

Morris, Todd

Article — Published Version

Re-examining female labor supply responses to the 1994 Australian pension reform

Review of Economics of the Household

Provided in Cooperation with:

Springer Nature

Suggested Citation: Morris, Todd (2021) : Re-examining female labor supply responses to the 1994 Australian pension reform, Review of Economics of the Household, ISSN 1573-7152, Springer US, New York, NY, Vol. 20, Iss. 2, pp. 419-445,
<https://doi.org/10.1007/s11150-021-09561-y>

This Version is available at:

<https://hdl.handle.net/10419/286978>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Re-examining female labor supply responses to the 1994 Australian pension reform

Todd Morris ¹

Received: 28 March 2020 / Accepted: 16 March 2021 / Published online: 15 April 2021
© The Author(s) 2021; This article is published with open access

Abstract

Many governments are aiming to extend working lives by raising the age at which people can claim retirement pensions. This makes it vital to understand how these policies affect retirement decisions. In this paper, I revisit the labor supply effects of a major Australian reform that increased women’s pension age from 60 to 65. Atalay and Barrett (2015) studied these effects using repeated household surveys and a differences-in-differences design in which male cohorts form the comparison group. They estimate that the reform increased female labor force participation by 12 percentage points. Using earlier data, I show that the parallel-trends assumption did not hold before the reform because of a strong female-specific trend in participation rates across the relevant cohorts. Accounting for this trend, the estimated effect on female participation falls by two-thirds and becomes statistically insignificant at conventional levels. This highlights the importance of carefully assessing and controlling for trends across cohorts when evaluating pension reforms, which are typically phased in across cohorts.

Keywords Retirement age · Labor supply · Cohort effects

JEL classification H55 · J26 · I38

An earlier version of this paper circulated with the title “Large response to delayed eligibility or a pre-existing trend in female participation? Re-examining an Australian pension reform”.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s11150-021-09561-y>.

✉ Todd Morris
toddstuartmorris@gmail.com

¹ Munich Center for the Economics of Aging, Max Planck Institute for Social Law and Social Policy, Amalienstrasse 33, 80799 Munich, Germany

1 Introduction

In response to aging populations, many governments are reducing the generosity of retirement pensions for younger cohorts through increases in eligibility ages or reductions in benefit amounts. A key aim of these reforms is to encourage older workers to remain in the labor force, making it vital to understand how these policies affect retirement decisions. While several studies find large effects of these reforms on retirement decisions (e.g., Börsch-Supan & Schnabel, 1998; Cribb et al., 2016; Geyer & Welteke, 2021; Hanel & Riphahn, 2012; Lalive et al., 2017; Snyder & Evans, 2006; Staubli & Zweimüller, 2013; Vestad, 2013), it has been difficult to disentangle the key mechanisms as these reforms may affect retirement behavior through several channels. For example, these reforms may encourage people to delay retirement due to (i) a decrease in individuals' lifetime pension wealth, known as the wealth effect, (ii) the opportunity to work for a longer period and thus accrue a larger pension, known as the accrual effect, (iii) a change in social norms, (iv) the presence of liquidity constraints, or (v) by delaying any income/earnings tests that only apply when receiving a pension, which make it unattractive for people to continue working after they become eligible for retirement benefits.

An important paper in this literature is Atalay and Barrett (2015). Atalay and Barrett (2015) study an Australian reform in 1994 that gradually raised women's eligibility age for the means-tested public retirement pension from 60 to 65 based on their birthdates (see Fig. 1). Atalay and Barrett (2015) argue that this reform is useful for understanding mechanisms because Australia's retirement pension is non-contributory. This means that there is no accrual effect — an effect which has been emphasized in the literature (Gruber & Wise, 2004; Samwick, 1998) — and any impact on retirement decisions must result from other mechanisms. Atalay and Barrett (2015) estimate the impact of

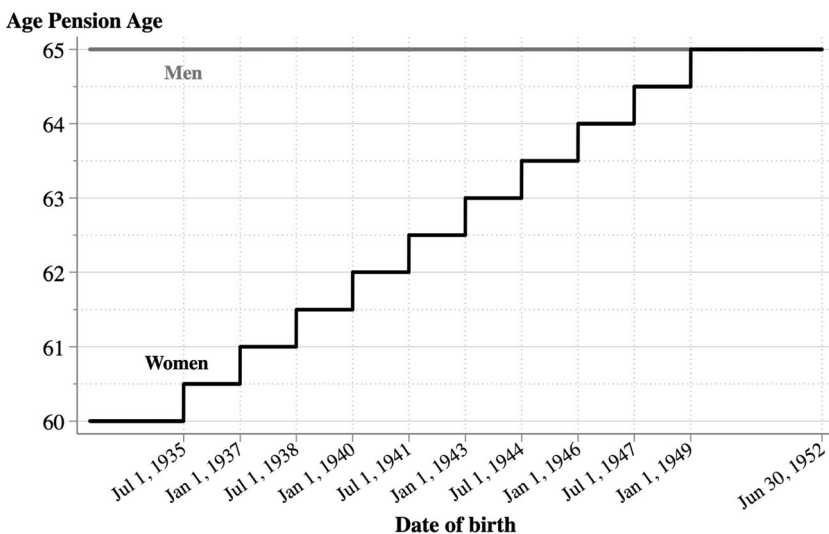


Fig. 1 Age pension age (APA) by date of birth and gender. Notes: This figure shows the eligibility age of Australian men and women for the Age Pension based on their date of birth. The phased increases in women's Age Pension Age (APA) from 60 to 65 were legislated in 1994.

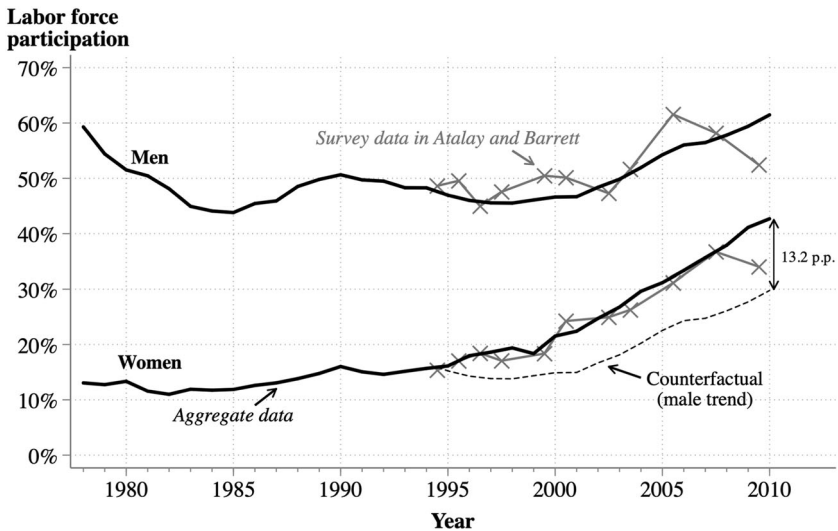


Fig. 2 Labor force participation rates of men and women aged 60–64 over time. Notes: This figure replicates Fig. 1 in Atalay and Barrett (2015). The black lines plot aggregate time series data from the Labor Force Survey, while the gray lines show the calculations based on the Income and Housing Costs Surveys (with the X indicating each survey). I also include a dashed line denoting the counterfactual change in female labor force participation (LFP) over Atalay and Barrett’s (2015) sample period (1994–2010); this line shows how female LFP would have changed if it followed the same trend as male LFP over the sample period. Evidently, female LFP increased considerably more than male LFP over this period.

the reform on female labor force participation at ages 60 to 64 using repeated household surveys ranging from 1994 to 2010 and a differences-in-differences design in which male cohorts form the comparison group. Their headline estimate is large, reflecting the stronger increase in female participation rates over this period (see Fig. 2). They estimate that each one-year increase in women’s eligibility age increased female labor force participation by 12 percentage points, an effect they interpret as evidence of strong wealth effects.¹

In this paper, I start by replicating Atalay and Barrett’s (2015) labor force participation estimates. Using the same source data, I apply the sample restrictions and methods described in Atalay and Barrett’s (2015) paper and find that their estimates can be successfully replicated. Namely, due to the clarity of Atalay and Barrett’s (2015) paper, I am able to closely match the sample and regression estimates without needing to use original program files. Then, I extend the analysis with earlier data, building on a literature that is increasingly scrutinizing difference-in-difference methods (e.g., Bertrand et al., 2004; Borjas, 2017; Cameron et al., 2008; Donald & Lang, 2007; Goodman-Bacon, 2018; Wolfers, 2006). In the Australian context, one concern with using men as the comparison group is that, as in many countries, the participation rate for women has risen considerably over recent decades, while the rate for men has been relatively stable. With these gender-specific time trends in

¹ Atalay and Barrett (2015) also estimate significant spillovers on other government programs, such as the Disability Support Pension.

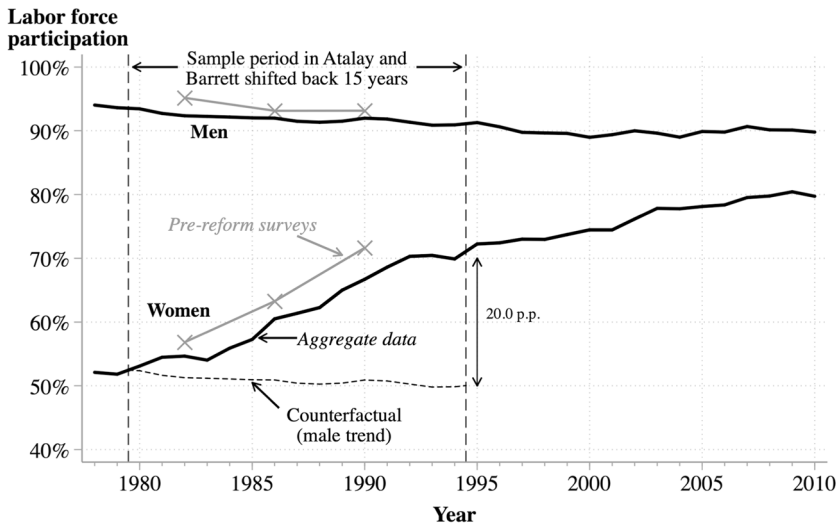
mind, later cohorts of women may have higher participation rates for reasons other than their higher pension age. If this is the case, Atalay and Barrett's (2015) estimates may be biased upwards. As Atalay and Barrett's (2015) sample period started after the reform was announced, they were unable to use their data to examine whether the parallel-trends assumption held before the reform for the cohorts in their sample.

I extend the analysis using two sources of data from before the reform was announced. First, I use aggregate data on participation rates by age and gender to examine whether the parallel-trends assumption held before the reform for the cohorts in Atalay and Barrett's (2015) sample. This builds on Fig. 1 in Atalay and Barrett (2015), which showed similar pre-reform trends in male and female participation rates near the retirement age among earlier cohorts. For the relevant cohorts, I find opposing trends in male and female participation rates before the reform (at younger ages), with a strong upward trend in female participation rates and a weak downward trend in male participation rates (see Fig. 3). For example, at ages 45 to 49, women's participation rate increased by 18.6 percentage points in the fifteen years before Atalay and Barrett's (2015) sample period (from 52.5% to 71.1%), while men's participation rate decreased by 2.4 percentage points (from 93.5% to 91.1%). Moreover, I find similar gender-specific trends among these cohorts at both older and younger ages, suggesting that women would have reduced the gender gap in participation rates at ages 60 to 64 over Atalay and Barrett's (2015) sample period even if the reform had not occurred. This indicates that Atalay and Barrett's (2015) estimates, which attribute the convergence in male and female participation rates to the reform, may be biased upwards.

Second, I use repeated household surveys from 1982 to 1990 to examine whether the different trends of men and women can be explained by the control variables in Atalay and Barrett's (2015) regressions. Specifically, I estimate placebo regressions using Atalay and Barrett's (2015) differences-in-differences approach to test for an impact on the participation rates of female cohorts before the reform was announced. In these regressions, positive estimates of the (placebo) treatment effect would indicate that Atalay and Barrett's (2015) estimates are likely to be biased upwards. I find that the estimates are positive, highly statistically significant, and broadly consistent with the size of Atalay and Barrett's (2015) estimates in percentage terms. Thus, Atalay and Barrett's (2015) estimates are likely to confound women's responses to the reform with a persistent cohort trend in female participation rates.

In the final part of the paper, I use the replication sample to re-estimate the effects of the reform on female labor force participation. I follow the empirical strategy of several recent studies examining phased increases in pension-eligibility ages (Cribb et al., 2016; Oguzoglu et al., 2020; Staubli & Zweimüller, 2013). This approach differs from Atalay and Barrett's (2015) approach by directly controlling for the trend in female participation rates across cohorts. These controls may partially or even fully absorb any wealth effects from the reform but not effects from other mechanisms, such as a change in social norms, liquidity constraints or the pension's income test. I find that the estimated increase in female labor force participation falls by two-thirds to 4.1 percentage points and becomes statistically indistinguishable from zero at the 10% level (with a p value of 0.109). This estimate provides only weak evidence that the reform had any effect on female labor force participation. This modest impact is consistent with the raw data, which shows little evidence of a

(a) Ages 45–49



(b) Ages 50–54

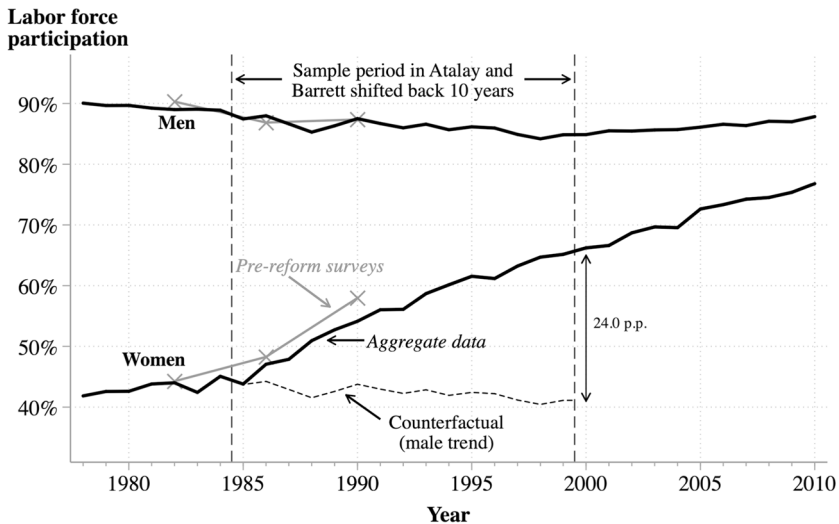


Fig. 3 Labor force participation rates of the relevant cohorts at younger ages. Notes: These figures plot labor force participation (LFP) rates from the Labor Force Survey (in black) and the Income Distribution Surveys (in gray). Between the vertical dashed lines, the cohorts in Atalay and Barrett's (2015) sample were aged 45–49 in (a) and 50–54 in (b). During these periods, male and female LFP rates follow opposing trends. Thus, the parallel-trends assumption did not hold for the relevant cohorts before the reform.

drop in female participation rates at the eligibility age (and only a small decrease in male participation rates at their constant eligibility age of 65). Moreover, the estimated effect is similar in magnitude to (unpublished) estimates of the reform by

Morris (2019) based on more detailed survey data.² These impacts are not only smaller than Atalay and Barrett's (2015) estimates but also, as discussed below, smaller than estimates of similar reforms in other countries.³

More generally, the results in this paper have implications for studies estimating the impact of retirement reforms. Over recent decades, female labor force participation rates have increased considerably across the developed world; from 1970 to 2016, the participation rate of women aged 25 to 64 in OECD countries increased from 46.3% to 68.1% (OECD, 2018). In the context of Australia, my results show that these trends can result in meaningful differences across cohorts in female participation rates near the retirement age. Even when these differences appear to reflect logical responses to a decrease in the generosity of retirement benefits for later cohorts — and even when differences across earlier cohorts are minimal near the retirement age — it is important, where possible, to examine trends across the relevant cohorts in earlier time periods. This recommendation applies generally and is not specific to studies focused on the effects on female labor supply; studies focused on other potentially correlated outcomes like welfare receipt or health should also be careful in attributing cohort differences in outcomes to the effects of pension reforms.

2 Background and replication of Atalay and Barrett (2015)

2.1 Institutional background

Australia's public retirement pension, the Age Pension, is received by around 70% of older Australians. Unlike in many other countries, employment history does not affect pension eligibility or payment rates. Instead, there is a means test based on the current income and assets of the household; from a maximum level, payments are gradually reduced to zero for households with higher income and assets. In practice, around two-thirds of pension recipients receive the maximum payment (Barrett & Tseng, 2008). In 2008, near the end of Atalay and Barrett's (2015) sample period, this payment was \$546.80 every two weeks for singles (~\$14,300 per annum) and \$913.60 for couples combined (~\$23,800 per annum).⁴ For perspective, the median equivalized household income in Australia in 2008 was \$692 per week. Thus, a single household receiving the maximum payment and no other income would have income equal to 40% of the median household income in the Australian population (44% for couples).⁵

² Using longitudinal survey data from 2001 to 2015, Morris (2019) estimates a 3.1-percentage-point increase in female labor force participation, an effect that is statistically significant at the 10% level. These estimates correspond to the average impact of the reform on a slightly different set of cohorts at slightly different ages.

³ Recent studies examining the impact of retirement reforms on female labor supply in other countries include the following: Hanel and Riphahn (2012), Staubli and Zweimüller (2013), Vestad (2013), Cribb et al. (2016), Lalive et al. (2017), and Geyer and Welteke (2021).

⁴ The average exchange rate in 2008 was 1 AUD = 0.85 USD (ranging 0.60–0.98 USD). Thus, at the average exchange rate, the maximum payment was equivalent to US\$12,155 per annum for singles and US\$20,230 for couples.

⁵ For this calculation (and other measures of household income in the paper), I adjust for household size using an equivalence scale of 1.5 for couples.

Households are eligible for the maximum payment if their assessable income and assets are below the respective income and asset disregards.⁶ Otherwise, payments are determined by either the income test or the assets test, whichever implies a lower payment. In 2008, each dollar of assessable income above the income disregard reduced Age Pension payments by 40 cents (20 cents each for couples) and each \$1000 of assessable assets above the asset disregard reduced payments by \$1.50. In practice, the income test binds for many more pensioners than the assets test (Chomik et al., 2018).

To be eligible for the Age Pension, individuals must reach the Age Pension Age (APA). The APA is most similar to an Early Retirement Age (ERA); it is the first age individuals can claim the Age Pension. However, there is a stronger incentive to start claiming at the APA because payment rates do not depend on the age at which individuals start claiming. Prior to the Social Security Legislation Amendment Act (No. 2) 1994, the APA was 60 for women and 65 for men. This reform, proposed in 1993, legislated that women's APA would increase to 65. As shown in Fig. 1, the increases were phased in based on women's date of birth. For example, the APA increased to 60.5 for women born between July 1935 and December 1936, and it increased to 61 for women born between January 1937 and June 1938. The APA continued increasing in this manner until it reached 65 — the APA of men — for women born after December 1948.

Other sources of retirement funding exist besides the Age Pension. For example, many Australians receive other government payments before they reach the APA. Oguzoglu et al. (2020) show that, depending on the cohort, around 30–45% of women receive a regular payment before they reach the APA, with the most common payment being the Disability Support Pension. Like the Age Pension, other payments are non-contributory and means tested, and many have the same payment rates. These payments have additional eligibility criteria but no age conditions. Typically, recipients of other payments are transferred onto the Age Pension at the APA (Oguzoglu et al., 2020). Australians can also fund their retirement privately through voluntary savings and a mandatory scheme called superannuation, in which employers make contributions to their employees' private retirement accounts.⁷ However, the Age Pension is the main source of retirement income for older women.⁸

2.2 Sample

Atalay and Barrett's (2015) sample comes from 11 cross-sections of the nationally representative Australian Bureau of Statistics Income and Housing Costs Surveys

⁶ Most forms of income are assessed. In 2008, the income disregard was \$138 for singles and \$240 for couples. The assets test excludes the value of the primary residence but assesses most other assets. Because the exclusion of the primary residence favors homeowners, their asset disregard is set at lower level. In 2008, the asset disregard for homeowners was \$171,750 for singles and \$243,500 for couples. For renters, the respective disregards were around \$120,000 higher.

⁷ The access age for superannuation was 55 throughout the sample period.

⁸ Most older women have little saved in superannuation. In 2013/14, the median woman aged 60–64 had just \$28,000 of superannuation wealth (Clare, 2015). Historically, older women would have had even less, since superannuation contributions only became mandatory in 1992. Moreover, superannuation wealth is likely to be highest among wealthier women who are not eligible for the Age Pension.

(IHCS). The first survey was in 1994/95 and the last in 2009/10, with surveys conducted over the course of a financial year and a gap of at most one year between surveys.⁹ Atalay and Barrett (2015) restrict their sample to men and women aged 60–64. They also exclude immigrants who have lived in Australia for less than ten years, who would not satisfy the residency condition for the Age Pension. Using the same source data and the information in Atalay and Barrett's (2015) paper, I am able to closely replicate their sample. Appendix Table B1 shows that the replication sample is similar to Atalay and Barrett's (2015) sample in terms of the number of observations and the means of the key variables.¹⁰ Overall, women in the sample have low rates of labor force participation (LFP), with just 27% of women participating compared to 52% of men. A relatively high proportion of women in the sample receive a government payment (56%),¹¹ with 39% of women receiving a payment before they are eligible for the Age Pension. Due to the means-testing of payments, payment recipients are typically in lower income households. In 2008 dollars, female payment recipients have average equivalized household income of \$441 (including pension income),¹² compared to \$860 for non-recipients.

2.3 Mechanisms affecting the impact on female labor force participation

Before describing Atalay and Barrett's (2015) estimates, it is worth discussing the relevant factors that may affect the size of the reform's impact on female LFP. The reform may increase female LFP at ages 60–64 for two main reasons. First, it may be optimal for some women to permanently leave the labor force (i.e., retire) exactly at the APA. This optimality may arise due to (i) the income test for the Age Pension, which results in an increase in the effective marginal tax rate on earnings of up to 40 percentage points at the APA, or (ii) liquidity constraints, since individuals cannot borrow against their future Age Pension wealth. The APA may also influence social norms by providing a signal as an appropriate age to retire (Cribb et al., 2016). Second, the reform reduces the present value of women's expected lifetime income from the Age Pension. This negative wealth effect is the key mechanism emphasized

⁹ Between 1994/95 and 2009/10, the IHCS was conducted in 11 out of 16 years. It was not conducted in 1998/99, 2001/02, 2004/05, 2006/07 and 2008/09.

¹⁰ One point where there may be minor differences is the exclusion of immigrants who arrived in Australia in the last ten years. There is some discretion in implementing this restriction as the variable for 'year of arrival in Australia' is often only available in five-year windows (e.g., 1985–89). To be conservative, I exclude anyone with a non-zero probability of arriving in Australia in the last ten years. Nonetheless, I still have a slightly different number of observations (88 more respondents, or 0.7% of the sample) than in Atalay and Barrett (2015).

¹¹ This proportion is slightly higher than reported in Atalay and Barrett (2015), which appears to be due to differences in the payments included in this definition (which are not listed in Atalay and Barrett (2015)). I include recipients of the following payments: the Age Pension (34.9% of women), the Disability Support Pension (7.8%), Partner Allowance (3.4%), Widow Allowance (2.4%), Carer Payment (2.1%), Newstart Allowance (1.8%), Service Pension (1.1%), Mature Age Allowance (0.7%), Veteran's Disability Pension (0.6%), War Widow's Pension (0.5%), Wife Pension (0.3%), Parenting Payment (0.05%), Sickness Allowance (0.03%), Special Benefit (0.02%), and 'other government cash pensions and allowances' (0.15%).

¹² Excluding women's income from government payments, payment recipients have equivalized household income of \$341.

by Atalay and Barrett (2015). Atalay and Barrett (2015) construct a model that indicates that women's optimal retirement age will increase as long as leisure is a normal good. The model implies that the reform will affect retirement decisions at all ages, including below the old APA and above the new APA. However, studies of comparable reforms indicate that most of the impact is likely to result from a shift in retirement decisions at ages in which pension eligibility is affected (Cribb et al., 2016; Geyer & Welteke, 2021; Staubli & Zweimüller, 2013).¹³

There are also several factors that may attenuate the impact of the reform on female LFP. First, women in these cohorts had very low rates of LFP, with many women having left (or never entered) the labor force before age 60. Second, other government payments provided a substitute form of retirement income for many women until they reached the new APA (Atalay & Barrett, 2015; Oguzoglu et al., 2020).¹⁴ Third, the substitution to other payments was particularly strong from single women (Oguzoglu et al., 2020), who are more likely to be financially constrained and faced a larger reduction in income from the retirement pension (due to the higher per-person payment rates for singles). Fourth, the Age Pension is unavailable to many older women due to the means test; Oguzoglu et al.'s (2020) estimates imply that the reform had no effect on Age Pension receipt for 48.1% of women.

Finally, there is a mismatch between women who are likely to claim the Age Pension at the APA and women who are most likely to be in the labor force at ages 60–64. Appendix Fig. B1 shows how Age Pension receipt and LFP varies across the distribution of household income for women in the sample. I sort women into quintiles based on their (current) equivalized household income, excluding their own income from the Age Pension (and other payments). In Appendix Fig. B1a, I present Age Pension receipt for women in the sample who are above the APA, and in Appendix Fig. B1b I present the LFP rates of women at age 59, before all women have reached the APA. As expected, Age Pension receipt is much higher for women in lower income households, while LFP is much higher for women in high-income

¹³ Using register data, Staubli and Zweimüller (2013) examine increases in the ERA (and stronger penalties for early claiming) for Austrian men and women. They find large effects on employment rates at ages where eligibility was affected due to a spike in labor force exits at the ERA. However, they find little effect on employment rates at younger or older ages (see Table 6 and Figs. 4 and 5 of their paper). Cribb et al. (2016) study a similar reform in the UK that raised women's ERA from 60 to 62. Despite clear effects on employment rates at ages 60 and 61, there is no sign of an effect on employment rates at ages 59 and 62 (see Fig. 3 of their paper). Moreover, the authors test for wealth effects by omitting cohort controls from their regressions and find that this does not materially affect their results. Finally, Geyer and Welteke (2021) exploit a sharp increase in the ERA in Germany from 60 to 63 for women born on or after January 1, 1952 using a regression discontinuity design. Although the reform was well anticipated and represented "a strong negative wealth shock" (Geyer & Welteke, 2017, p. 7), they find precisely estimated zero effects on employment rates prior to age 60 despite a large increase at ages 60–61 (14.4 percentage points).

¹⁴ Using administrative data on all welfare-receipt spells from 1998 to 2013, Oguzoglu et al. (2020) estimate that the reform decreased women's receipt of government payments by only 12.6 percentage points, since the 51.9-percentage-point decrease in Age Pension receipt was offset by a 39.3-percentage-point increase in other payments. Atalay and Barrett (2015) estimate a much larger decrease in women's receipt of government payments (27.0 percentage points). However, Oguzoglu et al. (2020, p. 76) note that Atalay and Barrett's (2015) estimates "may not have adequately dealt with cohort differences" in the receipt of government payments and come from "survey data that may under-report the receipt of disability and unemployment benefits".

households. Thus, the group of women who are more likely to remain in the labor force at ages 60–64 were likely to be relatively unaffected by the reform.

2.4 Atalay and Barrett's (2015) estimates

Atalay and Barrett (2015) estimate the causal effect of the reform on female LFP using a differences-in-differences design in which male cohorts form the comparison group. Figure 2 shows this approach graphically. Figure 2 replicates Fig. 1 in Atalay and Barrett (2015), showing the trends in male and female LFP at ages 60–64 over time. The black lines plot aggregate time series data from the Australian Bureau of Statistics Labor Force Survey, while the gray lines show the calculations based on the IHCS data (with the X indicating each survey).¹⁵ I also include a dashed line denoting the counterfactual change in female LFP; this line shows how female LFP would have changed over Atalay and Barrett's (2015) sample period if it followed the same trend as male LFP. Evidently, female LFP increased by 13.2 percentage points more than male LFP over the sample period. If men are a suitable control group, this comparison indicates that the reform has caused a substantial increase in female LFP.¹⁶

Atalay and Barrett (2015) formalise their approach using a regression framework. In their regressions, they estimate the impact of the reform on the LFP of women who were affected by the reform, those born after June 30, 1935 (the “after” cohort). Their approach compares the LFP of women in the after cohort with the LFP of women born before July 1, 1935 (the “before” cohort) — whose APA remained at 60 — using the equivalent male cohorts as the comparison group. Atalay and Barrett (2015) estimate the following regressions to implement this approach:

$$LFP_i = \beta x_i + \alpha_0 Female_i + \alpha_1 \Pr(CohortA_i^* = 1) + \delta[Female_i \times \Pr(CohortA_i^* = 1)] + u_i \quad (1)$$

where LFP_i is an indicator for individual i participating in the labor force; $Female_i$ is an indicator for individual i being female; $\Pr(CohortA_i^* = 1)$ is the probability that individual i is in the after cohort; and x_i includes controls for age, marital status, an indicator for having a bachelor's degree or higher, household size and state of residence.¹⁷ The key explanatory variable is $Female_i \times \Pr(CohortA_i^* = 1)$. Its coefficient, δ , identifies the effect of the reform on the LFP rates of women in the

¹⁵ As noted by Atalay and Barrett (2015), the LFP trends in the IHCS sample are broadly consistent with the aggregate data, and similar old-age LFP trends have been observed in many countries over recent decades. For example, Coile et al. (2018, p. 28) note a “U-shaped pattern for men that generally reaches its minimum in the mid-1990s and a much shallower U or continuously rising pattern for women”.

¹⁶ Figure 2 provides an opportunity to assess the plausibility of this assumption by comparing the LFP rates of men and women aged 60–64 before the reform. While male and female LFP rates followed different trends from 1978 to 1985, with a large decline in male LFP rates and little change in female LFP rates, the trends were broadly similar from 1985 to 1993, with male and female LFP rates both increasing by around 5 percentage points. However, this analysis corresponds to earlier birth cohorts that were not affected by the reform. I focus on the relevant cohorts in detail below.

¹⁷ The IHCS data includes people's age in years and the survey year and quarter. Hence, each person's date of birth lies within a fifteen-month window. To construct $\Pr(CohortA_i^* = 1)$, Atalay and Barrett (2015) assume that each person's birthdate is uniformly distributed within the possible window.

Table 1 Replication of baseline estimates.

	Atalay & Barrett (1)	Replication (2)
Treatment effect (Female \times After)	0.077*** (0.021)	0.078*** (0.020)
Female	-0.288*** (0.019)	-0.305*** (0.018)
After cohort	0.030* (0.017)	0.039** (0.017)
Mean: Women in after cohort		0.297
Implied percent increase in female participation		35.3%
R-squared		0.121
Observations	11,409	11,497

Notes: This table shows that I am able to successfully replicate Atalay and Barrett's (2015) baseline estimates on female labor force participation. The estimates (reported in Table 3 of Atalay and Barrett (2015)) come from estimating Eq. (1). The sample consists of men and women aged 60–64 from the eleven Income and Housing Costs Surveys between 1994/95 and 2009/10.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors, based on 999 replications, in parentheses.

after cohort under the assumption that, if not for the reform, the difference in the LFP rates of women in the before and after cohorts would have been equal to the corresponding difference among men.

Atalay and Barrett's (2015) estimate of δ is 7.7 percentage points ($p < 0.01$). This estimate indicates that, on average, the reform increased the LFP of women in the after cohort by 7.7 percentage points at ages 60–64. This estimate implies a large effect, given that it is an average impact among women in the affected cohorts and 41% of women in these cohorts had already reached the APA when surveyed. For example, if we were to assume that this impact only applied to women below the APA (consistent with the findings of similar studies), this estimate implies an increase in LFP of 13.1 percentage points.¹⁸ Overall, Table 1 shows that I am able to closely replicate this estimate. I find an estimate of δ of 7.8 percentage points, and the standard error is almost identical to the one reported in Atalay and Barrett (2015) to three decimal places.

In addition, I am able to closely replicate Atalay and Barrett's (2015) other LFP estimates (see Appendix Table B2); these estimates come from modified versions of Eq. (1) that allow the treatment effect to vary based on women's APA cohort (and APA cohort and age). For example, I am able to closely replicate Atalay and Barrett's (2015)

¹⁸ Some of Atalay and Barrett's (2015) estimates indicate that the reform significantly increased the LFP of women who had already reached the APA. This is inconsistent with the findings of studies examining similar reforms (see footnote 13) and difficult to reconcile with (i) the strong incentive to start claiming the Age Pension at the APA and (ii) the difficulty of supplementing pension income with labor income because of the income test. This suggests that Atalay and Barrett's (2015) estimates may be picking up factors other than the reform, such as cohort trends among women in high-income households.

headline estimate that each one-year increase in women's APA increased female LFP by 12.0 percentage points ($p < 0.05$).

For robustness, Atalay and Barrett (2015) also estimate their regressions without men as a control group (dropping the female dummy and interaction terms from their regressions). These estimates are larger, which Atalay and Barrett (2015) argue validates the use of male cohorts as a control group. However, this analysis provides no evidence that male cohorts are *correctly* accounting for any unobserved cohort and time factors affecting female LFP. Atalay and Barrett (2015) also estimate regressions on the subsample of single men and women.¹⁹ Atalay and Barrett (2015) motivate this analysis as allowing them to assess the robustness of the results to potential within-family spillovers. Atalay and Barrett (2015) estimate larger effects on this subsample (replicated in Appendix Table B2 for Eq. (1)). I examine whether single men are a suitable comparison group for single women in the next section.

From here on, I focus on the validity of the baseline specification (Eq. (1)). I do this because at least one of the identifying assumptions underpinning the expanded specifications will not hold if the baseline parallel-trends assumption does not hold. See Appendix A.1 for more details.

3 Robustness to cohort trends in female participation

In this section, I extend Atalay and Barrett's (2015) analysis with earlier data. Starting with aggregate time series data on labor force participation rates by age and gender, I examine whether the parallel-trends assumption held before the reform for the cohorts in Atalay and Barrett's (2015) sample.

Figure 3a and b show the trends in male and female LFP at ages 45–49 and 50–54 respectively. For now, focus on the black lines, which plot aggregate time series data from the Australian Bureau of Statistics Labor Force Survey. These lines show opposing LFP trends for men and women. For example, at ages 45–49, female LFP increased from 52.1% to 79.7% over the period from 1978 to 2010, while male LFP decreased from 94.0% to 89.8%. Figure 3a and b also include vertical dashed lines to denote Atalay and Barrett's (2015) sample period shifted back fifteen and ten years respectively. Between these lines, the sample consists of individuals born between 1929 and 1950, the cohorts in Atalay and Barrett's (2015) sample. During these periods, male and female LFP rates follow very different trends, with a large increase in female LFP rates and a small decrease in male LFP rates. Therefore, the parallel-trends assumption did not hold for the relevant cohorts before the reform.²⁰

¹⁹ Although not explicitly discussed, Atalay and Barrett (2015) appear to define their single group as all persons who are currently single. I use this definition and closely match the proportion of single men and women in their sample.

²⁰ I do not focus on the trends at ages 55–59 because the reform was announced before most of the relevant cohorts reached this age. As such, trends across female cohorts at this age could partially reflect responses to the reform. Nonetheless, the gender-specific trends across cohorts at this age are similar (see Appendix Fig. B2), with much higher LFP rates for younger female cohorts and fairly constant rates for male cohorts. Moreover, these trend differences are apparent for the relevant cohorts before the reform was announced in 1993 (and legislated in 1994).

A critical question is whether these gender-specific trends are driven by transitory factors or factors likely to persist through to Atalay and Barrett's (2015) sample period. To assess this question, I compare the trends in LFP rates across the relevant cohorts at several points in time. In Fig. 4a, I present LFP rates at ages 50–54 over the period from 1983 to 1994 against the five-year lag of LFP rates at ages 45–49 (here, both series correspond to individuals born between 1928 and 1944). In Fig. 4b, I present LFP rates at ages 45–54 over the period from 1978 to 1994 against the ten-year lag of LFP rates at ages 35–44 (here, both series correspond to individuals born between 1923 and 1949).²¹ In each figure, there are strong positive correlations for men and women between (i) the LFP rate in a given year and (ii) the lagged LFP rate of the same cohorts at a younger age. For example, there are strong upward trends in (i) female LFP rates at ages 45–54 from 1978 to 1994 and (ii) female LFP rates at ages 35–44 from 1968 to 1984, while the corresponding trends for men are weaker and in the opposite direction. Hence, persistent cohort factors can explain a large part of the gender-specific trends in Fig. 3.

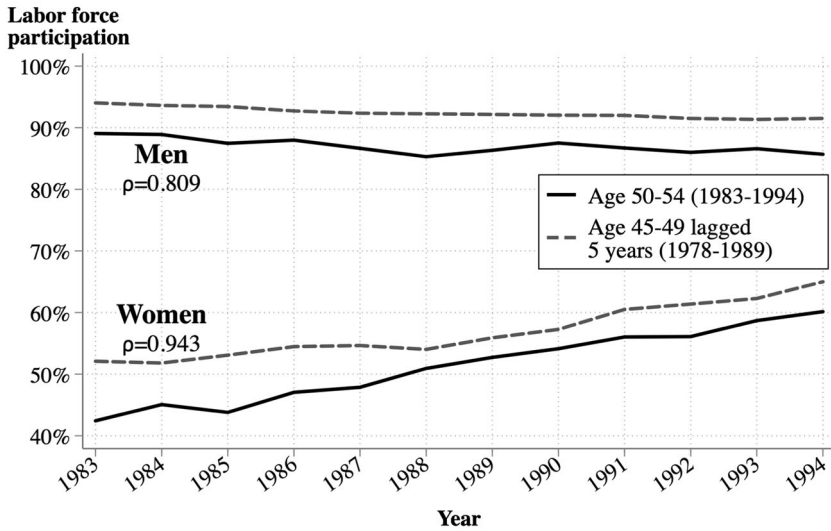
To further understand whether these trends would have persisted if not for the reform, it is useful to understand the characteristics of women behind the rise in female LFP. For example, if the rise in female LFP was entirely driven by women in low-income households, many of these women may have dropped out of the labor force once they turned 60 and became eligible for the Age Pension. To examine this, I use three earlier surveys administered by the Australian Bureau of Statistics: the 1982, 1986 and 1990 Income Distribution Surveys. These surveys are appropriate for this analysis because they (i) provide a nationally representative snapshot of LFP by age and gender at several points in time prior to the reform and (ii) have information on household income. Below, I also use these surveys to check whether the different trends of men and women can be explained by the controls in Atalay and Barrett's (2015) regressions and whether single men are likely to be an appropriate comparison group for single women.²² As outlined by Siminski et al. (2003), these surveys differ from the surveys used by Atalay and Barrett (2015) in important ways. For example, age information in the pre-reform surveys is in five-year groups rather than single years, and the pre-reform surveys were conducted over a shorter period of time (less than three months) than the post-reform surveys, which were conducted over the course of a financial year. Nonetheless, I show in Appendix A.3 that the placebo regression estimates below are only marginally affected by these and other differences in the pre-reform surveys.

With these surveys, I exclude immigrants who have lived in Australia for less than ten years (to be consistent with Atalay and Barrett (2015)), and I restrict the sample

²¹ To construct Fig. 4, I combine data from the Labor Force Historical Timeseries, the precursor to the Labor Force Survey, over the period from 1968 to 1977 with data from the Labor Force Survey over the period from 1978 to 1994. I cannot present LFP rates by age in five-year groups in Fig. 4b because they were not reported until 1978.

²² Other surveys administered by the Australian Bureau of Statistics during this period, namely the Household Expenditure Surveys in 1984 and 1988–89, lack information on relevant control variables such as education. I do not use the 1993–94 round of the IHCS as the LFP rates in this survey could have been affected by the announcement of the reform, though the results are robust to the inclusion of this survey (available on request).

(a) Participation rates at ages 50–54 and lagged rates at 45–49



(b) Participation rates at ages 45–54 and lagged rates at 35–44

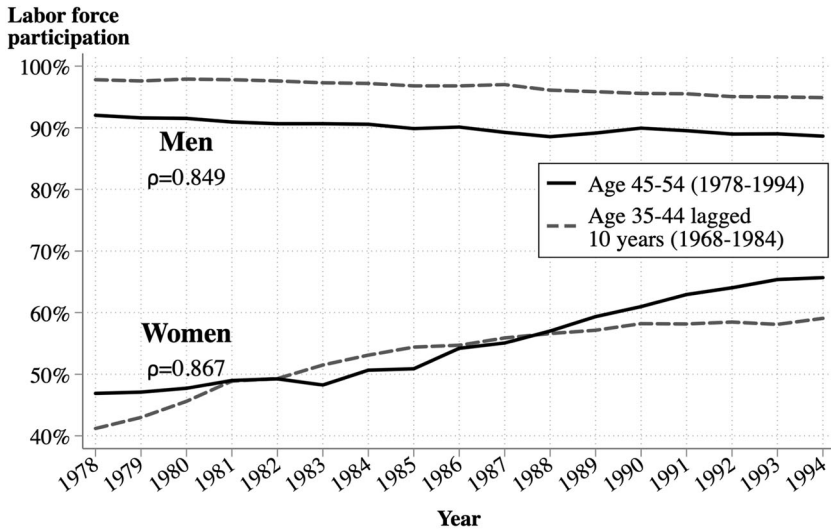


Fig. 4 Persistence of labor force participation trends across cohorts. Notes: These figures combine data from the Labor Force Historical Timeseries (1968–1977) with the Labor Force Survey (1978–1994). **a** plots labor force participation (LFP) rates at ages 50–54 from 1983 to 1994 against the five-year lag of LFP rates at ages 45–49 (here, both series correspond to the 1928–1944 cohorts). **b** plots LFP rates at ages 45–54 from 1978 to 1993 against the ten-year lag of LFP rates at ages 35–44 (here, both series correspond to the 1923–1949 cohorts).

to people aged 45–49 or 50–54.²³ The gray lines in Fig. 3a and b plot the LFP rates of men and women in each survey (with the X indicating each survey). As the graphs demonstrate, the LFP rates in these surveys are broadly consistent with the aggregate data at the time of each survey.²⁴

Figure 5 shows that the increase in female LFP rates has come from women in both high- and low-income households. Using the pre-reform surveys, I divide women in half based on the median equivalized household income among women of the same age in the relevant survey.²⁵ Figure 5 shows that although women in high-income households had much higher baseline rates of LFP, they exhibited a similar (and slightly larger) percentage-point increase in LFP rates between 1982 and 1990 at ages 45–49 and 50–54.²⁶ Assuming that most of these women would have remained in high-income households, they would have had limited access to government support at ages 60–64 even if the reform had not occurred. As such, it seems likely that women would have reduced the gender gap in LFP rates at ages 60–64 over Atalay and Barrett's (2015) sample period even if the reform had not occurred.

One key question remains: can the control variables in Atalay and Barrett's (2015) regressions explain the different trends of men and women across the relevant cohorts? To examine this question, I use the pre-reform sample to estimate placebo regressions with Atalay and Barrett's (2015) baseline regression approach. That is, I estimate Eq. (1) on the pre-reform sample to test whether the LFP of women in the after cohort was affected before the reform. To estimate these regressions, I have to define $\Pr(\text{Cohort}A_i^* = 1)$, the probability that individual i is in the after cohort. With age information in five-year groups and surveys conducted over a three-month period, each individual's date of birth lies within a window of five years and three months. This relatively large window makes it less appealing to follow Atalay and Barrett's (2015) assumption that each individual's date of birth is uniformly distributed within the possible window. Hence, I construct $\Pr(\text{Cohort}A_i^* = 1)$ in a more refined way that accounts for age-specific mortality and differences in the size of individual birth cohorts: using annual population counts by gender and single year of age from the Australian Bureau of Statistics (see Appendix A.2 for details).²⁷

²³ I also exclude the 2.1% of individuals who live in non-private dwellings because the post-reform surveys used by Atalay and Barrett (2015) only include residents of private dwellings. The estimates are similar without this restriction (available on request).

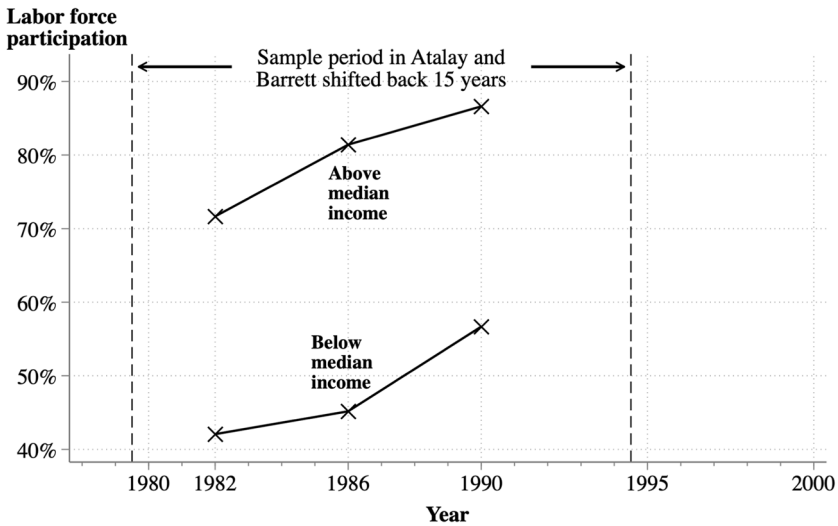
²⁴ The LFP rates in the survey data are also consistent with those in the sample files for the 1981, 1986 and 1991 Censuses, which contain a 1% random sample of all records in each Census (see Appendix Fig. B3).

²⁵ Since female LFP can directly determine whether women have household income above or below the median, I also examine the LFP trends of women based on the income of other members of their household. Specifically, I focus on partnered women and divide them in half based on the median income of their spouses in the relevant year. Appendix Fig. B4 shows that women with high-income spouses have significantly contributed to the increase in female LFP rates and have stronger LFP increases than women with low-income spouses.

²⁶ Unsurprisingly, men in high-income households did not experience the same increase in LFP rates at these ages, with LFP rates close to 100% in all three surveys (not shown).

²⁷ As it turns out, the estimates are similar if I assume that birthdates are uniformly distributed within the possible windows (see Appendix Table B3).

(a) Ages 45–49



(b) Ages 50–54

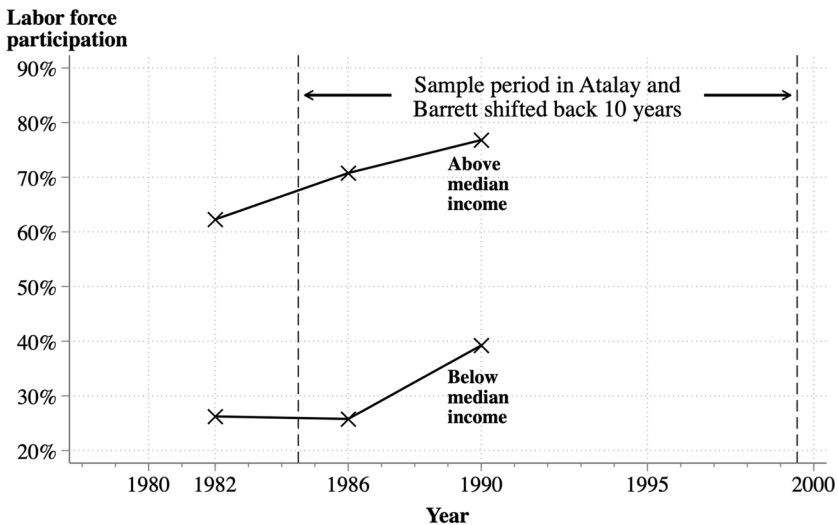


Fig. 5 Pre-reform female labor force participation rates by household income. Notes: These figures plot the labor force participation (LFP) rates of women from the 1982, 1986 and 1990 Income Distribution Surveys. I divide women in half based on the median equivalized household income among women of the relevant age in the relevant survey. Between the vertical dashed lines, the cohorts in Atalay and Barrett's (2015) sample were aged 45–49 in (a) and 50–54 in (b). These graphs show that women in low- and high-income households contributed to the higher LFP rates among later cohorts prior to the reform. See Section 3 for more details.

This approach is validated by the fact that these surveys are strongly representative of the Australian population in terms of age (see Appendix Fig. B5).²⁸

For these regressions, positive and statistically significant estimates of δ , the placebo treatment effect, would indicate that Atalay and Barrett's (2015) estimates are likely to be biased upwards by unobservable gender-specific differences in LFP across the relevant cohorts. Table 2 presents the results. In Panel A, I present the results for the full sample of men and women. Columns (1) and (2) show positive, large and statistically significant estimates of δ at ages 45–49 and 50–54 of 24.6 and 15.7 percentage points respectively (both $p < 0.01$).²⁹ In column (3), I remove anyone with a positive probability of being misclassified into the before and after cohorts. While this restriction precludes estimation on the age 45–49 sample (as there is no one who is definitely in the before cohort), the estimated placebo effect at ages 50–54 remains similarly large and statistically significant (16.1 percentage points, $p < 0.01$). In Panel B, I present the results when the sample is restricted to men and women with household income above the median (for their age and gender in the relevant year). These results are useful because high-income women would have had limited access to government payments at age 60 even if the reform had not occurred (and thus the pre-reform trends for this group may be especially likely to reflect counterfactual LFP trends at ages 60–64). The results are similar, with large and highly statistically significant placebo estimates ranging from 12.7 to 24.2 percentage points. Overall, this analysis suggests that Atalay and Barrett's (2015) estimates are likely to be biased upwards.

Using the pre-reform sample, I also examine whether single men are likely to be an appropriate control group for single women. Appendix Fig. B6 replicates Fig. 3 for this sample. I also present LFP rates from the Labor Force Survey for unmarried men and women, since the Labor Force Survey does not report LFP rates by single status. While 'unmarried' is an imperfect proxy for single status, this data provides a continuous time series over a longer period (and allows the cohorts in Atalay and Barrett's (2015) sample to be matched exactly). Evidently, the relevant cohorts of single men and women had very different LFP trends before the reform. Single women's LFP rates increased by approximately 15 percentage points at ages 45–49 and 50–54 between the 1982 and 1990 surveys, while there was little change in the LFP rates of single men at ages 45–49 and a decrease of 10 percentage points at ages 50–54. There are also strong gender-specific trends for the unmarried sample over the relevant years. Moreover, the gender-specific trends cannot be explained by the control variables in Atalay and Barrett's (2015) regressions. For the pre-reform sample of single men and women, the

²⁸ In Appendix Fig. B5, I calculate the fraction of male and female respondents in different age groups in each survey. The surveys allow for a consistent definition of 12 age groups, including 45–49 and 50–54. Then, I use estimates of the total population by sex and single year of age from the Australian Bureau of Statistics to construct the fraction of the 15+ population in each age group in the relevant years (1982, 1986, 1990). The figure shows that the population shares at each age closely align with the shares in each survey for men and women. The fit is especially good at the key ages of 45–49 and 50–54. As such, it is likely that the survey would also be representative within these five-year age groups. This justifies the use of population data to assign the probability that individuals are in the after cohort.

²⁹ Note that these regressions do not include individual-level controls for age, unlike Atalay and Barrett's (2015) regressions, as the pre-reform surveys only report age in five-year groups. However, in Appendix A.3, I show that the age distribution changed in very similar ways over time for men and women and, as such, the estimates are similar with controls for the average age of men and women at the time of the relevant survey.

Table 2 Placebo regression estimates on the pre-reform sample.

	Age 45–49 sample (1)	Age 50–54 sample 1 (2)	Age 50–54 sample 2 (3)
Panel A: All men and women			
Placebo effect (Female \times After)	0.246*** (0.038)	0.157*** (0.027)	0.161*** (0.027)
Female	–0.492*** (0.033)	–0.452*** (0.016)	–0.460*** (0.019)
After cohort	–0.057*** (0.019)	–0.032** (0.014)	–0.034** (0.015)
Mean: Women in after cohort	0.688	0.580	0.580
Implied percent increase in female participation	55.6%	37.0%	38.6%
R-squared	0.163	0.206	0.207
Observations	5,085	4,763	3,733
Panel B: Men and women with household income above median			
Placebo effect (Female \times After)	0.242*** (0.045)	0.127*** (0.030)	0.136*** (0.032)
Female	–0.382*** (0.040)	–0.338*** (0.020)	–0.351*** (0.023)
After cohort	–0.017 (0.012)	0.001 (0.011)	0.002 (0.011)
Mean: Women in after cohort	0.848	0.768	0.768
Implied percent increase in female participation	39.8%	19.8%	21.5%
R-squared	0.128	0.176	0.183
Observations	2537	2377	1862

Notes: This table presents placebo regression estimates on female labor force participation using data from the 1982, 1986 and 1990 Income Distribution Surveys. Column (2) includes the full sample of 50–54-year-olds. In column (3), I remove anyone with a positive probability of being misclassified into the before and after cohorts, which restricts the sample to individuals in the 1982 and 1990 surveys. In Panel B, I restrict the sample to men and women with equivalized household income above the median (for their age and gender in the relevant survey). See Eq. (1) and Section 3 for details on the regression specification.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors, based on 999 replications, in parentheses.

placebo estimates are positive, large and mostly statistically significant (ranging from 16.2 to 24.6 percentage points, see Appendix Table B4). Thus, Atalay and Barrett's (2015) estimates on single women are also likely to be biased upwards.

3.1 Discussion

It is worth making several points here. First, while the placebo estimates in Table 2 are larger than Atalay and Barrett's (2015) estimates of women's actual labor supply response at ages 60–64, women had much higher LFP rates at ages 45–49 and 50–54. Thus, in percentage terms, the placebo estimates imply a similar increase in female LFP to Atalay and Barrett's (2015) estimates (37.0–55.6% compared to 35.3%).

Second, the probabilistic assignment of the after cohort means that the estimates in column 1 for the age 45–49 sample are particularly sensitive to differences in LFP rates between the 1982 survey (cohorts 1932–37) and subsequent surveys. In the 1982 survey, there is a 48.4% chance that individuals are in the after cohort, and this uncertainty magnifies the placebo estimate. A conservative way to address this is to assume that everyone in the 1982 survey is in the before cohort (i.e., by reducing the probability that these individuals are in the after cohort to 0). Appendix Table B5 shows that this adjustment reduces the placebo estimate by 44% to 13.7 percentage points. However, as the standard error also falls by the same percentage, the t-statistic is unchanged (6.3). Note that this approach may if anything be conservative because the aggregate data in Fig. 3a shows that female LFP rates are slightly higher for the 1932–37 cohorts than the 1930–35 cohorts, while the reverse is true for male LFP rates; this is evident by comparing the LFP rates in 1982 (cohorts 1932–37) and 1980 (cohorts 1930–35). Thus, if it were possible to exactly match the before cohort, it is likely that the placebo estimate would be similar to 13.7 percentage points.

Third, there is little change in the placebo estimates with the inclusion of controls. As shown in Appendix Table B6, the placebo estimates fall by less than 5% with controls. Hence, the control variables in Atalay and Barrett's (2015) regressions explain only a small fraction of the different trends of men and women prior to the reform.³⁰

Fourth, as Australia experienced recessions in 1982 and 1990, a possible concern is that the placebo estimates may be affected by macroeconomic shocks. However, the aggregate time series data in Fig. 3a and b shows no evidence of any meaningful break in the LFP trends of the relevant cohorts in any of the three survey years. Moreover, Figs. 3 and 4 show that there is nothing specific about the 1982 to 1990 period. The aggregate data in Fig. 3a shows different trends in the LFP rates of men and women in the relevant cohorts both throughout and beyond these years, and Fig. 4b shows similar gender-specific trends for the relevant cohorts ten years earlier, during a different time period.

Fifth, as the pre-reform surveys are less frequent and contain age information in five-year groups rather than single years, it is not possible to exactly match the cohorts in Atalay and Barrett's (2015) sample. In the pre-reform sample, the age 45–49 sample consists of individuals born from 1932 to 1945, and the age 50–54 sample consists of individuals born from 1927 to 1940. Atalay and Barrett's (2015) sample consists of individuals born from 1929 to 1950. However, an exact match is possible in the aggregate data; between the vertical dashed lines in Fig. 3a and b, where the cohorts match those in Atalay and Barrett's (2015) sample, men and women have opposing trends. As such, we would expect to find similar results if it were possible to exactly match the cohorts in Atalay and Barrett's (2015) sample in the survey data.

Sixth, given the slightly different cohorts in Atalay and Barrett's (2015) sample, controls may have a greater effect on their estimates. However, controls have a

³⁰ Similarly, control variables explain only a small fraction of the rise in female LFP prior to the reform. Estimating the placebo regressions without men as the control group, I find that controls only reduce the placebo estimates by around 15%, and the estimates remain positive and highly statistically significant (ranging from 11.5 to 19.8 percentage points, see Appendix Table B7).

similarly minor impact on Atalay and Barrett's (2015) estimates, reducing their baseline estimate by just 10% (see Appendix Table B8).

Finally, in Appendix A.3, I verify the robustness of the placebo estimates further. Specifically, I show that the estimates are only marginally affected by other limitations of the pre-reform surveys, which are discussed in detail by Siminski et al. (2003) and include (i) changes in the measurement of labor force status across surveys, (ii) seasonality in LFP rates, and (iii) the lack of individual-level age controls.

4 Re-estimating the impact on female participation

In this section, I re-estimate the reform's impact on female LFP using a different approach. I follow the approach of several recent studies examining phased increases in pension-eligibility ages (Cribb et al., 2016; Oguzoglu et al., 2020; Staubli & Zweimüller, 2013). Specifically, I compare the LFP of women who are just below the Age Pension Age (APA) to those who are just above, with fixed effects for women's age and cohort. The benefit of this approach is that it directly controls for differences in female LFP across cohorts, which, as shown above, cannot be explained by controls.

For women in the replication sample, I estimate regressions based on the following specification:

$$LFP_i = \beta x_i + \delta \mathbf{1}(age_i < APA_i) + FE_age_yrs_i + FE_APA_Cohort_i + u_i \quad (2)$$

where $FE_age_yrs_i$ is a set of fixed effects for woman i 's age in years; $FE_APA_Cohort_i$ is a set of fixed effects for woman i 's APA cohort; $\mathbf{1}(age_i < APA_i)$ is an indicator variable for woman i being below her APA; and x_i includes the same variables as in Eq. (1) (except age). As age is only available in years, there is potential misclassification for some women of (i) the cohort fixed effects, $FE_APA_Cohort_i$, and (ii) the key variable, $\mathbf{1}(age_i < APA_i)$. I address this issue by replacing these variables with (i) the probabilities that woman i is in each APA cohort, $Pr(APA_Cohort_i)$, and (ii) the probability that woman i is below her (expected) APA, $Pr(age_i < \mathbb{E}[APA_i])$.³¹ In these regressions, δ estimates the effect on female LFP of women remaining below the APA because of the reform if the following assumption holds: different cohorts would have had identical age-related trends in LFP if not for the reform.

Figure 6 shows graphical evidence in support of this assumption; it shows the age-related trends in LFP for different cohorts of women. While there are considerable differences in LFP rates across cohorts, the trends with age are similar at ages 55–59, when all cohorts are below the APA. Moreover, in Appendix A.4, I show that — after controlling for age, cohort and individual factors — there is no evidence of any trend in female LFP based on the number of years until the APA (and no significant change at the APA).

³¹ To construct these variables, I first calculate the probabilities that woman i is in each APA cohort, $Pr(APA_Cohort_i)$, under the assumption that woman i 's birthdate is uniformly distributed within the possible fifteen-month window, as per Atalay and Barrett's (2015) approach. Then, I use these probabilities to calculate woman i 's expected APA, $\mathbb{E}[APA_i]$. Finally, I calculate $Pr(age_i < \mathbb{E}[APA_i])$ under the assumption that each woman's exact age is uniformly distributed on the interval $[age_y, age_y + 1)$, where age_y is her age in years.

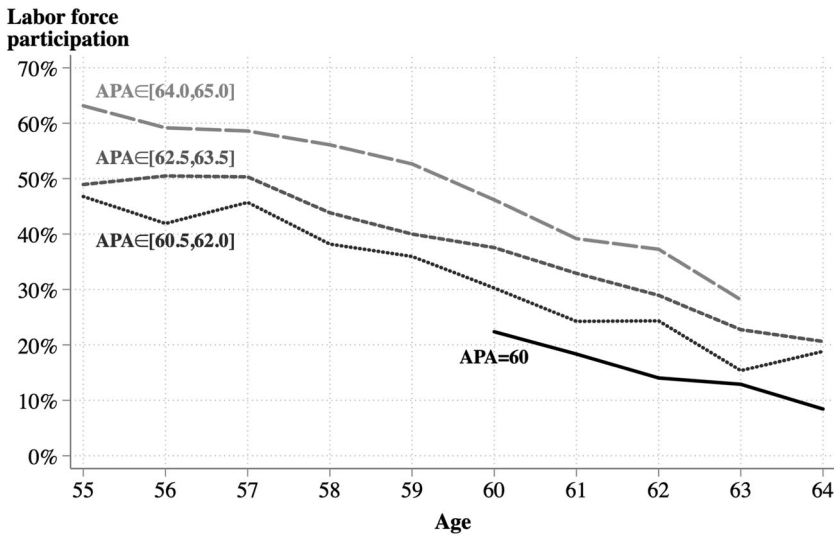


Fig. 6 Female labor force participation rates by age and cohort. Notes: This figure shows female labor force participation rates by age and cohort. I assign women to their most likely cohort. The sample comes from the eleven Income and Housing Costs Surveys used by Atalay and Barrett (2015) from 1994 to 2010.

Table 3 presents estimates of the treatment effect from Eq. (2). Column 1 shows a similar estimate to Atalay and Barrett's (2015) preferred estimate when cohort controls are excluded (13.4 vs. 12.0 percentage points). However, when cohort controls are included, column 2 shows that the estimate falls to 4.1 percentage points and becomes statistically indistinguishable from zero at the 10% level ($p = 0.109$).

Note that the cohort controls may partially or even fully absorb any wealth effects of the reform, the main mechanism offered by Atalay and Barrett (2015) to explain their estimates.³² Specifically, the cohort controls will absorb any wealth effects on female LFP at ages where women's pension eligibility is not affected, which, as discussed above, are likely to be small based on studies of similar reforms (Cribb et al., 2016; Geyer & Welteke, 2019; Staubli & Zweimüller, 2013).³³ However, the treatment effect may capture wealth effects at ages where women's pension eligibility is affected. For example, the estimates are likely to capture the response of a woman who decides to retire at age 63 rather than 60 because of the wealth effects of her APA increasing from 60 to 63. Given that changes in pension wealth may be particularly salient at ages where women's

³² It is worth pointing out that even if Atalay and Barrett's (2015) identification strategy were valid, their estimates cannot isolate the wealth effects of the reform from other mechanisms. As discussed above, there are several other mechanisms that could make it optimal for women to retire exactly at the APA, and Atalay and Barrett (2015) do not control for whether women have reached the APA. In Appendix Table B9, I add such controls and find that Atalay and Barrett's (2015) baseline estimate falls by 54% and becomes statistically indistinguishable from zero at the 10% level.

³³ These effects are also likely to be small in the Australian context due to (i) the low LFP rates and high reliance on other government payments among women in low-income households prior to the APA, (ii) the strong incentive to start claiming the Age Pension at the APA, and (iii) the difficulty of supplementing pension income with labor income because of the income test.

Table 3 Re-estimating the impact on female labor force participation.

	No cohort controls (1)	Preferred estimate (2)
Treatment effect (below pension age)	0.134*** (0.016)	0.041 (0.026)
Cohort controls		X
R-squared	0.086	0.093
Observations	5,863	5,863

Notes: This table re-estimates the effects of the reform on female labor force participation using Eq. (2). Column (1) excludes the cohort controls. Column (2) includes them and shows the preferred estimate. The regression sample consists of all women in the replication sample: women aged 60–64 in the eleven Income and Housing Costs Surveys between 1994/95 and 2009/10.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors, based on 999 replications, in parentheses.

pension eligibility is affected, the estimates may capture a sizable proportion of any wealth effects from the reform.

I also show that the estimates remain similar if I use time fixed effects rather than cohort controls in Eq. (2). Specifically, I add fixed effects for the quarter-year that the interview took place. This specification reduces measurement error (since women's exact cohort is unknown) and potentially allows δ to better capture differences in LFP rates across cohorts stemming from wealth effects. Cribb et al. (2016) use a similar approach to examine the importance of wealth effects in their context. Appendix Table B10 presents the estimates. The estimates are similar and actually slightly smaller than the baseline estimates (3.9 percentage points, $p = 0.075$). The estimates are also similar with survey-wave or year fixed effects (not shown).

I also assess the robustness of the results to other concerns. First, I examine whether the lack of exact information on age causes attenuation in the effects. In this context, there are two sources of measurement error: (i) in the assignment of women's APA cohort and (ii) in the assignment of whether women are below the APA. In Appendix Table B10, I examine the sensitivity of the estimates to each source of measurement error in columns 3–5. In column 3, I restrict the sample to women who can be classified as either above or below the APA with certainty, which reduces the sample size by 13%. The treatment effect remains similar and falls slightly compared to the baseline estimates (from 4.1 to 3.8 percentage points). In column 4, I restrict the sample to women who can be assigned to their APA cohort with certainty, which reduces the sample size much more, by 55%. The estimated treatment effect remains at 4.1 percentage points but becomes less precise. In column 5, I combine both restrictions, which reduces the sample size slightly more (56%). In this case the treatment effect falls to 3.1 percentage points. Thus, the estimates are relatively stable and there is little evidence that measurement error leads to attenuated estimates.

Second, since Atalay and Barrett (2015) interpret their LFP estimates as a decline in the probability of retirement, it is worth discussing under what circumstances the estimated effects on LFP are likely to reflect a change in retirement decisions. Based on the analysis in Section 3, one concern with using “not in the labor force” as a proxy for retirement is that women in earlier cohorts were more likely to have never worked (or dropped out of the labor force well before the APA), as documented in Figs. 3 and 4. This is problematic for Atalay and Barrett’s (2015) identification strategy and suggests that their estimates are more likely to reflect pre-existing differences in LFP rates across cohorts rather than changes in retirement decisions near the APA. In contrast, Eq. (2) relies on within-cohort changes in LFP around the APA. This variation is more likely to result from decisions to retire, given the low re-entry into the labor market at these ages.³⁴

Third, a further valid concern is reform endogeneity. There are two potential sources of endogeneity. First, it is clear that the reform was implemented in such a way that there was a strong positive correlation between women’s pre-reform rates of LFP and the size of the increase in APA. Unlike Atalay and Barrett’s (2015) estimates, which rely on cross-cohort comparisons, my preferred estimates account for this source of endogeneity by relying on within-cohort changes in female LFP around the APA. A second possible concern is that the reform may have been implemented in response to changes in social norms about the appropriate age to retire. This would imply that later cohorts would have retired later even if the reform had not occurred, biasing both estimates upwards. However, the trends in LFP rates prior to age 60 in Fig. 6 show little evidence that later cohorts systematically favored later retirement (conditional on their LFP rates at age 55). It also seems unlikely that my estimates are strongly biased upwards, given that (i) the 90% confidence interval for my estimate contains zero and (ii) economic theory suggests we should expect at least some increase in female LFP rates.

4.1 Discussion

The modest effects on female LFP are consistent with other evidence. First, the raw data shows little evidence of a drop in female LFP at the APA. Figure 7 plots the average LFP rate of women with respect to the number of years until their pension age.³⁵ While female LFP decreases with age, there is little evidence of a stronger drop at the APA. Moreover, Figure 4c of Morris (2019) shows a near-identical

³⁴ I assess this further in Appendix A.5 by examining whether the decrease in LFP rates around the APA coincides with an increase in the receipt of government payments. Specifically, I show that, around the APA, women mainly transition from being in the labor force and not receiving a government payment to being out of the labor force and receiving a government payment. Meanwhile, there is little change in the proportion of women who are in the labor force and receiving a government payment. This indicates that within-cohort changes in LFP around the APA are likely to mainly reflect decisions to retire with the support of government payments.

³⁵ I assign each woman’s number of years to pension age based on which group is most likely, assuming birthdates are uniformly distributed within the possible 15-month windows. I omit 16% of women who are equally likely to be in neighboring groups. The remaining women can be assigned to the correct group with an average probability of 0.83.

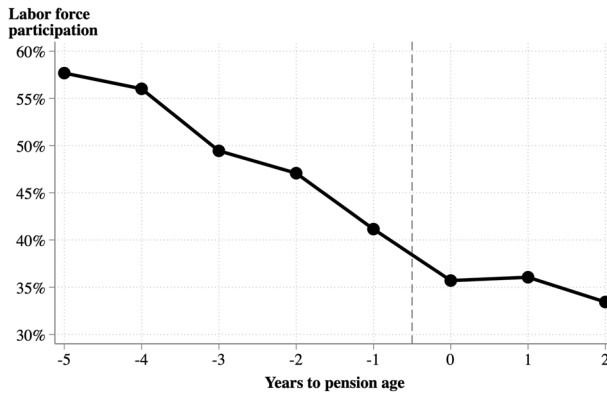


Fig. 7 Female labor force participation around the pension age. Notes: This figure shows the average labor force participation of women by the number of years to their pension age. I assign women's number of years to pension age based on their most likely group, assuming birthdates are uniformly distributed within the possible 15-month windows. The sample comes from the eleven Income and Housing Costs Surveys used by Atalay and Barrett (2015) from 1994 to 2010.

pattern in female LFP based on longitudinal survey data from 2001 to 2015 that contains precise information on birthdates and survey dates.³⁶ Second, with this data, Morris (2019) finds an increase in female LFP of similar magnitude (3.1 percentage points, $p < 0.1$) to the preferred estimate in Table 3 (4.1 percentage points). Although the two estimates correspond to the average impact of delayed pension eligibility on slightly different birth cohorts at slightly different ages, both estimates suggest that the reform had modest effects. Third, these impacts are consistent with similarly modest changes in male LFP at the APA. Using the same data as Morris (2019), Appendix Fig. B7 shows that the decrease in male LFP from ages 64 to 65 is only 5.5 percentage points larger than the decreases from 62 to 63 and 63 to 64.³⁷ This decline is smaller than the decline observed at the ERA in most countries.

Overall, the estimates suggest that the Australian reform had a smaller impact on female LFP than similar reforms in other countries. For example, studies of similar reforms in the U.K. (Cribb et al., 2016), Austria (Staubli & Zweimüller, 2013) and Germany (Geyer et al., 2020; Geyer & Welteke, 2021) estimate increases in female employment of 6.4 to 14.4 percentage points. The smaller impacts found here may partly result from the fact that around one third of women do not satisfy the means test — Atalay and Barrett (2015) estimate this proportion to be 37.5%. If we assume that all of the increase in LFP occurs among women who satisfy the means test, the estimated effect on LFP increases from 4.1 to 6.6 percentage points, similar to Cribb et al.'s (2016) estimate of 6.4 percentage points for the impact of raising women's State Pension Age from 60 to 62. However, it is worth remembering that my preferred estimate is statistically indistinguishable from zero at the 10% level.

³⁶ Morris (2019) uses waves 1–14 of the Household Income and Labor Dynamics in Australia (HILDA) survey.

³⁷ Unfortunately, it is not possible to construct a similar graph using the IHCS data, as age is only available in five-year groups at ages 65 and above.

It is somewhat surprising that the Australian reform did not have a larger impact given that the means test discourages work beyond the APA. This lies in contrast to the reform examined by Cribb et al. (2016) in which there is no income test for the pension and thus a much weaker incentive to retire at the State Pension Age. While the absence of an accrual effect is one explanation for the modest effects, it may not be the most important one. Relative to other reforms, the Australian reform targeted a group of women with much weaker attachment to the labor market and a very high reliance on other government transfers. As shown by several studies (Atalay and Barrett, 2015; Morris, 2019; Oguzoglu et al., 2020), many women extended their receipt of other means-tested government transfers in response to the reform. Recent estimates by Oguzoglu et al. (2020) with administrative data emphasize how large these responses were; Oguzoglu et al.'s (2020) estimates indicate that 76% of women who were delayed from receiving the Age Pension ended up receiving another government payment instead. These spillovers are much larger than have been found in other countries and help explain the relatively modest labor supply responses.

5 Conclusion

In this paper, I replicate Atalay and Barrett's (2015) labor supply estimates and extend their analysis. Atalay and Barrett (2015) estimate the effects of an Australian reform in 1994 that gradually raised women's pension-eligibility age from 60 to 65. Using a differences-in-differences design in which male cohorts form the comparison group, Atalay and Barrett (2015) estimate that each one-year increase in women's eligibility age increased female labor force participation by 12 percentage points. While I am able to successfully replicate this estimate, I show that the underlying parallel-trends assumption did not hold before the reform for the relevant cohorts. Specifically, there is a strong upward trend in female participation rates and a weak downward trend in male participation rates. After accounting for these gender-specific trends, I find that the estimated increase in female labor force participation falls by two-thirds to 4.1 percentage points and becomes statistically indistinguishable from zero at the 10% level. This suggests that the Australian reform had a smaller impact on female labor supply than estimates of similar reforms in other countries.

More generally, the results have implications for other studies examining the impact of retirement reforms on female labor supply or other (potentially) inter-dependent outcomes such as welfare receipt and health. This paper demonstrates the importance of cross-cohort trends in female labor supply in the Australian context, and given the widespread increase in female labor supply over recent decades, similar trends may also exist in other countries. Hence, studies evaluating the impact of retirement reforms on female outcomes should examine trends across cohorts before these reforms and control for them if necessary.

Acknowledgements I wish to especially thank my advisors — John de New, Timothy Moore, and Jenny Williams — for their helpful comments and suggestions, and I am grateful for comments from Garry Barrett, Bill Evans, Cain Polidano, Peter Siminski, Stefan Staubli, Roger Wilkins and four anonymous referees. I gratefully acknowledge the support of the Faculty of Business and Economics

and the Department of Economics at the University of Melbourne. This paper uses unit-record data from the Household Income and Labor Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services and is managed by the Melbourne Institute of Applied Economic and Social Research.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data used in this study can be obtained from the Australian Bureau of Statistics and the Australian Government's Department of Social Services, but restrictions apply to the availability of these data. Instructions for how other researchers can obtain the data, and all the information needed to proceed from the raw data to the results of the paper (including code) are available from the author upon request.

Compliance with ethical standards

Conflict of interest The author declares no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Atalay, K., & Barrett, G. F. (2015). The impact of age pension eligibility age on retirement and program dependence: Evidence from an Australian experiment. *Review of Economics and Statistics*, 97(1), 71–87. https://doi.org/10.1162/REST_a_00443.
- Barrett, G. F., & Tseng, Y. P. (2008). Retirement saving in Australia. *Canadian Public Policy*, 34(4), S177–S193.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?. *Quarterly Journal of Economics*, 119(1), 249–275.
- Borjas, G. J. (2017). The wage impact of the Marielitos: A reappraisal. *ILR Review*, 70(5), 1077–1110.
- Börsch-Supan, A., & Schnabel, R. (1998). Social security and declining labor-force participation in Germany. *American Economic Review*, 88(2), 173–178.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), 414–427.
- Chomik, R., Graham, S., Bateman, H., Piggott, J., & Yan, S. (2018). Retirement income in Australia: Part II - Public support. The ARC Centre of Excellence in Population Ageing Research (CEPAR).
- Clare, R. (2015). Superannuation account balances by age and gender. The Association of Superannuation Funds in Australia (ASFA), <https://www.superannuation.asn.au/policy/reports> (Accessed on Aug 25, 2020).
- Coile, C., Milligan, K. S., & Wise, D. A. (2018). *Introduction to social security programs and retirement around the world: Working longer*. University of Chicago Press.
- Cribb, J., Emmerson, C., & Tetlow, G. (2016). Signals matter? Large retirement responses to limited financial incentives. *Labour Economics*, 42, 203–212.

- Donald, S. G., & Lang, K. (2007). Inference with difference-in-differences and other panel data. *Review of Economics and Statistics*, 89(2), 221–233. <https://doi.org/10.1162/rest.89.2.221>.
- Geyer, J., & Welteke C. (2017). Closing routes to retirement: How do people respond? IZA Discussion Paper No. 10681.
- Geyer, J., & Welteke, C. (2021). Closing routes to retirement for women: How do they respond? *Journal of Human Resources*, 56(1), 311–341.
- Geyer, J., Haan, P., Hammerschmid, A., & Peters, M. (2020). Labor market and distributional effects of an increase in the retirement age. *Labour Economics*, 65, 101817. <https://doi.org/10.1016/j.labeco.2020.101817>.
- Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. NBER Working Paper No. 25018.
- Gruber, J., & Wise, D. (2004). *Social security programs and retirement around the world: Micro estimation*. Chicago: University of Chicago Press.
- Hanel, B., & Riphahn, R. T. (2012). The timing of retirement—New evidence from Swiss female workers. *Labour Economics*, 19(5), 718–728. <https://doi.org/10.1016/j.labeco.2012.05.013>.
- Lalive, R., Magesan, A., & Staubli, S. (2017). Raising the full retirement age: Defaults vs incentives. NBER Working Paper orrc17–12.
- Morris, T. (2019). The unequal burden of retirement reform: Evidence from Australia. SSRN Working Paper No. 3416304.
- OECD. (2018). Labour force statistics by sex and age: Indicators. *OECD Employment and Labour Market Statistics* (database). <https://doi.org/10.1787/data-00310-en>.
- Oguzoglu, U., Polidano, C., & Vu, H. (2020). Impacts from delaying access to retirement benefits on welfare receipt and expenditure: Evidence from a natural experiment. *Economic Record*, 96(312), 65–86.
- Samwick, A. A. (1998). New evidence on pensions, social security, and the timing of retirement. *Journal of Public Economics*, 70(2), 207–236.
- Siminski, P., Saunders, P., Waseem, S., & Chalmers, J. (2003). Assessing the quality and inter-temporal comparability of ABS household income distribution survey data. Discussion Paper No. 123, Social Policy Research Centre, University of New South Wales.
- Snyder, S. E., & Evans, W. N. (2006). The effect of income on mortality: Evidence from the social security notch. *Review of Economics and Statistics*, 88(3), 482–495.
- Staubli, S. E., & Zweimüller, J. (2013). Does raising the early retirement age increase employment of older workers?. *Journal of Public Economics*, 108, 17–32. <https://doi.org/10.1016/j.jpubeco.2013.09.003>.
- Vestad, O. L. (2013). Labour supply effects of early retirement provision. *Labour Economics*, 25, 98–109. <https://doi.org/10.1016/j.labeco.2013.04.001>.
- Wolfers, J. (2006). Did unilateral divorce laws raise divorce rates? A reconciliation and new results. *American Economic Review*, 96(5), 1802–1820.