Crawford School of Public Policy

TTPI

Tax and Transfer Policy Institute

Spatial heterogeneity in welfare reform success

TTPI - Working Paper 19/2023
December 2023

Barbara Broadway
Melbourne Institute: Applied Economic & Social Research
The University of Melbourne

Anna Zhu
RMIT University, IZA

Abstract

We investigate if neighbourhood characteristics matter to the success of an exogenous change in a country's institutional settings. We examine the causal impact from one of the largest welfare reforms in Australia, which used the levers of reducing Income Support payments and increasing participation requirements, to reduce welfare dependency and to improve employment outcomes among single mothers. Using a new administrative dataset, which captures the full universe of single mothers targeted by this reform, along with information from five other data sources, we find significant heterogeneity in the reform effects across local areas. The reform did not have the intended effect in geographic regions that were relatively disadvantaged. The effect of the reform for all the local labour market in Australia is estimated with Regression Discontinuity models and correlated with the characteristics of the local labour market region. Our aim is to ask: is there spatial heterogeneity in the local reform effects? And if so, can we find patterns that describe how the reform's effectiveness varies with local conditions such as employment opportunities, access to services, and community characteristics?

Keywords: welfare reform, spatial heterogeneity, single mothers

* *We acknowledge the support of the Australian Research Council Linkage Project (LP170100472) 'The Impact of Income Support Design on the Outcomes of Children and Youth'. We thank Bruce Bradbury, Robert Breunig, Ashley Craig, Denzil Fiebig, Tim Robinson and numerous seminar and conference participants for helpful comments. We would also like to thank Yin-King Fok, Michael Duffield and Tessa Loriggio for their excellent research assistance. The paper uses administrative data from the Data On Multiple INdividual Occurrences" (DOMINO) dataset, provided by the Australian Government Department of Social Services (DSS). It also uses the restricted release file of the Household, Income and Labour Dynamics in Australia (HILDA) survey. HILDA is funded by the Australian Government Department of Social Services (DSS) and managed by the Melbourne Institute. The findings and views reported in this paper are those of the authors alone and should not be attributed to either DSS or the Melbourne Institute. Author contacts: b.broadway@unimelb.edu.au; anna.zhu@rmit.edu.au*

Tax and Transfer Policy Institute

Crawford School of Public Policy

College of Asia and the Pacific

+61 2 6125 9318

tax.policy@anu.edu.au

The Australian National University

Canberra ACT 0200 Australia

www.anu.edu.au

The Tax and Transfer Policy Institute (TTPI) is an independent policy institute that was established in 2013 with seed funding from the federal government. It is supported by the Crawford School of Public Policy of the Australian National University.

TTPI contributes to public policy by improving understanding, building the evidence base, and promoting the study, discussion and debate of the economic and social impacts of the tax and transfer system.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

1. Introduction

Recent economic recessions (during the financial crisis and the COVID-19 pandemic) have thrown into question how welfare policy ought to be implemented. Economic hardship is not evenly distributed across different areas, and neither is the added pain from a recession – with lasting effects on the regions that were hit the worst (Hershbein and Stuart, forthcoming). And yet the rules of nation-wide policies are often uniformly applied. For example, in many countries, national policy reforms to the amount of government support and/or the stringency of job-activation rules apply uniformly to all regions regardless of their different economic conditions. If residents have a better chance to connect to job or retraining opportunities in some areas, or face greater barriers to employment in others, then a-priori, we may expect individuals to respond differently to an aggregate welfare policy depending on where they live. However, little research examines how the effects of a nation-wide policy reform may vary across different localities and how such local reform responses varies with the characteristics of the place of residence. Addressing these questions can be a first step to understanding how we might leverage neighbourhood strengths to complement nation-wide policies or how to better support areas with a propensity to respond weakly to welfare reform.

In this paper, we examine one of the most significant welfare reforms in Australia: the 2006 Welfare-to-Work reform. The reform significantly reduced the amount of income support provided to low-income single mothers. Its aim was to reduce welfare dependency and to increase earnings from employment. Similar reforms in other countries have been shown to reduce caseloads and increase employment, and there is some evidence to suggest that the children of affected mothers also benefited (Blank 2002; 2006; Smolensky and Gootman 2003; Auerbach, Card and Quigley 2005; Gong and Breunig 2014; Broadway, Loriggio, Ryan and Zhu 2021; Suziedelyte and Zhu 2021; Hartley, Lamarche and Ziliak 2022). Fisher and Zhu (2019) also show that single mothers are more likely to repartner soon after separation when the amount of income support reduces, while Aizer et al., (forthcoming) find no effect of cash transfers on family formation.

The reform was implemented uniformly across the country, and the States and Territories did not have any prerogative to vary the timing, intensity or scope of the policy. However, the same reform rules can engender different impacts in different geographic areas due to variations in the strength of the local labour market, income levels, local infrastructure and organisational factors. Studies from all over the world show that the consequences of spatial inequality are felt by individuals in every sphere of life. Chyn and Katz (2021) provide an overview. In a nutshell: there is a large body of evidence that shows that spatial inequality is causally connected to poverty (Ludwig et al. 2013) as well as employment outcomes and income (Deryugina, Kawano, and Levitt, 2018; Deryugina and Molitor, 2020; Collins and Wanamaker, 2014; Boustan; 2016), and these spatial inequalities in economic outcomes can be reinforced of offset by the urban development patterns it produces (Modai-Snir and van Ham, 2018). Many studies find that growing up in more advantaged areas has a causal impact on children's educational attainment, above and beyond family and individual characteristics (Chetty, Hendren, and Katz 2016; Chyn 2018; Chetty and Hendren 2018; Laliberté 2021).

Australia is no exception to these phenomena. Its geographic traits, such as the hyper-concentration of the population in the coastal cities and the discernible rural-urban distinction, has similarities to those of many other countries including Canada and the United States. If there are large and persistent disparities in the economic context in which individuals take their decisions, it stands to reason that they will respond differently to changes in policy and institutions (Becker, Egger, Ehrlich, 2013). As expected, there are large differences in income (Henman, 2008), income in equality (Biddle and Montaigne 2017) and education (Smith et al. 2019).

To examine to what degree these differences in economic context matter to reform success or failure, we estimate the reforms' impact on welfare dependency and employment earnings up to six years after separation, separately for 87 distinct local labour markets, each with a population of about 300,000 individuals. Then, we examine which characteristics of the local labour market correlate most strongly with the strength of the reform response in the area.

A key obstacle to producing local estimates of reform effects and analyses so far, has been the lack of large, long and representative panel datasets. We use a very rich and novel administrative dataset that includes *the full universe* of single mothers in the welfare system, as well as nearly all Australian families with children on low to medium incomes. The "Data On Multiple Individual Occurrences" (DOMINO) dataset contains bi-weekly records of welfare benefit receipt and earnings over a two-decade period. The enormous size of this dataset, together with its precise information on date of separation, allows us to use a Regression Discontinuity Design RDD (Local Linear Regression) approach to estimate the causal impact of the reform within each area with sufficient statistical power. We verify that mothers do not manipulate the date of separation and address potential anticipatory behaviour associated with the treated group being more likely to self-select into areas with better employment prospects at the time of separation. We also address selective sorting after separation as we count mothers for the entire analysis (up to six years after separation) at the place where they resided at the time of the separation, thus counting moving as a potential mechanism for the reform effect within a region.

We turn to six different data sources[1] to describe the geospatial characteristics of each local labour market. We compiled information on the local economic conditions, values towards women working, ethnic networks, potential relationship-re-partnering opportunities, cultural profile, accessibility to services and amenities and the demographics. Our aim is to examine spatial heterogeneity in reform effects. We begin with a univariate analysis to find patterns between area characteristics and local reform impacts, using simple pairwise correlations. In a second step, we use Machine Learning (ML) techniques to enable a multivariate analysis. This approach allows us to select the characteristics that are most predictive of the local reform effect size, to then create an index that is highly predictive of reform elasticities. We chose this path because standard regression techniques would lead to overfitting issues, when working with few observations (one per local labour market) and a very broad range of features describing these labour markets. This second part of the analysis is descriptive in nature and does not allow us to uncover causal relationships between a specific area characteristic and its exact impact on reform success. It does, however, allow us to gauge the extent of the heterogeneity, and to discern which types of places the reform is most likely to succeed or fail.

This study makes two important contributions. First, we show that the success or failure of a welfare reform depends crucially on the economic context in which it is implemented. We find that the exact same change in policy parameters can have effects of opposite sign when implemented in different circumstances – and in addition, we find that weaker intended reform impacts (smaller reductions in length of time on Income Support and smaller increases in private income) occurred in the most disadvantaged places. This is an important contribution to the literature evaluating welfare reforms, which typically evaluates country-wide effects. We show that *on average* targeted welfare recipients responded to financial incentives in the intended way, however, this high-level result masks substantial spatial heterogeneity in effect sizes. The large variation we find in effect sizes in different economic contexts across space,

[1] Including three administrative datasets – Data On Multiple INdividual Occurrences (DOMINO), Australian Children's Education & Care Quality Authority (ACECQA), National Centre for Vocational Education Research (NCVER) and two survey-based datasets – Household, Income and Labour Dynamics in Australia (HILDA) and Australian Bureau of Statistics (ABS) Census and the ABS Employee Earnings, Benefits and Trade Union Survey. We have 61 variables in total.

suggest that variation in effect sizes may also be expected if the economic context for a certain set of policy settings varies over time.

Our second contribution is to the literature on place-based economic policies. By following people who move areas, previous research has shown that place affects their outcomes. We add to these findings by showing that place also affects their *behavioural response* to an exogenous change in institutional settings.

The remainder of the paper is organised as follows: section 2 discusses the reform, and which behavioural adjustments we should expect based on the change in policy settings. Section 3 describes i) the estimation approach for the causal estimation of reform impacts in each local labour market, and ii) the approach for finding correlation patterns between local area characteristics and the strength and direction of reform impacts. Section 4 describes the data, and section 5 shows results, including a discussion of sensitivity analyses. Section 5 also discusses a number of issues relating to identification, especially selective sorting into place of residence and its implications for interpretation. Section 6 concludes.

2. The Reform Context

The Australian Government provides financial support to principal carers of a young child if they are on a low income, paid out through its federal agency *Centrelink* that also administers all other income support payments (such as payment in case of unemployment or disability). *Parenting Payment* is, in principle available to partnered (Parenting Payment Partnered – PPP) as well as to single parents (Parenting Payment Single – PPS). However, they are two separate payments with separate conditions for eligibility, separate income tests and separate payment amounts. Most importantly, the partner's income is taken into account when determining eligibility for PPP, while the former partner's income is irrelevant for eligibility for PPS. Because of that, a separation often marks the onset of eligibility for income support for carers of young children, typically a mother. PPS underwent a major reform in 2006, when eligibility criteria were tightened, and payment amounts were reduced for most parents. This reform, and its effect on mothers' employment outcomes, are the focus of this analysis.

Historically, PPS had been available to principal carers of a child under the age of 16; just before the reform, recipients were paid up to A$489 per fortnight if their other income did not exceed A$152 per fortnight; once they exceeded that threshold, the payment was reduced by 40ct in the dollar, so that no payment was made when the carer had an income of A$1372 per fortnight. The payment provided 24% of average full-time earnings for those with no other income, and some support for anyone with an income of up to 63% of average full-time earnings.[2] This system was in place for all applicants who separated on or before 30 June 2006.

Applications made for separation that occurred on or after 1 July 2006 had to meet a different standard: now PPS was only available to parents of a child who was 8 years or younger. However, most carers who were ineligible for PPS because their youngest child was older than the age threshold, could apply for an alternative income support payment: New Start Allowance (NSA), a payment generally aimed at people who were unemployed.[3] This payment does not require the recipient to have any caring responsibilities for children, regardless of their age. It does, however, differ from PPS in three important ways: i) NSA requires the recipient to participate in training and job search activities; ii) the maximum payment amount in July 2006

[2] In May 2006, average total full-time earnings across all industries were A$2180 per fortnight (Australian Bureau of Statistics, 2006).

[3] New Start Allowance was replaced early in the Covid-19 pandemic (in March 2020) by the Jobseeker Payment, with significantly higher payment rates. These increased payment rates have since been reverted back close to the pre-2020 payment level, with mostly the name change remaining in place.

was about 15% lower than for PPS, at $421 per fortnight; and iii) the taper rate was higher: the payment was reduced by 50ct for every dollar of income above A$62 per fortnight, and by 60ct for every dollar of income above $250 per fortnight; the payment was thus reduced to zero when a person's income exceeded $793 per fortnight – only 58% of the earnings threshold for receipt of PPS.

Note that applications based on separation that happened or before 30 June 2006 were grandfathered beyond the reform date; once deemed eligible, the applicant remained eligible even if their youngest child had been eight years or younger before the reform and turned 9 after the reform. This creates two cohorts of parents who are eligible for different income support payments, only depending on the date when they first registered their separation with Centrelink.[4] Table 1 summarises the payments a single parent can access, depending on their date of separation.

Table 1 – Eligibility for income support for single parents, by family circumstances and date of separation

	Youngest child younger than 8 years at separation	Youngest child 9 or older at separation
Separated on or before 30 June 2006	eligible for PPS until youngest child is 16	eligible for PPS until youngest child is 16
Separated on or after 1 July 2006	eligible for PPS until youngest child turns 8; then eligible for NSA if participation requirements are met	eligible for NSA if participation requirements are met

There are three ways in which this reform can affect individuals' welfare reliance and labour supply. First, at every level of earnings up to $1372 per fortnight (the threshold where the old payment cut out completely), families affected by the reform will have less total income (as a sum of earnings and income support) than grandfathered families. Taking individual labour supply as a given, this should lead to savings for the government in a purely mechanical way. Especially pronounced is this effect in the earned income region between $793 and $1372 per fortnight, where families before the reform would have been eligible for a partial payment of PPS but are now ineligible for the new payment NSA. Secondly, part of the logic of the reform

[4] There are some groups of single parents who are exempt from the new, tighter welfare regime and remain eligible for Parenting Payment Single regardless of their separation date; these are primarily applicants who are separating due to family violence or intimate partner violence, who are foster carers, who are carers to more than four children, or who are carers for a severely disabled child.

was that exactly this income effect should *increase* affected families' labour supply, as they are trying to make up lost welfare payments with earned income. If this desired behavioural adjustment does indeed occur, income support expenditure should drop further (in addition to the mechanical reduction mentioned before) while (some of) the income losses for families would be offset. However, there is a third channel through which the reform can change behaviour, income and government expenditure, and it works in the opposite and much less desirable direction: the tapering out of payments with increased earned income is an effective tax, and the reform drastically increased this tax. For mothers who earn between $62 and $152 per fortnight, the effective tax increased from zero before the reform to 50ct per dollar earned afterwards. Between $152 and $250, it increased from 40ct to 50ct per dollar earned, and from $250 to $793 it increased from 40ct to 60ct per dollar earned. These are very large changes in effective marginal tax rates, that could potentially lead to a large and unintended *reduction* in labour supply. Many studies from around the world show that the most promising way to improve employment outcomes for welfare recipients is to *increase* the payment with earned income (like, for example, the Earned Income Tax Credits in the U.S. does); the reform analysed in this paper, unfortunately, implemented the exact opposite policy. In combination, if the reform turned out to reduce labour supply *and* the level of income support paid at any given level of labour supply, this would mean that families are a lot worse off and potentially pushed below the poverty line.

One further important institutional feature to mention concerns a couple's reporting choice: a couple can, in theory, report a separation that did not actually occur, or report a separation on an incorrect date. Here it is important to note that both before and after the reform, a couple seeking to maximise the payments they receive, is always better off reporting a separation than reporting cohabitation, and the financial pay-off from misreporting is very similar in both policy regimes (Suziedelyte and Zhu, 2021). Even though misreporting may occur, there is no institutional reason why the extent of misreporting should change because of the reform.[5]

3. Estimation approach

The analysis is conducted in two parts: in the first part we estimate the effect of income support conditions that single mothers face, on their welfare receipt and earnings from employment in the years following the separation. We estimate the reform effect for each local area separately,

[5] In addition to misreporting a separation that did not occur, couples who are separated could also misreport the date when this happened. We will test for bunching in separations around the reform date to test whether this is an issue empirically.

using a regression discontinuity design approach. In the second part of the analysis, we use a machine learning approach to find correlation patterns between the local reform effect and an area's demographic, economic and social characteristics.

4.1 Estimating the reform effect on mother's welfare receipt

Our aim in the first part of the analysis is to get an unbiased estimate of the effect of reducing welfare benefits for single mothers, on their welfare dependency and earned income. As described in the last section, if parents with a youngest child between age 8 and age 15 separated, and applied for income support, on or before 30 June 2006, they were eligible for PPS until their youngest child turned 16; but if parents with a youngest child in the same age range separated and applied for income support on or after 1 July 2006, they immediately were eligible only for NSA – which has a lower maximum benefit, a larger taper rate, and more participation requirements. This reform created a sharp discontinuity in eligibility for income support depending on the dater of separation, and a sharp change in work incentives.

We exploit this discontinuity to identify the effect of the reform on mothers' welfare receipt and earned income in the years following the separation. The assumption underlying this estimation approach is, that the characteristics of mothers who separate just before or just after the date when the reform was implemented, are on average the same - except for their eligibility for PPS versus NSA. This assumption would be violated if mothers choose their separation date to ensure that they are subject to the institutional regime they prefer. We test this assumption using the McCrary test (see Section 6.1) and remove a small number of local areas where there is some evidence that manipulation of the separation date could be present in our data.

The running variable sep_i is the date of separation for mother *i*, normalised to be zero on the date when the reform took effect (1 July 2006). A dummy variable $post_i$ shows the mothers treatment status and takes on value 1 if the mother separated on or after 1 July 2006, and 0 otherwise. Six separate outcome variables $Y1_i, Y2_i, Y3_i, Y4_i, Y5_i\ and\ Y6_i$ (in what follows, for simplicity reduced to Y_i) measure total months of welfare receipt regardless of payment type for mother *i*, in the first one to six years after she separated. In a second set of results, $Y1_i, Y2_i, Y3_i, Y4_i, Y5_i\ and\ Y6_i$ will represent annual income earned.

We then estimate:

$$Y_i = \beta_0 + \beta_1 \cdot g(sep_i) + \beta_2 \cdot post_i + \beta_3 \cdot g(sep_i) \cdot post_i + \eta_i$$

where β_2 measures the reform effect. The function $g(sep_i)$ specifies how many observations left and right of the cut-off date are used for the analysis, and how much weight they are given. It is empirically determined using local linear regression (LLR, see Lee and Lemieux, 2010). There are two conflicting factors that need to be balanced when choosing the estimator's bandwidth, that is, how many observations are included in the estimation how far away from the cut-off point. First, the larger the bandwidth, the larger the risk that the main underlying assumption (that mothers who separate before and after the cut-off only differ in their treatment status and not in other relevant characteristics) is violated and the estimate is biased. Second, the smaller the bandwidth, the smaller the sample that can be used to estimate the reform effect, and the greater the estimate's variance. We follow Cattaneo et al. (2020) and select as optimal bandwidth which minimizes the mean squared error, separately before and after the implementation of the reform. Within these two bandwidths, triangular kernels are used to assign weights to all observations that enter the analysis with non-zero weight; the kernel weight is larger the closer to the cut-off date the mother separated.

This equation is estimated not on the national level, but separately for each local area. A 'local area' is defined at the so-called Statistical Area Level 4 (SA4), a geographic unit of analysis created and defined by the Australian Bureau of Statistics. SA4s are designed to represent functional labour markets, reflecting clusters where people both live and work. In major cities they comprise between 300,000 and 500,000 persons, and outside of major cities between 100,00 and 300,000 persons (Australian Bureau of Statistics, 2011). There are 87 SA4s in Australia. That is, the entire analysis, from testing for bunching in separation dates over determining the bandwidth to assigning kernel weights and estimating β_2 for the six outcome variables, is carried out 87 times. This process yields (87)*6 estimates of local reform effects for each outcome (welfare receipt an annual income earned), each applicable to a different area and a different point in time after separation. In Section 6.1, we discuss potential threats to identification and how they have been addressed.

4.2 Finding correlation patterns between local reform effects and local conditions

After carrying out the estimations needed to find local area effects, we will examine how the size of the reform effect varies across local areas. This part of the analysis will seek to answer two questions: first, how much variation in the reform effect is there? Is the geographic heterogeneity *large* enough that evaluations that rely on just a national average effect, would cover up an important phenomenon? And secondly, *what type* of areas are likely to experience a large or a small, a desirable or undesirable policy impact? We aim to construct an index that allows for the prediction of an earnings and welfare receipt elasticity per SA4, based on the area's characteristics. For this piece of exploratory research, the selection of indicators that describe a local area, is not theory driven. Instead, we have selected a broad range of characteristics from many different domains. We loosely follow the framework by Tanton et al. (2021), who develop an index to describe Australian communities and their relative levels of multi-layered disadvantage.

Notes on identification and interpretation

At this point, it is important to discuss what this study can reveal – and what it cannot. Our analysis purely uncovers correlation patterns and cann*ot* establish causal effects between any given local area characteristic, and the reform impact experienced in the area. There are two reasons for this limitation: if we find an area characteristic to be correlated with reform success or failure, (1) it could reflect other unmeasured differences in the characteristics of the region, and (2) it could reflect differences in the characteristics of residents rather than place-based attributes.

On the first issue, correlated variables cause a problem when we try to identify which dimensions of place matter. For example, areas that face high economic deprivation, are also more likely to exhibit a clustering of other characteristics that point to the deep and multi-layered nature of their disadvantage (Tanton et al 2021). As a result, if we find that a place-based factor is correlated with reform success this relationship could be causal – or it could be driven by its correlation to another important place-based factor.

The second issue is selective sorting. A single mother's decision concerning where to live is not random. i.e. place bundles the individual characteristics of those who reside there and that of the place itself. The people who reside in a place may have selectively moved to a region or they may have selectively stayed in a region (Oreopoulos 2003; Lindahl 2011; Hedman et al., 2017). For example, low-income individuals tend to live in relatively disadvantaged areas because high housing costs are often a barrier to entry into better resourced areas. Income levels may be correlated to unmeasured characteristics of the individual in a way that conflates individual and spatial disparities. Such sorting means that differences in reform effects between regions can simply be a manifestation of the inequality in how individuals respond to the reform. For example, regions may differ in their level of economic activity because manufacturing industries occupy different shares or because the residents have different levels of education. If our heterogeneity analysis finds that areas with stronger local economic activity tend to be where the reform effect size is the largest, we cannot say if it is because of the share of manufacturing in the local economy or because of the type of people who reside there. Thus, our heterogeneity analysis cannot inform policy makers about which type of place-based policies might work.

In short: the aim of the heterogeneity analysis is not to understand *why* some areas perform better than others in terms of the size of their response to the reform. Instead, we simply use heterogeneity analysis to a) get an idea of *the extent* of heterogeneity in reform effects and b) point to the most important dimensions along which reform effects vary.

Estimation Process: Searching for Spatial Heterogeneity - Univariate Analysis

We first estimate pairwise correlations of the local reform response (RDD estimates) and the characteristics of the local labour market.[6] Where we find a significant correlation between a given area characteristic – such as, for example, the population share with post-school qualifications in the area – and that area's reform response, we compare average reform effects for areas that show high versus low values of that characteristics to gauge the economic significance of the variation in reform impacts.

While the univariate analysis is straightforward, a number of empirical issues may arise. The first is that it can fail to detect important predictors that would otherwise be detected through a multivariate analysis. For example, the correlations along one dimension might be counteracted by correlations along another dimension: individually they are not significant, however, the partial correlations would reveal both features as being important. In the next section, we estimate partial correlations but minimise potential overfitting issues through a dimension reduction exercise. The second potential issue is multiple hypothesis testing from running 61 separate pairwise regressions. We address this by using a sample splitting approach where we estimate the model with one part of the sample and evaluate the model with another part of the sample.

[6] We partial out the effects of the age structure and remoteness status of the area through a Frisch-Waugh-Lovell procedure, both for the local reform effects and the local area characteristics. We use four categories for the age structure (based on the share of the population who are aged 0 to 14 years, 15 to 24 years, 25 to 64 years and above 65 years) and three categories for remoteness status (cities, regional areas and remote areas). We then partial out the level effects of age (omitting the category: 25-64) and remoteness (omitting the category: cities) along with the interactions between age and remoteness status.

Estimation Process: Searching for Spatial Heterogeneity - Multivariate Analysis

We complement this univariate analysis with a multivariate approach, by examining heterogeneity with an index of characteristics that is highly predictive of welfare reform effects in the first year after the policy implementation. A Machine Learning (ML) algorithm selects the characteristics that are most predictive of the local reform effect size. We then create an index based on an Ordinary Least Squares regression of the actual reform effects one year after the reform on those ML-selected features. We use the linear, Least absolute shrinkage and selection operator (LASSO) algorithm.[7] As the ratio of our sample size to number of features is relatively small, overfitting issues are more likely to occur. This is why we used cross-validation to select the model.

Cross-validation

Cross-validation is a resampling method whereby data is partitioned into 'testing' and 'training' data over multiple iterations. In each iteration a model is trained on the 'training' data, and the results are validated using the 'testing' dataset. Cross-validation is primarily used to gain insight into the predictive capabilities of the model, including how well it might generalise to new, previously unseen data, and to flag potential issues such as overfitting and multiple hypothesis testing.

We used Leave-One-Out cross-validation to determine the best hyperparameters for LASSO. For LASSO we optimised lambda. To optimise these, we performed a number of cross-validation steps, each using a distinct set of hyperparameters (grid-search cross-validation). The hyperparameters producing the lowest error were then selected. The hyperparameter optimisation step utilised mean squared error (MSE) as the error function to be minimised.

We performed hyperparameter selection leave-one-out (LOO) cross-validation. Leave-one-out cross-validation is a form of k-fold cross-validation where the number of folds equals the total number of data points. In each iteration 1 data point is used as test data, and the rest are used for training. When using LOO CV there is one iteration performed for each data point in the data set. This had the advantage of allowing every observation to contribute to hyperparameter determination, without having a dependence on grouping of the data as in x-fold cross-validation.

[7] In our sensitivity section, we also present results from estimating non-linear machine learning algorithms.

As the dataset already has all SA4s across Australia, we were less concerned with generating a generalisable result than with ensuring that the parameters selected as contributing significantly to the outcome measure were robust. Our further analysis therefore consisted of multiple workflows.

Data pre-processing

All variables were standardised (to have mean 0 and variance 1) prior to further processing. Standardization of the data ensures that features in the dataset have similar scales and that features which are numerically greater do not dominate other features in the regression.

Algorithms and Analysis Methods

LASSO is a linear regression model with L1 regularisation. This enables the algorithm to be utilised for not only regularisation, but also feature selection. Using the whole dataset, hyperparameters for the LASSO estimator was selected using a LOO process, minimising MSE, as described above.

Adjustment for the Estimation of the Outcome Variable

The dependent variable is based on the estimated reform effect from the RDD regression. We adjust for the fact that the estimates have differing levels of precision by weighting our regressions with the inverse of the variance of the estimates. However, we should note that the primary concern when estimating models with an estimated dependent variable, is heteroscedasticity and inconsistently estimated standard errors (Lewis and Linzer, 2005.) This does not constitute a major problem for this analysis, as we do not use the standard errors generated in the machine-learning process for statistical inference.

Spatial autocorrelation

We examine if spatial overfitting or spatial autocorrelation is an issue. We show there is no autocorrelation in the outcome. We test for spatial autocorrelation in the error terms between neighbouring areas, using Moran's I with a weighted matrix that assigns a weight of one to all combinations of local labour markets (SA4s) that were neighbouring each other geographically, and zero to all others (Moran, 1950). We do not find evidence of spatial autocorrelation.[8]

[8] We obtain the error terms from a regression of the income support reform effect outcome (and separately for the private earnings reform effect outcome) on the features that were chosen in the LASSO-procedure described above. We compare the

4. Data

5.1 Data used to estimate causal reform effects

We use novel data of the full population of Social Security System enrolees. The dataset is called "Data On Multiple INdividual Occurrences" (DOMINO). These data have several key benefits. First, they are high frequency with daily information of income support receipt status from 2000 to 2019. These repeat observations enable an analysis of the dynamics of welfare exit and entry. Second, they include both welfare recipients and individuals receiving other government transfers, so they also include individuals who did not receive income support payments. Third, they include rich information on over 32 million persons who had contact with the social security system anytime from 2000 to 2019, therefore representing a large dataset ideal for estimating RDD regressions.

Population

Australian federal social security records from 2000 to 2019 form the basis of our dataset. All social security payments are administered by the national welfare agency called Centrelink. There are over 32 million persons in these data who had any contact with the Centrelink system between 2000 and 2019. All registrants are 15 years or above as this is the minimum age of eligibility. The financial circumstances of these registrants vary greatly: some have high levels of financial needs, such as those who are in receipt of highly targeted income support payments; others have higher incomes and register with the social security system because they receive one of the non-income support payments described in the previous section, such as one-off government bonuses or cost-of-children payments.

Each individual is tracked over time on a highly frequent basis. For our main variable of welfare receipt status, we know the precise start and end date associated with payment receipt. A key advantage of this data structure for our study is that we can construct a precise picture of the duration and dynamics of welfare receipt over a long period of time.

The reason we observe such high-(daily)-frequency data is because income support (or welfare) payments are highly targeted. This means that recipients' eligibility for payments are assessed regularly, and recipients are required to report changes (such as to relationship status, earnings or living conditions) within 14 days of the change. The start and end dates of payment receipt

cross-products in the assigned weights (1 for SA4s that share a border and 0 for those that do not) with the cross-products of the errors from 300 matrices of randomly assigned weights. We do not find that the sum of the cross-products in the former case is statistically different from than the sum of the cross-products in the latter case.

are then recorded in our data. Specifically, recipients are required to report their financial circumstances and living arrangements to Centrelink on a regular (bi-weekly) basis by filling in a 14-34-page form that elicits information about the recipients' (and if applicable, partners') basic information (name, address, contact details, gender, date of birth, ethnicity, language, citizenship, arrival information), marital status and relationship event history, demographic information about their dependent children, accommodation details, employment and study details.

Sample

Our first step is to identify the universe of individuals who received any type of payment from Centrelink and who separated between 2001 and 2011. We use 5-year windows on either side of the cutoff date of 1 July 2006 to ensure there are enough observations to estimate each RDD regression.

At the time of separation, we select women who had at least one child under their care. Table 2 describes our sample by the date of separation – whether a mother separated 'pre' or 'post' the policy cutoff date. In total, we have 459,249 mothers in the sample. Columns (1) – (3) cover the overall sample: all mothers who had separated within a 10-year window around 1 July 2006. Columns (4) – (5) cover the sample close to the policy cutoff date. We manually restricted it to 3 months on either side of the cutoff. The purpose of this is to show that RDD, which automatically selects a small bandwidth around the cutoff plays an important role in ensuring that the characteristics of treated and control group (or 'pre' and 'post' groups) are balanced. Standard deviations are presented in brackets, and the standard errors of the differences are shown in parentheses.

Mothers, on average, separated when they were aged 33-34 years old, they had approximately 2 children each, their youngest child was roughly 4-5 years old at the time of separation, and the youngest child was just as likely to be female versus male. The majority of the sample are Australian-born, with a small percentage (approximately 6 percent) being Indigenous.

Table 2 illustrates the usefulness of an RDD approach that focuses estimation on the sample of mothers who separated within a smaller window around the cutoff. We use automatic bandwidth selection in our estimations – rather than manually restricting it to 6 months on either side of the cutoff (which we only do in the descriptive analysis for illustrative purposes). We see that mothers who separated before and after the cutoff do not differ along the pre-treatment variables that exist in the data (in Column 6). Whereas in the full sample (Column 3), differences appear for the number of children, the Indigenous status and the ages of the youngest child and the mother.

Table 2 – Descriptive Statistics

	Overall Sample			5-month window		
	Pre	Post	Change	Pre	Post	Change
	(1)	(2)	(3)	(4)	(5)	(6)
Children (number)	2.300***	2.199***	-0.110***	2.279***	2.229***	-0.055
	[1.361]	[1.313]	(0.008)	[1.373]	[1.299]	(0.035)
N	213483	220946	378383	20932	19593	34777
Child sex (female)	0.492***	0.492***	0.001	0.497***	0.491***	-0.003
	[0.5]	[0.5]	(0.001)	[0.5]	[0.5]	(0.003)
N	203460	213426	363214	20089	18838	33404
Child age	4.339***	4.924***	0.544***	4.869***	4.834***	-0.036
	[3.252]	[4.309]	(0.024)	[3.845]	[3.946]	(0.05)
N	225333	233916	397462	22065	20769	36549
AU-born (mother)	0.768***	0.768***	0.002	0.771***	0.770***	0.001
	[0.422]	[0.422]	(0.003)	[0.42]	[0.421]	(0.003)
N	225314	233890	397425	22065	20767	36547
Indigenous	0.066***	0.063***	-0.002*	0.064***	0.066***	0.003
	[0.248]	[0.243]	(0.001)	[0.244]	[0.248]	(0.003)
N	215614	224067	380688	20588	19552	34278
Mother age	33.886***	34.633***	0.619***	34.682***	34.506***	-0.196
	[7.526]	[8.537]	(0.054)	[7.997]	[8.126]	(0.128)
N	225333	233916	397462	22065	20769	36549

* p<0.1, ** p<0.05, *** p<0.01

Source: DOMINO data, authors own calculations. Notes: All age variables are measured at the time of relationship separation. All child characteristics are based on the youngest child in the family.

5.2 Data used to analyse the link between the reform effects and local area characteristics

We draw on five different data sets to describe a wide range of local area characteristics. First, we use the DOMINO dataset described above to create indicators of welfare receipt in the local area. Our second, and most important data source for local area characteristics is the 2006

Australian Census. The census is collected by the Australian Bureau of Statistics. It includes basic demographics such as age and household structure, and a broad range of household and individual characteristics such as income, education, employment, languages spoken at home, housing situation and health.

We chose the 2006 Census as our main data source for two reasons. First, it is possible to describe even relatively small areas well given that it includes the entire population. And secondly, the collection date on 8 August 2006 is very close to the reform implementation date. As mentioned, we follow the framework by Tanton et al 2021, who develop an index to describe Australian communities using 37 indicators. However, their analysis is primarily based on the 2016 Census, and not all of these items were included in the 2006 Census. We use the Census primarily to create measures of the economic situation in the local area: we can calculate the population share who is employed, unemployed or not in the labour force, and what sector and broad occupation employed people work in. The census also contains information on income and allows us to derive the population share that live below the poverty line, and we can describe the housing market in terms of the prevalence of renting and rent levels. The census also contains information on transport modes used by commuters, which we use as proxies for access to public transport. We also used the census for: languages spoken at home, self-reported ancestry and place of birth. We derive measures of the share within a region that exhibit these characteristics to create a measure of diversity. Specifically, we calculate the population share who speak English at home, the share who is born in Australia, and the share who reports their ancestry to be "Australian", respectively. In addition, we calculate the probability that two randomly drawn individuals in an SA4 (who are both not part of the respective "majority group"), speak the same language/have the same ancestry/are born in the same country as each other. And lastly, we know the population share who regularly does volunteer work, which we use as a proxy for community engagement.

We created some additional characteristics that were not covered by the Census using the Household, Income and Labour Dynamics in Australia (HILDA) Survey. HILDA is a long-standing household panel study that started in 2001 and surveys households annually. The study covers a very broad range of topics, including many that cannot be included in the Census. Relevant for our study, the survey includes information on attitudes towards working mothers, experiences of community, and accessibility and affordability of childcare. We construct measures of residents' satisfaction with their place of residence, which was recorded as self-reported satisfaction with the neighbourhood overall, with their perceived safety in the local area, and with their feeling part of the local community. Local area residents' average, self-reported difficulties with finding childcare in the right location, for the right hours or at an acceptable cost, was used as measures of access to affordable care. HILDA records respondents' attitudes to working mothers with a battery of statements with which respondents can agree or disagree; we used average levels of agreement in an area to measure beliefs held in the community. We also used this data set to create an indicator for the prevalence of families in an area relying on consumption loans to cover everyday expenses as a measure of financial hardship. Crucial to our purpose, the HILDA survey data are of excellent quality and is representative of the Australian population (Watson, 2012). We use data from wave 5, which was collected mostly in September 2005 and thus predates the reform but is still close to the reform date.

We measure access to childcare in a more objective fashion than via self-reported difficulties. Specifically, we created a measure of excess supply of places per child by combining data from Australian Children's Education & Care Quality Authority (ACECQA), and the 2006 census. ACECQA is a national authority tasked with administering the National Quality Framework (NQF) for children's education and care; they provide information on all registered childcare providers, how many places they are approved for and what quality rating the service has. This allows us to calculate the number of approved places per local area. Combining this with the number of preschool age children in the local area, which is recorded in the census, gives excess places per child (which may be negative or positive).[9]

We also use the Employee Earnings, Benefits and Trade Union Membership survey collected by the ABS in September 2006, to include a measure of women's average weekly earnings and the gender earnings gap. And finally, we use information from the National Centre for Vocational Education Research (NCVER) for access to post-school education institutions in the area. The full set of indicators included in the analysis are included in Appendix A.

[9] Unfortunately, historical data on approved childcare places is not available; the earliest data collection started from 2020. We turn to HILDA data to obtain information on which SA4s at the point of the reform can be plausibly described by the ACECQA data, and which cannot. We measure reported difficulties in finding childcare in 2005 and in 2020 for all local areas and convert both into deciles (based on HILDA). We then convert excess childcare places in 2020 (ACECQA) into deciles, thus mirroring the HILDA data construction. Where an area's excess childcare places in 2020, reported difficulties in 2020 and reported difficulties in 2005 "match up" (that is, deviate by at most four deciles from one another), we assume that the 2020 objective childcare supply data contains valuable information. For areas where a) there is no strong link between excess childcare places in 2020 and reported difficulties in 2020 (deviation of more than four deciles), or b) no strong link between reported difficulty in 2020 and reported difficulty in 2005, we assume that the 2020 childcare supply measure does not contain valuable information for the situation in 2005. We interact excess childcare places with a dummy variable that indicates for which area we have useful information (reliable, dummy=1) and for which ones we do not (unreliable, dummy=0).

5. Results

6.1 The local reform effect

Threats to identification

Before showing the reform effect itself, we test whether the identifying assumption – that mothers who separate just before and just after the cut-off differ only in their eligibility for the more generous welfare payment PPS and not in other, relevant characteristics – might be violated because of self-selection into the policy-regime. Compositional differences in treatment and control group mothers may arise from two potential sources: manipulation of the timing of separation and through selective sorting. We proceed in this section to test if these are problems.

Manipulation of the timing of separation

The only plausible way to self-select into the policy regime is by manipulating the date of separation. If separating couples do manipulate the timing of their separation to remain in the earlier, more generous policy regime, we should see more separations just before the cut-off date, and fewer separations just after.

We perform the standard test suggested by McCrary (2008). To perform the test, we use a local linear regression to smooth a histogram of separations over calendar time and obtain a density function. We can then test if there is a discrete jump in the density of separations at the point of the reform implementation. In addition, we use the test based on Cattaneo et al. (2018) who suggest a local polynomial estimation of the density. As each estimation is carried out separately for each local area, this tests also needs to be carried out separately, for each local area. We find no evidence for bunching of separations prior to the reform implementation, in 77 local areas. There are 87 SA4s in total in Australia. In 10 local areas, there is indeed a discontinuity. Although, in only 6 of these 10 do we find there to be more separations before the cut-off, than afterwards; in the other cases, more separations occurred after the cut-off.[10] We also test for balance in the pre-existing characteristics of mothers at the cutoff. Across six observable characteristics that exist in the data (number of children, age and sex of the youngest child, and the mother's own age, immigration status and indigeneity), multiplied by the 87 SA4 locations, we find only 28 instances where there is a statistically significant discontinuity (at

[10] Upon inspection, there is no discernible pattern to the local areas that failed the density-discontinuity test. There is a mix of regional, rural and urban areas, as well as a mix of relatively advantaged and disadvantaged areas according to income and poverty.

the 5% level) in the characteristics of mothers who sperate before or after the reform date (results of the balance tests in observable characteristics are included in Appendix B in Figure B.3). We would expect roughly 28 type-one error cases to arise at the 5 percent level. Nonetheless, we follow the conservative approach and only use reform effects of areas that exhibited no discontinuity.[11]

Selective sorting

Another potential source of bias for the RDD regressions for each local labour market is that mothers with a stronger attachment to the labour market may select themselves into areas where there are better prospects of finding and keeping employment, compared to those with a weaker attachment. This selective sorting may be more pronounced among mothers affected by the reform, compared to those who were unaffected. If so, there could be some areas where a change in earnings or welfare receipt conflates behavioural response enabled by an area's characteristics and a change in the composition of the area's resident population.

We take a number of steps to address this potential issue. First, we consider each mother's place of residence to be the area in which she resides at the time of the separation, regardless of whether she moves local labour markets later on. In doing so, we keep the population attributed to a local area constant over time and circumvent problems from a change in the population composition that arise after separation (and due to the policy). In attributing mothers' outcomes consistently to their source location, we thus include the mechanism of improving one's outcomes by moving to another local labour market, if such a phenomenon exists, as part of the total policy effect.

We also address selective sorting that may have occurred prior to and in anticipation of separation, and which may be more pronounced for treated group mothers compared to control group mothers.[12] Specifically, we analyse the density of separations at the cutoff date. Anticipatory sorting might be exhibited by a discontinuity in the density at the reform cutoff date. For example, treated group mothers, in the new policy environment, may seek to leave areas with relatively weak labour markets as they are confronted with a new and increased need to complement their income support payments with earnings form employment. Areas with

[11] We also checked for sensitivity of the results when we used the estimates for all local areas, including the ones that failed the density test (See Appendix Table A4 – section 2).

[12] The reform was announced on 15 May 2005 and implemented on 1 July 2006. Thus, there is a chance that mothers could anticipate the reform and selectively sort into an area 'at' the point of relationship separation.

better employment prospects may disproportionally attract single mothers with relatively high earnings potential or those with greater need to find employment, thus making it unclear the direction of the potential bias. We subsequently delete any local labour market areas where we observe any bunching or discontinuity at the RDD cutoff date. Furthermore, we test if mothers' observable characteristics in a local area have changed around the RDD cut-off date and find scant evidence of such selective sorting (see our strategy to deal with manipulation of separation dates, and Figure B.3).

Last, we directly test if the reform induced any changes in families' probability of moving between local areas. We repeat the RDD-estimation described before, but with a dependent variable that measures if a mother has changed her place of residence from one SA4 to another SA4 in the 1^{st}, 2^{nd}, …, and 6^{th} year after separation, as a function of the running variable separation date. We then focus our analysis primarily on years after separation in which no effect of the reform on moving decisions was found (which is the case for year 1, year 2 and year 3 after separation). A graphical representation of the results of the relocation analysis is included in Appendix B, Figure B2.

A likely reason for the lack of any significant movement across local labour markets (SA4 regions) in the first three years after the reform is because the sample of analysis is based on a relatively immobile population – single mothers with young children and who have a partner that they recently separated from. The unit of analysis (the SA4) is large enough that a move between areas would plausibly affect parenting arrangements. In such a scenario, separated parents in Australia are typically not legally allowed to relocate their children without a court intervention or the other parent's permission; this should significantly reduce the population's mobility at least in the short- to medium-term. Furthermore, relocation costs alone (as an addition to the financial shock of relationship separation), as well as the length of rental contracts (which are usually at least one year in length) could minimise the chance that single mothers in our sample engage in selective sorting in the short term. In the longer term (4 years from the reform), we begin to see evidence of movements across local labour markets. By then, mothers may face fewer financial and legal challenges to moving.

Results of RDD estimations for local areas

Typically, an analysis using a regression discontinuity design presents a graphical representation of the outcome variable against the running variable, to show the discontinuity occurs where it should occur. As this is not practical for 1,386 estimates[13], we additionally estimated the reform effect on the national level to show a graphical representation.[14] The overall reform effects on employment earnings, and welfare receipt are shown in Figure 1, for the first, second, …, to sixth year after the date of separation. Panel (a) shows that on average, the reform reduced welfare receipt in the first 365 days after separation by 1.13 months or 34 days. In the second (and third) year, the reform-induced reduction in welfare receipt is an additional 0.76 (0.78) months or 24 (24) days. The added reform effect then remains relatively stable, with only a slight decrease; however, as mentioned before at this point estimates may be affected by mothers' moving between labour markets.

A decreasing marginal reform effect is to be expected for two reasons: first, as time after separation progresses, more and more mothers leave welfare receipt, and once they have done so, any further reform effect in subsequent years can only operate through the channel of reduced re-entries. If a single mother has a high chance of remaining independent from income support once she has alternative income sources – usually by finding employment, re-partnering or both – the scope for further reform effect is limited. And secondly, while mothers with the youngest child who is eight years old, find themselves in different policy regimes for up to seven years depending on their date of separation, mothers whose oldest child is 15 years at separation, are exposed to the treatment of the reform for only one year. That is, the pool of mothers for whom there is a difference in policy regimes depending on their separation date, shrinks every year, as their youngest children turn 16. Combined, these two mechanisms lead to a decreasing marginal reform effect.

In terms of welfare receipt, the reform achieved the intended effect. However, the reform also aimed to increase employment income among mothers with young children. We find evidence to suggest that this second goal did not materialise. In fact, the reform resulted in a significant *reduction* in earnings on average, especially in the first three years after separation (see Figure 1, panel (b)). This unintended reform effect is large. In the first year after separation, mothers

---

[13] 77 local areas times three outcomes at six points in time. We display these separate local labour market-level RDD reform effects in the form of 1,386 RDplots, which can be found at this website: https://annawzhu.github.io/Broadway_Zhu_Welfare_Geographies/ On this site, we display the results for each local labour market (SA4), grouped by the State or Territory. Within each SA4, we show the RDPlots for three outcomes (Income Support, Earnings and Employment) and within each outcome, the plots for each of the 6 years post-reform.

[14] For the national estimate, we also performed the same McCrary density test; no evidence of manipulation of separation dates was found. A graphical representation of the test is included in Figure B1 in the Appendix.

subjected to the new rules earn $800 less than grandfathered mothers did; in the second and third year, this earnings loss further increases to $1,100.

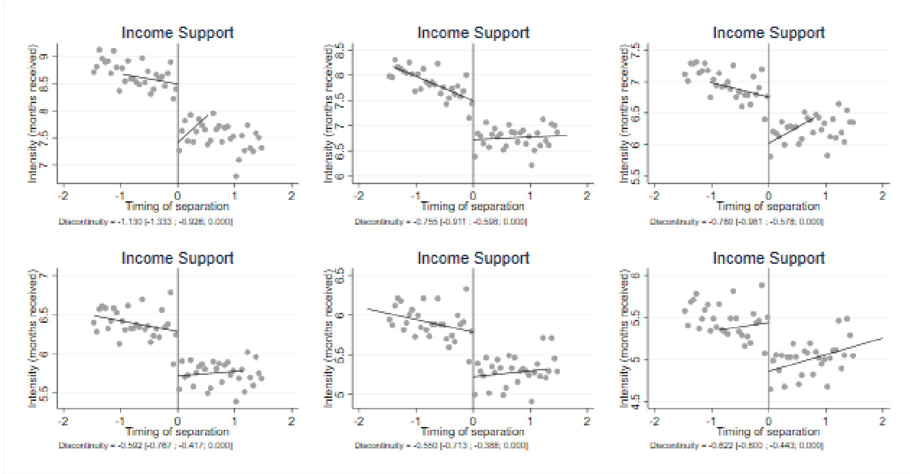
Thus, we show that the reform's effect on reduced welfare reliance does not stem from improvements in employment earnings. A possible explanation is that while the payment level is lower after the reform (expected to increase employment via reduced non-labour income), the taper rate is higher. This could discourage employment – and hence independence from welfare receipt. While PPS was reduced by 40ct per dollar of income earned, the new payment NSA is reduced by 60ct per dollar, and this reduction in benefits begins at a lower threshold. In other words, the reform increased the effective marginal tax rate faced by mothers with young children who receive income support, by a full 20 percentage points if they earned more than $152 per fortnight, and by 60 percentage points if they earned between $62 and $151 per fortnight. The disincentives set by these very high marginal tax rates can greatly exceed any increased incentives to work from reduced income. This finding aligns with international findings on the effects of welfare reform: reductions in effective marginal tax rates are more effective in setting incentives to work than a reduction in overall benefit levels. In fact, income tax credits, or an increase in payments for every dollar earned, has been shown to be superior to other policy designs (Broadway et al. 2021). Instead, the Australian Welfare-to-Work reform took the opposite approach and greatly increased effective marginal taxes on the lowest income earners.[15]

Behind this sizable reform effect on the national level, is large variation in reform effects at the local level. Figure 2 shows the distribution of reform effects across 77 local areas. In panel (b) we can see that the national average reduction in earnings by $800/year, ranges from a $6000 income *loss* to a $4000 income *gain*.

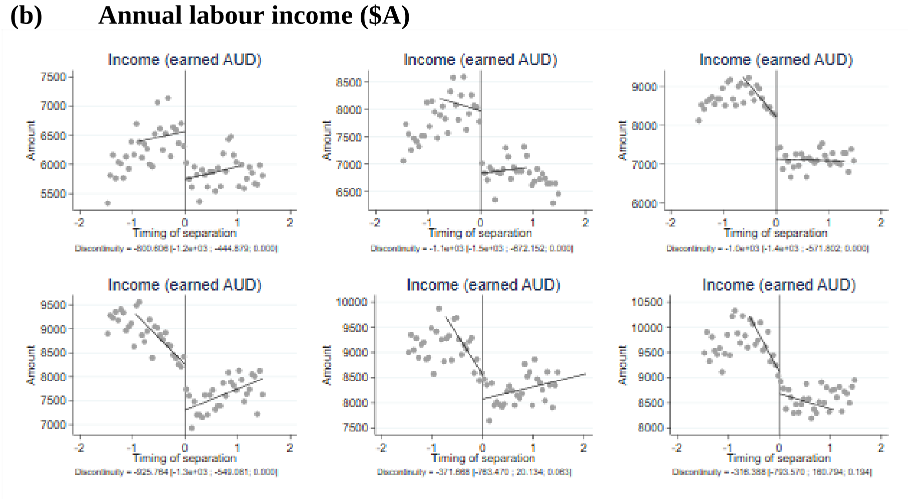
Figure 1 Discontinuity in outcomes 1 to 6 years after separation, before and after reform

(a) Income support receipt (months)

[15] There are two other possible pathways for increased welfare reliance in conjunction with decreased earnings: 1) It is possible that single mothers who also have a disability, would usually not apply for a disability support pension (DSP) as the application process is significantly more onerous than that for the PPS, but the financial benefits were identical. After the reform, single mothers with a disability do have a financial incentive to apply for DSP over PPS as PPS is now paid at a lower rate. At the same time, DSP can 'lock' recipients into welfare receipt: a recipient who takes up employment only to discover that they are not able to keep that activity level, could lose eligibility in the process and have to undergo the stringent medical testing again. 2) If the reform induces single mothers to re-partner in order to escape poverty – instead of entering a potentially lengthy search for employment – but the romantic relationships they enter are inherently less stable than an employment relationship would have been, welfare reliance some months or even years after the initial re-partnering decision could be higher than they otherwise would have been.



(b) Annual labour income ($A)



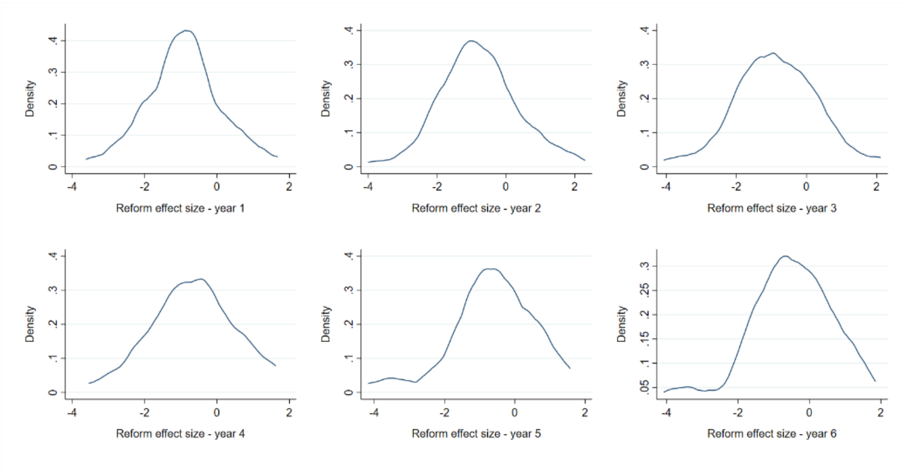
Source: DOMINO, authors own calculations

Notes: The figure shows months of welfare receipt (any benefit type) (panel a), and total annual labour income (panel b) in the first, second, …, sixth year after separation, depending on separation date. Day zero is the day when the reform took effect (1 July 2006). The estimation sample includes mothers of children below age 15, who separated within five years of the reform date.

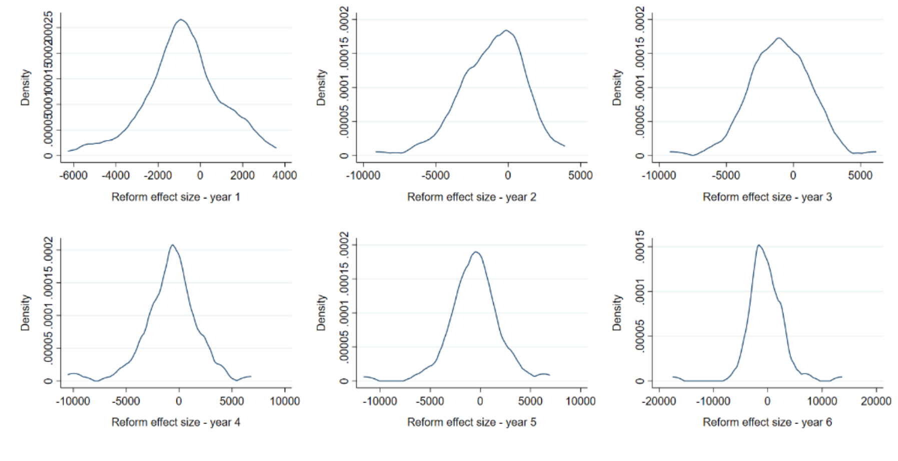
Likewise, in the second year after separation, the majority of SA4s showed a negative and unintended response to the reform, but we also did find improved earnings outcomes in a sizable minority of cases. Welfare receipt decreased in most areas (panel (a)) but increased in others. In some places, the unintended reform impacts on employment earnings are so severe, that the targeted population is now more likely to rely on welfare payments, at now lower payment levels and with less additional earnings.

Figure 2 Reform effect on outcomes by local area (kernel density)

(a) Income support receipt (months)



(b) Annual labour income ($A)



Source: DOMINO, authors own calculations

Notes: The figure shows the kernel density of reform effects on outcomes in the first, second, …, sixth year after separation, across local areas. The kernel is based on 77 estimated local effects. See notes to Figure 1.

To be sure that the spread in estimated local area reform responses measures real heterogeneity in effects, we generate the reform effect estimates for each local area using dates that were not the true policy cutoff date. Using this ‘placebo’- test yields a distribution of estimates that should result purely from random noise. We find that this distribution of values has a mean of zero and is far narrower, than the distribution of real effects we find at the actual policy implementation date. This applies to both the outcome of income support and earnings. We will also show in the following section, that the heterogeneity in reform impacts can be meaningfully connected to local area characteristics – which should not be possible if the heterogeneity in effects were purely the result of random chance.

5.2 Local reform effects and local area characteristics

Univariate heterogeneity analysis

We first test the correlation between a given area characteristic and the area's reform effect, separately for each local area characteristic described in Table 3 and the two outcomes. All variables were standardised to the average age structure and remoteness status "major city", but beyond this, it is a univariate analysis. Table 4 shows local area characteristics that we found to be (negatively or positively) associated with a reform effect on welfare receipt in the first three years after separation, at the 10%-level. [16]

The table shows the difference in RDD-estimated reform effects for areas above the 50th-percentile of the characteristic in question, compared to areas below the 50th -percentile. For example, take the result for "average weekly earnings, women": it shows the difference in predicted reform effects for areas assuming women earned on average more than $642.38/week in 2006, compared to areas where they would have earned on average less than $642.38/week (=the 50th-percentile of areas ranked by average women's earnings, corrected for age structure and remoteness). The results quantify the magnitude of important features and their relationship with reform effects and give an indication of their economic significance.

Note that the more 'successful' the reform was in reducing welfare receipt the more likely we are to see a negative sign and a large absolute value. A stronger *negative* effect for one group of areas compared to the other, thus means that those local areas had a *stronger* welfare-reducing reform effect.

We primarily find that there is a stronger reduction in income support receipt in areas with better economic opportunities. In areas where women's earnings are higher instead of lower than the 50th percentile, the reform reduced Income Support receipt by an additional 19, 22 and 18 days in the first, second and third year after separation; compared to an overall reform effect of 34 days in the first year and 24 in the second and third year. We also find a stable link between the size of the welfare-reducing reform effect and an areas' economic advantage based on four indices constructed by the Australian Bureau of Statistics.[17] In areas above the median of each index (where higher values always indicate more advantaged areas), based on our

[16] Results for the fourth, fifth and sixth year after separation (that could be affected by selective geographic relocation processes) are included in Appendix C – Additional results.
[17] The four indices are Index of Relative Socio-Economic Disadvantage (IRSD), Index of Socio-Economic Advantage and Disadvantage (IRSAD), Index of Education and Occupation (IEO) and Index of Economic Resources (IER).

estimates we would expect welfare receipt to be reduced by an additional 14 (12 and 16) days in the first (second and third) year after separation, compared to areas below the median. While we cannot disentangle exactly which aspect of economic advantage is the main driver behind this phenomenon (and it might well be a combination of many correlated factors), it is a remarkably stable pattern that we find across a wide range of proxies for “advantage”. For example, in areas with higher household income, less poverty, a smaller population share who receive income support or unemployment benefits, more high-skilled and less low-skilled occupation, or higher women’s earnings, we are more likely to see the intended reform effect of reducing welfare receipt; in contrast, in poorer areas with fewer economic opportunities this was less likely to happen.

In addition, we find that areas with more GPs per head of population appear to perform “better”, which could be a real response to better access to health services or a statistical artifact, where GPs relative to population density, soak up an effect of population density not sufficiently dealt with by our remoteness indicators: areas with very low population density tend to have more GPs per 100,000 persons than high-density urban areas. We also find stronger reform effects in areas where there is more agreement with the statement “Many working mothers seem to care more about being successful at work than meeting the needs of their children”, however, this effect only appears in one year.

The results for the reform’s effect on annual income point tell a similar story: it is areas with better economic opportunities that experience a less negative effect of the new welfare rules. In areas where the share of the employed population with a postschool qualification is 42.5% or higher (the 50th-percentile cutoff) the reform’s effects on annual earnings is $521.19 more positive than in areas where fewer people have a postschool qualification. Keep in mind that this is relative to an average reform effect of -$800 in the first year: the reform effect in both types of areas is overall negative, but the negative shock is less pronounced in areas where more people have postschool qualifications. We find variation in the reform effect of a similar magnitude for areas with more high-skilled – rather than low-skilled – occupations. However, we do not find the strong link between *poverty* or *disadvantage* with changes in annual income, that we found for welfare receipt. Whether or not areas have more severe unintended results in terms of income appears more closely connected to human capital, and specifically to qualifications and occupations.

Table 4 Differences in reform effects on welfare receipt, for areas above instead of below the median on a given characteristic (First, second and third year after separation)

	Year 1		Year 2		Year 3	
	Diff.	p-value	Diff.	p-value	Diff.	p-value
Access to and need for services						
Health services and demand						
Number of GPs working in SA4, per 100,000	.184	.060	.365	.006	.332	.003
population share that needs help with core activities			.444	.050		
Opportunities						
Earnings						
average weekly earnings, women	-.632	.042	-.709	.012	-.599	.025
Occupation level						
share of population with a low-skilled occupation			.400	.080		
share of population with a high-skilled occupation			-.607	.075		
Rental market						
average weekly rent			-.202	.076		
Income/poverty						
average equivalised household income			-.086	.060		
share of households with equivalised income below national poverty line			.013	.096		
Income support receipt						
population share who received any income support for at least one day in last year			.305	.027	.276	.042
population share who received unemployment benefits for at least one day in last year			.293	.071		
population share who received any income support for at least six months in last year			.183	.041	.230	.058
population share who received any income support for all of last year			.366	.047	.438	.065
Overall (dis)advantage						
Index of Relative Socio-Economic Disadvantage (IRSD)	-.465	.005	-.387	.052	-.514	.020
Index of Socio-Economic Advantage and Disadvantage (IRSAD)	-.465	.005	-.387	.057	-.514	.022
Index of Education and Occupation (IEO)	-.465	.004	-.387	.051	-.514	.021
Index of Economic Resources (IER)	-.455	.004	-.315	.054	-.456	.021
Community						
Values						
Average agreement [1-7]: Many working mothers seem to care more about being successful at work than meeting the needs of their children	-.287	.065				

Source: DOMINO and other data sources described in Section 5. Authors own calculations

Notes: The table shows how the reform impact on months of welfare receipt (any benefit type) in the first, second, and third year after separation, differs for two types of areas: those where a local area characteristic has a value above the median, compared to those where the characteristic is below the median. The analysis is univariate and does not hold other local area characteristics constant. The table includes results for every characteristic for which we found a statistically significant correlation with the estimated local area reform impact. For information on how local area reform impacts were estimated, see section 4, and for information on the quality of the estimates see section 6. For information on underlying data sources describing local area characteristics, see section 5.

Table 5 Differences in reform effects on annual earnings, for areas above instead of below the median on a given characteristic (First, second and third year after separation)

	Year 1		Year 2		Year 3	
	Diff.	p-value	Diff.	p-value	Diff.	p-value
Access to and need for services						
Means of transportation/ Commuting modes						
population share that works from home	1047.696	.005				
Childcare						
Excess childcare places per child (preschool age only)	656.755	.037	520.062	.098	673.329	.030
Excess childcare places per child (preschool age only), if highly reliable	279.672	.092				
Health services and demand						
Number of GPs working in SA4, per 100,000	351.577	.040	876.127	.017		
Opportunities						
Education level						
share of population with a post-secondary qualification	521.194	.068	1067.840	.025	1068.275	.008
Occupation level						
share of population with a low-skilled occupation	-746.056	.078	-957.680	.029		
share of population with a high-skilled occupation	646.382	.040	841.697	.053	715.457	.035
Sector						
share of population who works for local government					-415.459	.063
Rental market						
average weekly rent	552.163	.038				
average weekly rent, relative to average local income	264.561	.067				
Community						
Values						
Average agreement [1-7] Mothers who don't really need the money shouldn't work	-297.842	.094	-568.851	.078	-689.686	.068
Average agreement [1-7] As long as the care is good, it is fine for children under 3 years of age to be placed in childcare all day for 5 days a week [reversed]	898.457	.099			1048.067	.059
Community engagement						
population share who did unpaid voluntary work in the last twelve months	667.855	.031				
Reported satisfaction [0,10] with how safe you feel					-644.617	.033

Source: DOMINO and other data sources described in Section 5. Authors own calculations

Notes: The table shows how the reform impact on annual earnings from employment in the first, second, and third year after separation, differs for two types of areas: those where a local area characteristic has a value above the median, compared to those where the characteristic is below the median. The analysis is univariate and does not hold other local area characteristics constant. Also see Notes to Table 4.

Access to services also seems to play a role for changes in income that it did not play for changes in welfare receipt: areas with better access to childcare have significantly lower unintended reform impacts on annual earned income. The difference in areas along the dimension of “childcare access” is about as large as what we find when comparing areas according to their human capital.

We also find a relatively clear pattern that the reform’s effect varies by attitudes and values. In areas where agreement with the statement “Mothers who don't really need the money shouldn't work” is high, the reform has a significantly worse effect than in areas where it is low. Where many people agree that “childcare is fine for children under 3”, however, the reform had a smaller unintended effect on overall earnings.

It appears that while the reduction in welfare receipt was mostly tied to an area’s financial position, the effect on earnings is much more closely related to individual attitudes, skills and their surrounding infrastructure that enables employment. To reduce the set of results presented, in what follows, we will focus our attention on the reform’s effects on earnings rather than welfare receipt. Reform impacts on welfare receipt are included in Appendix C.

Multivariate heterogeneity analysis based on a machine-learning generated index of reform success

We then move from the univariate analysis to a multivariate one. We use the machine learning process described in section 4 to find a model that is still able to detect where one variable might have associations with local reform effects above and beyond the correlations shared with another variable, while avoiding overfitting. Almost all variables ended up being assigned a positive weight in the index. The only exceptions were: average equivalised household income, population share who received unemployment benefits for all of last year, two of the indices of disadvantage (IRSD and IER) and one of the indicators of “values”.

The top panel of Table 6 then shows the degree of heterogeneity in reform effects that we can see along the predictive index. We rank SA4s by their predicted reform impact on earnings, based on local area characteristic; we then split the SA4s in quartiles. We report the average of the *RDD-estimated* reform impacts on earnings within each quartile. This yields a measure of the quality of the predictive index and the degree of heterogeneity in reform effects we can observe along the most relevant socioeconomic dimensions. We find that in the SA4s with predicted reform success in the bottom two quartiles, we estimated a reduction of welfare receipt by $3061.40 and $1027.39 in the first year after the reform. In contrast, in the quartile of SA4s that were predicted to have the second lowest reduction in annual earnings, according to their area characteristics, the estimated reform impact is only -$632.67, and in the top quartile we see an increase in earnings by $1433.24. All this is in comparison to an average national effect of -$800/year.

The third panel in the table gives three indicators of the predictive quality of the index created by the machine-learning model. First the average deviation of an area's rank in terms of reform impact as predicted by the characteristics-based index, from its rank according to the actual, RDD-estimated reform impact. On average, an area's predicted rank deviates from its RDD-estimated rank by 7.6 in the first year – which is substantially lower than the 25.66 that would be expected if the predictive index were random and contained no systematic information. Secondly, we show the share of SA4s whose predicted rank was less than 5 ranks from their observed one (to be expected for only 10.6% of all SA4s if ranks were randomly allocated), which is high at 62% in the first year, but the reduces to just over 40% in the second and 34% in the third year. And finally, we report the R^2 of a simple linear regression of observed, RDD-estimated reform impact, on the area-characteristics based predictive index, which is very high at first but then quickly declines.

Table 6 Heterogeneity of RDD-estimated reform effects on annual earnings, along the distribution of predictive index of reform success

	Average RDD-estimated reform effect by index prediction					
	Year 1		Year 2		Year 3	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-3061.40	1350.50	-3670.37	2037.07	-2987.46	2171.83
Second quartile	-1027.39	600.34	-1155.20	1329.73	-1155.67	2090.06
Third quartile	-632.67	660.98	-659.90	1356.15	-658.51	1720.09
Top quartile	1433.234	1024.47	1009.10	1294.79	731.40	1955.25
	Relationship between index predicted and RDD-estimated local reform effects					
	Year 1		Year 2		Year 3	
Average absolute deviation between predicted rank and observed rank	7.61		11.17		14.28	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.62		0.47		0.35	
R2 of regression of RDD-estimated reform effect on index prediction	0.87		0.57		0.32	

Source: DOMINO and other data sources described in Section 5. Authors own calculations

Notes: The table shows how the reform impact on annual earnings from employment in the first, second, and third year after separation, differs for areas with high and low levels of the index predicting reform success, generated by the Machine - Learning process described in Section 4.2. For information on how local area reform impacts were estimated, see section 4, and for information on the quality of the estimates see section 6. For information on underlying data sources describing local area characteristics, see section 5.

Overall, our analysis draws out two insights. First, behind the national average reform effects (which deemed this policy as successful in achieving its intended effect) hides the more nuanced narrative that the reform was successful in some areas and less successful in others. Policymakers ought to be aware that the story experienced on the ground, locally, can vary considerably from that reported in typical, nationwide evaluations. Second, this variation in reform impacts can be meaningfully connected to the degree of economic disadvantage in a place.

5.3 Sensitivity analysis

We confirm that our results are not meaningfully changed to the following changes in the estimation procedure:

i) if results are calculated using all 87 local labour markets, including the ones that failed the McCrary density test

ii) if results are calculated using a simple local linear regression (OLS) , that assigns equal weight to all individuals in our window of observation.

iii) if results are calculated using a small sub-population of public housing residents, for whom geographic location can be plausibly assumed to be exogenous due to specific policy settings

iv) if we use a more flexible algorithm to develop the index of reform success (Gradient Boosting regression).

No sensitivity check revealed a substantial change in results or conclusions to be drawn. Details are included in Appendix D.

6. Conclusion

Geography matters to the success or failure of an exogenous change in a country's institutional settings (from a nation-wide welfare reform). Using a new dataset which captures the near-universe of single mothers targeted by one of the largest welfare reforms in Australia, it shows that this reform did not have the intended effect in geographic regions that were relatively disadvantaged.

We find significant heterogeneity in the reform effects across the country. In regions with stronger economic opportunities (such as higher household income, less poverty, a smaller population share who receive income support or unemployment benefits, more high-skilled and less low-skilled occupation, or higher women's earnings), we find stronger intended reform effects - in terms of reduced income support receipt. This contrasts with the impacts in the more disadvantaged regions, which exhibited far weaker responses (even increases in income support receipt).

The all-round lack-lustre impacts of the reform in the most disadvantaged geographic regions is an unintended effect of the policy. Another unintended effect is along the margin of employment income, which also appear in the estimates of the impact at the national level. While the reform achieved its direct aim of reducing total reliance (months) on Income Support receipt, it did not achieve the indirect aim of increasing employment income. The reform created an incentive for mothers to reduce earnings because it increased the taper rates, which mechanically increased the effective marginal tax rates for earnings. This finding is in line with the evidence in the literature showing Earned Income Tax Credits (EITC) (which lowers effective marginal tax rates) has the opposite effect of incentivising employment earnings.

Heterogeneity in the effects of the policy on employment earnings also emerge. However, gaps in the reform effects are most pronounced between regions with high versus relatively low levels of human capital (in terms of education and income), as well as access to childcare services and according to attitudes. For the outcome of employment income, unlike for income support levels, the level of poverty or disadvantage in an area is not an important predictor of the size of the welfare reform response. Importantly, the local region effect sizes are substantively and statistically different to the national average. For example, the national average reduction in earnings in the first year was by $800/year, however, this ranges from a $6000 income loss to a $4000 income gain across different geographic areas. Thus we conclude that estimating reform effects at the national, average level masks important heterogeneity.

We also find heterogeneity in reform effects based on more complex indices of advantage and disadvantage. We used indices often-used (and created by the Australian Bureau of Statistics) in Australian studies of spatial analysis and our own Machine-Learning (ML) derived indices. The latter index measures the propensity of reform response for each geographic region. These ML-driven indices may help policy makers when targeting place-based policies because success in such policies requires us not only to know which areas are most in need, but also which areas are most likely to respond. More understanding here enables scarce resources to be used effectively.

Overall, we provide evidence that stricter welfare policies (reduced generosity and increased participation requirements) are least successful in relatively disadvantaged areas. This finding thus raises serious equity concerns, as it suggests an exogenous change in the institutional settings (from a nation-wide welfare reform) hits hardest those regions that are already severely disadvantaged. To address such equity concerns, rather than directing place-blind policies, policy-makers could tailor policies to an area's strengths and weaknesses. For example, job-search requirements (number of jobs welfare clients are required to apply for) could be adjusted according to the level of job opportunities in the area. In addition, further supports could be implemented in relatively disadvantaged areas to complement nation-wide policies.

Further research is needed to explain the source of the heterogeneity. In particular, spatial heterogeneity conflates differences in people and differences in area characteristics. Also, future research can demonstrate mobility and job opportunities on the local level to be causally connected to reform success. Policies that can support an individual's mobility and allow people to move to where such job opportunities exist are a potential policy lever for achieving greater equity in the outcomes of welfare reform.

References

Aizer, A., Cho, S., Eli, S. and Lleras-Muney, A. (2023) [Forthcoming]. The Impact of Cash Transfers to Poor Mothers on Family Structure and Maternal Well-Being, American Economic Journal: Applied Economics.

Auerbach, A., Card D., and Quigley, Quigley J.M. (2006) Public Policy and the income Distribution. Russell Sage Foundation.

Australian Bureau of Statistics (May 2006) 'Table 2: Average weekly earnings, Australia (dollars) – seasonally adjusted' [data set] , Australia, accessed 28 February 2023.

Australian Bureau of Statistics (July 2011) 1270.0.55.001 - Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2011 , ABS Website, accessed 28 February 2023.

Becker, Sascha O., Peter H. Egger, and Maximilian v. Ehrlich. 2013. "Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects." American Economic Journal: Economic Policy 5 (4): 29–77.

Biddle, N., & Montaigne, M. (2017). Income Inequality in Australia–Decomposing by City and Suburb. Economic Papers: A journal of applied economics and policy, 36(4), 367-379.

Blank, Rebecca M. (2002). "Evaluating Welfare Reform in the U.S." Journal of Economic Literature 40, no. 4 (December 2002): 1105-66.

Blank, Rebecca M. (2006). "What Did the 1990s Welfare Reforms Accomplish?" In Poverty, the Distribution of Income and Public Policy, eds. 2006.

Boustan, Leah Platt. (2016). Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets. Princeton: Princeton University Press.

Broadway, B., LoRiggio, T., Ryan, C., and Zhu, A (2021). Literature Review on the Impact of Welfare Policy Design on Children and Youth. Journal of Economic Surveys, 36(4), 809-840.

Cattaneo, M.D., Jansson, M., Ma, X., (2018). Manipulation testing based on density discontinuity. Stata J. 18 (1), 234–261.

Cattaneo, M., Idrobo, N., & Titiunik, R. (2020). A Practical Introduction to Regression Discontinuity Designs: Foundations (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge: Cambridge University Press. doi:10.1017/9781108684606

Chetty, Raj, and Nathaniel Hendren. (2018). "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." The Quarterly Journal of Economics 133.3: 1107-1162.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. (2016) "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment." American Economic Review 106.4: 855-902.

Chyn, Eric (2018). "Moved to opportunity: The long-run effects of public housing demolition on children." American Economic Review 108.10: 3028-3056.

Chyn, Eric, and Lawrence F. Katz (2021). "Neighborhoods matter: Assessing the evidence for place effects." Journal of Economic Perspectives 35.4: 197-222.

Collins, William J., and Marianne H. Wanamaker (2014). "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data." American Economic Journal: Applied Economics 6 (1): 220–52.

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. (2018). "The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns." American Economic Journal: Applied Economics 10 (2): 202–33

Deryugina, Tatyana, and David Molitor. (2020). "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina." American Economic Review 110 (11): 3602–33

Fisher, H. and Zhu, A.,(2019). The effect of changing financial incentives on repartnering. The Economic Journal, 129(623), 2833-2866.

Friedman, J.H. (2001). Greedy function approximation : a gradient Boosting machine. Annals of Statistics 29(5):1189–1232.

Gong, X. & Breunig, R. (2014), `Channels of labour supply responses of lone parents to changed work incentives', Oxford Economic Papers 66(4), 916-939.

Hartley, Robert Paul, Carlos Lamarche and James P Ziliak. (2022). "Welfare reform and the intergenerational transmission of dependence." Journal of Political Economy 130(3):523–565.

Hedman L., van Ham M. & Tammaru T. (2017) Three generations of intergenerational transmission of neighbourhood context. IZA Discussion Paper No. 11218 (www.iza.org).

Henman, Paul. (2008) "Equivalent salaries in five Australian capital cities." Australian Journal of Social Issues 43.4: 615-630.

Hershbein, B. and Stuart B. (2023) [Forthcoming], The Evolution of Local Labor Markets After Recessions, American Economic Journal: Applied Economics.

Laliberté, Jean-William. (2021). "Long-Term Contextual Effects in Education: Schools and Neighborhoods." American Economic Journal: Economic Policy 13 (2): 336–77.

Lee, D.S., Lemieux, T., (2010). Regression discontinuity designs in economics. J. Econ. Lit. 48 (2), 281–355

Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income - evidence from Sweden, The Journal of Economic Inequality 9(2): 207–226.

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. (2013). “Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity.” American Economic Review 103 (3): 226–31.

McCrary, J., (2008). Manipulation of the running variable in the regression discontinuity design: a density test. Journal of Econometrics 142 (2), 698–714.

Moran, P. (1950). Notes on Continuous Stochastic Phenomena. Biometrika, 37 (1), 17–23.

Modai-Snir, Tal, and Maarten van Ham (2018). "Neighbourhood change and spatial polarization: The roles of increasing inequality and divergent urban development." Cities 82 : 108-118.

Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. The quarterly journal of economics, 118(4), 1533-1575.

Smith, Crichton, Nick Parr, and Salut Muhidin (2019). "Mapping schools' NAPLAN results: A spatial inequality of school outcomes in Australia." Geographical Research 57.2: 133-150.

Smolensky, Eugene, and Jennifer Appleton Gootman, eds.(2003). Working Families and Growing Kids: Caring for Children and Adolescents, Chapter 7. National Academies Press.

Suziedelyte, A. and Zhu, A. (2021), The intergenerational impact of reduced generosity in the social safety net, Journal of Economic Behavior & Organization 192(1), 1-24.

Tanton, R., Dare, L., Miranti, R., Vidyattama, Y., Yule, A. and McCabe, M. (2021), Dropping Off the Edge 2021: Persistent and multilayered disadvantage in Australia, Jesuit Social Services: Melbourne.

Watson, N. (2012), Longitudinal and Cross-sectional Weighting Methodology for the HILDA Survey, HILDA Technical Paper Series 2/12, Melbourne Institute: Applied Economic and Social Research, Melbourne, Australia.

Appendix A – Full list of local area characteristics included in analysis

Characteristic	Data source	Level	Measure
Access to and need for services			
Means of transportation/ Commuting modes	ABS Census 2006	SA4	Population: age 15+, employed ● population share that takes, train, bus, ferry or tram to work ● population share that walks or cycles to work ● population share that works from home ● population share that relies on private transport
Education institutions	NCVER	SA4	● Number of post-secondary accredited education institutions (without universities)
Childcare			
Difficulty finding childcare, including outside school hours care	HILDA, wave 2005	SA4	Population: HILDA respondents (age 15+) with child <15 years in household, who used or thought about using childcare Reported difficulty from 0 to 10, average over SA4, in deciles ● Finding a childcare centre in the right location ● Getting care for the hours you need ● The cost of childcare ● All of the above combined
Excess childcare places per child (preschool age only)	Enrolments: ABS, Pre-school education 2021 Approved places: ACECQA (Sept 2022) Pre-school population: Census 2021	SA4	Approved childcare places minus enrolments in SA4, relative to population below school age in SA4

Reliability indicator	HILDA wave 2005, HILDA wave 2020 and excess childcare places per child	SA4	tracks i) combined difficulty finding childcare 2005 (in deciles), ii) equivalent combined difficulty finding childcare 2020 (in deciles) and iii) excess childcare places in 2021/22. Variable takes on zero, if discrepancy exceeds four deciles (for example, if SA4 is in very undesirable decile on perceived difficulty in 2005 and 2020, but has very high excess supply), and one otherwise. Enters model as interaction with excess childcare supply.[18]
Health services	ABS Census 2006	SA4	Population: full ● Number of GPs working in SA4, per 100,000 ● Share of population who needs help with core activities
Opportunities			
Options for re-partnering	ABS Census 2006	SA4	Population: age 25 to 54 ● potential partners: population share who is male, has no university degree, is not legally/de facto married ● potential competition: population share who is female, has no university degree not legally/de facto married over potential partners
Economic activity	ABS Census 2006	SA4	Population: age 15 to 64, no university education ● Men: share employed, share unemployed and share not in the labour force (adds to 100%) ● Women: share employed full-time, share employed part-time, share unemployed, share not in the labour force (adds to 100%)
Earnings	ABS: Employee Earnings, Benefits and Trade Union Membership survey	SA4	Population: 15+ • Average weekly earnings (part-time and full-time), women, 2006 • Gender earnings gaps (part-time and full-time), 2006

[18] The purpose of this variable is to deal with the timing of the excess childcare variable. Data from 2006 for approved places was not available, and areas with "good" or "bad" (relative to other areas) childcare supply in 2021/22 might have been in a very different position (again, relative to other areas) in 2006. If that is the case, the variable does not contain meaningful information to explain variation in local reform effects. However, the much later measure could contain useful information for the most "stable" local areas, even if it does not for areas where childcare supply has improved or worsened a lot since the reform. We thus include excess childcare supply in two variables: i) the "full" variable and ii) interacted with the reliability indicator that is 1 only for "stable" areas (where perceived difficulty in 2005, perceived difficulty in 2020 and objective childcare supply in 2020 are similar, relative to other areas) and 0 for areas that have undergone more change or where we see a disconnect between perceived difficulty and objective excess supply.

Education	ABS Census 2006	SA4	Population: age 15+ ● population share with highest schooling Year 12 ● population share with postschool qualification
Occupation level	ABS Census 2006	SA4	Population: age 15+, employed ● population share who are managers, professionals or technicians ● population share who are community service workers or administrative workers ● population share who are sales workers, machine operators or labourers
Sector of firm	ABS Census 2006	SA4	Population: age 15+, employed ● population share who work for the State or Federal Government (public) ● population share who work for the Local Government Authority (lga) ● population share who work for the private sector
Rental market			
Prevalence of renting	ABS Census 2006	SA4	Population: age 15+ population share that rents for pay
Cost of renting	ABS Census 2006	SA4	Population: age 15+, rents for pay (for income: age 15+) ● Average rent per week ● Average rent per week relative to average income (SA4-level) per week ● Average rent per week relative to poverty line (SA4-level) per week
Income/poverty	ABS Census 2006	SA4	Population: households ● Equivalised household income (ABS definition), average over all households in SA4 ● Share of households in SA4, with equivalised household income below Australia-wide poverty line
Income support receipt	DOMINO	SA4	● population share who received any income support for at least one day in last year ● population share who received unemployment benefits for at least one day in last year ● population share who received any income support for at least six months in last year ● population share who received unemployment benefits for at least six months in last year ● population share who received any income support for the full last year ● population share who received unemployment benefits for the full last year

Overall (dis)advantage	DOMINO	SA4	● Index of Relative Socio-Economic Disadvantage (IRSD) ● Index of Relative Socio-Economic Advantage and Disadvantage (IRSAD) ● Index of Education and Occupation (IEO) ● Index of Economic Resources (IER).
	HILDA	SA4	● Population share who acquired a loan for the purpose of consumption goods
Community			
Values	HILDA, wave 2005	SA4	Agreement with following statements, average response in SA4 [1 to 7] ● "Many working mothers seem to care more about being successful at work than meeting the needs of their children" ● "Mothers who don't really need the money shouldn't work" ● "As long as the care is good, it is fine for children under 3 years of age to be placed in childcare all day for 5 days a week"
Diversity			
Main Language spoken at home	ABS Census 2006	SA4	Population: full ● population share who speak mainly English at home ● for two random residents with other language mainly spoken at home: probability that language is the same; average over total non-English-speaking population in SA4
Country of birth	ABS Census 2006	SA4	Population: full ● population share who was born in Australia ● for two random residents with other country of birth: probability of having same country of birth; average over total not Australian-born population in SA4
Ancestry	ABS Census 2006	SA4	Population: full ● population share who describe own ancestry as 'Australian' ● for two random residents with other ancestry: probability of having same ancestry; average over total population with non-Australian ancestry in SA4

Community engagement	ABS Census 2006	SA4	Population: age 15+ ● population share who spent time doing unpaid voluntary work for an organisation or group in the last twelve months
Satisfaction with neighbourhood	HILDA, wave 2005	SA4	Population: Population: HILDA respondents (age 15+) Reported satisfaction [0,10] with… , average in SA4: ● …the neighbourhood in which you live ● …feeling part of your local community ● …how safe you feel
Demographic controls			
Age structure			Population: full ● Male and female combined: population share age 0 to 14 (preschool and school age), population share 15 to 24 (youth transitions), 25 to 64 (prime-age and older workers) and 65+ (retirement age) All combined add up to 100%
Geographic controls			
City , Regional and Remote			Population: full ● Binary variables for city region, regional area or remote area.

Appendix B – Additional information on validity of RDD estimates

Figure B1. McCrary Density Test – National level

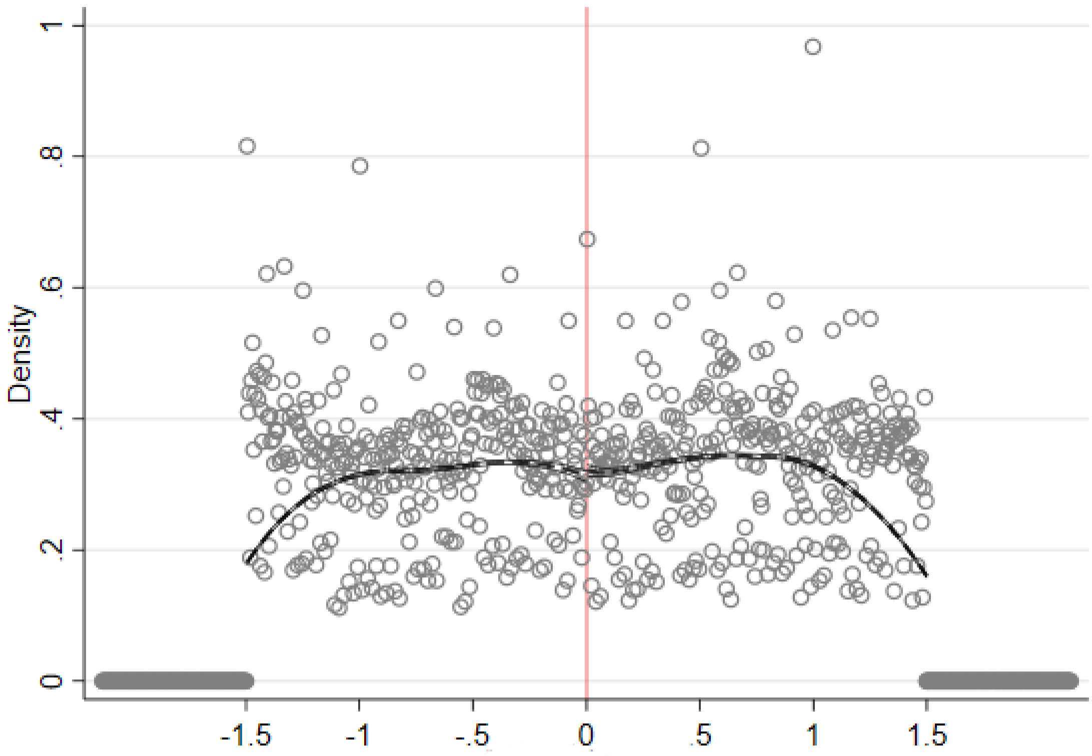


Figure B.2 – Reform effect on relocation 1 to 6 years after separation

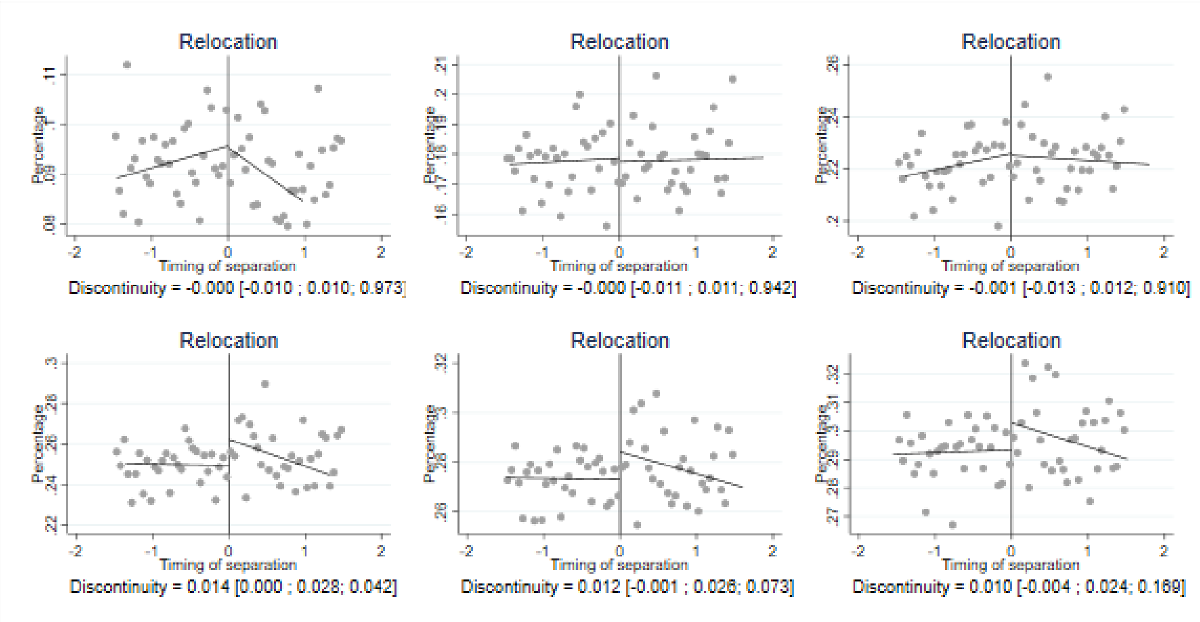
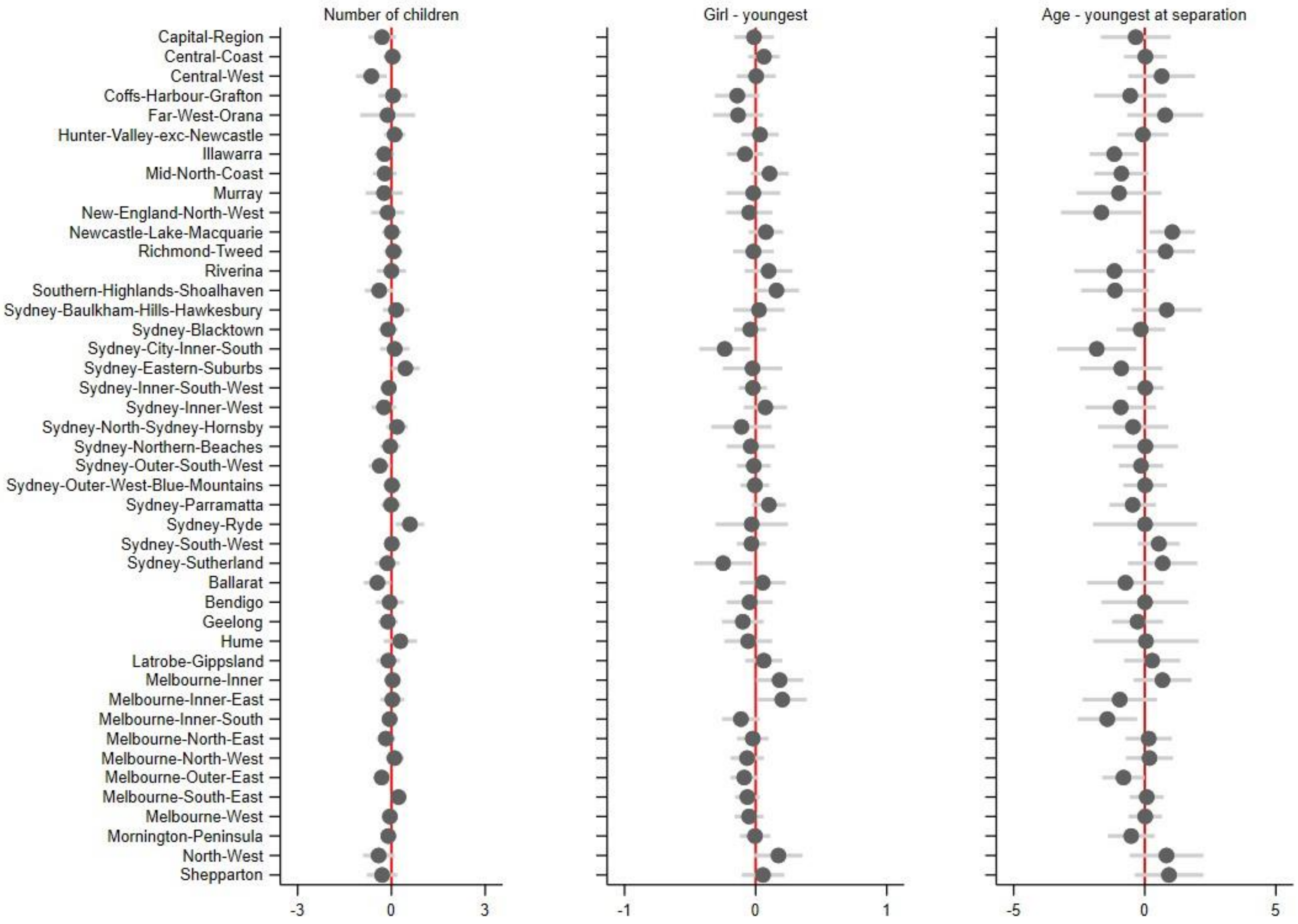
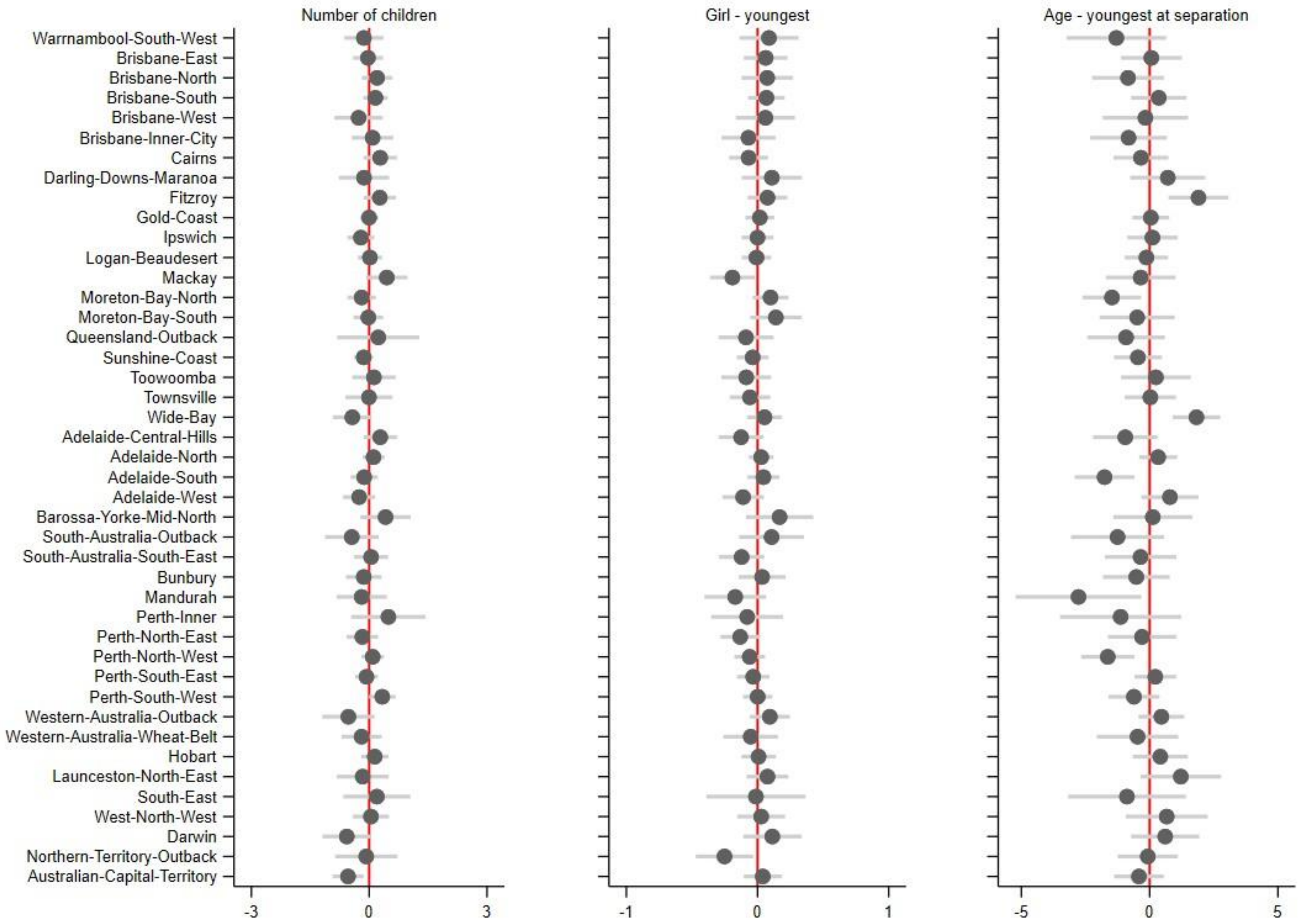


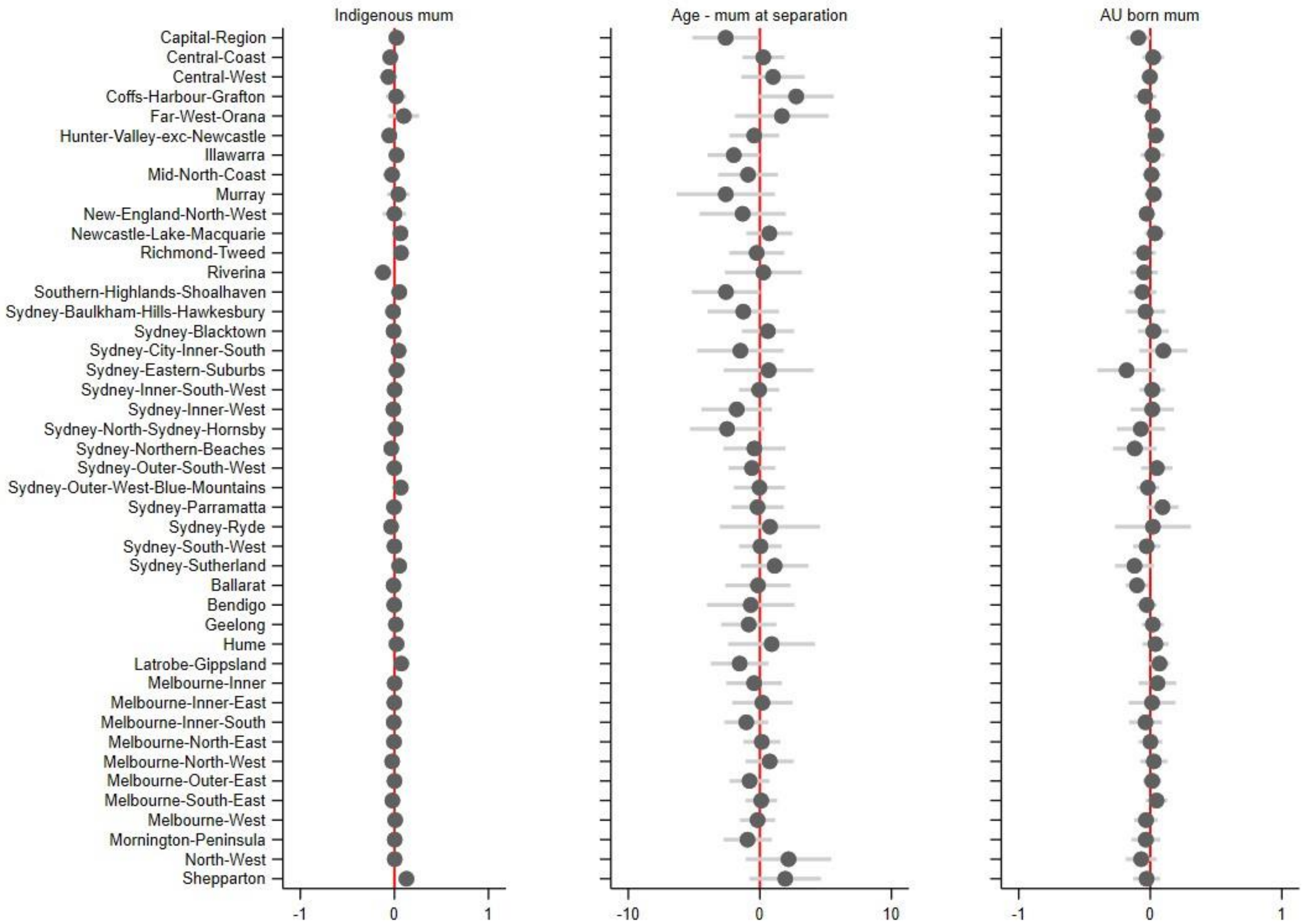
Figure B.3 Balance of observed characteristics for mothes separating before and aftre cut-off date



Note: grey bounds are 95% confidence intervals



Note: grey bounds are 95% confidence intervals



Note: grey bounds are 95% confidence intervals

Appendix C – Additional results

Table C.1 Differences in reform effects on annual earnings, for areas above instead of below the median on a given characteristic (Fourth, fifth and sixth year after separation)

	Year 4		Year 5		Year 6	
	Diff.	p-value	Diff.	p-value	Diff.	p-value
Access to and need for services						
Childcare						
Excess childcare places per child (preschool age only)	1084.840	.006	1193.202	.003		
Health services and demand						
Number of GPs working in SA4, per 100,000			1418.137	.008	1292.171	.000
Opportunities						
Education level						
share of population with a post-secondary qualification	945.662	.032	1101.084	.020	1484.787	.026
Occupation level						
share of population with a low-skilled occupation	-969.904	.026	-753.090	.036		
share of population with a high-skilled occupation	887.669	.038	749.723	.052	759.326	.099
Sector						
share of population who works for local government	-92.150	.091				
Overall (dis)advantage						
Index of Relative Socio-Economic Disadvantage (IRSD)	-573.534	.079	-638.817	.085		
Index of Socio-Economic Advantage and Disadvantage (IRSAD)	-573.534	.091				
Index of Education and Occupation (IEO)	-573.534	.060	-638.817	.055		
Index of Economic Resources (IER)	-620.454	.069	-674.680	.069		
Community						
Community engagement						
Reported satisfaction [0,10] with feeling part of your local community					838.814	.019

Source: DOMINO and other data sources described in Section 4. Authors own calculations

Notes: The table shows how the reform impact on annual employment earnings in the fourth, fifth and sixth year after separation, differs for two types of areas: those where a local area characteristic has a value above the median, compared to those where the characteristic is below the median. The analysis is univariate and does not hold other local area characteristics constant. The table includes results for every characteristic for which we found a statistically significant correlation with the estimated local area reform impact. For information on how local area reform impacts were estimated, see section 3, and for information on the quality of the estimates see section 5. For information on underlying data sources describing local area characteristics, see section 4.

Table C2. Differences in reform effects on welfare receipt, for areas above instead of below the median on a given characteristic (Fourth, fifth and sixth year after separation)

	Year 4		Year 5		Year 6	
	Diff.	p-value	Diff.	p-value	Diff.	p-value
Access to and need for services						
Means of transportation/ Commuting modes						
population share that takes, train, bus, ferry or tram to work	-.178	.097				
Health services and demand						
Number of GPs working in SA4, per 100,000	.096	.064	.093	.067		
Opportunities						
Earnings						
average weekly earnings, women	-.482	.034	-.465	.064	-.473	.098
Education level						
share of population with a post-secondary qualification			-.380	.073	-.456	.049
Occupation level						
share of population with a low-skilled occupation			.369	.056		
Sector						
share of population who works for local government			.158	.069	.108	.062
Rental market						
average weekly rent	-.286	.069	-.398	.037		
average weekly rent, relative to average local income					-.461	.091
Community						
Diversity						
for two random residents with country of birth other than Australia: probability that country of birth is the same			.454	.009	.438	.008

Source: DOMINO and other data sources described in Section 4. Authors own calculations

Notes: The table shows how the reform impact on months of welfare receipt (any benefit type) in the first, second, and third year after separation, differs for two types of areas: those where a local area characteristic has a value above the median, compared to those where the characteristic is below the median. The analysis is univariate and does not hold other local area characteristics constant. Also see notes to table C1.

Table C3. Heterogeneity of RDD-estimated reform effects on welfare receipt, along the distribution of predictive index of reform success

Variables included in Index
Access to and need for services
Health services and demand
Number of GPs working in SA4, per 100,000
Opportunities
Options for re-partnering
potential competition: population share who is female, has no university degree not legally/de facto married over potential partners
Economic Activity
share of female population who is full-time employed
Earnings
average weekly earnings, women
Overall (dis)advantage
Index of Education and Occupation (IEO)
Community
Values
Average agreement [1-7]: Many working mothers seem to care more about being successful at work than meeting the needs of their children

	Average RDD-estimated reform effect by index prediction					
	Year 1		Year 2		Year 3	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Top quartile	-1.516	.863	-1.495	1.126	-1.606	1.192
Third quartile	-1.410	.753	-1.156	.714	-1.098	.763
Second quartile	-.782	.736	-.567	.899	-.456	.877
Bottom quartile	.077	.968	.092	1.123	-.243	1.336

	Relationship between predicted and RDD-estimated local reform effects		
	Year 1	Year 2	Year 3
Average absolute deviation between predicted rank and observed rank	14.115	15.149	15.954
Proportion of SA4s with |predicted rank - observed rank| <=5	.312	.429	.416
R2 of regression of observed reform effect on index prediction	.406	.305	.222

Source: DOMINO and other data sources described in Section 4. Authors own calculations

Notes: The table shows how the reform impact on welfare receipt in the first, second, and third year after separation, differs for areas with high and low levels of the index predicting reform success, generated by the Machine -Learning process described in Section 3.2. Also see notes to Table 6.

Appendix D - Sensitivity analysis

We perform three sensitivity checks. First, we check if our results change when we keep the ten local areas where the density test pointed to a discontinuity at the cutoff. Is the distribution of local reform effects and its connection to area characteristics sensitive to the inclusion and exclusion of areas that may be affected by selective sorting or manipulation of the timing of separation? Second, we check how the results are affected by the weights assigned to different observations in the RDD procedure. How sensitive is our conclusion to this aspect of the estimation procedure? We test this by repeating the analysis with a local linear regression (OLS) estimator, that assigns equal weight to all individuals in our window of observation.

And third, we dig deeper into the issue of mothers potentially self-selecting into an area of residence with better employment prospects. As discussed before, such self-selection a) could bias the RDD-estimate of reform effects if self-selection is more common among mothers who are subject to the reform, and b) would change the appropriate interpretation of the connection between "place" and the reform effect. We turn to a specific sub-population of single mothers who are very immobile: public housing residents. A detailed description of the Australian system of public housing is included in Appendix E. For the purpose of this analysis, the key points to note are that i) there is a severe shortage of public housing, ii) those who do receive it have a strong financial incentive to keep their allocated dwelling and iii) the system allows for – practically – little choice in the location of the allocated dwelling. As a result, the place of residence for public housing residents can be plausibly treated as exogenous.

However, it is necessary to be cautious when interpretating this sensitivity check: the population of public housing residents is relatively disadvantaged, compared to our main sample of mothers. For example, they are less attached to the labour market, experience a higher incidence of disability, and are more likely to be in receipt of welfare for a longer period of time. While estimates for this subgroup are plausibly unaffected by selective sorting into location, they also might have a truly different behavioural response to the reform than the general population. Any differences in the distribution of effect sizes when making this change, thus cannot be interpreted as merely the result of removing any potential bias from selective sorting. Instead, we only use this test to confirm that the general conclusions – that heterogeneity is large and tied to the degree of disadvantage in an area – also shows up in a population that is very immobile.

Lastly, for the second part of the analysis only, we also test if the results are sensitive to the specific ML-algorithm used and present the multivariate heterogeneity analysis based on a Gradient Boosting Regression (GBR)-generated index rather than the LASSO-generated index.

Sensitivity of RDD-estimated reform impacts

Table D1 shows the mean and standard deviation, the 25th-percentile and 75th percentile, as well as the skew of the distribution of RDD estimated reform impacts. The first panel refers to the base results presented in this paper as a benchmark. In year 1, the average overall local reform effect is a $823.66 reduction in annual earnings (unsurprisingly, this is very close to the effect resulting from a single, nationwide estimation of a reduction in annual earnings by $800). Behind this average is, however, a broad range of heterogenous effects ranging from much more negative to positive, as can be seen from the 25th and 75th percentile. The distribution is asymmetric and skews negative. The estimated effects are similar in Year 1, Year 2 and Year 3 and then begin to decrease on average and increase in heterogeneity.

This pattern changes very little if we include the ten local areas back into the analysis, that had been removed because they failed the McCrary density test. The effects are slightly less negative on average and have higher positive results. This is consistent with mothers having selectively sorted into the area because of better employment prospects, leading to an overly positive, biased estimate. While the results are overall very similar whether the ten areas are included or removed from the analysis, the nature of the small difference reassures us that the conservative approach of removing these areas was the appropriate choice.

Using an OLS estimate instead of the RDD approach, turns the estimated effects slightly more negative, makes positive estimates rarer, reduces the standard deviation of the distribution of estimated effects, and makes it a little more symmetrical. At the same time, the estimates the estimates over time are more similar to each other than is the case for the RDD estimates. A more symmetrical distribution that changes less over time, is in line with the OLS estimates being subject to more random noise than the RDD estimates are. Again, while the results are broadly similar, the small differences reassure us in our choice of preferred estimator.

Finally, we turn to the results based on the public housing sample. Here, the mean effect is smaller, but the standard deviation of effects is much larger. This last test confirms that our first main result – that behind a nationwide average, there is large variation on the local level ranging from very positive to very negative effects – still holds in a population for which the phenomenon of selective sorting is less likely to occur.

Table D1. Selected Moments of the distribution of RDD-estimated reform effects on employment earnings – base model versus alterations to estimation process

BASE RESULTS						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Mean	-823.66	-1112.46	-1011.21	-810.14	-599.63	-414.60
Std. Dev.	1847.79	2247.27	2352.21	2671.47	2624.55	3582.89
25th-pctl	-1690.66	-2376.75	-2328.15	-1904.72	-1875.05	-1924.23
75th-pctl	185.65	421.79	420.83	372.80	634.04	1553.91
Skewness	-0.24	-0.68	-0.22	-0.73	-0.58	-0.46
FULL SAMPLE (including SA4s that failed the density test)						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Mean	-693.29	-1029.58	-901.51	-662.46	-495.79	-402.87
Std. Dev.	1865.08	2250.48	2367.61	2663.84	2578.07	3556.04
25th-pctl	-1680.43	-2376.75	-2277.14	-1904.72	-1784.39	-2020.91
75th-pctl	315.56	594.78	474.99	641.02	722.64	1489.71
Skewness	-0.18	-0.63	-0.21	-0.65	-0.58	-0.26
ORDINARY LEAST SQUARES (OLS) (instead of RDD estimates)						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Mean	-1056.08	-1169.80	-1499.52	-1091.25	-886.03	-1227.99
Std. Dev.	1572.71	1611.47	1750.42	1980.01	1955.93	1719.11
25th-pctl	-1822.11	-2063.22	-2606.61	-2067.47	-1678.00	-2107.49
75th-pctl	-36.59	-151.58	-362.79	-3.59	99.80	-255.66
Skewness	-0.44	-0.31	0.11	0.14	-0.02	-0.10
PUBLIC HOUSING SAMPLE						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Mean	-424.57	-1040.58	-2140.86	-377.09	-639.56	-402.87
Std. Dev.	4065.46	4833.86	8225.10	5047.94	5426.85	3556.04
25th-pctl	-2920.08	-3446.23	-4080.99	-3307.81	-4619.17	-2020.91
75th-pctl	2396.12	1142.48	931.66	1366.25	2520.81	1489.71
Skewness	-0.24	-0.89	-4.48	1.28	0.69	-0.26

Source: DOMINO and other data sources described in Section 4. Authors own calculations

Notes: The table shows how estimates of geographic variation in reform impacts on annual earnings from employment, vary with changes in estimation methods. Every panel shows moments of the distribution of estimated effects and how these moments change with a change in estimation method. For information on how local area reform impacts were estimated, see section 4, and for information on the quality of the estimates see section 5. For information on underlying data sources describing local area characteristics, see section 4.

Next, we check if the second important result – that heterogeneity is not only large but also meaningfully tied to local area characteristics – holds when we vary certain aspects of the estimation procedure. The top panels of Table D2 and D3 (where d3 refers to year 4, 5 and 6) repeats base results shown in Table 6, for ease of comparison. The next two panels show the results when i) the ML-index is created on the full sample of local areas, including the ones that failed the McCrary density test and ii) when the ML-index is created based on estimated reform effects generated by an OLS estimation rather than an RDD procedure. As before, we find that differences are small. Using only public housing residents, we find again that the distribution of estimated effects has a wider spread, a smaller (more negative) mean; however, the general conclusions still hold and are, if anything, more pronounced than in our base results. Lastly, we perform the heterogeneity analysis using a non-linear ML-procedure, based on a Gradient Boosting Regression (GBR) (Friedman, 2001). The GBR iteratively estimates an ensemble of trees and a benefit of the GBR over the LASSO is that it can identify feature importance based on non-linearities. In Table D2 and D3, we present the results from the index that is based on the predicted values from the GBR algorithm estimated on the full sample. As with the LASSO model, we have trained the model on a separate train set and evaluated it on a test set, using LOO cross-validation, and then estimated the final model on the full sample. Taking into account the non-linearities in the features appears to make very little difference to the results, although it is likely that given the relatively limited samples, our GBR model has been pruned substantially to avoid overfitting.

Table D2 Heterogeneity of RDD-estimated reform effects along the distribution of predictive index of reform success – base model versus alterations to estimation process (first, second and third year after separation)

BASE RESULTS						
	Average RDD-estimated reform effect by index prediction					
	Year 1		Year 2		Year 3	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-3061.40	1350.50	-3670.37	2037.07	-2987.46	2171.83
Second quartile	-1027.39	600.34	-1155.20	1329.73	-1155.67	2090.06
Third quartile	-632.67	660.98	-659.90	1356.15	-658.51	1720.09
Top quartile	1433.234	1024.47	1009.10	1294.79	731.40	1955.25
	Relationship between predicted and RDD-estimated local reform effects					
	Year 1		Year 2		Year 3	
Average absolute deviation between predicted rank and observed rank	7.61		11.17		14.28	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.62		0.47		0.35	
R2 of regression of RDD-estimated reform effect on index prediction	0.87		0.57		0.32	
FULL SAMPLE (including SA4s that failed the density test)						
	Average RDD-estimated reform effect by index prediction					
	Year 1		Year 2		Year 3	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-2628.00	1544.08	-3253.37	2139.61	-2675.31	2340.11
Second quartile	-1121.11	891.23	-1040.36	1472.13	-1035.19	1868.65
Third quartile	-70.90	1121.21	-238.92	2027.82	-301.44	1968.62
Top quartile	1121.36	1464.07	462.55	1418.52	444.03	2238.13
	Relationship between predicted and RDD-estimated local reform effects					
	Year 1		Year 2		Year 3	
Average absolute deviation between predicted rank and observed rank	12.62		17.10		19.84	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.44		0.31		0.21	
R2 of regression of RDD-estimated reform effect on index prediction	0.69		0.46		0.29	
ORDINARY LEAST SQUARES (OLS) (instead of RDD estimates)						
	Average RDD-estimated reform effect by index prediction					
	Year 1		Year 2		Year 3	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-2880.12	1323.72	-2838.77	1303.98	-3001.82	1291.00
Second quartile	-1484.58	598.72	-1072.67	914.11	-1201.37	1576.92
Third quartile	-573.37	655.31	-865.86	1298.40	-1266.31	1466.62
Top quartile	682.59	860.85	29.17	1442.99	-588.39	1775.04
	Relationship between predicted and RDD-estimated local reform effects					
	Year 1		Year 2		Year 3	
Average absolute deviation between predicted rank and observed rank	7.89		11.75		13.82	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.60		0.45		0.42	
R2 of regression of RDD-estimated reform effect on index prediction	0.82		0.48		0.31	

PUBLIC HOUSING SAMPLE						
	Average RDD-estimated reform effect by index prediction					
	Year 1		Year 2		Year 3	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-5013.69	2914.44	-3832.51	3188.80	-3722.68	3196.87
Second quartile	-1416.58	1732.63	-2079.53	6902.32	-5304.13	14661.74
Third quartile	727.84	2465.96	-595.58	2630.04	-1814.75	3139.46
Top quartile	4278.49	1875.73	2563.32	3380.13	2369.72	4570.86
	Relationship between predicted and RDD-estimated local reform effects					
	Year 1		Year 2		Year 3	
Average absolute deviation between predicted rank and observed rank	6.37		12.28		14.55	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.65		0.47		0.36	
R2 of regression of RDD-estimated reform effect on index prediction	0.85		0.23		0.08	
GBR						
	Average RDD-estimated reform effect by index prediction					
	Year 1		Year 2		Year 3	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-2733.83	1632.68	-3050.22	2519.30	-2418.84	2593.70
Second quartile	-1097.54	731.54	-1419.87	1252.91	-689.62	2509.53
Third quartile	-489.61	916.15	-605.12	1655.80	-970.62	1945.70
Top quartile	1547.82	1195.62	1028.58	1322.59	385.32	1495.69
	Relationship between predicted and RDD-estimated local reform effects					
	Year 1		Year 2		Year 3	
Average absolute deviation between predicted rank and observed rank	8.85		12.55		16.64	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.57		0.39		0.30	
R2 of regression of RDD-estimated reform effect on index prediction	0.75		0.45		0.18	

Source: DOMINO and other data sources described in Section 4. Authors own calculations

Notes: The table shows the link between reform impact on annual earnings from employment, and levels of the index predicting reform success as generated by the Machine -Learning process described in Section 3.2. Every panel shows how the estimated link changes with a change in estimation method. For information on how local area reform impacts were estimated, see section 4, and for information on the quality of the estimates see section 5. For information on underlying data sources describing local area characteristics, see section 4.

Table D3. Sensitivity Analysis of the Heterogeneity of RDD-estimated reform effects along the distribution of predictive index of reform success – base model versus alterations to estimation process (fourth, fifth and sixth year after separation)

1. BASE RESULTS

	Average RDD-estimated reform effect by index prediction					
	Year 4		Year 5		Year 6	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-2393.36	2785.90	-2152.01	3255.81	-2578.63	4051.04
Second quartile	-1369.89	2459.60	-999.29	1799.56	-915.39	2310.38
Third quartile	-574.67	2267.24	-242.54	2144.19	-20.79	2424.31
Top quartile	1079.37	2090.83	925.7	2290.72	1956.84	3881.36
	Relationship between predicted and RDD-estimated local reform effects					
	Year 4		Year 5		Year 6	
Average absolute deviation between predicted rank and observed rank	15.20		16.00		18.78	
Proportion of SA4s with |predicted rank - observed rank| <=5	.35		.33		.21	
R2 of regression of RDD-estimated reform effect on index prediction	.25		.19		.15	

2. FULL SAMPLE (

	Average RDD-estimated reform effect by index prediction					
	Year 4		Year 5		Year 6	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-2327.04	2590.60	-2069.40	3010.90	-2080.48	3874.26
Second quartile	-1100.26	2251.69	-636.12	1828.18	-1100.51	2253.18
Third quartile	-106.91	2278.95	-146.90	1951.96	223.85	2912.69
Top quartile	927.93	2587.73	869.54	2609.22	1487.13	4081.20
	Relationship between predicted and RDD-estimated local reform effects					
	Year 4		Year 5		Year 6	
Average absolute deviation between predicted rank and observed rank	20.28		20.53		25.56	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.19		0.25		0.16	
R2 of regression of RDD-estimated reform effect on index prediction	0.25		0.18		0.12	

3. ORDINARY LEAST SQUARES

	Average OLS-estimated reform effect by index prediction					
	Year 4		Year 5		Year 6	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-2448.55	1773.63	-1941.56	1341.15	-1800.49	1000.21
Second quartile	-793.45	2052.06	-487.88	2056.16	-1053.71	1795.82
Third quartile	-774.07	1428.92	-738.17	1478.38	-1369.10	1485.46
Top quartile	-405.73	2109.33	-458.63	2499.24	-732.37	2312.87
	Relationship between predicted and OLS-estimated local reform effects					
	Year 4		Year 5		Year 6	
Average absolute deviation between predicted rank and observed rank	15.40		17.17		18.28	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.34		0.31		0.34	
R2 of regression of RDD-estimated reform effect on index prediction	0.20		0.09		0.05	

4. PUBLIC HOUSING SAMPLE

	Average RDD-estimated reform effect by index prediction			
	Year 4		Year 5	
Index prediction:	Diff.	SE	Diff.	SE
Bottom quartile	-2232.93	3680.25	-1833.23	5355.97
Second quartile	84.49	7137.69	-748.40	6882.57
Third quartile	-1105.22	3094.59	-1526.41	3740.22
Top quartile	1937.06	4949.59	1629.14	5219.80
	Relationship between predicted and RDD-estimated local reform effects			
	Year 4		Year 5	
Average absolute deviation between predicted rank and observed rank	16.55		17.84	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.34		0.27	
R2 of regression of RDD-estimated reform effect on index prediction	0.10		0.08	

5. Gradient Boosting Regression (GBR)

	Average RDD-estimated reform effect by index prediction					
	Year 4		Year 5		Year 6	
Index prediction:	Diff.	SE	Diff.	SE	Diff	SE
Bottom quartile	-2060.01	3082.80	-1769.63	3373.00	-2280.55	4016.69
Second quartile	-568.31	2294.61	-370.70	2346.89	-556.71	2703.34
Third quartile	-982.25	2732.29	-625.99	2289.05	38.55	2677.56
Top quartile	807.27	1597.81	637.91	1800.27	1676.25	4167.71
	Relationship between predicted and RDD-estimated local reform effects					
	Year 4		Year 5		Year 6	
Average absolute deviation between predicted rank and observed rank	16.97		18.62		18.74	
Proportion of SA4s with |predicted rank - observed rank| <=5	0.26		0.26		0.29	
R2 of regression of RDD-estimated reform effect on index prediction	0.12		0.10		0.13	

Source: DOMINO and other data sources described in Section 4. Authors own calculations

Notes: The table shows the link between reform impact on annual earnings from employment, and levels of the index predicting reform success as generated by the Machine -Learning process described in Section 3.2. Every panel shows how the estimated link changes with a change in estimation method. Also see notes to Table D2, which shows analogous results for Year 1, Year 2 and Year 3.

Appendix E – Policy Background: Public Housing in Australia

In Australia, low-income households can receive housing support by renting state government-owned residential properties below market rent at a price that is determined by the tenant's income. This is called public housing. At the time of the Welfare to Work reform, on 30 June 2006, Australia's state governments owned a total of 337k tenantable public housing dwellings. 71% of these dwellings were located in major cities and 17% in inner regional areas (SCRGSP, 2007, Table 16A.1).

While public housing is the responsibility of the states, the general principles of public housing policy are fairly uniform across all of Australia. Arguably the most important characteristics of public housing for the purpose of this study, is the almost complete absence of tenant choice when it comes to their location of residence (Productivity Commission, 2017; p.171). This is in marked contrast to, for example, policies such as housing vouchers like they are used in the U.S. The lack of choice results from i) the extreme scarcity of public housing, ii) the relative attractiveness of public housing for those who do get an offer, and iii) the allocation process that leads to offers being effectively made on a take-it-or-leave-it basis (Productivity Commission, 2017; p.180).

In principle, anyone with a household income below a certain income threshold is eligible for public housing. However, the number of state-owned properties is well below demand. In the year leading up to 30 June 2006, about 28k new households were allocated a dwelling, while 187k eligible applicants remained unhoused (SCRGSP, 2007, Table 16A.1). Waiting times were thus long: of those who had been allocated housing in that year, 30% had waited 2 years or more (Australian Institute of Health and Welfare, 2008, p.ix).

When an application is lodged, general eligibility based on income is determined first, and applicants who are deemed eligible join a waitlist. At this point, the minimum number of bedrooms the applicant needs is determined (based on current household composition), as well other requirements such as necessary accommodations for a disability. The applicant can also nominate several preferred locations. Once the applicant reaches the top of the waitlist and a property with the required specifications becomes available, the government determines if the applicant still meets all eligibility criteria, and if so, an offer is made that the applicant may refuse or accept. While the minimum number of bedrooms and disability accommodations must be met, location preferences are only a guideline for offers. Typically, the applicant can refuse only two offers (and in some states only one offer), before they are forced to rejoin the waitlist at the back of the queue. This allocation process means that a public housing offer is, effectively, made on a take-it-or-leave it basis (Productivity Commission, 2017, p. 180).

If the tenant accepts the offer, they enter a very long-term lease agreement. At the time of writing this paper, state governments have moved towards offering 5-year or 10-year lease agreements; however, at the time of the reform analysed in this paper (2006), the vast majority of head tenants would have entered a lifetime lease (Productivity Commission, 2017; p. 181). By contrast, in the private rental market 1-year leases or ongoing leases with very little tenant protection are the dominant form of contract. Public housing thus offers the tenant an extraordinary degree of stability. Financially, public housing is also extremely attractive. Once per year, the state government determines the market rent for the dwelling. Twice per year, the tenants' household income is determined. The tenant then receives a rebate on the market rent that caps their out-of-pocket rent payment at a percentage of their household income or at the market rent, whichever is lower (for example, the maximum rent in Victoria is 25% of household income; Victorian State Department of Families, Fairness and Housing (2023), p.11). Even if a tenant no longer meets the income test that originally made them eligible for public housing, they can thus remain in the property until their lease expires (which may mean for life) unless they choose to exit. However, the financial benefit of remaining in public housing decreases, the more the tenant's situation improves. In 2006-07, the year of the reform studied in this paper, 87% of households in public housing received a rebated rent (SCRGSP, 2007, Table 16A.1).

A side effect of these two very large benefits (low rent payments tied to the family's ability to pay, and high stability) is that public housing tenants have very strong incentives not to move. If a tenant wishes to move – be it to change locations, or because a change in household composition makes the property unsuitable – they must apply for a public housing transfer. A tenant-initiated transfer application is treated identically to new applications, and they join the same waitlist. While the tenant has some limited influence on their location of residence when they first apply by nominating a (non-binding) preferred location, the cost to moving locations once the tenant has been housed is thus very high, and often prohibitively high: the tenant either has to join a waitlist for potentially several years, or leave a very financially attractive and stable rental agreement with the state government, in order to re-enter the much more expensive and less reliable private market. The result is that the population of public housing tenants is very geographically immobile. While this situation poses many challenges from a social policy perspective, it allows us to construct a sample of analysis for which the reform is very unlikely to have induced selection into a location of residence based on the local area's characteristics.