Crawford School of Public Policy

TTPI

Tax and Transfer Policy Institute

Place, jobs, peers and the teenage years: exposure effects and intergenerational mobility

---

TTPI - Working Paper 10/2018
June 2018

Nathan Deutscher
Crawford School of Public Policy,
Australian National University

Abstract

I show that where a child grows up has a causal effect on their adult income, but that place matters most in the teenage years. I use variation in the age at which Australian children move to identify this pattern of place exposure effects. I explore two potential explanations. First, this pattern is partly explained by the fact that spending more years in a place in adolescence lifts the probability of entering the associated local labor market and earning any corresponding wage premium. Second, I identify long-lasting peer effects using cross-cohort variation in peer parental income among permanent postcode residents.

Keywords: Intergenerational mobility, neighborhood effects, local labor markets, peer effects

**The author has been supported by a Sir Roland Wilson scholarship, funded by the Australian Treasury and the Australian National University. The helpful suggestions and advice of Robert Breunig, Lorenzo Cappellari, Tue Gorgens, Nathan Hendren, Bhash Mazumder, Pat Sharkey and Jan Stuhler have been greatly appreciated. I would like to thank Julia Neville, Thomas Abhayaratna, Bruce Bastian, Matt Power and Julia Rymasz for their invaluable expertise, advice and support accessing the intergenerational data. This research uses data from the Australian Taxation Office. All findings, opinions and conclusions are those of the author and do not necessarily represent the views of the Australian Government or any of its agencies. This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Survey was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to the Australian Government, DSS or the Melbourne Institute. The research plan was approved by the Australian National University Human Research Ethics Committee, protocol number 2017/832. Author contact details: nathan.deutscher@anu.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 in the Crawford School of Public Policy has been established to carry out research on tax and transfer policy, law and implementation for public benefit in Australia. The research of TTPI focuses on key themes of economic prosperity, social equity and system resilience. Responding to the need to adapt Australia's tax and transfer system to meet contemporary challenges, TTPI delivers policy-relevant research and seeks to inform public knowledge and debate on tax and transfers in Australia, the region and the world. TTPI is committed to working with governments, other academic scholars and institutions, business and the community.

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

A new wave of studies highlights that where a child grows up influences their adult income, education and social outcomes (Chetty and Hendren (2018a); Chetty et al. (2016); Chyn (2018); Damm and Dustmann (2014)). Previously, correlations in the outcomes of neighboring children sat alongside more tenuous evidence for an underlying causal relationship. Yet knowing when and why place matters is just as important as knowing if place matters. How much could we gain by intervening earlier in childhood?

I find that where an Australian child grows up has a causal effect on their adult outcomes, but place matters most in the teenage years. I show this using de-identified Australian intergenerational tax data. I explore two explanations for the observed place effects — the fact that where you grow up influences where you end up (local labor market conditions) and who you grow up with (peer effects). This paper provides further evidence for causal place effects, but highlights the importance of the teenage years. It also suggests a role for mechanisms – local labor markets and peers – that may be more challenging to redress than simply the quality of local institutions and policy settings.

I begin with the empirical framework introduced in the landmark study by Chetty and Hendren (2018a). I use variation in the *age* at which children move within Australia to identify the causal effect of exposure to place. This methodology avoids the selection bias inherent in simple comparisons between those choosing to live in different places or move between them. The key identifying assumption is that selection effects do not vary with the age at which children move.

The ability to observe children moving from infancy onwards, and hence detect when place matters most, is the key advantage of the Australian data. Before age 11, the outcomes of children who move converge to those of the permanent residents of their destination at a rate of around 1.1% for each year spent in the destination. After age 11, this rate of convergence is around 4.2%. These findings have implications for the interpretation and use of the estimates from the influential work of Chetty and Hendren (2018a,b). As those authors note, their estimates by necessity relied on incomplete child histories, capturing moves from age nine onwards.[1] This paper's

[1] Chetty and Hendren (2018a) use outcomes at age 24 for the estimation of the average causal effects of place, which limits them to moves from age nine. They use outcomes at age 26 for the estimation of the causal effects of specific places, which limits them to moves from age 11. For earlier

findings suggest that extrapolating back from their estimates could overstate the effect of place.[2] In the Australian setting, a child moving at birth to a place with one percentile higher incomes for permanent residents would receive around 70% of that effect themselves. Extrapolating back the exposure effects observed from age nine onwards yields a higher (biased) estimate of around 100%. On the other hand, the relative effects of specific places identified in Chetty and Hendren (2018b) could well capture the period in which location matters most — providing more comfort that they would be robust to complete childhood histories.

The pronounced sensitivity of the teenage years complements the existing human capital literature, which has long identified the importance of early childhood (e.g. Case et al. (2005); Cunha and Heckman (2007); Currie and Almond (2011)). For example, shocks to fetal health and targeted policy interventions in early childhood have been found to have long-lasting consequences. This paper finds that the causal effects of the places considered here are most potent later in childhood — perhaps reflecting differences between these mechanisms versus those most frequently examined in prior literature. The second part of the paper examines two such mechanisms that may result in a causal effect of place and explain the importance of adolescence.

First, place can matter because where you grow up is often where you end up working. Controlling for a child's location in adulthood, interacted with their cohort, halves the estimated exposure effects. This may overstate the role of local labor markets as it controls for an endogenous outcome — place effects may operate in part by giving children the skills required in the stronger labour markets. To provide a clearer test, I examine whether the outcomes of those who move better reflect those who grew up in a place, or those who end up there — and hence subject to the associated local labor market but not necessarily the childhood environment. These exogenous controls also lower the estimated exposure effects. Depending on the specification, anywhere between 15-55% of the effect of exposure to place in the teenage years can be explained by local labor markets, with the least restrictive specification consistent with the midpoint of this range.[3]

outcomes, moves at earlier ages were captured.

[2]Chetty and Hendren are clear about the assumptions inherent in such extrapolations of average or specific causal effects of place. They also frequently take a conservative approach, for example by extrapolating over less than the full period over which exposure to place may matter.

[3]These tests for a role for local labor market conditions go beyond those in Chetty and Hendren (2018a). In their work, the described controls for a child's location in adulthood are time invariant and thus capture only persistent differences in local labor market conditions. The exogenous controls

Second, place can matter because of who you grow up with. I identify positive and statistically significant peer effects among permanent residents of postcodes using cross-cohort variation in peer parental income rank. A 10 percentile rank point increase in the mean parent income rank of your postcode-cohort peers is associated with a 0.2-0.3 point increase in your own household income rank at age 24. In contrast, in placebo tests that replace an individual's peers with those from the financial years either side, the estimated peer effects are small and statistically insignificant. Finally, the peer effects appear to be driven by interactions between same-sex peers, which makes some nonsocial explanations far less likely. Given the large psychological literature on the increasing importance of peers during the teenage years (Brown and Larson (2009)), such peer effects may play a role in explaining the observed pattern of exposure effects. While I find that simple controls for peers cannot explain the observed effects of place, this may reflect the limitations of working with indirect measures of peers.

A final contribution of this paper is in validating and extending the research design of Chetty and Hendren (2018a). I show their approach can be applied in Australia — a country with less than a tenth of the population of the United States, and less geographic variation in intergenerational mobility. I am also able to replicate many of their validation exercises, with largely comforting results. The observed exposure effects remain under a range of specifications and sample selection criteria, and when comparing same-sex siblings. The approach may thus be able to yield fresh insights from further applications in a wide variety of settings.

2 Related literature

Many studies have documented correlations in the economic outcomes of children growing up in the same neighborhood (Black and Devereux (2011); Sharkey and Faber (2014)). However, the extent to which this reflected sorting by families into neighborhoods remained an open question. Early experimental studies typically found little evidence of neighborhoods affecting economic outcomes (Katz et al. (2001); Oreopoulos (2003); Ludwig et al. (2008)). Using family fixed effects, Aaronson (1998)

included in this paper, based on those ending up in a location, are also new. That said, we might expect a stronger role for local labor market conditions in determining Australian place effects for a number of reasons, which are discussed in more detail in Section 6.1

found evidence for a causal role of neighborhoods, but small sample sizes were a constraint on this and other early quasi-experimental work.

In their seminal work, Chetty and Hendren (2018a) introduce a new and ingenious approach — using variation in the age at which children move locations — and apply it to administrative data in the United States to identify compelling causal effects of place. Crucially, these effects vary according to the time *exposed* to the location. A separate paper revisits the Moving to Opportunity experiment and also finds that location matters according to time exposed, with beneficial outcomes for those moving to a lower-poverty neighborhood before the teenage years (Chetty et al. (2016)).

A recurring theme in Chetty and Hendren (2018a) and Chetty et al. (2016) is that these exposure effects appear relatively constant — the outcomes of children who move to a better neighborhood improve *linearly* in proportion to the amount of time they spend growing up there. However, as Chetty and Hendren note, the observed linearity is based on children moving at ages nine and above, and extrapolating from this is "a strong assumption that should be evaluated in future work".[4] Depending on how their estimates are extrapolated, a child moving at birth to a new location in the United States could be expected to pick up anywhere from 56% to 92% of the difference in permanent residents' expected outcomes between their origin and destination.[5] Further, their causal estimates for specific places are based on moves made from age 11 onwards, and thus potentially miss differences, if any, in the causal effects of places in early childhood.

There is no reason to expect that each year of childhood in a better location will be equally important.[6] Indeed, the literature most closely related in methodology generally finds this is not the case. Age at migration to a country often matters most in the teenage years, with negligible or marginal gains to English language proficiency, years of schooling, earnings, height and cognitive skills from an extra year in the destination country prior to adolescence (e.g. Bleakley and Chin (2004); Van den Berg et al. (2014); Basu (2018)).[7] Exposure to place within a country can

[4]See pages 4 and 53 of Chetty and Hendren (2018a).

[5]Depending on whether their 4% annual rate of convergence is applied only to the 14 years that they observe it over, from age 9 to 23 years, or to the full 23 years.

[6]I occasionally refer to locations with higher predicted total household income ranks as having 'better' outcomes or as 'better' locations. This is a convenient shorthand and is not intended as a definitive normative statement, as many features of place, well beyond those considered here, will be relevant for individual wellbeing.

[7]This is more typically framed in terms of there being a critical (pre-teen) age, beyond which

also matter more within the teenage years for outcomes such as teen parenthood (Wodtke (2013)). Indeed, such patterns are also apparent in Chetty and Hendren (2018a) when college attendance and teen birth are the outcomes considered.[8]

This paper adds to the literature by showing that exposure to place, at least on the scale considered here, matters most in the teenage years. Note this need not contradict the large literature on the sensitivity of early childhood (e.g. Case et al. (2005); Cunha and Heckman (2007); Currie and Almond (2011)) — it may simply be the large neighborhoods examined here mask substantial variation in what matters in the early years, while exhibiting variation in what matters most during the teenage years. The shocks entailed by moving may simply differ to those examined in the literature on early childhood.

So why might place matter? An important observation in the economic literature on neighborhood effects has been that where you grow up might matter because it influences the local labor market you end up in. In Page and Solon (2003a), most of the correlation in the adult earnings of neighboring boys in the United States is explained by the large earnings premium in urban areas and the high correlation between childhood and adult urbanicity.[9] A potential role for local labor markets has also been found in studies of neighborhood effects in Norway and Australia (Raaum et al. (2006); Overman (2002)). I find that up to half the exposure effects identified in this paper can be explained by local labor market conditions.[10]

Another explanation for a causal role of place is the influence of peers. Peer relationships become more salient and complex from adolescence (Brown and Larson (2009)), and potentially explain part of the sensitivity of the teenage years. Peer effects have received a lot of attention in the education literature (Sacerdote (2011)), reflecting both their potential role in school settings and the availability of data with clearly identified cohorts. School-based studies have also gone well beyond student achievement, to broader behaviors with potentially long-lasting consequences, includ-

one's ability to catch up to those born in the country is progressively more limited.

[8]Chetty and Hendren (2018a) note on teen birth: "the gradient is especially steep [exposure effects are especially high] between ages 13 and 18, suggesting that a child's neighborhood environment during adolescence may play a particularly important role in determining teen birth outcomes" (page 52). This paper suggests the same could be said for broader economic outcomes.

[9]The same authors find the same is true of neighboring girls (Page and Solon (2003b)).

[10]If the influence of an individual's initial local labor market fades with time, then this may also help explain Danish findings that neighborhoods account for a large share of the variance in permanent earnings between siblings early in the working life, but a negligible share beyond age 30 (Bingley et al. (2016)).

ing the decision to drop out (e.g. Gaviria and Raphael (2001)). Explicit studies of long-run effects are rare, but one notable exception is Black et al. (2013), where cross-cohort variation is used to study the causal effects of teenage peers on long-run outcomes in Norway.[11] I similarly use cross-cohort variation to demonstrate the presence of statistically significant, large and lasting peer effects in Australia.

3 Data

I use de-identified Australian Taxation Office (ATO) intergenerational data drawn from federal income tax returns from the 1991 to 2015 financial years. Parent-child links primarily come from linking children to adults living at the same address when the child registers for a Tax File Number (TFN): a unique personal identifier issued by the federal government. The links are also informed by a short period in which children were claimed on tax returns. Children are linked to a 'primary parent', the adult considered most likely to be a parent based on shared address, shared last name and age at the child's birth.[12] Children are also linked to the primary parent's first reported spouse over the period 1991-2015.[13] More details on the construction of the data is in Appendix E.[14]

[11]Black et al. (2013) find girls' outcomes benefit from having a higher proportion of female peers, while boys' outcomes are harmed. They find little evidence that peers' mothers' matter, but peers' fathers' income seems to matter for boys. Their study differs in ways that make it difficult to directly compare the results. Most notably, they define peers as those who were in the ninth grade in the same year and school — a more direct but also much narrower definition.

[12]Intergenerational links based on shared residential addresses are behind the Canadian data used in a number of influential studies (Corak and Heisz (1999); Oreopoulos (2003); Corak and Piraino (2011)). However, the Australian data has the notable advantage of drawing on earlier registration addresses and additional information such as claims for children on tax returns. The data most closely resembles that used by Chetty et al. (2014) in its quality and coverage.

[13]Provided the youngest child in the family is no older than 24 at the time the parent reported the spouse. Questions on spouse details have always appeared on tax returns over this period, but answers have not always been required. Linking to the earliest reported spouse rather than the spouse at a fixed point in time or in childhood attempts to cater for the resulting underreporting of spouses. Nonetheless, a restriction based on the age of the youngest child is used to avoid linking to spouses that were not present during childhood.

[14]The intergenerational data was constructed by ATO staff. The author wrote code which was submitted to the ATO to be run internally on the resulting datasets to produce the results presented in this paper.

3.1 Sample definitions

I initially restrict attention to a sample of those who: 1) were born in Australia between the 1978 and 1991 financial years; and 2) remained resident in Australia through to 2015. Earlier cohorts cannot be reliably restricted to those born in Australia and later cohorts are too young to observe at age 24, the main outcome measure in this paper. The resulting families closely mirror population benchmarks for family structure, median parental age at birth and family size (Table 1). There are slightly more children in couple families and fewer in lone mother families, likely reflecting re-partnering of lone mothers that is captured in our panel but not in the population cross-section. For the 1978-1991 birth cohorts considered here, I have parent links for around 3.1 million children, or 92% of those in the data.[15]

Table 1: Family characteristics in sample and population

Birth cohort	Full sample 1978-1991	Population Various
Family structure (%)		
Couple	86	80
Lone mother	9	16
Lone father	5	4
Median parental age at birth (years)		
Mother	27	27
Father	30	30
Family size(%)		
1	12	13
2	38	40
3	30	30
4	13	12
5	4	3
6	2	1
7 or more	1	1
Mean family size	2.7	2.6
Number of children	3,376,800	3,185,400
Number of children linked to parents	3,108,000	NA
Number of families	1,834,300	1,772,300

Notes: Population estimates are based on: Family Characteristics Survey 2003, Australian Bureau of Statistics (2017b) (family structure, 1979-93 birth cohorts); Births, Australian Bureau of Statistics (2010) (median parental age at birth, 1978-91 birth cohorts); and the 1991 Census, Australian Bureau of Statistics (1991) (family size, 1978-91 birth cohorts).

[15] This is on par with the 91% link rate achieved by Chetty et al. (2014) for the 1980-1988 cohorts used in Chetty and Hendren (2018a) — see Online Appendix Table 1 in Chetty et al. (2014).

I split the full sample in two based on residential history — creating permanent resident and mover subsamples. I use Statistical Area 4 (SA4), as defined by the Australian Bureau of Statistics (ABS), as my unit of geography. There are 107 SA4 covering Australia, typically with populations in the range of 100,000-500,000. The choice of SA4 is motivated by the methodology — these units strike the best balance between having a large permanent resident population (and thus more precise predicted outcomes for those who move) and having a large population of moves (and thus a larger sample).[16] Permanent residents are those whose primary parent files from only one SA4 from 1991 through to the year the child turned 35. Movers are those whose primary parents file from multiple SA4 — I focus on those moving once (with parents filing from two distinct SA4).

I limit the influence of measurement error in the presence and timing of moves in three ways. First, I consider only moves between SA4 where the primary parent filed in at least two years from both. Second, I consider only moves where the parent's first year filing from the destination immediately follows their last year filing from the origin — this ensures the child's age at move is known to within a year. Third, I restrict attention to those moving at least 15 kilometers, thus dropping short moves that just happen to cross SA4 boundaries.[17]

To investigate peer effects, I begin by restricting attention to those growing up in the same postcode. I create a set of permanent postcode residents whose primary parent files from only one postcode from 1991 through to the year the child turned 20. Large neighborhoods are no longer required by the methodology, and this is closer to the common conception of a neighborhood and the scale on which social interactions take place. It also more closely resembles the existing literature, where peer effects are commonly examined in the context of relatively small groups (e.g. workplaces, grades or classrooms). Nonetheless, I check for differences in estimated peer effects

[16]The next unit up in the main structure of the Australian Statistical Geography Standard is the state/territory level — Australia has only eight of these and only 27% of the moves in the sample cross state/territory lines. The next unit down in the structure is the Statistical Area 3 level, with populations in the range of 30,000-130,000 — too small to generate precise predicted outcomes based on permanent residents.

[17]In Chetty and Hendren (2018a) only moves of 100 miles or more are considered. A less restrictive condition is used here as Australia's population is highly concentrated in a small number of major cities — around a third of Australians live in Sydney or Melbourne. Moves within these cities are thus of policy interest and empirically important. In the unrestricted mover subsample, 44.3% of moves are within the same city while 26.3% are within Sydney or Melbourne.

by postcode size.[18]

In all analyses, individuals with strictly negative parent or child adult total pre-tax family income are dropped, as negative income is typically associated with high wealth and hence a poor indicator of actual economic wellbeing.

3.2 Variable definitions

Individual income is defined as total pre-tax income.[19] In years where an individual has filed a tax return, this is their reported total income or loss. In years where an individual has not filed a tax return, it is the sum of individual salary and wages reported by employers through annual pay-as-you-go (PAYG) payment summaries, and taxable government allowances, benefits and pensions. This income measure includes labor and capital income, and taxable government payments such as unemployment and study benefits. It is prior to any tax deductions or offsets. Income variables are measured in 2015 dollars, adjusted for inflation using the headline consumer price index published by the Australian Bureau of Statistics (2017c).

3.2.1 Parent income

Parent family income is the sum of the income of the identified parent(s) in each year, averaged from 1991 to 2001 inclusive. This window is chosen to strike a balance between potential attenuation biases arising from measuring parental income over too short a period versus too late in life when income is typically more volatile (Mazumder (2005)).

3.2.2 Child income

Child adult family income is the income of the child and their most recently reported spouse. I follow Chetty and Hendren (2018a) in measuring child income at age 24 or, more precisely, in the financial year in which they turn 24. The intent here is not to proxy for a child's lifetime income, but rather to gauge whether, how and why place might have a causal effect on later outcomes. As such, measuring outcomes at

[18] The postcode identifiers are randomized in the data, and are thus unable to be aggregated to higher level geographies to examine the effect of neighborhood size.

[19] Individual income is clearly identified in Australian tax returns as the individual is the primary unit of taxation. While family-based income tests exist for some taxes and many family benefits, income is reported on an individual basis.

this age is not as obviously problematic as it would be in more typical intergenerational mobility studies, where life-cycle bias has received significant attention (for a recent examination, see Nybom and Stuhler (2017)). Nonetheless, key results are also examined with outcomes measured at later ages for a subset of individuals.

3.2.3 Location

In constructing the de-identified file, parents were assigned to the residential postcode listed on their tax return in each year.[20] Parental postcode was mapped to higher-level geographies using the 2011 Australian Statistical Geography Standard released by the Australian Bureau of Statistics (2011). In particular, the ABS postcode correspondence was used to map postcodes to Statistical Areas; a hierarchy broadly designed to range from local labor markets (SA4) to local communities (SA2).[21] Move distances were calculated based on the longitude and latitude centroids from the ABS postal areas corresponding to the origin and destination postcodes.

3.3 Sample comparisons and summary statistics

Table 2 presents key features of the sample compared against Chetty and Hendren (2018a). The key advantages of the Australian data are the longer panel of tax returns and longer span of birth cohorts linked to parents. This allows outcomes to be observed for children moving from age two (rather than from age nine). Despite the longer panel, sample sizes are significantly smaller, reflecting Australia's population. The samples of permanent residents and 1-time movers are 1,683,800 and 313,900.

A challenge to identifying causal effects of place in Australia is that these effects are likely smaller than in the United States. Australia is a relatively centralized federation, with the federal government controlling the individual tax and transfer systems and exerting significant influence in health and education policy. Potentially reflecting this, there is less variation in outcomes across Australian regions. Panel

[20]For non-filing parents the postcode was recorded as missing. Postcodes were interpolated across periods of missing observations — if a parent lists the same postcode either side of such a gap they were assumed to have been in the one location the whole time.

[21]The ABS postcode correspondence assigns postcodes to areas, roughly in proportion to population. A postcode may sit entirely within an area or be split across multiple areas. Postcodes are assigned to the area containing the largest split. Around 80% of postcodes sit within the one SA4. While some individuals will be misallocated, this misallocation is consistent by postcode and hence simply a caveat on the definition of the geographical units under consideration.

C in Table 2 shows the distribution across regions of the expected household income rank of a child born into the 25^{th} percentile of the income distribution. This is the 'absolute mobility' measure reported by Chetty et al. (2014). There is less regional dispersion in these expected ranks in Australia — the difference between the regions at the 90^{th} versus the 10^{th} percentile is only 7.8 rank points in Australia versus 14.7 in the United States.

Finally, the geographic units used in this paper are smaller on average, and much more uniform in size (Panel D). This reflects the differing treatment of major cities. In the United States, major cities such as New York and Chicago are represented by a single commuting zone, whereas here they are represented by several SA4. In the Australian context of a smaller and more concentrated population this is an advantage, as within-city variations can be exploited. However, smaller geographic units also limit my ability to conduct some validation exercises.

In Appendix Table B.1 summary statistics are presented for the permanent resident and 1-time mover samples. Moving children tend to be from slightly higher income families and are more often from couple families; they also tend to end up with higher incomes themselves. However, these differences are small and mask substantial variation.[22] In Appendix Table B.2 summary statistics are presented for the origin and destination of the 1-time movers — there is a slight tendency to move to places with lower parent income ranks, fewer permanent residents and worse predicted outcomes for children.

4 Empirical framework

4.1 Estimation of causal place effects

I use the identification strategy introduced in Chetty and Hendren (2018a). This exploits variation in the *age* of children when their primary parent moves to identify the causal effect of *exposure* to place.

In the first step, I use the sample of permanent residents to estimate the relationship between parent and child household income ranks (p_i and y_i) in each SA4. The

[22] In the United States, the moving families have higher mean but lower median incomes, and their children end up with lower mean and median incomes. See Table 1 of Chetty and Hendren (2018a).

Table 2: Comparison of data with Chetty and Hendren (2018a)

	Deutscher (2018)	Chetty and Hendren (2018a)
Panel A: Sample size and time span		
Birth cohorts	1978-1991 (14 years)	1980-1988 (9 years)
Permanent residents	1,683,800	19,499,662
1-time movers	313,900	1,553,021
Panel B: Location and income information		
Data range	1991-2015 (25 years)	1996-2012 (17 years)
Potential range of age at move	1-39 years	9-32 years
Analysis range of age at move	2-34 years	9-30 years
Panel C: Regional distribution of expected adult rank of child born into bottom half		
Mean region	47.1	43.9
p10 region	44.0	37.3
Median region	46.6	43.3
p90 region	51.8	52.0
Panel D: Regional distribution of average annual cohort size		
Mean region	2,200	4,596
p10 region	900	251
Median region	1,700	1,512
p90 region	3,300	10,358

Notes: Panels A and B are based on this paper and Chetty and Hendren (2018a). Range of age at move assumes outcomes are measured at age 24. The distributions of child rank outcomes and children are calculated from the author's calculations and Chetty et al. (2014) (online data 5) for the 1978-82 and 1980-82 birth cohorts respectively.

following parsimonious linear specification is used:

$$y_i = \alpha_{ls} + \beta_{ls} p_i + \varepsilon_i \tag{1}$$

where l and s denote the child's SA4 location and financial year of birth cohort respectively. Rank-based measures are less sensitive to attenuation and life-cycle bias — a point made in a number of studies, but particularly clear in the lifetime income histories available in Swedish data (Nybom and Stuhler (2017)). In many countries, the linear specification is also a reasonable approximation to the relationship between parent and child household income ranks (Bratberg et al. (2017)). To illustrate the suitability of the specification, Appendix Figure A.1 plots the mean child household income rank at age 24 by parent household income decile for the 18 largest SA4 birth cohorts for 1978: both the general linearity of the relationship and relatively small deviations from the national relationship are striking. The fitted model from equation 1 then generates two predicted ranks for each child in the sample of 1-time movers — denoted $\bar{y}_{ops}$ and $\bar{y}_{dps}$ — based on two counterfactuals, namely that they were permanent residents of their origin and destination respectively.

In the second step, I use the sample of 1-time movers to estimate plausibly causal place effects. Specifically, I estimate a child's eventual household income rank as a function of their predicted origin outcome ($\bar{y}_{pos}$) and the 'shock' to this coming from their destination ($\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$), interacted with their age at move m. In the most general model, I estimate:

$$y_i = \alpha_{qosm} + \sum_{s=1978}^{1990} I(s_i = s)(\kappa_s \Delta_{odps}) + \sum_{m=2}^{34} I(m_i = m)(b_m \Delta_{odps}) + \varepsilon_i \tag{2}$$

where α_{qosm} is a set of fixed effects for the child's parent income decile q, origin o, cohort s and age at move m. The coefficients b_m capture the expected increase in rank associated with moving at age m to a destination with a 1 percentile rank higher predicted outcome than your origin for permanent residents. The coefficients κ_s allow for measurement error arising from potential mis-measurement of the child's origin — some children may have been born somewhere other than their parent's first recorded location, and this is more of a concern with the earlier cohorts.[23] Similar estimates of b_m are obtained from a more parsimonious and less computationally burdensome model:

$$\begin{aligned} y_i = &\sum_{s=1978}^{1991} I(s_i = s)(\alpha_s^1 + \alpha_s^2 \bar{y}_{pos}) + \sum_{m=2}^{34} I(m_i = m)(\zeta_m^1 + \zeta_m^2 p) \\ &+ \sum_{s=1978}^{1990} I(s_i = s)(\kappa_s \Delta_{odps}) + \sum_{m=2}^{34} I(m_i = m)(b_m \Delta_{odps}) + \varepsilon_i \end{aligned} \tag{3}$$

where the first sum captures origin effects, the second disruption effects, the third cohort controls and the fourth the coefficients of interest.

The b_m coefficients capture both the causal effect of moving at age m (β_m) and a selection effect (δ_m). The selection effect captures the idea that parents who move may have other attributes that affect child outcomes, for better or worse, such as differing levels of wealth, job or relationship security. Indeed, as shown in Table B.1 movers are more likely to come from high-income and couple families. To identify causal effects, I follow Chetty and Hendren (2018a) in making the additional strong assumption that selection effects do not vary with a child's age at move ($\delta_m = \delta$ for all m). This will be discussed in further detail later. However, with this assumption, the causal effect of place can be identified. First, the selection effect δ can be identified from b_m where m is greater than the age at which the outcome is measured. Second,

[23] Since our data begins in 1991, it is only for the most recent cohort that we have location over the child's full childhood.

the causal effect β_m of moving at age m can be identified by subtracting the selection effect δ from b_m. The causal effect of an additional year of exposure at age m can be identified as $\gamma_m := b_m - b_{m-1}$. If these exposure effects are a constant $\gamma_{\tilde{m}}$ over age groups $\tilde{m} \in M$ then we can further parametrize equation 3 as:

$$\begin{aligned} y_i = &\sum_{s=1978}^{1991} I(s_i = s)(\alpha_s^1 + \alpha_s^2 \bar{y}_{pos}) + \sum_{m=2}^{34} I(m_i = m)(\zeta_m^1 + \zeta_m^2 p_i) \\ &+ \sum_{s=1978}^{1990} I(s_i = s)(\kappa_s \Delta_{odps}) + \sum_{\tilde{m} \in M} (\delta_{\tilde{m}} + \gamma_{\tilde{m}} e_{\tilde{m}}) \Delta_{odps} + \varepsilon_i \end{aligned} \tag{4}$$

where $e_{\tilde{m}}$ is the number of years in age group $\tilde{m}$ that the child was exposed to the destination for. I will mostly use the age groups $\{2, ..., 11\}$, $\{12, ...24\}$ and $\{25, ..., 34\}$.[24]

Finally, the above models fall into the general class of two-step estimation, where some regressors in the model of interest (one of equations 2-4) are generated from an earlier model (equation 1). A resulting econometric concern is that not accounting for the uncertainty in the generated regressors can lead to inconsistent standard errors (Pagan (1984)). I address this in two ways. First, I use the estimated standard error on the Δ_{odps} terms to restrict the sample to moves with more precisely estimated regressors. Specifically, I consider moves where the standard error on Δ_{odps} is less than 2 percentile rank points. This is a more direct means of ensuring precision than the population restriction used in Chetty and Hendren (2018a), where the baseline analysis considers only commuting zones with populations of 250,000 or more in 2000. Second, Murphy and Topel (1985) provide a formula for calculating asymptotically correct standard errors in fairly general settings, and these Murphy-Topel standard errors are presented where possible throughout this paper.[25] Appendix D describes in more detail the calculation of Murphy-Topel standard errors, and shows key findings to be robust to varying the precision-based sample restriction.

[24]For the post-outcome age group, we should have $\delta_{\tilde{m}} = \delta$ identifying the selection effect. I set $\delta_{\tilde{m}} = \delta_{\tilde{m}'}$ for the pre-outcome age groups, allowing only the exposure effect $\gamma_{\tilde{m}}$ to vary by age.

[25]Bootstrapping standard errors was another option but was not computationally feasible. Murphy-Topel standard errors cannot be calculated where the model of interest has regressors from multiple earlier models. In this case naive standard errors are report — this is mentioned in the table notes where applicable. The differences between naive and Murphy-Topel standard errors in the regressions in this paper tend to be relatively small.

4.2 Estimation of causal peer effects

Where a child grows up may matter simply because it determines who a child grows up with. A child's peers may influence their accumulation of skills, behaviors and aspirations — all feeding into eventual outcomes. There are many models of peer effects, ranging from roles for an entire peer group's mean ability, to the effect of individual 'bad apples' or 'shining lights' (Sacerdote (2011)). The adults in a child's neighborhood may also matter. Again, many mechanisms could drive this. For example, adults may serve as role models or facilitate job market search. This fits into the broader class of human capital externalities, which has a long history in the mobility literature.[26] I begin with a broad conception of peer effects, that includes all these potential pathways.[27]

Compelling causal identification of peer effects is notoriously difficult (for a recent discussion, see Angrist (2014)). Common shocks to a peer group — such as particularly good teachers or local labor markets — will generate spurious correlation in outcomes and the appearance of peer effects where none may exist. A common way to address this is to focus on ex ante variation in peer characteristics, and this is the approach taken here. Indeed, Angrist and Pischke (2009) suggest this is the 'best shot' at a causal investigation of peer effects; an approach that uses 'some measure of peer quality that predates the outcome variable and is therefore unaffected by common shocks' (page 196).

I exploit variation in the mean parent income ranks of financial-year-of-birth cohorts within a postcode.[28] Mean parent income rank will almost certainly vary in a systemic way *across* postcodes and be correlated with unobservable factors influencing child outcomes. Similarly, there may well be trends in mean parental income rank correlated with trends in unobservable factors. However, idiosyncratic differences in birth timing will also generate plausibly exogenous differences between peer groups

[26]For example, Borjas (1992) shows the outcomes of a generation to relate not only to their parent's outcomes but also to the average outcomes of their parents' ethnic group.

[27]Manski (2000) describes three reasons members of the same group may end up with similar outcomes: 1) *endogenous interactions* whereby individual behavior varies with group behavior; 2) *contextual interactions* whereby individual behavior varies with group background; and 3) *correlated effects* whereby individuals behave similarly simply due to similar individual or institutional characteristics. I identify the combined effect of (1) and (2), rather than (1) alone.

[28]Financial year of birth was chosen as the time span of interest as it was readily available. However, school entry cut offs also mean it is a good conceptual choice, as most Australian children born in the same financial year of birth end up in the same grade in school. The same is not true of calendar years.

within a postcode. This approach is similar to that taken by Black et al. (2013) where cross-cohort variations in the ninth grade peer groups of Norwegians is exploited to identify peer effects. In this case, I estimate:

$$y_i = \alpha + \beta p_i + \eta c_{zs(i)} + \zeta \bar{c}_{zs(i)} + \varepsilon_i \tag{5}$$

where $c_{zs(i)}$ is the mean parent income rank in postcode z, birth cohort s, and $\bar{c}_{zs(i)}$ is the 3-, 5- or 7-year moving average of the same. The moving average controls for factors that may influence child outcomes, and be correlated with mean parent income rank within the postcode. I use leave-one-out means to exclude variation in peer group parent rank driven by an individual's own parents. In keeping with the earlier investigation of the causal effects of place, y_i is child household income rank at age 24 and p_i is their parent household income rank. Equation 5 is simply the canonical intergenerational mobility regression for the rank-rank (Spearman) correlation β, with two added terms — the peer mean parent income rank, with associated peer effect η, and the moving average of the same as a control.[29] As apparent in Table 3, there is significant variation in peer mean parent income ranks, both across postcode-year peer groups and across years within postcodes, the latter being the source of variation exploited here.

Table 3: Distribution of peer group size and mean parent income rank

	Mean	p10	p50	p90	N
Number of peers	134	22	99	288	1,339,300
Peer mean parent income rank	52	38	51	66	1,334,500
Peer mean parent income rank...					
minus 3-year moving average	-.01	-3.2	.01	3.2	1,143,600
minus 5-year moving average	-.02	-3.3	-.02	3.2	954,500
minus 7-year moving average	.01	-3.4	.02	3.4	767,000

Notes: For the sample of permanent postcode residents, shows distributional statistics for the size and mean parent income rank of the peer groups defined by shared permanent postcode and financial year of birth. The moving averages are taken within the postcode, across the 3, 5 or 7 financial year of birth cohorts centred around the individual's financial year of birth.

A potential criticism of the above is that families may be more (less) likely to leave a postcode if their child ends up in a particularly poor (rich) cohort. This selection process could result in unobserved differences between the families observed in the poor or rich peer groups. However, moving is costly and disruptive, so this

[29] I also estimated specifications that combined these two terms, adding only the deviation in peer mean parent income rank ($c_{zs(i)} - \bar{c}_{zs(i)}$) to the regression, with similar results.

may be less of an issue than endogenous school choice is in the more typical setting when peer groups are defined by school grades or classes. Nonetheless, I reestimate Equation 5 with family fixed effects to test the robustness of the results to fixed family unobservables.

5 The causal effect of exposure to place

I now turn to the results. In this section, I discuss the estimated causal effects of exposure to place in Australia, including the heightened importance of place in the teenage years. In the following section, I examine potential explanations for these patterns, including the role of local labor markets and peer effects.

I begin by examining Figure 1, which shows the estimated coefficients b_m from Equation 3. These represent the expected boost to an individual's household income rank at age 24 associated with a move at age m to a place with 1 percentile rank higher outcomes for permanent residents. Three patterns are of interest — the evidence of selection effects, positive exposure effects and the more pronounced sensitivity of the teenage years.

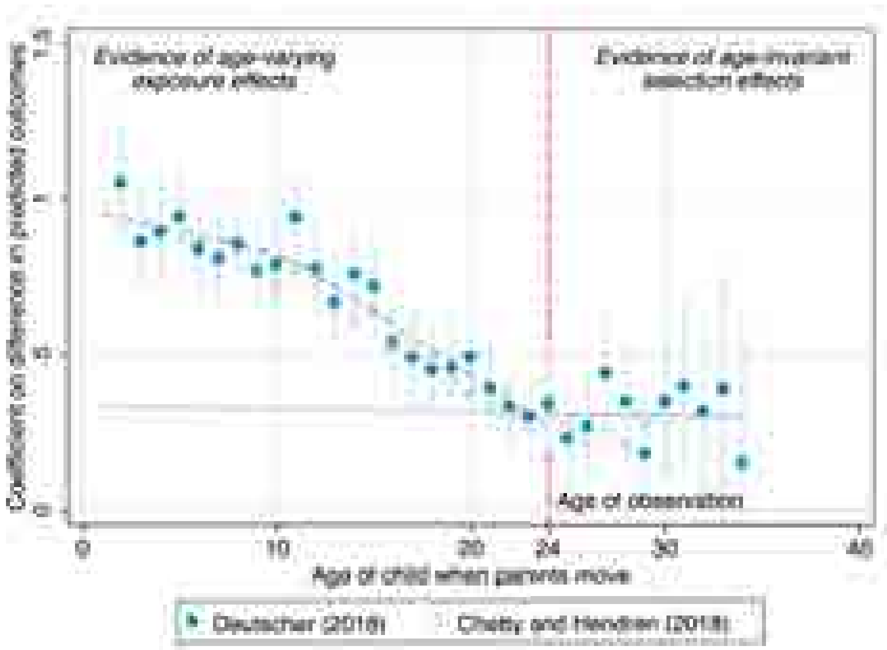
First, there is evidence of positive but age-invariant selection. Children do better at age 24 even if their parents only later move to a place with better outcomes. Comfortingly this selection effect appears invariant to the child's age at move: if I fit a line to the estimated regression coefficients b_m for $m \in \{25, ..., 34\}$ I get a slope coefficient of -0.00097 (s.e. 0.0085).[30] The mean selection effect is 0.308, slightly higher than the 0.226 found for the United States in Chetty and Hendren (2018a). This is not surprising, as if Australia's lower regional variation in mobility compared to United States reflects less of a causal role of place this is exactly what one would expect.

Second, there is evidence of positive exposure effects — the benefits of a move to a place with better outcomes are greater the earlier one moves. Third, the exposure effects ($\gamma_m = b_m - b_{m+1}$) appear to vary systematically by age, with larger exposure effects in the teenage years. All up, a child whose parent moves at their birth would

[30]Selection effects are not considered for $m = 24$ as the child's outcome is potentially affected by moves at this age. Outcomes are measured in the income year the child turns 24. Age at move is determined by how old the child turns in the income year for which the primary parent files a return from a new location. While the parent will file the return after that income year, the move may well have occurred during it and affected child outcomes.

be expected to pick up about 70% of the difference in predicted outcomes between their origin and destination.[31]

Figure 1: Place exposure effect estimates for child income rank in adulthood



Notes: Estimated coefficients b_m from equation 3, with 95% confidence intervals. The b_m capture the expected boost to an individual's household income rank at age 24 from moving at age m to a place with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their age at parent move m with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. This replicates Panel B, Figure IV from Chetty and Hendren (2018a). The point estimates from that paper are also shown.

In Appendix Table B.3, I provide exposure effect estimates and model fit statistics for several versions of the parametrized model in equation 4. In column (1), I assume constant exposure effects as in Chetty and Hendren (2018a). In columns (2)-(8) I allow a kink, with exposure effects that are allowed to vary between early and late childhood. Consistent with the visual impression left by Figure 1, the best fitting

[31] The fact that most of the differences in outcomes between places in Australia appears to be causal does not imply that place explains a large portion of the variation in outcomes between individuals. As already noted, the geographic variation in outcomes in Australia is relatively modest, and it may be that Danish findings that neighborhoods account for a large share of the variance in permanent earnings between siblings early in the working life, but a negligible share beyond age 30 (Bingley et al. (2016)) would also replicate here, though this is beyond the scope of this paper.

parametrized model allows a kink at age 11.[32] An additional year in a place with 1 percentile rank point higher outcomes is associated with a gain at age 24 of 0.011 (s.e. 0.007) rank points before age 11 and 0.042 (s.e. 0.003) rank points after age 11; the p-value on a test of equivalence of the early and late childhood exposure effects is 0.00068. This model is the baseline model for the analysis that follows.

The ability to measure early childhood exposure effects is the primary advantage of the Australian data, and the finding of smaller place effects in these years is a key contribution of this paper. It also seems unlikely this finding will be unique to Australia. As noted earlier, the age-at-migration literature has generally found additional years of exposure to the host country matter most in the teenage years for skills such as language acquisition. Further, there are hints in Chetty and Hendren (2018a) that a similar pattern may be observed in the United States once more data becomes available. For example, as the authors themselves have noted, the teenage years seem more important than earlier years in determining teen births and college attendance.[33]

The sensitivity of the teenage years is not driven by any particular subpopulation. In Appendix Table B.4, exposure effects are estimated separately for various subpopulations — across all specifications the exposure effects are higher in late childhood. There is, however, some notable heterogeneity. Late childhood exposure effects are significantly larger for boys than for girls — 0.049 (s.e. 0.005) versus 0.032 (s.e. 0.005). Moves by individuals from poorer families or to better destinations are nearer to having constant exposure effects, though the teenage years still appear more sensitive.

Finally, the sensitivity of the teenage years is not confined to place exposure effects — the fixed costs associated with moving also appear larger in these years.[34] Appendix Figure A.2 shows the expected household income rank at age 24 of a child born in 1991, with parent income rank of 50, and moving at age m between an

[32]This model is preferred to the other potential kink points on all the model selection criteria. It is also preferred to a model with constant exposure effects on all but the BIC, which places the highest penalty on the additional parameter.

[33]Also, in Figures IV and V in Chetty and Hendren (2018a) the coefficients b_m for $m \in \{9, ..., 12\}$ do not generally display as strong a linear relationship as the later years. These coefficients are also likely not as precisely estimated given they are based on only the later cohorts for which moves at this age can be observed.

[34]'Fixed' in the sense that they do not vary with the difference in predicted outcomes between the origin and destination (Δ_{odps}).

origin and destination where their predicted outcome is also an income rank of 50 (implying $\Delta_{odps} = 0$). Positive selection effects are apparent — children exceed their predicted ranks at age 24 even if their parents only move later — but comfortingly these also appear age-invariant.[35] If these selection effects can also be extrapolated back then the chart suggests a fixed cost of moving between places. These fixed costs are also larger during the teenage years: they are largest at age 17 — the age at which students enter their final year of schooling and prepare for university entrance exams. The fixed costs of moving are relatively large, and may well outweigh the benefits. If we subtract the assumed selection effects, then in the sample of moves in the data, only 15.6% of moves (23.9% of moves to places with better outcomes) carry an expected benefit to the child based on their age at move m, parent rank p and the difference in outcomes of permanent residents Δ_{odps}.[36]

5.1 Validation exercises

A caveat on these results is the strong underlying assumption that selection effects are age-invariant. There appear to be good grounds to doubt this assumption — both intuitively and based on the data. Appendix Figure A.3 shows moves in late childhood are much less common and slightly skewed towards lower income and sole parent families. While parent income rank is controlled for in the regressions, differences in other background traits may lead to biased estimates of the causal effect of place.

In Appendix C I replicate where possible and discuss in detail the validation exercises conducted in Chetty and Hendren (2018a). The results are generally comforting. Using a more general set of controls or a later age of observation does not alter the conclusions above. Adding family fixed effects results in only a very modest reduction in the estimated exposure effects, and the sensitivity of the teenage years remains apparent. Finally, an event study illustrates the best predictor of a mover's outcomes comes from looking at their cohort, rather than neighboring cohorts.

[35] This positive selection into moving is consistent with Appendix Table B.1, where movers were found to come from slightly higher income families, on average.

[36] This discussion is intended as illustrative only, as I don't present arguments in favor of a causal interpretation of these fixed costs. It is simply a useful caveat to bear in mind — the regressions here do not imply that moving to a better place is invariably associated with better expected outcomes. Further, fixed costs of moving are not without precedent. In an altogether different setting Chetty et al. (2016) find slightly negative effects for those moving in adolescence as part of the Moving to Opportunity experiment (but positive effects for those moving when young), which they suggest may reflect disruption effects.

There are two cases where the validation exercises fail. In particular, the exposure effects for moves in the top decile of relative postcode outflows are significantly attenuated. The idea behind this exercise is that these moves seem more likely to have been forced — due to mass layoffs or natural disasters — rather than chosen. This is a potential caveat on the external validity of the results, but would also be consistent with the finding of Chyn (2018) that treatment effects may fundamentally differ between those choosing to move versus those forced to move. Finally, the mover's probabilities of falling into the top or bottom deciles are better predicted by the expected ranks of the permanent residents (rather than expected probabilities of the same outcomes), but this could readily be due to much smaller sample sizes, as discussed further in the Appendix.

In sum, the tests provide good evidence in support of a causal interpretation of the estimates in the Australian setting. Any omitted variable giving rise to the exposure effects must operate within the family in proportion to the time exposed to the destination, and replicate the outcomes of an individual's particular birth cohort. However, these tests are also consistent with a number of potential explanations for a causal effect of place — including potential roles for local market conditions and peer effects as considered in the following section.

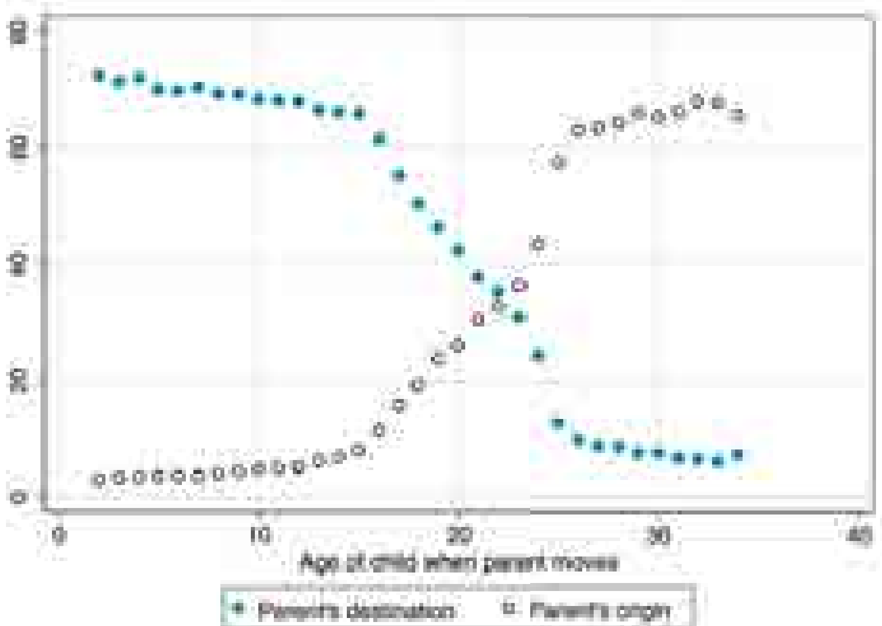
6 Why the teenage years?

6.1 The role of local labor market conditions

Where a child grows up might matter simply because it relates to where they end up working. If a child ends up in a stronger local labor market they will likely do better even without their location having had any effect on their development.

If local labor markets drive the observed effects, then why would time spent in a location matter? Why would we see an exposure effect pattern? Figure 2 provides an explanation — a child's end location looks like a function of their exposure (through their primary parent) to locations growing up. For those moving once, the earlier a child's parent moves, the more likely the child is to end up in the destination by age 24. This is particularly the case for moves after 15, and there is no age at which a parent's move creates a large, obvious discontinuity in the child's probability of ending up in the destination location. A similar pattern is also apparent if controlling for family fixed effects (Appendix Figure A.4). And for those moving twice, the more years spent in the initial destination the more likely the child is to end up there (Appendix Figure A.5). There are intuitive explanations for a child's location being a function of their exposure to locations while growing up. A longer period of exposure in childhood might generate stronger ties to the location through social ties and preference formation. It may also indicate stronger exogenous ties to the location, such as the presence of other family or community members.[37]

Figure 2: Moving once: % of children in a given location at age 24, by age at parent move



Notes: Based on the 1-time movers sample. For children of a parent who moves once, shows the % those in a given location at age 24 — the parent's origin or destination — by age at move. Those in a different location are not shown for clarity, while those without a known location are excluded. The child's location is known for all those lodging a tax return in the year they turn 24 with a valid postcode.

[37]Mis-measurement of child location is also a possibility given older children are less likely to follow their parents. As noted in Chetty and Hendren (2018a) this possibility of children not moving with their parents means the exposure effects estimated on the basis of parent moves should be interpreted as an intention-to-treat effect rather than an effect of treatment on the treated. I do not attempt to calculate the latter, as the Australian tax data does not reveal whether children moved with their parents. However, Appendix Figure A.6 shows that in a similarly constructed 1-time mover sample from the Household Income and Labour Dynamics in Australia (HILDA) survey all those aged 16 or less when their parent moved followed their parent. Even among these individuals, the child's end location is a function of their exposure to the destination. This suggests mis-measurement cannot be the primary explanation of the pattern observed in Figure 2.

These charts refute some arguments against a role for local labor market conditions. Figure 2 shows that the pattern of exposure effects observed in Figure 1, with no obvious discontinuity in place effects just prior to labor market entry, could in fact still occur even if local labor markets explained all the observed place effects. Similarly, even if exposure to an interim destination matters for outcomes (as found by Chetty and Hendren (2018a) for those whose parents move multiple times), then this too is consistent with effects entirely driven by local labor markets.

In Table 4, I present estimated exposure effects with a variety of different controls for a child's end location. These are 'bad' control variables, in that they control for an endogenous outcome — an individual's location may indeed influence their income, but income may also influence location. Nonetheless, this exercise is useful as an initial investigation. In column (2) I add controls for a child's location at age 24. This partially attenuates the estimated exposure coefficient, which falls by around 40% from 0.042 to 0.026.[38] Since we would expect some attenuation anyway, given we are controlling for an endogenous outcome, it is tempting to read this as a lower bound for the exposure effect net of local labor market effects. However, local labor markets have cycles, and will generally differ in their strength over time. A more comprehensive set of local labor market controls would have the same richness as the place effects, with differences by cohort and parent rank.

Allowing local labor market controls to vary with time is important. In column (3) I interact child location with child cohort and the exposure effect estimate is now more than halved relative to the baseline, falling to 0.020. In column (4) I interact this with parent rank as well, but with negligible further effect. Table 4 suggests local labor market conditions could play a significant role in generating exposure effects. However, the question still remains as to how much this attenuation is simply a side-effect of controlling for an endogenous outcome.

An alternative and more compelling test of the importance of local labor market conditions is a placebo test along the lines of those used in validating the research design. If local labor markets explain the effects observed then the outcomes of

[38]In Chetty and Hendren (2018a) the estimated exposure effect falls by only around 20%, from 0.040 to 0.031 (Table II). This is actually based on a specification that controls for child location in 2012, rather than at age 24, interacted with cohort. This could lead to either lower or greater attenuation relative to what would be observed if the specifications in column (2) and (3) were replicated in the US data. In practice however, the resulting coefficients are at most slightly smaller than the 0.031, and remain above those observed in the Australian data (personal communication with Nathan Hendren).

Table 4: Exposure effects with local labor market controls

	Baseline	With child location controls		
	(1)	(2)	(3)	(4)
Early exposure	0.011	0.011	0.009	0.010
	(0.007)	(0.006)	(0.006)	(0.006)
Late exposure	0.042	0.026	0.020	0.020
	(0.003)	(0.003)	(0.003)	(0.003)
Post-outcome exposure	0.008	0.015	0.009	0.011
	(0.013)	(0.012)	(0.012)	(0.012)
Controls				
child location ...		X	X	X
... X child cohort ...			X	X
... X parental rank				X
N	264,500	264,500	264,500	264,500

Notes: Estimates of the exposure effects $\gamma_{\tilde{m}}$ from equation 4 for early ($m \in \{2, ..., 11\}$), late ($m \in \{12, ..., 24\}$) or post-outcome ($m \in \{25, ..., 34\}$) exposure with various controls for child location. These represent the expected boost to an individual's household income rank at age 24 associated with an additional year at this stage of life in a destination with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their time exposed to the destination in each stage of life $e_{\tilde{m}}$ with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. Child location controls cover: none (1); child location fixed effects (2); child location-cohort fixed effects (3); and child location-cohort fixed effects and interactions with parental rank (4). Child location is simply the corresponding SA4 for those filing a tax return with a valid postcode. Murphy-Topel standard errors are in parentheses.

children who end up residing in a location should be a better predictor of movers' outcomes than the outcomes of those who grew up there. To test this, I consider the subset of the permanent resident children who file a tax return from a valid postcode. This subsample can then be used to generate two sets of predictions for the movers — one based on where the child "grew up" (their primary parent's location) and one based on where they "ended up" (their filing location).

The choice of specification is a key consideration in conducting this exercise. I begin with the more parsimonious linear kink specification in equation (4), before moving to less restrictive specifications. Earlier, I chose age 11 as the kink point separating early and late childhood. However, this was chosen to maximize the ability of those growing up in a place to explain the outcomes of those moving places. A kink at age 11 may thus tilt this test in favor of those growing up in a place over those ending up there. On the other hand, age 15 is a very clear kink point in the relationship between where you grow up and where you end up (Figure 2). Using this kink point may thus tilt the test in the other direction, maximizing the predictive content of the regressors based on those ending up in a place. I initially conduct this test using two specifications, one with a kink point at age 11 and one with a kink point at age 15, before further relaxing these constraints.

The results of this exercise suggest a clearer role for local labor markets. In Table 5 I present the estimated late childhood exposure effects for a mover's rank and probability of making the top decile. Estimates are shown using, first separately and then simultaneously, the predictions based on either where the children of permanent residents grew up ("Growers") or where they ended up ("Enders"). The separate regressions produce exposure effect estimates that are similar to those from the baseline specification. When estimated simultaneously the results are highly sensitive to the specification. With a kink at age 11, the exposure effect based on the Growers falls, but only by a modest 15% when the outcome is mean income rank. If I instead locate the kink at age 15 then the predictions based on the Enders dominate, with the exposure effect based on the Growers falling by 55% when the outcome is mean income rank. These regressions leave open the possibility that where you grow up matters mostly because it influences where you end up, and that the endogenous controls in Table 4 are not, in fact, overstating the role of local labor markets.

Finally, I further relax the specification, returning to one that is more agnostic as to how exposure effects may vary over childhood. Specifically, I modify equation 3 to also include the predicted outcomes, and differences in predicted outcomes between origin and destination, based on the 'Enders'.[39] The resulting coefficients are shown in Figure 3. While not definitive, a few useful observations can be made. At almost all ages there is predictive power in the experiences of both those growing up and those ending up in a place. The predictive content of both $\Delta_{odps}^{Growers}$ and Δ_{odps}^{Enders} tends to fall with age, particularly in the teenage years. The precise fall in predictive

[39] The precise specification is:

$$\begin{aligned} y_i = & \sum_{m=2}^{34} I(m_i = m)(\zeta_m^1 + \zeta_m^2 p) + \sum_{s=1978}^{1990} I(s_i = s)(\kappa_s \Delta_{odps}^{Growers}) \\ & + \sum_{s=1978}^{1991} I(s_i = s)(\alpha_s^1 + \alpha_s^2 \bar{y}_{pos}^{Growers} + \alpha_s^3 \bar{y}_{pos}^{Enders}) \\ & + \sum_{m=2}^{34} I(m_i = m)(b_m^{Growers} \Delta_{odps}^{Growers} + b_m^{Enders} \Delta_{odps}^{Enders}) + \varepsilon_i \end{aligned} \tag{6}$$

Table 5: Exposure effects using predicted outcomes of permanent versus eventual residents

	Mean rank			Top 10%		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Kink at age 11						
Growers	0.037		0.031	0.017		0.010
	(0.004)		(0.008)	(0.003)		(0.003)
Enders		0.029	0.005		0.017	0.012
		(0.003)	(0.007)		(0.002)	(0.003)
Panel B: Kink at age 15						
Growers	0.044		0.020	0.024		0.013
	(0.006)		(0.012)	(0.004)		(0.005)
Enders		0.039	0.024		0.023	0.018
		(0.005)	(0.010)		(0.004)	(0.004)
N	221,700	215,600	209,800	210,500	206,800	176,200

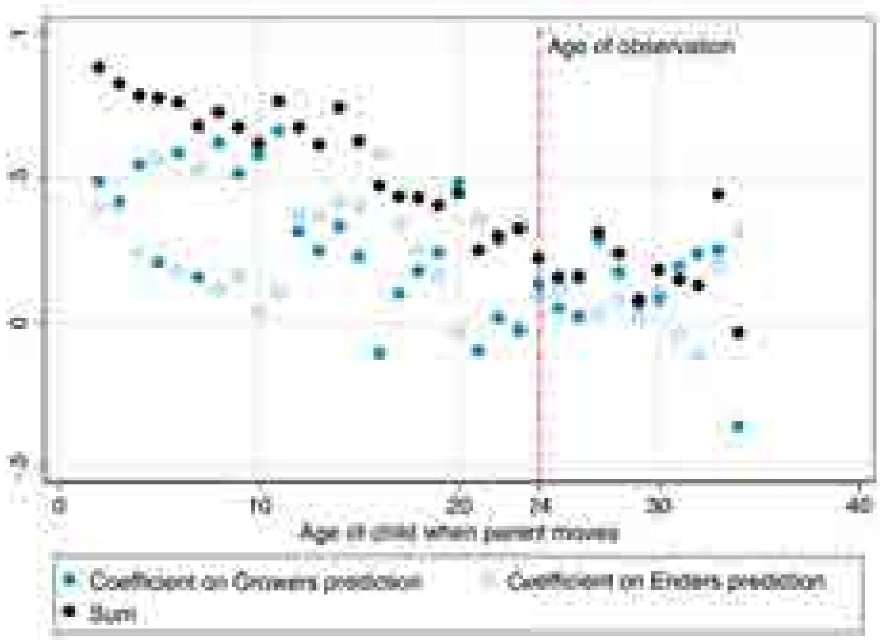
Notes: Estimates of the exposure effects $\gamma_{\tilde{m}}$ from equation 4 for late childhood for different outcomes and based on different predictions for origin and destination outcomes. The outcomes of those whose parents move once in their childhood provide the dependent variables. In columns (1)-(3) the outcome is the individual's household income rank at age 24, in columns (4)-(6) it is the event that this rank is in the top decile. As in Chetty and Hendren (2018a), I add the square of parent income rank to equation 1 before estimating the predicted probabilities of being in the top decile. When the outcome is being in the top decile, rather than mean rank, I use the precision threshold of $\Delta_{odps} < 10$ in order to restrict consideration to a similarly sized subsample of moves with the most precisely estimated differences between origin and destination. In all regressions the predicted outcomes for the origin and the destination, which provide a counterfactual for the moving individual, are based on a sample of those whose parents are permanent residents of a location, and who file a tax return in the year they turn 24 with a valid postcode (movers are also required to do this). In columns (1) and (4) the predicted outcomes for the origin and destination are based on the sample of individuals who grew up there, based on the location of their primary parent. In columns (2) and (5) the predicted outcomes for the origin and destination are based on the sample of individuals who file a tax return from there. In both these cases the specification is as in equation 4. In columns (3) and (6) both sets of generated regressors — those based on where individuals grew up and those based on where they ended up — are included in the regression. Standard errors are in parentheses (naive for columns (3) and (6) and Murphy-Topel otherwise).

content is quite sensitive to how it is measured. If we look simply at the difference in the average coefficients before age 11 and after age 24, then the coefficients on the Growers fall by around 2.7% a year, while those on the Enders fall by around 1.4% a year. This would suggest where an individual ends up explains around a third of the total causal effect of place, a story between the two extremes emerging from Table 5.

These tests for a role for local labor market conditions go beyond those provided in Chetty and Hendren (2018a). In theory, these results could reflect specific features of the Australian economy over this period, such as the mining boom resulting from Chinese demand for Australian resources.[40] The mining boom is certainly an

[40]The mining boom saw employment in the industry increase well over three-fold from 2002 to 2012 and was concentrated in two states — Western Australia and Queensland. Together these

Figure 3: Place exposure effect estimates for child income rank in adulthood, based either on permanent residents who grew up or ended up in a location



Notes: Estimated coefficients $b_m^{Growers}$ and b_m^{Enders} from equation 6, and their sum. The b_m capture the expected boost to an individual's household income rank at age 24 from moving at age m to a place with 1 percentile rank higher expected outcomes for permanent residents — either those who grew up in the locations or those who ended up there. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their age at parent move m with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o, calculated separately based on the permanent residents who grew up or ended up in the locations. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with $\Delta_{odps}^{Growers}$ to capture potential mis-measurement of the origin. Due to the high degree of correlation between $\Delta_{odps}^{Growers}$ and Δ_{odps}^{Enders} this last term is included based on the growers only. The results are not meaningfully different if this term is excluded all together.

important factor in recent experiences of intergenerational mobility in Australia: Appendix Figure A.7 shows 'absolute mobility' in mining states has followed economic conditions relatively closely. However, in Appendix Table B.5 I show this paper's key findings remain when restricting to moves within the non-mining states. Even for those moving within the non-mining states, exposure effects are larger in the teenage years and significantly attenuated when including local labor market controls. The same is true when restricting to moves within major metropolitan areas (Appendix Table B.6).

states accounted for around a third of Australian employment but two-thirds of mining employment over the period of observation (Australian Bureau of Statistics (2017d)).

The relative importance of local labor markets in Australia may simply reflect a lesser role played by other factors. As noted in Section 3.3, Australia has much less geographic variation in intergenerational mobility than the United States. This is perhaps unsurprising, as Australia is a much more centralized federation, with less geographic variation in policy settings. As a consequence, even if local labor markets are no more important in absolute terms than in the United States, they could account for a greater share of the causal effects of place observed here.

6.2 The role of peers

The results above suggest place effects and the sensitivity of the teenage years is partly — but not wholly — driven by local labor market conditions. A prime candidate for some of the residual effect is a role for peer effects. The psychological literature has long noted that the transition from early childhood to adolescence is marked by the increasing salience and complexity of peer relationships (Brown and Larson (2009)).

The Australian data shows clear evidence of peer effects. To identify peer effects I regress an individual's household income rank on their parents' rank and their peers' mean parent rank. Peers are those sharing a postcode and a financial year of birth. I include the moving average of peer mean parent ranks for that postcode as a control. Intuitively, if a child just happens to be born into a wealthier or poorer birth cohort for their postcode, do they do any better or worse? Children could be influenced either directly by their peers' parents — as potential role models, for example — or indirectly — insofar as they proxy for the abilities and behaviors of the peers themselves.

In Table 6 I display the coefficients on parent rank and peer parent rank. Across a range of specifications, an increase of 10 percentile ranks in the peer parent rank is estimated to increase a child's household income rank at age 24 by between 0.2 and 0.3 percentile rank points. The influence of peers is around a fifth that of the influence of parents at that age — a large yet plausible effect size. The addition of family fixed effects has little effect on the point estimates, though their statistical significance is lost. While the estimate is notably smaller with the shortest window width, this would be consistent with individuals having some peers drawn from the neighboring birth cohorts. In this case, having a higher mean peer parent rank would be expected to be more beneficial when it is conditional on the wider moving averages that place less weight on cohorts that may also exert some influence on outcomes.[41] Given this possibility, I use a window width of 7 years in the remaining analysis.

Table 6: Parent and peer influences on household income rank at age 24

	(1)	(2)	(3)	(4)	(5)	(6)
Parent rank	0.129	0.129	0.129			
	(0.002)	(0.002)	(0.002)			
Peer parent rank	0.017	0.028	0.025	0.019	0.028	0.030
	(0.008)	(0.009)	(0.011)	(0.020)	(0.023)	(0.027)
Specification						
Window width	3	5	7	3	5	7
Family fixed effects				X	X	X
N	1,126,200	939,700	754,900	1,126,200	939,700	754,900

Notes: coefficients from equation 5 — the regression of a child's household income rank at age 24 on: their parent household income rank; their peers mean parent rank; and the 3-, 5- or 7-year moving average of peer mean parent rank. Peers are defined by postcode and financial year of birth and exclude the individual in question. The moving average of the peer mean parent rank is taken across adjacent birth cohorts in each postcode. A peer's primary parent must have been a permanent resident of the postcode — not filing from outside it — from 1991 to the year in which the child turned 20. Robust standard errors, clustered by postcode, are in parentheses.

As a further test of the results, I conduct a placebo test, looking for any effect of the peers born in the years either side of an individual's own cohort.[42] In particular, I re-estimate the baseline specification from column (3) of Table 6 as if an individual's financial year of birth was $s + l$ rather than s, where $l \in \{-6, ..., 6\}$. Figure 4 presents the resulting peer effect estimates. The largest and only statistically significant estimate is that on an individual's own cohort. This suggests the results are not driven by inadequate controls for trends within postcodes that may generate spurious correlation between mean peer parental rank and individual outcomes.

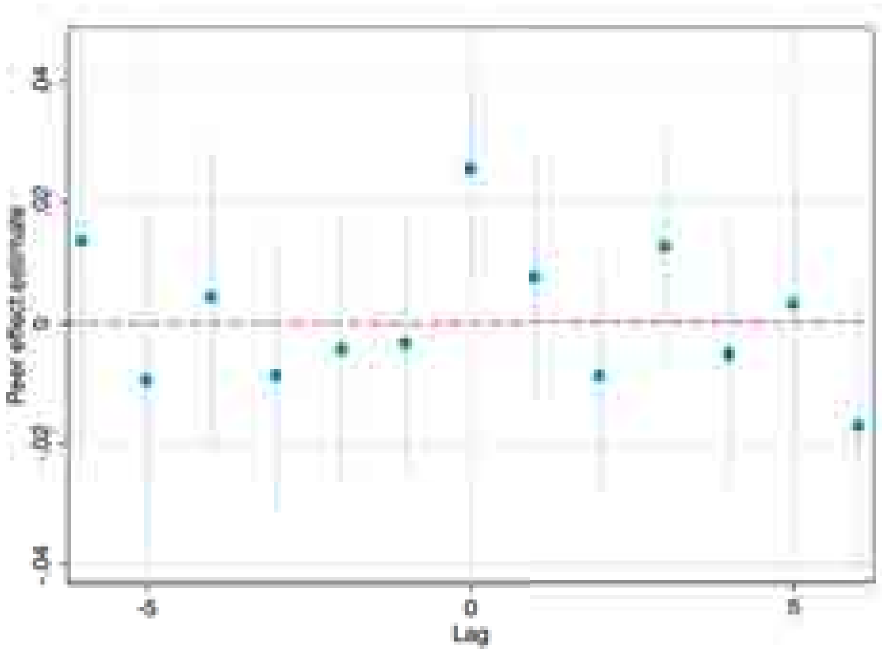
Finally, the identified peer effects are not driven by particularly small or large postcodes. While those in smaller postcodes are perhaps more likely to interact with one another, they may also be a less complete picture of an individual's social network. In Appendix Figure A.8 I present the estimated peer effects restricting attention to those in progressively smaller peer groups. The estimates are less precise, but range between 0.017 and 0.026 and all include the baseline estimates from Table 6 within their confidence intervals.

Peer effects do not appear to fade with time, and rather tend to increase with age.

[41] Black et al. (2013) also note this possibility, stating that they view estimates of theirs based on a 3-year moving average as a "lower bound on true peer effects."

[42] I thank Nathan Hendren for suggesting a test along these lines.

Figure 4: Peer effect estimates: placebo test



Notes: Based on permanent postcode residents. Shows the coefficients (and 95% confidence intervals) from a regression of household income rank at age 24 on own parent household income rank and the mean parent household income rank of peers (defined by shared permanent postcode and a financial year of birth that is shifted by a lag l relative to the individual's own). A 7-year moving average of the mean parent rank of peers is included as a control, in line with the specification in column (3) of Table 6.

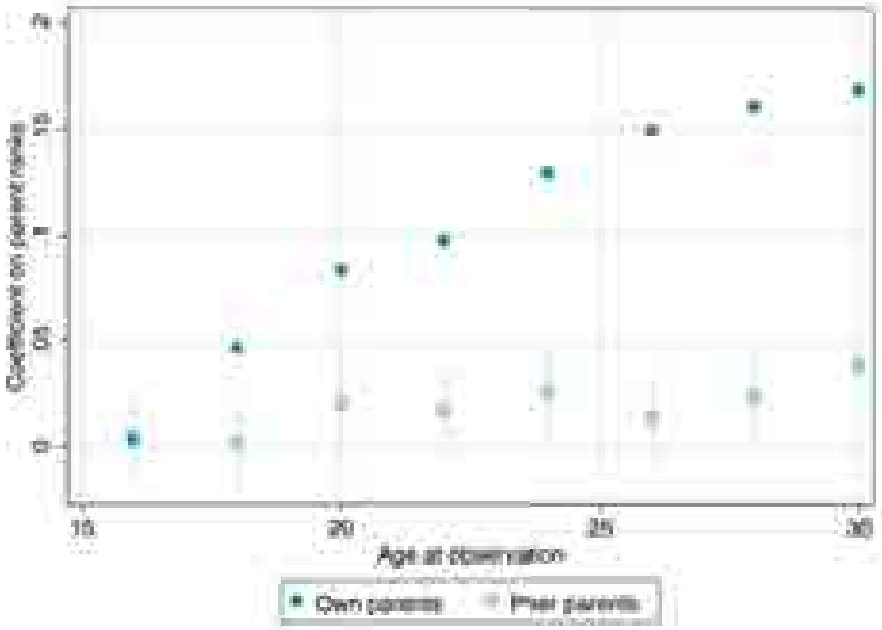
Figure 5 shows the coefficients on parent and peer parent rank lifting the age at which child outcomes are observed from 16 to 30, using the specification in column (3) of Table 6.[43] In this case, peer effects measured at age 30 are slightly larger than those measured at age 24: though they remain around a fifth of the parent effect. This suggests the peer effects are lasting; they may even be subject to a modest version of the life-cycle bias effects apparent in the estimated coefficients on parent rank and well known in the intergenerational mobility literature (Black and Devereux (2011)).

6.2.1 Why might peers matter?

A strength and a weakness of the peer effects identified here is that they may reflect many underlying mechanisms. These include the influences of the peers themselves, and potentially their parents, arising from social interactions. But they may also include nonsocial externalities — for example, an individual may benefit indirectly if

[43]Appendix Table B.7 replicates Table 6 in full with outcomes measured at age 30.

Figure 5: Influence of own and peer parents on household income rank at various ages



Notes: Based on permanent postcode residents. Shows the coefficients (and 95% confidence intervals) from a regression of household income rank at various ages on own parent household income rank and the mean parent household income rank of peers (defined by shared permanent postcode and financial year of birth). A 7-year moving average of the mean parent rank of peers is included as a control, in line with the specification in column (3) of Table 6.

their peers' parents lobby for better teachers, more resources or more opportunities for their peer group cohort.

To better understand the likely mechanisms at play, I reestimate equation 5 separately for men and women, and with separate controls for female and male peers. Table 7 presents the estimated peer effects. For both women and men, it is the rank of their same-sex peers' parents that matters most. The results rule out some otherwise plausible explanations for the peer effects. A simple "boy/girl next door" explanation, whereby peers matter because some individuals marry their peers, would require the opposite-sex peers to matter most. Further, it appears less likely that nonsocial externalities drive the results, as they would then have to operate in a gender specific manner only — for example, richer parents only lobbying for teachers, resources or opportunities that specifically benefit their child's gender.

Finally, data limitations make it difficult to provide a compelling test of whether peers are more important at particular points in life, such as the teenage years. As a suggestive test, I examined how estimates varied if the definition of the peer group was

Table 7: Parent and peer influences on household income rank at age 24 — by individual and peer sex

	(1)	(2)	(3)	(4)
Panel A: Men				
Parent rank	0.133	0.134	0.133	0.134
	(0.002)	(0.002)	(0.002)	(0.002)
Peer parent rank for...				
...all peers	0.028			
	(0.014)			
...male peers		0.019		0.024
		(0.010)		(0.011)
...female peers			0.001	0.001
			(0.010)	(0.011)
N	388,100	383,900	381,900	380,300
Panel B: Women				
Parent rank	0.123	0.123	0.123	0.123
	(0.002)	(0.002)	(0.002)	(0.002)
Peer parent rank for...				
...all peers	0.023			
	(0.015)			
...male peers		0.005		0.007
		(0.011)		(0.012)
...female peers			0.019	0.017
			(0.011)	(0.011)
N	366,800	361,700	362,200	359,600

Notes: coefficients from equation 5 — the regression of a child's household income rank at age 24 on: their parent household income rank; their peers mean parent rank; and the 3-, 5- or 7-year moving average of peer mean parent rank — separately by individual and peer sex. Peers are defined by sex, postcode and financial year of birth and exclude the individual in question. The moving average of the peer mean parent rank is taken across adjacent birth cohorts in each postcode. A peer's primary parent must have been a permanent resident of the postcode — not filing from outside it — from 1991 to the year in which the child turned 20. Robust standard errors, clustered by postcode, are in parentheses.

expanded to include those moving into a postcode-cohort area before the cohort hit the teenage years. However, insufficient variation between these competing peer group measures meant it was not possible to draw firm conclusions. An added difficulty is the absence of direct data on peer relationships — it could be that peer relationships are formed in early childhood but only influence outcomes from adolescence onwards.

It is also difficult, given the data limitations, to nest peer effects in the model of exposure-to-place effects in a compelling way. For example, it is straightforward to include controls for the peer shock that movers experience.[44] However, these fail to

[44]Specifically, I construct controls for the mean parent income rank of the permanent residents of the origin ($\bar{p}_{os}$) and the difference between this and the same for the destination ($\Delta_{ods} := \bar{p}_{ds} - \bar{p}_{os}$). I then re-estimate equation 4, including controls for $\bar{p}_{os}$ and Δ_{ods} that are identical to those for $\bar{y}_{pos}$ and Δ_{odps}. This allows for an exposure-to-peers effect. Whether these measures are based on the

explain the patterns observed — the late childhood exposure-to-place effect remains steady at 0.04, while the exposure-to-peers effects are small and statistically insignificant. It could be that peers do not, in fact, explain any of the causal effects of place examined earlier in the paper. But another possibility is that these simple controls fail to capture the true peers of movers, or their influence on movers' outcomes, any better than the predicted outcomes of the permanent residents do.[45] In the absence of direct measures of peer relationships, their influence on movers seems likely to remain embedded in the black box of the exposure-to-place effect observed here.

Further research on when peers influence outcomes and if they can explain the causal effect of place would be valuable. Also of interest is whether the effects identified here (for permanent residents) operate through the peers themselves or their parents. For example, in a recent contribution Bell et al. (2017) find striking relationships between an individual's patent rate in a specific technology class and the patent rates of not only their father, but their father's colleagues and their neighbors — such findings seem suggestive of a place for role model effects.

7 Conclusion

The seminal work of Chetty and Hendren (2018a) provided fresh evidence that where a child grows up matters for their later life outcomes. In this paper I explore when and why place matters — critical questions for those seeking to redress the inequalities arising from causal place effects. I find place matters most in the teenage years, with a clear role for local labor markets, and evidence for peer effects at a more localised scale.

Exposure-to-place effects are largest in the teenage years, and generally small and not statistically significant in early childhood. This is consistent with age-at-migration studies finding the benefits to language acquisition from migrating a year earlier are generally largest in adolescence as well. It also accords with the findings of Chetty and Hendren (2018a) for earlier life outcomes, such as teenage births and

origin and destination SA4, or on the origin and destination postcodes, does not materially change the results.

[45]The predicted outcomes of the permanent residents in theory include the influence of the full set of peers that lived in that location, even if only temporarily. The controls for peers included here are based only on the permanent residents. A time-varying measure of peers was explored but is made difficult by the gaps in the location histories inherent in the tax data.

college attendance. As such, it seems unlikely this finding is particular to Australia. This finding does not suggest that early childhood is unimportant. Rather, what matters in early childhood may be factors — such as the family or more localized environmental influences — where most variation is within rather than between the large neighborhoods examined here.

Any explanations for causal place effects should also seek to explain the sensitivity of the teenage years. Two possibilities are explored in this paper — that where a child grows up matters because it influences where they end up (local labor market conditions) and who they grow up with (peer effects).

Local labor market conditions can partly explain the observed patterns, based on tests with endogenous and exogenous controls. When using exogenous controls based on those ending up in a place (rather than those growing up there), anywhere from 15-55% of the effect of exposure to place in late childhood is explained away. This is not driven by idiosyncratic factors, such as the mining boom which coincided with the period over which child outcomes are observed. Rather, the results point to a more general potential for local labor markets to generate exposure effect patterns — as where a child ends up is itself a function of their exposure to places growing up.

Sizable peer effects are also present among the permanent residents of Australian postcodes. Given the psychological literature on the heightened importance of peer relationships in adolescence (Brown and Larson (2009)), these may provide an explanation for some of the residual influence of place. While simple controls for peers fail to explain the exposure-to-place effects, this may reflect a need for better data on actual peer relationships. Further promising directions for research may include more detailed examinations of when peers matter, whether peers' parents matter directly (as role models or job contacts, say) or indirectly (as a measure of peer 'quality') and looking beyond the linear-in-means model of peer effects examined here.

Finally, this paper serves as a validation of the research design introduced in Chetty and Hendren (2018a), and its broader applicability in settings with smaller populations and less geographic variation in intergenerational mobility. Future applications of this design in settings capturing child outcomes even later in life, and with the potential for more accurate measures of peer relationships, may yield further insights into why place matters.

References

Aaronson, D. (1998). Using sibling data to estimate the impact of neighborhoods on children's educational outcomes, *Journal of Human Resources* **33**(4): 915–946.

Angrist, J. D. (2014). The perils of peer effects, *Labour Economics* **30**: 98–108.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press, Princeton, N.J.

Australian Bureau of Statistics (1991). Census of Population and Housing, 1991, *Cat. no. 2101.0*, ABS, Canberra.

Australian Bureau of Statistics (2010). Family Characteristics Survey, Australia, 2009-10, *Cat. no. 4442.0*, ABS, Canberra.

Australian Bureau of Statistics (2011). Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, July 2011, *Cat. no. 1270.0.55.001*, ABS, Canberra.

Australian Bureau of Statistics (2017a). Australian Demographic Statistics, June 2017, *Cat. no. 3101.0*, ABS, Canberra.

Australian Bureau of Statistics (2017b). Births, Australia, 2016, *Cat. no. 3301.0*, ABS, Canberra.

Australian Bureau of Statistics (2017c). Consumer Price Index, Australia, March 2017, *Cat. no. 6401.0*, ABS, Canberra.

Australian Bureau of Statistics (2017d). Labour Force, Australia, Detailed, November 2017, *Cat. no. 6291.0.55.001*, ABS, Canberra.

Basu, S. (2018). Age-of-arrival effects on the education of immigrant children: A sibling study, *Journal of Family and Economic Issues (forthcoming)* .

Bell, A. M., Chetty, R., Jaravel, X., Petkova, N. and Van Reenen, J. (2017). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. NBER Working Paper No. 24062.

Bingley, P., Cappellari, L. and Tatsiramos, K. (2016). Family, community and long-term earnings inequality. IZA Discussion Paper No. 10089.

Black, S. E. and Devereux, P. J. (2011). Recent developments in intergenerational mobility, *Handbook of Labor Economics*, Vol. 4, Elsevier, chapter 16, pp. 1487–1541.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2013). Under pressure? The effect of peers on outcomes of young adults, *Journal of Labor Economics* **31**(1): 119–153.

Bleakley, H. and Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants, *Review of Economics and Statistics* **86**(2): 481–496.

Borjas, G. J. (1992). Ethnic capital and intergenerational mobility, *The Quarterly Journal of Economics* **107**(1): 123–150.

Bratberg, E., Davis, J., Mazumder, B., Nybom, M., Schnitzlein, D. D. and Vaage, K. (2017). A comparison of intergenerational mobility curves in Germany, Norway, Sweden, and the US, *The Scandinavian Journal of Economics* **119**(1): 72–101.

Brown, B. B. and Larson, J. (2009). Peer relationships in adolescence, *Handbook of Adolescent Psychology*, Vol. 2, John Wiley & Sons, chapter 3.

Case, A., Fertig, A. and Paxson, C. (2005). The lasting impact of childhood health and circumstance, *Journal of Health Economics* **24**(2): 365–389.

Chetty, R. and Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects, *The Quarterly Journal of Economics (forthcoming)* .

Chetty, R. and Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility II: County-level estimates, *The Quarterly Journal of Economics (forthcoming)* .

Chetty, R., Hendren, N. and Katz, L. F. (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.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States, *The Quarterly Journal of Economics* **129**(4): 1553–1623.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children, *American Economic Review (forthcoming)* .

Corak, M. and Heisz, A. (1999). The intergenerational earnings and income mobility of Canadian men: Evidence from longitudinal income tax data, *Journal of Human Resources* **34**(3): 504–533.

Corak, M. and Piraino, P. (2011). The intergenerational transmission of employers, *Journal of Labor Economics* **29**(1): 37–68.

Cunha, F. and Heckman, J. (2007). The technology of skill formation, *American Economic Review* **97**(2): 31–47.

Currie, J. and Almond, D. (2011). Human capital development before age five, *Handbook of Labor Economics*, Vol. 4, Elsevier, pp. 1315–1486.

Damm, A. P. and Dustmann, C. (2014). Does growing up in a high crime neighborhood affect youth criminal behavior?, *American Economic Review* **104**(6): 1806–1832.

Gaviria, A. and Raphael, S. (2001). School-based peer effects and juvenile behavior, *Review of Economics and Statistics* **83**(2): 257–268.

Greene, W. H. (2003). *Econometric analysis*, 5th edition edn, Prentice Hall, Upper Saddle River, N.J.

Hardin, J. W. (2002). The robust variance estimator for two-stage models, *Stata Journal* **2**(3): 253–266.

Hole, A. R. (2006). Calculating Murphy-Topel variance estimates in Stata: A simplified procedure, *Stata Journal* **6**(4): 521–529.

Katz, L. F., Kling, J. R. and Liebman, J. B. (2001). Moving to opportunity in Boston: Early results of a randomized mobility experiment, *The Quarterly Journal of Economics* **116**(2): 607–654.

Ludwig, J., Liebman, J. B., Kling, J. R., Duncan, G. J., Katz, L. F., Kessler, R. C. and Sanbonmatsu, L. (2008). What can we learn about neighborhood effects from the Moving to Opportunity experiment?, *American Journal of Sociology* **114**(1): 144–188.

Manski, C. F. (2000). Economic analysis of social interactions, *Journal of Economic Perspectives* **14**(3): 115–136.

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data, *Review of Economics and Statistics* **87**(2): 235–255.

Murphy, K. M. and Topel, R. H. (1985). Estimation and inference in two-step econometric models, *Journal of Business & Economic Statistics* **3**(4): 370–79.

Nybom, M. and Stuhler, J. (2017). Biases in standard measures of intergenerational income dependence, *Journal of Human Resources* **52**(3): 800–825.

Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood, *The Quarterly Journal of Economics* **118**(4): 1533–1575.

Overman, H. G. (2002). Neighbourhood effects in large and small neighbourhoods, *Urban Studies* **39**(1): 117–130.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors, *International Economic Review* **25**(1): 221–247.

Page, M. E. and Solon, G. (2003a). Correlations between brothers and neighboring boys in their adult earnings: The importance of being urban, *Journal of Labor Economics* **21**(4): 831–855.

Page, M. E. and Solon, G. (2003b). Correlations between sisters and neighbouring girls in their subsequent income as adults, *Journal of Applied Econometrics* **18**(5): 545–562.

Raaum, O., Salvanes, K. G. and Sørensen, E. Ø. (2006). The neighbourhood is not what it used to be, *The Economic Journal* **116**(508): 200–222.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?, *Handbook of the Economics of Education*, Vol. 3, Elsevier, chapter 4, pp. 249–277.

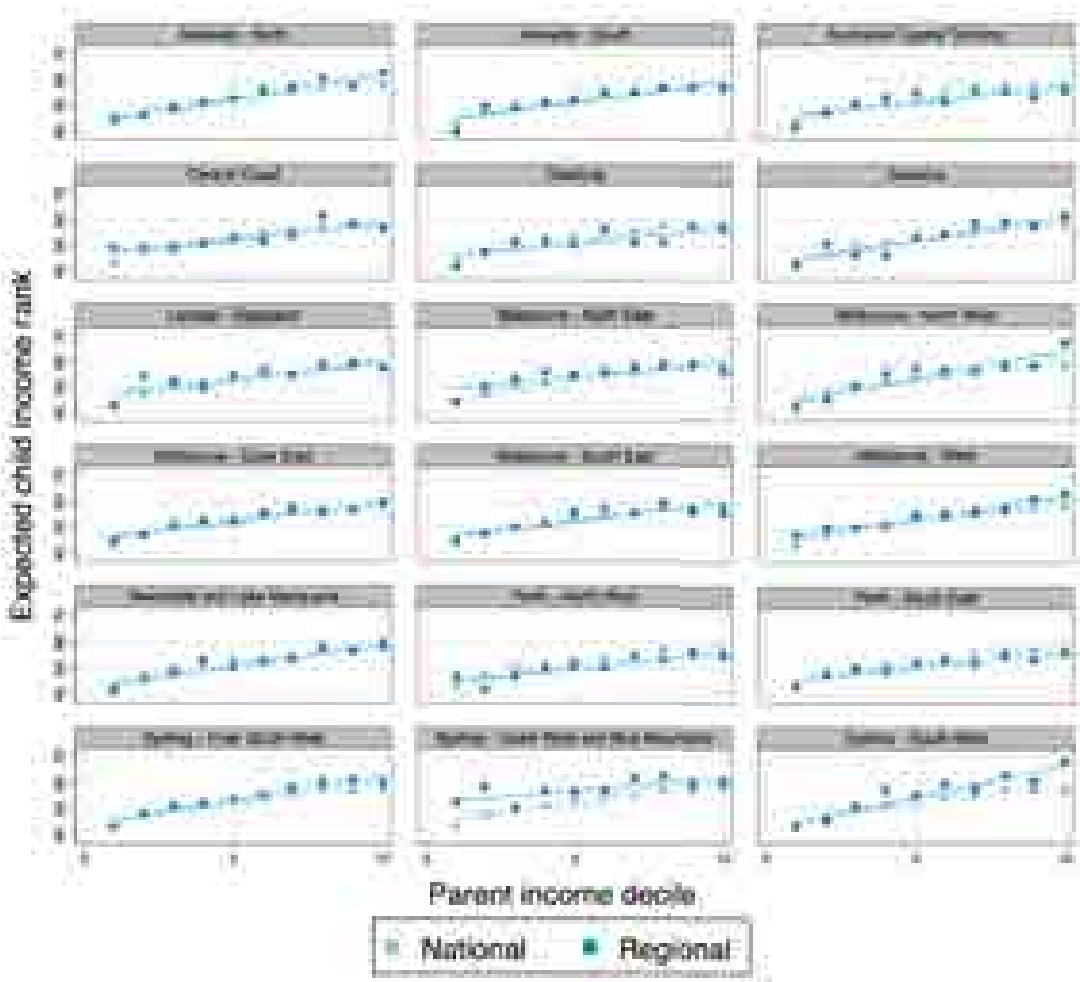
Sharkey, P. and Faber, J. W. (2014). Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects, *Annual Review of Sociology* **40**: 559–579.

Van den Berg, G. J., Lundborg, P., Nystedt, P. and Rooth, D.-O. (2014). Critical periods during childhood and adolescence, *Journal of the European Economic Association* **12**(6): 1521–1557.

Wodtke, G. T. (2013). Duration and timing of exposure to neighborhood poverty and the risk of adolescent parenthood, *Demography* **50**(5): 1765–1788.

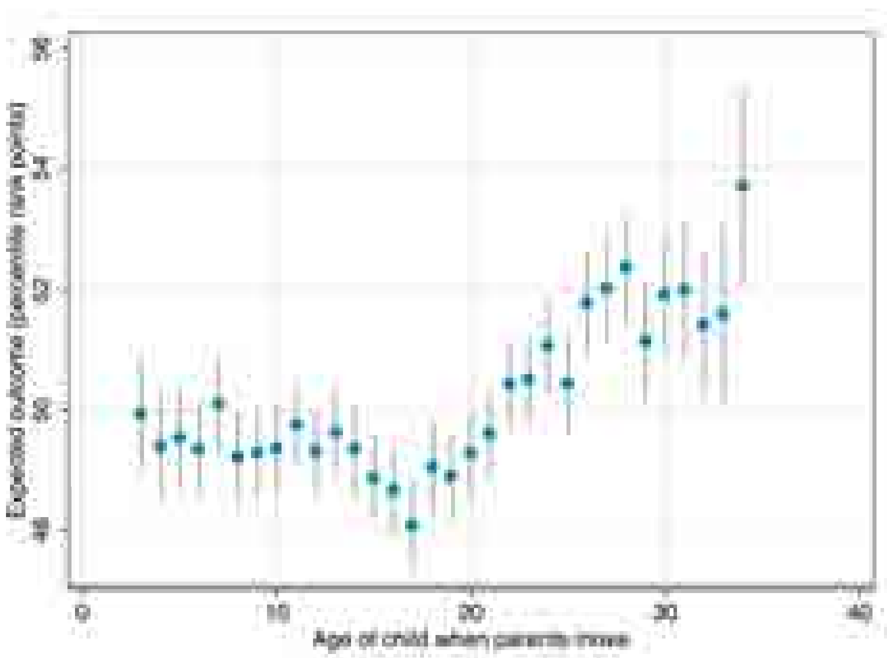
A Additional charts

Figure A.1: National and regional relationships between parent and child income ranks: permanent residents born in 1978



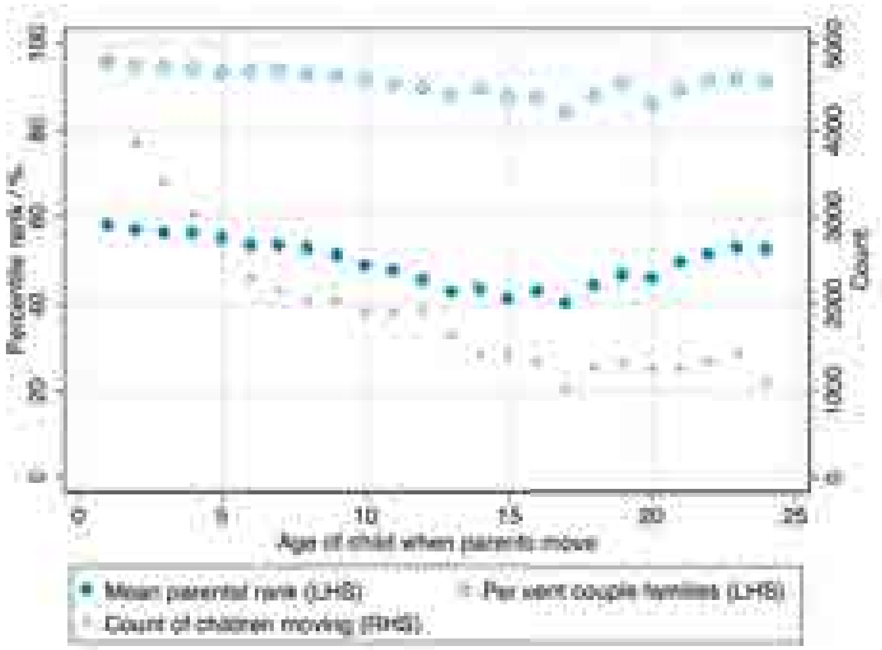
Notes: Based on the sample of permanent residents. Chart illustrates the mean household total income rank at age 24, by parent income decile, for children born in 1978 and in one of the 18 largest SA4.

Figure A.2: Expected outcome of child: born in 1991, to parents with median income, and moving between places where similar permanent residents end up with median income



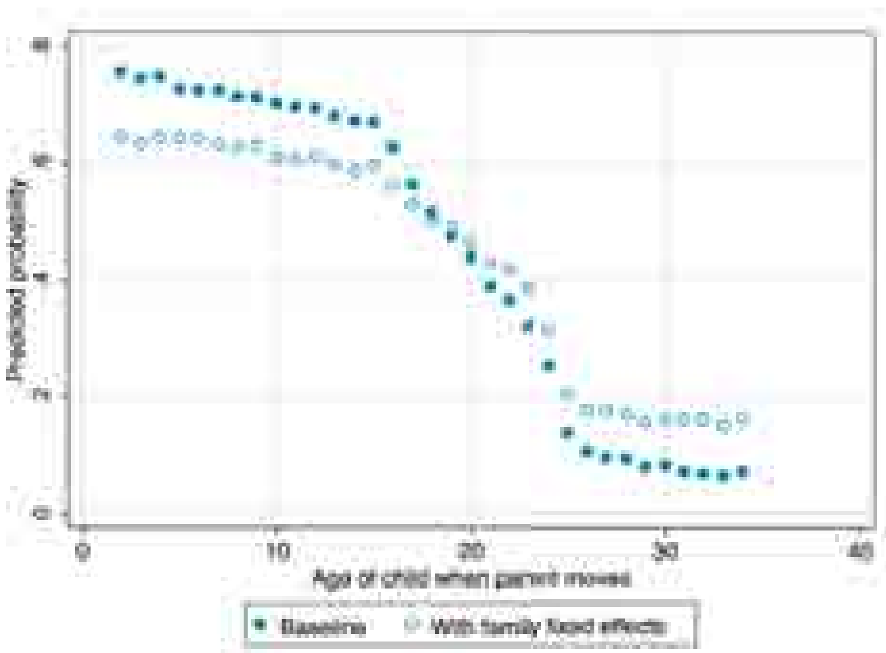
Notes: Estimated linear combination $\alpha^1_{1991} + 50\alpha^2_{1991} + \zeta^1_m + 50\zeta^2_m$ of coefficients from equation 3. This captures the expected household income rank at age 24 for a child: born in 1991; with parents at the 50^{th} percentile of the income distribution; and moving at age m between an origin and destination where their predicted outcome based on permanent residents is also the 50^{th} percentile. The full regression regresses the adult ranks y_i of those whose parents move once in their childhood on the interaction of their age at parent move m with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin.

Figure A.3: Family characteristics by age at move: 1991 cohort



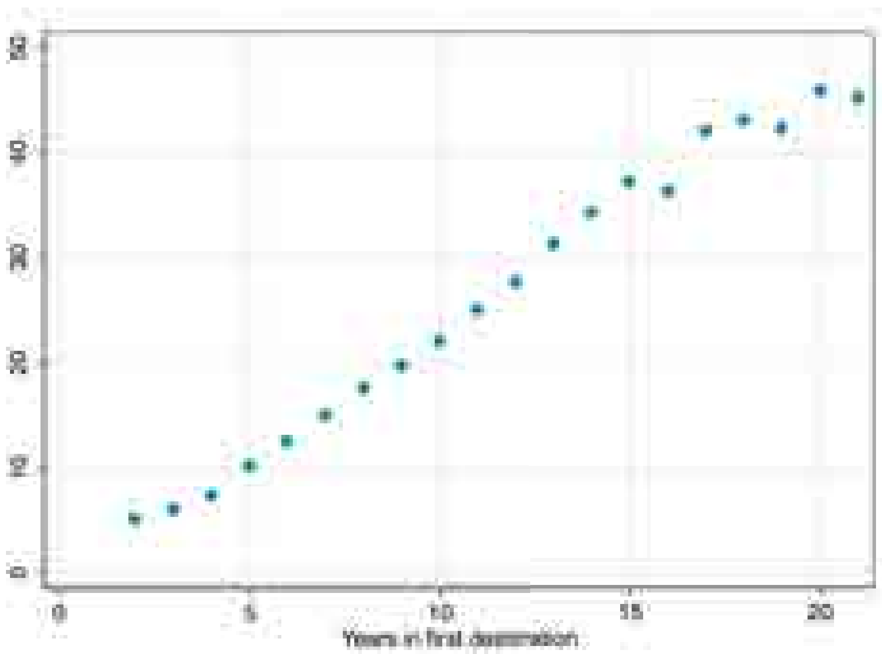
Notes: For the individuals born in the 1991 financial year whose parents move once, shows the mean parent rank, proportion in couple families and sample size by the individual's age at move.

Figure A.4: Moving once: predicted probability of child being in destination at age 24, by age at parent move



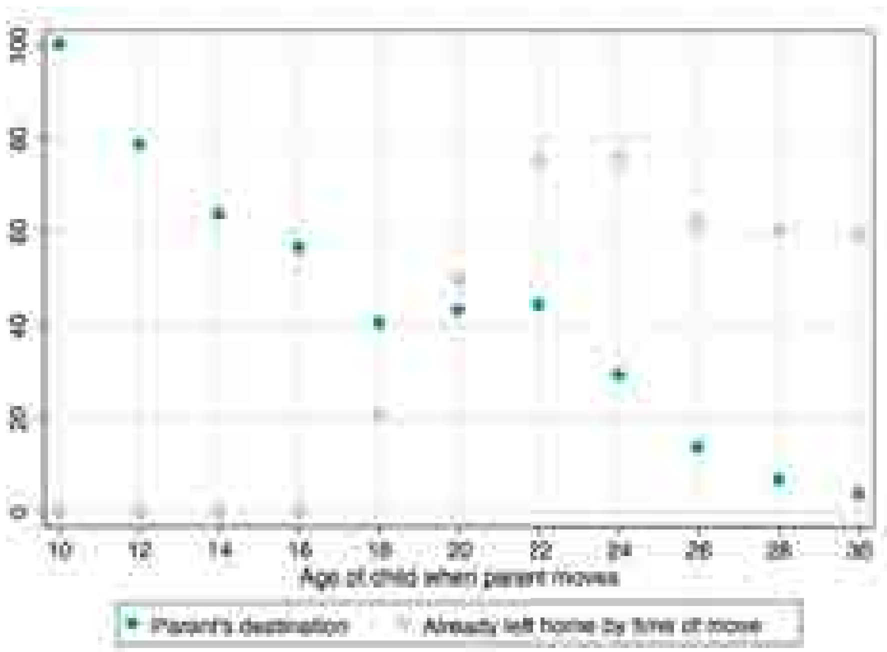
Notes: Based on the 1-time movers sample, restricted to those filing a tax return from a known location at age 24. The figure shows the predicted probability a child is filing from their parent's destination at age 24, by their age at move. Predicted probabilities are generated first from a linear regression of the binary outcome that a child is in their parents destination on age at move and birth cohort dummies (baseline), then from the same regression allowing for family fixed effects.

Figure A.5: Moving twice: % of children in first destination at age 24, by years spent there



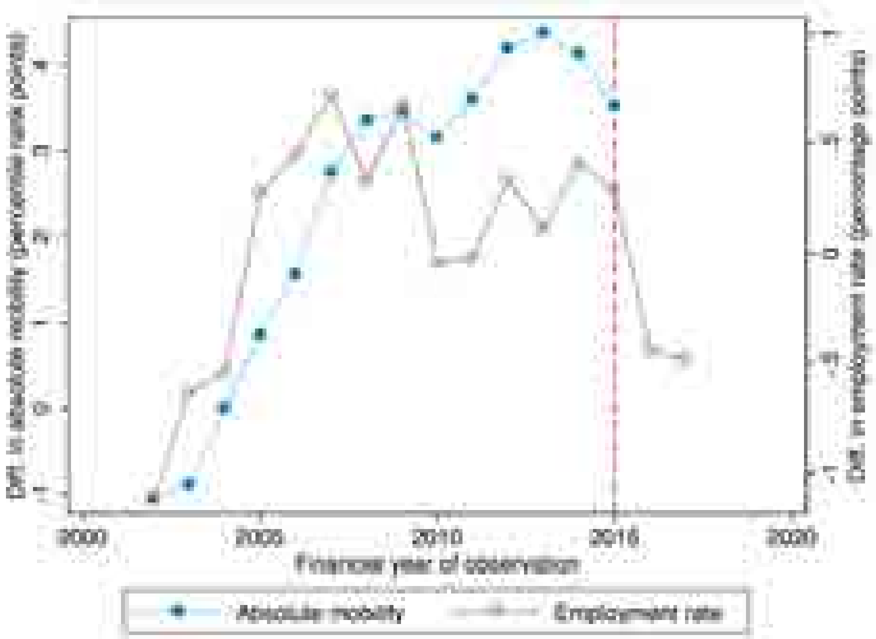
Notes: Based on a 2-time movers sample, defined analogously to the 1-time movers sample. For children of a parent who moves twice, shows the % of those in the first destination at age 24 by the time the parent spent in the first destination. The child's location is known for those lodging a tax return in the year they turn 24 with a valid postcode.

Figure A.6: Moving once, survey data: % of children in destination at age 24, by age at parent move



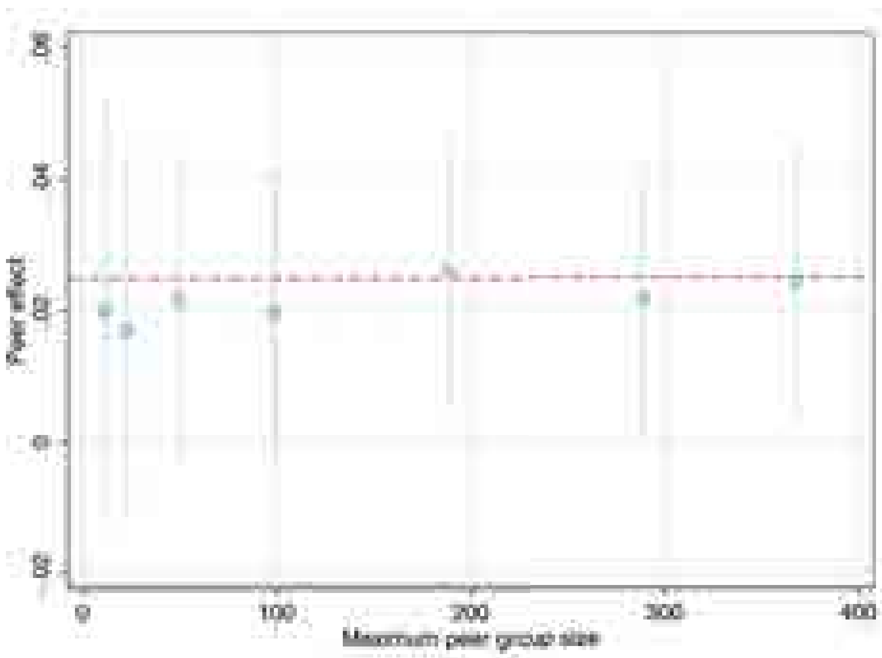
Notes: Replicates the 1-time mover sample as closely as possible using the Household Income and Labour Dynamics in Australia (HILDA) survey. I link respondents to their biological mother (or father, if the mother is missing). I restrict attention to respondents whose linked parent lived in two SA4 from 2001 (the first year of HILDA) to the year in which the respondent turned 35. Once again one SA4 episode must immediately precede the other. The chart shows the per cent share of respondents living, at age 24, in the parent's destination by their age at move. Also shown are the per cent share who had already left home by the time their parent moved.

Figure A.7: Intergenerational mobility and economic conditions: deviation from national average in resource-rich states



Notes: Illustrates the deviation from the national average in an individual's expected household income rank, given their parents were permanent residents of the resource-rich states of Queensland and Western Australia and were at the 25^{th} percentile of the national parent income distribution. This is shown alongside the deviation of those states' unemployment rates from the national average, from Australian Bureau of Statistics (2017d).

Figure A.8: Effect of peer parents on household income rank at age 24: by peer group size



Notes: Based on permanent postcode residents. Shows the peer effect estimates (and 95% confidence intervals) from a regression of household income rank at age 24 on own parent household income rank and the mean parent household income rank of peers (defined by shared permanent postcode and financial year of birth). A 7-year moving average of the mean parent rank of peers is included as a control, in line with the specification in column (3) of Table 6. This regression is run for individuals in progressively smaller cohorts, with the chosen thresholds corresponding to the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 95^{th} percentiles of cohort size in the sample. The dashed line indicates the baseline estimate on the full sample from column (3) of Table 6.

B Additional tables

Table B.1: Summary statistics for permanent residents and one-time movers

	Permanent residents			1-time movers		
	Mean	Std. dev.	Median	Mean	Std. dev.	Median
Panel A: Family background						
Parent income ($)	79,300	71,600	72,600	86,100	80,700	77,700
Parent income rank	50.7	28.5	51	54.8	27.3	56
Indicator, in a couple family	0.87	0.33	1	0.91	0.29	1
Family size	2.7	1.2	3	2.7	1.2	2
Panel B: Outcomes						
Child income ($)	61,800	46,400	54,100	62,600	45,300	54,800
Child rank	52.2	28.4	53	52.8	28.6	54
N		1,683,800			313,900	

Notes: The full sample consists of those children born between 1978-91, remaining resident in Australia through to 2015 and linked to parents. The permanent residents are those children whose primary parent files from only one SA4 from 1991 through to the year the child turned 35. The 1-time movers are those whose primary parent filed from two SA4 from 1991 through to the year the child turned 35, filed from each at least twice, began filing in the destination the year after they ceased filing in the origin, and moved at least 15 kilometres (based on postcode centroids). Parent income is the average household total pre-tax income from 1991-2001 in 2015 dollars. Child income is the household total pre-tax income in the year the child turns 24. Ranks are calculated separately for each birth cohort.

Table B.2: Difference between destination and origin: 1-time mover subsample

	Mean	Std. dev.	Median
Mean permanent resident parent rank	-1.06	9.98	-.83
Number of permanent residents	-95	1,000	-57
Predicted child rank	-.083	5.09	-.088
N		313,900	

Notes: Shows differences in the characteristics of the 1-time movers destinations and origins. These characteristics of place are based on the permanent residents. The difference in the predicted child rank is simply the difference in predicted values for a child in birth cohort s and with parent income rank p for a permanent resident of the origin o versus the destination d, that is $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$. The difference in the mean permanent resident parent rank is the difference in the means of the parent ranks p of the permanent residents in the same cohort s in the origin o and destination d, that is $\bar{p}_{os} - \bar{p}_{ds}$.

Table B.3: Exposure effect estimates and model fit statistics: by model specification

	Linear	Piecewise linear with kink at age...						
		10	11	12	13	14	15	16
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.033							
	(0.002)							
Early		0.010	0.011	0.015	0.018	0.020	0.023	0.026
		(0.008)	(0.007)	(0.006)	(0.005)	(0.005)	(0.004)	(0.004)
Late		0.039	0.042	0.043	0.043	0.045	0.045	0.045
		(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)
Post-outcome	0.008	0.008	0.008	0.008	0.009	0.009	0.009	0.009
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
$(R^2 - R^2_{max})10^6$	-10	-2	0	-1	-2	-2	-4	-6
$(aR^2 - aR^2_{max})10^6$	-9	-2	0	-1	-2	-2	-4	-6
$AIC_{min} - AIC$	-10	-3	0	-1	-3	-3	-5	-7
$BIC_{min} - BIC$	0	-4	-1	-2	-3	-3	-6	-8
N	264,500	264,500	264,500	264,500	264,500	264,500	264,500	264,500

Notes: Exposure effect estimates and model fit statistics for competing models of exposure effects — a constant exposure effects model as in Chetty and Hendren (2018a) and a piecewise linear model with the kink at varying ages. Model fit statistics are transformed as described to aid readability — higher values indicate better fits. Statistics are estimated from equation 4 for early ($m \in \{2, ..., k\}$), late ($m \in \{k, ..., 24\}$) or post-outcome ($m \in \{25, ..., 34\}$) exposure for varying values of the kink k (columns (2)-(8)) or assuming constant exposure effects in early and late childhood (column (1)). The coefficients represent the expected boost to an individual's household income rank associated with an additional year at this stage of life in a destination with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their time exposed to the destination at each life stage with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. Murphy-Topel standard errors are in parentheses.

Table B.4: Exposure effect estimates: by population subgroup

	Baseline	Gender		Parent		Destination	
		Male	Female	Poorer	Richer	Worse	Better
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Early	0.011	0.011	0.011	0.026	0.003	0.009	0.019
	(0.007)	(0.009)	(0.010)	(0.011)	(0.009)	(0.016)	(0.014)
Late	0.042	0.049	0.032	0.035	0.045	0.039	0.037
	(0.003)	(0.005)	(0.005)	(0.006)	(0.004)	(0.008)	(0.008)
Post-outcome	0.008	0.031	-0.015	0.038	-0.006	0.001	-0.018
	(0.013)	(0.018)	(0.019)	(0.022)	(0.016)	(0.028)	(0.030)
N	264,500	135,100	129,400	124,000	140,500	132,300	132,100

Notes: Estimates of the exposure effects $\gamma_{\tilde{m}}$ from equation 4 for early ($m \in \{2, ..., 11\}$), late ($m \in \{12, ..., 24\}$) or post-outcome ($m \in \{25, ..., 34\}$) exposure for the full sample and various subpopulations. These represent the expected boost to an individual's household income rank associated with an additional year at this stage of life in a destination with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their time exposed to the destination at each life stage with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. The subpopulations considered are, in order: males and females; those with parental income rank ≤ 50 (or not); and those with $\Delta_{odps} < 0$ (or not). Murphy-Topel standard errors are in parentheses.

Table B.5: Moves within non-mining states: exposure effects with local labor market controls

	Baseline	With child location controls			
	(1)	(2)	(3)	(4)	(5)
Early	0.014	0.017	0.014	0.000	0.021
	(0.010)	(0.009)	(0.009)	(0.028)	(0.017)
Late	0.041	0.027	0.023	0.029	0.025
	(0.005)	(0.005)	(0.005)	(0.011)	(0.016)
Endogenous controls					
child location ...		X	X		
... X child cohort ...			X		
Exogenous controls					
Ender predictions				X	
Ender predictions, kink at 15					X
N	160,200	160,200	160,200	126,500	126,500

Notes: Estimates of the exposure effects $\gamma_{\bar{m}}$ from equation 4 for early ($m \in \{2, ..., 11\}$), late ($m \in \{12, ..., 24\}$) or post-outcome ($m \in \{25, ..., 34\}$) exposure with various controls for child location. These represent the expected boost to an individual's household income rank associated with an additional year at this stage of life in a destination with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their time exposed to the destination at each life stage with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. Child location controls cover: none (1); child location fixed effects (2); child location-cohort fixed effects (3); and exogenous controls (4) and (5). Child location is simply the corresponding SA4 for those filing a tax return with a valid postcode. Standard errors are in parentheses (Murphy-Topel for (1)-(3), naive otherwise).

Table B.6: Moves within cities: exposure effects with local labor market controls

	Baseline	With child location controls			
	(1)	(2)	(3)	(4)	(5)
Early	0.015	0.015	0.016	0.031	0.037
	(0.014)	(0.013)	(0.013)	(0.036)	(0.021)
Late	0.041	0.027	0.024	0.024	0.014
	(0.007)	(0.007)	(0.007)	(0.014)	(0.021)
Endogenous controls					
child location ...		X	X		
... X child cohort ...			X		
Exogenous controls					
Ender predictions				X	
Ender predictions, kink at 15					X
N	83,200	83,200	83,200	68,500	68,500

Notes: Estimates of the exposure effects $\gamma_{\bar{m}}$ from equation 4 for early ($m \in \{2, ..., 11\}$), late ($m \in \{12, ..., 24\}$) or post-outcome ($m \in \{25, ..., 34\}$) exposure with various controls for child location. These represent the expected boost to an individual's household income rank associated with an additional year at this stage of life in a destination with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their time exposed to the destination at each life stage with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. Child location controls cover: none (1); child location fixed effects (2); child location-cohort fixed effects (3); and exogenous controls (4) and (5). Child location is simply the corresponding SA4 for those filing a tax return with a valid postcode. Standard errors are in parentheses (Murphy-Topel for (1)-(3), naive otherwise).

Table B.7: Parent and peer influences on household income rank at age 30

	(1)	(2)	(3)	(4)	(5)	(6)
Parent rank	0.170	0.170	0.168			
	(0.002)	(0.002)	(0.002)			
Peers	0.022	0.040	0.038	0.007	0.024	0.029
	(0.011)	(0.011)	(0.013)	(0.035)	(0.039)	(0.049)
Specification						
Window width	3	5	7	3	5	7
Family fixed effects				X	X	X
N	716,500	606,300	499,700	716,500	606,300	499,700

Notes: coefficients from equation 5 — the regression of a child's household income rank at age 24 on: their parent household income rank; and their peers mean parent rank expressed as a deviation from the 3-, 5- or 7-year moving average. Peers are defined by postcode and financial year of birth and exclude the individual in question. A peer's primary parent must have been a permanent resident of the postcode — not filing from outside it — from 1991 to the year in which the child turned 20. Robust standard errors, clustered by postcode, are in parentheses.

C Validation exercises

This Appendix replicates validation exercises conducted by Chetty and Hendren (2018a), with largely comforting results. The first set of tests considers the robustness of the estimates to more general specifications and later ages of observation, the remainder examine in more detail the key identifying assumption — that selection effects do not vary with the age at move of the child.

C.1 Specification and age at observation

In Figure C.1 I show that the patterns of exposure effects observed in Figure 1 emerge even if using the more general specification in equation 2. This more general specification replaces parametric controls for origin and disruption effects with fixed effects for each combination of parent income decile, cohort, origin and age at move. Age-invariant selection effects, positive exposure effects and the pronounced sensitivity of the teenage years all remain apparent.

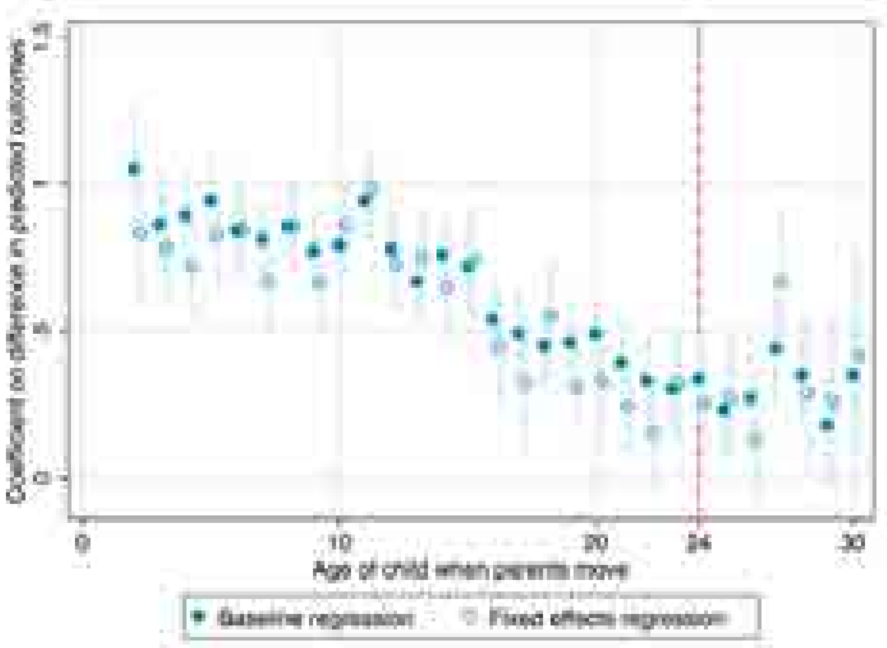
In Table C.2 I switch attention to the models in which place effects are explicitly modeled as a function of exposure to place. Once again moving from the baseline model (column (1)) to one where parametric controls for origin and disruption effects are replaced by fixed effects (column (2)) has little effect on the estimates — if anything the sensitivity of the teenage years is even more pronounced. Lifting the age at which income is measured from 24 to 26, 28 or 30 also leaves the general conclusions unchanged.

C.2 Family fixed effects

The key identifying assumption behind the methodology here, and in Chetty and Hendren (2018a), is that selection effects do not vary with the age at move of the child. This seems unlikely to be true in a strict sense — certainly observables appear to differ slightly by age at move (Appendix Figure A.3) — but it remains unclear whether the extent of any variation is sufficient to meaningfully bias the results.

An obvious place to begin testing this assumption is through the addition of family fixed effects to control for any fixed differences between families moving with children at different ages. I also consider family-sex fixed effects given the evidence in Table B.4 of heterogenous exposure effects by child sex. In these fixed effect tests, identification

Figure C.1: Place exposure effect estimates for child income rank in adulthood.



Notes: Estimated coefficients b_m from equations 2 and 3. The b_m capture the expected boost to an individual's household income rank at age 24 from moving at age m to a place with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their age at parent move m with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls vary across the specifications. Equation 2 includes indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin capture, alongside fixed effects for each combination of parent income decile, origin, cohort and age at move. Equation 3 discards the fixed effects and includes instead: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); and disruption effects (via indicators for age at move and their interaction with parental rank). This replicates Figure IV from Chetty and Hendren (2018a).

comes from comparing siblings in different cohorts who thus differ in both their length of time exposed to the destination (the e_m) and in the predicted outcomes of their destination relative to the origin Δ_{odps} (since these are allowed to vary by birth cohort s). This requires a greater degree of precision in the measurement of the predicted outcomes to avoid attenuation bias, so more stringent sample restrictions on the estimated precision in Δ_{odps} are also considered.

The results are comforting. With family fixed effects, the estimated exposure effect falls modestly from 0.042 to around 0.03. With family-sex fixed effects, the fall is even less pronounced, with the estimates remaining at around 0.04. This suggests heterogeneity by child sex is important in the Australian setting. I also examine the selection effect — the expected boost to an individual's household income rank from

Table C.1: Exposure effect estimates: more general specification and later ages of observation

	Baseline	General	Later age of observation		
	(1)	(2)	(3)	(4)	(5)
Early	0.011	-0.008	0.001	0.001	-0.013
	(0.007)	(0.009)	(0.012)	(0.023)	(0.052)
Late	0.042	0.052	0.044	0.045	0.044
	(0.003)	(0.005)	(0.005)	(0.006)	(0.007)
Post-outcome	0.008	0.013	-0.010	-0.004	0.033
	(0.013)	(0.019)	(0.015)	(0.016)	(0.021)
Age of observation	24	24	26	28	30
N	264,500	264,500	221,000	181,900	142,200

Notes: Estimates of the exposure effects $\gamma_{\bar{m}}$ from equation 4 for early ($m \in \{2, ..., 11\}$), late ($m \in \{12, ..., 24\}$) or post-24 ($m \in \{25, ..., 34\}$) exposure, with either a more general set of controls (2) or for a later age of observation (3)-(5). These represent the expected boost to an individual's household income rank at the given age associated with an additional year at this stage of life in a destination with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their time exposed to the destination at each life stage with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. In (1) and (3)-(5) controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. In (2) all but the last control is replaced by a much larger set of fixed effects for each combination of parent decile, origin, cohort and age at move. Murphy-Topel standard errors are in parentheses.

having their parent move to a destination with 1 percentile rank higher expected outcomes *after* the child turns 24. With family-sex fixed effects this selection effect is halved and no longer statistically significant. It falls further towards zero as the sample is restricted to moves where the difference in origin and destination predicted outcomes is more precisely estimated.

C.3 Exogenous moves

A remaining concern is that there may be time-varying differences between families moving with children at different ages. Relationship breakdown, job loss or promotion could all give rise to moves, and themselves matter for outcomes in proportion to the time a child is exposed to them. The next test considers subsamples of moves that are more plausibly exogenous — moves out of locations in years with unusually large outflows for that location — and then re-estimates the exposure effects.

Let k_{pt} be the number of families leaving postcode p in financial year t as a proportion of the average number of families leaving the same postcode from 1991

Table C.2: Exposure effect estimates: more general specification and family fixed effects

	Baseline	Family fixed effects			Family-sex fixed effects		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Early	0.011	-0.025	-0.012	-0.005	-0.021	-0.002	0.001
	(0.007)	(0.010)	(0.013)	(0.013)	(0.015)	(0.013)	(0.013)
Late	0.042	0.028	0.032	0.030	0.039	0.035	0.040
	(0.003)	(0.006)	(0.008)	(0.008)	(0.043)	(0.008)	(0.009)
Post-outcome	0.008	0.040	0.011	0.025	0.018	-0.036	-0.028
	(0.013)	(0.017)	(0.043)	(0.022)	(0.028)	(0.022)	(0.026)
Selection	0.292	0.365	0.293	0.287	0.140	0.097	0.047
	(0.068)	(0.104)	(0.145)	(0.133)	(0.361)	(0.123)	(0.140)
Sample s.e. on Δ_{odps}	< 2	< 2	< 1.75	< 1.5	< 2	< 1.75	< 1.5
N	264,500	263,100	228,300	175,400	263,100	228,300	175,400

Notes: Estimates of the exposure effects $\gamma_{\bar{m}}$ from equation 4 for early ($m \in \{2, ..., 11\}$), late ($m \in \{12, ..., 24\}$) or post-outcome ($m \in \{25, ..., 34\}$) exposure, with either family or family-sex fixed effects. These represent the expected boost to an individual's household income rank at age 24 associated with an additional year at this stage of life in a destination with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their time exposed to the destination at each life stage with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank and cohort in the destination versus the origin. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. Attention is restricted to families with five or fewer children. Murphy-Topel standard errors are in parentheses.

to 2014. As in Chetty and Hendren (2018a), many of those postcode-years with the highest relative outflows k_{pt} are associated with external shocks (such as mine closures in the Australian setting).[46] As noted by Chetty and Hendren (2018a), while moves in subsamples with high values of k_{pt} may be more often for exogenous reasons, the destinations may still reflect endogenous choices. I follow them in instrumenting for Δ_{odps} and y_{ops} by $E[\Delta_{odps}|p, q]$ and $E[y_{ops}|p, q]$ — the mean Δ_{odps} and y_{ops} for all movers in the sample from postcode p and in parental income decile q. I also present OLS estimates that do not account for endogenous choice of destination.

Figure C.2 shows the estimated late childhood exposure effect and its 95% confidence interval for subsamples drawn from moves that were part of progressively larger relative outflows from a postcode. I consider moves where k_{pt} was above its median value, 55^{th} percentile and so on to the 95^{th} percentile. The results are mixed. Below

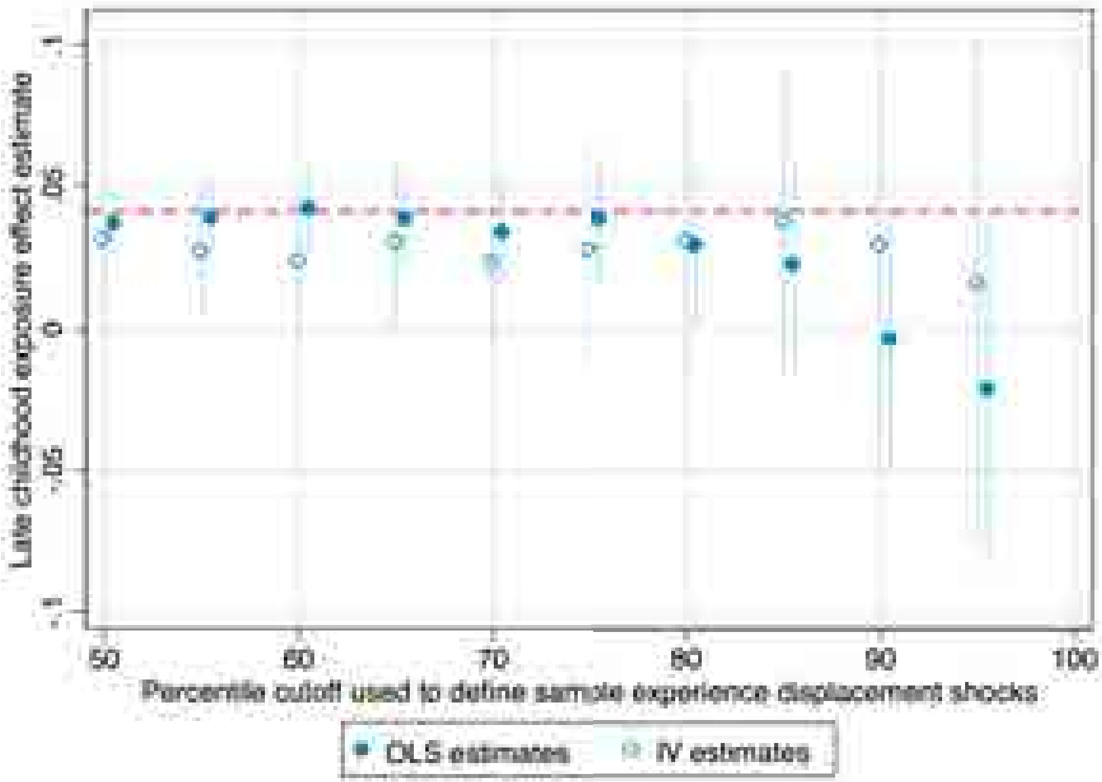
[46] Postcode-years with less than ten families leaving are dropped to avoid have high relative outflows that are driven by small underlying populations. I use the same threshold as in Chetty and Hendren (2018a), purely to remain as close as reasonable to their specification.

the 80^{th} percentile of relative postcode outflows the OLS exposure effect estimates are relatively close to the baseline estimate of 0.042. Beyond that point the estimates fall substantially, with negative point estimates and large standard errors for moves in the top decile of relative outflows. The IV estimates are more stable, but less precisely estimated. The average IV exposure effect estimate is 0.027, an attenuation of 30% relative to the baseline, with a less pronounced fall in point estimates in the top decile of relative outflows. In their (IV) estimates, Chetty and Hendren (2018a) see a similar attenuation of around 20% on average, but if anything less attenuation of point estimates for the top decile.

Figure C.2 provides some comfort that the results are not driven by other factors correlated with *moderately* large relative postcode outflows, but the same cannot be said for the largest outflows. This validation exercise is thus less conclusive in the Australian setting than it appeared in the United States. One explanation is the failure of the identifying assumption — perhaps selection effects do vary with age. That said, the other validation exercises make this explanation more challenging to uphold. A more benign explanation may that the largest relative postcode outflows in Australia tend to be coupled with other factors that mitigate the effects of exposure to the destination. Indeed, fundamental differences in the treatment effects experienced by those choosing to move versus those forced to move are apparent in Chyn (2018).[47] This would be a threat to the external rather than the internal validity of the baseline estimates. For example, I argue in Section 6.1 that the observed pattern of exposure effects can partly be explained by the fact that increased time exposed to the destination increases the probability an individual ends up working there, in the associated local labor market. It is quite plausible that being forced out of an origin due to job loss or natural disaster breaks this relationship between childhood exposure to a place and adult location. This would result in the attenuation or even disappearance of the exposure effects in adult outcomes, depending on the extent to which local labor markets are responsible for such patterns.

[47] While Chyn (2018) finds larger treatment effects for those forced to move, this need not contradict the attenuation apparent in Figure C.2 if, as seems plausible, the appropriate specification of the treatment effect changes alongside its magnitude for exogenous shocks.

Figure C.2: Place exposure effect estimates for progressively larger displacement shocks.



Notes: Estimates of the exposure effects $\gamma_{\tilde{m}}$ and 95% confidence intervals from equation 4 for late childhood exposure, for subsamples of those moving out of postcodes in years with progressively higher relative outflows. These are identified by first calculating, for each postcode p and financial year t, the number of families leaving the postcode divided by the average annual number of families leaving the postcode from 1991 to 2014 (call it k_{pt}). Each individual in the 1-time mover sample is thus associated with a value of k_{pt} that indicates whether they were part of a relatively small $k_{pt} << 1$ or large outflow $k_{pt} >> 1$. The chart estimates the exposure effects for those with values of k_{pt} above its median value, its 55^{th} percentile and so on. OLS estimates are presented, alongside IV estimates where the origin and destination outcomes are instrumented for as described in the text. The IV estimates replicate Figure VI from Chetty and Hendren (2018a).

C.4 Placebo test

A final test shows the outcomes of movers converges to those of permanent residents in a manner that picks up more than just the *persistent* differences in outcomes between the destination and origin. Rather, movers converge to the *cohort-specific* outcomes of permanent residents.

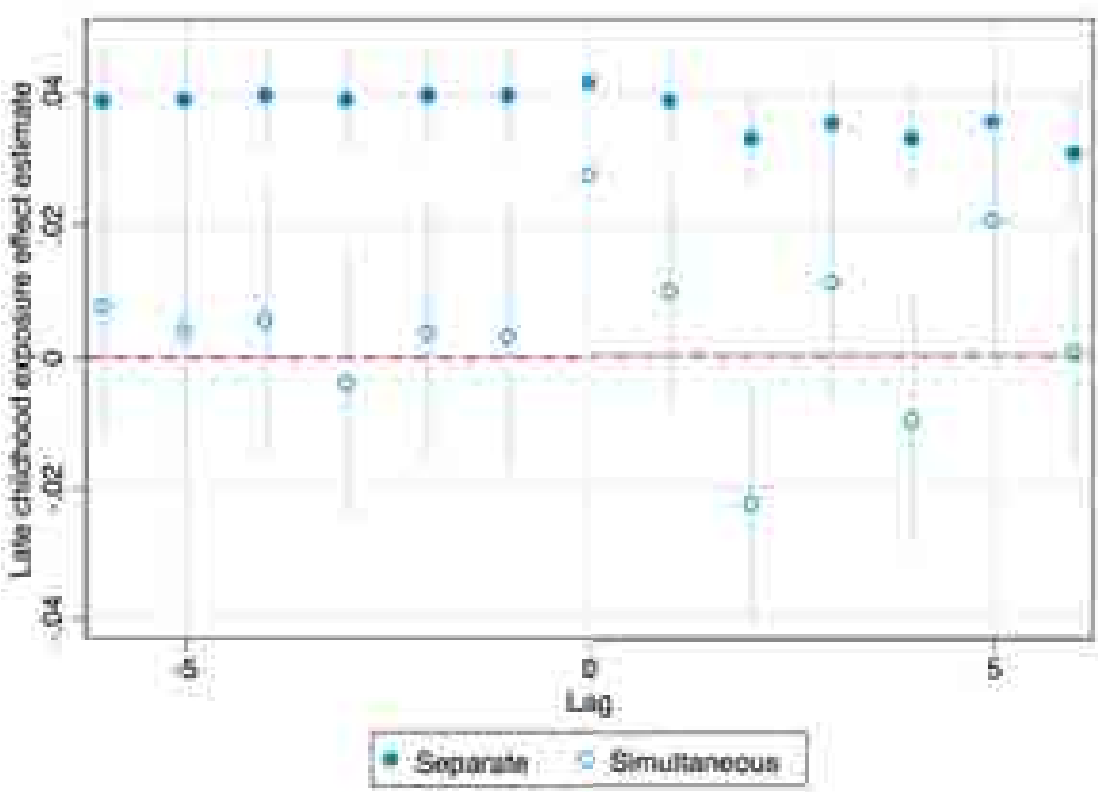
This greatly limits the potential for unobserved factors to explain away the exposure effects. For example, it seems unlikely that unobserved shocks when parents move — such as to income, wealth or family status — are correlated with the as-yet-unobserved *cohort-specific* predicted outcomes for a child. Such shocks seem far more likely to be correlated, if at all, with the persistent features of a place. A fuller and more formal discussion of these issues can be found in Chetty and Hendren (2018a).

First, I show the best predictor of a mover's outcome is based on the experience of movers in their cohort, rather than those of surrounding cohorts. Following Chetty and Hendren (2018a) I run two sets of regressions. In the first thirteen regressions I re-estimate the baseline specification in equation 4 as if an individual's financial year of birth was $s+l$ rather than s, where $l \in \{-6, ..., 6\}$. The resulting late childhood exposure effect estimates γ are in the solid dots in Figure C.3. Reflecting high serial correlation in a location's predicted outcomes, the exposure effects are all around the baseline estimate of 0.04. In the second single regression I re-estimate the baseline specification but include the lags and leads for the origin and difference terms. Where these lags or leads fall outside the sample window, the predicted outcomes are set to zero and an indicator I_l for the absence of that lag or lead is set to one. This gives rise to the specification below:

$$\begin{aligned} y_i = & \sum_{s=1978}^{1991} I(s_i = s)(\alpha_s^1 + \alpha_s^2 \bar{y}_{pos}) + \sum_{m=1}^{30} I(m_i = m)(\zeta_m^1 + \zeta_m^2 p_i) \\ & + \sum_{s=1978}^{1991} I(s_i = s)(\kappa_s \Delta_{odps}) \\ & + \sum_{l \in \{-6,-5,\ldots,5,6\}} \left(\sum_{\tilde{m} \in M} \delta_{\tilde{m}} + \gamma_{\tilde{m}} e_{\tilde{m}} \right) \Delta_{odp,s+l} \\ & + \sum_{l \in \{-6,-5,\ldots,5,6\}} \alpha_l \bar{y}_{po,s+l} + \omega_l I_l + \varepsilon_i \end{aligned} \tag{7}$$

The results are in the hollow dots in Figure C.3, and support a causal interpretation of the exposure effect estimates. The exposure effect estimate for the true cohort is only slightly attenuated. Further, while this estimate is statistically different from zero (with a p-value of 0.0057), the lags and leads are jointly insignificant (with a p-value of 0.20 on the joint test). It follows that any selection process giving rise to the observed exposure effects must do so in a way that is correlated not just with the persistent features of a place, but its cohort-specific features — a far more onerous requirement.

Figure C.3: Place exposure effect estimates: event study.



Notes: Estimates of the exposure effects $\gamma_{\tilde{m}}$ and 95% confidence intervals from equations 4 and 7 for late childhood exposure, when predicted outcomes are derived from a birth cohort that is not necessarily your own (solid dots) or when predicted outcomes for your birth cohort are included alongside those for neighboring cohorts (hollow dots). Thus the solid dots represent coefficients from thirteen separate regressions, using the predicted outcomes for those in financial year of birth cohort $s+l$ rather than an individual's actual birth cohort s, where $l \in \{-6, ..., 6\}$. The hollow dots run a single regression that includes the origin and difference in predicted outcome terms for all neighboring cohorts as in equation 7. Both these specifications allow for cohort effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. This replicates Figure VII from Chetty and Hendren (2018a).

In Chetty and Hendren (2018a) they go further, and show movers outcomes converge to those of permanent residents in distribution and in a gender-specific manner as well. I attempted but failed to replicate the first of these. Whereas Chetty and Hendren (2018a) find that the best predictor of a mover being in the top or bottom decile of the income distribution is the predicted probability of the same outcome, rather than the predicted mean outcome, I found the opposite.[48] However, I feel the most likely explanation of this lies in the lower precision of the Australian predictions

[48]Results available on request. Specifications were as detailed in Chetty and Hendren (2018a).

for permanent residents due to smaller geographic units and reduced geographic variation. The predicted probabilities of making the top or bottom decile are particularly imprecise and thus, if they capture more noise than signal, it is quite plausible that the predicted mean ranks may give a better indicator of the likely distributional outcomes of movers. As was noted by Chetty and Hendren (2018a), an advantage of their data was the ability to conduct validation exercises requiring both large samples and significant variation.

C.5 Summary

The results outlined in this section provide comfort as to the internal validity of the research design introduced in Chetty and Hendren (2018a), both generally and in the Australian setting. Any unobserved factor explaining the observed exposure effects would need to operate within the family in proportion to time exposed and be able to replicate the cohort-specific outcomes of permanent residents. The examination of exogenous moves left a question mark over external validity, but could be consistent with one of the explanations for exposure effects put forward in this paper — that time spent growing up in a location matters largely because it influences where you end up working.

D Generated regressors, precision and valid inference

The equations estimated in this paper (and in Chetty and Hendren (2018a)) fall into the more general class of two-step estimation, where regressors in the model of interest are generated from an auxiliary model. In particular, in the first step, the expected outcomes y for children born into a particular location l, cohort s, and parental household income rank p are predicted based on the sample of permanent residents of that location:

$$y_i = \alpha_{ls} + \beta_{ls} p_i + \varepsilon_i \tag{8}$$

This model provides predicted values for the movers — denoted $\bar{y}_{ops}$ and $\bar{y}_{dps}$ — where we take their location l to be either their origin o or destination d respectively. Let $\hat{\beta}_1$ be the vector of estimated coefficients and X_{1o} and X_{1d} the matrices of observations indicating a mover's origin or destination respectively, along with their cohort and parent rank.

In the second step, these predicted values are used to generate regressors — $\bar{y}_{ops}$ and $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — for inclusion in a model for the outcomes of the movers:

$$\begin{aligned} y_i &= g(x_{2i}, \beta_2, \bar{y}_{dsp}, \bar{y}_{osp}) + \varepsilon_i \\ &= g(x_{2i}, \beta_2, x_{1d}\hat{\beta}_1, x_{1o}\hat{\beta}_1) + \varepsilon_i \end{aligned} \tag{9}$$

This is a classic example of the use of generated regressors. As noted in Pagan (1984), generated regressors pose a number of potential econometric issues. Perhaps most notably, while coefficients estimated from Equation 9 are generally consistent, the standard errors will not be, as they fail to account for uncertainty in the generated regressors. Perhaps reasonably, given they restrict attention to commuting zones with populations over 250,000, where the generated regressors are fairly precisely estimated, Chetty and Hendren (2018a) do not consider this issue. However, given the Australian data is marked by smaller geographies and less geographic variation, this issue seems worth considering in more detail here.

D.1 Valid inference

Murphy and Topel (1985) provide a procedure for calculating asymptotically correct standard errors in the fairly general circumstances. From the presentation in Greene

(2003) the Murphy-Topel estimated covariance matrix for the model, given the two steps are estimated on different samples, is:

$$M = \hat{V}_2 + \hat{V}_2 \hat{C} \hat{V}_1 \hat{C}^T \hat{V}_2 \tag{10}$$

where $\hat{V}_1$ and $\hat{V}_2$ are the estimated covariance matrices for models 1 and 2 respectively and:

$$\hat{C} = \sum_{i=1}^{n} \left(\frac{\partial \ln f_{i2}}{\partial \hat{\beta}_2} \right) \left(\frac{\partial \ln f_{i2}}{\partial \hat{\beta}_1^T} \right) \tag{11}$$

where f_{i1} and f_{i2} are the contributions of observation i to the likelihood functions of models 1 and 2 respectively. Now, we can follow the presentation in Hole (2006) and apply the chain rule to observe that:

$$\begin{aligned}
\frac{\partial \ln f_{i2}}{\partial \hat{\beta}_2} &= \frac{\partial \ln f_{i2}}{\partial (x_{i2}\hat{\beta}_2)} \frac{\partial (x_{i2}\hat{\beta}_2)}{\partial \hat{\beta}_2} \\
&= \frac{\partial \ln f_{i2}}{\partial (x_{i2}\hat{\beta}_2)} x_{i2} \\
&= \frac{\partial \ln f_{i2}}{\partial \hat{y}_{mover}} x_{i2}
\end{aligned}$$

and:

$$\begin{aligned}
\frac{\partial \ln f_{i2}}{\partial \hat{\beta}_1} &= \frac{\partial \ln f_{i2}}{\partial (x_{i1o}\hat{\beta}_1)} \frac{\partial (x_{i1o}\hat{\beta}_1)}{\partial \hat{\beta}_1} + \frac{\partial \ln f_{i2}}{\partial (x_{i1d}\hat{\beta}_1)} \frac{\partial (x_{i1d}\hat{\beta}_1)}{\partial \hat{\beta}_1} \\
&= \frac{\partial \ln f_{i2}}{\partial (x_{i2}\hat{\beta}_2)} \left(\frac{\partial (x_{i2}\hat{\beta}_2)}{\partial (x_{i1o}\hat{\beta}_1)} x_{i1o} + \frac{\partial (x_{i2}\hat{\beta}_2)}{\partial (x_{i1d}\hat{\beta}_1)} x_{i1d} \right) \\
&= \frac{\partial \ln f_{i2}}{\partial \hat{y}_{mover}} \left(\frac{\partial \hat{y}_{mover}}{\partial \bar{y}_{ops}} x_{i1o} + \frac{\partial \hat{y}_{mover}}{\partial \bar{y}_{dps}} x_{i1d} \right)
\end{aligned}$$

In both equations the first term is simply the score vector for model 2 — for simplicity denote its elements s_{i2}. The second equation includes derivatives in the brackets that simply pick up the estimated coefficients on the predicted values. The resulting estimate of $\hat{C}$ is as follows:

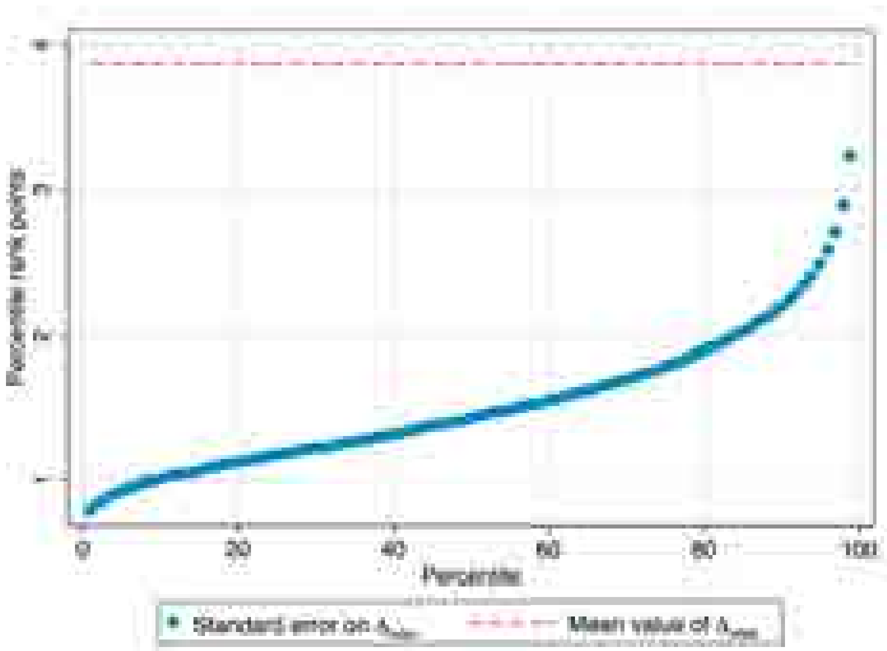
$$\hat{C} = X_2^T \mathrm{Diag} \left\{ s_{i2}^2 \frac{\partial \hat{y}_{mover}}{\partial \bar{y}_{ops}} \right\} X_{1o} + X_2^T \mathrm{Diag} \left\{ s_{i2}^2 \frac{\partial \hat{y}_{mover}}{\partial \bar{y}_{dps}} \right\} X_{1d} \tag{12}$$

The above easily extends to the case where predicted values for neighboring cohorts are also included in the regression. The implementation of these standard errors in STATA has been outlined in Hardin (2002) and simplified in Hole (2006).

D.2 Precision-based sample restrictions

Finally, throughout this paper, analysis is restricted in to those for whom the difference in predicted outcomes Δ_{odps} is more precisely estimated. The distribution of the standard error in Δ_{odps} for the 1-time movers sample is shown in Figure D.1. For most of the analysis, I require $\Delta_{odps} < 2$, thus restricting attention to around the 80% of the sample for whom Δ_{odps} is most precisely estimated.

Figure D.1: Distribution of standard error in difference in predicted outcomes for permanent residents of the destination and the origin



Notes: For the 1-time mover sample, shows the distribution of the estimated standard errors on the key generated regressor: Δ_{odps}. Also shows the mean value of this regressor.

Key findings are robust to this precision-based sample restriction. In Table D.1, exposure effect estimates are shown for the baseline case, and for increasing levels of precision in Δ_{odps}. The results are not particularly sensitive to the choice of the precision-based sample restriction, with the late childhood exposure effect estimates all close to the baseline estimate of 0.042 and always larger than the early childhood exposure effect estimate.

Table D.1: Exposure effect estimates: varying levels of precision in Δ_{odps}

	Baseline	Increasing levels of precision				
	(1)	(2)	(3)	(4)	(5)	(6)
Early	0.011	0.006	0.008	0.011	0.030	0.010
	(0.007)	(0.006)	(0.006)	(0.007)	(0.009)	(0.023)
Late	0.042	0.039	0.040	0.042	0.040	0.046
	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)	(0.013)
Post-outcome	0.008	-0.001	0.001	0.008	0.015	0.052
	(0.013)	(0.021)	(0.012)	(0.013)	(0.018)	(0.056)
Sample restrictions						
s.e. on Δ_{odps}	< 2	none	< 2.5	< 2	< 1.5	< 1
N	264,500	312,900	297,800	264,500	176,300	30,200

Notes: Estimates of the exposure effects $\gamma_{\bar{m}}$ from equation 4 for early ($m \in \{2, ..., 11\}$), late ($m \in \{12, ..., 24\}$) or post-outcome ($m \in \{25, ..., 34\}$) exposure for the full baseline sample and larger or smaller samples based on varying restrictions on the standard error on Δ_{odps}. These represent the expected boost to an individual's household income rank at age 24 associated with an additional year at this stage of life in a destination with 1 percentile rank higher expected outcomes for permanent residents. They are estimated by regressing the adult ranks y_i of those whose parents move once in their childhood on the interaction of their time exposed to the destination at each life stage with $\Delta_{odps} = \bar{y}_{dps} - \bar{y}_{ops}$ — the difference between the expected outcomes for permanent residents of the same parent percentile rank and cohort in the destination versus the origin. the difference between the expected outcomes for permanent residents of the same parent percentile rank p and cohort s in the destination d versus the origin o. Controls capture: cohort and origin effects (via indicators for cohort and their interactions with predicted outcomes for permanent residents of the origin); disruption effects (via indicators for age at move and their interaction with parental rank); and indicators for cohort interacted with Δ_{odps} to capture potential mis-measurement of the origin. Murphy-Topel standard errors are in parentheses.

E Intergenerational data construction

This Appendix describes the creation of the Australian Taxation Office's (ATO) de-identified intergenerational dataset. It is based on information provided by the ATO and those involved in the construction of the dataset.

E.1 Overview

The dataset begins with the universe of federal tax returns from the 1991 to 2015 financial years, linked across individuals. This provides comprehensive information on individual incomes — the key challenge is linking parents and children.

Australia does not have two sources of parent-child links commonly used internationally. Birth register information is held by state and territories, and there is no national register as there is for Nordic countries. Further, parents are generally not required to provide identifying information for their children on tax returns, as family benefits are administered separately as cash transfers.[49] This rules out the methodology underlying Chetty et al. (2014), which uses the fact that parents' tax returns in the United States report their children's social security numbers.

Instead, parent-child links were formed by matching individuals to parents based on their reported residential addresses. Individuals report a residential address when they register for a tax file number — a unique personal identifier that is the closest Australian analogue to a social security number. The vast majority of individuals do this before they turn 17. These individuals are then linked to their likely parents based on residential addresses reported in tax returns. These links are disciplined by a set of more direct links available for a subset of individuals.

Address matching is behind the Statistics Canada dataset used in numerous widely-cited studies of intergenerational mobility (e.g. Corak and Heisz (1999); Oreopoulos (2003); Corak and Piraino (2011)). Yet the Australian institutional background, described below, means the ATO intergenerational dataset delivers a much higher match rate. Corak and Heisz (1999) report that they have parent links for around 49% of their selected Canadian cohorts. For the Australian cohorts studied in this paper, the link rate is around 92%, in line with that achieved by Chetty et al.

[49]Linking tax returns to this separate administrative database would have failed to provide complete parent-child links, as cash transfers have been and remain highly targeted, rather than universal.

(2014).

E.2 Institutional background

Address matching delivers high quality parent-child links in Australian tax data because most individuals register for a tax file number (TFN) with the ATO while still young and living in the family home. This reflects strong incentives to do so. Since its introduction in 1989, or shortly afterwards, a TFN has been needed to:

- avoid paying higher withholding tax rates on labor and capital income;
- apply for unemployment, disability or family benefits; and
- apply for concessional loans for higher education.

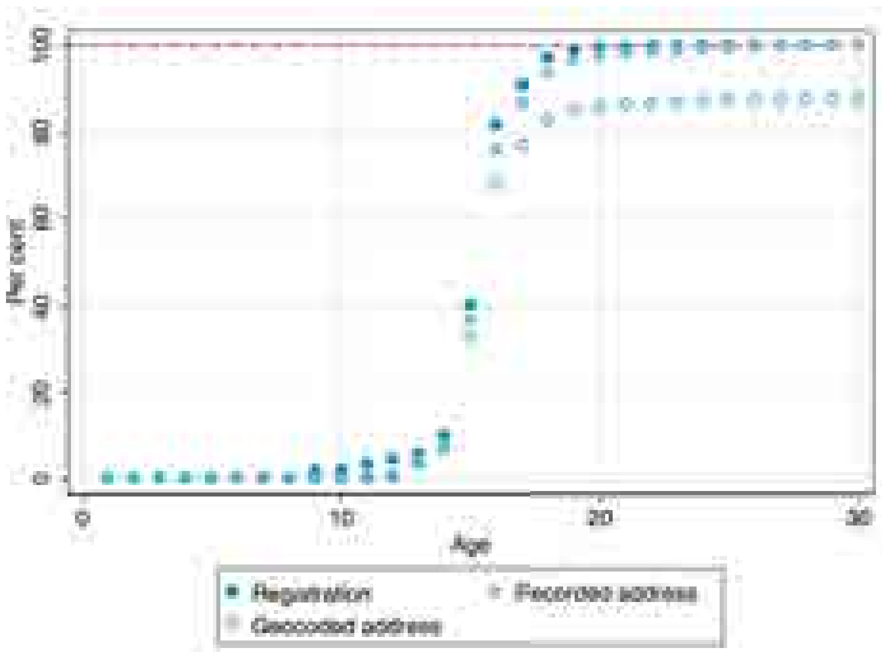
As a result most transitions from childhood to independence — be it work, welfare or higher education — reward or require registering for a TFN. For example, of those born in Australia in the 1980 financial year and with a TFN by the time they were 30, over 90% had registered by age 17, and over 99% by age 20 (see Figure E.1). Importantly, a residential address is captured for most of these children at the point of registration, and is typically of sufficient quality to match to a geocoded address.

E.3 Family linking procedure

The ATO dataset focuses on those born between the 1970 and 2000 income years (inclusive). Those born earlier are difficult to link to parents as many will have left the family home before the tax return panel begins in 1991. Similarly, many of those born later were yet to register for a TFN at the time the dataset was constructed.

Family links were generated for all individuals in the relevant birth cohorts, whether or not they were born in Australia. However, for the file used in this research, attention was restricted to those born in Australia. Country of birth is not directly observed in the tax data, but a good proxy for those born in Australia was derived based on other administrative information. From the 1978 birth cohort onwards this proxy performs particularly well, with the resulting Australian-born annual birth cohorts deviating by at most 1.5% from the population benchmark for the 1978-1991 cohorts.

Figure E.1: Proportion registered for a TFN by age (Australia-born, 1980 birth cohort)



Notes: Darker blue dots show the proportion of registered clients born in Australia in the 1980 financial year who had registered by the given age. The lighter blue and hollow dots show the respective proportions with a recorded address and an address matched to a geocoded address by that age.

E.4 Family Tax Assistance links

Between the 1997 and 2000 income years, Family Tax Assistance (FTA) allowed low-to-middle income families to claim a higher effective tax free threshold. Low income families could claim the entire benefit through the payments system. However, middle income families had to provide the given names and dates of birth of their children on their tax returns. This provides a relatively direct source of family links for a subset of the child population.[50] These direct links then informed the algorithm for generating family links from the more widely available address links.

Initially, the details of all children a parent claimed between 1997 and 2000 were collected. This included a child's first name, date of birth and potential last names — while a child's actual last name is not listed, potential last names as inferred from those of their claiming parent and that parent's spouse. Duplicate claims were

[50]FTA claims do not necessarily imply a biological parent-child relationship, though in most cases the claimant will be a biological parent or primary carer.

dropped and the remaining claims formed a base population of FTA children. FTA children were then linked to their adult selves among individuals registered for a TFN. A sequence of matches was performed, with only unmatched children passed to the next stage:

- *Perfect matches*: the first name, last name and date of birth match a unique individual;
- *First name error*: the last name and date of birth match a unique individual, where the two first names to have a levenshtein string edit distance of at most two;
- *DOB error*: the first name and last name match a unique individual, where the two years of birth are the same;
- *First name and DOB error*: the last name matches a unique individual, where the two first names have a levenshtein string edit distance of at most two and the two years of birth are the same;
- *Last name error*: the first name and date of birth match a unique individual.

Well over 70% of claimed children in each year were perfectly matched to an adult client. Fuzzy but unique matches allowed over 85% of claimed children in each year to be matched. These matches are spread across the birth cohorts of interest, with large numbers of those born from the 1980s onwards having FTA links.

E.5 Address links

As FTA could only provide family links for a selective subset of the child population, the primary source of family links was based on shared residential addresses. As a first step, children were linked to all individuals who had ever lived at an address the child had lived at.[51] This forms the set of potential siblings and parents.

E.5.1 Siblings

First, individuals were linked as siblings if:

[51] The address was not required to be concurrent, as address histories in the tax data have gaps, and non-concurrent shared addresses in tax data may have been concurrent in reality.

- they had been at the same address within five years of one another;
- they both lived at that address before they turned 20;
- they had less than a 13 year age gap; and
- they had the same earliest last name.

These links were 'filled' out to ensure transitivity.[52] At the end of the parent linking process individuals were also linked as siblings if they shared the same parents.

E.5.2 Parents

Individuals were then linked to parents. First, potential parents who were particularly young at the birth of the child (under 15 years of age) or old (45 years of age for women, 55 years of age for men) are dropped. Then the subsample of children who were *perfectly* matched as FTA children and have parent links as a result was isolated. A logistic regression was run on this subsample on the outcome that a potential parent is an FTA parent. The independent variables used in this regression were:

- potential parent sex interacted with an indicator for whether the potential parent and child share a last name;
- potential parent sex interacted with a quartic in parental age at birth either side of the median age at birth for that sex (29 for men, 27 for women)[53];
- an indicator for whether the potential parent and child address histories imply they were at the same address at the same time;
- the length of overlap (in years) of a concurrent address episode;
- the distance (in years) separating a non-concurrent address episode; and
- the age of the child when first at the address (categorical between 13 and 25), interacted with whether the address episode was concurrent or not.

[52] That is, if Alice is Bob's sibling and Bob is Charlie's sibling then Alice is Charlie's sibling. As a result children with more than a 13 year age gap may be identified as siblings if, for example, they share a sibling in common.

[53] Over the period 1975-1990, as calculated from Australian Bureau of Statistics (2017b).

In the final step children were linked to their most probable parent, based on the logistic model's out-of-sample predictions for the probability a potential parent was an FTA parent. Each child was linked to the potential parent with the highest predicted probability of being an FTA parent, conditional on that probability being greater than 0.5. At the chosen threshold, less than 4% of the address-derived parents for the FTA subsample failed to match the FTA parent. Given the FTA links are not infallible, this seems reasonable. FTA link parents were then used for those children with no parent.

E.5.3 Postcode links

Address matching is limited by the absence of complete address histories. Further back in the panel tax filers are less likely to have a recorded residential address. However, residential postcodes are reliably recorded — they are captured for the vast majority of tax filers in each of the years between 1991 and 2015. To exploit this, a set of supplementary links is based on residential postcode histories.

For all children, the postcode of their first address was extracted, typically their address when registering for a TFN. Children were also assigned their earliest recorded last name. Children were then linked to all individuals in the same postcode in the same year and with the same last name. This formed the set of potential parents. As in the address matching, potential parents who are particularly young at the birth of the child (under 15 years of age) or old (45 years of age for women, 55 years of age for men) were dropped. Given the large number of potential postcodes and last names, most children end up with a relatively small set of potential parents.

In the next step the subsample of children who perfectly matched FTA children and have parent links as a result is again isolated. Once again, a logistic regression was run on the binary outcome that a potential parent is an FTA parent. The independent variables used in this regression were:

- potential parent sex interacted with quartics in parental age at birth either side of the median age at birth for that sex (29 for men, 27 for women)[54];
- an indicator for if the potential parent's spouse is among the alternatives, also interacted with a categorical variable for the number of potential parents for the child (top coded at ten);

[54] Over the period 1975-1990, as calculated from Australian Bureau of Statistics (2017b).

- the age of the child when first at the postcode (categorical between 13 and 25).

In the final step children were linked to their most probable parent in the same manner as for the address matching. The exception is that here a slightly more conservative threshold was set — children were only linked to parents if the estimated probability of the potential parent being an FTA parent was greater than 0.75. The accuracy of this algorithm was only a little worse than the address matching. At the chosen threshold, only 8% of the supplementary parents for the FTA subsample fail to match the FTA parent.

E.5.4 Sibling links

In this step, individuals with siblings were linked to the most probable parent for their family. The steps were as follows:

- look across groups of siblings (families);
- identify the most probable parent for each family — that is, the potential parent with the highest estimated probability of being a true parent[55]; and
- match individuals to the resulting most probable parent for their family.

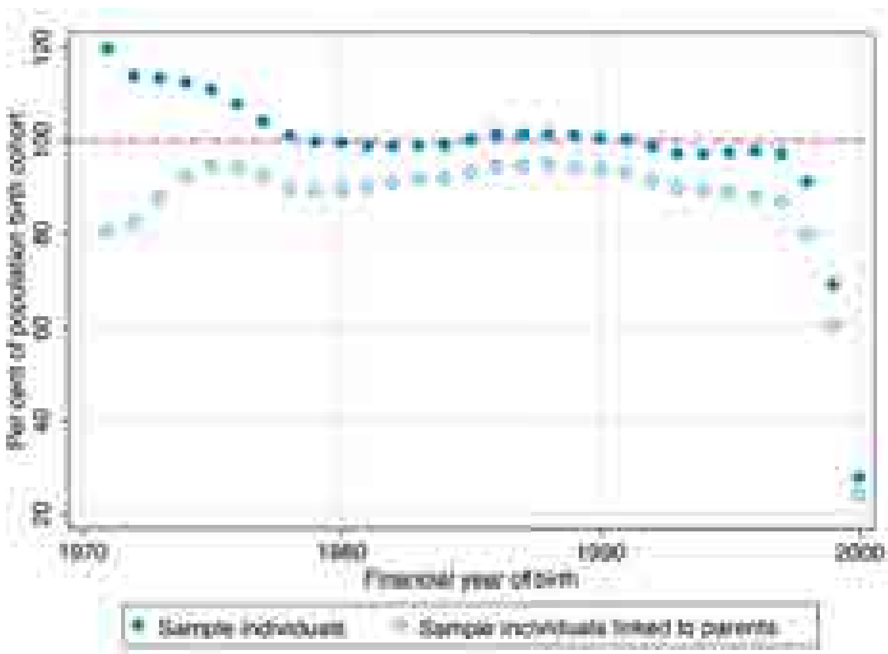
Reassuringly, this process showed a great deal of consistency in the parent-child links. For children already linked to a parent, that parent is not replaced, or replaced by their spouse, in 90% of cases. Once children are matched to their most probable parent, those parents are matched to their earliest reported spouse over the period 1991 to 2015.

E.5.5 Resulting parent-child links

The parent-child links resulting from this process are shown in Figure E.2. From 1978 to 1991 onwards the sample closely matches the size of the Australian-born population, deviating from the population benchmark by at most 1.5%. The proportion of the population linked to parents averages around 92% over this period as well. For the birth cohorts examined in this paper, 88% of links are derived from shared residential address, 4% from FTA, 2% from postcodes and 6% from siblings.

[55]FTA parents were assigned a probability of 1 and supplementary parents were all assigned to 0.75.

Figure E.2: Sample coverage rates relative to the population of interest (%)



Notes: Shows the number of individuals in the sample, and the number linked to parents, as a percentage of the relevant population of interest. The population of interest is taken as the number of births in Australia (Australian Bureau of Statistics (2017b)), or for financial years prior to 1976 (where this data is not available) the estimated resident population aged zero on the last day (30 June) of the relevant financial year (Australian Bureau of Statistics (2017a)). Where both series are available the deviate by at most 2%.