0	0
	Max	613,992.94	20	3	3	4	1
	# obs.	3,367	3,367	3,367	3,367	3,367	3,367
1994	Mean	189,942.49	11.68	0.27	0.21	0.74	0.78
	Std. dev.	86,939.76	2.58	0.53	0.45	0.91	0.42
	Min	76.31	6	0	0	0	0
	Max	621,120.69	20	3	3	5	1
	# obs.	3,402	3,402	3,402	3,402	3,402	3,402
1995	Mean	197,970.04	12.02	0.27	0.21	0.75	0.77
	Std. dev.	91,960.38	2.71	0.53	0.45	0.91	0.42
	Min	148.36	6	0	0	0	0
	Max	669,919.88	20	3	2	5	1
	# obs.	3,487	3,487	3,487	3,487	3,487	3,487
1996	Mean	201,701.99	12.09	0.27	0.21	0.76	0.76
	Std. dev.	94,491.86	2.74	0.52	0.44	0.92	0.43
	Min	147.21	6	0	0	0	0
	Max	700,311.56	20	3	2	4	1
	# obs.	2,943	2,943	2,943	2,943	2,943	2,943
1997	Mean	205,355.30	12.11	0.28	0.23	0.78	0.74
	Std. dev.	100,804.37	2.64	0.54	0.47	0.96	0.44
	Min	355.72	6	0	0	0	0
	Max	724,097.88	20	3	3	5	1
	# obs.	2,165	2,165	2,165	2,165	2,165	2,165

Table A2. (Continued)

Year	Statistic	Real non-labor income[a]	Length of schooling	No. of children aged 0–3 years	No. of children aged 4–6 years	No. of children aged 7–18 years	Dummy for densely populated area
1998	Mean	212,958.14	12.10	0.28	0.22	0.78	0.76
	Std. dev.	98,141.60	2.65	0.53	0.45	0.96	0.43
	Min	1,118.65	6	0	0	0	0
	Max	768,439.69	20.00	3	2	6	1
	# obs.	1,911	1,911	1,911	1,911	1,911	1,911
1999	Mean	217,723.04	12.38	0.28	0.21	0.75	0.75
	Std. dev.	105,422.57	2.73	0.53	0.44	0.95	0.43
	Min	615.90	6	0	0	0	0
	Max	817,057.06	20	3	3	8	1
	# obs.	1,866	1,866	1,866	1,866	1,866	1,866
2000	Mean	221,043.72	12.31	0.25	0.21	0.78	0.76
	Std. dev.	107,755.84	2.71	0.50	0.45	0.97	0.43
	Min	797.76	6	0	0	0	0
	Max	853,378.13	20	3	2	5	1
	# obs.	1,902	1,902	1,902	1,902	1,902	1,902
2001	Mean	229,121.26	12.55	0.28	0.23	0.77	0.75
	Std. dev.	111,758.60	2.72	0.53	0.47	0.96	0.43
	Min	776.41	6	0	0	0	0
	Max	883,351.38	20	3	3	7	1
	# obs.	1,960	1,960	1,960	1,960	1,960	1,960
2002	Mean	234,870.94	12.61	0.26	0.21	0.78	0.77
	Std. dev.	115,676.31	2.72	0.51	0.44	0.97	0.42
	Min	640.23	6	0	0	0	0
	Max	857,033.50	20	3	3	4	1
	# obs.	1,971	1,971	1,971	1,971	1,971	1,971
2003	Mean	241,066.05	12.73	0.26	0.20	0.80	0.78
	Std. dev.	120,388.44	2.75	0.53	0.43	0.99	0.42
	Min	612.42	6	0	0	0	0
	Max	920,717.44	20	3	2	6	1
	# obs.	1,803	1,803	1,803	1,803	1,803	1,803
2004	Mean	249,191.90	12.84	0.19	0.19	0.81	0.76
	Std. dev.	130,986.18	2.75	0.44	0.43	0.96	0.43
	Min	130.94	6	0	0	0	0
	Max	1,008,552.13	20	2	2	4	1
	# obs.	1,888	1,888	1,888	1,888	1,888	1,888
2005	Mean	255,669.29	12.91	0.26	0.20	0.79	0.78
	Std. dev.	135,794.08	2.76	0.52	0.44	0.96	0.42
	Min	368.55	6	0	0	0	0
	Max	1,033,620.25	20	3	2	4	1
	# obs.	1,859	1,859	1,859	1,859	1,859	1,859
2006	Mean	268,907.12	12.97	0.24	0.20	0.85	0.77
	Std. dev.	132,819.68	2.71	0.50	0.45	1.00	0.42
	Min	890.59	6	0	0	0	0
	Max	1,066,316.50	20	3	2	4	1
	# obs.	1,771	1,771	1,771	1,771	1,771	1,771

Table A2. (Continued)

Year	Statistic	Real non-labor income[a]	Length of schooling	No. of children aged 0–3 years	No. of children aged 4–6 years	No. of children aged 7–18 years	Dummy for densely populated area
2007	Mean	272,823.54	13.68	0.27	0.19	0.81	0.78
	Std. dev.	131,669.67	2.59	0.54	0.43	0.99	0.41
	Min	465.88	6	0	0	0	0
	Max	902,558.50	20	3	2	5	1
	# obs.	1,365	1,365	1,365	1,365	1,365	1,365
2008	Mean	291,271.22	13.86	0.25	0.21	0.87	0.78
	Std. dev.	144,369.18	2.56	0.51	0.45	1.01	0.41
	Min	588.94	6	0	0	0	0
	Max	1,094,694.63	20	3	3	5	1
	# obs.	3,125	3,125	3,125	3,125	3,125	3,125

[a] NOK (in constant 1998-prices).

Appendix B: Explanatory power of choice and aggregate friction variables at the micro and macro level

Table B1. The importance of choice variables at the micro level

	Model Class			
Models	NU	FU	MU	BU
One-stage selection				
Sample I				
Optimal model[a]				
Log-likelihood	-5,364.46787	-5,350.52639	-5,354.73833	-5,354.40890
AIC	10,758.936	10,723.053	10,731.477	10,728.818
Optimal model without choice variables				
Log-likelihood	-5,656.94585	-5,610.78807	-5,620.14993	-5,614.20649
AIC	11,327.892	11,227.576	11,246.300	11,232.413
Sample II				
Optimal model[a]				
Log-likelihood	-5,417.33133	-5,408.10473	-5,414.51270	-5,408.10473
AIC	10,860.663	10,836.210	10,851.025	10,836.210
Optimal model without choice variables				
Log-likelihood	-5,684.80831	-5,643.13075	-5,658.23496	-5,643.13075
AIC	11,379.617	11,290.262	11,322.470	11,290.262
Two-stage selection				
Sample I				
Optimal model[a]				
Log-likelihood	-5,353.39941	-5,345.52977	-5,346.27091	-5,347.81444
AIC	10,746.799	10,717.060	10,722.542	10,719.629
Optimal model without choice variables				
Log-likelihood	-5,646.20066	-5,608.42200	-5,614.09293	-5,609.42178
AIC	11,316.401	11,226.844	11,242.186	11,226.844
Sample II				
Optimal model[a]				
Log-likelihood	-5,397.87492	-5,394.10982	-5,410.013	-5,396.57824
AIC	10,835.750	10,818.220	10,844.026	10,821.157
Optimal model without choice variables				
Log-likelihood	-5,661.86769	-5,632.33597	-5,653.36821	-5,634.71096
AIC	11,347.735	11,278.672	11,314.736	11,281.422

[a] Cf. tables 3 and 5-7.

In the last column of Table A1 we report the empirical unemployment rates within Sample I.[12] Let the time series we obtain when dividing these rates by 100 be denoted u_t. How do the different models perform with respect to explaining the variation in this unemployment rate? If we look at a specific model class and at one of the selection methods, we may consider three different models, Model 1-Model 3. Model 1 contains only an intercept, Model 2 contains all variables according to the optimally selected model except the eight choice variables. Finally, Model 3 corresponds to the selected model with both chance and choice variables. Let $\hat{u}_{it}^j$ denote the predicted probability that individual i is unemployed in period t according to Model j, where $t \in \{1988q2, 1988q3, \ldots, 2008q4\}$. Let J_t denote the set of females included in the sample in period t, and n_t the corresponding total number of females.

The predicted unemployment rate in period t according to Model j is given by

$$\hat{u}_t^j = \frac{1}{n_t} \sum_{i \in J_t} \hat{u}_{it}^j. \tag{B1}$$

As a measure of how well the different models are capable of explaining the variation in the sample unemployment rate we employ the following measure

$$D_j = 100\sqrt{\frac{1}{n_t}(\hat{u}_t^j - u_t)^2}\text{, } j = \text{Model 1, Model 2 and Model 3.} \tag{B2}$$

[12] Corresponding empirical unemployment rates are available for Sample II.

Table B2. The explanatory power of different models at the macro level according to Model Class and selection method

Model Class/Selection method	Measures of deviation		
	$D_{Model\ 1}$	$D_{Model\ 2}$	$D_{Model\ 3}$
Sample I			
One stage selection			
NU	0.965	0.848	0.671
FU	0.965	0.586	0.580
MU	0.965	0.640	0.609
BU	0.965	0.601	0.596
Two stage selection			
NU	0.965	0.783	0.589
FU	0.965	0.561	0.532
MU	0.965	0.595	0.541
BU	0.965	0.572	0.556
Sample II			
One stage selection			
NU	0.946	0.852	0.674
FU	0.946	0.617	0.619
MU	0.946	0.692	0.656
BU	0.946	0.617	0.619
Two stage selection			
NU	0.946	0.731	0.539
FU	0.946	0.523	0.500
MU	0.946	0.658	0.623
BU	0.946	0.528	0.504