A Regionally Focussed Survival Analysis into Re-employment Rates in New Zealand.

Matthew Steiner

2021

A dissertation submitted to Auckland University of Technology in partial fulfilment of the requirements for the degree of Master of Business (MBus).

Supervisors: Matthew Ryan, Alexander Plum, Gail Pacheco.

Abstract

I apply survival analysis techniques to analyse regional variations in the probability that a jobseeker will transition from unemployment into employment. Controlling for variations in individual characteristics, I aim to identify whether an individual's job search is affected by the New Zealand region in which they reside. Statistics New Zealand's IDI dataset is used to link 2013 census responses to MSD benefit spell and residential address data. Unemployment spells are identified using administrative Unemployment and Jobseeker Support benefit spell records. The IDI provides a rich dataset on individual characteristics and allows for individual benefit spells to be tracked over long periods with precise information on regional assignment and spell dates. MSD data in the IDI allows for the isolation of specific benefit types ensuring that the identification of unemployed individuals is robust. Earlier New Zealand datasets were not capable of achieving this. Unemployment spells that begin during the full four calendar months either side of the 2013 census are tracked until data is unavailable after July 2020. Regional variation is commonly observed in foreign literature; however, the New Zealand literature is scarce. I conclude that regional variation does exist in New Zealand. Jobseekers in the Southland and Canterbury regions are most likely to transition into employment. Contrarily, Wellington and Northland exhibit the lowest transition probabilities. Jobseekers in these regions are on average 10% less likely to find work at any spell duration compared to a jobseeker in Auckland. Findings in this research are robust to various specifications and robustness models.

Contents

Abstract 2
List of Figures 5
List of Tables 5
Acknowledgements 6
Disclaimers 7
Attestation of Authorship 7
Output produced from the IDI: 7
Outputs produced from 2013 Census data: 7
Disclaimer for Inland Revenue tax data: 7
Chapter 1: Introduction 8
Chapter 2: Literature Review 11
2.1 Overview of Unemployment Theory 11
2.2 New Zealand Literature 13
2.3 Global Regional Literature 15
2.4 New Zealand Regional Indicators 18
Chapter 3: Empirical Strategy 20
3.1 Survival Analysis Theory 20
3.2 Cox Proportional Hazards Model 21
3.3 Discrete Time 22
3.4 Censoring Mechanisms 23
3.5 Partial Likelihood Estimation 24
Chapter 4: Data and Sampling 27
4.1 Data Set 27
4.2 Census Data and Covariates 27
4.3 Identifying the Unemployed 31
4.4 Regional Treatments 34
4.5 Validating the Identification of Unemployed Individuals 35
4.6 Verifying Re-employment Transitions 36
Chapter 5: Empirical Results 38
5.1 Initial Model 38
5.2 Primary Model Specification 39
5.2a Regional Effects 39
5.2b Covariate Effects 43
5.3 Proportional Hazard Assumption Tests 47
5.4 Region Specific Models 50
Chapter 6: Robustness Checks 52
6.1 Adjusted Start Day 52

6.2 Sample Period Test 53
6.3 Additional Regional Specifications 55
6.4 Compare Region at End to Start 56
6.5 Macro model 58
Chapter 7: Discussion 61
7.1 Implications 61
7.2 Limitations 63
Chapter 8: Conclusion 66
References 68
Appendices 73
Appendix A: Economic Conditions Graphs 73
Appendix B: New Zealand Regions Map 74
Appendix C: Primary Model Output Table 75
Appendix D: Underlying Variable Counts 78
Appendix E: Region Specific Models Output Tables 79
Appendix F: Aggregated Regions Output Table 81
Appendix G: Underlying Counts for Additional Variable Specifications 82
Appendix H: Robustness Models Output Table 83
Appendix I: Comparing Region at Start and End of Spell Output Table 84
Appendix J: Underlying Regional Mechanism Output Table 85
Appendix K: Spell Length Histogram 86
Appendix L: Covariate Proportional Hazard Assumption Test 87
Appendix M: Regional Survivor Functions 90

List of Figures

Figure 1. Survival Function per Romanian County (Ciucă & Matei, 2011)……16

Figure 2. IDI Table Linkages………………………..……………………..…27

Figure 3. Regional Hazard Functions……………………………………………..42

Figure 4. Regional Proportional Hazards Test…………………………………48

List of Tables

Table 1. Summary of Regional Effects in the Literature……………………...15

Table 2. Regional Macroeconomic Data……………………………………...19

Table 3. How Benefit Abatements are Applied………………………………32

Table 4. Initial Model Regional Hazard Ratio Estimations……….………..38

Table 5. Regional Hazard Ratio Estimations…………………………………40

Table 6. Covariate Effects (1)…………………………………………………44

Table 7. Covariate Effects (2)…………………………………………………45

Table 8. Covariate Effects (3)…………………………………………………47

Table 9. Proportional Hazards Test……………………………………………49

Table 10. Regional Models with a Redefined Start Day………………………..53

Table 11. Sample Period Test…………………………………………………...55

Table 12. Additional Regional Specifications…………………………………..56

Table 13. Comparing Region at the End of a Spell to at the Start..……………..57

Table 14. Macroeconomic Indicator Variable Hazard Ratios………………….. 59

Acknowledgements

I would like to thank my academic supervisors, Matthew Ryan, Alexander Plum, and Gail Pacheco for their outstanding support throughout this entire process. Their experience and guidance has ensured that I have completed my work to a level that is far beyond what I thought I was capable of. The quality of their feedback has allowed me to make significant strides in developing my research, analytical, and academic writing skills. Their support has kept me on track and filled me with the confidence I needed throughout the long and difficult process that has been my first original research project. I am exceptionally grateful for their support during my dissertation. Thank you!

Disclaimers

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signed: Matthew Steiner
21/07/2021
16949385

Output produced from the IDI:

These results are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) which is carefully managed by Stats NZ. For more information about the IDI please visit https://www.stats.govt.nz/integrated-data/.

Outputs produced from 2013 Census data:

Access to the data used in this study was provided by Stats NZ under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the author, not Stats NZ or individual data suppliers.

Disclaimer for Inland Revenue tax data:

The results are based in part on tax data supplied by Inland Revenue to Stats NZ under the Tax Administration Act 1994 for statistical purposes. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Chapter 1: Introduction

This research analyses regional level variations in the probability that an individual will transition from unemployment into employment in New Zealand[1]. This transition event will be referred to as re-employment[2]. Semi-parametric survival analysis techniques are used to model these probability variations. Conclusions can be drawn regarding whether regional differences in re-employment probabilities exist. Global literature has observed such regional variations in other nations. There is a need to collate empirical evidence regarding the existence, magnitude, and significance of such effects in New Zealand.

Statistics New Zealand's Integrated Data Infrastructure (IDI) is used for this analysis. The IDI contains microdata on New Zealand individuals and households held by different government agencies (Statistics NZ, n.d.). Different datasets within the IDI can be linked to create bespoke datasets using precise administrative data. I identify individuals as being unemployed if they are receiving a New Zealand Unemployment or Jobseeker Support benefit (the benefit(s)) [3]. The rich data included in the IDI resolves one of the larger limitations of earlier research. Moore (2006) discusses the inability to distinguish between benefit types; not all benefit types require that a recipient be unemployed. The IDI includes data directly from the Ministry of Social Development (MSD) who provide benefit payments. Specific benefit types can be isolated to ensure that the identification of unemployed people is robust.

Regional level variations in re-employment probabilities are estimated using survival analysis techniques. Individuals begin in one state: unemployed. Duration data is then used to model the probability that an individual will transition into another state: employed (Cameron & Trivedi, 2005). An individual experiences a 'spell' if they are in the unemployed state. They then 'die' or 'fail' when they transition into employment. Modelling predicts the hazard and survivor functions for each New Zealand region. Survivor functions describe the probability that an individual will remain unemployed for longer than a given time (Cameron & Trivedi, 2005). The hazard function states the instantaneous probability that an individual will transition into employment, conditional on them remaining unemployed until that duration[4]. Estimation tables show the hazard ratios; this ratio states the relative risk of transitioning into employment compared to a reference group.

[1] New Zealand is divided into 16 regions as displayed in Appendix B.
[2] This analysis excludes transitions from being out of the labour force into employment.
[3] The Jobseeker Support benefit replaced the Unemployment benefit in July of 2013 (MSD, 2013).
[4] For example, the hazard rate at a spell length of 5 days states the probability that an individual who remains in the unemployed state until day 5 will find employment on day 5 of their unemployment spell.

To develop effective labour market policy, it is important to understand the re-employment probabilities that unemployed individuals face. Greater unemployment reduces economic efficiency and creates a deadweight loss in the economy (Lucas, 1978). Unemployed labour does not contribute to production which reduces the productive capacity of the economy. Greater re-employment probabilities improve an individual labourer's likelihood of returning to work; returning jobseekers to employment will reduce any deadweight loss. Individuals that exhibit lower re-employment probabilities are also at a greater risk of experiencing longer unemployment spells or even long-term unemployment[5]. Individuals who face longer unemployment spells experience reduced well-being, reduced incomes, and tend to earn reduced future incomes (Nichols et al., 2013). Children of the unemployed also tend to perform worse academically. These negative effects extend to local communities (Nichols et al., 2013). Communities where long-term unemployment persists tend to experience increased crime and worse overall wellbeing. These negative effects are more likely to be experienced in regional labour markets where re-employment probabilities are lowest. I aim to diagnose the key factors in determining an unemployment spell's length. Policy responses can address these factors and reduce the likelihood that individuals and communities experience the adverse effects of unemployment.

Supplementing the policy implications, I am further motivated by the scarcity of New Zealand literature. Global literature has identified regional re-employment probability variations, particularly in European labour markets. New Zealand literature has not comprehensively focussed its scope on regional effects creating a gap in the literature. Prior New Zealand studies have limitations in their design including Moore's (2006) dataset's inability to identify specific benefit types. Benefit type and monthly income source data in the IDI precisely identifies unemployed individuals, resolving this limitation of earlier New Zealand research.

I observe regional variation to exist in New Zealand. Southland and Canterbury jobseekers are identified as having the highest re-employment probabilities over the length of an unemployment spell. Re-employment probabilities are, respectively, 32% and 15% greater than for a comparable jobseeker in Auckland. Wellington and Northland jobseekers experience the worst re-employment probabilities. Both regions exhibit re-employment probabilities that are 10% lower relative to someone in Auckland. Beyond regional effects, I observe that male, younger, more educated, fully abled, and married

[5] Long-term unemployment is defined as experiencing unemployment for over 12 months (OECD, 2016).

jobseekers experience greater re-employment probabilities. I also observe that having fewer prior spells on any MSD benefit also increases this probability. I observe that Pacifica experience the greatest re-employment probabilities relative to other ethnicities; Maori and 'Other' ethnicities are not statistically different from NZ Europeans. Asian jobseekers experience the lowest re-employment probabilities. Additional model specifications are used; these verify the robustness of the findings in this analysis. It can be stated with confidence that regional re-employment probability variation does exist in the New Zealand labour market.

Chapter 2: Literature Review

2.1 Overview of Unemployment Theory

Both jobseekers and employers are engaged in a search process (McCall, 1970). Re-employment requires both a job offer and the acceptance of this offer (Oberholzer-Gee, 2008). Employers offer employment if a jobseeker is perceived to be the best fit for the position. Jobseekers accept the offer if they perceive that the compensation[6] which is offered outweighs the opportunity cost of remaining unemployed and continuing their job search[7]. Re-employment occurs when the quality of the labour supply matches the requirements of the labour demand, and when the compensation offered matches the jobseekers' expectations.

The quality of a jobseeker can be characterised by their level of human capital. Human capital is comprised of an individual's skills, experience, and productive knowledge (Melin, 2001). Improved human capital increases productivity which increases an individual's value to an employer. In turn, this should improve one's ability to find work as they should rank higher than other candidates supplying the labour market.

A counter argument does exist that enhanced human capital should increase that individual's expected wage (Melin, 2001). This manifests through the scarcer supply and greater marginal revenue generation of higher skilled workers. Higher wage expectations may price these workers outside of some employer's budgets or negate any additional revenues generated through enhanced productivity. These individuals may also become more selective in their job search as discussed in Tansel & Taşçi (2010). Higher skilled workers are unlikely to compete for the same positions as those with significantly fewer skills. Therefore, a separation of labour supply exists. Greater human capital should result in an individual ranking higher than less skilled workers. But, if separate labour markets are supplied by individuals with similar human capital levels, this may dilute the benefit of having greater human capital[8].

Human capital is rarely fully observable. Employers must use observable characteristics to estimate an applicant's true ability (Kroft et al., 2013; Kauermann & Khomski, 2009; Forbes & Barker, 2017; Oberholzer-Gee, 2008). Some characteristics

---

[6] Compensation in this context encompasses pay, benefits, title, job satisfaction, company fit etc.

[7] Receiving lower income (predominantly from the benefit) in the hope of finding a better employment opportunity in the future.

[8] As an example, a highly skilled medical doctor will most likely only compete with other similarly skilled doctors, not someone with no/lower qualifications. Therefore, being a skilled doctor does not increase their re-employment probability in that labour market; all jobseekers in that labour market are highly skilled medical doctors.

can be informative; work history and educational achievement partially indicate human capital levels. Additional observable characteristics including age, sex, and unemployment duration are used to estimate what remains unobserved. Observable characteristics, however, provide imperfect information about unobservable human capital levels; this results in imperfect hiring decisions.

Of interest to us is employers' use of spell durations to identify unobservable human capital (Kroft et al., 2013; Oberholzer-Gee, 2008; Eriksson & Rooth, 2014; Lockwood, 1991). Heckman & Borjas (1980) add to this discussion by stating that firms also use the number of prior unemployment spells an individual has experienced based on similar intuitions to using spell durations. Literature suggests that costs exist to maintain skills over an unemployment spell; because of these costs employers assume that human capital will depreciate while unemployed (Eriksson & Rooth, 2014; Pavoni & Violante, 2007). Acemoglu (1995) observes that originally identical applicants at shorter unemployment durations deviate in their *ex-post* human capital levels when only one is hired; the rejected candidate experiences an extended spell and further human capital depreciation. Hiring firms also infer the decisions of other firms based on spell durations; this is referred to as rational herding (Oberholzer-Gee, 2008). Employers assume that if an individual had the necessary ability, they would have been hired earlier in their unemployment spell. This thought process is based on matching models in which firms imperfectly test jobseekers' abilities before making hiring decisions (Lockwood, 1991). By rejecting an applicant, firms transmit information to other firms regarding the level of human capital that was observed in their testing process. Because of these two mechanisms, firms rank applicants by spell length to decide between otherwise equivalent applicants (Acemoglu, 1995; Blanchard & Diamond, 1994). Vishwanath (1989) further discusses the effect of stigma against those with longer unemployment spells due to the imperfect nature of information gathering by firms.

As a result of the above factors, re-employment probabilities are unlikely to remain constant over the length of a spell. This property is referred to as duration dependence. Evidence suggests a negative duration dependence in most labour markets; that is, a decreasing hazard function (D'Agostino & Mealli, 2000; Eubanks & Wiczer, 2016; Fernández-Blanco & Preugschat, 2018; Hyslop et al., 2004; Kroft et al., 2013; Eriksson & Rooth, 2014; Oberholzer-Gee, 2008; Acemoglu, 1995). Spell lengths are negatively linked with the probability of transitioning into employment.

Duration dependence can be explained at an individual level and at a population level. 'True' duration dependence describes individual-level effects. Human capital depreciates over the length of a spell as people do not maintain their skills while out of work—the human capital argument (Eubanks & Wiczer, 2016; Eriksson & Rooth, 2014; Pavoni & Violante, 2007; Acemoglu, 1995). Employers also perceive that unobservable human capital depreciates as a spell extends—ranking models (Jarosch & Pilossoph, 2018; Oberholzer-Gee, 2008; Kroft et al., 2013). Interviews are viewed as a costly process, so firms use unemployment durations when offering interviews to reduce the likelihood that they engage in a costly interview for an unsuitable candidate (Jarosch & Pilossoph, 2018).

At the population level, aggregation causes duration dependence even when an individual's human capital is assumed to be fixed and observable (Eubanks & Wiczer, 2016). At shorter spell lengths, the pool of unemployed individuals includes both high and low skilled jobseekers. Those with higher levels of human capital exhibit higher fixed re-employment probabilities; they should transition into employment after shorter durations. At longer spell durations, the pool of unemployed individuals is comprised of predominantly lower skilled individuals. Negative duration dependence manifests because of these discrepancies in the human capital composition of the pool of unemployed individuals at different spell lengths.

These theories of duration dependence solely describe decisions on the demand side; additional theories exist. Upon initial entrance to the labour market, a jobseeker is presented with the full pool of available positions in the market (Coles & Smith, 1998). Once a worker has not found employment with this original pool of available positions, jobseekers are limited by both the quantity and frequency of new job postings. This pool is significantly smaller compared to when first entering the labour market; re-employment probabilities decline as a consequence. Supplementing the flow of additional jobs to the market is the attitudes of jobseekers (Kroft et al., 2013). Extended unemployment spells can discourage jobseekers causing them to reduce the intensity of their job search.

2.2 New Zealand Literature

Similar literature regarding the factors that affect re-employment probabilities in New Zealand is scarce. The most relevant domestic studies are survival analyses conducted by Moore (2004; 2006), and an OLS model by Hyslop et al. (2004). Regional effects are only briefly discussed in these studies. This research aims to fill this gap in the New Zealand literature.

All three studies do identify regional effects. Moore (2004) specifically identifies the negative effects of living north of the Auckland region. Hyslop et al. (2004)

observes that Aucklanders are between 2–10 percent less likely to be re-employed than those living in other regions. They further observe that the Taranaki, Wellington, West Coast/Tasman/Nelson/Marlborough, and Southland regions exhibit the highest re-employment probabilities. Regional effects are observed to be magnified by the duration of a spell. All regions experience re-employment probabilities that are 2–6% greater than in Auckland 3 months after beginning a benefit spell; this difference increases to 5–10% after 18 months (Hyslop et al., 2004).

Negative duration dependence has also been identified in New Zealand. Hyslop et al. (2004) identified that an individual experiencing a spell lasting over 6 months is 5–10% less likely to be off a benefit compared to someone with a 3-month spell. More specifically, individuals with spell durations ranging between 6–12 months experience 4% reductions in job finding rates; this increases to 10–15% for spells that last over one year. International authors, including Jarosch & Pilossoph (2018), Kroft et al. (2013), and Oberholzer-Gee (2008), have also quantified negative duration dependence. These authors observed significantly larger negative duration dependence effects; duration dependence may be less prominent in New Zealand.

While Hyslop et al. (2004) assigns a larger focus to regional effects, Moore's (2004; 2006) use of survival analysis techniques in a New Zealand context is most relevant to this research. Both authors use benefit-to-work transitions as a proxy for re-employment transitions. Moore (2006) discusses how earlier New Zealand datasets could only identify those receiving an income tested benefit. These datasets were incapable of distinguishing between benefit types; not all benefits require that recipients be unemployed[9], producing uncertainty in Moore's research design. Administrative benefit data in the IDI can make this distinction, resolving this data limitation. Moore (2006) also discusses the importance of including additional covariates in future research and specifically recommends ethnicity. Moore only includes age, sex, region, industry, and the month someone becomes unemployed. Additional covariates (including ethnicity) are identified in the literature and are included in this analysis. Specific covariates and the supporting literature are discussed in Chapter 4.

[9] Benefits are available for additional reasons including urgent and unexpected costs, health/disability, child-care, and senior services (WINZ, n.d.). These do not require a recipient to be unemployed.

2.3 Global Regional Literature

It is common in the literature to include a region variable when analysing the factors that affect re-employment probabilities. As Chapter 2.2 outlines, Hyslop et al. (2004) and Moore (2004; 2006) have identified regional effects in New Zealand. Regional effects are, however, only briefly discussed in New Zealand literature; global literature must also be reviewed to gain a more complete understanding of the current literature. Table 1 briefly summarises the key sources.

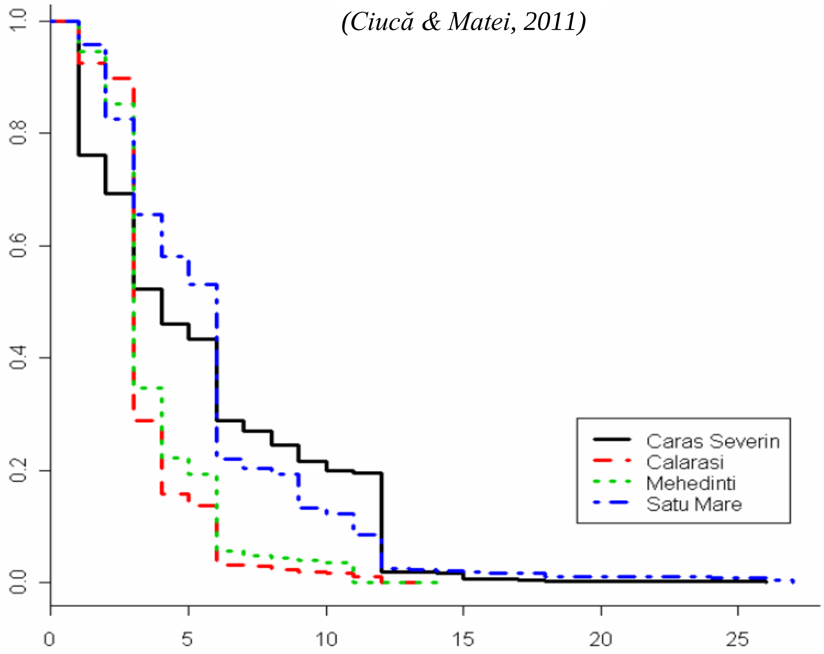
Table 1

Authors	Year	Region	Regional Re-employment Probability Variation
Moore	2004	New Zealand	Probabilities are lowest North of Auckland.
Moore	2006	New Zealand	Regional variation is observed.
Hyslop et al.	2004	New Zealand	Auckland has the lowest probabilities. Taranaki, Wellington, West Coast/Tasman/Nelson/Marlborough, and Southland have the highest probabilities.
D'Agostino & Mealli	2000	EU	Regional variation is observed. Unemployment rates cause such variation.
Imbens & Lynch	2006	USA	Regional variation is observed. Unemployment rates cause such variation.
Babucea et al.	2009	EU	Regional variation is observed that is most prominent in Italy. Some countries experience reduced regional effects.
Tansel & Taşçi	2010	Turkey	Regional variables are mostly insignificant. Only one of seven regions exhibit significance.
Ciucă & Matei	2011	Romania	Regional variation is observed.
Forbes & Barker	2017	Australia	Regional variation is observed. Unemployment rates cause such variation. Individual characteristics are more informative than regional variables.

Imbens & Lynch (2006) estimated regional re-employment probabilities for young people in America by including three broad regions. Regional variation was identified, and the authors found that unemployed youth in the Western region had greater re-employment probabilities than the remaining two American regions. One exception existed for men in the Southern region, suggesting the importance of including additional covariates. The authors also identify that regional economic performance variables are more significant than individual characteristics in determining re-employment probabilities. Region variables in this research intend to capture such economic performance mechanisms; Imbens & Lynch's (2006) finding should manifest as regional variation if such economic factors are also significant in New Zealand.

Ciucă & Matei's (2011) findings support those of Imbens & Lynch. Survivor functions in Figure 1 demonstrate regional variation across different Romanian counties. The survivor function essentially indicates the proportion of the initial sample that remains unemployed after a given duration of unemployment. Survivor functions that begin steeper indicate greater re-employment probabilities at shorter unemployment durations. Variations in the shape of each county's survivor function, combined with the authors' statistical tests, confirm that regional re-employment probability variation does exist in Romania.

Figure 1: **Survival Function per County**



This graph has been taken from Ciucă & Matei (2011) to visually demonstrate the authors findings for regional variation in re-employment probabilities in Romania. This graph plots the survivor functions for four different Romanian counties. This graph is the work of the original authors.

D'Agostino & Mealli (2000) and Babucea et al. (2009) both identify regional effects within their national level studies in Europe. D'Agostino & Mealli (2000) observe that regional effects vary in magnitude in different countries. Regional effects are most pronounced in Italy where those living in southern regions are about half as likely to exit unemployment compared to other Italian regions. Babucea et al. (2009) also observes regional re-employment probability variation in Europe. In Slovenia, regions that are more economically developed exhibit the largest re-employment probabilities. In Croatia, one region is observed to experience re-employment probabilities that are 121% greater than another region. This is similarly observed in Macedonia where two regions are half as likely to transition into employment compared to the region with the highest average re-employment probability.

Many authors including D'Agostino & Mealli (2000) replace region variables with that region's unemployment rate. Regions on their own are simply labels that capture underlying mechanisms affecting re-employment probabilities; of highest interest to us are socio-economic mechanisms. Forbes & Barker (2017), Imbens & Lynch (2006), and D'Agostino & Mealli (2000) all capture regional variations using a region's unemployment rate. Imbens & Lynch (2006) observe that a one percentage point increase in the unemployment rate is associated with a 3% decrease in job finding rates[10]. Some of the authors' models even suggest that this negative effect is as high as 7% for men. Forbes & Barker (2017) similarly observe that a 1 percentage point increase in the unemployment rate reduces an individual's re-employment probability by 4% on average in Australia.

The unemployment rate is not the sole mechanism affecting regional re-employment probabilities. Babucea et al. (2009) discusses how more economically developed regions exhibit improved job-finding rates in Europe. This effect is most prominent in Slovenia. Imbens & Lynch (2006) observe a similar outcome in the United States. It is also identified in New Zealand literature that an individual's industry and occupation before unemployment is significant (Moore, 2006). Industry can be captured at an aggregated regional level to indicate the types of available jobs in each region. Individual occupations are not controlled for in the model so that major industries in a regional labour market are captured by region variables. Moore's identification of industry differences as a mechanism of regional variation is supported by Dixon & Stillman (2008) and Forbes and Barker (2017).

Kroft et al. (2013) further observes that tighter labour markets experience stronger duration dependence. Stronger duration dependence implies that re-employment probabilities decline more quickly as an unemployment spell extends in length; hazard functions are steeper. Market tightness describes the balance of demand and supply in a labour market; tighter markets approach full employment (Brigden & Thomas, 2003). Tighter markets usually exhibit greater re-employment probabilities at shorter unemployment durations (Kroft et al., 2013). Market tightness, however, has little effect on re-employment probabilities at longer unemployment durations (Kroft et al., 2013). This suggests that in tighter labour markets where probabilities are initially high, these probabilities must decrease by a larger magnitude to reach a long-term re-employment probability that is consistent across levels of market tightness. Duration dependence is therefore more pronounced, or stronger, in tighter labour markets.

Regional effects are not always observed to be significant. Tansel & Taşçi (2010) use regional covariates in a gender focussed study of Turkish re-employment rates. Only one of seven regions—South-East Anatolia—exhibits a statistically significant regional effect. Estimates suggest that this effect only exists for males and is positive. South-East Anatolia is one of Turkey's poorer regions. Observing improved re-employment

[10] High collinearity between region variables and that region's unemployment rate makes it difficult to test whether regional variables remain significant after including regional unemployment rates in a model.

probabilities appears to contradict Imbens & Lynch (2006) and Babucea et al.'s (2009) findings that better performing regional economies exhibit improved re-employment probabilities. Tansel & Taşçi (2010) note the importance of unique regional labour market policy mechanisms. The South-Eastern Anatolia Project (GAP) is one of the Middle East's largest ever regional development projects (Bilgen, 2020). This policy has resulted in new construction projects, irrigation improvement, urbanization, and improved literacy courses for children in the region. Tansel & Taşçi's (2010) finding suggests the impact that effective government policy can have on re-employment probabilities: hence the significance of identifying the New Zealand regions that require additional policy intervention.

Forbes & Barker (2017) analyse how local labour market conditions affect unemployment durations in Australia. The authors' key finding is that regional effects are not as significant as other variables. Individual characteristics including age, education, and experience are found to have larger effects on re-employment probabilities. Region variables are therefore only a minor determinant of unemployment durations in Australia. This starkly contradicts the findings of Imbens & Lynch (2006) who identify that these economic effects hold greater explanatory power in the United States.

2.4 New Zealand Regional Indicators

Regions are simply labels and regional re-employment probability variation is caused by underlying mechanisms within that region. Socio-economic mechanisms are one component of this; these mechanisms are the focus of this research. This section briefly discusses how economic indicators varied across New Zealand regions in 2013. The sample for this research is comprised of unemployment spells that begin near to the 2013 census, hence using 2013 data. These indicators can be observed in Table 2.

New Zealand regional economic indicators did exhibit variation in 2013 as seen in Table 2 (MBIE, 2021). Based on the discussions in Chapter 2.3, it can be hypothesised that re-employment probabilities should also vary regionally. The socio-economic mechanism that is most commonly discussed in the literature is the unemployment rate (Forbes & Barker, 2017; Imbens & Lynch, 2006; D'Agostino & Mealli, 2000). Unemployment rates do exhibit variation. Many of the lowest rates exist in the South Island; higher re-employment probabilities and stronger duration dependence can be hypothesised in these regions. South Island unemployment rates are generally under 5% while most North Island rates are greater than 6.5%—two are even greater than 8%.

Regions with higher unemployment rates also tend to experience higher deprivation indexes. The deprivation index is derived using census data to assign a decile rating for individual meshblocks[11]. The index captures income, employment, housing, and the availability of services. Regional indexes are derived from weighted

Table 2

Regional macro data for 2013							
Region	Unemployment Rate	GDP per capita	Primary Industry	Deprivation Index	Level 3 Leavers	Underutilization Rate	Avg Household Income
Auckland	6.8	52.15	Manufacturing	5.1	78	15.7	87.4
Northland	8.5	32.38	Manufacturing	7	52	20.8	64.2
Waikato	5.5	42.84	Manufacturing	5.7	59	13.5	75.4
Bop	7.1	40.78	Manufacturing	6.1	65	16.1	72.3
Gisborne	7.5	33.66	Natural Resources	7.3	61	18	63.8
Hawkes	8.3	39.55	Manufacturing	6.1	72	16	66.8
Taranaki	4.9	76.46	Natural Resources	5.6	56	11	79.8
Manawatu	7.8	37.6	Manufacturing	6.4	59	16.6	64.6
Wellington	6.4	61.88	Professional	4.9	69	14.2	92
Westcoast	4.6	48.06	Natural Resources	5.7	35	15.4	70.3
Canterbury	4.3	49.15	Manufacturing	4.2	66	11.2	84.2
Otago	4.7	44.36	Construction	4.7	63	12.3	75.2
Southland	4.1	50.3	Agriculture	5	54	11	75.2
Tasman	4.6	32.52	Manufacturing	4.7	46	10.5	70.3
Nelson	5.3	47.04	Manufacturing	5.3	87	12.4	71.4
Marlborough	4.6	48.8	Manufacturing	5.1	49	11.3	71.3
Based on data from MBIE's Regional Economic Activity Web Tool							

averages of meshblocks contained within that region (MBIE, 2021). Most regions fall in the 4.7–5.3 range; however, several demonstrate much greater deprivation. GDP per capita and average household incomes also exhibit variation. The highest average household incomes are in the three largest regions by population: Auckland, Canterbury, and Wellington. Regarding GDP per capita, the Taranaki region performs the best. Some regions exhibit GDP per capita values that are twice as high as in others. Average household incomes range from $63,800 in Gisborne to $92,000 in Wellington. Based on the discussions of Babucea et al. (2009) and Imbens & Lynch (2006), regions with higher average household income, higher GDP per capita, and lower deprivation indexes should exhibit greater re-employment probabilities.

[11] Meshblocks are Statistics NZ's smallest geographical areas containing in the range of 60–110 people (Atkinson et al., 2014).

Chapter 3: Empirical Strategy

3.1 Survival Analysis Theory

Survival analysis techniques model transition events using duration data (Cameron & Trivedi, 2005). Originating in biostatistics, the technique later evolved to test equipment life expectancies during World War II; in modern times the method has social science applications (Mukhopadhyay & Singh, 2011). Terminology in survival analysis is reminiscent of this biostatistical origin. Medical trials would often test patient durations until death, relapse, or adverse effect after exposure to some form of treatment. In the present scenario, the outcome of interest is a positive event—exiting unemployment—but the terminology reflects the negative connotations of the model's origin. Survival analysis methods classify individuals into different states; in this case unemployed and employed. An individual experiences a 'spell' if they are in the unemployed state. Spell duration (or length) specifies the length of time that an individual is unemployed. Spell start and end dates determine these durations. Death or failure describes the transition event into employment from the unemployed state.

An individual's unemployment spell duration is given by the random continuous variable *T* (Cameron & Trivedi, 2005). Lower case *t* refers to a particular value for the random variable *T*. Unemployment data is usually not continuous and is instead captured in discrete intervals. Chapter 3.3 discusses how discrete time data is addressed in the model. The cumulative distribution function is defined as:

$$F(t) = Pr[T \leq t] \tag{1}$$

This function states the probability that an individual's unemployment spell length *T* will be at most a given duration *t*. The density function is the first derivative of the cumulative distribution function with respect to *t*:

$$f(t) = \frac{dF(t)}{dt} \tag{2}$$

Research into re-employment probabilities is most concerned with those who remain unemployed. As such, it is more useful to consider instead the probability that T is greater than a given value for *t*. This function is referred to as the *survivor function*:

$$S(t) = \Pr[T > t] = 1 - F(t) \tag{3}$$

Survivor functions are monotonically decreasing functions from 1 to 0. This is due to the function being a decumulative distribution function. It is assumed that $S(\infty) = 0$ because, if given infinite time, all jobseekers will transition into work.

While the survivor function plots the probabilities of people remaining unemployed, the hazard function represents transition events (Cameron & Trivedi, 2005). This function plots the instantaneous probability that an individual will transition at a time *t,* conditional on surviving (i.e., remaining unemployed) to that time *t*. Hazard functions may be better understood with a numerical example. If an individual has remained unemployed for 10 days, the hazard function at *t=10* states the probability that they will find employment on that 10th day. This function is defined as:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr[t \leq T < t + \Delta t | T \geq t]}{\Delta t} = \frac{f(t)}{S(t)} \tag{4}$$

The function can also be rewritten to express the hazard function as the rate of change of the log survivor function:

$$\lambda(t) = -\frac{d\ ln\big(S(t)\big)}{dt} \tag{5}$$

By integrating $\lambda(t)$ and using the condition that $S(0) = 1$, the survivor function can be derived in terms of the hazard rate:

$$S(t) = \exp\left(-\int_0^t \lambda(u)\ du\right) \tag{6}$$

3.2 Cox Proportional Hazards Model

Survival analysis functions can be modelled in numerous ways. Parametric models allow for regressors but require that the distribution of survival times be specified accurately (Cameron & Trivedi, 2005; Columbia Public Health, 2021). Non-parametric models do not require any such specification, but do not allow regressors to be included. This analysis uses the semi-parametric Cox proportional hazards model. Using a semi-parametric model allows regressors to be included without needing to fully specify the distribution of failure times (Cameron & Trivedi, 2005). The Cox model is common in the literature and is used by Ciucă & Matei (2011), D'Agostino & Mealli (2000), Moore (2006), and Tansel & Taşçi (2010). Cameron & Trivedi (2005) also detail the commonness of the Cox model in similar studies.

The Cox model is a proportional hazards model, so it requires the proportional hazard assumption. It is assumed that covariates transpose the entire hazard function so that the log of the hazard functions should be parallel. In other words, covariates only vertically shift the log hazard functions; they do not affect the slope of the curves. At all durations, the deviation of the log hazard from a reference group that a certain covariate value induces is constant.

Under a Cox proportional hazards model, the hazard function is expressed as (Cameron & Trivedi, 2005):

$$\lambda(t|\boldsymbol{x},\boldsymbol{\beta}) = \lambda_0(t)\phi(\boldsymbol{x},\boldsymbol{\beta}) \tag{7}$$

The function $\phi(\boldsymbol{x},\boldsymbol{\beta})$ is generally specified as:

$$\phi(\boldsymbol{x},\boldsymbol{\beta}) = exp(\boldsymbol{x}'\boldsymbol{\beta}) = exp(\beta_1 x_1 + \cdots + \beta_k x_k) \tag{8}$$

The hazard function in equation (7) is a function of the non-parametric baseline hazard function $\lambda_0(t)$, and a parametric function of regressors and coefficients $\phi(x,\beta)$[12]. The baseline hazard is simply the hazard function if all regressors are equal to zero. It can be interpreted in a similar manner to an intercept term in linear regression models (Charan, 2020). The vector $\boldsymbol{\beta}$ in the model represents the transformation to the hazard function from the baseline function caused by the vector of covariates $\boldsymbol{x}$. The proportional hazards assumption ensures that each covariate's β value remains constant at any time *t*. The letter *k* represents the number of covariates that are included in the model.

The Cox model survivor function is (Cameron & Trivedi, 2005):

$$S(t|\boldsymbol{x},\boldsymbol{\beta}) = S_0(t)^{\phi(\boldsymbol{x},\boldsymbol{\beta})} \tag{9}$$

The function $\phi(x,\beta)$ has the same specification as for the hazard function. In this case, $S_0(t)$ is the baseline survivor function. The baseline survivor function is the survivor function if all covariate values are equal to 0.

3.3 Discrete Time

While some survival analysis models can be adjusted for discrete time, the Cox model assumes that time is continuous. Unemployment spell data cannot be captured continuously. An individual's labour market state is captured in discrete intervals based on the regularity of data capture. For the purposes of this study, time is measured in discrete daily intervals that are derived from the start and end dates of an individual's unemployment spell.

Because the Cox model assumes continuous time, it does not allow for failure ties (StataCorp, 2019). The model compares an individual who fails to a pool of individuals that are still at risk (the risk pool), which is done by ranking individuals by failure time. The problem arises of how the model should rank individuals with equal failure times.

[12] The model is semi-parametric as it contains both a non-parametric and a parametric component.

The size of the risk pool should remain constant for all those who fail at the same time. When using discrete time, if the Cox model forces a ranking this results in the failure pool not being equal for spells who share a failure time. For example, suppose two individuals both find work after *t* days of unemployment, but they are ranked as failures *n* and *n+1*. The risk pool for spell *n* will include spell *n+1*, however, the risk pool for spell *n+1* will not include spell *n*. Lower ranked spells will therefore be compared to a smaller risk pool than a higher ranked spell even if they both transition simultaneously in discrete time.

When using discrete data, ties are inevitable and must be managed. The Breslow method, which is the simplest method for handling failure ties, is used in this study (StataCorp, 2019). The method essentially clusters spells by failure time and assigns the risk pool of the first ranked failure with the same failure time. Using the example in the previous paragraph, all spells that transition after the same unemployment duration as spell *n* will be assigned the same risk pool as spell *n*[13].

3.4 Censoring Mechanisms

Unemployment data is often incomplete; a spell start or end date may be missing. Survival analysis tools are sophisticated in that they can include spells with incomplete duration data and still draw accurate conclusions (Cameron & Trivedi, 2005). The sampling method used for this analysis requires that a start date be available [14], so it is solely spell end dates that may be missing. Using administrative data means that when a spell does end, that date is included in the dataset. When an end date is not available it indicates an ongoing unemployment spell; the date is not ‘missing’, rather the data is incomplete.

Spells that are incomplete must be right-censored to be included in the sample. Data for spell end dates is available until the *31/07/2020*. Therefore, it is known that if a spell requires censoring, the true spell end date lies in the range (0*1/08/2020*, ∞). Censoring data above a date is termed Type I right-censoring (Cameron & Trivedi, 2005). For survival analysis techniques to be valid, censoring must be independent. In other words, the distribution of censoring times must not be informative of the distribution of actual failure times. If censoring is independent, then censoring can be assumed to be exogenous. In this case the Cox model can accommodate censored data using standard techniques (Cameron & Trivedi, 2005). A binary censoring indicator δ_i is included in the model; a value of 1 indicates that a spell has failed, a value of 0 indicates a censored spell.

[13] Where spell *n* is the first ranked spell to transition from unemployment into employment after a given duration of unemployment *t*.

[14] Refer Chapter 4.3 for a description of the sampling method.

3.5 Partial Likelihood Estimation

To estimate parameter values in the model, a partial likelihood function is maximised (Cameron & Trivedi, 2005). First, spells are ranked in order of failure time. A risk set R_j is the set of all individuals who are at risk of failing just before the j^{th} ordered individual. The risk set includes all spells that have not died or been censored at t_j. A set D_j includes all those who die at a time t_j. The cardinality of D_j is captured by d_j; this denotes the number of individuals in the set D_j. This can be summarised as:

$$\begin{aligned} R(t_j) &= [l:\ t_l \geq t_j] \\ D(t_j) &= [l:\ t_l = t_j] \\ d_j &= \sum\nolimits_l \mathbf{1}(t_l = t_j) \end{aligned} \tag{10}$$

The probability that spell j is the next spell to transition is equal to the conditional probability that any spell in $R(t_j)$ fails.

$$\Pr\left[T_j = t_j \middle| R(t_j)\right] = \frac{\lambda_j(t_j|\boldsymbol{x}_j, \boldsymbol{\beta})}{\sum_{l\in R(t_j)} \lambda_l(t_j|\boldsymbol{x}_l, \boldsymbol{\beta})} = \frac{\phi(\boldsymbol{x}_j, \boldsymbol{\beta})}{\sum_{l\in R(t_j)} \phi(\boldsymbol{x}_l, \boldsymbol{\beta})} \tag{11}$$

The hazard function has been simplified from equation (4) so that $T_j = t_j$, because these specifications are written in terms of discrete data. The baseline hazard function drops out of this equation due to the proportional hazards assumption (Cameron & Trivedi, 2005).

Due to failure ties, equation (11) must be rewritten:

$$\Pr\left[T_j = t_j \middle| j \in R(t_j)\right] \simeq \frac{\prod_{m\in D(t_j)} \phi(\boldsymbol{x}_m, \boldsymbol{\beta})}{\left[\sum_{l\in R(t_j)} \phi(\boldsymbol{x}_l, \boldsymbol{\beta})\right]^{d_j}} \tag{12}$$

Equation (12) uses the Breslow method as discussed in Chapter 3.3. The partial likelihood function becomes the product of all these probabilities over all ranked failure events, k (Cameron & Trivedi, 2005). There are k spells that fail, while there are N spells in the sample—including censored spells.

$$L(\beta) = \prod_{j=1}^{k} \frac{\prod_{m\in D(t_j)} \phi(\boldsymbol{x}_m, \boldsymbol{\beta})}{\left[\sum_{l\in R(t_j)} \phi(\boldsymbol{x}_l, \boldsymbol{\beta})\right]^{d_j}} \tag{13}$$

The Cox model estimates β values by maximising the log of equation (13):

$$\ln L = \sum_{j=1}^{k} \left[\sum_{m\in D(t_j)} ln\phi(\boldsymbol{x}_m, \boldsymbol{\beta}) - d_j\, ln\left(\sum\nolimits_{l\in R(t_j)} \phi(\boldsymbol{x}_l, \boldsymbol{\beta}) \right) \right] \tag{14}$$

Equation (14) can be adjusted for censored data by including the censor mechanism, δ_i (Cameron & Trivedi, 2005). This is a binary variable that is 0 if the spell is censored and 1 if the spell ends. The censoring mechanism only contributes to the risk pool, rather than $D(t_j)$ as censored spells are not observed to fail; it is only known when they are censored. Equation (14) becomes:

$$\ln L(\beta) = \sum_{i=1}^{N} \delta_i \left[ln\phi(\boldsymbol{x}_i, \beta) - \ln\left(\sum\nolimits_{l \in R(t_j)} \phi(\boldsymbol{x}_l, \boldsymbol{\beta}) \right) \right] \tag{15}$$

Using equation (8), the first order condition becomes:

$$\frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^{N} \delta_i \left[\boldsymbol{x}_i - d_j\, \boldsymbol{x}_i^*(\boldsymbol{\beta}) \right] = 0 \tag{16}$$

$x_i^*(\beta)$ is a weighted average of the regressors $\boldsymbol{x}_l$ for individuals that are at risk at failure time t_i. This is equal to:

$$x_i^*(\beta) = \frac{\sum_{l \in R(t_i)} \boldsymbol{x}_l \exp(\boldsymbol{x}_l' \boldsymbol{\beta})}{\sum_{l \in R(t_i)} \exp(\boldsymbol{x}_l' \boldsymbol{\beta})} \tag{17}$$

Using the first order condition—Equation (16)—the Cox model estimates the parameter values β (Cameron & Trivedi, 2005).

It is not, however, conventional to report these parameters in output tables. Instead, the hazard ratio, or relative risk, is displayed. The proportional hazards assumption assumes that the percentage difference between the hazard rate of a category and a reference group is constant at all spell durations. Time subscripts can therefore be dropped. If a categorical variable has *z* categories, then $i=0,1,\ldots,z-1$ represents the different possible categories the variable captures; $i=0$ is the reference category. Using the region variable as an example, $i=0$ represents the reference region which is Auckland and $i=\{1,2,\ldots,15\}$ represents the 15 remaining New Zealand regions. The hazard ratio for category *i* of a given variable is expressed as:

$$Hazard\ Ratio_i = \frac{\lambda_i}{\lambda_0} = \exp(\beta_i) \tag{18}$$

Hazard ratios are used to estimate the relative risk of a category at any spell length compared to the reference category for that variable (Zwiener et al., 2011). As an example, a hazard ratio of 1.5 implies that a job seeker in category *i* is 50% more likely to find employment at any unemployment spell duration compared to the reference category for that variable. Similarly, a hazard ratio of 0.7 implies that a jobseeker in that category is 30% less likely to find employment compared to the reference category. This is how categorical variables—including the region variable—are interpreted. For continuous variables, the hazard ratio states the effect of a one-unit change in the variable.

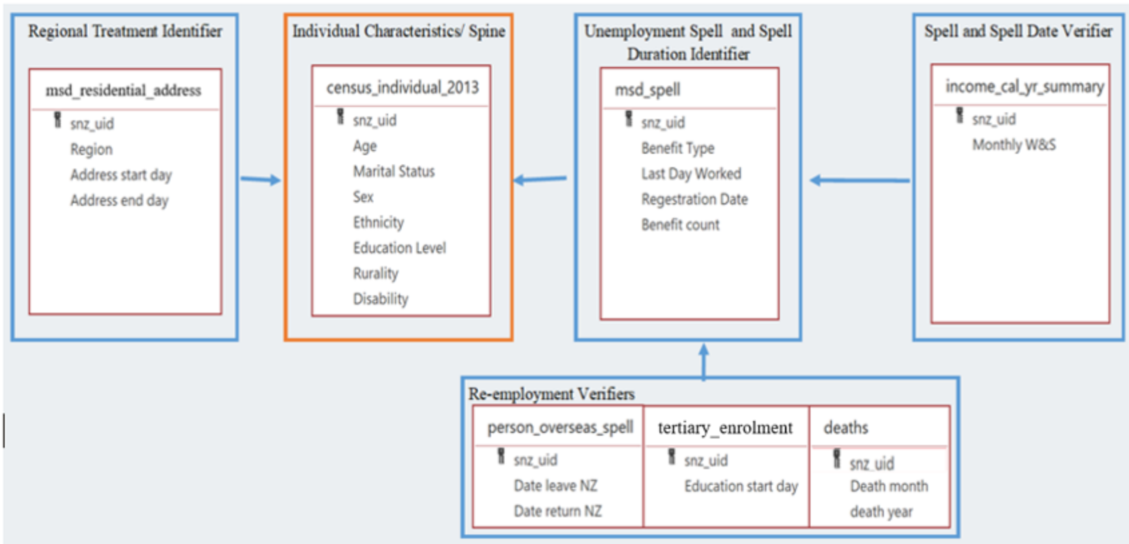
The β parameters that are estimated in Equation (16), estimate the marginal effects of that category when other covariate values are equal to zero. This isolates the relative risk of that category compared to the reference category of that variable. Because values for other covariates are fixed when estimating the parameters in equations (16), equation (18) does not need to be adjusted for a multivariate analysis.

Chapter 4: Data and Sampling

4.1 Data Set

This research uses Statistics New Zealand's (Stats NZ) Integrated Data Infrastructure (IDI). The IDI contains linked microdata on New Zealand individuals and households (Stats NZ, n.d.). Data is sourced from administrative datasets held by both government and non-government organizations. Further data is collected through Stats NZ surveys. Available data covers life events including employment, education, census responses, and income. Data is deidentified to ensure that data remains confidential. Unique identifiers[15] allow datasets to be linked across an individual to create custom samples. Figure 2 below summarises how different IDI datasets have been linked to create the dataset for this research. Each dataset is discussed in greater detail throughout this section. Once sampling is applied, a total of 23,970 unemployment spells are included in the sample.

Figure 2 **Table Linkages**



This diagram summarises the linkages and purpose of each IDI dataset that has been used. Each is discussed in the following sections.

4.2 Census Data and Covariates

A spine of 2013 census data is used for this analysis as shown in Figure 2. The census is run every five years and is the official count of both people and dwellings in New Zealand (Stats NZ, 2020a). The census is the only population-wide survey in New Zealand. Census data is used to create characteristic variables for individuals including sex, age, and ethnicity. An individual's characteristics are defined as their response to the 2013 census on the 5th of March 2013. Covariates are fixed to these values. Each unemployment

[15] snz_uid in Figure 2 uniquely identifies an individual within the IDI. Using this identifier, an individual's data can be linked across IDI datasets.

spell is a single observation; time variation in covariates is ignored for simplicity. While the 2018 census is more recent, the 2013 census allows for longer spells to be modelled. 2018 data would limit spell lengths to approximately 30 months before censoring is required; the 2013 census allows for spells up to 8 years long to be modelled.

The purpose of this study is to identify whether an individual's re-employment probability is different in each New Zealand region. Regional re-employment probabilities can be thought of as being the product of regional socio-economic and demographic compositions. Demographics generate variation in the actual and perceived quality and fit of the labour supply. Socio-economic factors instead influence labour demand; regional labour markets with a greater abundance of appropriate jobs tend to exhibit higher re-employment probabilities (Hane-Weijman, 2018). The literature below discusses how individual characteristics affect re-employment probabilities. If not controlled for, these individual level variables are captured by a region variable as that region's demographic composition. To capture an individual's counterfactual re-employment probability if their job search were in a different region, individual characteristics must be independent of regional assignment. In other words, an individual's characteristics cannot depend on the region they reside in. Controlling for such characteristics isolates the socio-economic determinants of regional re-employment probabilities. Socio-economic factors do capture a counterfactual job search in a different region. Numerous covariates are derived from 2013 census data to control for regional demographic compositions including age, educational achievement, and sex.

Higher education is linked to improved levels of human capital which enhances productivity and increases the likelihood of an individual finding employment (Melin, 2001). This outcome is well established in the literature (Forbes & Barker, 2017; Moore, 2004; Tansel & Taşçi, 2010; D'Agostino & Mealli, 2000; Babucea et al.,2009; Ciucă & Matei, 2011). An individual's education level is captured by the highest education level they had achieved by the 2013 census. Non-New Zealand qualifications are equated to the comparative New Zealand level[16].

Gender is also an important covariate to include. Moore (2004; 2006) and Hyslop et al. (2004) identify gender effects in New Zealand. Hyslop et al. (2004) quantified this effect; New Zealand women are 1–2% less likely to be in employment than men. Improved re-employment probabilities for men in New Zealand are largest for those under the age of 35 (Moore, 2006), however, variation has been declining since 1990

[16] For example, Level 3 represents an individual receiving an equivalent qualification to completing the final year of a New Zealand high school.

(Moore, 2004). Hyslop et al. (2004) identified that gender effects are most pronounced at shorter spell lengths; Ciucă & Matei (2011) observed a similar effect in Romania. Tansel & Taşçi (2010), Eubanks & Wiczer (2016), Babucea et al. (2009), and D'Agostino & Mealli's (2000) additionally observe that males enjoy higher re-employment probabilities in European nations and the United States. Forbes & Barker (2017) instead identify an insignificant effect in Australia; instead, unemployed women are 30% more likely to exit the labour force entirely. Sex is captured by a dummy variable derived from 2013 census responses. A value of one represents an individual being female.

Regarding the inclusion of age in such an analysis, D'Agostino & Mealli (2000) identify that age can affect re-employment probabilities in two ways. 'Parabolic' suggests that probabilities peak in the middle of one's career, while 'monotonic' implies that age has a constant directional effect. 'Monotonic' effects are most commonly observed; younger jobseekers are generally observed to experience higher re-employment probabilities (D'Agostino & Mealli, 2000; Ciucă & Matei, 2011; Eubanks & Wiczer, 2016; Forbes & Barker, 2017; Tansel & Taşçi, 2010; Babucea et al., 2009; Moore, 2004; Moore, 2006; Hyslop et al, 2004). In New Zealand those aged between 30–55 are on average 2–3% less likely to find work than a 20–24 year-old (Hyslop et al., 2004). This effect grows to 7% and 20% for 55–59 and 60–65 year-olds, respectively. Only D'Agostino & Mealli (2000) discusses 'parabolic' age effects; the authors observe this effect in the United Kingdom, Spain, and Italy. Age in this study captures an individual's age range[17] on the date of the 2013 census. Only those aged between 18–60 are included. The benefit is unavailable to those under the age of 18 and those who become unemployed after the age of 60 may exit the labour force into retirement[18]. Transitions into retirement are difficult to capture accurately, so these individuals are excluded to ensure that they do not bias the results.

The effect of marriage has been observed to improve re-employment probabilities in New Zealand and Australia (Moore, 2004; Forbes & Barker, 2017). In Australia, marriage improves re-employment probabilities by 30% and reduces transitions out of the labour force by 20% (Forbes & Barker's, 2017). Moore (2004) similarly observes improved re-employment probabilities for married jobseekers in New Zealand[19]. Ambiguous outcomes are instead observed in Europe (Tansel & Taşçi, 2010; D'Agostino

---

[17] The age ranges used in this study are 18–23, 24–29, 30–35, 36–40, 41–45, 46–50, 51–55, and 56–60.

[18] The retirement age in New Zealand is 65 years old (New Zealand Government, 2020).

[19] Moore (2004) does not state the specific effect in New Zealand. The author only states that marriage is observed to have a positive effect.

& Mealli, 2000). Tansel & Taşçi (2010) specifically observe that marriage positively affects male, and adversely affects female, jobseekers. In this research, marital status is captured by 2013 census responses. Marital status is aggregated into three categories: marriage/civil union, never being in a marriage/civil union, and being divorced/separated/widowed.

New Zealand regions include both urban and rural labour markets; 10 regions had over 25% of its population living rurally in 2013 (MBIE, 2013). In the West Coast, Tasman, and Northland regions this proportion was over 40%. Urban versus rural labour markets can be assumed to exhibit structural differences; however, the question remains as to whether these differences affect re-employment probabilities. The literature does not suggest a definitive outcome. Imbens & Lynch (2006) observe inconsistent effects for American youths. Tansel & Taşçi (2010) identify an insignificant effect in Turkish labour markets. Rurality will again be defined as an individual's response to the 2013 census which asks if an individual's residential address is in an urban or rural setting.

Researchers capture ethnicity differently depending on the country context of the study. Imbens & Lynch (2006) use generic ethnicity categories. Forbes & Barker (2017) categorise ethnicity as whether someone is Australian born, and if not, whether their country of birth is English speaking. Kroft et al. (2013) instead produces fictitious names that are purposefully informative of ethnicity. Each study identifies that re-employment probabilities are greatest for White ethnicities. Moore's (2004) New Zealand study finds that NZ Europeans experience improved re-employment probabilities while Pacifica experience reductions. Pacheco & Webber (2016) find a somewhat apposing outcome. Maori and Pacifica experienced slightly greater shifts into full-time employment relative to NZ Europeans between 1996 to 2006. The authors do discuss that NZ Europeans may have already had higher rates of full-time employment. This finding may simply indicate that it is more difficult to increase this proportion if the proportion was already high. Ethnicity is captured in this research using a prioritised ethnicity variable derived from 2013 census responses. The census allows individuals to classify themselves into multiple ethnic groups; five are used in this research[20]. Prioritised ethnicity takes the two ethnicities an individual identifies with most and assigns them to the highest priority ethnicity of the two[21] (Education Counts, 2021).

---

[20] The different ethnicity classifications are Maori, Pacifica, Asian, Other, and NZ European. The order they are stated in is the order used for priority ranking.

[21] The census asks individuals to rank the ethnicities they provide in order of their affiliation to that ethnicity. Prioritisation uses the order stated in footnote 20 and assigns an individual to the highest ranked of the two ethnicities that an individual has the highest affiliation to. Someone who identifies as

Forbes and Barker (2017) include a variable that captures those with long-term health conditions. They observe that health conditions drastically reduce labour market success in Australia. Those with conditions experience a 43% reduction in their re-employment probability compared to someone with no conditions. Unemployed people with disabilities are also over twice as likely to exit the labour force entirely. Disability status is captured as a dummy variable. A value of one indicates that someone suffers from a disability. Disability status is self-reported in the 2013 census. The census question asks if an individual has a long-term disability that stops them from doing everyday things that others can do[22].

4.3 Identifying the Unemployed

Census data is then linked to Ministry of Social Development (MSD) benefit data as shown in Figure 2. Unemployed people are identified as those who receive a New Zealand Unemployment or Jobseeker Support benefit. The former became the latter due to Welfare Reform changes on the 15th of July 2013 (MSD, 2013). Moore (2006) also used benefits to identify unemployed people in his similar study. I aim to build on Moore's analysis by focussing on regional effects. Moore's (2006) study had the limitation that he was unable to identify specific benefit types; not all benefits require recipients to be unemployed. MSD data in the IDI resolves this limitation by allowing specific benefit types to be identified; unemployed people can be identified with a high degree of confidence.

After the Welfare Reform changes, benefit eligibility extended past the unemployed to include part-time workers seeking further employment (underemployed) and those working limited hours due to health conditions (WINZ, 2021). Further conditions on beneficiaries includes a willingness to accept reasonable employment, be over 18 years old, and be a New Zealand citizen or a permanent resident residing in New Zealand for at least two years.

The International Labour Organization (ILO) defines unemployment as being without work, being available to work, and actively seeking work (ILO, n.d.). Beneficiaries are required to be actively seeking and willing to accept work; they meet the 'availability' and 'actively seeking' criteria. As the benefit does not require an individual to be unemployed, the 'without work' criterion is not guaranteed. Chapter 4.5 discusses additional sampling methods to ensure that this condition is met. After these

NZ European and Maori, will be classed as Maori because Maori is a higher priority ethnicity. Only those with no other ethnic affiliations are classed as being NZ European.

[22] Long-term for disability status refers to a period of over 6-months.

techniques are applied, the three ILO conditions of unemployment can be assumed; the length of a benefit spell can be used to estimate the duration of an unemployment spell.

Once unemployment can be assumed, two caveats remain. Firstly, a benefit spell can end for non-job finding reasons. Chapter 4.6 discusses methods to resolve this issue. The second caveat is that the benefit is means tested (Community Law, 2020). Additional household income earned above a threshold value results in an abatement of the benefit. Whether an individual has a partner, has children[23], and their partners benefit status determines how the abatement is applied. Table 3 summarises this. An individual loses benefit eligibility if the abatement nullifies the entire benefit payment. Means testing results in the scope of this study adjusting to only those whose household income while unemployed is below that which would result in a full abatement. An individual must also register for a benefit spell, eliminating unemployed people who do not register for a benefit from the sample.

Table 3

Scenario	**Household Income Threshold**	**Abatement per Dollar Above Threshold Value**
Single without children	$90 gross per week	70 cents
Single with dependent children	$115 gross per week (may ignore up to $20 for childcare)	$115-$215- 30 cents $215+ - 70 cents
Partner also on benefit	$90 gross per week	35 cents
Partner not on benefit	$90 gross per week	70 cents

(Community Law, 2020)

As explained earlier, the population spine, which consists of 2013 census data, is linked to MSD benefit spell data as shown in Figure 2. MSD is responsible for the provision of social welfare including social housing, student loans, and most relevant to this research, New Zealand benefits (MSD, n.d.). A benefit spell is defined as the period of time in which an individual receives an MSD benefit. MSD benefit spell data in the IDI contains administrative records for spells across all MSD benefit types. This includes data on which benefit is being received, the total number of benefit spells an individual has experienced, and the dates associated with a benefit spell. Sampling begins by identifying unemployed people as those receiving either an Unemployment or Jobseeker Support benefit.

[23] Having children only affects those without a partner.

Census variables may change over time[24]. Census variables need to be accurate for the first day of an individual's unemployment spell. To ensure that this is true, only spells that begin near to the 2013 census can be included in the sample. A sample period around the 2013 census is determined; unemployment spells that begin during this sample period are included in the sample. This period must also be large enough to capture a sufficient sample size of unemployment spells. The period of the 1st of November 2012 to the 31st of July 2013 is chosen as this sample period[25]; this period captures all benefit spells that begin in the full four calendar months either side of the March 2013 census.

To ensure that spells captured by the sample period occur during comparable economic circumstances, Appendix A displays macroeconomic indicators for New Zealand[26]. The graphs are derived from Reserve Bank of New Zealand (RBNZ) 2020 data. The highlighted regions represent the sample period. It cannot be expected that no economic variation will occur during the sample period; however, GDP growth, the CPI, and the current account balance demonstrate stability over the sample period. It must also be noted that RBNZ kept the OCR constant at 2.5%. The unemployment rate does however fall by 0.5 percentage points between December and March. The NZD trade weighted index and NZD/USD exchange rates demonstrate larger than ideal variation too.

Spell dates are determined by administrative MSD benefit spell data. Two dates in the MSD dataset are used to estimate an individual's spell start date. The beginning of a spell is initially defined as the last date that an individual worked before an unemployment spell. This variable is often missing in MSD's dataset. When this is the case, the day that an individual registered with MSD for a benefit spell is used. Constructing the spell start day using two variables increases the sample size as one of these dates is available for all MSD benefit spells. The end date for an unemployment spell is simply the date stated in MSD data as the end date for a benefit spell. All concluded benefit spells are recorded in the data; an unavailable end date indicates an ongoing spell rather than missing data.

One final covariate that is included in this analysis is the number of MSD benefit spells that an individual has experienced. This is consistent with Tansel & Taşçi's (2010) inclusion of a variable that indicates previous unemployment spells. The authors identified that first-time jobseekers perform worse than those with prior unemployment spells. This indicates that job search experience improves re-employment probabilities in

[24] For example, increased age range, gain additional qualifications, or change marital status.

[25] This period is selected relatively arbitrarily. Models were first attempted using a smaller sample period; this produced too small a sample size.

[26] Chapter 6.2 empirically tests this further.

Turkey. Heckman & Borjas (1980) instead discuss how prior spells have the same negative effect on re-employment probabilities as longer spells. Employers use previous unemployment spells as indicators of human capital deficiencies that previous employers identified. Employers also assume that human capital depreciates during an unemployment spell (Eriksson & Rooth, 2014; Pavoni & Violante, 2007). Previous spells therefore supplement spell durations in employer ranking models. A spell count variable is derived from MSD benefit spell data. This variable is inclusive of all MSD benefit types and is not limited to just the Unemployment and Jobseeker Support benefits. Estimations cannot be interpreted as the effect of an additional unemployment spell on re-employment probabilities, rather the effect of an additional MSD benefit spell.

4.4 Regional Treatments

Regional identification is essential to the research question. As shown in Figure 2, MSD residential address data is merged with 2013 census and MSD benefit spell data to identify an individual's region. A regional treatment is assigned to an individual based on their residential address on the day that they become unemployed[27]. MSD residential address data captures the geographic information associated with an MSD client including the region of New Zealand that their residential address is located in. Additionally, the start and end dates of an individual residing at an address is available. If an individual's benefit spell begins while they resided at an address, then the region corresponding to that address is assigned to that individual. This region variable is coded as a single categorical variable.

MSD data divides New Zealand into 16 regions:

- Auckland
- Northland
- Waikato
- Bay of Plenty
- Gisborne
- Hawkes Bay
- Taranaki
- Manawatu
- Wellington
- West Coast
- Canterbury
- Otago
- Southland
- Tasman
- Nelson
- Marlborough

These 16 regions become the different possible treatments in this analysis. Auckland, as the largest region by population, is selected as the reference region. Geographical boundaries for each region are displayed in Appendix B.

[27] An additional specification uses an individual's residential address on the last day of their unemployment spell. This specification can be found in Chapter 6.4.

4.5 Validating the Identification of Unemployed Individuals

Using benefit data to identify the unemployed raises caveats as discussed in Chapter 4.3. Underemployed individuals maintain their Jobseeker Support benefit eligibility; this research is not interested in such individuals. Inland Revenue Department (IRD) data includes information on an individual's wages and salary (W&S) income in a calendar month. IRD is responsible for collecting the majority of the Crown's revenue, predominantly through taxation (IRD, n.d.). This data is used to further verify that an individual is without work.

The following criteria is used to identify those who are without work.

1. Earned no W&S in the first full calendar month that they are receiving a benefit.
2. Earned no W&S in the month of, or month after, starting a benefit spell.
3. Earned at most 30% of their pre-benefit months' W&S in either the month of, or month after starting a benefit spell.

Conditions (2) and (3) are used to identify those whose unemployment spell does not last a full calendar month. Condition (3) also aims to capture those who may gain short-term employment. Employment such as at a single event, or a one-off days' work that pays wages and salary does not fulfil the intention of the re-employment definition. Short-term employment would result in an immediate additional unemployment spell following the work; such transitions are not of high economic significance. The 30% threshold allows such individuals to be included in the sample while excluding those who do not meet the 'without work' criterion.

Administrative data and the sampling methods used often result in multiple records for what is a single spell. This can occur due to temporary benefit eligibility loss, benefit data changes, and if an individual has multiple records in additional datasets that are linked to MSD benefit data. In instances where one benefit spell ends then another immediately commences, the records are combined into one continuous unemployment spell. A similar procedure is used for overlapping spells. When benefit records overlap, the record with the earliest start date captures the start date and individual covariates; the spell end date is taken from the overlapping record with the latest end date. Duplicate spells are deleted.

When the Unemployment benefit became the Jobseeker Support benefit, MSD created an additional administrative record for ongoing spells. The initial Unemployment benefit spell ends, and a new Jobseeker Support benefit spell begins. This transition was not precisely recorded; if it were, the methods in the above paragraph would have resolved the issue. IRD W&S data is used to estimate a new spell start date for spells that begin on the date of the transition. The midpoint (15th) of the month in which an individual last earned W&S before earning no W&S becomes the new spell start date. If this method does not change the month someone became unemployed, their original start date is used. Individuals are excluded from the sample if this new start date does not fall within the sample period: 1st of November 2012 to the 31st of July 2013.

4.6 Verifying Re-employment Transitions

Finding work is not the sole determinant of a benefit spell concluding. Anyone who no longer meets MSD's benefit criteria can have their spell terminated. As Table 2 shows, three additional tables are used to exclude those whose spell likely ended without a transition into employment. These tables include records on transitions into tertiary education, extended overseas travel, and death.

Ministry of Education (MoE) tertiary course enrolment data includes records of all enrolments to every tertiary education provider in New Zealand. Individuals who enrol in further education are unlikely to be actively seeking or available for full-time employment. Other social welfare mechanisms are available to these individuals, and they will likely transition from the benefit to student financial support provided by StudyLink[28]. Ministry of Business Innovation and Employment (MBIE) person overseas spell data captures border movements to and from New Zealand. Records are available for those who go overseas for more than 187 days. Beneficiaries lose eligibility if they go overseas for an extended period. Lastly, Department of Internal Affairs (DIA) life events data is used to identify those who have passed away. DIA death records include all deaths that are registered in New Zealand. Records are available for New Zealand born individuals who pass away overseas.

Enrolling in further education, travelling overseas, and death all result in the conclusion of a benefit spell without transitioning into employment[29]. If an individual has a record in any of these tables within one week of the end date of their spell, then they are excluded from the sample.

To validate that any remaining spells solely transition into employment, IRD income data is used to verify that an individual earns wages and salary after the conclusion of their spell. The spell end date specified in MSD data is used if that individual earns wages and salary in the month their spell is said to end, or the 2 following months. If not,

[28] StudyLink is the MSD department that provides student financial support including student loans and allowances. This is a separate MSD department to WINZ; WINZ provides the Unemployment/Jobseeker Support benefits.

[29] This is not an exhaustive list of non-employment reasons for benefit termination.

then MSD's date is considered inaccurate. The spell end date is adjusted to the midpoint (15th) of the month that they begin to earn wages and salary after an unemployment spell. Only the 12 months after MSD's stated spell end date are considered. If no month is identified, then the data is deemed to be too uninformative, and these individuals are excluded from the sample.

Chapter 5: Empirical Results

5.1 Initial Model

The primary aim of this research is to identify the counterfactual re-employment probabilities a jobseeker would face in each New Zealand region. I begin by analysing the effects of regional variables without the inclusion of covariates in the model. This provides an initial understanding of regional treatment effects. Without covariates, regional variables capture both socio-economic and demographic compositional effects on re-employment probabilities. By excluding covariates, estimations in Table 4 assume that an individual's characteristics are dependent on the region in which they are seeking employment. As Chapter 4.2 discusses, this assumption is untrue[30]; interpretations must be made with caution. Estimations in Table 4 display the regional hazard ratios of this model. Hazard ratios state the relative risk compared to a reference category[31]; Auckland is the reference region.

Table 4

VARIABLES	Hazard Ratio	Standard Error
Auckland	*Reference*	
Northland	0.865***	(0.0312)
Waikato	0.955*	(0.0233)
BOP	0.975	(0.0262)
Gisborne	0.906*	(0.0471)
Hawkes	0.996	(0.0300)
Taranaki	0.926	(0.0461)
Manawatu	0.909***	(0.0270)
Wellington	0.902***	(0.0216)
West Coast	0.902	(0.0683)
Canterbury	1.044*	(0.0269)
Otago	1.024	(0.0345)
Southland	1.159***	(0.0409)
Tasman	0.979	(0.0747)
Nelson	0.849***	(0.0505)
Marlborough	0.922	(0.0700)
N	23,970	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Source: IDI. Authors own calculations

[30] The characteristics of an individual who begins a job search in another region do not change. For example, beginning a job search in a region with a younger and less educated population does not decrease one's age and level of educational achievement. When demographic effects are captured by a regional variable (as in Table 4), the model assume that individual characteristics are dependent on the region in which they reside.

[31] Refer Chapter 3.5 for more details on interpreting hazard ratios.

Table 4 indicates that regional variation may occur on average for spells that begin in the period of 01/11/2012–31/07/2013. Five regions exhibit statistically different re-employment probabilities from Auckland under this specification. Only Southland experiences greater re-employment probabilities: re-employment probabilities are on average 16% greater than for those in Auckland. Other highly statistically significant regions include the Northland, Manawatu, Wellington, and Nelson regions; each is 10–15% less likely to find work than someone in Auckland.

5.2 Primary Model Specification

5.2a Regional Effects

Covariates are now included to disentangle regional demographic compositions and isolate regional effects. Table 5 displays regional hazard ratios[32]; Model (1) analyses regional variations while Model (2) uses aggregated regional definitions. Aggregation combines regions to ensure that each aggregated region includes a similar number of unemployment spells and ensures that each region has a large enough sample size[33].

The primary conclusion is that yes, regional re-employment probability variation does exist in New Zealand even after controlling for regional demographic variations. Probability variation from Auckland is highly statistically significant in the Northland, Hawkes Bay, Wellington, Canterbury, and Southland regions. Re-employment probabilities are greatest in Southland; on average, they are 32% higher than in Auckland. Canterbury and the Hawkes Bay regions similarly exhibit greater re-employment probabilities. Each experiences re-employment probabilities that are on average 15% and 9% higher than Auckland, respectively. On the contrary, Wellington and Northland both exhibit statistically significant reductions in their re-employment probabilities. A jobseeker in these regions is on average 10% less likely to find work in any given time-period than a jobseeker in Auckland.

Variation is not limited to these regions. Other regions do exhibit re-employment probability variations at lower significance levels. Re-employment probabilities in the Otago and Bay of Plenty regions are on average 6.4% and 5% higher than in Auckland, respectively, at the 10% significance level. Re-employment probabilities are on average 5% lower in the Manawatu region at the same level of statistical significance.

[32] The full estimation table is available in Appendix C.
[33] Refer Appendix D for region and aggregated region sample counts.

Table 5

	(1)		(2)
Region		**Aggregated Region**	
Auckland	*Reference*	*Auckland*	*Reference*
Northland	0.902***	Northland	0.901***
	(0.0347)		(0.0346)
Waikato	1.011	Waikato	1.010
	(0.0251)		(0.0250)
BOP	1.048*	BOP	1.046
	(0.0293)		(0.0292)
Gisborne	0.958	East Coast	1.051*
	(0.0503)		(0.0307)
Hawkes	1.086**	Taranaki-Manawatu	0.964
	(0.0353)		(0.0265)
Taranaki	1.007	Wellington	0.897***
	(0.0519)		(0.0217)
Manawatu	0.952*	West Coast-Tasman	1.019
	(0.0286)		(0.0404)
Wellington	0.898***	Canterbury	1.148***
	(0.0217)		(0.0308)
West Coast	1.003	Southland	1.161***
	(0.0754)		(0.0317)
Canterbury	1.149***		
	(0.0309)		
Otago	1.064*		
	(0.0361)		
Southland	1.322***		
	(0.0496)		
Tasman	1.133		
	(0.0933)		
Nelson	0.932		
	(0.0616)		
Marlborough	1.084		
	(0.0791)		
N	23,970	***N***	23,970

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Complete table in Appendix C
Source: IDI. Authors own calculations

Regional hazard ratios in Table 5 generally appear to have increased in value relative to Table 4. When covariates are excluded the model overestimates negative, and underestimates positive, regional effects on re-employment probabilities. Compared to Table 4, Table 5 observes a greater number of regions to exhibit statistically significant re-employment probability improvements compared to Auckland. A model that excludes covariates does not sufficiently estimate changes to an individual's re-employment probability if their job search were in a different region.

New Zealand regions differ largely in the size of their population. The number of spells available for each region also demonstrates this variability[34]. Regional aggregation aims to ensure that each aggregated region has a similar number of spells within the sample[35]. Aggregation in Model (2) aims to verify the initial findings of Model (1). As in Model (1), the Southland region (which now includes Otago) experiences greater re-employment probabilities than Auckland. The reduced magnitude of the hazard ratio captures the difference between the individual regional effects of Southland and Otago. The Northland, Wellington, and Canterbury regions maintain their significance from Model (1); however, the Bay of Plenty and East Coast regions lose significance when aggregated.

Hazard ratios only provide a static analysis. These ratios are based on the proportional hazards assumption which requires covariate variations to be constant at all spell durations. While this assumption allows the versatility of the semi-parametric Cox model to be applied, the assumption is unlikely to be perfect. Observing the hazard functions for each region allows for a more in-depth understanding of each regional labour market. Figure 3 plots the hazard functions for each region both overlayed and on separate axis[36]. The associated regional survivor functions can be found in Appendix M. The hazard function states the instantaneous probability that an individual in a region will transition into employment, conditional on them still being unemployed at that duration of unemployment.

The slope of a hazard function indicates duration dependence. Generally, the functions are downward sloping implying that New Zealand regions do exhibit negative duration dependence. This is consistent with observations in the wider literature[37]. The steepest functions tend to be those regions that have the highest re-employment probabilities. Steeper functions imply stronger duration dependence. This is consistent with Kroft et al.'s (2013) finding that re-employment probabilities tend to converge at longer spell durations. For regional re-employment probabilities to converge at longer spell durations, regions with higher initial probabilities must exhibit greater duration dependence.

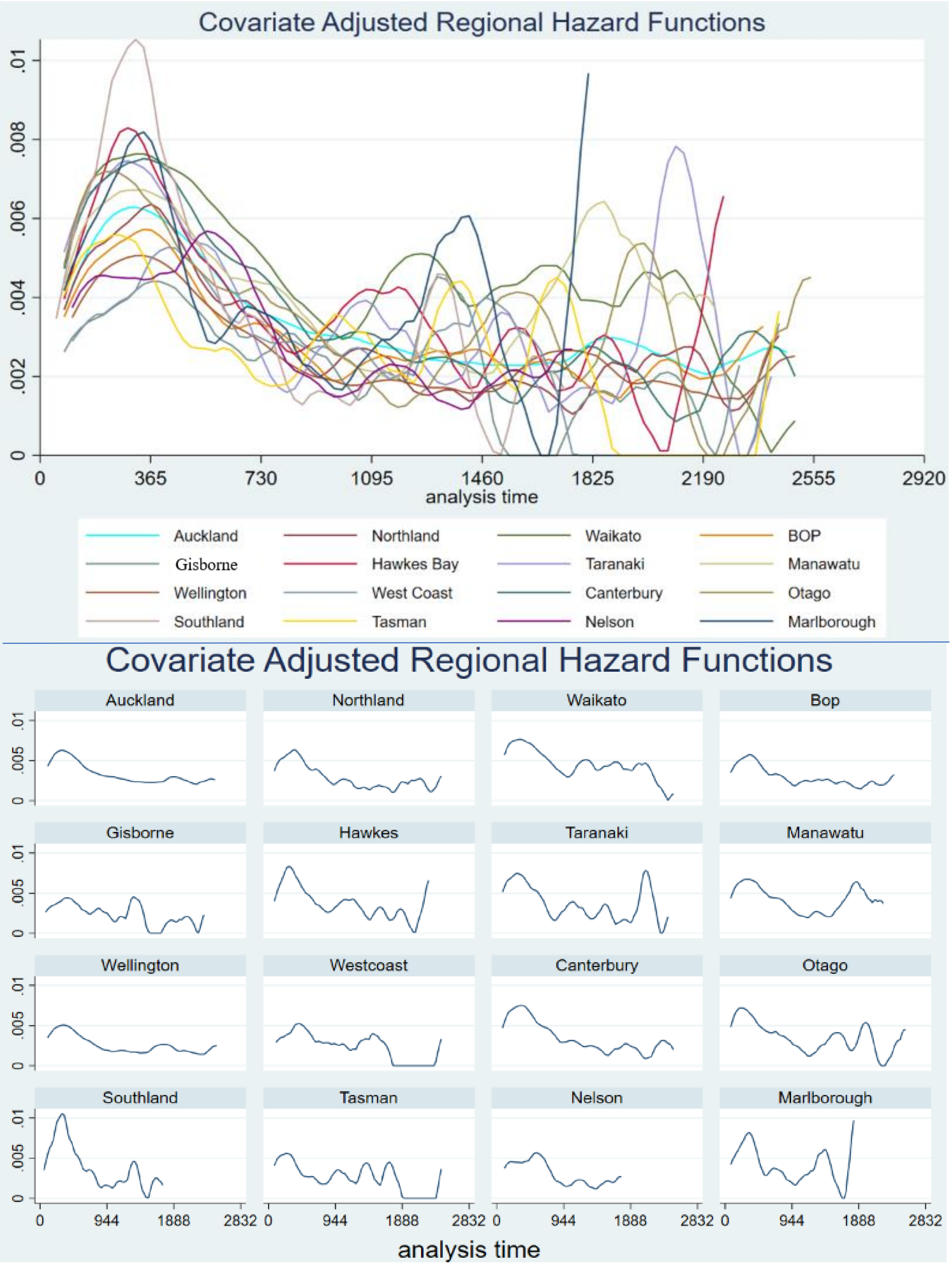
[34] Regional/aggregated region sample counts can be found in Appendix D.

[35] Appendix B demonstrates how the regions are aggregated.

[36] Functions in Figure 3 control for the covariates in the model and are based on Model (1) in Table 5/Appendix C. This means that the functions are estimated using the proportional hazards assumption which minimises the variation in duration dependence. However, by plotting the functions against spell duration, rather than estimating a single parameter, regional re-employment probability variation at different spell durations can be better understood.

[37] This is discussed in Chapter 2.

Figure 3



Graphs are derived from the primary model reported as Model (1) in Table 5/Appendix C.

Duration dependence is not consistently negative; re-employment probabilities are observed to improve in approximately the first year of a spell. The relationship between re-employment probabilities and spell length could be described as parabolic in nature: re-employment probabilities initially increase and then decrease after approximately one year of unemployment. Positive duration dependence may be due to the time it takes to reach the stage of accepting a job offer. It takes a jobseeker time after becoming unemployed to prepare for job search; the hiring process then takes time to match a jobseeker to a hiring firm. The duration of this process may manifest as positive duration dependence at shorter spell durations.

Wide variation in the shape of each regional hazard function can be observed. Regarding regional sample sizes, larger regions exhibit much smoother functions. Hazard rates for smaller regions even fall to 0 for an extended period of potential spell durations. This is likely indicative of the regional sample not including spells that end at these potential durations. This further motivates the inclusion of an aggregated region specification (Model (2)) to increase the number of spells in each region. Most regions exhibit a similar shaped function in the first three years of a spell; spells greater than five years start to lose reasonable interpretability when graphically represented. It is likely that only a small number of spells in each region lasted for this length of time. As a result, the functions become less smooth and exhibit greater volatility.

Observing the overlayed functions allows for a better understanding of the hazard ratios in Table 5. Regions tend to experience the deviation that their hazard ratio suggests at shorter spell lengths. These functions appear to converge or cross Auckland's reference function at longer spell lengths. Regional re-employment probability variations appear to be most pronounced at shorter spell durations. Kroft et al.'s (2013) discussion supports this finding. The authors found regional market tightness—which is one of the mechanisms the region variable aims to capture—has little effect on hazard rates at longer spell durations.

5.2b Covariate Effects

Covariates are included to control for demographic composition variability that would otherwise be captured by region variables. By including covariates in the model, hazard ratios are estimated so that the effect of each covariate can be analysed. The tables in this section are excerpts from Appendix C. Models (1) and (2) are the same models as are in Table 5.

Both models in Table 6 suggest that men in New Zealand experience higher re-employment probabilities. Female re-employment probabilities are on average 12% lower than their male counterparts. This is consistent with Moore (2004; 2006) and Hyslop et al.'s (2004) findings of gender effects in the New Zealand labour market.

Re-employment probabilities are observed to decrease with age. Table 6 suggests that there is a monotonic, rather than a parabolic, relationship between age and re-employment probabilities. Hazard rates vary by a smaller degree between older age ranges; the relationship can be described as a negative exponential relationship. Table 6 suggests that younger people can expect shorter unemployment spells; the youngest group experiences re-employment probabilities that are on average 37% greater than the oldest age range. This outcome could indicate several mechanisms. For example, employers may have a preference for younger workers or there may be a greater stigma attached to unemployment when older. Future researchers may wish to analyse the specific mechanisms causing older jobseekers to exhibit lower re-employment probabilities.

Table 6

VARIABLES	(1)	(2)
Sex	0.878***	0.877***
(Male=Reference)	(0.0125)	(0.0125)
Age Range		
18–23	*Reference*	*Reference*
24–29	0.880***	0.881***
	(0.0187)	(0.0187)
30–35	0.823***	0.825***
	(0.0223)	(0.0223)
36–40	0.766***	0.766***
	(0.0241)	(0.0241)
41–45	0.728***	0.727***
	(0.0211)	(0.0211)
46–50	0.713***	0.714***
	(0.0213)	(0.0213)
51–55	0.657***	0.658***
	(0.0211)	(0.0211)
56–60	0.628***	0.630***
	(0.0228)	(0.0228)
N	23,970	23,970

Robust standard errors in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Full table in Appendix C
Source: IDI. Authors own calculations

It can be seen in Table 7 that Pacifica and Asian ethnicities exhibit re-employment probabilities that are statistically significant in their difference from NZ Europeans. Pacific Peoples are on average 6.5% more likely to find work; Asians experience 6% reductions in their re-employment probabilities. Identifying no statistical difference for Maori and improved probabilities for Pacifica is a surprising outcome. MBIE'S (2020a) Maori labour market trends report discusses how Maori are over-represented in both the unemployment and underutilization rates, with nearly one third of Maori youth not being in any formal education, employment, or training (NEET). MBIE'S (2020b) Pacific Peoples report shows similar NEET and underutilization rate trends for Pacifica. MBIE's data and previous literature (Imbens & Lynch, 2006; Forbes & Barker, 2017; Kroft et al., 2013; Moore's, 2004) suggests an expectation of worse labour market outcomes for these ethnicities. Future research may wish to investigate the mechanisms behind this outcome.

Table 7

VARIABLES	(1)	(2)
Ethnicity		
NZ European	*Reference*	*Reference*
Maori	0.977	0.978
	(0.0165)	(0.0164)
Pacifica	1.065**	1.067***
	(0.0263)	(0.0264)
Asian	0.943*	0.943*
	(0.0296)	(0.0296)
Other	0.953	0.954
	(0.0421)	(0.0421)
Disability	0.699***	0.699***
(No disability=Reference)	(0.0153)	(0.0154)
Education Level		
No Education	*Reference*	*Reference*
Level 1	1.059***	1.059***
	(0.0234)	(0.0234)
Level 2	1.131***	1.128***
	(0.0259)	(0.0259)
Level 3	1.167***	1.164***
	(0.0259)	(0.0258)
Level 4	1.189***	1.187***
	(0.0312)	(0.0311)
Level 5 Diploma	1.268***	1.265***
	(0.0469)	(0.0464)
Level 6 Diploma	1.303***	1.304***
	(0.0579)	(0.0577)
Bachelor's degree	1.444***	1.432***
	(0.0400)	(0.0397)
Postgrad/Honours	1.658***	1.648***
	(0.0949)	(0.0939)
Masters	1.358***	1.355***
	(0.0823)	(0.0817)
PhD	1.584***	1.553***
	(0.264)	(0.258)
N	23,970	23,970

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Full table in Appendix C
Source: IDI. Authors own calculations

Outcomes for those suffering from disabilities is consistent with the literature. Table 7 shows that disabled people experience 30% reductions in their re-employment probabilities relative to someone who is fully abled. This is not as drastic as Forbes and Barker (2017) who observe a 43% reduction in Australia; however, a 30% reduction is still substantial. Disabled people experiencing worse re-employment probabilities is likely the result of a reduced quantity of suitable jobs, and some combination of perceived and real productivity loss due to their condition. Employers may also consider it costly to ensure that the workplace can accommodate their disability and may fear legal issues for failure to establish a suitable work environment. This may further reduce the demand for disabled workers.

As human capital literature suggests, increased education has a positive effect on re-employment probabilities in New Zealand. Hazard ratios in Table 7 suggest that improvements are greatest between larger milestones[38]. However, as also observed by Babucea et al. (2009), a master's degree results in a lower hazard ratio compared to a bachelors or honours degree[39]. This ratio increases for a PhD; however, honours graduates experience the highest re-employment probabilities. Re-employment probabilities for honours graduates are on average 66% higher than for someone with no qualifications. While this may be the result of smaller sample sizes for higher level qualifications[40], a similar effect to that described by Tansel & Taşçi (2010) may exist. The authors discuss how those with higher qualifications become more selective in their job search and apply for more specialized positions. Rather than their higher qualifications providing a human capital edge, they instead have achieved the necessary educational requirements for specialized roles. These individuals compete against candidates with similar qualifications eliminating any human capital advantage derived from their qualification.

Marriage is observed to improve re-employment probabilities as was suggested by Moore (2004) and Forbes & Barker (2017) in New Zealand and Australian studies. Marriage's effect is not ambiguous as European authors suggested (Tansel & Taşçi, 2010; D'Agostino & Mealli, 2000). Table 8 shows that those whose marriage has ended or have never been married experience re-employment probabilities that are 10% lower than someone who is married or in a civil union. Hazard ratios suggest that never being married slightly improves one's re-employment probability relative to someone whose marriage has ended. Standard errors suggest that these values may not be statistically different from each other; this is not statistically tested though.

Table 8 suggests that there is no statistically significant effect of living in an urban or rural dwelling. Wider literature observes little to no or ambiguous effects of living

[38] For example: no qualification to level 1, level 1 to level 2 (usually associated with university entrance), Level 4 to a tertiary diploma, diploma to a bachelor's degree, bachelor's to a postgraduate degree.

[39] It is not tested whether the hazard ratios are statistically lower for a masters or PhD graduate compared to a bachelors or honours graduate. Solely hazard ratio values are used to make this claim.

[40] Particularly for those with PhD's. Refer Appendix D for underlying counts.

rurally (Imbens & Lynch, 2006; Tansel & Taşçi, 2010). Observing no statistically significant effect in New Zealand is as the literature suggested.

Table 8

VARIABLE	(1)	(2)
Marital Status		
Married/Civil Union	*Reference*	*Reference*
Divorced/Widowed/ Separated	0.899***	0.899***
	(0.0249)	(0.0249)
Never Married	0.908***	0.908***
	(0.0216)	(0.0216)
Rural	1.029	1.031
(Urban=Reference)	(0.0236)	(0.0235)
Spell Count	0.989***	0.989***
	(0.00137)	(0.00136)
N	23,970	23,970
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 Full table in Appendix C Source: IDI. Authors own calculations		

The spell count variable in Table 8 captures spells on any MSD benefit. This variable is significant; additional benefit spells reduce one's re-employment probability by 1.1%, on average. This finding contradicts Tansel & Taşçi's (2010) finding that those with prior unemployment spells exhibit improved re-employment probabilities. The finding is instead consistent with Heckman & Borjas (1980). These authors discuss how employers use previous unemployment spells in a similar manner to unemployment durations in ranking models. The variable does, however, capture other benefit types. It would be interesting in future research to attempt to isolate unemployment spell counts[41]. This outcome can only be interpreted in terms of previous MSD benefit spells rather than specifically unemployment spells.

5.3 Proportional Hazard Assumption Tests

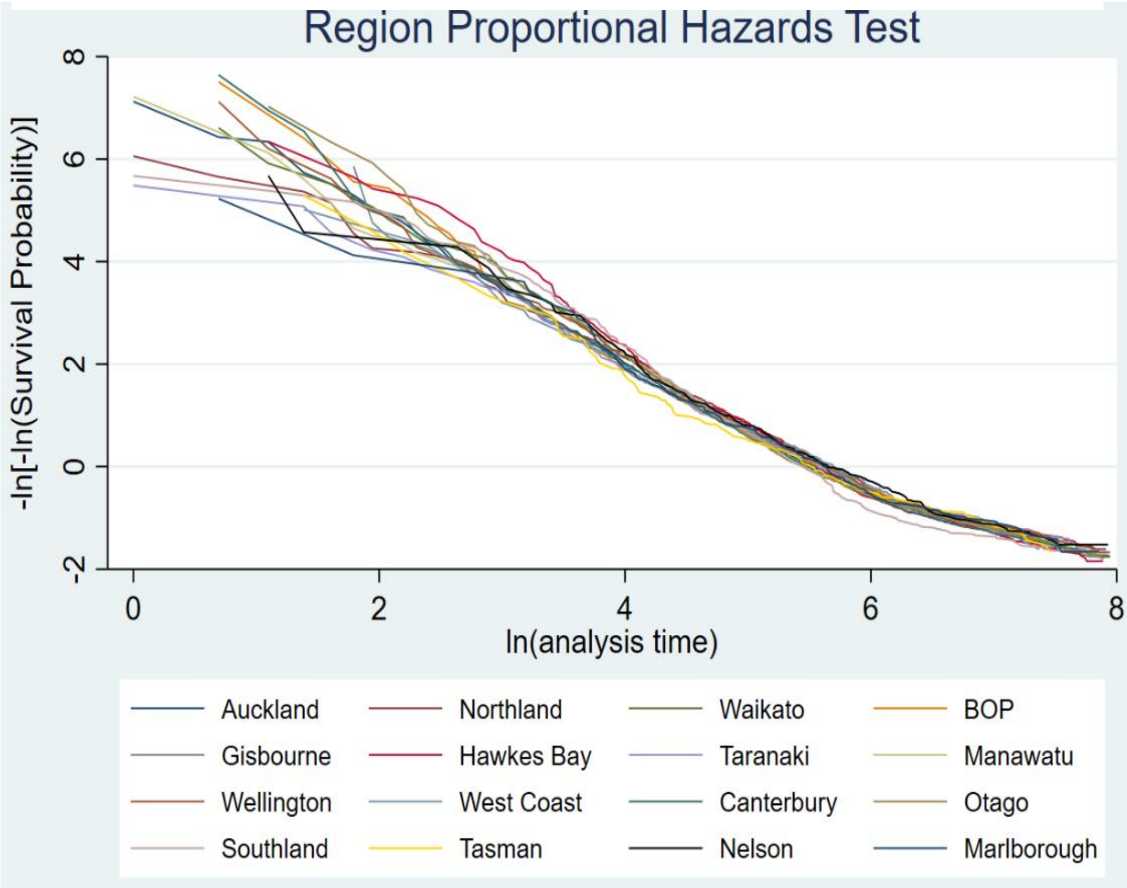
All proportional hazards models, including the Cox model, assume that the log hazard functions are proportional (Cameron & Trivedi, 2005). Variables only affect a baseline hazard by a scale factor that is independent of time. For this assumption to be true, the log-log hazard plots for each covariate must produce parallel functions (Stata, n.d.). Figure 4 plots this for each region.

The regional functions all follow a similar shape. While functions that cross do breach the proportional hazards assumption, the similarity of regional functions provides confidence that the assumption will produce an accurate estimation of the true effects of

[41] Such a variable was not easily available in the IDI.

regional treatments. Regional functions appear to best demonstrate parallel trends at longer spell durations.

Figure 4



Appendix L displays covariate proportional hazards assumption tests. All variables demonstrate similarly shaped functions. The proportional hazards assumption cannot be rejected using graphical evidence. It can be observed that a number of these functions do cross at points. This is especially prominent in variables with many categories. In these variables' cases, it is unlikely to have no crossing functions using real world data.

Graphical representation alone is not sufficient for determining the success of the proportional hazards assumption. These graphs simply provide a visual method for testing the assumption. The assumption can also be tested statistically; Table 9 reports the findings of this test. The null hypothesis is that the proportional hazards assumption does hold. A variable (or category) fails the test if the p-value suggests that the null hypothesis should be rejected.

Importantly, most regions pass the test. Only the Taranaki and Tasman regions fail at the 1% significance level; Southland and Hawkes Bay fail at the 5% level. All other regions cannot have the null hypothesis of proportional hazards rejected. This implies that regional variables are approximated well using the proportional hazards assumption.

Additional covariates exhibit a similar combination of categories that pass and fail the test. Most noticeably, the sex and disability variables both fail at the 1% significance level. Disability failing is surprising as the graphical representation implies strong proportionality exists. Only the Pacifica ethnicity and the 36–40 age range fails the test; otherwise, ethnicity and age variables cannot have the proportional hazards assumption rejected. The proportional hazards assumption can be rejected for half of the education level categories; one level fails at the 10% level, and another at the 5% level. The remaining three fail at the 1% level.

Table 9

PH Test using ln(Analysis Time)							
Variable	**rho**	**chi2**	**Prob>chi2**	**Variable**	**rho**	**chi2**	**Prob>chi2**
Auckland	.	.	.	NZ European	.	.	.
Northland	-0.0075	1.26	0.2611	Maori	0.0079	1.39	0.2382
Waikato	0.0073	1.11	0.2919	Pacifica	0.0121	2.92	0.0875
BOP	-0.0059	0.78	0.3768	Asian	-0.0029	0.19	0.6650
Gisborne	0.0033	0.21	0.6501	Other	0.0105	2.37	0.1240
Hawkes	0.0176	6.45	0.0111	Disability	0.0395	29.52	0.0000
Taranaki	-0.0177	7.86	0.0051	Married/ Civil Union	.	.	.
Manawatu	0.0002	0.00	0.9776	Divorced/ Separated/ Widowed	0.0142	4.60	0.0320
Wellington	-0.0005	0.01	0.9387	Never Married	0.0112	2.82	0.0929
West Coast	-0.0009	0.01	0.9034	No Education	.	.	.
Canterbury	-0.0027	0.17	0.6810	Level 1	-0.0066	0.92	0.3370
Otago	0.0032	0.23	0.6317	Level 2	-0.0136	3.88	0.0488
Southland	0.0166	6.15	0.0132	Level 3	-0.0192	7.70	0.0055
Tasman	-0.0169	7.91	0.0049	Level 4	-0.0239	12.43	0.0004
Nelson	-0.0034	0.30	0.5837	Level 5 Diploma	0.0003	0.00	0.9626
Marlborough	-0.0013	0.03	0.8530	Level 6 Diploma	-0.0101	2.34	0.1259
Sex	-0.0315	22.09	0.0000	Bachelor's Degree	-0.0267	16.69	0.0000
18-23	.	.	.	Postgrad/Honours	-0.0112	3.03	0.0817
24-29	-0.0005	0.01	0.9396	Masters	-0.0057	0.75	0.3873
30-35	-0.0062	0.87	0.3520	PhD	-0.0057	0.81	0.3681
36-40	-0.0139	4.51	0.0338	Rural	0.0078	1.34	0.2463
41-45	-0.0031	0.21	0.6466	Spell Count	0.0029	0.20	0.6578
46-50	-0.0020	0.09	0.7630				
51-55	-0.0017	0.06	0.8011	Global Test		151.19	0.0000
56-60	-0.0042	0.40	0.5274				

Hypothesis test	
H0	The proportional hazard assumption holds
H1	The proportional hazard assumption does not hold

Due to the high number of categories and variables, it is not surprising that the global test fails. Additional model specifications that exclude the covariates that fail the proportional hazards test still fail the global test. Proportionality is an estimation method and does not need to be perfect (Stata, n.d.); however, this does present a limitation of the Cox proportional hazards model. The model may not properly be capturing the heterogeneity of the data. Variables may also be affecting the slope as well as the vertical shift of the function. Future researchers may wish to explore fully parametric survival analysis techniques that do not require the proportional hazards assumption to be true.

5.4 Region Specific Models

A single parameter value is estimated for each covariate. For covariates to accurately control for regional demographic compositions, the parameter needs to be accurate for all regions. To test this, models are run for each individual region; only that region's sample of unemployment spells are included. Beyond assessing the effectiveness of the covariates, these specifications can provide interesting estimations of regional heterogeneity in covariates' effects on re-employment probabilities. For simplicity, age range, education, and spell count covariates are aggregated[42].

Hazard ratios in Appendix E suggest that female jobseekers are disadvantaged in each New Zealand region. Some smaller regions, however, exhibit insignificant gender effect. Insignificant covariates are commonly observed for smaller regions in this section due to small regional sample sizes. The larger Northland and Taranaki regions do, however, also exhibit insignificant sex effects. Gisborne also acts as an outlier with women 31% less likely to find work over an unemployment spell compared to men.

Age exhibits a large range of outcomes. Some regions including Gisborne, Southland, Otago, and the upper South Island regions are insignificant. Others range from Canterbury's 9% reductions in re-employment probabilities by increasing one's aggregated age range, to 24% reductions in the West Coast region. Most regions exhibit hazard ratios for age that are between 0.91–0.84.

Ethnicity tends to be insignificant at a region-specific level. The primary model suggested 6% decreases for Asian jobseekers, and 6.5% improvements for Pacifica. These are relatively small variations, and the reduced regional samples exhibit larger standard errors; insignificance is not surprising in these specifications. The same reasoning is likely why marital status is also largely insignificant. Only a few regions exhibit statistically significant ethnicity and marital status effects.

The primary model suggests that education's effect is most prominent after major milestones motivating the greater aggregation of the variable in this section. Most regions are relatively consistent, falling in the 10–20% improvement range for increasing education by one aggregated level. The Taranaki and West Coast regions act as outliers with increases of over 30% from achieving an additional aggregated education level.

[42] The new variables are summarised in Appendix G.

Rurality is insignificant in most regions; this is consistent with the primary model. Cantabrians, however, experience re-employment probabilities that are 24% higher in rural areas. Disability status is generally consistent with the primary model in individual regions. Only the Otago region deviates significantly; re-employment probabilities are on average half that of a fully abled person. Spell counts are now aggregated to capture the fact that individuals that have experienced a similar number of benefit spells are unlikely to be too dissimilar. Region specific models demonstrate that additional spells continue to have a negative effect on re-employment probabilities.

Regions with smaller underlying sample sizes generally produce insignificant covariate estimations. To negate this effect, the models have also been run for aggregated regions. The output table for these models are in Appendix F. Larger sample sizes do generate significant results that are closer to what would be expected from the full sample. Education, marital status, sex, and spell count are now closer to the primary model estimates. Ethnicity variables, however, generally remain insignificant.

Chapter 6: Robustness Checks

6.1 Adjusted Start Day

Two variables are used to estimate the start date of an unemployment spell. The last date that someone worked is used first and provides the most accurate date for an individual becoming unemployed. If this date is missing in the data, the date an individual registered for a benefit spell is used. The limitation of using two potential start dates is that the definition of an unemployment spell is inconsistent. An individual is unlikely to immediately register for a benefit upon job loss creating a disparity between unemployment spells and benefit spell durations. Section 225.1 of the *Social Security Act 2018* further enforces a 13-week non-entitlement period. Anyone who becomes voluntarily unemployed without sufficient reason or is dismissed for misconduct cannot register for a spell for 13 weeks after becoming unemployed. This creates additional uncertainty as to whether the registration date accurately captures the start of an individual's unemployment spell.

Models (3) and (4) in Appendix C are iterations of Models (1) and (2) where the start of a spell is redefined as solely the last date that someone worked; Table 10 summarises the regional findings. As this date is often missing, the sample size almost halves[43]. Using a consistent spell start date definition I can verify the findings of the primary model. Region variables under this new specification do not deviate in their directional effect on re-employment probabilities from Table 5. This suggests that the start of a spell does not lose accuracy when including registration dates in the definition.

Some hazard ratios do, however, deviate in magnitude from those in the primary model. The Northland region now is only significant at the 10%, rather than 1%, level and the Bay of Plenty region gains significance. The Bay of Plenty now experience 8.5% re-employment probability improvements relative to Auckland, compared to 5% improvements in the primary model. Under this specification, the Hawkes Bay, Otago, and Southland regions exhibit greater re-employment probabilities compared to Model (1). Model (4) exhibits similar outcomes. This is as would be expected due to aggregated regions being derived from the individual regions.

[43] Motivating my inclusion of registration dates to create the spell start date variable.

Table 10

(3) Region		(4) Aggregated Region	
Region		**Aggregated Region**	
Auckland	*Reference*	*Auckland*	*Reference*
Northland	0.914*	Northland	0.912*
	(0.0479)		(0.0477)
Waikato	0.996	Waikato	0.994
	(0.0329)		(0.0329)
BOP	1.085**	BOP	1.083**
	(0.0395)		(0.0393)
Gisborne	1.054	East Coast	1.104***
	(0.0742)		(0.0393)
Hawkes	1.121***	Taranaki-Manawatu	0.963
	(0.0430)		(0.0328)
Taranaki	0.993	Wellington	0.916***
	(0.0613)		(0.0294)
Manawatu	0.955	West Coast-Tasman	1.016
	(0.0358)		(0.0488)
Wellington	0.916***	Canterbury	1.143***
	(0.0294)		(0.0397)
West Coast	0.981	Southland	1.250***
	(0.0942)		(0.0418)
Canterbury	1.144***		
	(0.0398)		
Otago	1.154***		
	(0.0501)		
Southland	1.376***		
	(0.0572)		
Tasman	1.164		
	(0.118)		
Nelson	0.954		
	(0.0716)		
Marlborough	1.003		
	(0.0927)		
N	13,908	*N*	13,908

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Full table in Appendix C
Source: IDI. Authors own calculations

6.2 Sample Period Test

A wide sample period from 01/11/2012 to 31/07/2013 was selected to ensure that the sample size is sufficiently large enough to draw valid conclusions. It must be assessed whether the sample period ensures that individuals remain comparable; the date that someone becomes unemployed cannot influence their re-employment probability. As Appendix A shows, New Zealand does experience some variation in key economic indicators over the sample period. This variation is not large compared to the wider graphs; it can be hypothesised that economic conditions remained relatively homogeneous during the sample period. To provide further descriptive evidence, the correlation between the start date of a spell and the duration of a spell is calculated to be -0.0419. This implies that there is only a very weak negative correlation.

So far, it can be hypothesised that the date an individual becomes unemployed within the sample period does not drastically affect their re-employment probability[44]. To further test this hypothesis a survival analysis is run that includes a covariate capturing the date an individual became unemployed. This is the same start date variable that is used to calculate an individual's unemployment duration. Model (1) in Table 11 runs this specification; Model (2) aggregates start dates into categories[45].The full estimation table is available in Appendix H. If start date variables are statistically significant then it cannot be concluded that a spell's start date does not affect an individual's re-employment probability.

Model (1) suggests that re-employment probabilities are unaffected by the start day of a spell. Model (2) concurs this finding in the first 100 days and the 21/05/13–09/07/13 range of the sample period. The remaining ranges of the sample period exhibit re-employment probability decreases of between 5–9%. The second 100 days of the sample period, do not appear to be statistically different from each other, however, this is not statistically tested. Considering that the first 100 days are during the New Zealand summer, seasonal effects may partially explain this outcome. In 2013, the agricultural industry contributed to 6.45% of national employment (Statista, 2021). At the 2013 census, agriculture, forestry, and fisheries were New Zealand's seventh largest employing sector (Figure.NZ, 2013). These industries are seasonally dependent and are particularly labour intensive in spring/summer months. In other months, these industries[46] reduce their labour demand, possibly contributing to the disparities seen in Table 11.

Model (1) and the correlation between spell duration and spell start dates suggest that the sample period is successful. Re-employment probabilities are not drastically affected by the date that an individual loses their job. Model (2) does produce some doubt as some periods experience decreased re-employment probabilities. Appendix A suggests that this is not unexpected due to variation existing across economic indicators. Any

---

[44] E.g., seasonal effects did not significantly affect re-employment probabilities, or large economic variation did not occur during the sample period.

[45] Appendix G displays the count of spells that begin in each aggregated start day range for Model (2)

[46] Agriculture is simply an example. Other industries e.g., tourism is likely to exhibit increased labour demand in summer months.

variation over the sample period reduces homogeneity in the sample and potentially creates some level of bias. While variation is not ideal, the sample needs to be large enough to minimise standard errors. This results in a trade-off between desiring a smaller sample period and a larger sample size. Further research may be needed to identify and control for seasonal effects to minimise this variation.

Table 11

	(1) Start Day	(2) Start Range
Spell Start Day	1.000 (8.49e-05)	
Start Day Range		
(01/11/12 - 20/12/12)		*Reference*
21/12/12 - 08/02/13		0.987 (0.0228)
09/02/13 - 31/03/13		0.945*** (0.0202)
01/04/13 - 20/05/13		0.958** (0.0202)
21/05/13 - 09/07/13		1.019 (0.0224)
10/07/13 - 31/07/13		0.913*** (0.0237)
N	23,970	23,970
Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1 Full table in Appendix H Source: IDI. Authors own calculations		

6.3 Additional Regional Specifications

Two additional specifications are used to analyse further geographic re-employment probability variation in New Zealand. Model (3) in Table 12 replaces the region variable with a dummy variable capturing whether an individual resides in a major New Zealand centre at the beginning of their unemployment spell. Model (4) replaces the region variable with a dummy variable representing whether an individual resides in the North or South Island of New Zealand. The full output table is available in Appendix H.

A major centre is defined as residing at an urban address[47] in a region that contains New Zealand's 5 largest cities[48]. The major centre variable is a dummy variable with a reference category of not living in a major centre by this definition. The hazard ratio in

[47] An individual is identified to live in an urban setting using the urban/rural covariate derived from 2013 census data.

[48] The Auckland, Waikato, Wellington, Canterbury, and Otago regions, aiming to capture jobseekers in the Auckland, Hamilton, Wellington, Christchurch, and Dunedin cities.

Table 12 shows that living in a major centre is detrimental to an individual's job search. These individuals experience re-employment probabilities that are on average 3% lower than those not in a major centre.

Table 12

VARIABLES	(3) Major Centre	(4) NZ Island
Major Centre *(non-major centre=Reference)*	0.968** (0.0154)	
NZ Island *(North Island=Reference)*		1.143*** (0.0196)
N	23,970	23,970
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 Full table in Appendix H Source: IDI. Authors own calculations		

Model (4) in Table 12 replaces the region variable with whether an individual resides in the North or South Island of New Zealand. The reference category is the North Island. Those residing in the South Island experience re-employment probabilities that are 14% larger than in the North Island. This outcome was hypothesised in Chapter 2.4 based on the unemployment rates in Table 2. Forbes & Barker (2017), Imbens & Lynch (2006), and D'Agostino & Mealli (2000) discussed how regions with lower unemployment rates experience higher re-employment probabilities; the expected outcome is observed in New Zealand.

6.4 Compare Region at End to Start

Regional assignment is determined by the region associated with an individual's residential address on the day that they became unemployed. As Appendix D shows, mean spell lengths are approximately 290 days, with a standard deviation that suggests wide duration variability[49]. It is plausible that some individuals will move to a new region while unemployed. These individuals likely would have engaged in a job search in this new region. Individuals that move may differ in their unobservable characteristics; these may include motivation levels, search intensities, and human capital composition. This section does not aim to identify the specific unobservable characteristics affecting variations between those who do and do not move. The aim is simply to identify whether re-employment probabilities are different between the groups.

The same method for identifying individual regional assignments at the start of a spell is used to identify regional assignments at the end. As Appendix G shows, over 3000

[49] Also refer Appendix K.

people did in fact move to a different region by the end of their unemployment spell. The full estimation table can be found in Appendix I. Model (1) establishes reference hazard ratios by running a simplified specification of the primary model based on regional assignment at the start of a spell[50]. Hazard ratios are not exact matches to those in Appendix C but are close estimates under a simplified model specification that is also used in models (2) and (3). Model (2) instead assigns individuals to the region that they resided in at the end of their unemployment spell. Model (3) only includes individuals who did not move region; re-employment probabilities are therefore only affected by one regional labour market.

Table 13

VARIABLES	(1) Region at Spell Start	(SE)	(2) Region at Spell End	(SE)	(3) Region if Didn't Move	(SE)
Region						
Auckland	Reference		Reference		Reference	
Northland	0.899***	(0.0344)	0.851***	(0.0320)	0.891***	(0.0374)
Waikato	1.003	(0.0249)	0.993	(0.0245)	1.014	(0.0270)
BOP	1.045	(0.0292)	1.040	(0.0291)	1.046	(0.0317)
Gisborne	0.959	(0.0501)	0.913*	(0.0478)	0.958	(0.0548)
Hawkes	1.079**	(0.0350)	1.085**	(0.0363)	1.082**	(0.0378)
Taranaki	1.009	(0.0514)	1.013	(0.0500)	1.007	(0.0547)
Manawatu	0.943*	(0.0283)	0.925***	(0.0280)	0.946*	(0.0312)
Wellington	0.898***	(0.0217)	0.882***	(0.0213)	0.888***	(0.0228)
West Coast	0.976	(0.0734)	0.979	(0.0712)	1.005	(0.0866)
Canterbury	1.139***	(0.0305)	1.129***	(0.0305)	1.150***	(0.0329)
Otago	1.063*	(0.0359)	1.048	(0.0366)	1.062	(0.0393)
Southland	1.303***	(0.0483)	1.261***	(0.0456)	1.297***	(0.0491)
Tasman	1.140	(0.0933)	1.081	(0.0988)	1.230**	(0.127)
Nelson	0.927	(0.0608)	0.945	(0.0602)	0.964	(0.0739)
Marlborough	1.075	(0.0793)	1.101	(0.0854)	1.098	(0.0973)
N	23,970		23,829		20,916	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Full table in Appendix I

Model (2) in Table 13 does not show any drastic changes from Model (1). This is expected as only a small proportion of the sample move to a different region. Gisborne gains a slight level of significance with a 9% reduction in re-employment probabilities compared to Auckland; conversely, Otago loses its statistical significance. A 7.5%

[50] This model differs from Model (1) in Appendix C/Chapter 5.2 in that it aggregates to a greater degree. Hazard ratio therefore differ slightly in magnitude. Model (1) is run so that Models (2) and (3) can be better compared to the primary model by using the same, simpler, covariate specifications.

reduction in the Manawatu is now highly statistically significant based on regional assignment at the end of a spell. Each of these changes is only marginal.

Model (3) only analyses those who do not change their regional assignment. Most regional hazard ratios do not drastically deviate from the hazard ratios in Model (1). Tasman is an exception. Those in the Tasman region who do not move for work are 23% more likely to find work at any spell duration than an Aucklander who remains in the Auckland region. Tasman was insignificant in Model (1).

So far, little variation can be concluded between those who do and do not move in terms of hazard ratios. The models in Table 13 show that regional variations do not change drastically when regions are respecified as the region someone ends their job search in, or the region someone stays in. Model (5) in Appendix H, however, replaces the region variable with a dummy variable representing if an individual moved. This model identifies that those who do not move region exhibit a 24% greater re-employment probability, on average, compared to someone who does move. This implies that regional variations remain relatively constant for those who move and do not, however those who do move are negatively impacted overall. Further research is needed to identify the mechanisms behind this result.

6.5 Macro model

Many authors discuss replacing region variables with macroeconomic indicators (Forbes & Barker, 2017; Babucea et al., 2009; Imbens & Lynch, 2006; Kroft et al., 2013; Moore, 2006; D'Agostino & Mealli, 2000). Controlling for regional demographic variation with covariates ensure that regional estimates isolate socio-economic effects. Replacing regional variables with socio-economic variables allows for the identification of the specific mechanisms causing regional variation. Understanding the underlying mechanisms behind regional variation provides a basis for estimating regional effects in future socio-economic contexts. Three model specifications are run using two variables in each. The three models are run using one of the unemployment rate, the underutilization rate, and GDP per capita. Industry is included as the second variable in each model. Industry was identified to be significant in New Zealand by Moore (2006). Industry captures a region's primary industry in terms of contribution to regional GDP. Regional economic data is the same as in Table 2; this data is derived from MBIE's Regional Economic Activity Web Tool. Hazard ratios for these mechanisms are reported in Table 14. Full model estimations can be found in Appendix J.

Literature suggests that at higher regional unemployment rates, re-employment probabilities should decline (Forbes & Barker, 2017; Imbens & Lynch, 2006; Kroft et al., 2013; D'Agostino & Mealli, 2000). The same outcome is observed in New Zealand. A one percentage point increase in a regions unemployment rate is, on average, associated with a 3% reduction in re-employment probabilities in that region. This outcome is consistent with observations in the literature. When the unemployment rate increases by one percentage point, a 3% probability decrease is observed in the USA (Imbens & Lynch, 2006) and a 4% decrease is observed in Australia (Forbes & Barker, 2017).

Table 14

VARIABLES	(1)	(2)	(3)
Regional Indicators			
Unemployment Rate	0.971***		
	(0.00618)		
Underutilization Rate		0.982***	
		(0.00337)	
GDP per capita			1.002**
(In thousands)			(0.00103)
Industry			
Manufacturing	*Reference*	*Reference*	*Reference*
Natural Resources	0.947	0.952	0.946
	(0.0316)	(0.0316)	(0.0325)
Professional	0.879***	0.869***	0.855***
	(0.0193)	(0.0192)	(0.0232)
Construction	0.990	0.992	1.044
	(0.0333)	(0.0327)	(0.0332)
Agriculture	1.191***	1.187***	1.264***
	(0.0457)	(0.0448)	(0.0449)
N	23,970	23,970	23,970

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Full table in Appendix J
Sources: MBIE Regional Economic Activity Web Tool and IDI. Authors own calculations

Model (2) identifies that the unemployment rate has a greater effect than underutilization; a one percentage point increase in the underutilization rate is associated with only a 2% decrease in re-employment probabilities. Underutilization captures both those who are unemployed and underemployed. Underutilization having a smaller negative effect than unemployment implies that those who are part-time employed are more likely to find further employment than the unemployed are at finding a job.

Model (3) shows that GDP per capita has very little impact on re-employment probabilities. There is a small positive effect; a $1000 increase in GDP per capita results in a 0.02% increase in one's re-employment probability. This effect can be concluded as being essentially negligible.

Industry variables are used more as a control. Table 2 shows that there is little variation in primary industry; most regions are classed as manufacturing regions. Improved probabilities for the agricultural industry are likely highly correlated to the Southland region having the highest re-employment probabilities. A similar observation is made for the professional industry which is the primary industry in Wellington; Wellington exhibits the lowest regional re-employment probabilities. Only one region is assigned to these industries, likely biasing the hazard ratios. Future researchers may wish to use a micro-level industry variable to gain greater variation to minimise the uncertainty associated with these estimations.

The key outcome of these models is that New Zealand does align with the wider literature regarding the negative effect of higher unemployment rates on re-employment probabilities. It is also identified that underutilization can be used in a similar manner. Economic intuition would suggest these outcomes. Higher unemployment is generally indicative of a surplus labour supply. Therefore, more individuals are competing for an insufficient number of jobs. As a result, it is less likely that an individual will rank high enough to receive one of the limited available jobs.

Chapter 7: Discussion

7.1 Implications

Regional re-employment probability variation is observed to exist in New Zealand. Particularly the Wellington and Northland regions exhibit the lowest probabilities of returning to work. Most hazard ratios are greater than one indicating that the Auckland region is also underperforming compared to other New Zealand regions. It is further observed that South Island regions outperform those in the North Island; re-employment probabilities are on average 14% higher in the South Island. Three in four New Zealanders live in the underperforming North Island with the median centre of population moving further northward since 1921 (Stats NZ, 2017). In 2020, the Auckland, Wellington, Northland, and Waikato (which is not statistically different from Auckland) regions accounted for approximately three-fifths of New Zealand's population (MBIE, 2021). The majority of the nation's population resides in regions that disadvantage their job search.

Tansel & Taşçi (2010) observe that successful regionally targeted policy can drastically improve re-employment probabilities in that region. The GAP programme in Turkey generated the sole statistically different region; that region experienced the highest re-employment probabilities. Policy is needed to address New Zealand's weaker labour markets; this is magnified by the fact that a large proportion of New Zealand's population resides in these regions.

Regional variables in this study only capture the socio-economic mechanisms affecting re-employment probabilities for spells that begin between the 1st of November 2012 to the 31st of July 2013. Unemployment rate variation can, however, be used to estimate regional re-employment probability variations for future socio-economic contexts. I identify that a one percentage point increase in a region's unemployment rate reduces re-employment probabilities in that region by 3% on average. Similar findings are observed in the wider literature[51]. Policy makers can use this estimation method when formulating future labour market policy. Further research into other mechanisms, including observing micro-level industry variables, would allow for even greater accuracy in future estimations without the need to repeat a full survival analysis.

Underutilization is another possible mechanism for estimating future regional variations. On average, a one percentage point increase in the underutilization rate decreases re-employment probabilities by 2%. The Covid-19 pandemic has adversely

[51] Refer Chapter 2.3

affected underutilization more than unemployment due to the success of New Zealand's Wage Subsidy scheme (RBNZ, 2021). Unemployment as of March 2021 is only 0.6 percentage points above pre-Covid levels [52] (Stats NZ, 2021b), and labour force participation has only fallen by 0.3 percentage points between March of 2020 and 2021 (Stats NZ, 2021c). Underutilization, however, is 2.1 percentage points above pre-Covid levels (Stats NZ, 2021d). The hazard ratio for underutilization may better estimate re-employment probability variation to inform policy responses to Covid-19.

As suggested by Moore (2006), ethnicity is included, and unexpected results are observed. Imbens & Lynch (2006), Forbes & Barker (2017), Kroft et al. (2013), and Moore (2004) find that European/White ethnicities experience the highest re-employment probabilities. Moore (2004) specifically identifies that Pacifica experience reduced re-employment probabilities, while NZ Europeans experience the highest probabilities in New Zealand. I instead observe that Pacifica experience the highest re-employment probabilities and that Maori and 'Other' ethnicities are not statistically different from NZ Europeans. The Pacifica category is, however, observed to breach the proportional hazards assumption in Table 9; failing this test may bias the hazard ratio upwards. Future researchers may still find it interesting to test this outcome further to identify the mechanisms behind this surprising outcome for ethnicity.

Hazard functions in Figure 3 suggest a parabolic duration dependence exists in New Zealand. Re-employment probabilities tend to improve over approximately the first year of an unemployment spell. After the first year, duration dependence turns negative and re-employment probabilities start to decline. The OECD (2016) defines an individual as being long-term unemployed after one year of unemployment; negative duration dependence appears to manifest as a spell approaches the long-term unemployment threshold. Policy makers therefore do not require significant urgency to address shorter spells. These spells may extend due to the time required to appropriately match a worker to a firm; these individuals are of little policy significance. Policy is most necessary to assist the long-term unemployed into work as their re-employment probability is declining as their spell extends. While these individuals are out of work, the economy is operating under a deadweight loss (Lucas, 1978), and as their spell extends, they are less likely to return to work.

Policy can minimise the prevalence of long-term unemployment and reduce the strength of negative duration dependence experienced by the long-term unemployed.

[52] Where pre-Covid levels refers to data reported in December 2019.

Literature suggests that new policies should address any skills gap in the labour supply (Lindsay & Sturgeon, 2003), incentivise private sector job creation, and incentivise job search (Kluve, 2010). Long-term unemployment accentuates any skills gap due to human capital depreciation while unemployed (Eriksson & Rooth, 2014; Pavoni & Violante, 2007). Lindsay & Sturgeon (2003) state that policy makers must collaborate with local employers to provide work placement programmes. Such programmes bridge employment history gaps and provide opportunities to develop relevant skills, reducing any skills gap in the labour supply. Kluve (2010) also discusses the need to incentivise job search and job creation. Policies to incentivise job creation includes wage subsidies and fiscal policy (e.g., building new infrastructure). Policies that incentivise job search include funding easy-to-use job search websites/services and imposing benefit reductions for insufficient search intensities. These authors discussions are not specific to a New Zealand context. They act as policy suggestions for addressing long-term unemployment based on international evidence.

7.2 Limitations

Using benefit spells to identify unemployment spells presents some limitations. Firstly, the scope of the analysis is narrowed. Those who are ineligible to receive a benefit are excluded from the sample. Predominantly, this excludes those whose remaining household income results in an abatement that nullifies the benefit. Future researchers may wish to identify unemployed individuals using methods that do not restrict the research scope in such a way.

Self-selection bias exists in the sample too. Individuals must register for a benefit spell. The sample excludes those who may expect a short spell as they have already accepted a new position or anyone who fears the stigma associated with a benefit spell. Additionally, individuals may self-select into the region that they reside. Some individuals can be assumed to have been randomly assigned into a region, in that they remain in the region that they were born. Others may have explicitly chosen to live in a specific region. Self-selection distorts the homogeneity of both the sample as a whole and the regional samples; in other words, they are not representative of the full population (Lavrakas, 2008). As a result, bias may exist in the regional hazard ratio estimations. Covariates aim to minimise heterogeneity that exists in the sample; however, it cannot be eliminated completely.

Regional structural differences, particularly regarding industry composition, may minimise cross-region comparability. An individual may reside in a region with a poor hazard ratio; however, their specific industry may experience the greatest re-employment probabilities in that region. Relocating to another region may reduce their re-employment probability even if the hazard ratio would suggest the opposite effect. Future researchers may wish to include a micro-level variable to capture an individual's pre-unemployment industry. This could be interacted with region variables to identify which industries experience the greatest re-employment probabilities in each region.

Benefit spells require that an individual be a participant in the labour force. As a result, benefit eligibility ends when an individual exits the labour force. Overseas travel, education, and death records are used in conjuncture with wages and salary data to eliminate those exiting the labour force from the sample. Future researchers may wish to introduce additional data to ensure that the end of a benefit spell represents a transition into employment. Health, criminal, and child-birth records are possible examples of other non-employment transitions that may terminate a benefit spell.

The model does not capture whether an individual is rejecting job offers. Mortensen (1970) and McCall (1970) both discuss how wage expectations affect an individual's decision to accept or reject a job offer. A job may be offered to an individual who does not deem the compensation to be satisfactory and rejects the position. This research does not distinguish between someone remaining unemployed due to not receiving job offers or because they are rejecting unattractive offers. Future researchers may wish to disentangle the demand and supply side effects on re-employment probabilities. Deficiencies on either side require a different policy response, creating an interesting direction for future research.

Labour markets are dynamic; the re-employment probability estimations in this research are accurate for spells that began in the period of 01/11/2012–31/07/2013. Identifying the effect of unemployment rates does provide a basis for estimating probability variations under future contexts. It would, however, be useful for future researchers to conduct a similar study using another sample period. In particular, after sufficient time to develop a suitable sample of re-employment transitions, the effects of the Covid-19 pandemic would present an interesting scope for future iterations of this research. Post-Covid research would allow for the analysis of each regional labour market's ability to respond to an economic shock.

Finally, using the proportional hazards assumption presents additional limitations. Firstly, it is only an approximation method due to the assumption being imperfect when using real world data. Secondly, the model requires additional estimation measures to handle discrete data. Uncertainty is created as the Breslow method—which handles failure ties due to discrete data—has been observed to underestimate parameter values (Hertz-Picciotto & Rockhill, 1997). Thirdly, proportional hazards models are incapable of accurately observing duration dependence variation. This limitation further persists as it is observed in Chapter 5.3 that some variables do not exhibit proportion hazards at all.

To resolve the limitations of using a proportional hazards model, future studies may wish to apply parametric techniques. These require the specification of the distribution of survival times; this is difficult to specify accurately, and parametric models are not robust to misspecification (Columbia Public Health, 2021). Parametric models are uncommon in the literature because of this difficulty. A discrete data parametric model would allow for estimations of duration dependence and would overcome the limitations of the Breslow method. Future researchers should consider attempting a parametric survival analysis in New Zealand.

Chapter 8: Conclusion

Using survival analysis techniques, I have analysed the factors that affect an individual's re-employment probability with a focus on regional variations in New Zealand. An individual that receives either the New Zealand Unemployment or Jobseeker Support benefit is taken as being analogous to experiencing an unemployment spell. Dates representing the first and last day of a benefit spell are used to estimate an individual's unemployment duration. Covariates are included to control for demographic variations to isolate socio-economic mechanisms of regional re-employment probability variation. This ensures that estimations can be interpreted as the change in an individual's re-employment probability if they begin a job search in another region. These covariates' effect on re-employment probabilities can also be estimated due to their inclusion in the model.

The primary contribution of this research to the New Zealand literature is the estimation of re-employment probability variations for each region. Earlier researchers only briefly discuss regional effects while this research provides a comprehensive analysis. Most notably, Southland and Canterbury experience the highest re-employment probabilities that are on average 32% and 15% greater than for a jobseeker in Auckland, respectively. Wellington and Northland experience the worst re-employment probabilities; both regions' probabilities are on average 10% lower than for a jobseeker in Auckland. South Island regions also outperform North Island regions; South Island jobseekers are on average 14% more likely to return to work at any spell duration than a jobseeker in the North Island.

A further contribution is the estimation of the effect of unemployment rate variation on re-employment probability variation. A one percentage point increase in a region's unemployment rate is associated with a 3% reduction in re-employment probabilities. The underutilization rate can also be used; a one percentage point increase reduces re-employment probabilities by 2%. Regional economies and labour markets are dynamic; future regional variations will unlikely match those which have been identified here. Knowledge of these relationships allow future regional re-employment probability variations to be estimated.

New Zealand is observed to exhibit a parabolic duration dependence; this differs from the strictly negative duration dependence discussed in global literature. New Zealand jobseekers are observed to enjoy positive duration dependence in approximately the first year of their unemployment spell. Duration dependence turns negative as a jobseeker approaches long-term unemployment. Policy makers should therefore be most concerned with addressing the long-term unemployed; those experiencing shorter unemployment spells are more likely to return to work as their spell extends in the first year.

Previous literature has analysed covariate effects in New Zealand and abroad. My findings regarding these effects generally align with these sources. Male, higher educated, married, fully abled, and younger jobseekers all experience greater probabilities of finding work. However, ethnicity is seen to deviate. Higher re-employment probabilities are not observed for those of NZ European decent as literature would suggest. Instead, Pacifica are on average 6.5% more likely to return to work over the duration of a spell compared to NZ Europeans. It is similarly surprising that Maori experience re-employment probabilities that are not statistically different from NZ Europeans. This deviation from the literature presents an interesting area for future researchers to investigate.

The primary motivation to fill these gaps in the New Zealand literature is the policy implications. It has now been identified for policy makers which regional labour markets require targeted policy to improve the likelihood that jobseekers in that region return to work. Tansel & Taşçi (2010) observed how region-specific policy can improve regional re-employment probabilities; it is possible to achieve similar improvements in New Zealand. Covariate controls further identify for policy makers which demographics are experiencing the lowest re-employment probabilities. Observing parabolic duration dependence further demonstrates the importance of focusing policy on long-term unemployment. Unemployment adversely affects wellbeing, not just for the unemployed, but for their families and communities. Policy makers can use the evidence in this research to formulate policy that addresses poorer labour market success rates to achieve full employment targets. Such policy will improve labour market outcomes and wellbeing for some of the most vulnerable people and communities in New Zealand.

References

Acemoglu, D. (1995). Public Policy in a Model of Long-Term Unemployment. *Economica, 62(246), 161–178.* https://doi.org/10.2307/2554901

Atkinson, J., Salmond, C., & Crampton, P. (2014). NZDep2013 Index of Deprivation. *Dunedin: University of Otago. Ministry of Health.* health.govt.nz/publication/nzdep2013-index-deprivation

Babucea, A., Bićanić, I., Bohm, B., Boršič, D., Dănăcică, D., Kavkler, A., Tevdovski, D., & Toševska, K. (2009). Cox Regression Models for Unemployment Duration in Romania, Austria, Slovenia, Croatia, and Macedonia. *Romanian Journal of Economic Forecasting, 10(2), 81–104.*

Bilgen, A. (2020). Turkey's Southeastern Anatolia Project (GAP): A Qualitative Review of the Literature. *British Journal of Middle Eastern Studies, 47(4), 652–671.* https://doi.org/10.1080/13530194.2018.1549978

Blanchard, O., & Diamond, P. (1994). Ranking, Unemployment Duration, and Wages. *Review of Economic Studies, 61(3), 417–434.* https://doi.org/10.2307/2297897

Brigden, A., & Thomas, J. (2003). What does economic theory tell us about labour market tightness?. *Bank of England working papers 185,* Bank of England.

Cameron, C., & Trivedi, P. (2005). *Microeconometrics: Methods and Applications.* Cambridge University Press. Chapter 17, 573–608.

Charan, R. (2020). The Cox Proportional Hazards Model: A Regression Model for Survival Data. *Towards Data Science.* https://towardsdatascience.com/the-cox-proportional-hazards-mode l-35e60e554d8f

Ciucă, V., & Matei, M. (2011). Survival Rates in Unemployment. *International Journal of Mathematical Models and Methods in Applied Sciences, 2(5), 362–370*

Coles, M., & Smith, E. (1998). Marketplaces and Matching. *International Economic Review, 39(1), 239–254.* https://doi.org/10.2307/2527239

Columbia Public Health. (2021). Time-To-Event Data Analytics. *Population Health Methods.* https://www.publichealth.columbia.edu/research/population-health-methods/time-event-data-analysis

Community Law. (2020). *Benefit Rates: How Much You'll Get, And How Much You Can Earn: How Earning Money Will Affect Your Benefit ("Abatement").* https://communitylaw.org.nz/community-law-manual/chapter-23-dealing-with-work-and-income/benefit-rates-how-much-youll-get-and-how-much-you-can-earn/how-earning-money-will-affect-your-benefit-abatement/#:~:text=you%20can%20earn%20up%20to,benefit%20(or%20incentive%20payments).

D'Agostino, A., & Mealli, F. (2000). *Modelling short unemployment in Europe.* ISER Working Paper Series, No. 2000-06, University of Essex, Institute for Social and Economic Research (ISER), Colchester.

Dixon, S., & Stillman, S. (2008). The Impact of Firm Closure on Workers' Future Employment and Earnings. *Statistics New Zealand.*

Education Counts. (2021). Ethnic Group Codes. *Ministry of Education.* https://www.educationcounts.govt.nz/data-services/code-sets-and-classifications/ethnic_group_codes

Eriksson, S., & Rooth, D. (2014). Do Employers Use Unemployment as a Sorting Criterion When Hiring? Evidence from a Field Experiment. *American Economic Review, 104(3), 1014–1039.* https://dx.doi.org/10.1257/qer.104.3.1014

Eubanks, J., & Wiczer, D. (2016). *Duration Dependence and Composition in Unemployment Spells.* Federal Reserve Bank of St. Louis Review.

Fernández-Blanco, J., & Preugschat, E. (2018). On the Effects of Ranking by Unemployment Duration. *European Economic Review, 104 (2018), 92–110*. https://doi.org/10.1016/j.euroecorev.2018. 02.003

Figure.NZ. (2013). *Industries of Employment for People Residing in New Zealand: 2013 Census, Number of People Employed.* https://figure.nz/chart/RRCkp3KercDFyFfW-wFX9PMsXVqbEW70d

Forbes, M., & Barer, A. (2017). Local Labour Markets and Unemployment Duration. *Economic Record, 93(301), 238–254*. https://doi.org/10.1111/1475-4932.12318

Hane-Weijman, E., Eriksson, R., & Henning, M. (2018). Returning to Work: Regional Determinants of Re-employment After Major Redundancies. *Regional Studies, 52(6), 768–780*. https://doi.org/10.1080/00343404.2017.1395006

Heckman, J. & Borjas, G. (1980) Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence. *Economica, 47(187), 247–283.* https://doi. org/10.2307/2553150

Hertz-Picciotto, I. & Rockhill, B. (1997). Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression. *Biometrics, 53(3), 1151–1156.* https://doi.org/10.2307/2533573

Hyslop, D., Stillman, S., & Crichton, S. (2004). The Impact of Employment Experiences and Benefit Spell Duration on Benefit-to-Work Transitions. *Statistics New Zealand. LEED Project Technical Paper.*

Imbens, G., & Lynch, L. (2006). Re-employment Probabilities over the Business Cycle. *Portuguese Economic Journal, 5(2), 111–134.* https://doi.org/10.1007/s10258-006-0006-5

Inland Revenue Department (IRD). (n.d.). Who We Are. *About Us.* https://www.ird.govt. nz/international-tax/who-we-are#:~:text=We%20are%20a%20lean%2C%20multidisciplinary,with%20New%20Zealand%20tax%20law.

International Labour Organisation (ILO). (n.d.) *Unemployment Rate.* https://www.ilo.org /ilostat-files/Documents/description_UR_EN.pdf.

Jarosch, G., & Pilossoph, L. (2018). Statistical Discrimination and Duration Dependence in the Job Finding Rate. *Review of Economic Studies*.

Kauermann, G., & Khomski, P. (2009). Full Time or Part Time Re-employment: A Competing Risk Model with Frailties and Smooth Effects Using a Penalty Based Approach. *Journal of Computational and Graphical Statistics, 18(1), 106–125.* https://doi.org/10.1198/jcgs.2009.0007

Kluve, J. (2010). The Effectiveness of European Active Labor Market Programs. *Labour Economics, 17(6), 904-918.* Htttps://doi.org/10.1016/j.labeco.2010.02.004

Kroft, K., Lange, F., & Notowidigdo, M. *(2013).* Duration Dependence and Labour Market Conditions: Evidence from a Field Experiment. *The Quarterly Journal of Economics, 128(3), 1123–1167.* https://doi.org/10.1093/qje/qjt015

Lavrakas, P. (2008). *Encyclopaedia of Survey Research Methods*. Sage Research Methods. https://dx.doi.org/10.4135/9781412963947.n526

Lindsay, C. & Sturgeon, G. (2003). Local Responses to Long-Term Unemployment: Delivering Access to Employment in Edinburgh. *Local Economy, 18(2), 159-173. https://doi.org/10.1080/0269094032000061387*

Lockwood, B. (1991). Information Externalities in the Labour Market and the Duration of Unemployment. *The Review of Economic Studies, 58(4), 733–753.* https://doi.org/10.2307/2297830

Lucas, R. (1978). Unemployment Policy. *The American Economic Review, 68(2), 353–357. Papers and Proceedings of the Nineteenth Annual Meeting of the American Economic Association.*

McCall, J. (1970). Economics of Information and Job Search. *The Quarterly Journal of Economics, 84(1), 113–126.* https://doi.org/10.2307/1879403

Melin, M. (2001). Education – A Way Out of Poverty? *Swedish International Development Cooperation Agency.* New Education Division Documents no. 12. https://www.sida.se/contentassets/eb19ecb36def443f911678ca5cbffeed/13405.pdf#page=14

Ministry of Business Innovation and Employment (MBIE). (2019). Unemployment Rate in New Zealand. *Regional Economic Activity Web Tool.* https://www.mbie.govt.nz/business-and-employment/economic-development/regional-economic-development/activity-tools/web-tool/

Ministry of Business Innovation and Employment (MBIE). (2020a). Maori Labour Market Trends- December 2020. https://www.mbie.govt.nz/business-and-employment/employment-and-skills/labour-market-reports-data-and-analysis/other-labour-market-reports/maori-labour-market-trends/

Ministry of Business Innovation and Employment (MBIE). (2020b). Pacific Peoples in the Labour Market- December 2020. https://www.mbie.govt.nz/dmsdocument/13562-pacific-peoples-in-the-labour-market-december-2020-quarter-unadjusted

Ministry of Business Innovation and Employment (MBIE). (2021). *Regional Economic Activity Report.* https://www.mbie.govt.nz/assets/ca6e7a0c91/regional-economic-activity-report-2013.pdf

Ministry of Social Development (MSD). (n.d.). About MSD. *MSD.* https://www.msd.govt.nz/about-msd-and-our-work/about-msd/index.html

Ministry of Social Development (MSD). (2013). *Changes to Benefit Categories From 15 July 2013.* https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/statistics/benefit/benefit-factsheet-changes-2013.html

Moore, T. (2004). Longitudinal analysis of labour force data. *Labour, Employment, and Work in New Zealand. Proceedings of the 11th Conference* 205–210.

Moore, T. (2006). Survival Analysis of Transitions from Benefit to Work Using Administrative Data. *Labour, Employment, and Work in New Zealand.* https://doi.org/10.26686/lew.v0i0.1323

Mortensen, D. (1970). Job Search, the Duration of Unemployment, and the Phillips Curve. *The American Economic Review, 60(5), 847–862.*

Mukhopadhyay, K., & Singh, R. (2011). Survival Analysis in Clinical Trials: Basics and Must Know Areas. *Perspectives in Clinical Research.* https://doi.org/10.4103/2229-3485.86872

New Zealand Government. (2020). Income After You Turn 65. https://www.govt.nz/browse/tax-benefits-and-finance/managing-your-money-in-retirement/income-after-you-turn-65/

New Zealand Immigration. (2021). *Our Regions and Cities.* https://www.newzealandnow.govt.nz/regions-nz

Nichols, A., Mitchell, J. & Lindner, S. (2013). Consequences of Long-Term Unemployment. *Washington, DC: The Urban Institute.*

Oberholzer-Gee, F. (2008). Nonemployment Stigma as Rational Herding: A Field Experiment. *Journal of Economic Behaviour and Organization, 65(1), 30–40.* https://doi.org/10.1016/j.jebo.2004.05.008

Organization for Economic Co-operation and Development (OECD). (2016). Long-Term Unemployment. *OECD Factbook 2015-2016: Economic, Environmental and Social Statistics.* https://doi.org/10.1787/factbook-2015-en

Pacheco, G., & Webber, D. (2016). *Changes in Intra-City Employment Patterns: A Spatial Analysis. International Journal of Social Economics, 43(3), 263–283.* https://doi.org/10.1108/IJSE-02-2014-0027

Pavoni, N., & Violante, G.L. (2007). Optimal Welfare-to-Work Programs. *The Review of Economic Studies, 74(1), 283–318.* https://doi.org/10.1111/j.1467-937X.2007.00422.x

Reserve Bank of New Zealand (RBNZ). (2020). *Key Graphs*. https://www.rbnz.govt.nz/statistic s/key-graphs/

Reserve Bank of New Zealand (RBNZ). (2021). *Monetary Policy Statement: May 2021.*

Social Security Act 2018. s. 225.1. https://www.legislation.govt.nz/act/public/2018/0032/latest/whole.html

Stata. (n.d.). *stcox PH-assumption tests.* https://www.stata.com/manuals/ststcoxph-assumptiontests.pdf

StataCorp. (2019). *Stata Survival Analysis Reference Manual: Release 16.* https://www.stata.com/manuals/st.pdf

Statista. (2021). *New Zealand: Distribution of Employment by Economic Sector From 2010 to 2020.* Statista.com. https://www.statista.com/statistics/436457/employment-by-economic-sector-in-new-zealand/

Statistics NZ (Stats NZ). (n.d.). *Integrated Data Infrastructure.* https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/

Statistics NZ (Stats NZ). (2017). *Three in Four New Zealanders Live in the North Island.* https://www.stats.govt.nz/news/three-in-four-new-zealanders-live-in-the-north-island

Statistics NZ (Stats NZ). (2020a). Census of Population and Dwellings. *Statistics NZ.* https://www.stats.govt.nz/census/

Statistics NZ (Stats NZ). (2020b). *Unemployment Rate.* https://www.stats.govt.nz/indicators/unemployment-rate?gclid=CjwKCAjwt8uGBhBAEiwAayu_9WpO8Z5Y0t8quyaGIdYwZVRFHA5Yd_bBl054wuCOEb2ZVDFEUzOD5RoCKMAQAvD_BwE

Statistics NZ (Stats NZ). (2020c). *Labour Market Statistics: March 2021 Quarter*. https://www.stats.govt.nz/information-releases/labour-market-statistics-march-2021-quarter

Statistics NZ (Stats NZ). (2020d). *Underutilization Rate.* https://www.stats.govt.nz/indicators/underutilisation-rate

Tansel, A., & Mehmet Taşçi. H. (2010). Hazard Analysis of Unemployment Duration by Gender in a Developing Country: The Case of Turkey. *IZA Discussion Paper No. 4844.*

Vishwanath, T. (1989). Job Search, Stigma Effects, and Escape Rate from Unemployment. *Journal of Labour Economics, 7(4), 487–502.* https://doi.org/10.1086/298218

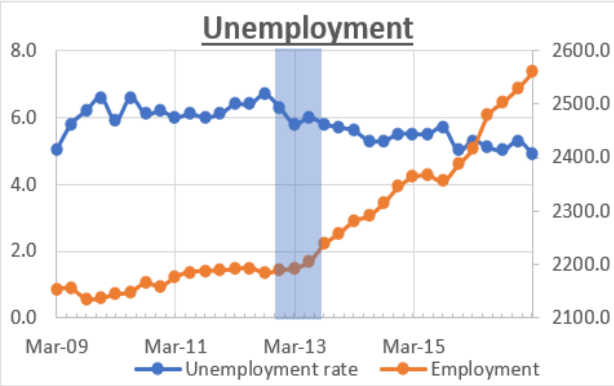
Work and Income New Zealand (WINZ). (n.d.). Benefits and Payments. https://www.workandincome.govt.nz/eligibility/index.html

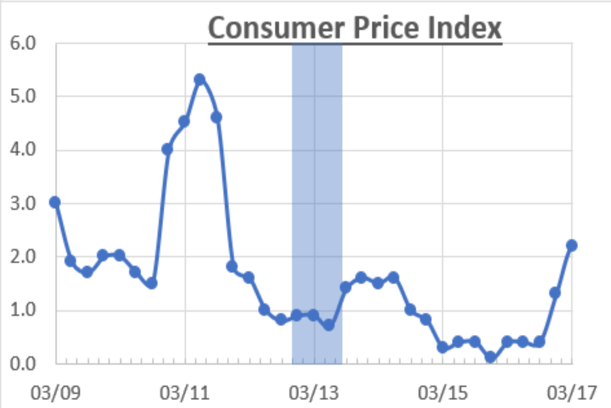
Work and Income New Zealand (WINZ). (2021). Jobseeker Support. https://www.workandincome.govt.nz/products/a-z-benefits/jobseeker-support.html#null

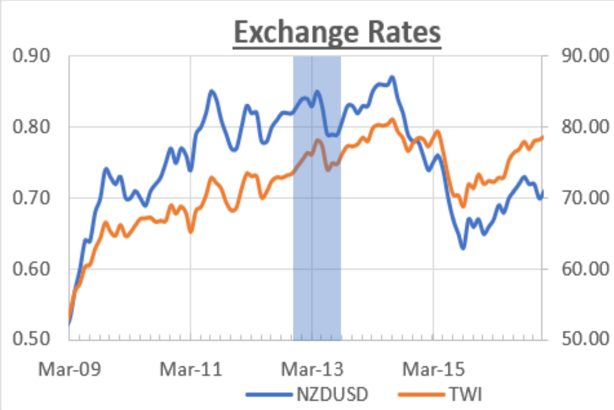
Zwiener, I., Blettner, M., & Hommel, G. (2011). Survival Analysis. *Deutsches Arzteblatt International, 108(10), 163–169.* https://doi.org/10.3238/arztebl.2011.0163

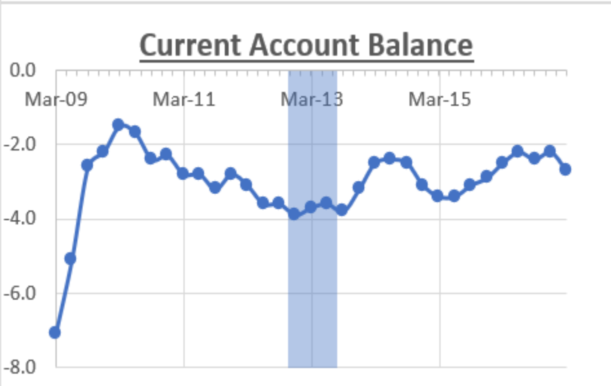
Appendices

Appendix A: Economic Conditions Graphs









The blue shaded region represents the sample period used in this research. The aim of these graphs is to demonstrate the underlying economic conditions in New Zealand affecting unemployment spells in the sample.

Appendix B: New Zealand Regions Map

Maps derived from regional maps available from New Zealand Immigration (2021).

Appendix C: Primary Model Output Table

VARIABLES	Standard Models		Last Day Worked= Start Day	
	(1) hazard ratio	(2) hazard ratio	(3) hazard ratio	(4) hazard ratio
Region				
Auckland	*Reference*	*Reference*	*Reference*	*Reference*
Northland	0.902***		0.914*	
	(0.0347)		(0.0479)	
Waikato	1.011		0.996	
	(0.0251)		(0.0329)	
BOP	1.048*		1.085**	
	(0.0293)		(0.0395)	
Gisborne	0.958		1.054	
	(0.0503)		(0.0742)	
Hawkes	1.086**		1.121***	
	(0.0353)		(0.0430)	
Taranaki	1.007		0.993	
	(0.0519)		(0.0613)	
Manawatu	0.952*		0.955	
	(0.0286)		(0.0358)	
Wellington	0.898***		0.916***	
	(0.0217)		(0.0294)	
West Coast	1.003		0.981	
	(0.0754)		(0.0942)	
Canterbury	1.149***		1.144***	
	(0.0309)		(0.0398)	
Otago	1.064*		1.154***	
	(0.0361)		(0.0501)	
Southland	1.322***		1.376***	
	(0.0496)		(0.0572)	
Tasman	1.133		1.164	
	(0.0933)		(0.118)	
Nelson	0.932		0.954	
	(0.0616)		(0.0716)	
Marlborough	1.084		1.003	
	(0.0791)		(0.0927)	
Aggregated Region				
Auckland	*Reference*	*Reference*	*Reference*	*Reference*
Northland		0.901***		0.912*
		(0.0346)		(0.0477)
Waikato		1.010		0.994
		(0.0250)		(0.0329)
BOP		1.046		1.083**
		(0.0292)		(0.0393)
East Coast		1.051*		1.104***
		(0.0307)		(0.0393)
Taranaki-Manawatu		0.964		0.963
		(0.0265)		(0.0328)
Wellington		0.897***		0.916***
		(0.0217)		(0.0294)

West Coast-Tasman		1.019		1.016
		(0.0404)		(0.0488)
Canterbury		1.148***		1.143***
		(0.0308)		(0.0397)
Southland		1.161***		1.250***
		(0.0317)		(0.0418)
Sex	0.878***	0.877***	0.851***	0.850***
(Male= Reference)	(0.0125)	(0.0125)	(0.0158)	(0.0157)
Age Range				
18–23	*Reference*	*Reference*	*Reference*	*Reference*
24–29	0.880***	0.881***	0.918***	0.919***
	(0.0187)	(0.0187)	(0.0256)	(0.0256)
30–35	0.823***	0.825***	0.923**	0.923**
	(0.0223)	(0.0223)	(0.0315)	(0.0315)
36–40	0.766***	0.766***	0.838***	0.838***
	(0.0241)	(0.0241)	(0.0335)	(0.0334)
41–45	0.728***	0.727***	0.832***	0.831***
	(0.0211)	(0.0211)	(0.0297)	(0.0297)
46–50	0.713***	0.714***	0.825***	0.827***
	(0.0213)	(0.0213)	(0.0312)	(0.0312)
51–55	0.657***	0.658***	0.786***	0.788***
	(0.0211)	(0.0211)	(0.0312)	(0.0312)
56–60	0.628***	0.630***	0.766***	0.768***
	(0.0228)	(0.0228)	(0.0328)	(0.0328)
Ethnicity				
NZ European	*Reference*	*Reference*	*Reference*	*Reference*
Maori	0.977	0.978	0.964*	0.966
	(0.0165)	(0.0164)	(0.0207)	(0.0206)
Pacifica	1.065**	1.067***	1.053	1.055*
	(0.0263)	(0.0264)	(0.0340)	(0.0341)
Asian	0.943*	0.943*	0.910**	0.910**
	(0.0296)	(0.0296)	(0.0400)	(0.0400)
Other	0.953	0.954	0.986	0.985
	(0.0421)	(0.0421)	(0.0577)	(0.0579)
Disability	0.699***	0.699***	0.739***	0.740***
(No = Reference)	(0.0153)	(0.0154)	(0.0217)	(0.0219)
Marital Status				
Married/Civil Union	*Reference*	*Reference*	*Reference*	*Reference*
Divorced/ Widowed/	0.899***	0.899***	0.875***	0.875***
Separated	(0.0249)	(0.0249)	(0.0298)	(0.0298)
Never Married	0.908***	0.908***	0.889***	0.890***
	(0.0216)	(0.0216)	(0.0266)	(0.0267)
Education Level				
No Education	*Reference*	*Reference*	*Reference*	*Reference*
Level 1	1.059***	1.059***	1.055*	1.055*
	(0.0234)	(0.0234)	(0.0293)	(0.0293)
Level 2	1.131***	1.128***	1.094***	1.090***
	(0.0259)	(0.0259)	(0.0325)	(0.0324)
Level 3	1.167***	1.164***	1.118***	1.119***
	(0.0259)	(0.0258)	(0.0320)	(0.0320)
Level 4	1.189***	1.187***	1.099***	1.098***
	(0.0312)	(0.0311)	(0.0373)	(0.0373)

Level 5 Diploma	1.268***	1.265***	1.250***	1.248***
	(0.0469)	(0.0464)	(0.0614)	(0.0605)
Level 6 Diploma	1.303***	1.304***	1.220***	1.220***
	(0.0579)	(0.0577)	(0.0666)	(0.0666)
Bachelor's degree	1.444***	1.432***	1.294***	1.286***
	(0.0400)	(0.0397)	(0.0462)	(0.0457)
Postgrad/Honours	1.658***	1.648***	1.474***	1.465***
	(0.0949)	(0.0939)	(0.114)	(0.113)
Masters	1.358***	1.355***	1.241***	1.239***
	(0.0823)	(0.0817)	(0.100)	(0.0995)
PhD	1.584***	1.553***	1.368	1.347
	(0.264)	(0.258)	(0.280)	(0.274)
Rural	1.029	1.031	1.070**	1.073**
(Urban= Reference)	(0.0236)	(0.0235)	(0.0308)	(0.0306)
Spell Count	0.989***	0.989***	0.991***	0.992***
	(0.00137)	(0.00136)	(0.00166)	(0.00165)
N	23,970	23,970	13,908	13,908

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
(1) Primary model specification.
(2) Aggregated regional model.
(3) Model (1) where the start of a spell is defined as the last day that someone worked.
(4) Model (3) where the start of a spell is defined as the last day that someone worked.
Source: IDI. Authors own calculations

Appendix D: Underlying Variable Counts

Variable sample counts.

Region	Freq.
Auckland	6762
Northland	960
Waikato	2421
Bop	2043
Gisborne	393
Hawkes	1281
Taranaki	519
Manawatu	1524
Wellington	2727
Westcoast	177
Canterbury	2304
Otago	1185
Southland	933
Tasman	216
Nelson	315
Marlborough	210
Total	23970

Aggregated Region	Freq.
Auckland	6762
Northland	960
Waikato	2421
BOP	2043
East Coast	1674
Taranaki-Manawatu	2043
Wellington	2727
West Coast-Tasman	918
Canterbury	2304
Southland	2118
Total	23970

Ethnicity	Freq.
NZ European	12966
Maori	6765
Pacifica	2193
Asian	1464
Other	582
Total	23970

Education	Freq.
No Qualification	5883
Level 1	3642
Level 2	3456
Level 3	3984
Level 4	2238
Level 5 Diploma	1011
Level 6 Diploma	651
Bachelor's degree	2328
Postgrad/Honours	402
Masters	327
PhD	48
Total	23970

Failure*	Freq.
Censored	114
Failure	23856
Total	23970

Age	Freq.
18-23	9054
24-29	3711
30-35	2337
36-40	1752
41-45	1980
46-50	2025
51-55	1812
56-60	1299
Total	23970

Sex	Freq.
Male	13593
Female	10377
Total	23970

Disability	Freq.
Not Disabled	21648
Disabled	2322
Total	23970

* Referring to censoring in survival model (e.g., last time observed is end of 2020)

Marital Status	Freq.
Married/ Civil Union	2904
Divorced/ Widowed/ Separated	3603
Never Married	17463
Total	23970

Rural	Freq.
Urban	21363
Rural	2607
Total	23970

Variable	Mean	Std. Dev.
Spell Count	6.62	6.095349
Start Day	139.45	82.00061
Last Day Worked	141.70	79.98656
Death Day	432.86	355.5753
Spell Length	293.41	344.3546
Length Last day worked	288.81	317.9469

Appendix E: Region Specific Models Output Tables

Regional Models

VARIABLES	(1) Auckland	(2) Northland	(3) Waikato	(4) BOP	(5) Gisborne	(6) Hawkes Bay	(7) Taranaki	(8) Manawatu
Sex *(Male= Reference)*	0.922***	0.920	0.874***	0.925	0.686***	0.829***	0.916	0.862***
	(0.0239)	(0.0647)	(0.0384)	(0.0449)	(0.0743)	(0.0516)	(0.0967)	(0.0491)
Age Range	0.869***	0.900***	0.862***	0.901***	0.916	0.965	0.810***	0.846***
	(0.0108)	(0.0285)	(0.0163)	(0.0200)	(0.0495)	(0.0261)	(0.0384)	(0.0216)
Ethnicity								
(NZ European= Reference)								
Maori	1.030	1.006	0.959	0.944	1.030	0.997	0.902	0.978
	(0.0361)	(0.0730)	(0.0448)	(0.0470)	(0.117)	(0.0645)	(0.0913)	(0.0579)
Pacifica	1.058	1.262	1.179*	0.888	1.780*	1.002	1.079	0.923
	(0.0370)	(0.276)	(0.114)	(0.117)	(0.531)	(0.157)	(0.205)	(0.108)
Asian	0.979	1.240	0.805*	1.229		1.298		0.822
	(0.0405)	(0.359)	(0.0995)	(0.213)		(0.375)		(0.139)
Other	0.903	0.911	0.959	1.358*		1.244		0.978
	(0.0671)	(0.179)	(0.126)	(0.219)		(0.237)		(0.225)
Disability *(No= Reference)*	0.727***	0.827*	0.691***	0.638***	0.812	0.725***	0.856	0.712***
	(0.0334)	(0.0910)	(0.0433)	(0.0501)	(0.152)	(0.0732)	(0.100)	(0.0674)
Marital Status								
(Married/Civil Union= Reference)								
Divorced/ Widowed/ Separated	0.927	0.917	0.836**	1.014	1.033	0.767*	1.275	0.947
	(0.0469)	(0.124)	(0.0704)	(0.0930)	(0.262)	(0.106)	(0.258)	(0.113)
Never Married	0.959	0.904	0.800***	1.042	1.294	0.753**	0.868	0.784**
	(0.0398)	(0.116)	(0.0573)	(0.0803)	(0.291)	(0.0898)	(0.149)	(0.0805)
Education	1.127***	1.115*	1.130***	1.119***	1.194**	1.144***	1.326***	1.183***
	(0.0172)	(0.0671)	(0.0350)	(0.0387)	(0.104)	(0.0563)	(0.124)	(0.0485)
Rural *(Urban= Reference)*	0.956	0.877*	1.010	1.055	0.917	1.027	0.980	0.953
	(0.0707)	(0.0637)	(0.0616)	(0.0682)	(0.126)	(0.0937)	(0.123)	(0.0747)
Spell Count	0.979***	0.986**	0.977***	0.992*	1.000	0.995	0.981**	0.996
	(0.00298)	(0.00644)	(0.00428)	(0.00410)	(0.0116)	(0.00474)	(0.00779)	(0.00487)
N	6,762	960	2,421	2,043	393	1,281	519	1,524

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.
Models use aggregated Age Range, Education, and Spell Count variables.
Source: IDI. Authors own calculations.

VARIABLES	(9) Wellington	(10) West Coast	(11) Canterbury	(12) Otago	(13) Southland	(14) Tasman	(15) Nelson	(16) M'borough
Sex *(Male= Reference)*	0.912**	0.932	0.841***	0.870**	0.861*	0.926	0.896	0.869
	(0.0381)	(0.163)	(0.0388)	(0.0541)	(0.0723)	(0.127)	(0.115)	(0.142)
Age Range	0.903***	0.766***	0.911***	0.959	0.964	0.925	0.921	1.006
	(0.0169)	(0.0637)	(0.0189)	(0.0282)	(0.0307)	(0.0615)	(0.0505)	(0.0749)
Ethnicity								
(NZ European= Reference)								
Maori	0.942	0.924	0.831***	0.996	1.052	0.925	1.144	0.992
	(0.0481)	(0.215)	(0.0550)	(0.0842)	(0.0872)	(0.225)	(0.187)	(0.183)
Pacifica	1.071		1.132	1.209	1.136			
	(0.0679)		(0.131)	(0.194)	(0.124)			
Asian	0.893		0.945	1.189			1.053	
	(0.0807)		(0.104)	(0.281)			(0.254)	
Other	1.023		0.944	0.673**	1.048			
	(0.129)		(0.102)	(0.136)	(0.333)			
Disability *(No= Reference)*	0.787***	0.777	0.670***	0.500***	0.738***	0.711**	0.633**	0.373***
	(0.0511)	(0.168)	(0.0375)	(0.0503)	(0.0843)	(0.119)	(0.121)	(0.0873)
Marital Status								
(Married/Civil Union= Reference)								
Divorced/ Widowed/ Separated	0.821**	1.461	0.863*	0.796*	0.731**	1.064	0.845	0.667
	(0.0710)	(0.553)	(0.0771)	(0.104)	(0.0899)	(0.390)	(0.183)	(0.183)
Never Married	0.924	1.164	0.880	0.882	0.852*	1.126	0.919	0.915
	(0.0677)	(0.400)	(0.0768)	(0.100)	(0.0796)	(0.384)	(0.194)	(0.249)
Education	1.197***	1.310*	1.163***	1.085**	1.079	1.079	0.998	1.020
	(0.0287)	(0.196)	(0.0319)	(0.0376)	(0.0744)	(0.157)	(0.117)	(0.110)
Rural *(Urban= Reference)*	1.004	1.190	1.244***	1.163*	1.055	0.959	1.520*	1.365
	(0.112)	(0.203)	(0.0771)	(0.0995)	(0.118)	(0.141)	(0.363)	(0.371)
Spell Count	0.983***	1.002	0.984***	0.982***	0.995	0.987	0.996	0.970**
	(0.00379)	(0.0173)	(0.00404)	(0.0055)	(0.0056)	(0.0112)	(0.0084)	(0.0141)
N	2,727	177	2,304	1,185	933	216	315	210

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Models use aggregated Age Range, Education, and Spell Count variables.
Source: IDI. Authors own calculations.

Appendix F: Aggregated Regions Output Table

Aggregated Regional Models

VARIABLES	(1) East Coast	(2) Taranaki-Manawatu	(3) West Coast-Tasman	(4) Southland
Sex *(Male= Reference)*	0.801***	0.871***	0.920	0.859***
	(0.0433)	(0.0433)	(0.0663)	(0.0419)
Age Range	0.968	0.849***	0.915***	0.977
	(0.0232)	(0.0189)	(0.0304)	(0.0205)
Ethnicity				
(NZ European= Reference)				
Maori	1.006	0.965	1.050	1.053
	(0.0564)	(0.0491)	(0.101)	(0.0611)
Pacifica	1.066	0.934		1.154
	(0.161)	(0.0972)		(0.105)
Asian	1.105	0.831	0.985	1.016
	(0.342)	(0.132)	(0.186)	(0.205)
Other	1.352*	0.868	1.564**	0.841
	(0.236)	(0.164)	(0.337)	(0.153)
Disability *(No= Reference)*	0.742***	0.758***	0.643***	0.594***
	(0.0652)	(0.0566)	(0.0619)	(0.0450)
Marital Status				
(Married/Civil Union= Reference)				
Divorced/ Widowed/ Separated	0.809*	1.013	0.934	0.767***
	(0.0996)	(0.103)	(0.132)	(0.0683)
Never Married	0.839*	0.816**	0.972	0.859**
	(0.0891)	(0.0716)	(0.131)	(0.0633)
Education	1.143***	1.191***	1.053	1.063**
	(0.0488)	(0.0449)	(0.0716)	(0.0316)
Rural *(Urban= Reference)*	0.966	0.966	1.173*	1.104
	(0.0726)	(0.0637)	(0.108)	(0.0731)
Spell Count	0.934**	0.918***	0.898***	0.887***
	(0.0301)	(0.0255)	(0.0357)	(0.0246)
N	1,674	2,043	918	2,118

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Models use aggregated Age Range, Education, and Spell Count variables.
Source: IDI. Authors own calculations

Appendix G: Underlying Counts for Additional Variable Specifications

Additional variable counts for aggregated covariates and robustness model variables.

Aggregated Education	Freq.
High school	16968
undergraduate/diploma	3900
bachelors	2328
postgraduate/PhD	774
Total	23970

Aggregated Age	Freq.
16-23	9054
24-29	3711
30-40	4086
41-50	4005
51-60	3114
Total	23970

NZ Island	Freq.
North Island	18630
South Island	5340
Total	23970

Major Centre	Freq.
No	9669
Yes	14301
Total	23970

Start Day Range	Freq.
01/11/12 - 20/12/12	5406
21/12/12 - 08/02/13	3375
09/02/13 - 31/03/13	3927
01/04/13 - 20/05/13	4200
21/05/13 - 09/07/13	4167
10/07/13 - 31/07/13	2895
Total	23970

Aggregated spell count	Freq.
1st	3834
2-5	9588
6-10	5871
11-15	2568
16-20	1170
21-25	585
26-30	216
31+	138
Total	23970

Regional Shift	Freq.
Moved Region	3054
Didn't Move	20916
Total	23970

Industry	Freq.
Manufacturing	18036
Natural Resources	1089
Professional	2727
Construction	1185
Agriculture	933
Total	23970

Variable (2013 regional variables)*	Obs	Mean	Std. Dev.
Unemployment Rate	23970	6.31	1.262
underutilization Rate	23970	48.13	8.857
(GDP per capita)/1000	23970	5.38	0.729
Deprivation Index	23970	67.05	9.281
Proportion of school leavers attaining NCEA Level 3	23970	6.90	3.420
Mean Household Income/1000	23970	79.38	9.178

* Based on data from MBIE's Regional Economic Activity Web Tool

Appendix H: Robustness Models Output Table

Robustness checks

VARIABLES	(1)	(2)	(3)	(4)	(5)
Spell Start Day	1.000 (8.49e-05)				
Start Day Range *(1–50=Reference)*					
51–100		0.987 (0.0228)			
101–150		0.945*** (0.0202)			
151–200		0.958** (0.0202)			
201–250		1.019 (0.0224)			
251–272		0.913*** (0.0237)			
Major Centre *(non-major centre=Reference)*			0.968** (0.0154)		
NZ Island *(North Island=Reference)*				1.143*** (0.0196)	
Regional Shift *(Moved Region=Reference)*					1.241*** (0.0239)
Region†	0.985*** (0.00204)	0.985*** (0.00204)			
Sex *(Male=Reference)*	0.892*** (0.0125)	0.893*** (0.0126)	0.892*** (0.0126)	0.891*** (0.0125)	0.894*** (0.0126)
Age	0.895*** (0.00571)	0.895*** (0.00571)	0.894*** (0.00569)	0.896*** (0.00570)	0.891*** (0.00571)
Ethnicity *(NZ European=Reference)*					
Maori	0.974 (0.0161)	0.973 (0.0161)	0.950*** (0.0154)	0.976 (0.0161)	0.956*** (0.0155)
Pacifica	1.055** (0.0251)	1.055** (0.0250)	1.044* (0.0250)	1.069*** (0.0256)	1.024 (0.0244)
Asian	0.946* (0.0289)	0.946* (0.0290)	0.940** (0.0289)	0.954 (0.0292)	0.918*** (0.0281)
Other	0.960 (0.0425)	0.960 (0.0424)	0.961 (0.0425)	0.967 (0.0427)	0.950 (0.0427)
Disability *(No=Reference)*	0.700*** (0.0152)	0.701*** (0.0152)	0.707*** (0.0153)	0.700*** (0.0152)	0.705*** (0.0153)
Marital Status *(Married/ Civil Union=Reference)*					
Divorced/ Widowed/ Separated	0.897*** (0.0247)	0.899*** (0.0248)	0.902*** (0.0249)	0.899*** (0.0248)	0.905*** (0.0251)
Never Married	0.909*** (0.0217)	0.909*** (0.0217)	0.912*** (0.0218)	0.911*** (0.0217)	0.910*** (0.0218)
Education	1.131*** (0.0103)	1.130*** (0.0102)	1.133*** (0.0103)	1.134*** (0.0103)	1.134*** (0.0103)
Rural *(Urban=Reference)*	1.035 (0.0231)	1.035 (0.0230)	1.006 (0.0246)	1.024 (0.0228)	1.036 (0.0233)
Aggregated Spell Count	0.941*** (0.00558)	0.941*** (0.00558)	0.943*** (0.00559)	0.939*** (0.00557)	0.944*** (0.00558)
N	23,970	23,970	23,970	23,970	23,970

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Models use aggregated Age Range, Education, and Spell Count variables. Source: IDI. Authors own calculations.
† Region is ranked by 2013 unemployment rates. A value of 0 is given to the region with the lowest 2013 unemployment rate. Tied regions hold a shared value.

Appendix I: Comparing Region at Start and End of Spell Output Table

VARIABLES	(1) Region at Spell Start	(SE)	(2) Region at Spell End	(SE)	(3) Region if Didn't Move	(SE)
Region						
Auckland	Reference		Reference		Reference	
Northland	0.899***	(0.0344)	0.851***	(0.0320)	0.891***	(0.0374)
Waikato	1.003	(0.0249)	0.993	(0.0245)	1.014	(0.0270)
BOP	1.045	(0.0292)	1.040	(0.0291)	1.046	(0.0317)
Gisborne	0.959	(0.0501)	0.913*	(0.0478)	0.958	(0.0548)
Hawkes	1.079**	(0.0350)	1.085**	(0.0363)	1.082**	(0.0378)
Taranaki	1.009	(0.0514)	1.013	(0.0500)	1.007	(0.0547)
Manawatu	0.943*	(0.0283)	0.925***	(0.0280)	0.946*	(0.0312)
Wellington	0.898***	(0.0217)	0.882***	(0.0213)	0.888***	(0.0228)
West Coast	0.976	(0.0734)	0.979	(0.0712)	1.005	(0.0866)
Canterbury	1.139***	(0.0305)	1.129***	(0.0305)	1.150***	(0.0329)
Otago	1.063*	(0.0359)	1.048	(0.0366)	1.062	(0.0393)
Southland	1.303***	(0.0483)	1.261***	(0.0456)	1.297***	(0.0491)
Tasman	1.140	(0.0933)	1.081	(0.0988)	1.230**	(0.127)
Nelson	0.927	(0.0608)	0.945	(0.0602)	0.964	(0.0739)
Marlborough	1.075	(0.0793)	1.101	(0.0854)	1.098	(0.0973)
Sex *(Male= Reference)*	0.891***	(0.0126)	0.889***	(0.0126)	0.894***	(0.0136)
Age Range	0.895***	(0.0057)	0.894***	(0.0057)	0.893***	(0.0061)
Ethnicity *(NZ European=Reference)*						
Maori	0.970*	(0.0162)	0.972*	(0.0163)	0.983	(0.0178)
Pacifica	1.064**	(0.0262)	1.056**	(0.0260)	1.050*	(0.0271)
Asian	0.951	(0.0297)	0.949*	(0.0295)	0.934**	(0.0300)
Other	0.966	(0.0427)	0.958	(0.0425)	0.946	(0.0454)
Disability *(No= Reference)*	0.699***	(0.0153)	0.700***	(0.0153)	0.696***	(0.0163)
Marital Status *(Married/ Civil Union=Reference)*						
Divorced/Widowed/Separated	0.900***	(0.0248)	0.896***	(0.0248)	0.898***	(0.0263)
Never Married	0.913***	(0.0218)	0.915***	(0.0218)	0.918***	(0.0233)
Education Level	1.136***	(0.0104)	1.137***	(0.0105)	1.141***	(0.0111)
Rural *(Urban= Reference)*	1.026	(0.0234)	1.030	(0.0236)	1.030	(0.0261)
Spell Count	0.937***	(0.0056)	0.937***	(0.0056)	0.943***	(0.0061)
N	23,970		23,829		20,916	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: IDI. Authors own calculations.
Models use aggregated Age Range, Education, and Spell Count variables.

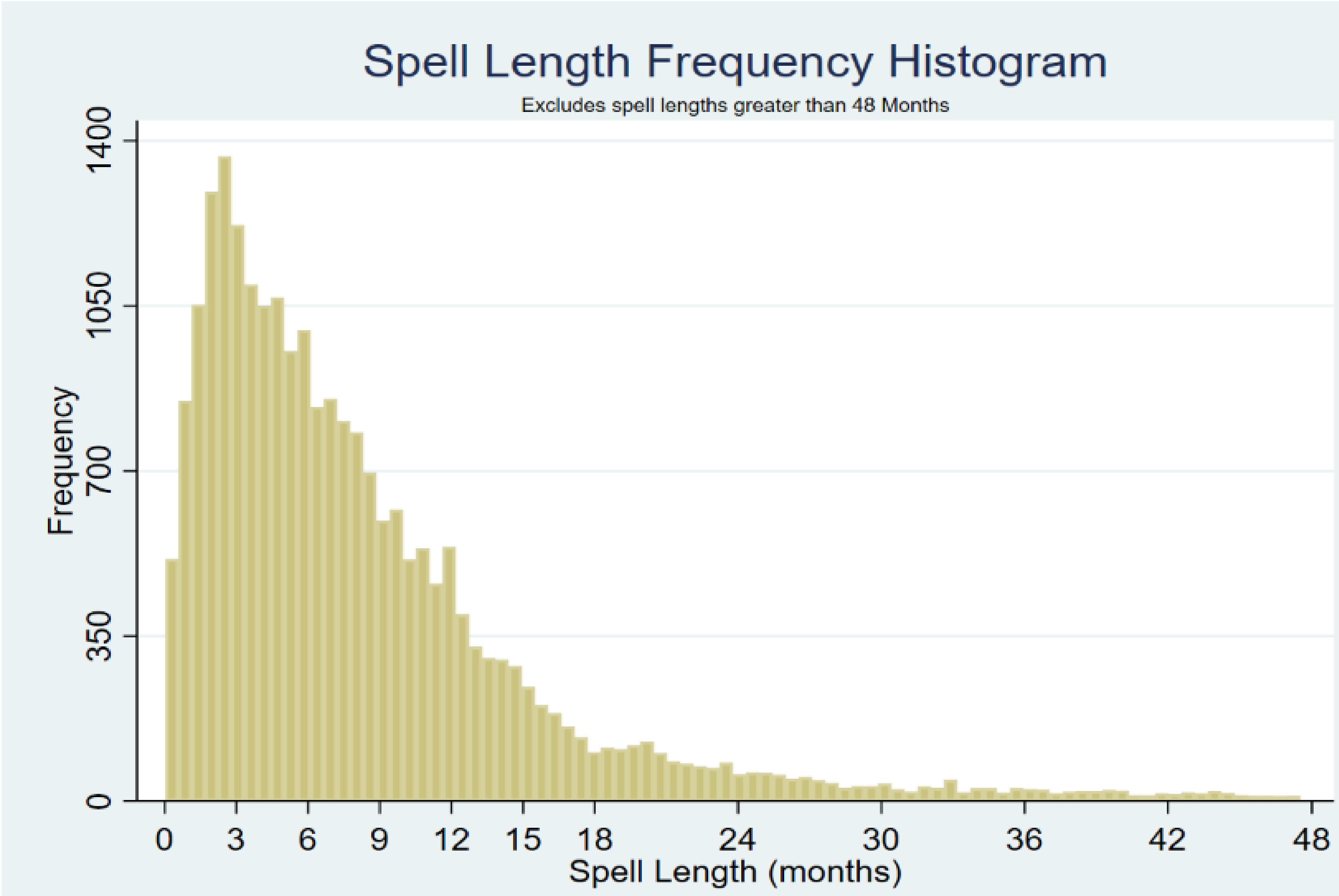
Appendix J: Underlying Regional Mechanism Output Table

Macroeconomic Variables Replace Region

VARIABLES	(1)	(2)	(3)
Regional Indicators			
Unemployment Rate	0.971***		
	(0.00618)		
Underutilization Rate		0.982***	
		(0.00337)	
(GDP per capita)/1000			1.002**
			(0.00103)
Industry *(Manufacturing= Reference)*			
Natural Resources	0.947	0.952	0.946
	(0.0316)	(0.0316)	(0.0325)
Professional	0.879***	0.869***	0.855***
	(0.0193)	(0.0192)	(0.0232)
Construction	0.990	0.992	1.044
	(0.0333)	(0.0327)	(0.0332)
Agriculture	1.191***	1.187***	1.264***
	(0.0457)	(0.0448)	(0.0449)
Sex *(Male= Reference)*	0.892***	0.892***	0.893***
	(0.0126)	(0.0126)	(0.0126)
Age	0.895***	0.895***	0.894***
	(0.00569)	(0.00570)	(0.00569)
Ethnicity (European Decent= *Reference*)			
Maori	0.971*	0.973*	0.964**
	(0.0161)	(0.0161)	(0.0157)
Pacifica	1.059**	1.062**	1.043*
	(0.0252)	(0.0253)	(0.0250)
Asian	0.943*	0.945*	0.933**
	(0.0290)	(0.0290)	(0.0288)
Other	0.963	0.965	0.959
	(0.0427)	(0.0428)	(0.0428)
Disability (None= *Reference*)	0.701***	0.701***	0.706***
	(0.0153)	(0.0153)	(0.0153)
Marital Status			
(Married/ Civil Union= *Reference*)			
Divorced/ Widowed/ Separated	0.901***	0.900***	0.903***
	(0.0248)	(0.0248)	(0.0249)
Never Married	0.914***	0.913***	0.915***
	(0.0218)	(0.0217)	(0.0218)
Education	1.136***	1.136***	1.135***
	(0.0104)	(0.0104)	(0.0103)
Rural (Urban= *Reference*)	1.020	1.026	1.022
	(0.0229)	(0.0231)	(0.0232)
Spell Count	0.939***	0.938***	0.940***
	(0.00559)	(0.00558)	(0.00559)
N	23,970	23,970	23,970

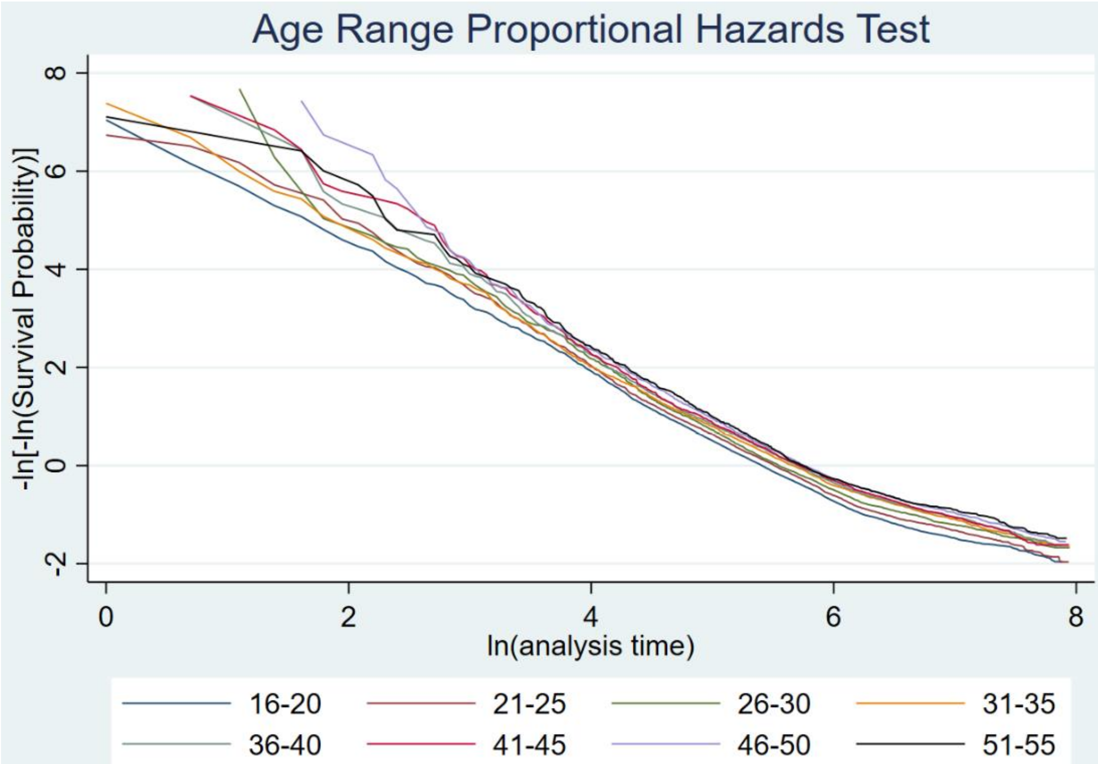
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Models use aggregated Age Range, Education, and Spell Count variables
Source: IDI. Authors own calculations

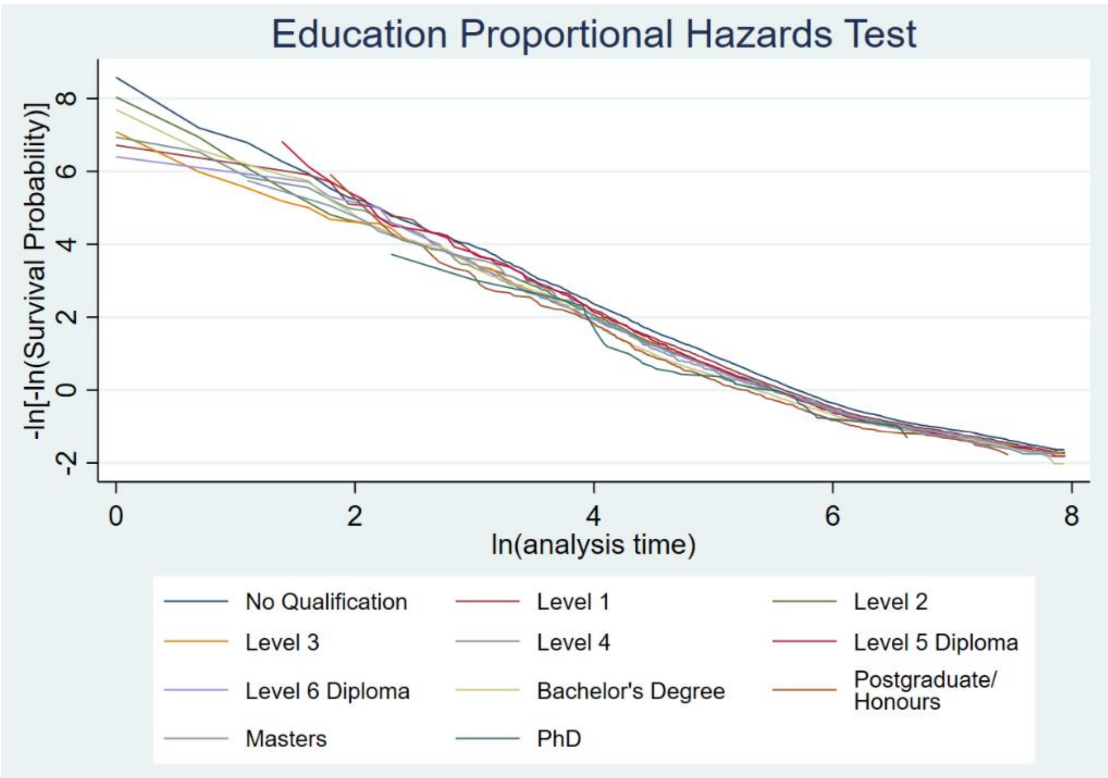
Appendix K: Spell Length Histogram

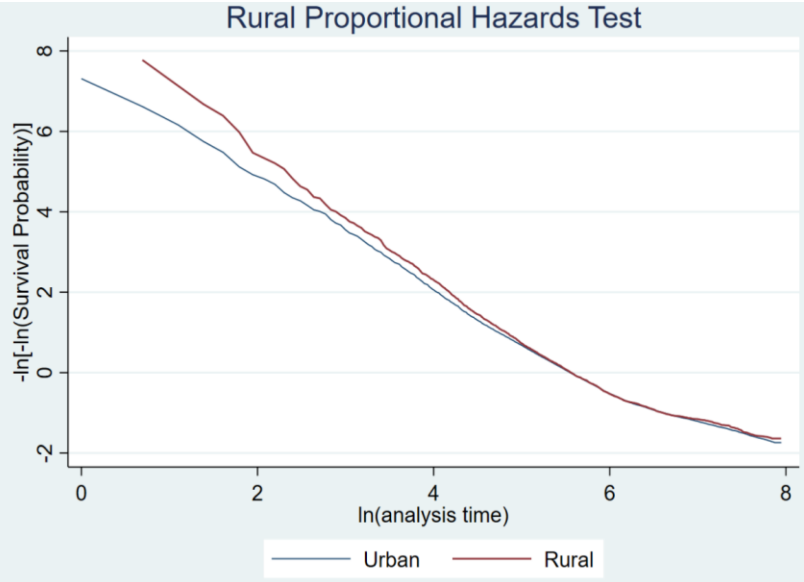


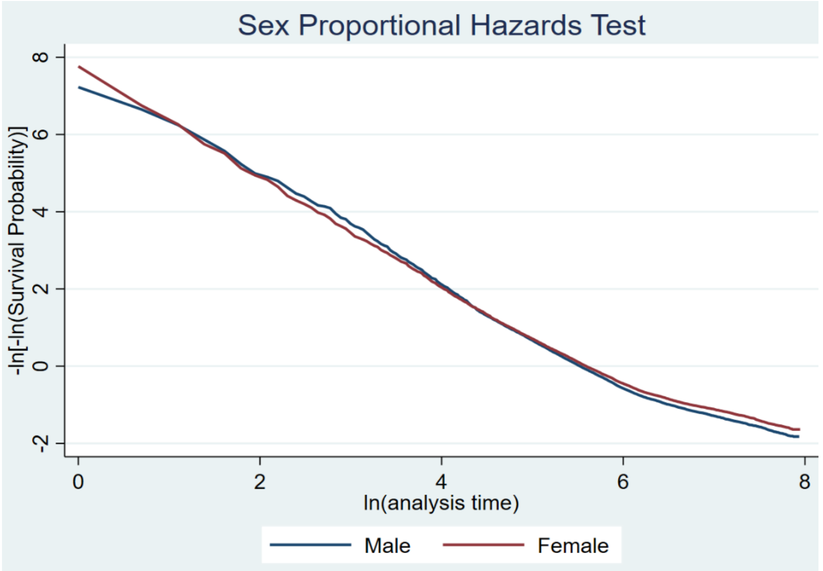
Frequency histogram 0f sell lengths in the sample. Spells greater than 48 months are excluded from this histogram due to small underlying counts of individual's experiencing spells at these lengths.

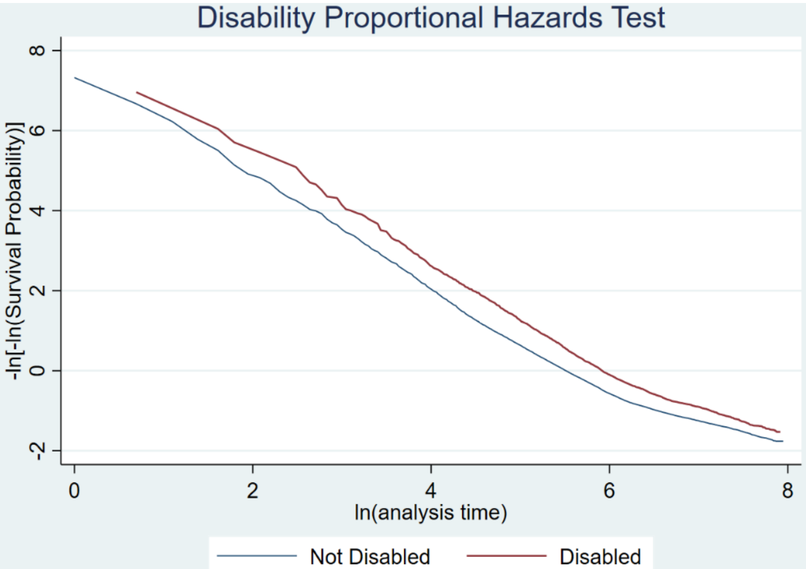
Appendix L: Covariate Proportional Hazard Assumption Test

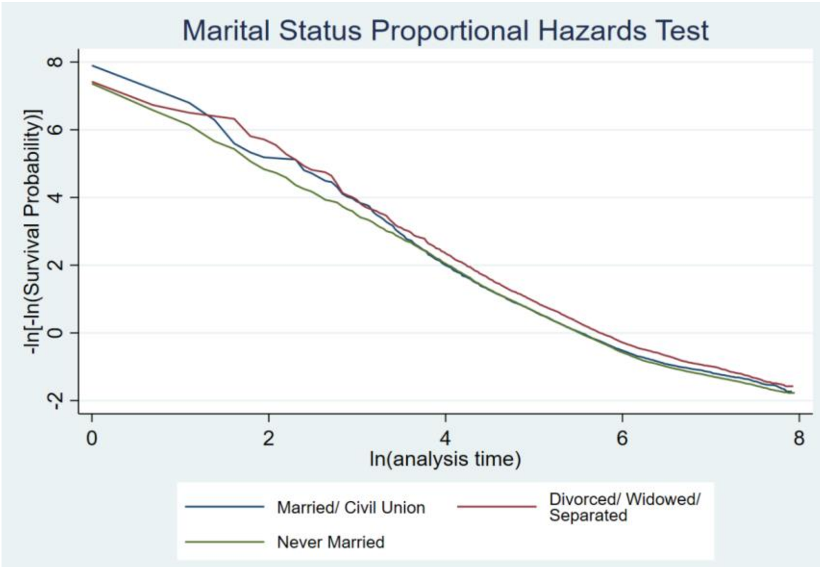


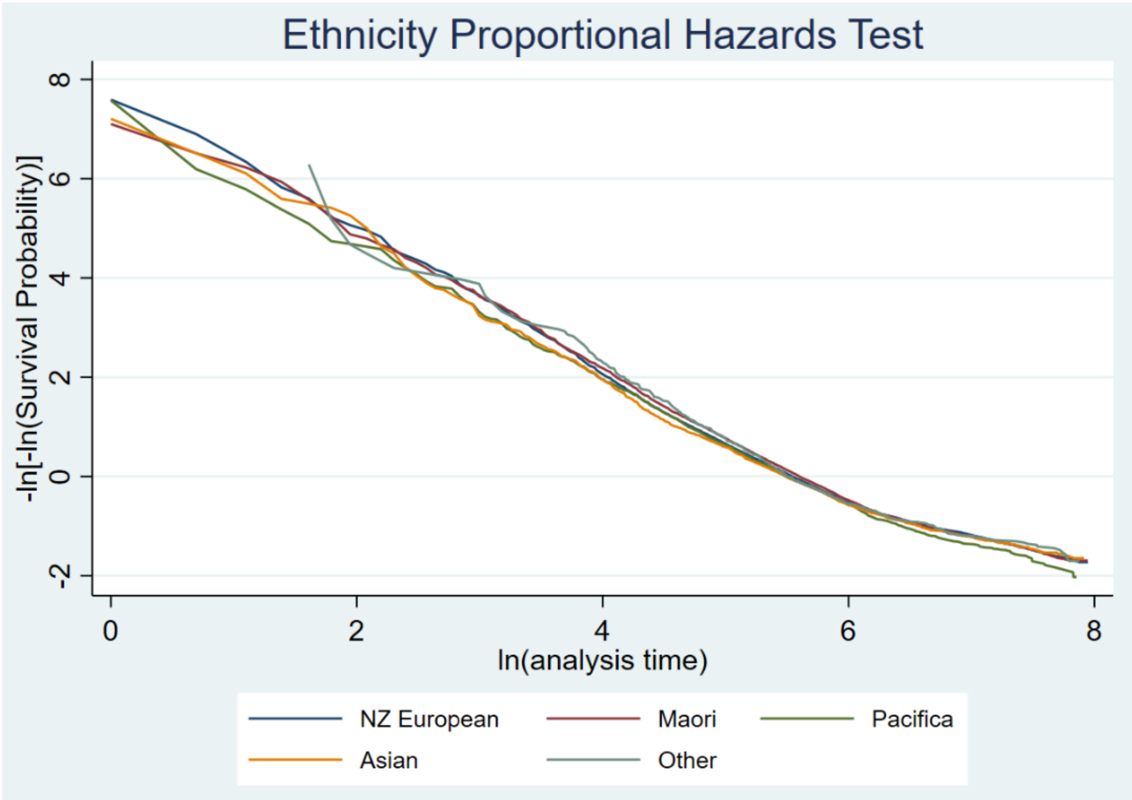


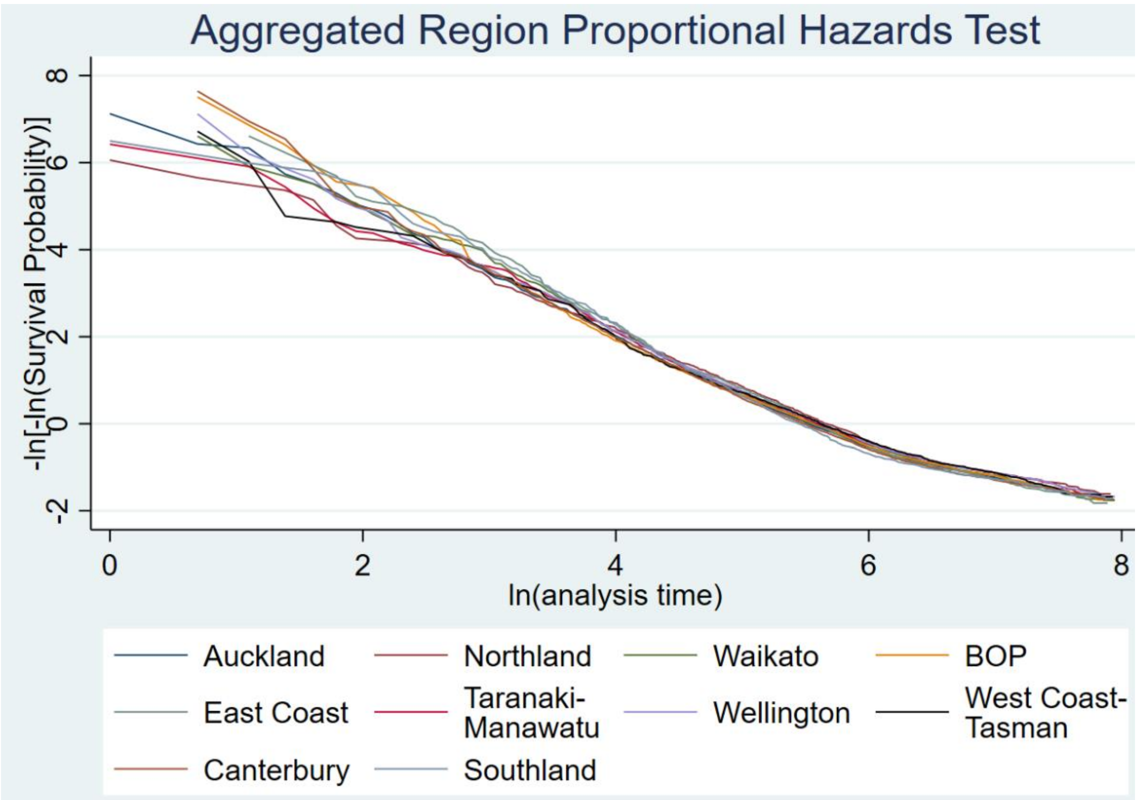

Rural Proportional Hazards Test
-ln[-ln(Survival Probability)]
-2
0
2
4
6
8
ln(analysis time)
Urban
Rural


Sex Proportional Hazards Test
-ln[-ln(Survival Probability)]
-2
0
2
4
6
8
ln(analysis time)
Male
Female

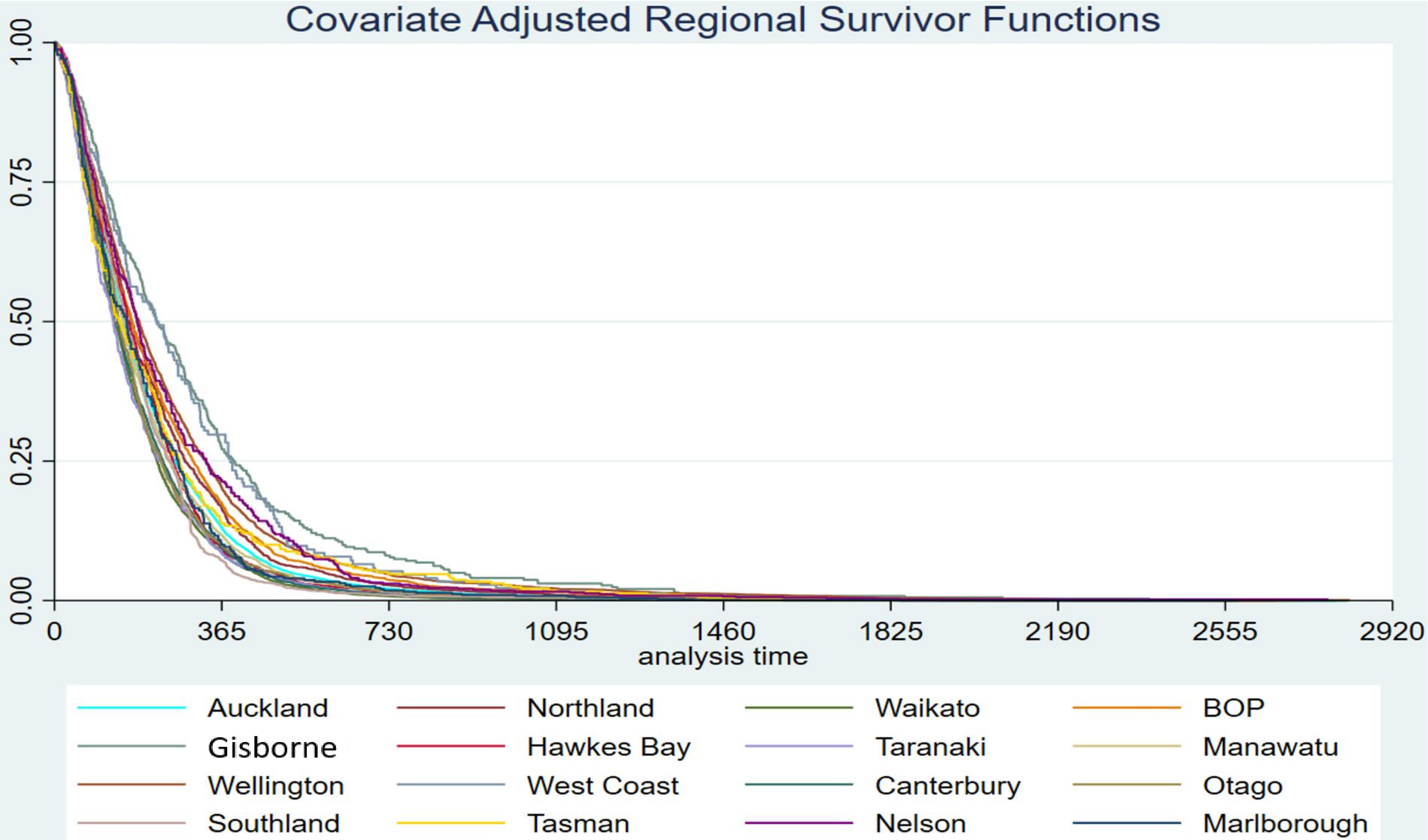

Disability Proportional Hazards Test
-ln[-ln(Survival Probability)]
-2
0
2
4
6
8
ln(analysis time)
Not Disabled
Disabled

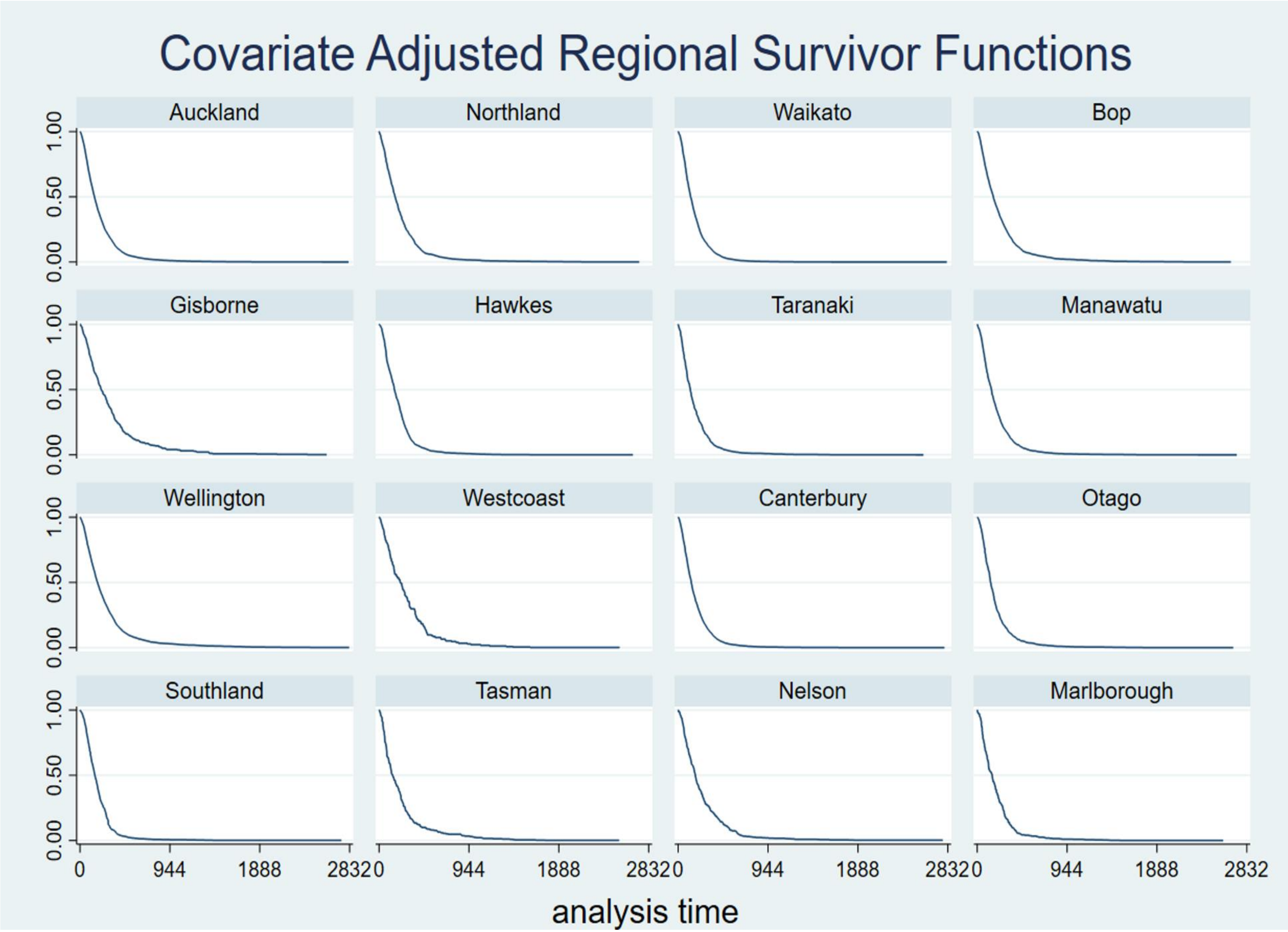

Marital Status Proportional Hazards Test
-ln[-ln(Survival Probability)]
-2
0
2
4
6
8
ln(analysis time)
Married/ Civil Union
Divorced/ Widowed/ Separated
Never Married


Ethnicity Proportional Hazards Test
-ln[-ln(Survival Probability)]
8
6
4
2
0
-2
0
2
4
6
8
ln(analysis time)
NZ European
Maori
Pacifica
Asian
Other


Aggregated Region Proportional Hazards Test
-ln[-ln(Survival Probability)]
8
6
4
2
0
-2
0
2
4
6
8
ln(analysis time)
Auckland
Northland
Waikato
BOP
East Coast
Taranaki-Manawatu
Wellington
West Coast-Tasman
Canterbury
Southland

Appendix M: Regional Survivor Functions




Covariate Adjusted Regional Survivor Functions
Auckland
Northland
Waikato
Bop
Gisborne
Hawkes
Taranaki
Manawatu
Wellington
Westcoast
Canterbury
Otago
Southland
Tasman
Nelson
Marlborough
0.00
0.50
1.00
0
944
1888
2832
analysis time