RESEARCH
SERIES
NUMBER 149
OCTOBER 2022

PREDICTING THE PROBABILITY OF LONG-TERM UNEMPLOYMENT AND RECALIBRATING IRELAND'S STATISTICAL PROFILING MODEL

SEAMUS MCGUINNESS, PAUL REDMOND, ELISH KELLY AND
KONSTANTINA MARAGKOU

ESRI ECONOMIC & SOCIAL
RESEARCH INSTITUTE

PREDICTING THE PROBABILITY OF LONG-TERM UNEMPLOYMENT AND RECALIBRATING IRELAND'S STATISTICAL PROFILING MODEL

Seamus McGuinness

Paul Redmond

Elish Kelly

Konstantina Maragkou

October 2022

RESEARCH SERIES

NUMBER 149

Available to download from www.esri.ie.

Whitaker Square, Sir John Rogerson's Quay, Dublin 2

https://doi.org/10.26504/rs149

ABOUT THE ESRI

The mission of the Economic and Social Research Institute is to advance evidence-based policymaking that supports economic sustainability and social progress in Ireland. ESRI researchers apply the highest standards of academic excellence to challenges facing policymakers, focusing on 12 areas of critical importance to 21st century Ireland.

The Institute was founded in 1960 by a group of senior civil servants led by Dr T.K. Whitaker, who identified the need for independent and in-depth research analysis to provide a robust evidence base for policymaking in Ireland.

Since then, the Institute has remained committed to independent research and its work is free of any expressed ideology or political position. The Institute publishes all research reaching the appropriate academic standard, irrespective of its findings or who funds the research.

The quality of its research output is guaranteed by a rigorous peer review process. ESRI researchers are experts in their fields and are committed to producing work that meets the highest academic standards and practices.

The work of the Institute is disseminated widely in books, journal articles and reports. ESRI publications are available to download, free of charge, from its website. Additionally, ESRI staff communicate research findings at regular conferences and seminars.

The ESRI is a company limited by guarantee, answerable to its members and governed by a Council, comprising 14 members who represent a cross-section of ESRI members from academia, civil services, state agencies, businesses and civil society. The Institute receives an annual grant-in-aid from the Department of Public Expenditure and Reform to support the scientific and public interest elements of the Institute's activities; the grant accounted for an average of 30 per cent of the Institute's income over the lifetime of the last Research Strategy. The remaining funding comes from research programmes supported by government departments and agencies, public bodies and competitive research programmes.

Further information is available at www.esri.ie.

THE AUTHORS

Seamus McGuinness is a Research Professor, and Paul Redmond and Elish Kelly are Senior Research Officers, at the ESRI. Konstantina Maragkou was a Postdoctoral Researcher at the ESRI when the research was undertaken. Professor McGuinness, Dr Redmond and Dr Kelly also hold adjunct research positions at Trinity College Dublin (TCD), and Professor McGuinness and Dr Kelly are research fellows at the Institute of Labor Economics (IZA) in the University of Bonn.

ACKNOWLEDGEMENTS

The research undertaken in this study was funded through a research programme with the Department of Social Protection. We would like to thank those individuals within the Department that assisted us during the course of the project. In particular, we are grateful to Barry Kennedy, Lisa Arnold, and Keith Hanley in the Activation and Case Management Unit, and to Dermot Corcoran, Pat Collins and Paul Baynes in the Statistics and Business Analytics Unit for their active engagement with the project. We would also like to thank the independent reviewers that provided comments on an earlier draft of the report.

TABLE OF CONTENTS

Acronyms iii
Executive summary v

Section 1 Background to the study 1
Section 2 Statistical profiling in other countries 3
2.1 The evolution of statistical profiling 3
2.2 Current international applications of statistical profiling 4
2.3 Statistical profiling in combination with other profiling mechanisms 10
Section 3 Data and sample selection 15
3.1 Analytical sample 15
Section 4 Results 19
Chapter 5 Summary and conclusions 35

References 39

Appendices 41

LIST OF FIGURES

Figure 4.1 Kaplan-Meier survival function: Exits from Live Register (2018 Data) 20
Figure 4.2 Kaplan-Meier survival function: Exits from Live Register (2006 Data) 20

LIST OF TABLES

Table 2.1 Countries implementing statistical profiling.. 7
Table 2.2 Countries implementing non-statistical profiling tools 14
Table 3.1 Sample attrition 16
Table 3.2 Sample characteristics 17
Table 4.1 The probability of exiting the Live Register in 2006 and 2018 by gender 22
Table 4.2 Reliability tests: 2018 and 2006 models by gender 25
Table 4.3 Pooled probit for exiting the Live Register prior to 12 months based on 2018 data 26
Table 4.4 Decomposition analysis on pooled model (males and females); Includes gender + variables with individual R-Squared statistics above 3 (8 questions)...................... 30
Table 4.5 Reduced PEX model (8 questions) 32
Table A.1 Excluded cases 41
Table A.2 Pooled model estimated using both a restricted and unrestricted sample 43

Acronyms

AI	Artificial intelligence
AMS	Arbeitsmarktservice, Austria’s public employment service
AST	Assessment Support Tool
CE	Community Employment scheme
DSP	Department of Social Protection
JA	Jobseeker’s Allowance
JB	Jobseeker’s Benefit
JCSI	Job Seeker Classification Instrument
JSS	Jobseeker Snapshot
KM	Kaplan-Meier (KM) survival function
OECD	Organisation for Economic Co-operation and Development
OES	Services Australia’s Online Employment Services
PES	Public employment services
PEX	Probability of Exit statistical profiling model
PRSI	Pay-related social insurance
SEA	State Employment Agency (Latvia)
STAR	Danish Agency for Labour Market and Recruitment
UWV	Dutch Institute for Employee Benefit Schemes
VDAB	Public employment service (Denmark)
WPRS	Worker Profiling and Re-employment Services

EXECUTIVE SUMMARY

- Ireland's Probability of Exit (PEX) statistical profiling model predicts the likelihood that a claimant will still be unemployed 12 months after the day that they make their initial unemployment benefit claim. The model was initially developed by the Economic and Social Research Institute (ESRI) in 2009. At that time, the ESRI highlighted that the development of a statistical profiling tool was not a once-off procedure but that continuous assessment and updating of the model were required in order to maintain and improve its accuracy levels.
- In 2021, the Department of Social Protection (DSP) commissioned the ESRI to recalibrate the PEX model. The researchers were also tasked with identifying the key characteristics used in the model to determine a claimant's risk of becoming long-term unemployed while, at the same time, retaining model accuracy.
- The data used to recalibrate the model consisted of individuals who initiated a claim between 1 August 2018 and 31 October 2018: in the analysis that follows, this is referred to as the *sample period*. As the aim of the study is to evaluate the determinants of long-term unemployment risk, we have information on how these claims progressed over a *follow-on period*, which extended to 1 February 2020.
- There was some variation in the restrictions imposed for the estimation of both the initial (2009) and present (2022) PEX models, and there has also likely been slight changes in the rules for accessing benefit over the time period covered by the study. Nevertheless, the approach adopted in estimating both models is broadly consistent.
- The descriptive analysis shows that the rate of exit from unemployment is relatively stable for the first 15 weeks of unemployment. After this, the rate at which individuals leave the Live Register slows down. By the time an individual reaches 50 weeks of unemployment, the probability of exiting the Live Register has declined substantially.
- Despite the fact that 11 years separate the two PEX model estimations, with a global recession occurring in the intervening years, the pattern of exit from unemployment in 2018 was similar to that estimated using a sample from 2006 for the initial PEX model. The determinants of long-term unemployment in 2018 were also found to be similar when compared to the earlier 2006 analysis. However, both 2018 and 2006 were characterised by low unemployment rates and it is possible that exit patterns from unemployment could be different in a recessionary period of high unemployment.
- In both 2006 and 2018, and for both genders, the probability of exiting the Live Register before 12 months duration declines with age, literacy problems, the presence of children, a previous spell of long-term unemployment and being casually employed. Conversely, for both periods and both genders, the probability of exiting the Live Register prior to 12 months was positively correlated with being in very good health.

- Despite having a common sign, the size of some of the estimated impacts vary across time and gender; for example, the coefficients on age tend to be substantially larger for both males and females in the models based on the most recent data, indicating that older workers were less likely to exit the Live Register prior to 12 months in 2018 compared to 2006.
- For males, recent job duration determines an exit from the Live Register in the 2018 model despite being largely irrelevant in the initial PEX model that used 2006 data. Conversely, while willingness to move for a job was a statistically significant factor for males in 2006, it had no predictive power in 2018. Finally, in 2018, location (village, town or city) was much less important, for both males and females, in predicting the probability of exiting the Live Register than it was in 2006.
- Marital status was an important factor for women in 2006 but was no longer relevant in 2018; conversely, while access to one's own transport was not important in the 2006 female model, it was a statistically significant factor in 2018.
- Comparing the diagnostic statistics of both models, it is clear that the 2018 model has greater predictive power compared to the 2006 models.
- Following a variance decomposition analysis, which was the method used to determine the key characteristics associated with long-term unemployment risk without losing model accuracy, we retained eight questions that, between them, account for 85 per cent of the variance in the data explained by the full model. The variables retained in order of importance were: (1) claiming Jobseeker's Allowance (JA); (2) recent employment history; (3) education; (4) self-perceived health; (5) a history of long-term unemployment; (6) having previously been on the Community Employment (CE) scheme; (7) having access to own transport; and (8) age.
- The performance of the 2018 model, based on just eight questions, is very similar to that found for the 2018 model based on the full data, suggesting that the reduction in the number of model variables does not come at the expense of model accuracy.

SECTION 1

Background to the study

This report provides the results from a study undertaken by the Economic and Social Research Institute (ESRI) for the Department of Social Protection (DSP) to re-estimate the statistical profiling model underlying Ireland's activation system – the Probability of Exit (PEX) model.[1] The primary objective of the project has been to recalibrate the PEX model. In doing so, we also identify the key characteristics associated with long-term unemployment risk while, at the same time, retaining maximum model accuracy.

Specifically, we:

- used recent profiling and administrative data covering the 2018 to 2019 time period to develop a recalibrated PEX model that estimates a jobseeker's risk of becoming long-term unemployed on the day that they make their initial unemployment claim;
- compared the results of the new model with those underlying the current PEX model that was developed in 2009 using 2006 data;
- used decomposition techniques to identify the key variables of importance in the new model for estimating a claimant's probability of exiting unemployment to the labour market before 12 months; and
- conducted a comparative analysis of the PEX model estimated on the full data with that estimated with a more limited set of characteristics to ensure that the reduced model, which is based on the key determinants of long-term unemployment risk, is not achieved at the cost of any substantial reduction in model accuracy.

It transpires that the recalibrated PEX model based on 2018 data focuses on human capital characteristics. Thus, going forward, it is very unlikely that the model's accuracy will be heavily influenced by cyclical factors. This will allow the DSP to adjust the model to changes in labour market conditions by applying additional weights if needed to cyclical variables, such as sector of employment.

In the next section of the report (Section 2), we discuss the evolution of statistical profiling and its current application within public employment services (PES) internationally. Section 3 outlines the data used in the study. The results from the

1 The PEX model uses statistical profiling to identify those at risk of becoming long-term unemployed. Those who are statistically identified as being most distant from the labour market, along with those under 30 years of age, are engaged with most frequently. The nature of interventions and supports offered are then informed by individual assessments by DSP Intreo case officers, who develop a broader profile of the jobseeker and agree a personal progression plan with them, which takes account of their full circumstances.

recalibration work are presented in Section 4, while Section 5 summarises the main findings from the study.

SECTION 2

Statistical profiling in other countries

2.1 THE EVOLUTION OF STATISTICAL PROFILING

The origins of statistical profiling to identify jobseekers at risk of becoming long-term unemployed dates back to the 1990s. In 1998, a study by the Organisation for Economic Co-operation and Development (OECD) found that Australia, Canada, the United Kingdom (UK) and the United States (US) had begun to design and introduce more formal systems to identify at-risk jobseekers. After this time period, advances in technology and computing power, along with the quality of administrative and survey data needed to develop robust statistical models, led to additional countries experimenting with various profiling tools. When Ireland's statistical profiling model was instigated in 2009, Australia, Denmark, Germany, and the US public employment services (PES) had already implemented fully operational systems. At that time, New Zealand, the Netherlands and South Korea were also experimenting with some forms of profiling, not necessarily statistical-based, as a means of targeting their employment services (see O'Connell et al., 2009). However, budgetary pressures experienced by PES after the Great Recession, along with more emphasis being placed on labour market activation, increased the attention given to, and implementation of, statistical profiling models across a number of countries in the last decade.

With the recent emergence of new data sources, such as 'click data' on job searches and 'big data', along with advanced machine learning techniques (van Landeghem et al., 2021; Desiere et al., 2019), statistical profiling tools have become even more widespread since Ireland's system was originally developed. According to van Landeghem et al. (2021), machine learning techniques are better than standard regression models at predicting someone's likelihood of becoming long-term unemployed. However, other research has found that models based on machine learning, such as Belgium's statistical profiling tool, are not any more accurate than those based on traditional statistical techniques, such as logistic regression (Desiere et al., 2019).

In addition to statistical profiling, three other methods that are mainly used by PES to identify those at risk of becoming long-term unemployed are: i) eligibility rules, ii) caseworker discretion, and iii) screening (Hasluck, 2008).[2] Some PES have

[2] Eligibility rules, which are set by PES, concern channelling jobseekers towards various forms of re-employment support on the basis of meeting certain criteria. Caseworker discretion is where the individual caseworker uses their own judgement to direct a jobseeker towards the type of intervention that they feel is most appropriate to meet the jobseeker's needs. Screening is where a caseworker attempts to score a claimant's employability using typically

implemented one of these approaches on their own, while others have used them in combination with statistical profiling. For example, Denmark and New Zealand combine statistical profiling with caseworker discretion. Austria implemented a statistical profiling tool in 2018, but they also have a rule that all jobseekers aged under 25 are assigned to a particular intervention stream regardless of their profiling score. In addition, caseworkers in Austria can ignore the results of the profiling model and assign jobseekers to a particular intervention stream that they think is most appropriate for the jobseeker. The Netherlands' statistical profiling tool determines when a jobseeker will first be invited to an interview with a caseworker, but they also have a rule that all jobseekers will be invited after six months even if their initial profiling score identified otherwise (Desiere et al., 2019).

In this section, we predominately focus on those countries that have solely implemented a statistical profiling tool. These countries are listed in the first panel of Table 2.1. We also give a brief overview of some of the PES that use statistical profiling in combination with caseworker discretion, eligibility rules and/or other profiling mechanisms: these countries are outlined in the second panel of Table 2.1, while those that use only non-statistical profiling methods are set out in Table 2.2.[3]

2.2 CURRENT INTERNATIONAL APPLICATIONS OF STATISTICAL PROFILING

The US and Australia were the first two countries to implement fully operational statistical profiling systems. Since 1993, each US state has been required by law to develop its own Worker Profiling and Re-employment Services (WPRS) system. All states have had fully operational WPRS systems in place since 1996, with implemented systems involving a partnership between states' unemployment insurance, public employment services and public job training bodies. Statistical models vary by state, with data collected from claimants during their initial claim and/or work registration process. Labour market information, such as the local unemployment rate, is also included in some models. According to Sullivan et al. (2007), which appears to be the only review of the WPRS since its inception, updating of statistical models also varies across states. This is despite the fact that a 1998 state and federal government review recommended that states should update and revise their profiling models regularly, and add new variables and revise model specifications when necessary (Sullivan et al., 2007). At the time of their review, Sullivan et al. (2007) found that some states had never updated or

psychological-based techniques: on the basis of a resulting ordinal employability score, the jobseeker is directed towards the intervention that is designed to meet their particular score.

3 See Desiere et al. (2019) for an overview of all profiling systems currently in use or under development in OECD countries.

revised their systems since originally implemented. They also found that, among those that *had* updated or revised their system, the main reason was to upgrade the occupational classification and industry classification systems used.[4] Again, in 2015, the US Department of Labor advised states to ensure that their WPRS statistical profiling models used appropriate information that had been shown to accurately predict the likelihood of unemployment insurance (UI) benefit exhaustion in their state, and for the coefficients to be updated as needed.[5]

In common with the US, Australia has been implementing a statistical profiling model since the early 1990s. Their current system, the Job Seeker Classification Instrument (JSCI), has been in place since 1998. This tool is operated by Centrelink, a subsidiary of Services Australia, an agency of the Australian government. Until recently, the data used in the JCSI statistical model were collected through a questionnaire-based interview conducted when a jobseeker first submitted a claim for income support.[6] When a claimant discloses new information, or there is a major change in their circumstances, Services Australia requires a reassessment of their JSCI score.[7] In 2018, a trial commenced to collect the JSCI data online, known as the Jobseeker Snapshot (JSS). Since April 2020, the JSS has been rolled-out as part of Services Australia's Online Employment Services (OES) platform.[8] The JSS captures the same information as the interview-based JSCI. However, the wording and sequencing of some questions have been modified for the online environment. Jobseekers who are unable to complete the JSS can provide the information needed for the JCSI statistical model through either a telephone or face-to-face interview. Since the JSCI statistical model was introduced in 1998, it has been re-estimated with each iteration of the mainstream employment services programme. For example, it was re-estimated in 2015 when Jobactive was introduced to replace Job Services Australia (Department of Education, Skills and Employment, 2021).[9] The inclusion of new factors, or sub-factors, into the JSCI is generally based on expert advice, consultation and academic research (Desiere et al., 2019).

In 2006, the Dutch Institute for Employee Benefit Schemes (UWV) established a project to create the Netherland's statistical profiling model, the Work Profiler (Wijnhoven and Havinga, 2014). The project, which consisted of three strands of

4 The occupational classification system was updated from DOT to SOC or O*Net, and the industry classification system from SICs to NAICS (Sullivan et al., 2007).

5 See US Department of Labor's 2015 advisory to state workforce agencies, at Employment and training administration advisory system (doleta.gov).

6 Most interviews occurred over the phone (75 per cent), with the remainder (25 per cent) conducted face to face (Department of Education, Skills and Employment, 2021).

7 Known as a Change of Circumstances Reassessment (CoCR).

8 The OES is a digital platform that allows job-ready jobseekers to self-manage job search and compliance activities.

9 Jobactive is a government-funded programme that helps jobseekers to find work. It also assists employers in finding suitable candidates for their vacancies. The programme is delivered by MTC Australia.

research,[10] and was run from 2006 to 2011, was undertaken by the UWV Centre for Knowledge and the School of Medical Sciences of the University Medical Centre Groningen (UMCG). The aim of the project was to identify the predictive characteristics of an individual and/or their personal situation for resuming work within 12 months of becoming unemployed. As with Ireland's Probability of Exit (PEX) statistical profiling model, the UWV wanted to identify the characteristics that are predictive at the start of the jobseeker's unemployment spell in order to identify their risk of becoming long-term unemployed. The project resulted in the identification of 11 predictive factors for their Work Profiler statistical model. The data for these factors, which are divided into hard (e.g., age, years employed in last job) and soft (e.g., views on return to work, job search behaviour, job search intention), are derived from 20 questions that jobseekers complete via an online questionnaire during their first three months of becoming unemployed.[11] The profiling tool also provides the UWV with insights into which factors their PES can influence at the start of a jobseeker's unemployment spell, so that the PES can identify the most appropriate computerised services to offer to jobseekers to enhance their probability of resuming work. During 2013, the Work Profiler was rolled out across 11 Dutch unemployment offices. This first version of the Netherland's statistical profiling tool was based solely on data captured in the province of North Holland. A second version of the model, which was updated during 2014, was based on data collected from 11 unemployment offices scattered throughout the Netherlands. The number of questions in the underlying profiling questionnaire during this updating process was extended from 20 to 55. The new predictive factors from this process, and their corresponding questions, form the second version of the Work Profiler statistical model, which was to be rolled out in the Netherlands in 2017 (Wijnhoven and Havinga, 2014).

---

10 The three strands, which were carried out one after the other, were: a literature review; a cross-sectional study; and a longitudinal study (see Wijnhoven and Havinga, 2014).

11 The longitudinal component of the Work Profiler project resulted in the number of factors to be included in the statistical profiling model being reduced from 155 to 20. This study was carried out in the province of North Holland between April 2008 and March 2009; its participants included all those who became unemployed in that period, a total of 3,618 individuals.

TABLE 2.1 COUNTRIES IMPLEMENTING STATISTICAL PROFILING

Country	Profiling system	Name of profiling tool	Outcome	Data source	Statistical model	Accuracy	Used for
I. Statistical profiling only							
Australia	Statistical profiling	Job Seeker Classification Instrument (JSCI)	Long-term unemployed (12 months)	Either online, phone or face-to-face questionnaire-based interview consisting of 49 questions, completed when a person first claims for income support. A minimum of 18 to be answered, with the remaining number depending on the jobseeker's level of disadvantage.	Logistic regression	–	Prioritising jobseekers: claimants are ordered by degree of labour market disadvantage, going from least disadvantaged (1) to most disadvantaged (76). They are then allocated to streams, based on their JCSI score, for varying levels of support.
Ireland	Statistical profiling	Probability of Exit (PEX) model	Probability of exit to employment within 12 months	Questionnaire as part of benefit claim process, administrative data	Probit regression	70%–86%	Determining minimum levels of engagement, which depend on distance from the labour market.
Italy	Statistical profiling		Long-term unemployed (12 months)	Administrative data	Logistic regression	–	Prioritising jobseekers; targeting; resource allocation.
Netherlands	Statistical profiling	Work Profiler	Probability of work resumption within a year.	Online questionnaire containing hard and soft questions, completed within the first three months of unemployment.	Logistic regression	70%	Prioritising jobseekers (selection and diagnosis); high and medium risk (0–50% chance of work resumption within a year) will receive personalised services (early intervention)
United States	Statistical profiling	Worker Profiling and Reemployment Services (WPRS)	Varies by state: mainly probability of exhausting unemployment insurance (UI) benefits.	Data collected from claimants during initial claim and/or work registration process; necessary labour market data.	Varies by state: logistic regression, linear multiple regression, tobit, neural network, discriminant analysis.[1]	–	Prioritising jobseekers: claimants are ranked highest to lowest in order of their probability of exhaustion of benefits.[2]

TABLE 2.1 (CONTD.) COUNTRIES IMPLEMENTING STATISTICAL PROFILING

Country	Profiling system	Name of profiling tool	Outcome	Data source	Statistical model	Accuracy	Used for
II. Statistical profiling combined with other profiling methods							
Austria	Statistical profiling, eligibility rules, caseworker discretion	*PAMAS: Personalisiertes Arbeitsmarktchancen-Assistenzsystem* (Personal Labour Market Opportunities – Assistance System)	Labour market integration probability (high: 3 months of unsubsidised employment within 7 months; low: 6 months of unsubsidised employment within 24 months; middle: all other outcomes).	Administrative data.	Logistic regression	80%–85%	Prioritising jobseekers.
Belgium (Flanders)	Statistical profiling, caseworker discretion	Next Steps	Long-term (>6 months) unemployed.	Administrative data; 'click data'.	Random forest model	Accuracy: 67% (AUC about 76%)	Prioritising jobseekers.
Denmark	Statistical profiling, caseworker discretion	*Profilafklaringsværktøjet* (The Profiling Tool)	Long-term (>26 weeks) unemployed.	Online questionnaire; Administrative data.	Big data model	>60%	Prioritising jobseekers.
New Zealand	Statistical profiling, caseworker discretion	Service Effectiveness Model (SEM), Liability Estimator Tool (LET)	Lifetime income support costs (LET), change in lifetime income support and staff costs from receiving a case management service (SEM).	SEM/LET are based on administrative data.	Random forest (LET), Gradient boosting (SEM)	AUC: 0.63 - 0.83	Allocating clients to intensive case management services that they are most likely to benefit from.

TABLE 2.1 (CONTD.) COUNTRIES IMPLEMENTING STATISTICAL PROFILING

Country	Profiling system	Name of profiling tool	Outcome	Data source	Statistical model	Accuracy	Used for
Latvia	Statistical profiling, personal motivational interview, labour market information		Long-term unemployed (12 month).	Online questionnaire, caseworker interview, administrative data.	Factor analysis	No data yet	Risk factors of unemployment; individual action plan (IAP) based on opportunities to find a job; motivation to search for a job; motivation to collaborate with SEA (Latvia PES); skills (soft & hard skills for job search); and prioritising of ALMPs.
Sweden	Statistical profiling, caseworker discretion	Bedömningsstödet (The Assessment Support Tool (AST))	Long-term unemployed (6 months).	Administrative data.	Logistic regression	–	Prioritising jobseekers.

Source: Adapted from Desiere et al. (2019), 'Statistical profiling in public employment services', OECD Social, Employment and Migration Working Papers No. 224, OECD Publishing, Paris.
Note: [1] Some state workforce agencies use the Characteristic Screen Model instead of a statistical model to identify appropriate claimants that need reemployment services.[12]
[2] If the Characteristic Screen Model is used, this gives rise to a list of claimants thought likely to exhaust their unemployment insurance (UI) benefits.

[12] See US Department of Labor's relevant webpage, Worker profiling and reemployment services.

2.3 STATISTICAL PROFILING IN COMBINATION WITH OTHER PROFILING MECHANISMS

Sweden developed its statistical profiling tool, called the Assessment Support Tool (AST), in 2011. Used in combination with caseworker discretion, AST allows the PES to identify those jobseekers at risk of becoming long-term unemployed so that they can intervene early to prevent this outcome. The Swedish government took the decision to develop this statistical profiling tool as it was felt that such a quantitative method was needed to correctly identify those jobseekers at risk of becoming long-term unemployed. When the AST was being developed, several lessons were learnt from the pilot phase. Specifically, there needed to be buy-in from PES management and caseworkers if the tool was to be implemented successfully. Second, the AST was identified as being an additional information stream to caseworkers, rather than a replacement of them. This is because the Swedish PES felt that labour market success for a jobseeker depended on other factors not easily captured through statistical models; for example, social networks, ambition, and mental and physical strength. Thus, in developing its statistical profiling tool, Sweden placed more weight on caseworker discretion in making a final decision on the support to be given to a jobseeker to assist them to reintegrate into the labour market. The AST was also only developed with the intention of profiling for segmentation and not for automatic targeting; it would only aim to differentiate between low-risk and high-risk jobseekers and would not be used to automatically assign them to a certain type of intervention aimed at assisting them to integrate back into the labour market. Data are gathered on 11 predictors and the information is captured by caseworkers during an initial interview held when a jobseeker registers as unemployed. The AST produces a risk estimate for a jobseeker, with jobseekers allocated to one of four different risk categories of long-term unemployment on the basis of their AST score. Caseworkers use this score to help them to make a final decision about segmentation and the most appropriate measures to assist a jobseeker reintegrate into the labour market (Loxha and Morgandi, 2014). No information is available in relation to whether the AST has been revised or updated, or how regularly this might occur.

Latvia's State Employment Agency (SEA) gradually introduced its profiling tool, which is applied to all registered unemployed, in 2013, and it was fully rolled out by November of that year.[13] Like Ireland's PEX model, this is used to identify those at risk of becoming long-term unemployed so that appropriate measures can be implemented. The Latvian profiling tool is based on a combination of: i) a statistical

---

[13] One of the reasons for the introduction of a profiling tool in Latvia was to ensure that the same services were provided to persons with similar needs across the different local PES offices and caseworkers.

model; ii) a SEA caseworker's assessment of the jobseeker's motivation to cooperate with the SEA and search for a job (assigned to one of 13 groups); iii) a jobseeker's self-assessment of their skills (assigned a low, medium or high skill score); and iv) current labour market conditions (assigned to one of three groups). All this information is combined to assign each jobseeker to one of 39 PES intervention groups. Specifically, the Latvian model indicates the relevant services and appropriate frequency of local PES office visits for each jobseeker. In relation to the statistical model, the data for this tool are obtained from a questionnaire that jobseekers complete online when they register as unemployed. Information gathered includes the jobseeker's age, educational attainment, language skills, employment history, place of residence, barriers to employment, family obligations, etc. The statistical tool also predicts a jobseeker's likelihood of finding a job by using information from a data system that contains the average length of unemployment for the groups of clients with the same demographic profile as the jobseeker. This 'average length of unemployment' data reflects the situation in the previous 27 months, and these data are updated every time a new client is registered, and a profile constructed for them. In general, after six months, the motivation of the jobseeker and their willingness to cooperate with the SEA are reassessed. However, this can be undertaken before this time point if there is a substantial change in the situation of the unemployed person.[14] As with Sweden's AST model, no information is available with regards to whether Latvia's statistical profiling model has been revised or updated, or how regularly this might take place.

In October 2018, Austria's public employment service, Arbeitsmarktservice (AMS), announced that it would roll out a statistical profiling system, specifically to identify jobseekers' prospects on the job market, both short-term and long-term. On the basis of their statistical model, known as the AMSA Model (AMS-Arbeitsmarktchancen-Modell),[15] individuals are categorised into three groups, with each group involving different supports to enter/re-enter the labour market (Allhutter et al., 2020). However, the profiling results are overridden for jobseekers aged less than 25, who are always assigned to the middle service stream. Caseworkers also have the discretion to ignore the results from the AMS model and assign a jobseeker to the service stream that they consider to be the most appropriate (Desiere et al., 2019). After a pilot phase in 2019, AMSA model was to be rolled out nationwide in July 2020 (Allhutter et al., 2020). The data used to estimate Austria's statistical profiling model come from two administrative data sources: i) data collected through self-reporting by jobseekers when they register online with the AMS and through interactions with the AMS; and ii) social security data from the Main Association of Austrian Social Security Institutions (Allhutter et

14 See OECD iLibrary | Home (oecd-ilibrary.org) .
15 Austria's statistical profiling model is also often referred to as the 'AMS algorithm' (Allhutter et al., 2020).

al., 2020). The information captured through these two data sources includes jobseekers' age, gender, nationality, education, health limitations, care responsibilities, prior work experience, frequency and duration of unemployment and participation in active labour market programmes (ALMPs), along with regional labour market developments. Labour market history information tends to be incomplete for young people, migrants and those who have spent long periods outside of the labour market, so different models are estimated for various subgroups (Desiere et al., 2019).

Flanders, in Belgium, was one of the first regions to develop an artificial intelligence (AI) statistical profiling model to identify those at risk of becoming long-term unemployed (Desiere and Struyven, 2020).[16] Its PES, the VDAB, founded an innovation lab in 2014 to focus on 'big data' analytics and, in the process, developed Next Steps, its statistical profiling model. This model is part of a new 'contract strategy' that was rolled out in Flanders in October 2018. The aim of this strategy is to screen all new jobseekers within six weeks of their registering at the PES. However, those jobseekers identified by Next Steps as being at high risk of long-term unemployment are prioritised for contacting (Desiere et al., 2019). The statistical tool assigns a jobseeker a profiling score regarding their likelihood of gaining employment within the next six months, 35 days after registration at their PES. On the basis of their score, jobseekers are assigned to one of four groups, based on their likelihood of resuming work. Those with the highest risk of long-term unemployment are contacted first by the PES (Desiere and Struyven, 2020). Next Steps uses a random forest model. The underlying data, which include information on a jobseeker's socio-economic characteristics and information on their labour market history, are collected and stored in a data warehouse. Information collected by caseworkers during previous and current unemployment spells are also included in the model, as are 'click data' capturing jobseekers' activity on the VDAB website, including clicking on job vacancies, as proxies for job search behaviour and motivation. Next Steps is built in a flexible way so that it can be updated regularly in order to remain accurate. Since its inception in 2018, the model has already been updated (February 2019) to reduce the number of explanatory variables included in the estimation process. This update was undertaken to simplify the model and to comply with privacy regulations and anti-discrimination law. Even with the reduction in the number of variables used to estimate the model, the most recent version of Next Steps was found to have a similar accuracy level to the original version developed in January 2018 (Desiere and Struyven, 2020). The tool is used to assist VDAB caseworkers in their decision making. Specifically, the tool only ensures that the most vulnerable jobseekers are

[16] Defined in Flanders as greater than six months.

contacted first: it has no impact on the subsequent referral decision that a caseworker makes (Desiere et al., 2019; and Desiere and Struyven, 2020).

As with Flanders, both Denmark and New Zealand have used advanced machine learning techniques to build their statistical profiling models. Denmark's PES, the Danish Agency for Labour Market and Recruitment (STAR), developed its profiling tool using the machine learning technique, 'decision tree classification'. This methodology identifies nine paths that predict a jobseeker's likelihood of becoming long-term unemployed.[17] A combination of administrative data and online survey data that capture behavioural information are used to estimate the model. The Danish profiling model supports caseworkers in making a decision on the most appropriate assistance for a jobseeker to enter/re-enter the labour market, as opposed to it being used as stand-alone instrument to determine the most appropriate labour market reintegration path for a jobseeker (Desiere et al., 2019). It is the same in New Zealand, where their statistical profiling models are estimated using administrative data and which employ random forest and gradient boosting techniques (Desiere et al., 2019).

[17] Defined as greater than 26 weeks in Denmark.

TABLE 2.2 COUNTRIES IMPLEMENTING NON-STATISTICAL PROFILING TOOLS

Country	Profiling system	Name of profiling tool	Outcome	Data source	Used for
Estonia	Caseworker based profiling				Allocating individuals between two different counselling levels.
Germany	Caseworker-based profiling	4 Phase Model			Categorising jobseekers.
Greece	Data assisted profiling, caseworker-based profiling			Online self-assessment questionnaire; administrative data	Case worker support; targeting; resource allocation.
Luxembourg[1]	Caseworker-based profiling				Targeting; resource allocation.
Norway	Rule-based profiling	*Forenklet oppfølging* (‘Simplified follow-up’)			Prioritising jobseekers; targeting; resource allocation.
Poland	Rule-based profiling (point-based system)		Assignment to one of three support type groups.	Online questionnaire; administrative data.	Targeting.
Slovenia	Caseworker-based profiling	Employability assessment			Tailoring services, prioritising jobseekers.
Switzerland	Caseworker-based profiling				Classifying jobseekers into easy, average, difficult to place into employment; determining the frequency of counselling meetings.
UK	Rule-based profiling				

Source: Adapted from Desiere et al. (2019), ‘Statistical profiling in public employment services’, OECD Social, Employment and Migration Working Papers No. 224, OECD Publishing, Paris.
Note: [1] Statistical profiling under development.

SECTION 3

Data and sample selection

The data for this study were provided to the ESRI by the Department of Social Protection (DSP), following a consultation process between the two bodies that sought to identify the data required to recalibrate the Probability of Exit (PEX) model. On the basis of this consultation, the derived dataset consists of individuals who initiated a claim between 1 August 2018 and 31 October 2018, which we refer to as the *sample period*. As the aim of our study is to evaluate the determinants of long-term unemployment, we also consider how these claims progressed over a *follow-on period*, which extends to 1 February 2020. Specifically, we know whether the claimant exited or remained on unemployment benefit over this period. Moreover, in cases where the claim was closed, we know the reason for the closure (the claimant moved into employment, on to another type of social welfare benefit or the claim was ended for other reasons), which we detail later on.

3.1 ANALYTICAL SAMPLE

The dataset consists of 36,076 individuals who initiated a claim between 1 August 2018 and 31 October 2018. We identify *stayers* as individuals who initiated a claim during the *sample period* and never exited the Live Register during the *follow-on* period of 12 months (exit=0 in the data). We define *leavers* as those who exited to, and stayed in, employment within one year of initiating their Jobseeker's Allowance (JA) or Jobseeker's Benefit (JB) claim.

From the initial sample of 36,076 individuals, we ended up with a sample of 13,671 individuals, consisting of 6,230 *stayers* and 7,441 *leavers*. We provide further detail on the categories that are excluded from our analysis below, with the information summarised in Table 3.1.

- We excluded all individuals who exited the Live Register for reasons other than to employment (a full list of the closure types is provided in Table A.1 in the appendix). There were 16,888 such cases. Of the dropped cases, the largest categories relate to exits to another benefit (14.4 per cent), unsigned claims (11.8 per cent), exits to a training programme (8.2 per cent) and exits to the Community Employment (CE) scheme (6.5 per cent).
- From the remaining sample, we omitted all individuals whose claim do not relate to either JA or JB. There were 1,613 such cases in the remaining sample.

- From the sample that is left, we excluded all individuals who exit the Live Register and re-entered at any time. There were 3,829 such cases in the remaining sample.
- Finally, we omitted all individuals for whom we had missing background information. There were 75 such cases in the remaining sample. The excluded categories are summarised below.

TABLE 3.1 SAMPLE ATTRITION

Profiling data	Excluded cases	Sample
Original population		36,076
Exited the Live Register for 'other' reasons	16,888	19,188
Non-JA and JB claims	1,613	17,575
Exited and re-entered Live Register	3,829	13, 746
Missing background information	75	13, 671

Source: Authors' analysis of DSP provided project data.

In Table 3.2, we compare the characteristics of leavers and stayers in the sample. The table presents the mean characteristics for the total sample, the leavers sample and the stayers sample. For instance, 41.2 per cent of the total sample are female, while in the leavers and stayers populations, the relevant figures are 43.7 and 38.3 per cent respectively. There are some notable differences between the leaver and stayer groups. Compared to those that stayed on the Live Register for at least 12 months (stayers), individuals who exited to employment (leavers) are more likely to be female, aged 25 to 34, married, have a spouse in employment, and educated to third level. Leavers were also far more likely to report being in very good health: 65 per cent of leavers report very good health compared to just 40 per cent of stayers. In addition, leavers were also more likely to have been previously employed, with longer job tenures, and were more likely to have access to their own transport.

There were notable differences in terms of reported literacy and numeracy problems, with just five per cent of leavers reporting these issues compared to 15 per cent of stayers. Perhaps unsurprisingly, given the differences in work histories between the two groups, leavers were far more likely to be in receipt of JB (as opposed to JA) compared to stayers.[18] It is important to note that JB is based on social insurance (PRSI) contributions,[19] and can only be collected for a maximum of nine months,[20] after which time recipients can apply for the means-tested JA payment. In this study, our models measure individuals' characteristics at the commencement of their claim. Therefore, those who were in receipt of JB at the

18 Jobseeker's Benefit is payable if the individual has the required PRSI contributions, whereas Jobseeker's Allowance is means tested.

19 Pay-related social insurance (PRSI).

20 This nine-month payment applies to JB claimants with 260 or more Class A, H or P PRSI paid contributions. It is paid for 6 months (156 days) to those with less than 260 Class A, H or P PRSI paid contributions.

beginning of their claim will have been in receipt of JA if, and when, they fell into long-term unemployment.

TABLE 3.2 SAMPLE CHARACTERISTICS

	Total	Leavers	Stayers
Gender			
Female	0.412	0.437	0.383
Age group			
18–24 years	0.182	0.185	0.178
25–34 years	0.323	0.357	0.282
35–44 years	0.242	0.254	0.228
45–54 years	0.166	0.151	0.184
55+ years	0.087	0.053	0.128
Health status			
Poor health	0.016	0.008	0.026
Fair health	0.084	0.044	0.131
Good health	0.365	0.298	0.446
Very good health	0.535	0.651	0.396
Relationship status			
Single	0.598	0.593	0.604
Married	0.268	0.301	0.228
Cohabiting	0.081	0.072	0.092
Separated/divorced	0.052	0.034	0.073
Widowed	0.001	0.001	0.003
Children	0.144	0.109	0.186
Spousal earnings			
Spousal earnings none	0.912	0.883	0.945
Spousal earnings €250	0.014	0.016	0.011
Spousal earnings €251–350	0.003	0.004	0.002
Spousal earnings €351+	0.071	0.096	0.042
Education			
Primary or less	0.070	0.025	0.124
Lower secondary	0.156	0.093	0.230
Upper secondary	0.290	0.262	0.323
Third level	0.485	0.620	0.322
Apprenticeship	0.105	0.099	0.112
Literacy / numeracy probs	0.091	0.046	0.146
English proficiency	0.848	0.889	0.800

TABLE 3.2 (CONTD.) SAMPLE CHARACTERISTICS

	Total	Leavers	Stayers
Employment			
Still in employment	0.083	0.030	0.147
Employed in last month	0.032	0.026	0.038
Employed in last year	0.473	0.583	0.343
Employed in last 5 years	0.273	0.290	0.254
Employed over 5 years ago	0.080	0.058	0.107
Never employed	0.058	0.013	0.112
Casually employed	0.007	0.005	0.009
Would move for a job	0.443	0.478	0.402
Never employed	0.080	0.028	0.142
Job duration <1month	0.054	0.049	0.061
Job duration 1–6 months	0.243	0.253	0.233
Job duration 6–12 months	0.171	0.178	0.163
Job duration 1–2 years	0.151	0.156	0.144
Job duration 2+ years	0.301	0.336	0.258
UE claim previous 5 years	0.604	0.517	0.708
Signing for 12 months+	0.174	0.084	0.280
CE scheme previous 5 years	0.153	0.076	0.245
On CE scheme for 12 months+	0.071	0.029	0.121
Jobseekers Allowance	0.637	0.424	0.891
Location			
Rural	0.204	0.205	0.204
Village	0.117	0.117	0.118
Town	0.250	0.221	0.284
City	0.429	0.457	0.395
Transport			
Own transport	0.511	0.612	0.39
Near transport	0.765	0.767	0.762
Sample	13,671	7,441	6,230

Source: Authors' analysis of DSP provided project data.

SECTION 4

Results

Figure 4.1 plots the Kaplan-Meier (KM) survival function for our final analytical sample of leavers and stayers. The KM survival function, the most commonly used descriptive statistic for durations, shows the proportion of individuals leaving the Live Register to employment during successive weeks. We estimate the function for the entire 78 weeks of the follow-on period during which we track individuals. The y-axis indicates the probability of being on the Live Register and the x-axis shows the time (in weeks) since the claim was initiated.

The downward sloping line indicates that the probability of remaining on the Live Register is decreasing in the number of weeks since the claim was initiated. The shape of the curve indicates that the probability of remaining on the Live Register decreases at a decreasing rate or, put differently, the probability of finding employment increases at a decreasing rate over time. For example, initially (for approximately the first 15 weeks), the rate of unemployment exit is quite high, as indicated by the steep downward slope. After this time point, the curve begins to flatten, indicating that the rate of exit to employment slows down considerably. From the initiation of the claim (0 weeks) up to 40 weeks, the probability of being on the Live Register declines by half, from 100 per cent to 50 per cent. However, comparing 40 weeks to 80 weeks, the probability of being on the Live Register declines from 50 per cent to just 45 per cent.

In Figure 4.2, we plot the KM survival function from the earlier study on which the original PEX model was based (O'Connell et al., 2009). The similarity between both KMs is notable despite the fact that the data periods on which they are based are 11 years apart, with a major global recession taking place in that period. Both figures indicate that approximately 50 per cent of claimants had exited the Live Register between 40 and 50 weeks. By week 78, approximately 55 per cent of claimants had exited the Live Register in both time periods.

FIGURE 4.1 KAPLAN-MEIER SURVIVAL FUNCTION: EXITS FROM LIVE REGISTER (2018 DATA)

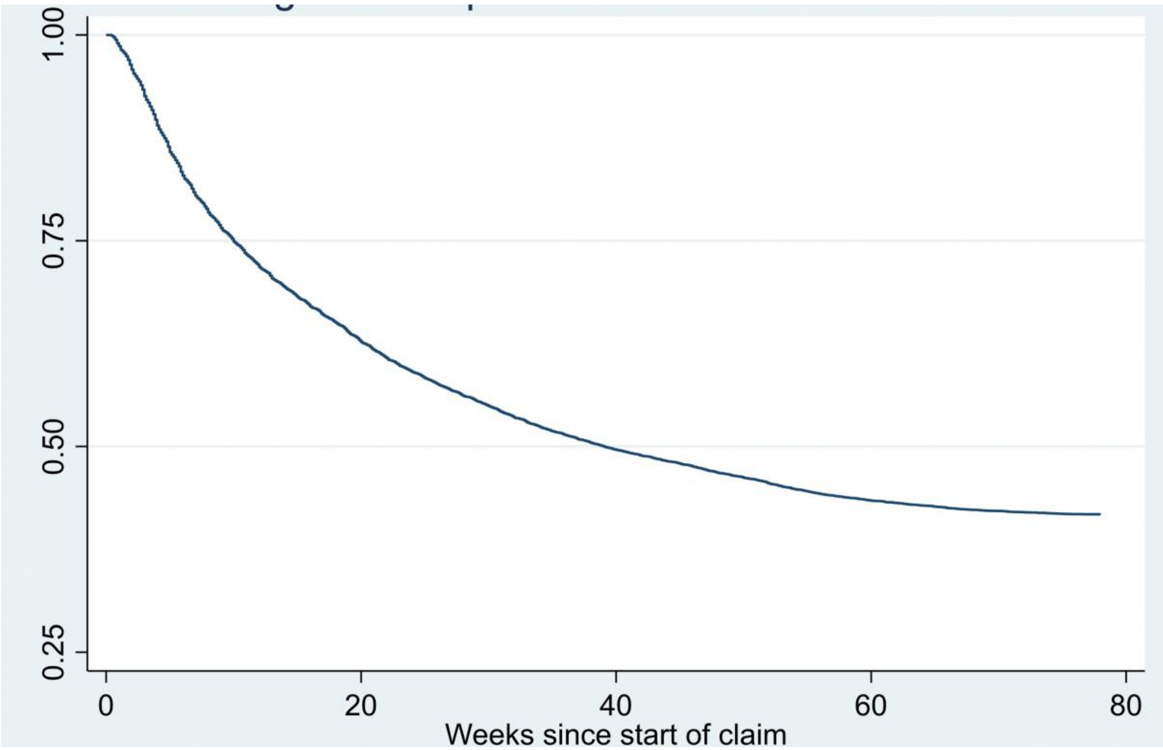
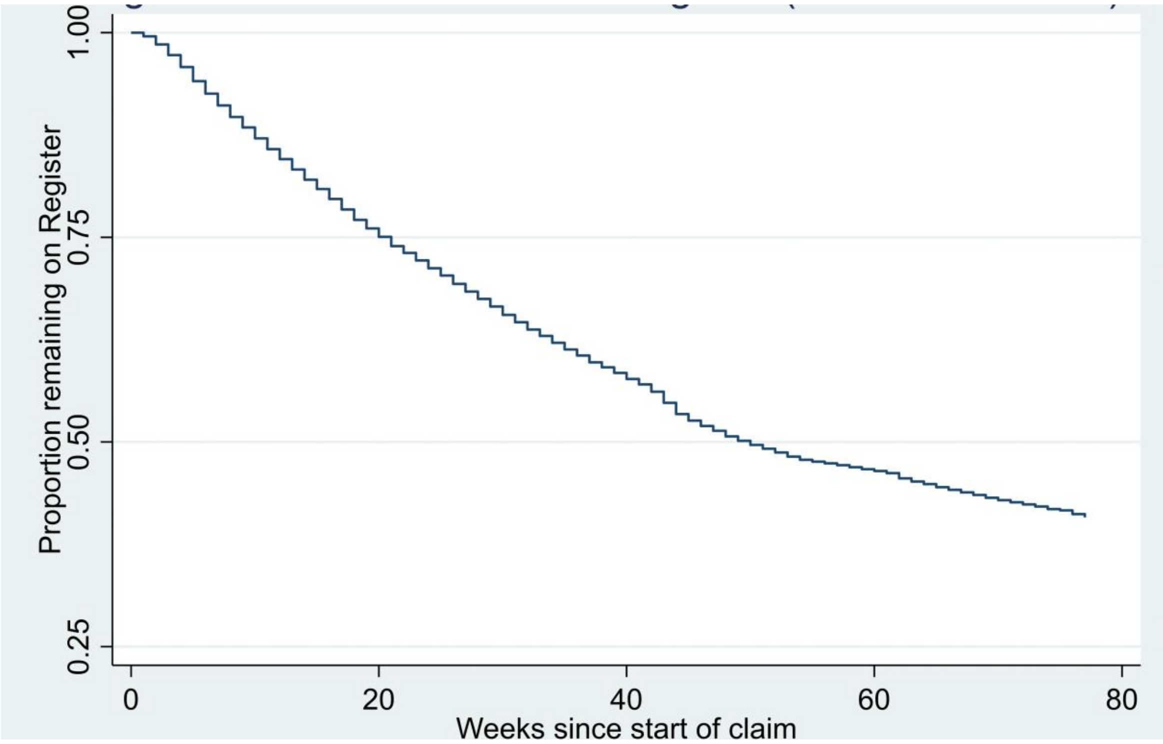


FIGURE 4.2 KAPLAN-MEIER SURVIVAL FUNCTION: EXITS FROM LIVE REGISTER (2006 DATA)



Source: Authors' analysis of DSP provided project data.

We next re-estimated the original PEX model on the new sample and compared these with the results from the previous study on which the current operational PEX model is based. This was undertaken in order to provide us with an insight into the extent to which the current PEX system remains relevant, and how factors influencing the probability of long-term unemployment may have changed over the period 2006 to 2018. As mentioned, the timeframe between the two data points saw a global recession that saw unemployment in Ireland rise to over 15 per cent and the return of net outmigration, particularly among young people.

There are some differences between the two sets of models that should be noted. Firstly, the 2006 male and female models contain location controls (county level), while the new 2018 models do not. Secondly, the 2018 models only contain individuals with active JA or JB claims, whereas the 2006 data included individuals in receipt of unemployment credits.[21] Finally, the 2018 data do not contain information on the number of claims a person has, information that was derived from administrative sources for the 2006 models. As the 2006 models were estimated separately by gender, this process is repeated in the current study, and the results are presented in Table 4.1.

Despite the passing of time and the occurrence of a major global recession, the model coefficients have remained relatively stable over time in terms of sign and significance. During both periods, and for both genders, the probability of exiting the Live Register before 12 months duration declines with age, literacy problems, the presence of children, a previous spell of long-term unemployment and being casually employed. Conversely, for both periods and both genders, the probability of exiting the Live Register prior to 12 months was positively correlated with being in very good health.

Nevertheless, despite having a common sign, some of the magnitudes of the estimated impacts vary across time and gender. For example, the coefficients on age tend to be substantially larger for both males and females in the models based on the most recent data, indicating that older workers were less likely to exit the Live Register prior to 12 months in 2018 compared to 2006. The positive impact of having a third-level education in achieving an exit from the Live Register in 2018 was also larger than was the case in 2006.

There are no consistent differences between both periods with regards to the impacts of previous periods of long-term unemployment or literacy problems. The impact of the presence of children is similar for women across both periods, but the effect is larger for men in 2018 compared to 2006.

For males, recent job duration determines an exit from the Live Register in the 2018 model despite being largely irrelevant in the 2006 model. Conversely, while

[21] These formed the reference category in the 2006 male and female models.

a willingness to move for a job was a statistically significant factor for males in 2006, it has no predictive power in 2018.

Finally, being located in a village, town or city is much less important, for both males and females, in predicting the probability of exit in 2018 compared to 2006. For females, marital status was an important factor in 2006 but it was not statistically significant in the 2018 time period. While access to one's own transport was not important in the 2006 female model, it was a statistically significant factor in 2018.

Comparing the diagnostic statistics of both models, it is clear that the pseudo R-squared statistics of the 2018 gender models are a number of times greater than those of the 2006 models, indicating that the 2018 models more accurately predict the outcome than the 2006 models.

TABLE 4.1 THE PROBABILITY OF EXITING THE LIVE REGISTER IN 2006 AND 2018 BY GENDER

	Males 2018	Males 2006	Females 2018	Females 2006
Age (Ref: 18–24 years)				
25–34 years	-0.133***	-0.031***	-0.137***	-0.034**
	(0.021)	(0.012)	(0.026)	(0.016)
35–44 years	-0.170***	-0.091***	-0.194***	-0.049***
	(0.024)	(0.014)	(0.030)	(0.018)
45–54 years	-0.245***	-0.110***	-0.212***	0.013
	(0.026)	(0.016)	(0.033)	(0.019)
55+ years	-0.376***	-0.216***	-0.409***	-0.069***
	(0.025)	(0.019)	(0.032)	(0.017)
Health (Ref: Bad/Very bad health)				
Fair health	-0.046	0.019	0.077	0.153***
	(0.054)	(0.040)	(0.078)	(0.047)
Good health	0.028	0.098**	0.152**	0.253***
	(0.050)	(0.038)	(0.074)	(0.042)
Very good health	0.124**	0.128***	0.262***	0.332***
	(0.050)	(0.039)	(0.076)	(0.047)
Marital status (Ref: Single)				
Married	0.116***	0.026**	0.019	-0.072***
	(0.021)	(0.013)	(0.022)	(0.017)
Cohabiting	0.031	-0.020	0.005	-0.000
	(0.026)	(0.032)	(0.030)	(0.037)
Separated/divorced	-0.009	-0.018	-0.024	-0.083***
	(0.037)	(0.026)	(0.035)	(0.032)
Widowed	-	0.043	-0.161	-0.057
	-	(0.053)	(0.162)	(0.041)
Children	-0.102***	-0.030***	-0.044*	-0.060***
	(0.022)	(0.006)	(0.023)	(0.010)
Spousal earnings (Ref: None)				
Spouse earnings €250	0.063	0.057**	-0.065	0.014
	(0.053)	(0.023)	(0.086)	(0.025)

TABLE 4.1 (CONTD.) THE PROBABILITY OF EXITING THE LIVE REGISTER IN 2006 AND 2018 BY GENDER

	Males 2018	Males 2006	Females 2018	Females 2006
Spouse earnings €251–€350	0.211*	0.009	-0.109	-0.032
	(0.113)	(0.044)	(0.112)	(0.084)
Spouse earnings €351+	0.015	0.029*	-0.001	-0.101***
	(0.029)	(0.017)	(0.035)	(0.017)
Education (Ref: Primary or less)				
Lower secondary	0.020	0.002	0.031	0.004
	(0.029)	(0.012)	(0.050)	(0.018)
Upper secondary	0.081***	0.063***	0.102**	0.034*
	(0.029)	(0.012)	(0.045)	(0.018)
Third level	0.192***	0.114***	0.224***	0.125***
	(0.028)	(0.013)	(0.046)	(0.018)
Apprenticeship	0.013	0.037***	-0.050*	-0.015
	(0.020)	(0.010)	(0.029)	(0.018)
Literacy/numeracy problems	-0.080***	-0.066***	-0.108***	-0.061**
	(0.024)	(0.015)	(0.037)	(0.025)
English proficiency	0.054***	-0.034	0.015	0.001
	(0.019)	(0.023)	(0.024)	(0.032)
Employment history (Ref: Never employed)				
Still In employment	-0.053	0.180***	0.102	0.244***
	(0.115)	(0.024)	(0.117)	(0.027)
Employed in last month	0.115	0.149***	0.310***	0.161***
	(0.107)	(0.027)	(0.113)	(0.033)
Employed in last year	-0.001	0.063**	0.220**	0.062*
	(0.109)	(0.026)	(0.108)	(0.033)
Employed in last 5 years	-0.030	0.029	0.137	-0.029
	(0.111)	(0.028)	(0.110)	(0.037)
Employed over 6 years ago	-0.160	-0.014	0.025	-0.136***
	(0.109)	(0.037)	(0.128)	(0.051)
Casually employed	-0.120	-0.094***	-0.185**	-0.160***
	(0.082)	(0.018)	(0.088)	(0.015)
Would move for a job	0.016	0.038***	0.044***	0.082***
	(0.014)	(0.008)	(0.017)	(0.011)
Job duration (Ref: Never employed)				
Less than a month	0.232**	-0.013	-0.017	0.021
	(0.091)	(0.027)	(0.132)	(0.034)
1–6 months	0.260***	0.011	0.084	0.069**
	(0.098)	(0.024)	(0.122)	(0.030)
6–12 months	0.277***	0.015	0.112	0.040
	(0.091)	(0.024)	(0.118)	(0.031)
1–2 years	0.299***	-0.037	0.091	0.041
	(0.086)	(0.026)	(0.120)	(0.031)
2+ years	0.251**	-0.065***	0.036	0.020
	(0.100)	(0.024)	(0.127)	(0.031)
UE claim previous 5 years	-0.100***	0.044***	-0.056***	0.126***
	(0.016)	(0.009)	(0.018)	(0.010)

TABLE 4.1 (CONTD.) THE PROBABILITY OF EXITING THE LIVE REGISTER IN 2006 AND 2018 BY GENDER

	Males 2018	Males 2006	Females 2018	Females 2006
Signing for 12 months+	-0.131***	-0.166***	-0.148***	-0.188***
	(0.018)	(0.012)	(0.025)	(0.016)
CE scheme previous 5 years	-0.098***	-0.070***	-0.095***	-0.074**
	(0.022)	(0.027)	(0.032)	(0.037)
On CE scheme for 12 months+	-0.047	-0.071**	0.004	-0.145***
	(0.033)	(0.035)	(0.044)	(0.044)
Jobseeker Payment (Ref: JB in 2018 and UE Credits in 2006)				
Jobseeker's Allowance	-0.394***	0.014	-0.387***	-0.115***
	(0.013)	(0.028)	(0.016)	(0.026)
Jobseeker's Benefit		0.194***		0.093***
		(0.027)		(0.024)
Number of claims[1]	-	-0.085	-	-0.332***
	-	(0.053)	-	(0.037)
Location (Ref: Rural)				
Village	-0.025	-0.035**	0.024	-0.024**
	(0.025)	(0.015)	(0.028)	(0.016)
Town	-0.026	-0.040***	-0.016	0.006
	(0.022)	(0.014)	(0.026)	(0.015)
City	0.019	-0.055***	0.053**	0.003
	(0.021)	(0.014)	(0.025)	(0.015)
Own transport	0.126***	0.058***	0.093***	0.015
	(0.015)	(0.009)	(0.018)	(0.011)
Near public transport	-0.005	0.019*	0.051**	-0.030**
	(0.019)	(0.011)	(0.022)	(0.012)
Observations	8,034	17,738	5,637	13,024
Pseudo R2	0.3145	0.1150	0.3312	0.1394

Notea: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
[1] Variable not available in the 2018 data.

In Table 4.2, we measure the accuracy of the 2018 model in identifying Live Register leavers and stayers 12 months following the initial claim. Using the results from Table 4.1, for each individual we can predict the probability that they leave the Live Register to employment, based on their characteristics. To evaluate the accuracy of the model, we begin with a 50 per cent cut-off point. This means that individuals with a predicted probability of exit above 50 per cent are classified as leavers, while those with a predicted probability of less than 50 per cent are classified as stayers. We can then compare the actual outcomes (leaver or stayer) with the predicted outcomes of the model (leaver or stayer). If actual outcomes match the predicted outcomes, the model is deemed to be an accurate predictor of Live Register exits. At the 50 per cent cut-off point, the model performs well, as it correctly predicts the actual outcomes of over 75 per cent of all cases. The prediction rates for leavers (78.7 per cent) were slightly higher than those for stayers (75.1 per cent).

It is also worth noting that the overall accuracy level of the 2018 male model (76.9 per cent) is considerably higher than that of the comparable 2006 model for males (69.2 per cent). A similar pattern emerges when we look at the female model: the overall model performance is somewhat better than that for males, with almost 78 per cent of outcomes correctly identified. Furthermore, as was the case for males, the 2018 female model appears to substantially outperform the comparable model based on the 2006 data.

TABLE 4.2 RELIABILITY TESTS: 2018 AND 2006 MODELS BY GENDER

	Males 2018	Males 2006	Females 2018	Females 2006
Correctly predicted	6,181	12,282	4,378	9,088
Total	8,034	17,738	5,637	13,034
Percentage	76.94%	69.24%	77.67%	69.72%
Stayers				
Correctly predicted (%)	75.15%	65.4%	74.93%	66.4%
Leavers				
Correctly predicted (%)	78.67%	79.6%	79.49%	71.1%

Given the similarity between the 2018 male and female models, it seems that a more practical approach would be to estimate a pooled model that contains a coefficient that will capture gender-based difference in the rate of exit from the Live Register.[22] The results from the pooled model are reported in Table 4.3 and are consistent with the gender-specific model. With regards to gender itself, females were found to be 2.4 percentage points less likely than males to exit the Live Register in the 12 months following the initiation of their unemployment claim. The probability of exit was negatively correlated with age, literacy/numeracy difficulties, casual employment, the presence of children, a history of previous long-term unemployment and receipt of JA. The probability of exit was found to be positively correlated with education, very good health, recent employment tenure, recent employment status, married status, willingness to move for a job and possessing one's own transport.

[22] Decision taken after consultation with the DSP.

The model diagnostics are again favourable and, in line with the gender-specific models, generate a pseudo R-squared statistic of 0.3182 and correctly predict 77 per cent of outcomes.

TABLE 4.3 POOLED PROBIT FOR EXITING THE LIVE REGISTER PRIOR TO 12 MONTHS BASED ON 2018 DATA

	All
Female	-0.024**
	(0.011)
Age (Ref: 18-24 years)	
25–34 years	-0.138***
	(0.016)
35–44 years	-0.183***
	(0.019)
45–54 years	-0.229***
	(0.020)
55+ years	-0.385***
	(0.020)
Health (Ref: Bad/very bad health)	
Fair health	-0.012
	(0.045)
Good health	0.062
	(0.042)
Very good health	0.161***
	(0.042)
Marital status (Ref: Single)	
Married	0.068***
	(0.015)
Cohabiting	0.018
	(0.020)
Separated/divorced	-0.018
	(0.025)
Widowed	-0.166
	(0.157)
Children	-0.070***
	(0.016)
Spousal earnings (Ref: None)	
Spouse earnings €250	0.033
	(0.045)
Spouse earnings €251-€350	0.015
	(0.086)
Spouse earnings €351+	0.016
	(0.022)

TABLE 4.3 (CONTD.) POOLED PROBIT FOR EXITING THE LIVE REGISTER PRIOR TO 12 MONTHS BASED ON 2018 DATA

	All
Education (Ref: Primary or less)	
Lower secondary	0.019
	(0.025)
Upper secondary	0.079***
	(0.024)
Third level	0.196***
	(0.024)
Apprenticeship	-0.001
	(0.016)
Literacy/numeracy problems	-0.087***
	(0.020)
English proficiency	0.039***
	(0.015)
Employment history (Ref: Never employed)	
Still in employment	0.020
	(0.085)
Employed in last month	0.206***
	(0.079)
Employed in last year	0.103
	(0.079)
Employed in last 5 years	0.048
	(0.081)
Employed over 6 years ago	-0.082
	(0.086)
Casually employed	-0.149**
	(0.060)
Would move for a job	0.028**
	(0.011)
Job duration (Ref: Never employed)	
Less than month	0.132*
	(0.077)
1–6 months	0.181**
	(0.076)
6–12 months	0.203***
	(0.073)
1–2 years	0.211***
	(0.071)
2+ years	0.158**
	(0.079)
UE claim previous 5 years	-0.079***
	(0.012)
Signing for 12 months+	-0.138***
	(0.015)
CE scheme previous 5 years	-0.101***
	(0.018)
On CE scheme for 12 months+	-0.024
	(0.026)
Jobseeker’s Allowance	-0.389***
	(0.010)

TABLE 4.3 (CONTD.) POOLED PROBIT FOR EXITING THE LIVE REGISTER PRIOR TO 12 MONTHS BASED ON 2018 DATA

	All
Location (Ref: Rural)	
Village	-0.005
	(0.019)
Town	-0.024
	(0.017)
City	0.031*
	(0.016)
Own transport	0.115***
	(0.011)
Near public transport	0.018
	(0.014)
Observations	13,671
Pseudo R2	0.3186
Reliability tests	
% correctly predicted (%, 50% cut-off)	77.35
% stayers correctly predicted (%, 50% cut-off)	75.34
% leavers correctly predicted (%, 50% cut-off)	79.00

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

While the existing profiling questions are relevant and generate a PEX model with good predictive strength, a secondary question of this research is to examine the key covariates driving long-term unemployment risk and, in so doing, identify the extent to which we can retain the PEX model's accuracy with fewer variables.

In terms of our empirical strategy adopted to address this issue, we begin by estimating the determinants of exiting the Live Register using a standard probit model, beginning with the specification used in Table 4.3, which includes a range of explanatory variables that can be broadly categorised into the following seven areas:

1. Gender and age;
2. Health;
3. Marital status, the presence of children and spousal income;
4. Human capital (education, training, literacy/numeracy, English proficiency);
5. Employment history, previous job duration and casual employment;
6. History of long-term unemployment; and
7. Location (town, village, city), access to public transport and willingness to move for a job.

In order to address this secondary research question, we use a decomposition of goodness of fit measure, developed by Huettner and Sunder (2012). Specifically, our initial model for decomposition can be written as a linear probability model, as in Equation 1:

$$Exit_i = \alpha + \beta X_i + \varepsilon_i \tag{1}$$

There are a number of variables grouped under each of the seven key categories detailed in Equation 1. The decomposition approach assigns an explanatory share to each group by examining changes in the marginal contributions to the R-squared statistic under every possible permutation of the seven sets of variables used. Specifically, their contribution to the R-squared in the final model of each category of variables is calculated using Equation 2:

$$share\left(x_j\right) = \frac{1}{k} \sum_{All\ permutations} \Delta MC \tag{2}$$

where K is the number of variable groupings (seven in the case of Equation 1), and $\sum \Delta MC$ is the sum of the marginal contribution to the R-squared statistic of the group of variables across all possible permutations that can be modelled with respect to the outcome variable. It should be noted that in order to ensure that the estimated marginal effects relating to specific model controls are between 0 and 100 per cent, the relationship between unemployment exits and the various groups of variables outlined in Equation 1 is estimated using a probit, rather than a linear probability model. However, the decomposition of the R-squared statistic is estimated using the linear probability model. Nevertheless, this is acceptable as we are not interested in the individual coefficients of the model but merely in the decomposition of its explained variance.

The results of our decomposition analysis are detailed in Table 4.4. Specifically, the table details the questions that account for over 3 per cent of the R-squared statistic. Consequently, we have retained just eight questions that, between them, account for 85 per cent of the variance in the data explained by the full model reported in Table 4.3.

The most important variable in determining an exit from the Live Register is having a claim activated for JA (as opposed to JB), which alone accounts for over one-third of the explained variance. The large negative impact is likely reflecting the fact that, compared to JB, claimants of JA will have much less recent experience of sustained employment. As mentioned previously, JB is based on social insurance (PRSI) contributions, whereas JA is a means-tested payment and will apply primarily to individuals who did not accumulate the required number of PRSI contributions. Related to this, recent employment history and, in particular, being employed in the month prior to the current claim, accounts for 13 per cent of the explained variance. Education accounts for over 10 per cent of the explained component, with the possession of a third-level qualification representing the most important single credential correlated with an exit from the Live Register. Self-perceived health, particularly reporting very good health, having a history of long-term unemployment (or having previously been on a CE scheme), having access to own transport and age are also retained as important factors.

TABLE 4.4 DECOMPOSITION ANALYSIS ON POOLED MODEL (MALES AND FEMALES)

Variable	R-squared
Jobseeker's Allowance (JA)	34.6
Employment history (Ref: Never employed)	
Still in employment	7.1
Employed in last month	0.4
Employed in last year	1.9
Employed in last 5 years	0.9
Employed over 6 years ago	2.9
Total employment history	**13.2**
Education (Ref: Primary or less)	
Third level	7.1
Upper secondary	1.3
Lower secondary	2.1
Total education	**10.5**
Own transport	5.2
Signing 12+ months	5.2
Health (Ref: Bad/very bad health)	
Very good health	4.1
Good health	1.7
Fair health	1.6
Total health	**7.4**
CE scheme last five years	3.5
Age (Ref: 18-24 Years)	
55+ years	3.1
45–54 years	0.7
35–44 years	0.4
25–34 years	0.5
Total age	**4.7**

Notes: Table includes gender and variables with individual R-squared statistics above 3 (8 questions). Where a dummy variable that is part of a group of variables has an R-squared statistic above 3, all of the other categories are also included, as they form part of the same question. For example, 'employed in last month' has an R-squared value of 7 and is part of a question about employment status that includes other variables that may not have as high an R-squared statistic. Therefore, each block of variables in Table 4.4 above is one question.

We re-estimate the new PEX model based on the eight questions identified through the decomposition approach and present the results in Table 4.5, along with the model reliability tests.

All of the coefficients have the expected sign and are statistically significant. The coefficients are somewhat larger, on average, than those based on the full model, which is to be expected. Relative to individuals in receipt of JB, claimants receiving JA are 40 percentage points less likely to exit the Live Register prior to 12 months. Individuals employed in the last month (year) are 38 (26) percentage points more likely to exit the Live Register within 12 months compared to claimants who have never been employed. Claimants with third-level (upper secondary) qualifications are 24 (11) percentage points more likely to exit than those with no qualifications. Having been previously long-term unemployed, or a CE scheme participant, reduces the probability of exit by more than 10 percentage points, while those aged 55+ (45–54) are 37 (22) percentage points more likely to fall into long-term unemployment relative to claimants aged 18–24. Finally, claimants reporting very good health are 22 percentage points more likely to leave the Live Register, while the impact for those with their own transport is 11 percentage points. With respect to the model performance metrics (the reliability statistics), these are very similar to that found for the model based on the full data, suggesting that the reduction in the number of model variables has not come at the expense of model accuracy.

TABLE 4.5 REDUCED PEX MODEL (8 QUESTIONS)

	All
Jobseeker's Allowance	-0.404***
	(0.009)
Employment history (Ref: Never employed)	
Still in employment	0.157***
	(0.029)
Employed in last month	0.367***
	(0.019)
Employed in last year	0.261***
	(0.019)
Employed in last 5 years	0.202***
	(0.021)
Employed over 6 years ago	0.080***
	(0.030)
Education (Ref: Primary or less)	
Lower secondary	0.045*
	(0.024)
Upper secondary	0.108***
	(0.023)
Third level	0.239***
	(0.022)
Own transport	0.110***
	(0.011)
Signing for 12 months+	-0.166***
	(0.014)
Health (Ref: Bad/very bad health)	
Fair health	-0.002
	(0.045)
Good health	0.075*
	(0.041)
Very good health	0.188***
	(0.041)
CES previous 5 years	-0.123***
	(0.015)
Age (Ref: 18–24 years)	
25–34 years	-0.142***
35–44 years	-0.187***
45–54 years	-0.224***
55+ years	-0.370***
Observations	13,671
Reliability tests	
% stayers correctly predicted (50% cut-off)	75.47%

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

In recalibrating the PEX model using 2018 data, we applied a number of restrictions to the data sample in order to be able to more accurately identify leavers and stayers. In particular, we dropped cases where claimants exited for reasons other than employment (e.g., to education/training or to other benefits), as well as those claimants who exited and re-entered the Live Register. These restrictions are slightly different to those applied when the original PEX model was constructed using 2006 data (O'Connell et al., 2009). For example, in developing the original PEX model, individuals who moved to other benefits were classified as stayers, and those that exited the Live Register for at least six weeks before re-entering the Live Register were classified as leavers. Given these differences between the original and recalibrated PEX models, it is useful, as a robustness check, to check that: i) our restrictions do not produce coefficients that are grossly inconsistent with our baseline sample; and ii) our restrictions serve to improve the PEX model's accuracy.

In Table A.2, we compare our sample restricted pooled model with one where excluded cases related to non-employment related exits and re-entrants are added back into the data. We can see that the sample size in the unrestricted model increases by a factor of around 2.5. Despite this, the models are similar in terms of the sign and significance of the model coefficients, which indicates that our restricted sample is reflective of the overall population of claimants. There are, however, three exceptions to this finding. These are: i) 'still in employment' (restricted model: 0.020 (NS); unrestricted model: -0.122***); ii) 'casually employed' (restricted model: -0.149**; unrestricted model: -0.019 (NS)); and iii) 'unemployment claim in previous 5 year' (restricted model: -0.079***; unrestricted model: 0.032***). Furthermore, the pseudo R-squared statistics for the restricted model are more than double that of the unrestricted model, confirming that our sample restrictions more accurately distinguished leavers from stayers in the data.

CHAPTER 5

Summary and conclusions

The objective of this study was to recalibrate the Probability of Exit (PEX) model underlying Ireland's activation system. The PEX model uses responses to a series of questions asked of each unemployment assistance claimant the day they first make their claim to give an estimated probability of them exiting the Live Register within 12 months. The current PEX model is based on coefficients calculated using 2006 data. Changes that have taken place in the Irish labour market since that period, including recovery from a global recession and a banking crisis, made it necessary to re-examine the model to identify whether it needed to be updated. A second objective of the study was to identify the key drivers of long-term unemployment risk while, at the same time, retaining maximum model accuracy. From a policy perspective, a supplementary outcome of the research is that it has helped to enhance our understanding of how the risk factors associated with falling into long-term unemployment have evolved over time.

The data for our study was provided to us by the Department of Social Protection (DSP), with the initial sample including all individuals who initiated an unemployment claim between 1 August 2018 and 31 October 2018; in the study, we refer to this as the *sample period*. As the aim of our study is to evaluate the determinants of long-term unemployment, we have information on how these claims progressed over a *follow-on period*, which extends to 1 February 2020 (approximately 16 to 18 months). From an initial sample of 36,076 individuals, after exclusions, our final sample consisted of 13,671 jobseekers. Of these, 6,230 stayed on the Live Register for a period of at least 12 months (*stayers)* and 7,441 exited the Live Register before 12 months (*leavers).*

In terms of descriptive statistics, there were some expected observable differences between leavers and stayers: leavers were more likely to be younger, report better health, be educated to third-level and have more developed labour market histories. Our Kaplan-Meier (KM) survival analysis, which plots the trajectory of exits from unemployment, showed that initially (for approximately the first 15 weeks) the rate of unemployment exit was quite high. After that time point, the curve began to flatten, indicating that the rate of exit to employment slowed down considerably. From the initiation of the claim (0 weeks) up to 40 weeks, the probability of being on the Live Register declined from 100 per cent to 50 per cent. However, comparing 40 weeks to 80 weeks following claim initiation, we saw that the probability of being on the Live Register declined from 50 per cent to just 45 per cent.

The KM function based on the 2018 data was similar to that estimated using the 2006 data that was used to develop the initial, and currently operational, PEX model. The similarity between both KM analyses is notable despite the fact that 11 years separate the two, a period that saw a major global recession and financial crisis.

Following our descriptive examination, we estimated probit models to identify characteristics that are important for determining long-term unemployment risk in 2018, and compared these with the results from the original PEX model estimated using 2006 data. Despite the passing of time and the occurrence of a major global recession, the model coefficients in the 2006 and 2018 models have remained relatively stable over time in terms of sign and significance. In both 2006 and 2018, and for both genders, the probability of exiting the Live Register before 12 months declines with age, literacy problems, the presence of children, a previous spell of long-term unemployment and being casually employed. Conversely, for both periods and both genders, the probability of exiting the Live Register prior to 12 months was positively correlated with being in very good health. Nevertheless, despite having a common sign, the magnitude of some of the estimated impacts vary across time (2006 and 2018) and gender. For example, the coefficients on age tend to be substantially larger for both males and females in the models based on the most recent data, indicating that in 2018 older workers were less likely to exit the Live Register prior to 12 months than they were in 2006.

Some factors differed in statistical significance between the 2006 and 2018 models. For males, recent job duration determines an exit from the Live Register in the 2018 model despite being largely irrelevant in the 2006 model. While a willingness to move for a job was a statistically significant factor for males in 2006, it has no predictive power in 2018. For females, marital status was an important factor in 2006 but by 2018 was no longer relevant. Conversely, access to one's own transport was not important in the 2006 female model, but was a statistically significant factor in 2018. Finally, being located in a village, town or city is much less important, for both males and females, in predicting the probability of exit in 2018 compared to 2006.

Comparing the diagnostic statistics of both the 2006 and 2018 models, it is clear that the pseudo R-squared statistics, and our statistical tests of model reliability, indicate that the 2018 models more accurately predict Live Register exit than the 2006 models.

The decomposition analysis that we undertook to identify the key drivers of long-term unemployment risk resulted in the retention of just eight profiling questions that, between them, account for 85 per cent of the variance in the data explained by the full model. The variables retained in order of importance were: i) claiming

JA; ii) recent employment history; iii) education; iv) self-perceived health; v) a history of long-term unemployment; vi) having previously been on the CE scheme; vii) having access to own transport; and viii) age. The performance of the PEX model based on just eight questions is very similar to that found for the model based on the full data, suggesting that the reduction in the number of model variables does not come at the expense of model accuracy.

In terms of policy, the decomposition analysis for identifying the key factors associated with long-term unemployment risk indicates that four of the eight questions relate to claimants having a relatively low level of labour market attachment in terms of their recent employment history. This suggests that activation measures that allow claimants to rebuild a labour market attachment through, for example, combinations of skills training and work placement, are likely to be effective in preventing jobseekers from becoming long-term unemployed.

REFERENCES

Allhutter, D., F. Cech, F. Fischer, G. Grill and A. Mager. (2020). 'Algorithmic profiling of job seekers in Austria: How austerity politics are made effective', *Frontiers in Big Data*, Vol. 3, No. 5, pp. 1-17.

Department of Education, Skills and Employment (2021). *Online job seeker classification instrument trial evaluation report*, Department of Education, Skills and Employment (Australia).

Desiere, S. and L. Struyvan. (2020). 'Using artificial intelligence to classify jobseekers: The Accuracy-Equity Trade-Off', *Journal of Social Policy*, Vol. 50, pp. 367-385.

Desiere, S., K. Langenbucher and L. Struyven. (2019). *Statistical profiling in public employment services: An international comparison*', OECD Social, Employment and Migration Working Papers No. 224, OECD Publishing, Paris, https://doi.org/10.1787/b5e5f16e-en.

Hasluck, C. (2008). 'The use of statistical profiling for targeting employment services: The international experience', in G. Di Domenico and S. Spattini (eds.), *New European approaches to long-term unemployment: What role for public employment services and what market for private stakeholders?*, Kluwer Law International, the Netherlands.

Loxha, A. and M. Matteo (2014). *Profiling the unemployed – A review of OECD experiences and implications for emerging economies*, World Bank Group Social Protection & Labor Discussion Paper No. 1424.

O'Connell, P.J., S. McGuinness, E. Kelly and J. Walsh (2009). *National profiling of the unemployed in Ireland*, ESRI Research Series Number 10, Economic and Social Research Institute, Dublin.

OECD (1998). *Early identification of jobseekers at risk of long-term unemployment: The role of profiling*, Organisation for Economic Cooperation and Development, Paris.

OECD (2019). *Evaluating Latvia's active labour market policies,* Organisation for Economic Cooperation and Development, Paris, https://doi.org/10.1787/6037200a-en.

Sullivan, W.F., L. Coffey, L. Kolovich, C.W. McGlew, D. Sanford and R. Sullivan. (2007). *Worker profiling and reemployment services evaluation of state worker profiling models, final report*, US Department of Labor, Employment and Training Administration, Office of Workforce Security, Washington DC.

van Landeghem , B., S. Desiere and L. Struyven. (2021). 'Statistical profiling of unemployed jobseekers', *IZA World of Labor*, doi: https://doi.org/10.15185/izawol.483.

Wijnhoven, M. and H. Havinga. (2014). 'The work profiler: A digital instrument for selection and diagnosis of the unemployed', *Local Economy*, Vol. 29, No. 6-7, pp. 740-749.

TABLE A.1 EXCLUDED CASES

Closure code	Closure reason	Treatment	Sample
CL001	UN-NOT SIGNED	Dropped	2,000
CL003	GEN-CLM BENX	Dropped	2,428
CL004	UN-TO CREDITS	Dropped	40
CL005	GEN-MNS NOT DIS	Dropped	60
CL006	GEN-MEAN INXS	Dropped	187
CL007	GEN-DISALLOWED	Dropped	91
CL008	UN-NOT AVAILABL	Dropped	542
CL009	GEN-TO CE SCH	Dropped	1,093
CL010	GEN-TO FAS CRSE	Dropped	1,378
CL011	GEN-TO PTJI	Dropped	9
CL012	GEN-TO RSS	Dropped	51
CL013	GEN - CTRL REVW	Dropped	2
CL014	GB-CWC IB CNTRL	Dropped	1
CL017	GEN-TO PPO	Dropped	1
CL018	GEN-TO RET PEN	Dropped	23
CL019	GEN-TO OAP/OACP	Dropped	421
CL020	GEN-TO INVAL	Dropped	7
CL021	GEN-CLMT ABROAD	Dropped	803
CL022	UN-TRFR UB-E303	Dropped	45
CL023	GEN-CLMT DECEAS	Dropped	30
CL024	GEN-AWD IN ERR	Dropped	8
CL025	GEN-CLM AUTO CL	Dropped	798
CL026	GEN-OTHER	Dropped	1,012
CL026 or CL133	GEN-OTHER or UN-GONE TO TUS	Dropped	1
CL027	GEN-BATCH EXP	Dropped	1,390
CL028	GEN-CLMT 66	Dropped	338
CL029	GEN-TO UN	Dropped	127
CL031	GEN-TO GB	Dropped	562
CL032	UN-NOT QUALIFD	Dropped	7
CL033	UN-B/X 156/U18	Dropped	2
CL034	GEN - OSSIG	Dropped	791
CL036	GB-FIRST/FINAL	Dropped	4
CL041	GB-MANL CLOSURE	Dropped	1
CL049	MAT-UNFT FR WRK	Dropped	3
CL055	GEN-O/S DOC/INF	Dropped	325
CL056	SWA-MORT PAID	Dropped	1
CL059	GEN-STUDENT	Dropped	561
CL060	SWA-ADDRESS	Dropped	19

TABLE A.1 (CONTD.) EXCLUDED CASES

CL061	SWA-INSTITUTION	Dropped	1
CL065	SWA-NEED MET	Dropped	1
CL066	GEN-PS	Dropped	4
CL067	GEN-WDRN/LAPSED	Dropped	59
CL069	SWA-CLIENT TRAN	Dropped	2
CL072	SWA-NEW CLAIM	Dropped	6
CL074	GB-PROPER TO DB	Dropped	166
CL077	GB-PROPR TO OIB	Dropped	2
CL083	GB-GONE TO MAT	Dropped	19
CL085	GEN-TRANS TO DA	Dropped	188
CL086	GEN-TRAN TO MAT	Dropped	11
CL087	GEN-TRAN TO LPA	Dropped	245
CL089	GEN-ADA SPS CLM	Dropped	114
CL090	GEN-IN CUSTODY	Dropped	168
CL095	BTW-COMPL BTW	Dropped	11
CL097	BTW-FRM NOT RTN	Dropped	1
CL098	GEN-LOC UNKNOWN	Dropped	28
CL103	GEN-OTH PENSION	Dropped	32
CL104	BTW-S/EMP > EMP	Dropped	71
CL105	BTW-EMP > S/EMP	Dropped	19
CL107	GEN-TO CARERS	Dropped	234
CL108	GEN-NOT SINGLE	Dropped	1
CL118	FIS-ONLY CLD 18	Dropped	1
CL123	DL MID CLAIM	Dropped	1
CL125	GEN-NOT HAB RES	Dropped	5
CL133	UN-GONE TO TUS	Dropped	333
CL136	GONE TO PCB	Dropped	1
CL142	IB < 7 DAYS	Dropped	2
Total			16,888

TABLE A.2 POOLED MODEL ESTIMATED USING BOTH A RESTRICTED AND UNRESTRICTED SAMPLE

	Restricted sample	Unrestricted sample
Gender		
Female	-0.024**	-0.006
	(0.011)	(0.005)
Age group		
25–34 years	-0.138***	-0.094***
	(0.016)	(0.008)
35–44 years	-0.183***	-0.134***
	(0.019)	(0.010)
45–54 years	-0.229***	-0.174***
	(0.020)	(0.012)
55+ years	-0.385***	-0.239***
	(0.020)	(0.014)
Health status		
Fair health	-0.012	-0.013
	(0.045)	(0.018)
Good health	0.062	0.012
	(0.042)	(0.017)
Very good health	0.161***	0.047***
	(0.042)	(0.017)
Marital status		
Married	0.068***	0.026***
	(0.015)	(0.007)
Cohabiting	0.018	-0.004
	(0.020)	(0.009)
Separated/divorced	-0.018	0.005
	(0.025)	(0.011)
Widowed	-0.166	0.027
	(0.157)	(0.050)
Children	-0.070***	-0.056***
	(0.016)	(0.008)
Spousal earnings		
Spouse earnings €250	0.033	0.037*
	(0.045)	(0.019)
Spouse earnings €251–€350	0.015	-0.024
	(0.086)	(0.044)
Spouse earnings €351+	0.016	0.011
	(0.022)	(0.010)
Education		
Lower secondary	0.019	-0.007
	(0.025)	(0.010)
Upper secondary	0.079***	0.005
	(0.024)	(0.010)
Third level	0.196***	0.041***
	(0.024)	(0.010)
Apprenticeship	-0.001	0.003
	(0.016)	(0.008)
Literacy/numeracy problems	-0.087***	-0.019**
	(0.020)	(0.008)
English proficiency	0.039***	0.017**
	(0.015)	(0.007)

TABLE A.2 POOLED MODEL ESTIMATED USING BOTH A RESTRICTED AND UNRESTRICTED SAMPLE

	Restricted sample	Unrestricted sample
Employment		
Still in employment	0.020	-0.122***
	(0.085)	(0.034)
Employed in last month	0.206***	-0.063**
	(0.079)	(0.025)
Employed in last year	0.103	-0.114***
	(0.079)	(0.028)
Employed in last 5 years	0.048	-0.147***
	(0.081)	(0.033)
Employed over 6 years ago	-0.082	-0.192***
	(0.086)	(0.035)
Casually employed	-0.149**	-0.019
	(0.060)	(0.025)
Would move for a job	0.028**	0.019***
	(0.011)	(0.005)
Job duration less than month	0.132*	0.157***
	(0.077)	(0.013)
Job duration 1–6 months	0.181**	0.202***
	(0.076)	(0.018)
Job duration 6–12 months	0.203***	0.174***
	(0.073)	(0.015)
Job duration 1–2 years	0.211***	0.153***
	(0.071)	(0.015)
Job duration 2+ years	0.158**	0.164***
	(0.079)	(0.020)
UE claim previous 5 years	-0.079***	0.032***
	(0.012)	(0.006)
Signing for 12 months+	-0.138***	-0.068***
	(0.015)	(0.007)
CE scheme previous 5 years	-0.101***	-0.064***
	(0.018)	(0.009)
On CE scheme for 12 months+	-0.024	0.012
	(0.026)	(0.010)
Jobseeker's Allowance	-0.389***	-0.234***
	(0.010)	(0.005)
Location		
Village	-0.005	-0.007
	(0.019)	(0.009)
Town	-0.024	-0.006
	(0.017)	(0.008)
City	0.031*	0.015**
	(0.016)	(0.007)
Transport		
Own transport	0.115***	0.041***
	(0.011)	(0.006)
Near public transport	0.018	0.009
	(0.014)	(0.007)
Pseudo R2	0.3186	0.1511
Observations	13,671	32,906

Whitaker Square,
Sir John Rogerson's Quay,
Dublin 2
Telephone **+353 1 863 2000**
Email **admin@esri.ie**
Web **www.esri.ie**
Twitter **@ESRIDublin**

ESRI ECONOMIC & SOCIAL RESEARCH INSTITUTE