



Effects of Taxes and Safety Net Pensions on life-cycle Labor Supply, Savings and Human Capital: the Case of Australia

by

Fedor Iskhakov

and

Michael Keane

ANU Working Papers in Economics and Econometrics
661

May 2018

JEL: J22, J24, J26, C63

ISBN: 0 86831 661 X

Effects of Taxes and Safety Net Pensions on life-cycle Labor Supply, Savings and Human Capital: the Case of Australia*

Fedor Iskhakov[†]

Michael Keane[‡]

May 16, 2018

Abstract

In this paper we structurally estimate a life-cycle model of consumption/savings, labor supply and retirement, using data from the Australian HILDA panel. We use the model to evaluate effects of the Australian aged pension system and tax policy on labor supply, consumption and retirement decisions. Our model accounts for human capital accumulation via learning by doing, as well as wealth accumulation and decumulation over the life cycle, uninsurable wage risk, credit constraints, a non-absorbing retirement decision, and labor market frictions. We account for the 'bunching' of hours by discretizing job offers into several hours levels, allowing us to investigate labor supply on both intensive and extensive margins. Our model allows us to quantify the effects of anticipated and unanticipated tax and pension policy changes at different points of the life cycle. Our results imply that the Australian Aged Pension system as currently designed is very poorly targeted, so that means testing and other program rules could be improved.

Keywords: Labor supply, human capital accumulation, retirement, pensions, taxes, structural model, anticipated and unanticipated policy changes, counterfactual simulations

JEL codes: J22, J24, J26, C63

*The research reported in this paper was supported by ARC grants FL110100247 and CE110001029. This paper uses unit record data from the Household, Income and Labor Dynamics in Australia (HILDA) Survey. The HILDA Project is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute. The findings and views reported in this paper, however, are those of the authors and should not be attributed to the ARC, FaHCSIA or the Melbourne Institute.

[†]Australian National University and ARC Center of Excellence in Population Ageing Research (CEPAR) fedor.iskhakov@anu.edu.au

[‡]University of New South Wales, School of Economics, and ARC Center of Excellence in Population Ageing Research (CEPAR) m.keane@unsw.edu.au

1 Introduction

In this paper we evaluate effects of income taxes and means tested safety net pensions on labor supply, consumption and retirement decisions using a structurally estimated life-cycle model. We use our model to analyze the Australian social security system. The Australian system is interesting, as it combines a defined contribution pension scheme, based on private accounts, with the safety net provided by the Aged Pension income support system for seniors with relatively low income/assets. This dual system has been praised as one of the best in the world by the OECD and other international organizations (Dixon, 2000; Mercer, 2015).

In contrast to the US Social Security system, the defined contribution aspect of the Australian system, known as “Superannuation”, takes Social Security taxes and payments off the government budget, as well as removing unfunded Social Security obligations from the intergenerational accounts.¹ However, in contrast to a completely privatized system, the Aged Pension provides fairly generous insurance for those who accumulate inadequate assets to finance retirement. A measure of the success of the privately run defined contribution scheme is that government spending on the Aged Pension is only 2.9% of GDP, less than half than OECD average (Chomik and MacLennan, 2014). Nevertheless, the Aged Pension is still a very large program in the Australian context. In 2014, the Aged Pension program cost \$60 billion, compared to total federal income taxes of \$180 billion.²

Our analysis is based on a rich life-cycle model that incorporates several key features: asset accumulation, liquidity constraints, human capital accumulation via learning by doing, and a discrete choice of hours involving six possible levels. Of course we also include the superannuation and Aged Pension rules, as well as retirement decisions. The model is novel, as there are no published life-cycle labor supply models that include both human capital and assets while also accounting for discreteness of hours, liquidity constraints and retirement decisions.³

The published work closest to our own is probably van der Klaauw and Wolpin (2008), who use a dynamic structural model to analyze the impact of the US Social Security (SS) system on

¹Participation in the defined contribution superannuation scheme is mandatory. Employers must contribute an amount equal to a fraction of workers’ salaries to the scheme (this is currently 9.5%) and employees are required to contribute at least a certain fraction as well. Contributions are managed by a large number of designated private investment funds (“super funds”), and workers can take the accumulated balance as a lump sum or annuity at retirement.

²Here and throughout the paper \$ refers to Australian dollar, \$AUD.

³Keane and Wasi (2016) was the first paper to incorporate human capital and assets as well as the extensive/intensive margin distinction in a complete life-cycle model. But they did not incorporate liquidity constraints or hours bunching at several levels. Also, their model was estimated on US data. A nice feature of our data is that the generosity up the Aged Pension was substantially increased in 2009-10. This change can plausibly be viewed as unanticipated. The change in rules provides us with a nice source of variation to help identify the model.

life-cycle labor supply and asset accumulation.^{4, 5} The key difference here is that, while SS is a universal system, the Australian Aged Pension is a means tested transfer (or “welfare”) program where participation is endogenous in the sense that whether one is eligible depends on earnings and assets. As Tran and Woodland (2014) point out, a means tested system may affect behavior through more channels than a universal system. Indeed, a universal system is a special case of a means tested system in which “taper rates” (i.e., the rates at which benefits are reduced with earnings and assets) are set at zero.

Our primary data source is the “Household, Income and Labor Dynamics in Australia Survey” (HILDA) which contains social and economic information on roughly 20 thousand households collected annually since 2001. The main focus of the survey is on family composition, income and labor supply, but it contains additional modules that collect information on wealth and superannuation. We use the first 16 waves of the panel to estimate our model.

We estimate the model via the method of simulated moments (MSM) using a sample of males aged 19 to 89. The moments we match include life-cycle profiles of labor supply (i.e., proportions choosing each of the six discrete hours levels at each age), profiles of wages and wage dispersion, wealth profiles, superannuation balances, and annual transition rates between employment and unemployment. All moments are conditional on education level. The simulated moments are based on the numerical solution of the dynamic programming problem.

In our model the continuous choice of consumption combined with discrete choice of hours renders the problem non-convex. There are kinks in the value functions, and the optimal consumption policy has discontinuities.⁶ To deal with this problem, we use the discrete-continuous generalization of the endogenous grid point method (DC-EGM) developed in Iskhakov et al. (2017). This paper is the first application of DC-EGM to an empirical model like ours. We make two other technical contributions: First, we develop a more convenient way to define the human capital state variable in dynamic models. Second, we present the first application of the smooth simulation algorithm of Bruins et al. (2018) to a dynamic structural model.

Our model implies a pattern of labor supply elasticities with age broadly consistent with the much simpler life-cycle model in Imai and Keane (2004).⁷ For instance, the Frisch elasticity is about 0.20 for college workers at age 30, but it grows to roughly 1.0 at age 65. Similarly, if we

⁴Keane and Wasi (2016) develop a similar life-cycle model to our own that incorporates US social security rules. But their focus is on how taxes affect labor supply, so they do not experiment with changes in SS rules.

⁵A limitation of van der Klaauw and Wolpin (2008) relative to our work is they treat assets and human capital at the first age when agents are observed, which is 50 to 60, as exogenously given, and hence invariant to SS policy rules. Since they don’t model the complete life-cycle they can’t analyze how changes in SS rules affect behavior at earlier ages. (On the other hand, they include health status in their model, which we do not).

⁶As a result, first order conditions alone can not be used to characterize optimal behavior.

⁷Imai and Keane (2004) were the first to fully structurally estimate a life-cycle model including both assets and endogenous human capital, but they assumed continuous hours and assumed interior solutions, ignored liquidity constraints, and did not model retirement or social security.

look at the effect of unanticipated permanent wage changes, Marshallian elasticities are quite small prior to age 45, but grow to about 0.50 at age 60, and 0.95 at age 65.

Our most interesting policy experiment is the complete elimination of the Aged Pension. This allows a lowering of income tax rates, so as to keep the government budget in balance. The experiment makes a very small fraction of people worse off ex post, particularly low education workers who also experience negative income shocks that leave them with low asset levels at retirement. But our welfare calculations indicate that, ex ante, all agents in the model would prefer to be born into a world without the Aged Pension (and with the lower tax rate). We take this to indicate that the Aged Pension is not well targeted at the relatively small population that benefits from the post-retirement consumption insurance that it provides. Indeed, the taper rate of the program is so low that benefits extend high up into the income distribution.⁸

The outline of the paper is as follows: In Section 2 we describe institutional features of the Australian social security system. In Section 3 we present our model. In Section 4 we discuss the DC-EGM solution method. In section 5 we describe the data. Section 6 discusses the estimation method and the moments that we fit. Section 7 presents our estimation results. In Section 8 we describe the fit of the model. Section 9 presents our policy simulations. Section 10 concludes.

2 The Social Security System in Australia

The Australian social security system was introduced in 1908 as an old-age and disability pension. It has since developed into a set of modern income support programs (Social Security Administration, 2015). Three characteristic features have been maintained throughout its growth; the prevalence of means tests, the use of funding from general revenue, and a strong emphasis on economic and social participation (Herscovitch and Stanton, 2008). The Australian model of social security differs markedly from many other countries, but is widely regarded as one of the best in the world.⁹

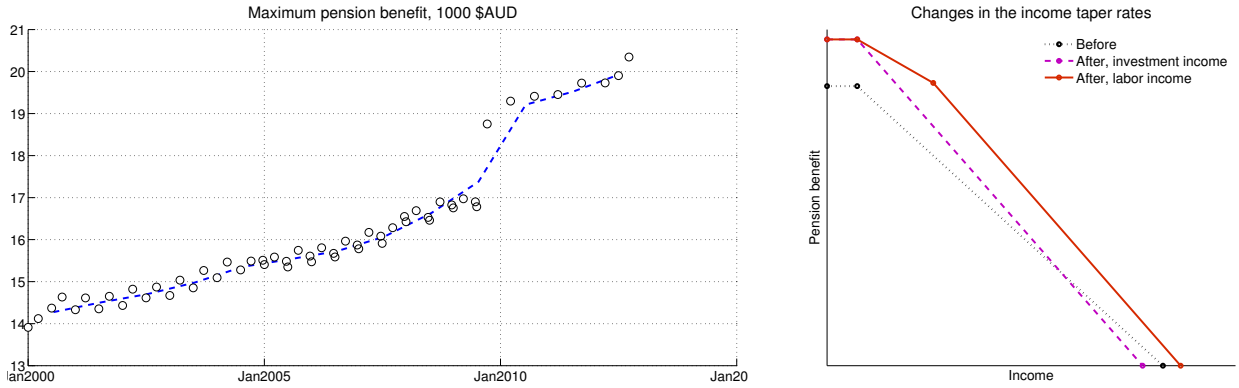
Australia's retirement income system consists of two parts. The first part is the "superannuation" system, which involves mandatory employer contributions paid into individual pension accounts, known as "super funds". The mandatory contribution rate was 9% from July 1, 2002 through June 30, 2013, which covers the bulk of our sample period.¹⁰ The private pension funds

⁸In ongoing work, we are experimenting with changes in design of the Aged Pension parameters that would improve the targeting of the program and perhaps improve ex ante welfare.

⁹The Mercer Global Pension Index, which compares adequacy, sustainability and integrity of retirement income systems around the world, gives Australia the third highest rating (Mercer, 2015).

¹⁰A guaranteed superannuation contribution rate for employers was introduced by the Keating government in 1992, at an initial rate of 3% that was set to increase gradually over time. By the period from July 1, 2000 through June 30, 2002 the rate was 8%. As noted in the text, the rate was then held fixed for 11 years, before being increased to 9.25% on July 1, 2013, and to 9.5% on July 1, 2014.

Figure 1: Changes in Age Pension rules, 2000-2012.



Notes: The left panel shows the maximum Age Pension benefit from 2000 to 2012. The dots are the actual levels, and the dotted line shows annual weighted averages. The right panel shows changes in taper rates and thresholds for the income test introduced on Sept. 1, 2009. Shown values are for single home owners.

that receive these contributions (i.e., the “super funds”) are closely regulated. There may also be additional voluntary contributions from employers, employees or the self-employed. These voluntary contributions are encouraged by tax concessions.

The second component of the retirement system is the means-tested “Age Pension” funded from general government revenue. The Age Pension is paid from the age of 65 (gradually rising to 67 by 2023) conditional on fulfilling residency requirements. Benefits, which are means tested based on income and assets, are calculated according to

$$\text{pension} = \max \{0, \text{full benefit} - \max \{ \text{income test}, \text{asset test} \} \}.$$

Thus, the full pension is reduced a certain amount, referred to as the “taper rate”, for every dollar of income above the income threshold, and every dollar of assets above the asset threshold,

$$\begin{aligned} \text{income test} &= \max \{0, \text{income taper rate} \cdot (\text{income} - \text{income threshold}) \}, \\ \text{asset test} &= \max \{0, \text{asset taper rate} \cdot (\text{wealth} - \text{asset threshold}) \}. \end{aligned}$$

The thresholds are rather high, so about 70% of age-eligible residents received at least a partial pension benefit in 2012 (FaHCSIA, 2012).¹¹

Age Pension benefits are indexed to the maximum of price and wage growth twice each year,

¹¹Different sources of income are treated differently under the income test: returns to financial investments are “deemed” at a fixed progressive rate, while incomes from long term income stream products (e.g. annuities) are reduced by returns of capital. Wealth is also treated differentially for the purpose of the asset test. Financial assets are assessed at their market value, while income stream products are assessed at their residual value. Most notably, a residential home is not included in the asset test. Our model largely disregards these details.

and the means-tests thresholds are adjusted annually to changes in the CPI or the Pensioner and Beneficiaries Cost of Living Index (PBCI), whichever is higher.

In 2009 the Australian government introduced several substantial changes in Age Pension rules. In particular, (a) the full pension benefit was increased by \$1689 per year, (b) the taper rate in the income test was increased from 40% to 50%, (c) the first \$500 earned fortnightly (\$13,000 annually) was subject to a lower taper rate of 25% (Commonwealth of Australia, 2009).¹² The three policy changes illustrated in Figure 1 came into effect on September 1, 2009.

At age 65, the amounts accumulated in the private superannuation accounts can be accessed. Although retirees can take superannuation in the form of annuities that provide retirement income streams, about half take the funds in a lump sum, while most others transfer the funds to phased withdrawal products that have no longevity protection and no upper limit on spending (Australian Prudential Regulation Authority, 2014).

3 The model

Here we present a life-cycle model of an individual who makes decisions on consumption and labor supply in each year of life after completing schooling. Retirement is endogenous and not necessarily an absorbing state. Working positive hours leads to accumulation of human capital through work experience (learning by doing), which affects the distribution of wage offers in future periods. The dynamics of the human capital accumulation process are specific to the individuals' education and unobserved heterogeneity type.

In each year starting at age $t = t_0$, and up to a maximum age $T = 100$, the agent chooses consumption c_t and hours of work h_t . The amount of consumption may not exceed the level of wealth M_t (measured at the beginning of the period) by more than a fixed amount a_0 denoting the credit constraint. Both M_t and a_0 are measured in \$1000, and we set $a_0 = 20$.

The choice of hours is restricted to six discrete levels given by

$$\begin{aligned}
 h_t \in H &= \{h^{(0)}, h^{(1)}, h^{(2)}, h^{(3)}, h^{(4)}, h^{(5)}\} & (1) \\
 &= \{0, 1000, 2000, 2250, 2500, 3000\} \text{ hours per year} \\
 &= \{0, 20, 40, 45, 50, 60\} \text{ hours per week.}
 \end{aligned}$$

We found that these six levels provide the best fit to the observed distribution of hours, using a k -median clustering algorithm with six clusters (see Section 5 for details).

¹²The reform also introduced a gradual increase of the eligibility age, which is planned to reach 67 in 2023, and some other minor changes that we leave out in this paper.

The deterministic part of the hourly offer wage rate is given by the product of human capital K_t and the rental price of a unit of human capital R_t . At the time the labor supply decision is made, workers are assumed to know the hourly offer wage rate up to a log-normally distributed idiosyncratic wage shock $\varepsilon_t^{wage} \sim \ln N(0, \sigma_t^{wage})$.¹³ Thus we have:

$$wage_t = K_t \cdot R_t \cdot \varepsilon_t^{wage}. \quad (2)$$

We assume the market for human capital is perfect, so all workers face the same human capital rental price R_t . We assume this is constant over time and so it is normalized to one. We also allow the variance of wage shocks change over the life cycle, $\sigma_t^{wage} = \varsigma_0 + \varsigma_1 t$.

Human capital K_t is a deterministic function of age and work experience \mathcal{E}_t , conditional on an individual's education $\tau_{edu} \in \{1, \dots, J^{edu}\}$ and, following Keane and Wolpin (1997), their unobserved skill endowment $\tau_{uh} \in \{1, \dots, J^{uh}\}$. To simplify notation, we denote compound *type* as $\tau = (\tau_{edu}, \tau_{uh})$. We estimate the model with $J^{uh} = 2$ unobserved types (“high” and “low”) and $J^{edu} = 3$ levels of education, namely high school dropouts (“dr”), high school graduates (“hs”) and college graduates (“clg”). Section 5 provides more details on how education levels are defined in the data. Both education and the unobserved type affect the stock of human capital at the start of one's career, and the human capital accumulation process.

Work experience $\mathcal{E}_t \in [0, 1]$ is defined as the fraction of total available time $t \cdot h^{(5)}$ that was devoted to work up to and including period $t - 1$. For computational reasons, it is more practical to use work experience $\mathcal{E}_t \in [0, 1]$ than human capital K_t as a state variable in the model because it spans the unit interval in every time period. State variable \mathcal{E}_t has a convenient deterministic rule of motion given by the following recursive expression:

$$\mathcal{E}_{t+1} = \begin{cases} \frac{h_t}{h^{(5)}}, & \text{if } t = 0, \\ \frac{1}{t+1} \left(\mathcal{E}_t t + \frac{h_t}{h^{(5)}} \right), & \text{if } t > 0. \end{cases} \quad (3)$$

Thus, the human capital production function is given by

$$K_{t+1}(\tau) = \exp \left(\eta_0(\tau_{edu}) + \eta_0(\tau_{uh}) + \eta_1(\tau_{edu}) \cdot t\mathcal{E}_t + \eta_2(\tau_{edu}) \cdot (t\mathcal{E}_t)^2 + \eta_3 t + \eta_4 t^2 \right) \quad (4)$$

where $t\mathcal{E}_t$ is total work experience (measured in maximum-hours-equivalent time periods), and the η are parameters to be estimated. The initial level of human capital is determined by the sum of the education specific intercept $\eta_0(\tau_{edu}) \in \{\eta_{0,clg}, \eta_{0,hs}, \eta_{0,dr}\}$, and the unobserved

¹³To allow for the wage shock to be integrated out together with other exogenous state transitions, we follow the standard technical trick of assuming that employment during period t earns wage income in period $t + 1$.

type specific intercept $\eta_0(\tau_{uh}) \in \{0, \eta_{0,\text{high}}\}$. The mapping from experience to human capital, governed by $\eta_1(\tau_{edu})$ and $\eta_2(\tau_{edu})$, also varies by education level.

At each age t the agent derives instantaneous utility $u(c_t)$ from consumption c_t and disutility $v_t(h_t, \tau_{uh})$ from working. The agent also takes into account the value of possibly leaving a bequest $B(M_t - c_t)$, discounted by the probability $1 - \delta_t$ of death before the start of period $t + 1$. Section 7.1 explains how the survival probability δ_t is specified. Preferences are additively separable within period and over time. Future utility is discounted with discount factor $\beta(\tau_{edu})$ which is education level specific.

Denote by $V_t(X_t)$ the highest attainable discounted expected utility over the remaining life cycle, given the state $X_t = (M_t, \mathcal{E}_t, \tau)$ of the decision maker at period t . The intertemporal decision problem is given by the *Bellman equation*:

$$V_t(X_t) = \max_{\substack{0 \leq c_t \leq M_t + a_0, \\ h_t \in H_t(\tau)}} \left\{ u(c_t) - v_t(h_t, \tau_{uh}) + \right. \\ \left. + \delta_t \beta(\tau_{edu}) E[V_{t+1}(X_{t+1}) | X_t, c_t, h_t] + (1 - \delta_t) \cdot B(M_t - c_t) \right\}, \quad (5)$$

The expectation in (5) is taken over the wage shocks, as well as discrete-hours specific taste shocks (introduced in section 4). The solution of (5) is given by the optimal decision rules

$$\begin{aligned} c_t^*(X_t) &: X_t \rightarrow [0, M_t + a_0], \\ h_t^*(X_t) &: X_t \rightarrow H_t(\tau) \subset H, \end{aligned} \quad (6)$$

which map the state X_t into optimal choices of consumption and labor supply. We impose some restrictions on the set of feasible hours $H_t(\tau)$ that distinguish it from the set H of all discrete hours levels. First, we assume working is not feasible after the maximum retirement age 85, i.e. $H_t(\tau) = \{h^{(0)}\}$ for $t \geq 85$. Second, we impose a minimum level of next period consumption $c_{min} = \$100$ per year, and disallow the choice of zero hours (i.e. $h^{(0)} \notin H_t(\tau)$) if optimal consumption in the next period would be below c_{min} . Finally, we assume that $t_0 = 19$ for high school graduates and dropouts, and $t_0 = 23$ for college graduates.

Letting $\mathbb{1}\{\cdot\}$ denote the indicator function, the intertemporal budget constraint is given by:

$$\begin{aligned} M_{t+1} &= (M_t - c_t)(1 + r) + \textit{AfterTax}(h_t \cdot \textit{wage}_{t+1}) \\ &\quad + tr \cdot \mathbb{1}\{t + 1 \leq 23\} \\ &\quad + \textit{pens}_{t+1} \cdot \mathbb{1}\{t + 1 \geq 65\} \\ &\quad + \textit{super}_{t+1} \cdot \mathbb{1}\{t + 1 = 65\}. \end{aligned} \quad (7)$$

The first component of (7) is assets carried over from the previous period at interest rate r , the second is labor earnings net of taxes, the third is transfers from parents, tr , which individuals receive up to age 23, the fourth is the Aged Pension, $pens_t$, received from age 65 onward, and the fifth is the superannuation lump sum payment, $super_t$, received at age 65.

In (7) we attribute to period $t + 1$ the wage income resulting from the labor supply decision made in period t . This is *equivalent* to assuming the labor supply decision is made before the wage shock is realized. Hence, the wage shock must be integrated out in the Bellman equation.

The non-labor income terms in (7) are specified in detail as follows. First, the transfer from parents at young ages tr is uniform across types, and is estimated as a single parameter. It is received in every period between t_0 and age 23, and thus enters into the initial assets given by $M_0 = tr + \tilde{M}_0$, where \tilde{M}_0 has a log-normal distribution $\tilde{M}_0 \sim \ln N(0, \varsigma_{w_0})$. The initial asset level M_0 can be interpreted as a one time bequest from parents received at $t = t_0$.

Next, the Aged Pension ($pens_t$) is modeled using a smoothed version of the means tested pension equation estimated from the data in Section 7.1:

$$pens_t = \mathcal{M}_\nu \left(10, 826.40 + 1249.67 \cdot \mathbb{1}\{year \geq 2010\} \right. \\ \left. - \mathcal{M}_\nu [0, \mathcal{M}_\nu(0.26906 \cdot wage_t, 0.00402 \cdot (M'_t - 140, 326.90))] \right), \quad (8)$$

where $\mathcal{M}_\nu(x, y) = \nu \cdot \log(\exp(x/\nu) + \exp(y/\nu))$ is a smooth maximum function.¹⁴ We fix the smoothing parameter at $\nu = 0.1$, resulting in the maximum error of $\log(2)\nu = \$69.3$ per year. The variable M'_t is wealth counted in the asset test, given by $M'_t = M_t - pens_t$, $t \geq 65$.

Finally, $super_t$ in equation (7) denotes the one-time withdrawal of funds from the superannuation account at age 65, the earliest age for withdrawal without tax penalties. Modeling other withdrawal options, such as annuities or phased withdrawal products, is beyond the scope of this paper.¹⁵ Our model also abstracts from voluntary contributions to superannuation accounts. We model the accumulated super account balance as a deterministic function of the amount of labor supplied throughout the career, which is summarized by the amount of accumulated human capital. Thus, we model the superannuation wealth as a linear function of human capital with a type specific slope:

$$super_t = \rho_0 + \rho_1(\tau_{edu}) \cdot K_t, t = 65. \quad (9)$$

The three parameters in $\rho_1(\tau_{edu}) \in \{\rho_{c\lg}, \rho_{hs}, \rho_{dr}\}$ are structurally estimated, while $\rho_0 = 0$ is

¹⁴ $\mathcal{M}_\nu(x, y) \rightarrow \max(x, y)$ when $\nu \rightarrow 0$.

¹⁵Our assumption of lump sum withdrawal is not too restrictive, as the vast majority of the Australian retirees either take out lump sums, or retain full control to draw down their accumulated pension wealth. We do not make a distinction between savings and such phased withdrawal accounts.

fixed at zero for normalization.

Our simple treatment of the superannuation system is internally consistent, as the payout at age 65 is taken into account by individuals when making their labor supply and consumption choices throughout the life cycle. Yet, because we assume away voluntary contributions, policies aimed at affecting this behavior obviously can not be simulated. Moreover, all of our counterfactual simulations should be interpreted as conditional on the assumption that voluntary contributions to the superannuation system are invariant to the experiment.¹⁶

We now turn to the specification of preferences. We use a constant relative risk aversion specification for the utility of consumption $u(c_t)$ and the bequest motive $B(b_t)$.

$$u(c_t) = \frac{c_t^{1-\zeta} - 1}{1-\zeta}, \quad (10)$$

$$B(b_t) = b_{scale} \cdot \frac{(b_t + a_0)^{1-\xi} - a_0^{1-\xi}}{1-\xi}, \quad (11)$$

where $b_t = M_t - c_t$ denotes bequeathed wealth, and $b_{scale} > 0$, $\zeta > 0$ and $\xi > 0$ are parameters to be estimated. Because we assume the credit constraint $c_t \leq M_t + a_0$ holds in every period, the power in $B(b_t)$ is taken for non-negative numbers only. Both $u(c_t)$ and $B(b_t)$ nest log specifications when $\zeta \rightarrow 1$ or $\xi \rightarrow 1$.

The disutility of work $v_t(h_t, \tau_{uh})$ is given by a vector of constants associated with each of the discrete levels of hours, $\gamma = (\gamma^{(1)}, \dots, \gamma^{(5)})$. These coefficients are allowed to vary with age and the unobserved heterogeneity type. We have:

$$v_t(h_t) = \mathbb{1}\{h_t > 0\} \cdot \kappa_{type}(\tau_{uh}) \cdot \kappa_{age}(t) \cdot \gamma(h_t), \quad (12)$$

where $\gamma(h_t) = \gamma^{(i)} \Leftrightarrow h_t = h^{(i)}$, $i \in \{1, \dots, 5\}$, and the type specific parameters are given by

$$\kappa_{type}(\tau_{uh}) = 1 + \kappa_1 \cdot \mathbb{1}\{\tau_{uh} = \text{low}\}, \quad (13)$$

$$\kappa_{age}(t) = 1 + \kappa_2 \cdot (t - 40)^2 \cdot \mathbb{1}\{t > 40\} + \kappa_3 \cdot (t - 25) \cdot \mathbb{1}\{t < 25\}. \quad (14)$$

As we abstract from job availability and involuntary job loss, estimates of the parameters $\gamma^{(1)}, \dots, \gamma^{(5)}$ will reflect not only disutilities of working the associated hours levels, but also the extent to which jobs at each hours level are available. These are fixed structural parameters

¹⁶Feng et al. (2014, p. 14) report that as coverage of superannuation has gradually increased since the introduction of the system in 1980s, to nearly universal coverage today, the participation rate in voluntary contribution “has fallen steadily at a rate of 2.5% per year from half of employees making voluntary contributions in 1993, to a little less than a quarter in 2007.” They also report a number of substantial changes in the rules governing voluntary contributions in the time period of our data, which “had an effect of introducing uncertainty around rule changes.” Modeling this would be very difficult, and is well beyond the scope of this paper.

from the point of view of an individual decision maker, although they are determined by technology and the equilibrium of the economy. We assume they are invariant to our experiments.

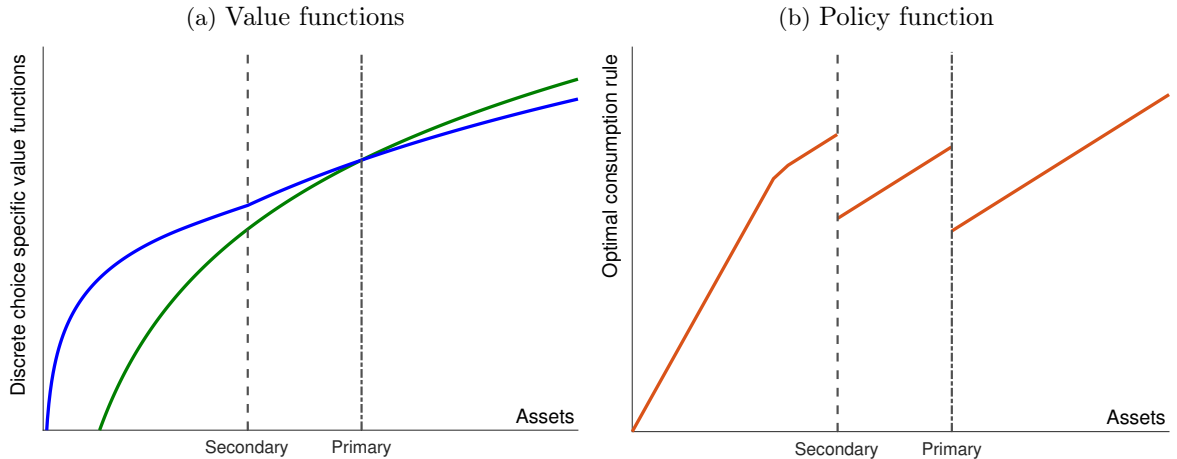
Our model includes discrete unobserved heterogeneity in the intercepts $\eta_0(\tau_{uh}) \in \{0, \eta_{0,\text{high}}\}$ in the human capital production function (4), and in the disutility of work parameter κ_1 in (13). We assume there are two types ("high" and "low") with type proportions $(p_{\text{clg}}, p_{\text{hs}}, p_{\text{dr}})$ that differ by education level.

4 Numerical solution method: The DC-EGM Algorithm

We solve the model using the DC-EGM algorithm developed by Iskhakov et al. (2017) which generalizes the endogenous grid point method (EGM) of Carroll (2006) to problems with discrete choices. As in the original EGM, the main idea of DC-EGM is to use an exogenous grid over the post decision state variable (savings) to construct an endogenous grid over the state space (wealth) such that the first order conditions in the Bellman equation are satisfied. In other words, instead of looking for optimal actions on a fixed grid over the state space, the method starts with a grid over the action space, which is mapped point-by-point into the *endogenous* grid over the state using the (analytically) inverted Euler equation. This approach avoids costly root finding operations, and because of its reliance on the exact solutions to the Euler equation, outperforms traditional methods in both speed and accuracy.

For concave maximization problems the first order conditions of the Euler equation are sufficient. But, as is well-known, problems containing mixed discrete and continuous choices are in general non-concave, and solution methods can not rely on first order conditions alone. Fella (2014) applies EGM to non-concave problems by first identifying the regions where first order conditions are not sufficient and using standard solution methods in such regions. Iskhakov et al. (2017) develop a generalization of EGM that automatically detects non-concave regions after studying their behavior in dynamic models. They show that in deterministic discrete-continuous dynamic choice problems non-concavities accumulate in the backward induction procedure, leading to kinks in value functions and discontinuities in the policy functions, which significantly complicates numerical solution (see Figure 2). Introduction of discrete choice specific random utility components in the tradition of McFadden (1973) and Rust (1987) alleviates the problem of the accumulation of kinks and discontinuities. As the model is not necessarily fully concave even with added smoothing, Iskhakov et al. (2017) show how the optimal policy can be efficiently computed in the remaining non-concave regions by comparing the solutions of the Euler equation for all local maxima in the Bellman equation. Filtering out and discarding the suboptimal solutions of the Euler equation is the primary innovation of the DC-EGM algorithm,

Figure 2: Kinks and discontinuities in discrete-continuous choice problems.



Notes: Using a simple consumption-savings model with binary choice, the plots illustrate the non-concavity of the value function in dynamic models with discrete and continuous choices. Comparing discrete choice specific value functions (panel a) leads to a *primary kink* in the value function at the level of assets where the decision maker is indifferent between the two discrete options. The primary kink in the value function is accompanied by a *primary discontinuity* of the optimal consumption function (panel b) because typically optimal consumption rules differ between discrete options. Moreover, the primary kink in period t gives rise to a *secondary kink* in the value function (panel a), and a *secondary discontinuity* of the optimal consumption function (panel b) in period $t - 1$. This is due to multiple local optima in the maximization in the Bellman equation, which arise because of general non-concavity of the problem caused by primary kinks. The secondary kink and discontinuity (both panels) resides at the level of assets that equate different local optima, or equivalently at the level of assets that leads to the primary kink location in next period assets under the optimal policy. In deterministic problems the number of kink points strictly increases in the process of backward induction. See Iskhakov et al. (2017) for more detail.

which ensures its speed and accuracy.¹⁷

Following Iskhakov et al. (2017) we assume that the choice of hours in our model is affected by additional state variables that are not observed by the econometrician. The influence of these state variables can be accounted for by adding random components to the deterministic utilities of hours, which we refer to as *taste shocks*. Denote these taste shocks $\varepsilon_t = (\varepsilon_t^{(0)}, \dots, \varepsilon_t^{(5)}) \in \mathbb{R}^6$, so that $\varepsilon_t^{(i)}$ is associated with supplying $h_t^{(i)}$ hours of labor. We assume that the random variables ε_t have independent multivariate extreme value distributed with a common scale parameter λ . They are also independent of wage shocks ε_t^{wage} and across time.

The intertemporal decision problem (5) modified by inclusion of taste shocks can then be written as

$$V_t(X_t) = \max_{h_t \in H_t(\tau)} \left\{ W_t(X_t, h_t) + \lambda \varepsilon_t(h_t) \right\}, \quad (15)$$

¹⁷A similar approach is extended to the multidimensional decision problems by Druedahl and Jørgensen (2017).

where discrete choice specific value functions $W(X_t, h_t)$ are given by

$$\begin{aligned}
W_t(X_t, h_t) &= \max_{0 \leq c_t \leq M_t + a_0} \left\{ u(c_t) - v_t(h_t, \tau_{uh}) + (1 - \delta_t)B(M_t - c_t) \right. \\
&\quad \left. + \delta_t \beta(\tau_{edu}) E \left[\max_{h_{t+1} \in H_{t+1}(\tau)} \{W_{t+1}(X_{t+1}, h_{t+1}) + \lambda \varepsilon_{t+1}(h_{t+1})\} \mid X_t, c_t, h_t \right] \right\} = \\
&= \max_{0 \leq c_t \leq M_t + a_0} \left\{ u(c_t) - v_t(h_t, \tau_{uh}) + (1 - \delta_t)B(M_t - c_t) \right. \\
&\quad \left. + \delta_t \beta(\tau_{edu}) E \left[\mathcal{LS}_\lambda (W_{t+1}(X_{t+1}, h_{t+1})) \mid X_t, c_t, h_t \right] \right\},
\end{aligned} \tag{16}$$

where $\varepsilon_t(h_t) = \varepsilon_t^{(i)} \Leftrightarrow h_t = h^{(i)}$, $i \in \{0, \dots, 5\}$, and $\mathcal{LS}_\lambda(x_1, \dots, x_k)$ is logsum function, which is a multivariate version of the smooth maximum function $\mathcal{M}_\nu(x, y)$ in (8). The overall value function in the modified problem is given by

Similar to (6), the solution to (16)-(15) gives rise to the optimal decision rules

$$\begin{aligned}
c_t^*(X_t) &: X_t \rightarrow [0, M_t + a_0], \\
h_t^*(X_t) &: X_t \rightarrow \left(P_t(h^{(0)}), \dots, P_t(h^{(6)}) \right), P_t(h_t) = 0 \forall h_t \notin H_t(\tau),
\end{aligned} \tag{17}$$

where $P_t(h^{(i)})$ is the choice probability of supplying $h^{(i)}$ hours of labor at period t , given by

$$P_t(h^{(i)}) = \frac{\exp(W_t(X_t, h^{(i)})/\lambda)}{\sum_{h'_t \in H_t(\tau)} \exp(W_t(X_t, h'_t)/\lambda)}. \tag{18}$$

The DC-EGM algorithm relies on the Euler equation which remains a necessary condition even in presence of discrete choice (Strub and Clausen, 2014). The Euler equation for the model is given by¹⁸

$$\begin{aligned}
u'(c_t) &= (1 - \delta_t)w'(M_t - c_t) \\
&\quad + \delta_t \beta(\tau_{edu}) E \left[\sum_{h_{t+1} \in H_{t+1}(\tau)} P_{t+1}(h_{t+1}) u'(c_{t+1}) [1 + r - R^{pens}] \mid X_t, c_t, h_t \right],
\end{aligned} \tag{19}$$

where $R^{pens} = \frac{\partial pens_{t+1}}{\partial M_{t+1}} \cdot \mathbb{1}\{t + 1 \geq 65\}$ is the effect of age pensions due to means testing. With out specification of the utility function $u(c_t)$ in (10), (19) can clearly be inverted analytically, which ensures the speed of our numerical solution.

In more detail, the DC-EGM algorithm iterates on Euler equation (19), and uses Bellman equation (16) to evaluate the choice specific value functions $W_t(X_t, h_t)$. In the terminal period T the optimal choice of consumption $c_T(X_T, h_T)$ conditional on h_T and choice specific value

¹⁸See derivation in the Appendix B.

functions are given by

$$c_T(X_T, h_T) = \arg \max_{0 \leq c \leq M_T + a_0} \left\{ u(c) + B(M_T - c) \right\}, \quad (20)$$

$$W_T(X_T, h_T) = u(c_T(X_T, h_T)) + B(M_T - c_T(X_T, h_T)). \quad (21)$$

Let $\{A_1, \dots, A_J\}$ denote the constant ascending exogenously fixed grid over savings $A_t = M_t - c_t$, with $A_1 = 0$ and $A_j < A_{j+1}$. In all preceding periods $t < T$ and for all $h(t) \in H_t(\tau)$ we perform the following steps, given the already computed optimal consumption function $c_{t+1}(X_{t+1}, h_{t+1})$ and choice specific value functions $W_{t+1}(X_{t+1}, h_{t+1})$:

1. For each point A_j of the grid over savings calculate the next period wealth M_{t+1} using (7). Note that M_{t+1} does not depend on M_t or c_t other than through $A_t = M_t - c_t$ which serves as a “sufficient statistic” for period t .
2. Using the policy function $c_{t+1}(X_{t+1}, h_{t+1})$ computed on the previous iteration, find the optimal levels of consumption for each h_{t+1} using one-dimensional interpolation over M_{t+1} .
3. Compute the right hand side (*RHS*) of the Euler equation (19) and the expected future value in (16) (*EV*) using Gaussian quadrature to evaluate the integral.
4. Compute the optimal level of current consumption as $c_t(X_t, h_t) = (u')^{-1}(\text{RHS})$ and plug it into (16) to also compute the current period $W_t(X_t, h_t)$.
5. Form the endogenous choice specific grids over current period wealth $\{M_1, \dots, M_J\}$ as $M_j = A_j + c_t(X_t, h_t)$. Period t iteration is complete, now both $c_t(X_t, h_t)$ and $W_t(X_t, h_t)$ can be computed by interpolation on the endogenous grids $\{M_1, \dots, M_J\}$.
6. If the endogenous grid $\{M_1, \dots, M_J\}$ is not monotone in spite of the fact that the grid over savings $\{A_1, \dots, A_J\}$ is ordered, call the *upper envelope* subroutine to eliminate the suboptimal points from the grid produced by the Euler equation solutions that correspond to local other than global maxima as described in (Iskhakov et al., 2017).
7. Continue with the next iteration $t - 1$ until reaching the initial period t_0 .

Notably, except for the terminal period, none of these steps above involve solving an optimization problem or computing solutions of an equation (due to the fact that the utility function in our model is analytically invertible). This is the primary advantage of the DC-EGM algorithm, which allows for fast and accurate solution of the dynamic programming problems with discrete and continuous choice.

5 Data

We estimate the model using the first 16 waves of the Household, Income and Labor Dynamics in Australia Survey (HILDA). This is a broad social and economic longitudinal survey which annually collects data on family and household dynamics, income and labor supply. The survey also includes several reoccurring modules on wealth, health and health insurance, retirement, fertility, literacy and numeracy among other topics (Wooden and Watson, 2007). The first wave of HILDA survey was administered in 2001 to a large national probability sample of Australian households which contained 19,914 people in total. New members of the selected households were added in subsequent waves. In 2011, 2,153 new households were added to replenish the sample.

We estimate the model using data on single and married men in Australia. We apply a number of screening filters related to the role of an individual in the household and the composition of the household to make sure the sample only contains male household heads, who are finished with their schooling and are part of the labor force. To ensure at least 100 data points at each age, we also restrict our data to the interval 19 to 89.

The resulting unbalanced panel contains 81,197 individual-age data points, which come from 10,133 individual households born between 1912 and 1997 and observed between 2001 and 2016. Because of the requirement that the key variables (education, hours of work) not to be missing, there can be gaps in the observed life cycles: 17.83% of life cycles were indeed observed with one or two of such gaps. The majority of the data comes from a single observation spell on individuals, of whom 17.99% were observed for all 16 years, and 12.40% were only observed once. However, our estimation approach does not require multiple observations on every person, so these observations were also included. Table 1 contains summary of the estimation sample, with detailed description of data preparation available in the Online Appendix.

Because we do not model schooling decisions, we make effort to exclude from the estimation sample the observations of the individuals who are still in school. We treat respondents as attending school at any age that is below the reported *age of leaving school*, or if the fraction of calendar time spent in full time education is above 50%. Moreover, for respondents under age 25 we allow for a “gap year” when they are treated as attending school if their schooling status in both the previous year and following year is positive or unknown.

HILDA has two records of hours of labor supply: number of hours worked in the week of the interview and in the “typical” week for the respondent. In addition, there is an extensive record of time use in the last financial year, which collects activity in the early, middle and late part of each month, for the period July through June. If a third of a month overlaps between

Table 1: Description of the structural estimation sample.

	College graduates	Highschool graduates	Highschool dropouts	All
Number of individuals	2,391 23.60%	5,254 51.85%	2,488 24.55%	10,133 100%
Number of data points	20,207 24.89%	41,965 51.68%	19,025 23.43%	81,197 100%
Average num. of obs. per individual	8.45	7.99	7.65	8.01
Num. of obs. age 16-30	2,828	7,458	2,110	12,396
Num. of obs. age 31-40	5,135	9,042	2,555	16,732
Num. of obs. age 41-50	4,957	9,405	3,633	17,995
Num. of obs. age 51-60	3,873	7,405	3,532	14,810
Num. of obs. age 61-70	2,135	4,735	3,127	9,997
Num. of obs. age 71-89	1,279	3,920	4,068	9,267
Non-missing <i>hours, education</i>	100.00%	100.00%	100.00%	100.00%
Non-missing <i>wage</i>	75.03%	65.77%	45.71%	63.38%
Non-missing <i>wealth</i>	22.68%	22.59%	22.14%	22.51%
Non-missing <i>super</i>	20.01%	17.66%	12.94%	17.14%

Notes: The table presents the summary statistics on the estimation sample. See Online Appendix for more detailed description of data preparation and filtering.

employment, unemployment or not in the labor force, then a half or a third is allocated to each as appropriate.

We adopt the following formula for calculating hours

$$\text{hours} = 50 \cdot (\text{typical week hours}) \cdot (\text{percent of time spent in all jobs during last year})/100.$$

In addition we censor the resulting value at 4200 (which corresponds to 12 hours of work per day 7 days per week) and correctly assign some of the missing values to zero hours (when the question was not asked because preceding replies indicated the respondent does not work).

The discrete levels of hours are derived by simple partitioning of observed distribution truncated below by $T_{min} = 500$ into K groups using K -median clustering algorithm. We chose $K = 6$ judging by how well the produced partitions matched the categorical labor market states reported in HILDA. The final discretization of hours and their correspondence to the categorical labor market states are presented in Table 2 and Table 3. Notice that zero hours observations include respondents in both out-of-labor-force (OLF) states, both unemployed states, but also employed states (when zero or little hours are reported).

We compute hourly wages on the individual level using the *last financial year gross wages*

Table 2: Discrete levels of hours of labor supply.

Hours in 6 levels			Median	Median		
per year	per week	N obs	per year	per week	Min	Max
0	0	15,305	0	0	0	500.00
1	1000	4,458	1200	24	500.03	1600.00
2	2000	15,894	2000	40	1600.02	2124.90
3	2250	5,460	2250	45	2133.36	2368.03
4	2500	8,725	2500	50	2375.10	2750.00
5	3000	6,248	3000	60	2750.10	4200.00

Notes: The table presents the discretized values of hours of labor supply. The discrete levels of hours were set to the medians in the intervals produced by the cluster analysis, with the exception of the part time hours which were set at 20 hours/week (1000 hours/year) instead of the median 1200 hours/year.

Table 3: HILDA classifications within hours levels.

Labor force status in HILDA	Hours in 6 discrete levels					
	0	1	2	3	4	5
Employed FT	244	864	15,778	5,460	8,725	6,248
Employed PT	1,254	3,594	116	0	0	0
Unemployed, looking for FT	1,245	0	0	0	0	0
Unemployed looking for PT	167	0	0	0	0	0
OLF, marginally attached	1,876	0	0	0	0	0
OLF, not marginally attached	10,518	0	0	0	0	0
Total	15,305	4,458	15,894	5,460	8,725	6,248

Notes: The table contains number of observations in each cross-tabulation cells. HILDA defines full-time work as 35 hours or more per week, based on usual hours. Marginal attachment to the labor force is determined by firstly establishing whether a person not in the labor force has a desire to work, and then by whether they have been actively seeking work or are available to start work within a short period of time.

and *salaries* variable in HILDA, for which missing values are imputed. We use the formula

$$\text{wage} = \begin{cases} \frac{\text{gross annual wage income}}{\text{hours}}, & \text{if hours} > 0, \\ \text{undefined} & \text{if hours} = 0. \end{cases}$$

In addition, we set wage to “undefined” if the calculated amount falls outside of the interval from \$5 to \$150.

Our wealth calculation relies on the HILDA wealth model and is based on the variables collected in the special module included in waves 2, 6, 10 and 14. We calculate wealth at the household level which includes the wealth of the spouse. We use the following formula

wealth = financial wealth + non-financial wealth – household debt – combined household super,

where the components are defined as (Summerfield et al., 2013, pp. 71-75).

The superannuation *profiles* are not used in the estimation. Because we make a simplifying assumption that accumulated superannuation is received as a lump sum payment at age 65, the only quantity we match in the data is the superannuation account balance at this age.

Finally, note that all monetary variables are converted to 2012 Australian dollars using the CPI index from the Australian Bureau of Statistics.

6 Estimation Method and Moment Conditions

To estimate the structural parameters of the model (see Table A in the Appendix A) we use the method of simulated moments (MSM), (McFadden, 1989). Assuming that the true parameter vector θ_0 lies in the interior of some parameter set $\Theta \in \mathbb{R}^{37}$, we obtain the estimate $\hat{\theta}$ by minimizing the sum of weighted distances between the points on the life cycle profiles for various variables that are observed in or computed from the data, and their simulated counterparts.

For each of the three education levels (high school, dropouts and college) we match the following life cycle profiles:

1. Fraction of population supplying positive amount of hours of labor (*working*);
2. Hours by those who are working (*hours conditional on working*);
3. Hourly wage for those working (*wage conditional on working*);
4. Variance of the wage for those working (*variance of wage*);
5. Skewness of the logarithm of wage earnings (*skewness of log-earnings*);
6. Fraction of population supplying 20 hours per week (*hours20*);
7. Fraction of population supplying 40 hours per week (*hours40*);
8. Fraction of population supplying 45 hours per week (*hours45*);
9. Fraction of population supplying 50 hours per week (*hours50*);
10. Total household wealth (*wealth*);
11. Fraction of population working positive hours in periods t and $t - 1$ (*work2work*);
12. Fraction of population not working in periods t and $t - 1$ (*nowork2nowork*).

In addition, we match one non-profile moment, namely the amount of superannuation at age 65. We aggregate wealth and higher order moments data in 5 years bundles, and disregard all the moments conditional on working after age 70. Moreover, we disregard age-education cells

Table 4: Number of empirical moments.

Moments	High school		Dropouts		College	
	Ages	N	Ages	N	Ages	N
<i>working</i>	19 - 86	67	19 - 88	70	23 - 89	64
<i>hours conditional on working</i>	19 - 70	48	19 - 70	48	23 - 70	44
<i>wage conditional on working</i>	19 - 70	48	19 - 70	48	23 - 70	44
<i>variance of wage</i> *	19 - 70	10	19 - 70	10	23 - 70	10
<i>skewness of log-earnings</i> *	19 - 85	13	19 - 85	13	23 - 85	13
<i>hours20</i>	19 - 86	67	19 - 86	68	23 - 89	64
<i>hours40</i>	19 - 82	61	19 - 84	64	23 - 79	57
<i>hours45</i>	19 - 77	55	19 - 83	56	23 - 76	51
<i>hours50</i>	19 - 76	58	19 - 88	66	23 - 77	53
<i>wealth</i> *	25 - 85	13	25 - 85	13	25 - 85	13
<i>work2work</i>	19 - 70	48	19 - 70	48	23 - 70	44
<i>nowork2nowork</i>	19 - 70	48	19 - 70	48	23 - 70	44
<i>super</i>	65	1	65	1	65	1
Total		537		553		502

Notes: The table presents the number of empirical moments by education for various life cycle profiles, and the age spans when these moments are observed. Note that within the given spans some moments may be missing. The total number of all moments in the data is 1977. *Wealth and higher moments data is aggregated in 5 year intervals.

that contain less than 10 observations of any given moment. In total we have 1592 moment conditions. The exact number of moments per one life cycle profile are shown in Table 4.

Denote $\mu(\theta) = \left(\mu_1^{hs}(\theta), \mu_1^{dr}(\theta), \mu_1^{cl}(\theta), \dots, \mu_{12}^{hs}(\theta), \mu_{12}^{dr}(\theta), \mu_{12}^{cl}(\theta) \right)$ the vector of 1592 moment conditions which is composed of stacked vectors of moments conditions $\mu_i^e(\theta)$ for all education levels $e \in \{\text{highschool, dropouts, college}\}$ and each of the $i \in \{1, \dots, 12\}$ life cycle profiles, with $i = 13$ referring to scalar superannuation moment. The MSM estimator is

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \frac{N}{1 + N/N_{sim}} \mu(\theta)' W \mu(\theta), \quad (22)$$

where N is the number of independent observed individual life cycles, N_{sim} is the number of simulated life cycles, and W is the square weighting matrix.

We solve the optimization problem (22) numerically using the following standard steps:

0. Compute the empirical moments, fix all random number generator seeds;
1. Given a trial value of parameter vector θ , analyze which components of θ have been updated and accordingly whether the dynamic programming model needs to be re-solved, and whether the initial conditions need to be re-generated;
2. Solve the dynamic programming model if needed, and update the decision rules $c_t^*(X_t)$

and $h_t^*(X_t)$, and corresponding choice probabilities $P_t(h^{(i)})$;

3. Simulate the initial conditions (initial wealth M_0) using dedicated set of random numbers, if needed;
4. Simulate life cycle profiles using the current initial conditions and current decision rules that constitute the solution of the dynamic programming problem for the values of relevant parameters in θ ;
5. Compute the moment conditions, i.e. the distance between the empirical and simulated moments, evaluate the objective function by computing the weighted sum (22), and check the stopping criteria. Unless converged, find next trial value of parameter vector θ in the outer loop minimization algorithm and return to step 1. Once the stopping criteria are satisfied, the current value of θ constitutes the MSM estimate $\hat{\theta}$.

The initial conditions other than initial wealth M_0 are fixed throughout the estimation. We use the distribution of educations observed in the data, and assume that the initial amount of work experience is zero, $\mathcal{E}_0 = 0$.

We use POUNDerS derivative free trust region optimization algorithm to compute the next trial value of the parameter vector in step 5 of the algorithm (Munson et al., 2012; Balay et al., 2018). The stopping criteria tolerance is 10^{-5} . Because the gradient information for the objective function is not available, POUNDerS relies on the gradient of the approximating model in the trust region to terminate the algorithm upon convergence.

In the absence of gradient information of the objective function in (22), the optimization algorithm relies exclusively on the functional value. Meanwhile, because the labor supply choices in the model are discrete, simulated moment conditions and therefore the objective function are piece-wise flat. To prevent the maximization loop from “stalling” in the flat areas, we modify the simulation algorithm in step 4 by adding small additional extreme value noise to the underlying values, which leads to both smoothing over discontinuities and to slight incline to the otherwise flat moment conditions. Effectively, we replace the maximum operators that produce simulated discrete choices with logsum functions that produces “smoothed choices” that can be represented by near integer numbers we term *smoothed weights*. Our approach is closely related to Bruins et al. (2018) with a notable innovation that the scale of the additional extreme value noise is expressed relative to the absolute values of the alternative utilities, and therefore does not have to be readjusted. Appendix C provides complete description of our approach.

The term “smooth weight” reflects the nature of the smoothly simulated discrete choice. Instead of simulated zeros and ones to denote particular chosen alternatives, the smooth simulator

produces near zeros and near ones. All quantities dependent on the chosen options are then interpolated using these smooth weights. Because the utilities of discrete choices calculated when the dynamic programming problem is solved are typically smooth functions of structural parameters, with the smooth simulator the objective function in (22) becomes smooth as well.

Assuming that the regularity conditions in McFadden (1989) and Pakes and Pollard (1989) hold, the MSM estimator $\hat{\theta}$ is consistent and asymptotically normally distributed, i.e.

$$\sqrt{N}(\hat{\theta} - \theta_0) \sim N(0, \Sigma). \quad (23)$$

The variance-covariance matrix Σ can be approximated by the sample analogues of the corresponding theoretical components

$$\hat{\Sigma} = (1 + N/N_{sim})(D'WD)^{-1}D'WSWD(D'WD)^{-1}, \quad (24)$$

where S is the variance-covariance matrix of the moment conditions $\mu(\theta)$, and $D = \frac{\partial \mu(\theta)}{\partial \theta}$ is the Jacobian matrix of the moment conditions computed at some consistent estimate $\hat{\theta}$, and W is the weighting matrix from (22).

The optimal choice of the weighting matrix is $W = S^{-1}$ in which case (24) collapses to $\hat{\Sigma} = (1 + N/N_{sim})(D'WD)^{-1}$, but because the finite sample approximation of W can be severely biased (Altonji and Segal, 1996), we follow the literature by using a “diagonal” weighting matrix which accounts for the variance of the moment conditions but ignores their correlation structure.

In addition, we correct for the excess variability over time of the moments that form life cycle profiles. Denote n_t the number of observations for particular period t of a given moment profile, d_t the empirical moment at period t and $s_t(\theta)$ its simulated counterpart. The moment condition is then given by $\mu_t(\theta) = d_t - s_t(\theta)$. Under the assumption of orthogonality between sampling errors and the simulation errors, the variance of the moment condition can be decomposed as

$$Var(\mu_t(\theta)) = Var(d_t) + Var(s_t(\theta)). \quad (25)$$

Asserting that $Var(s_t(\theta))$ is time-invariant, and summing over time, we have for every $t = 1, \dots, T$

$$\sum_{\tau=1}^T Var(\mu_{\tau}(\theta)) = \sum_{\tau=1}^T Var(d_{\tau}) + T \cdot Var(s_t(\theta)). \quad (26)$$

Solving (26) for $Var(s_t(\theta))$ and plugging the resulting expression back into (25) gives the formula

for corrected variance of moment condition

$$Var(\mu_t(\theta)) = Var(d_t) + \frac{1}{T} \sum_{\tau=1}^T Var(\mu_\tau(\theta)) - \frac{1}{T} \sum_{\tau=1}^T Var(d_\tau). \quad (27)$$

We estimate variances in the right hand side of (27) using cross-section sample variation, and use the corrected estimates of $Var(\mu_t(\theta))$ to construct the diagonal weighting matrix W .

7 Estimation results

7.1 First stage estimates

Some parameters of our model can be consistently estimated separately from the full structure, and then treated as known within the structural estimation. These are the survival curve, the age pension rule and the tax function. We summarize these "first stage" estimation results here, while the online Supplement provides full detail.

We use current Australian life tables Actuary (2009) to quantify the survival process in the model. However, to reduce computational burden, we assume that decision makers do not take the possibility of death into account before the age of 40. Therefore, we fit the following parametric model to the conditional death probabilities for males between ages 19 and 90 (standard error in parenthesis)

$$\delta_t = \begin{cases} 1 & \text{if } age_t < 40, \\ 1 - \frac{0.0006569}{(0.0000249)} \left[\exp \left(\frac{0.1078507}{(0.0008145)} (age_t - 40) \right) - 1 \right] & \text{if } age_t \geq 40, \end{cases} \quad (28)$$

The model is estimated with the method of non-linear least squares, the estimation errors are of the order 10^{-5} to 10^{-3} up to age 90 (Root mean squared error 0.001137). This is relatively good fit for the ages of primary concern for the labor supply model.

We use HILDA records on Australian age pension receipt, which is recorded on the individual level. To avoid the problem that pension formulas that vary slightly through the years (see Figure 1 in Section 2), we estimate a statistical model of pension payments using the available data and the covariates that enter the structural model. We attempted a number of parameterized functional forms that resembled the legislated pension equation. The preferred specification is the closest to the actual pension rules with separate asset and income tests

$$\text{pension} = \max \left\{ \theta_1 + \theta_2(\text{year}) - \max \left[0, \max \{ \theta_3(\text{wage} - \theta_5), \theta_4(\text{wealth} - \theta_6) \} \right], 0 \right\},$$

where variable names are self-explanatory, θ is a vector of parameters with $\theta_2(year)$ denoting a set of year fixed effects.

Because the wealth data essential for the estimation of the asset test in the pension equation is only collected in waves 2, 6, 10 and 14, only three fixed effects could be estimated. Following the description of the age pension reform in 2009 (see Section 2) we further restrict the model and only estimate one dummy coefficient for $year \geq 2010$. Also, parameter θ_5 had a tendency to take on negative values, so we fixed it at the economically sensible value $\theta_5 = 0$. The preferred estimated pension equation that is used in the structural model is (standard errors in parentheses) then:

$$\begin{aligned} \text{benefit}_{\max} &= 10,759.73 + 1,846.92(\text{when } year \geq 2010), & (29) \\ &\quad (183.96) \quad (173.52) \\ \text{pension} &= \max \left\{ \text{benefit}_{\max} - \max \left[\max \left\{ \underset{(0.020)}{0.27794} \text{ wage}, \underset{(0.0004)}{0.00499} (\text{wealth} - \underset{(46,895.27)}{117,082.60}) \right\}, 0 \right] \right\} \end{aligned}$$

The fact that the pension rule changed in 2010 is a major issue for structural estimation. To account for the change in the policy we solve the dynamic model separately for each policy regime, so that the shift in the benefit level can be treated as unanticipated in the simulations of the moments that enter moment conditions $\mu(\theta)$ and MSM objective function (22). The simulation is started with the optimal decision rules corresponding to the policy regime prior to 2010. Then, once the individual reaches the 2010 calendar year in their life-cycle, the simulator switches to the optimal decision rules corresponding to the policy regime after 2010. Obviously different people are at different ages when the shift occurs.

Similar to the pension equation, we estimate income tax function using the data available in HILDA instead of applying tax rules directly. We choose this approach for two reasons. First, as the model only addresses a limited set of socio-economic variables relevant for the computation of the income taxes, the complexity of the tax rules can not be replicated. Second, while there is some variability of tax rules over time, fortunately for us the tax structure was very stable over our sample period, both in terms of rates and the bracket structure.

Using non-linear least squares we estimated three bracket tax models using a selected sample of individuals with small non-labor earnings (less than \$500) and positive wage earnings not

Table 5: Estimates of the preference parameters.

Parameter	Description	Estimate	Std.Err.
ζ	CRRA coefficient in consumption	0.79488	0.07327
γ_1	Disutility of working 1000 hours (20 per week)	1.4139	0.38508
γ_2	Disutility of working 2000 hours (40 per week)	2.0088	0.59712
γ_3	Disutility of working 2250 hours (45 per week)	2.9213	0.78915
γ_4	Disutility of working 2500 hours (50 per week)	2.8639	0.80946
γ_5	Disutility of working 3000 hours (60 per week)	3.8775	1.05032
κ_1	Correction coefficient for low type with disutility of work	0.50321	0.17973
κ_2	Quadratic coefficient on age for older workers	0.00008	0.00004
κ_3	Linear coefficient on age for young workers	0.05083	0.01554
ξ	CRRA coefficient in utility of bequest	0.48834	0.34766
b_{scale}	Scale multiplier of the utility of bequest	0.68659	1.42044
β_{clg}	Discount factor, college	0.96963	0.00238
β_{hs}	Discount factor, highschool	0.96732	0.00189
β_{dr}	Discount factor, dropouts	0.96806	0.00138
λ	Scale of EV taste shocks	0.29950	0.08825

exceeding \$200,000. The preferred estimation results are given by

$$\text{tax} = \begin{cases} 0, & \text{if income} < \text{thld}_1 = 17.39184, \\ & \quad \quad \quad (0.10016) \\ 0.29907 \cdot (\text{income} - \text{thld}_1), & \text{if } \text{thld}_1 \leq \text{income} < \text{thld}_2, \\ (0.00096) & \\ 0.37930 \cdot (\text{income} - \text{thld}_2) + 0.29907 \cdot \text{thld}_1, & \text{if income} \geq \text{thld}_2 = 73.17661, \\ (0.00556) \quad \quad \quad (0.00096) & \quad \quad \quad (0.59292) \end{cases} \quad (30)$$

where the variable labels are self-explanatory.

7.2 Structural estimates

The estimates of the structural parameters are reported in Tables 5 to 7. Table 5 contains estimates of the preference parameters. The CRRA parameter for consumption in equation (11), ζ , is estimated to be 0.79. As this is slightly smaller than 1.0 it implies utility is slightly less concave than logarithmic in consumption. In simple versions of the life-cycle model this would imply that the substitution effect of wages on labor supply would slightly dominate the income effect. But this is not necessarily the case here, given that we have endogenous human capital and pensions that are a nonlinear function of earnings.

The disutility of work hours parameters from equation (13), $\gamma = (\gamma^{(1)}, \dots, \gamma^{(5)})$, imply that disutility is increasing in hours. But the increase is not monotonic, as the estimated disutility

Table 6: Estimates of the parameters of human capital accumulation process.

Parameter	Description	Estimate	Std.Err.
$\eta_{0,clg}$	Constant for college	2.78766	0.41169
$\eta_{0,hs}$	Constant for high school	2.56761	0.36634
$\eta_{0,dr}$	Constant for dropouts	2.45647	0.33269
$\eta_{0,high}$	Constant for high type	0.39311	0.41893
$\eta_{1,clg}$	Work experience for college	0.03041	0.00796
$\eta_{1,hs}$	Work experience for high school	0.02164	0.00768
$\eta_{1,dr}$	Work experience for dropout	0.01974	0.00682
$\eta_{2,clg}$	Work experience square for college	-0.00017	0.00021
$\eta_{2,hs}$	Work experience square for high school	-0.00002	0.00018
$\eta_{2,dr}$	Work experience square for dropout	0.00000	0.00010
η_3	Age (time index)	0.02676	0.00280
η_4	Age (time index) square	-0.00076	0.00004

of working 50 hours per week is slightly less than that of working 45. However, recall that in our model the γ reflect not only the disutility of working the associated level of hours, but also the extent to which jobs with each level of hours are available.

Next we report the κ parameters from equations (14)-(15). The estimate of κ_1 implies that the disutility of work is roughly 50% greater for the “low” type. The estimate of κ_2 implies the disutility of work is increasing with age after age 40, while κ_3 implies the disutility of work declines between the ages of 16 and 25.

The bequest function parameters in equation (12), b_{scale} and ξ , are imprecisely estimated, which is similar to what Imai and Keane (2004) found. This may be due the large amount of measurement error in assets, and the fact that behavior at younger ages is close to invariant over a wide locus of bequest function parameters. Nevertheless, the behavior of the model changes greatly if bequests are ignored entirely, suggesting bequests are important, but that the precise shape of the bequest function is very difficult to pin down.

Note that the discount factor is precisely estimated, with values of roughly .968 for the dropouts and high school types, and .97 for the college types. Finally, we report the scale λ of the extreme value taste shocks in equation (18). This parameter is important in distributing workers across the 6 discrete hours levels.¹⁹

Table 6 contains estimates of the human capital production function parameters. The wage functions intercepts for the three education types differ as expected. For example, the offer wage premium for college vs. HS, evaluated at zero work experience, is 0.22 log points, implying 25%

¹⁹Note that as λ approaches infinity workers are more equally distributed across hours levels .

Table 7: Estimates of the other parameters.

Parameter	Description	Estimate	Std.Err.
ς_0	St.dev. in shock distribution: constant	0.24485	0.24055
ς_1	St.dev. in shock distribution: age	0.00421	0.00935
tr	Transfer from parents	5.51308	1.43804
ρ_{clg}	Superannuation: human capital college	6.30347	2.58472
ρ_{hs}	Superannuation: human capital high school	5.43473	3.30737
ρ_{dr}	Superannuation: human capital dropouts	6.47838	3.95647
ς_{w_0}	Initial wealth sigma	0.14896	6.69399
p_{clg}	High type proportion college	0.90089	0.04952
p_{hs}	High type proportion high school	0.80130	0.04366
p_{dr}	High type proportion dropout	0.69306	0.04411

higher initial offer wages.

The intercept shift for the “high” skill type, $\eta_{0,high}$ is .393, implying $\exp(.393) = 48\%$ higher offer wages, *ceteris paribus*. But this parameter is very imprecisely estimated, which is quite unusual for these type of models (see Keane and Wolpin (1997) and other related mixture models, where offer wage differences between types are very precisely estimated). Thus, we spent a great deal of time exploring this phenomenon and came to the following conclusion: As will be clear later when we look at the income distribution, it has both fat left and right tails. We find that our model has some trouble capturing both tails with two types. One can vary $\eta_{0,high}$ over a fairly wide range (e.g., from about 0.30 to 0.50) with little impact on overall model fit, except that for lower values of $\eta_{0,high}$ one obtains a better fit to the left tail, while for higher values one obtains a better fit to the right tail. The high standard error reflects the fact that our search algorithm is unsure where to place $\eta_{0,high}$ within this large (rather flat) part of the objective function surface. But setting $\eta_{0,high} = 0$ would result in a serious deterioration in fit.²⁰

The next six rows of Table 6 report quadratics in work experience, which capture endogenous variation in human capital over the life-cycle arising from learning-by-doing. The effects of experience are allowed to differ for the three education groups. Consistent with results in Keane and Wasi (2016) and Imai and Keane (2004), the estimates imply that returns to work experience are greater for more educated workers. For example, the coefficient on experience is 0.0304 for the college type and only .0216 for HS type. Thus, endogenous human capital accumulation over the life-cycle is more important for more educated workers.

²⁰We are not too concerned about this phenomenon because (i) the fit to the overall wage and income distributions appear to be rather good, and (ii) it is obvious one needs at least two types with different mean offer wages to fit these distributions. Our search algorithm simply has trouble deciding how different to make offer wages for those two types so as to obtain the best fit.

The last two parameters in Table 6 are a quadratic in age, which captures exogenous variation in offer wages over the life-cycle. Notably, the pure quadratic in age is also highly significant and quantitatively important. It implies that offer wages follow an inverted U-shaped path in age (just as they do with work experience). Initially the age effect is positive 2.67% per year, but the age quadratic peaks at 18 years into the working life (and then declines). This age effect on offer wages may capture, in a reduced form way, a host of factors that are not picked up by work experience alone. For example, in the early part of the life-cycle, increasing age may reflect increases in general life skills that improve worker productivity. In the later part of the life-cycle the age quadratic may pick up declines in health and/or stamina that reduce productivity. As noted by Keane and Wasi (2016), it is important to control for age itself so as not to force the model to explain all life-cycle variation in wages by work experience alone, thus potentially exaggerating returns to experience (which in turn could distort estimates of labor supply elasticities).

Table 7 contains estimates of the remaining structural parameters. We start with the parameters ς that determine how the standard deviation of the wage shocks vary with age in equation (2). The estimates imply a U-shaped path. This is similar to the reduced form pattern found by Geweke and Keane (2000). Next, we report the estimated transfer from parents, which we assume individuals receive annually between t_0 and 23. It equals \$5.5k. The next three rows report the superannuation process parameters from equation (9). These imply, as expected, that the amount of superannuation accumulation is an increasing function of the level of accumulated human capital.

Table 7 also reports the estimate of ς_{w0} , which is the standard deviation of the (log of) initial wealth (in \$1000). This is 0.15 but very imprecisely estimated. This parameter is hard to pin down because we do not see the distribution of initial wealth directly. It is inferred implicitly from heterogeneity in assets at later ages (in particular, the “excess” heterogeneity that is difficult to explain given observed earnings).

Finally, Table 7 contains estimates of the type proportions conditional on education. Interestingly, the probability a person is the “high” type is about 90% among the college educated, 80% among the HS educated and 69% among the high school dropouts. The implication is skill heterogeneity is greater among the less educated. This enables the model to explain why some people with low education nevertheless obtain high wages.

8 Fit of the Model

In this section we describe the fit of the model to several key dimensions of the data. These are hours, employment, wages, earnings and assets.

Figure 3 shows the fit of the model to average annual hours over the life-cycle. The average includes both working and non-working individuals. Notice that the model provides a good fit to the life-cycle path of hours for high school workers. In particular it captures that hours rise at young ages, then stay very flat from the late 20s through the late 40s, and then decline steeply in the 50s and 60s.

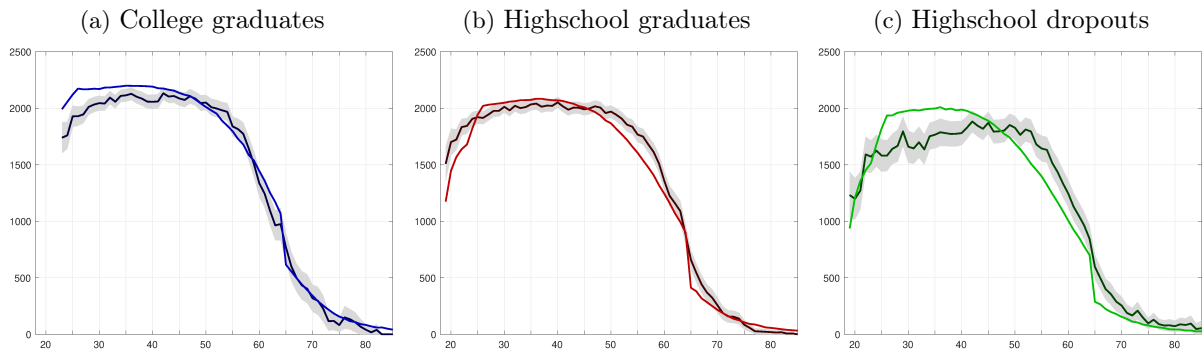
The fit for college workers is not quite as good because the model predicts their hours are too high at young ages. Similarly, the model predicts that the dropouts work too much from about age 25 to 40, less well than the other two groups. But despite these shortcomings the model matches the basic shape and levels of the life-cycle hours paths fairly well for all three groups.

Figures 4 and 5 decompose total hours into the extensive margin (i.e., employment) and the intensive margin (i.e., hours conditional on employment). The model captures the nature of this decomposition quite well for all three groups. As we see in Figure 5, hours conditional on employment are fairly flat over the life-cycle for all three education groups. For example, for high school workers, average hours are about 2250 per year at age 40, and they only drop to about 1800 hours per year at age 70. As we see in Figure 4 it is sharp drop in employment after age 50 that account for most of the declines in total hours that we see in Figure 3.

Recall that in our model agents can choose to work 20, 40, 45, 50 or 60 hours per week (where these values were chosen because they provide a good fit to the observed bunching of hours in the data). Figure 6 reports how the model fits the number of workers at each hours level, separately by education level. The model provides a very good fit to the distribution of workers to hours levels by age with one notable exception. Specifically, in the data we see that when people reach their early to mid 60s there is a jump in the number who work part-time (i.e., 20 hour per week). This pattern exists for all three education groups, but is most pronounced college graduates. Thus, it appears as if a sizable number of people transition to part-time work shortly before they retire. The model fails to generate this pattern, However, we did not feel this shortcoming was serious enough to complicate the model by adding parameters designed specifically to capture this feature of the data. Thus, we leave this as a topic for future research. The only other evident shortcoming of the model is that it seriously understates the fraction of high school dropouts who continue to work 60 hours per week in their 50s and 60s.

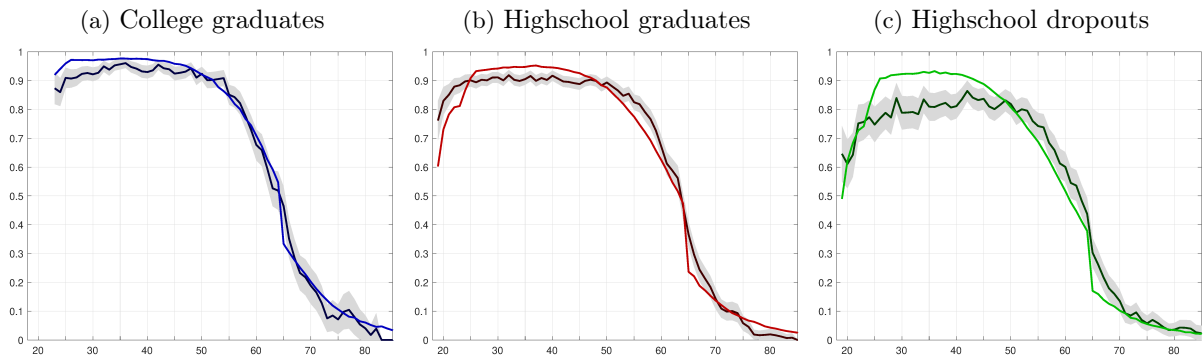
Figure 7 reports the fit of the model to the mean, variance and skewness of wages of employed

Figure 3: Fitted hours of labor supply.



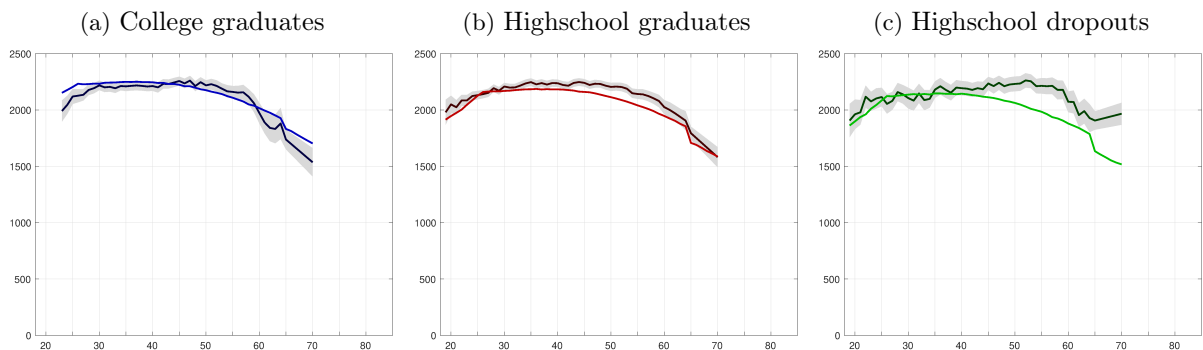
Notes: The plots show simulated life cycle profiles of hours of labor supply for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 4: Fitted labor force participation.



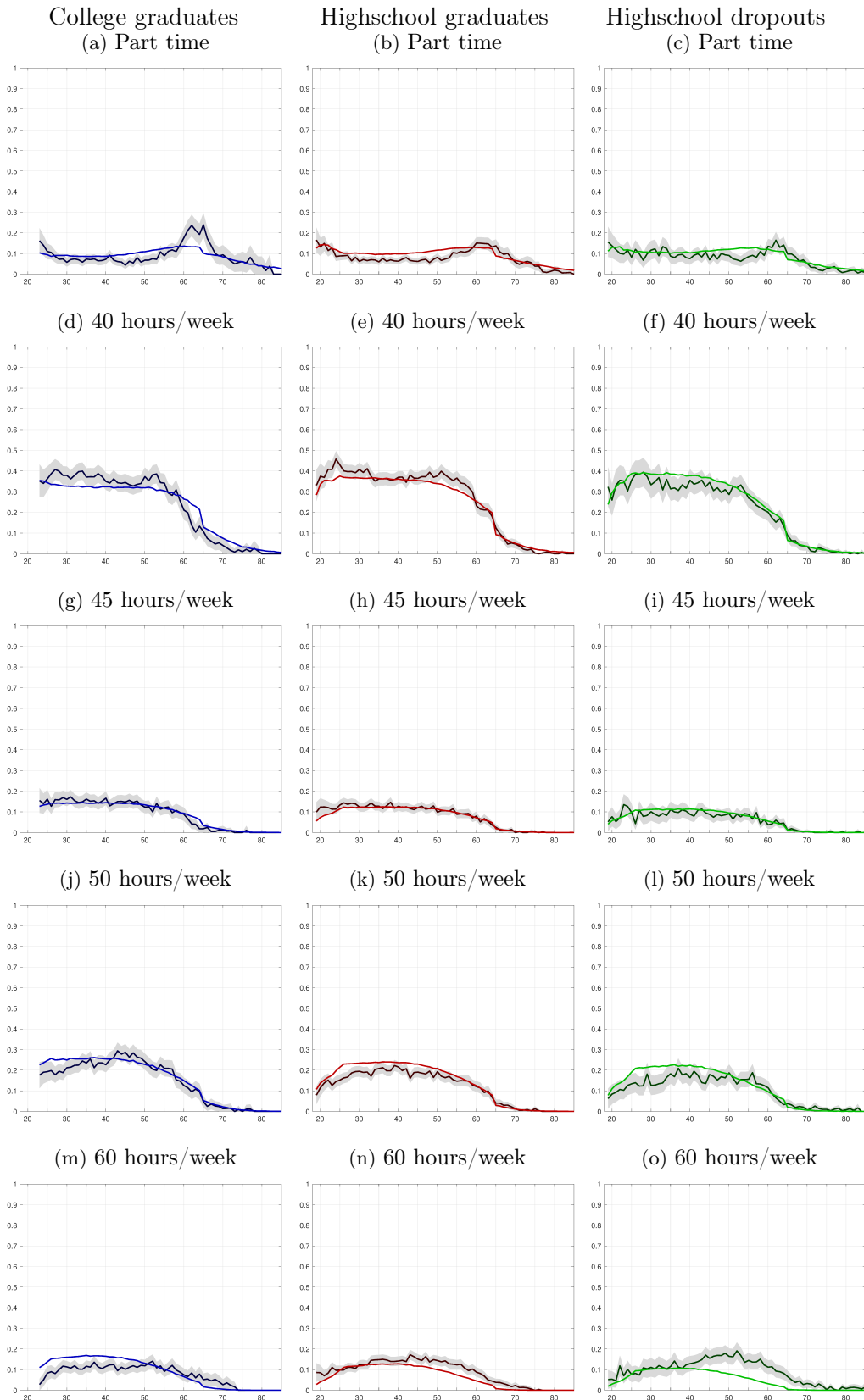
Notes: The plots show simulated life cycle profiles of fraction of working people for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 5: Fitted hours conditional on working.



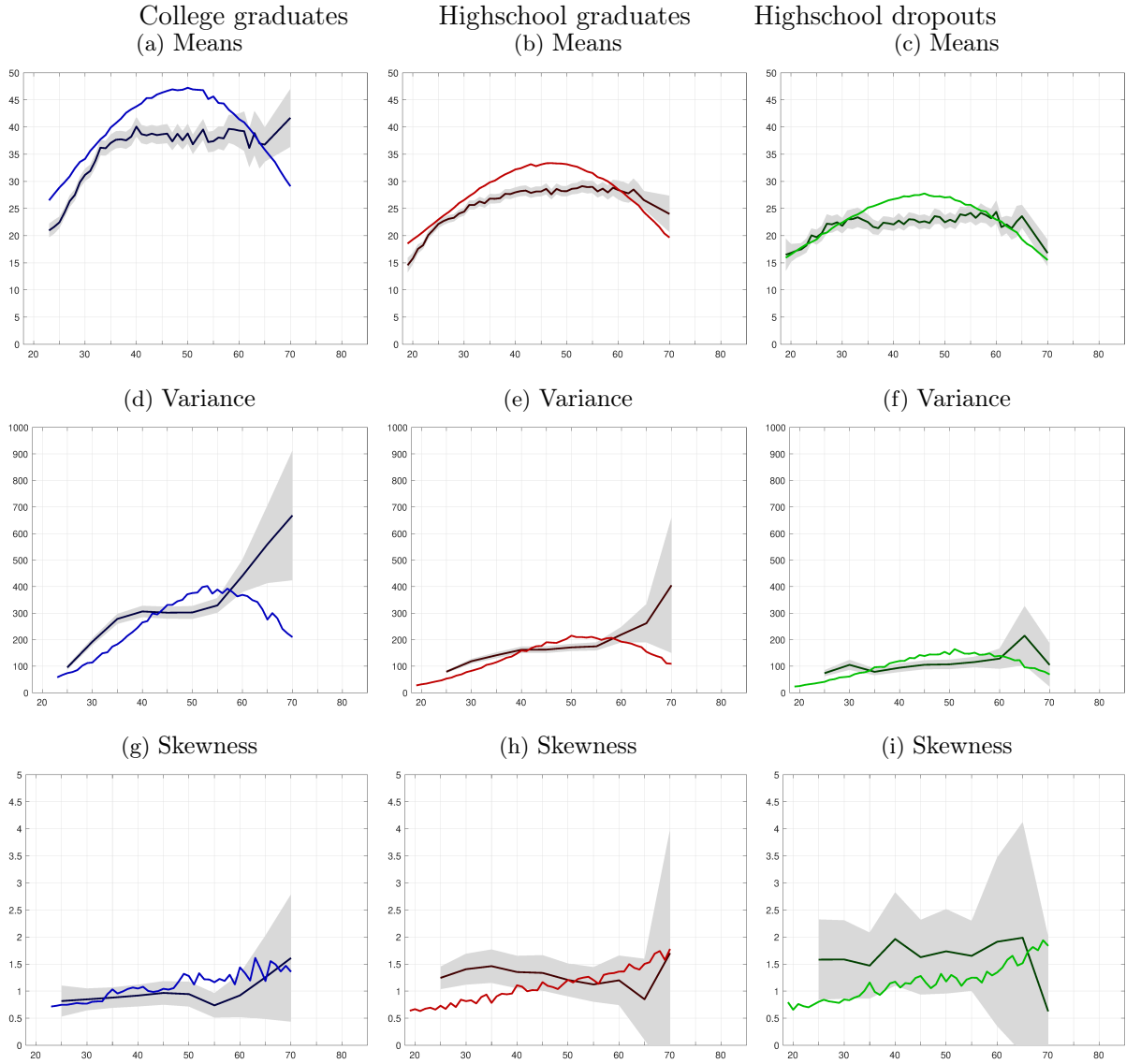
Notes: The plots show simulated life cycle profiles of hours conditional on working for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 6: Fitted fractions of population working particular levels of hours.



Notes: The plots show simulated life cycle profiles of fractions of people working particular number of hours for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 7: Fitted wage profiles.



Notes: The plots show simulated life cycle profiles of wages for the three education groups against the profiles observed in the data. Top row shows mean, second row shows variance, and the third row skewness of simulated wages. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean. Outliers in the top and bottom 1% of wage observations are omitted.

workers, broken down by education and age. In our view, if one wants to credibly evaluate the impact of social insurance programs like the aged pension, it is important that one's model matches not only means but also higher moments of the wage distribution.

The first row of Figure 7 reports results for average wage profiles. Clearly there are details of the life-cycle wage path for each education that the model fails to fit precisely. However, the big picture is that the life-cycle wage paths are very different for the three education groups. For dropouts wages only rise very slightly from when workers first enter the labor market until their life cycle peak. But for college workers wages roughly double. The high school workers are in between. The model does a very good job of capturing the broad differences between the three education groups. Obviously this is primarily because the estimates imply greater returns to experience for more educated workers.

On the other hand, it is striking that in the HILDA data all three education groups experience rapid wage growth at young ages, but this flattens out dramatically after about age 35. It thus seems that quadratics in age and experience do not capture wage paths in Australia as well as they do wage paths in the US or UK. We leave the fundamental source of this difference as a subject for future research.

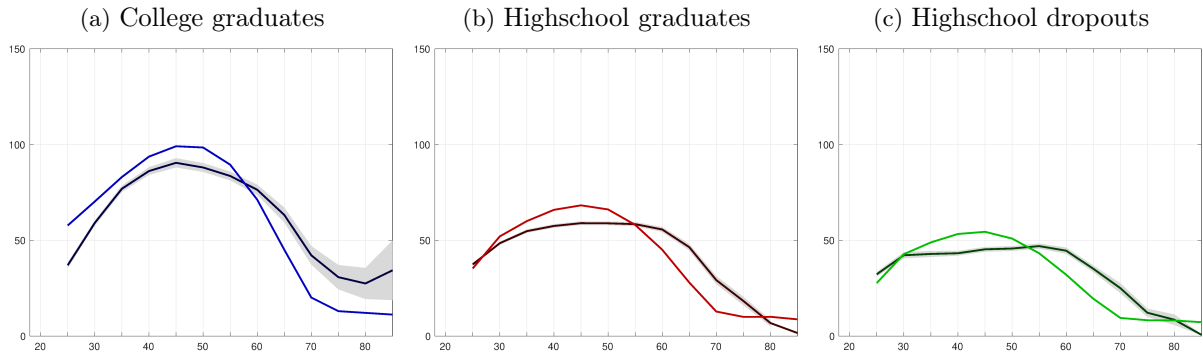
The second row of Figure 7 reports results for the variance of observed wages. As with means, one can see details of the life-cycle variance paths that the model fails to fit precisely. However, the big picture is that the life-cycle variance paths are very different for the three education groups. Variance rises with age for all three groups, but the increase is much sharper for the college education. The model successfully captures that difference, as well as the level differences across groups.

Finally, the third row of Figure 7 reports results for the skewness of observed wages. As third moments are difficult to estimate, the 2 standard deviation confidence bands for skewness in the actual data are wide. The model's predictions generally fall within these confidence bands, except that skewness appears to be understated for HS workers in their 20s and 30s.

Figure 8 reports on the evolution of average labor earnings over the life-cycle. The paths in the data differ sharply by education group. For college workers there is a very clear inverted U-shape over ages from 25 to 70, and the model captures this fairly well. For HS workers there is also a hump shape over the life-cycle, but it is much flatter. The model generates this much flatter profile. For the dropouts the earnings path is almost flat after the first few years, and the model approximates this with a very flat hump shape. These figures suggest that labor market returns to experience are much greater for more educated workers.

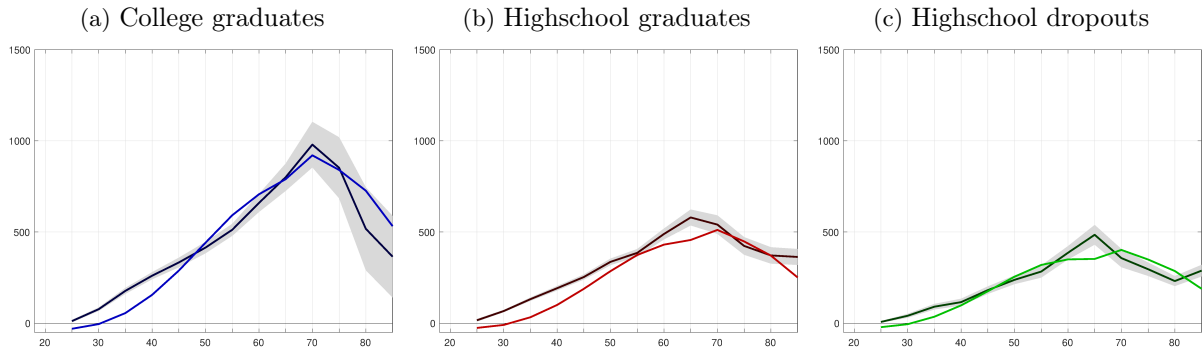
Figure 9 reports on the evolution of wealth over the wealth cycle. The model provides a

Figure 8: Simulated and observed wage earnings.



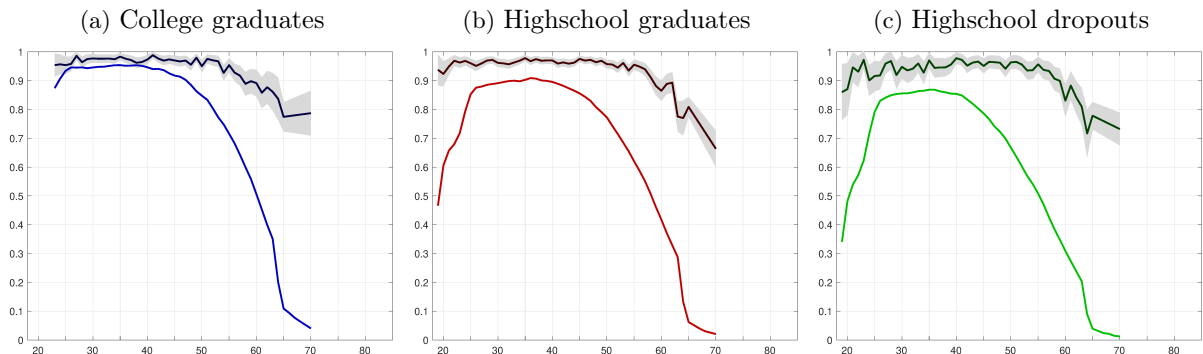
Notes: The plots show simulated life cycle profiles of wage earnings for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 9: Fitted wealth profiles.



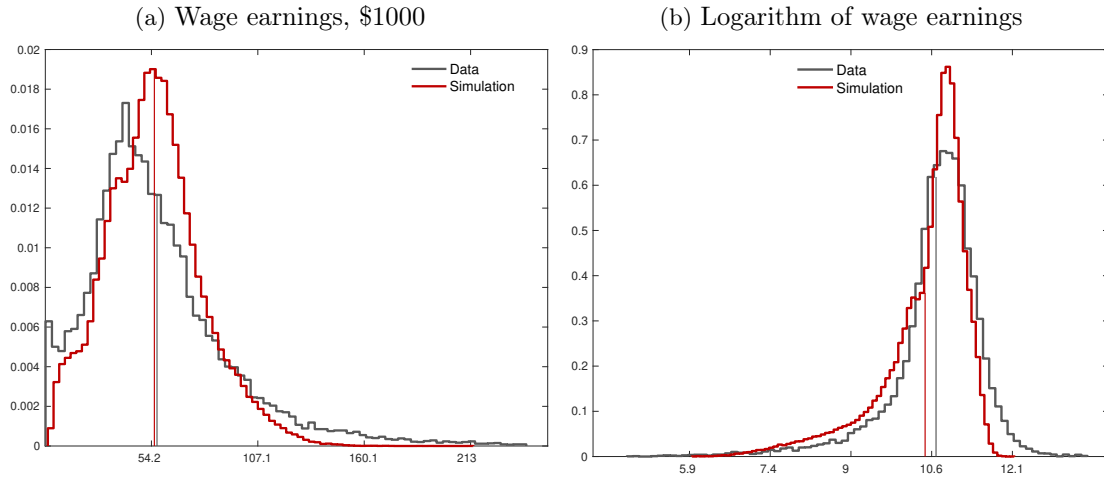
Notes: The plots show simulated life cycle profiles of assets for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 10: Fitted work to work transition rates.



Notes: The plots show simulated life cycle profiles of transition rates from work to work for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 11: Simulated and observed income distributions.



Notes: The plots show empirical probability density functions (histograms) for the observed and simulated before tax earnings in logs (panel b) and levels (panel a). Distribution of observed earnings in panel a is truncated at \$240,000.

remarkably good fit to the life-cycle wealth paths for all three education groups. At age 70 the mean wealth of college workers is roughly \$1M, that of high school workers is roughly \$500k, and that of dropouts is roughly \$400k. The model fits these level differences in wealth across education groups very precisely, as well as accurately capturing the shape of the life-cycle wealth paths. It is encouraging that we fit life-cycle wealth paths well, as this is obviously crucial for assessing the impact of the Aged Pension on older workers.

Figure 10 describes how the model fits the transition rate from work to work. The model does generate a high degree of persistence in the employed state, but nevertheless it consistently understates the very high degree of persistence in the data. Of course, it is well known since Keane and Wolpin (1997) that dynamic labor supply models have a tendency to understate persistence in employment, and also that one can fix this problem by adding features like additional heterogeneity and dependence of tastes for work and/or job offer probabilities on lagged employment. But such changes would substantially increase the size of the state space and hence computational burden. However, for our policy question, we feel that matching persistence in employment is of secondary importance compared to matching profiles of wealth and human capital. Put simply, we expect that the effect of changing aged pension rules on worker behavior could potentially differ greatly depending on the levels of assets and work experience that workers have at age 65, but we doubt that the degree of persistence in their employment history is likely to be an important factor. Thus, we chose not to complicate the model in an attempt to better match transition rates.

Finally, Figure 11 shows the population cross-sectional distribution of labor earnings, aggregated across all education levels and ages, both in the HILDA data and in our model. The left panel shows the distribution of the level of earnings, which has the expected Pareto shape, while the right panel shows the distribution of log income, which is strongly left-skewed relative to a normal. Our model generates a reasonably accurate approximation to the earnings distribution - viewed either way. We feel this is very important. In evaluating the impact of a social insurance program like the aged pension, face validity requires that one's model generate a baseline earnings distribution similar to that observed in the data.²¹

We can also see in Figure 11 that the actual earnings distribution has quite a complex shape, with both fat left and right tails. This is difficult to fit perfectly with just two types. As we mentioned in Section 7.2, the model estimates of skill differences between high and low types are a compromise that attempts to fill out both tails. In the levels graph we see that the model fails to generate the puzzling phenomenon of a small number of people with extremely low earnings (i.e., just a couple thousand dollars a year or less). But in the logarithmic graph we see that, overall, the model actually puts a bit too much mass in the left tail. We are not too concerned with the model's failure to explain the mass of people with extremely low incomes, as we suspect these arise largely from measurement error (e.g., people who only work a fraction of a year).

9 Policy simulations

In this section we report two sets of policy simulations. First, in Section 9.1, we examine the model's predictions with respect to changes in wage/tax rates. This enables us to gain a better understanding of the behavioral implications of our model, and to place its predictions in the context of the existing labor supply literature. Then, in Section 9.2, we turn to our main policy simulations designed to assess how the aged pension affects labor supply, savings and human capital over the life cycle. We also assess the impact of the aged pension on the welfare of the heterogeneous agents in the economy.

9.1 Labor Supply Simulations

In Figure 12, we report simulated responses to transitory wage changes. Specifically, we simulate 10% reductions in the wage rate that last for exactly one year. We compare effects of wage changes that occur at ages 30, 40, 50, 60, 65 and 70, to see how elasticities vary by age. Figure 12

²¹For example, if one's model generated far too few low income people compared to the actual data, a finding that social insurance programs are not welfare improving would lack any face credibility.

reports effects of fully anticipated transitory wage changes, so we obtain pure Frisch elasticities.²²

As we see in Figure 12, our model implies that Frisch elasticities increase with age, which is consistent with the pattern predicted by the much simpler life-cycle model in Imai and Keane (2004).²³ For instance, the Frisch elasticity is about 0.20 for college types at age 30, but it grows to roughly 1.0 at age 65.

Furthermore, Figure 12 also shows that the degree to which Frisch elasticities increase with age differs by human capital level, as predicted in Keane (2016) and found empirically in Keane and Wasi (2016). For high school dropouts, for whom human capital accumulation over the life cycle is not substantial, the Frisch elasticity is flat at about 0.65 at all ages. For high school graduates the Frisch elasticity grows from about 0.40 at age 30 to 0.90 at age 65, and, as we already noted, the growth for college graduates is even greater.

We also decomposed the Frisch elasticity into changes at the extensive margin (i.e., participation) vs. the intensive margin (i.e., hours conditional on work). We find the Frisch elasticity for hours is roughly 0.10 for all ages and education groups. Thus, most of the action is on the participation margin. Of course, this means the Frisch elasticity for employment is roughly 0.10 smaller than that for total hours, and it has about the same patterns with education/age discussed earlier.²⁴ These results are consistent with a large body of empirical work that has looked only at employed men and found very small Frisch elasticities (see, e.g., MaCurdy (1981)).

Next we simulate permanent 10% wage reductions. These wage changes are uncompensated, so we obtain Marshallian elasticities. One could think of these wage reductions as being induced by unanticipated 10 percentage point permanent increases in the tax rate on labor earnings, with the tax change going into effect at the indicated age, and the revenue discarded. Figure 13 plots the life-cycle impact of unanticipated 10% wage reductions that initially occur at ages 30, 40, 50, 60, 65 or 70 and persist for the rest of the life. For both HS dropout and HS graduate workers the initial impact of a wage reduction is negative. That is, the substitution effect dominates the income effect. Furthermore, the magnitude of the initial impact is greater for older workers. For example, for a HS graduate who experiences a 10% wage reduction at age 40, the decline in hours is roughly 1%, implying a Marshallian elasticity of only 0.10. But if the wage reduction occurs at age 50 the Marshallian elasticity is 0.30, and if it occurs at age 60 it is 0.55.

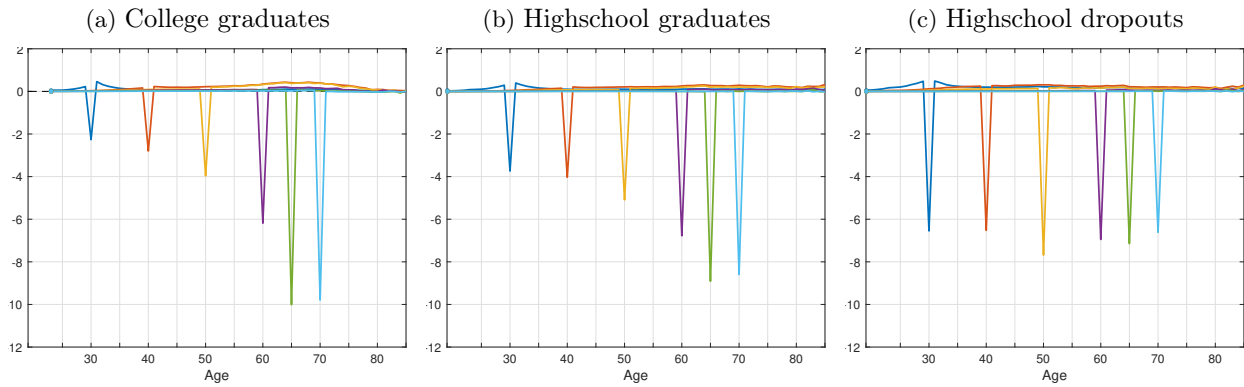
An interesting aspect of the results is that, for for HS dropout and HS graduate workers, the

²²We also calculated effects of unanticipated transitory changes. This gives approximate Frisch effects, as the transitory wage changes have very small effects on lifetime wealth (regardless of whether they are anticipated or not). Hence the elasticities with respect to unanticipated transitory wage changes are nearly identical to those for anticipated, so we do not report them. But they are available on request.

²³Theirs' was the first structural life-cycle model to include both assets and endogenous human capital, but they assumed continuous hours and assumed interior solutions, ignored liquidity constraints, and did not model retirement or social security

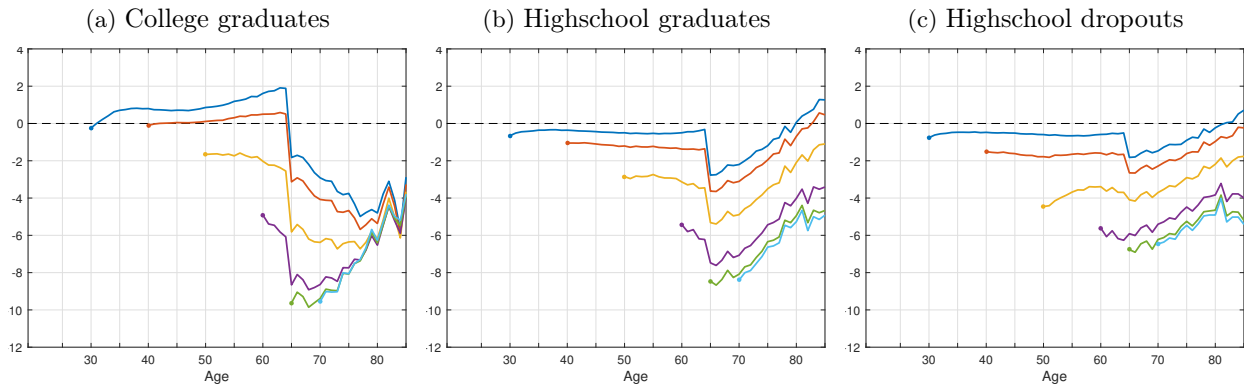
²⁴These results are available on request.

Figure 12: Responses of hours to transitory 10% wage reduction.



Notes: The plots show simulated percentage change in hours of labor supply induced by anticipated temporary 10% wage reduction at different points of the life cycle.

Figure 13: Responses of hours to permanent unanticipated 10% wage reduction.

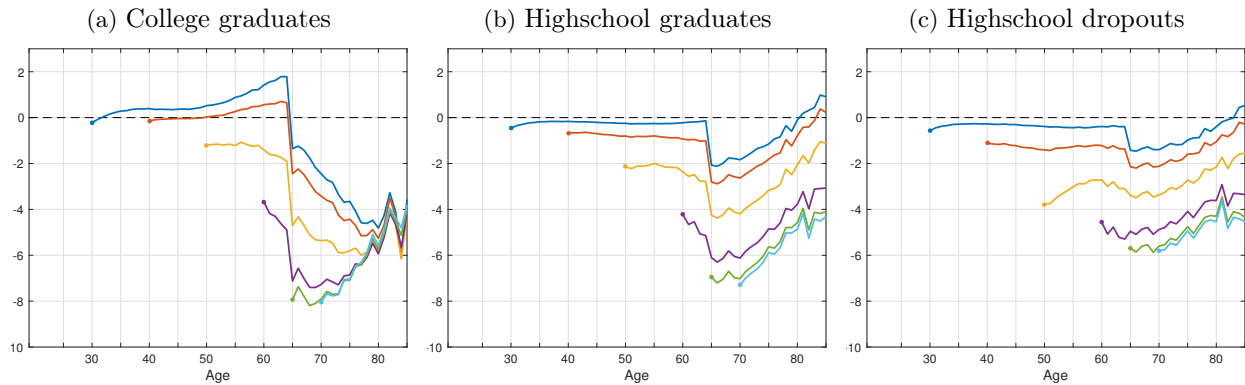


Notes: The plots show simulated percentage change in hours of labor supply induced by permanent unanticipated 10% wage reduction at different points of the life cycle.

impact doesn't vary much as the worker ages. For example, for a HS graduate who experiences a 10% wage reduction at age 40, the decline in hours is roughly 1% at ages 40 through 64, so the Marshallian elasticity is stable at about 0.10. But at age 65 the drop in hours is much greater. The reason this occurs is that at age 65 workers become eligible for retirement benefits (both superannuation and the aged pension). The presence of these retirement benefits amplifies the negative effect of the 10% wage reduction substantially at age 65, as more workers retire.

Figure 14 plots the effect of the permanent 10% wage reduction on participation. Two points stand out. First, the effects on participation are almost as large as the effects on total hours (Figure 13), implying again that most of the action is on the participation margin. Second, what Figure 14 clarifies the point that most of the large drops in hours at ages 65+ that we saw in Figure 13 are due to additional retirement, rather than reductions in hours.

Figure 14: Responses of participation to permanent unanticipated 10% wage reduction.



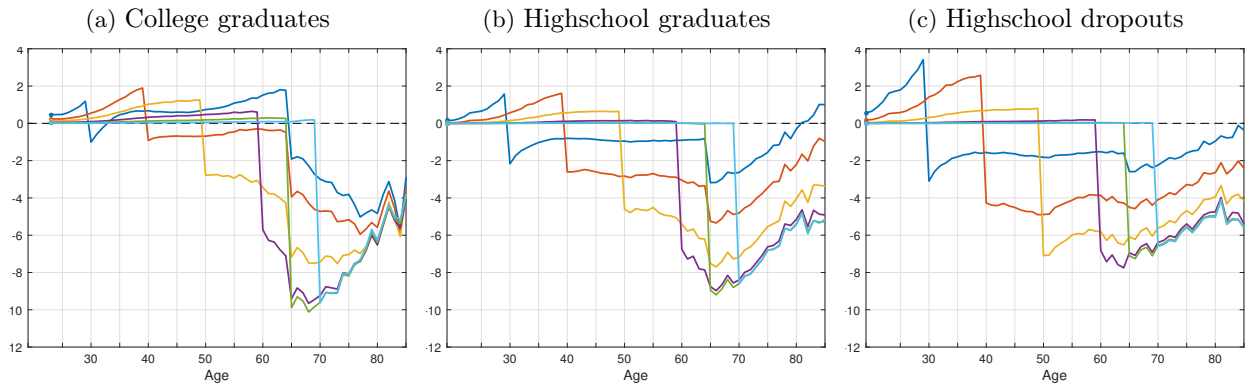
Notes: The plots show simulated percentage change in participation rates induced by permanent unanticipated 10% wage reduction at different points of the life cycle.

The results for college workers are rather different - see the left panels in Figures 13 and 14. The income effect dominates if college workers suffer a permanent wage decrease at age 30, and they work *more* until age 65. The initial impact is negligible, but they work about 1% to 2% more from age 35 to 64. However, just as with less educated workers, the lower wage makes college workers more likely to retire at age 65. If college workers experience a permanent wage decrease at age 40 the income and substitution effects roughly cancel for ages 40-64, but if the wage decrease happens at age 50 or older the substitution dominates (and the response of college workers looks fairly similar to less educated workers). These age patterns make sense, as income effects must be smaller for permanent tax increases that occur later in life.

Finally we consider permanent and fully anticipated 10% wage reductions. One could think of these as arising from anticipated permanent changes in tax rates. As we see in Figure 15, the anticipation of the wage change causes hours to increase up until the age when the tax increase goes into effect. Then, when the tax rate is increased, labor supply falls. But the magnitudes of the responses differ sharply by education, age of tax implementation, and age since tax implementation. For example, if a permanent tax increase goes into effect at age 30, labor supply drops by 3% at age 30 for dropouts, 2% for HS graduates, and only 1% for college graduates. We will refer to these as “immediate impacts” of permanent tax changes.

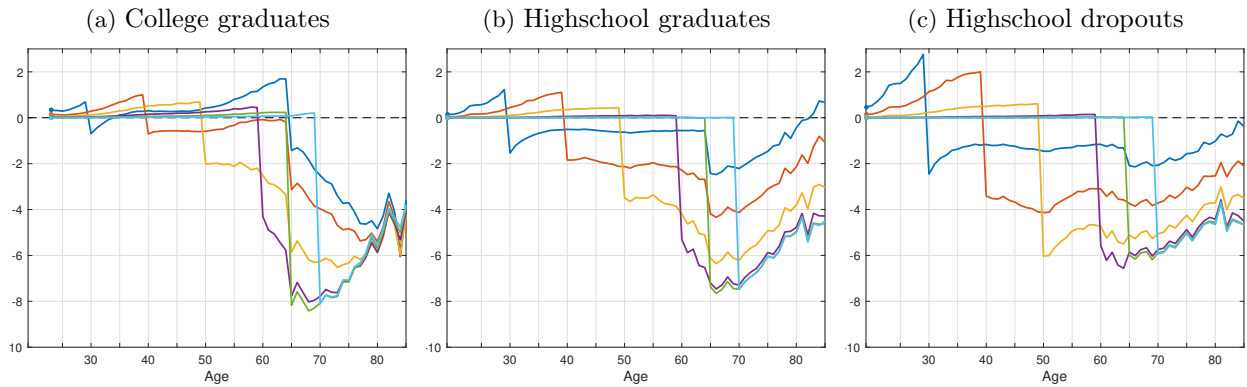
Figure 15 also shows that the immediate impact of anticipated tax increases grows with age (similar to how we found the Frisch elasticity grows with age). For example, for HS graduates the Marshallian elasticities in the year of tax implementation are 0.20, 0.30, 0.50 and 0.70 and 0.90, if the tax is increased at ages 30, 40, 50, 60 and 65, respectively. For college workers the pattern is 0.10, 0.10, 0.30, 0.60 and 1.00. So we see the increase in the Marshall elasticity (at

Figure 15: Responses of hours to permanent anticipated 10% wage reduction.



Notes: The plots show simulated percentage change in hours of labor supply induced by permanent anticipated 10% wage reduction at different points of the life cycle.

Figure 16: Responses of participation to permanent anticipated 10% wage reduction.



Notes: The plots show simulated percentage change in participation rates induced by permanent anticipated 10% wage reduction at different points of the life cycle.

the time of tax implementation) with age is steeper for the college graduates.

Figure 15 also allows us to examine how the Marshallian elasticity varies with age since tax implementation. The general pattern is that the Marshallian elasticity is rather small at younger ages, but then jumps up at ages 60 to 70. For example, for a HS graduate, when a 10% permanent tax increase is implemented at age 30, the immediate impact corresponds to an elasticity of 0.20, but this quickly falls to only about 0.10, and stays flat at that level from 35 to 64. However, at age 65 the Marshallian elasticity jumps to 0.30. This means that a tax rate increase causes workers to reallocate labor from older ages to younger ages (at least in relative terms, because labor supply falls more at ages 65+).

Keane and Wasi (2016) find a similar pattern whereby the Marshallian elasticity with respect to a permanently higher tax rate in effect for one's whole working life, and thus equivalent

to being fully anticipated, has effects that grow rapidly at older ages - see their Figure 14. Specifically, they find the Marshall elasticity is only about 0.10 at ages 30-45 but shoots up to about 1.0 at age 65 for HS and college types and about 0.70 for dropouts. It is comforting that our model generates fairly similar patterns, despite the substantial differences in specification, and even national context. This gives us more confidence that this is a robust finding.

Finally, Figure 16 reports the extensive margin responses to a permanent anticipated 10% tax increase implemented at different ages. Figures 15 and 16 are rather similar, again indicating that most of the action is on the extensive margin. In particular, the permanent tax increase increases the chance of retirement at age 65 substantially for all groups.

A caveat with regard to all the wage/tax experiments we have described is that the results are partial equilibrium. To the extent that a tax change has a large labor supply effect, we might expect it to also alter the equilibrium wage rate (i.e., the human capital rental rate). Keane and Rogerson (2015) derive conditions on aggregate technology such that human capital rental prices are fixed in these type of experiments.

9.2 Aged Pension Experiments

We now turn to our main policy simulations designed to assess how the aged pension affects labor supply, savings and human capital over the life cycle. In preliminary work we found that “tinkering” with aged pension rules (e.g., raising or lowering benefit levels by modest amounts) had rather minor effects on labor supply and other aspects of behavior. For example, we find that a 25% increase in the pension grant amount (holding means testing fixed) reduces labor supply of high school type workers by only 1% at age 60, and by even less at earlier ages. Even at age 65 and onward, the drop in labor supply is only about 5%.²⁵ Our model predicts the rather substantial change in the Aged Pension formula that occurred in 2010 (see Figure 1) was too small to have much impact on behavior, and our examination of data from before and after the policy change reveals no significant changes to labor supply or savings. While anti-climactic, our correct prediction of “no significant effect” is a type of model validation.

Thus, in our next experiment we assess the impact of the complete elimination of the Aged Pension program. Admittedly this is politically unrealistic, but it creates a much clearer picture of the program’s overall impact, and of the mechanisms through which it affects behavior.

Our simulations are conducted both with income taxes held fixed, and with taxes adjusted so policy changes are revenue neutral. As we noted in the introduction, the aged pension is a

²⁵Similarly, we find that a reduction in the wealth exemption for the aged pension has small effects on hours of work. For instance, a 50% cut in the exemption has essentially no effect on hours prior to age 65, and raises hours from age 65 onward by only about 0.6% to 1.6% depending on age and education. This is presumably because workers can respond by saving somewhat less prior to retirement.

large program. In 2014 it cost \$60 billion, compared to total income taxes of \$180 billion. Thus, given no behavioral response, we expect that income tax rates could be cut by roughly one-third if the aged pension were eliminated. In fact, as we show below, eliminating the aged pension has a strong positive labor supply effect. As a result, the income tax rate can be cut by 42% in our revenue neutral simulations.²⁶

We emphasize that even our revenue neutral simulations should be viewed as partial equilibrium. We repeat the caveat of the last section that policy experiments that generate large labor supply changes may alter equilibrium wages, and we do not account for this in our simulations. Such a large change as eliminating the Aged Pension might lead to other important changes in equilibrium as well (e.g., the possible creation on new insurance and annuity products). Nevertheless, our partial equilibrium experiments are useful to gain insights into the mechanisms through which the Aged pension affects *individual* life-cycle behavior, holding the wage rate and other features of the environment fixed.

9.2.1 Average Effects by Education Group

Figures 17 to 19 report the effects of eliminating the aged pension on total hours of work, consumption and assets, respectively. In each graph, the left panel holds taxes fixed, while the right panel reports the revenue neutral simulation. Each graph also reports results separately for the three education groups (college = yellow, HS = blue, dropouts = red).

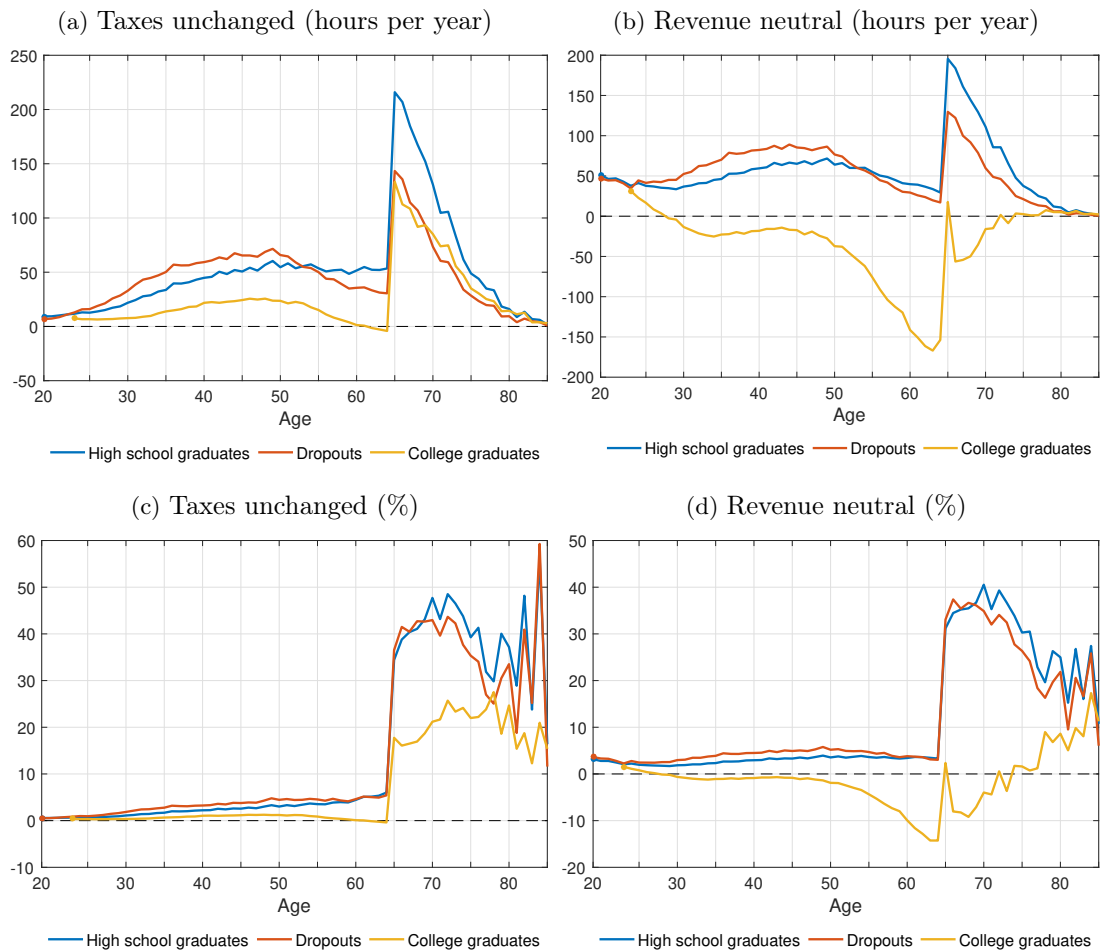
Consider the fixed tax rate simulation first. In Figure 17 we see that labor supply increases prior to age 65 for all three education groups. This is due to the income effect of the elimination of the pension benefit. Forward-looking agents know throughout their working lives that the pension will not be available to them when they retire, so they consume less (Figure 18) and accumulate more assets (Figure 19).

As we see in Figure 17, the largest positive effect on labor supply (prior to age 65) is for HS dropouts, followed by HS graduates and then college graduates. This is as expected, as the income effect should be proportional to expected pension benefits, and the less educated tend to receive more Aged Pension benefits. Interestingly, the increase in hours is negligible at young ages, as the receipt (or non-receipt) of pension benefits is far into the discounted future. The increase in labor supply peaks in about the 45 to 50 age range, at roughly 3.5% for HS dropouts, 2% for HS graduates, and 1.0% for college graduates (see panel c).

As we see in Figure 18, all three groups reduce consumption at the same time they are increasing labor supply. In levels it is the college group that reduces consumption the most, by

²⁶More precisely, the tax rates in both brackets can be cut by 42%, thus still keeping the tax system progressive.

Figure 17: Responses of hours of labor supply to elimination of Age Pension.

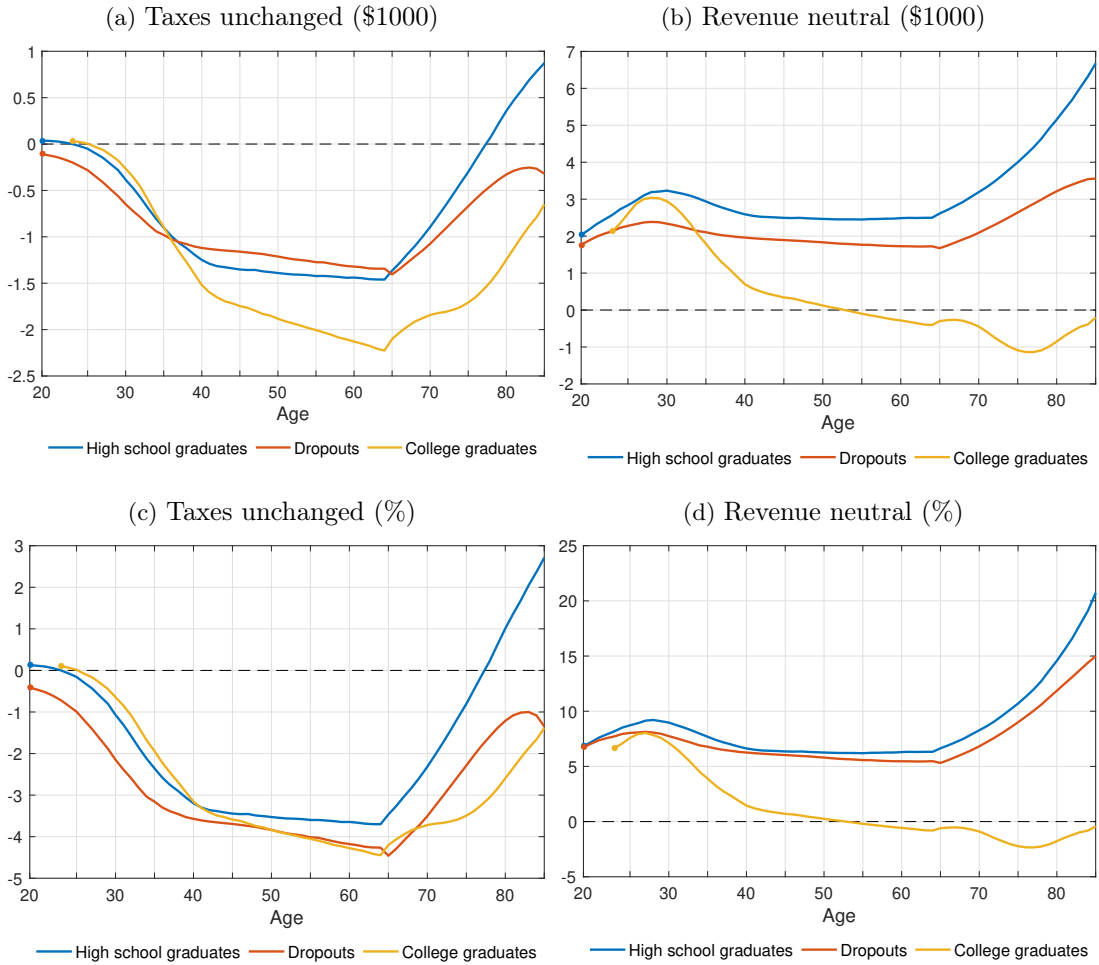


Notes: The plots show simulated absolute change in hours of labor supply induced by complete elimination of Age Pension with the same income tax (panels a and c) and with income tax rates lowered by 42% to keep the governmental revenue unchanged (panels b and d). Top row (panels a and b) show absolute changes, bottom row (panels c and d) show relative (%) changes.

about \$1.5K to \$2.2k per year in the 45 to 64 age range, while the dropout and HS graduate groups only decrease consumption by about \$1.2k to \$1.3 during that same age range. But in percentage terms (see panel c) the drop in consumption is similar for all three groups (i.e., about 3.5% to 4.5% per year in the 40 to 65 age range).

The next result of these changes is that all three groups accumulate more assets by age 64. As we see in Figure 19, the college groups accumulates \$150k more on average, while the dropout and HS graduate groups accumulate about \$175k more on average. It is to be expected that the additional wealth accumulation by the college group is somewhat less, as they are less likely to utilize the aged pension in the baseline economy. But in percentage terms (see panel c) the differences between educations are much more dramatic. The dropouts accumulate 120% more assets by age 64, while the college graduates accumulate only 25% more.

Figure 18: Responses of consumption to elimination of Age Pension.

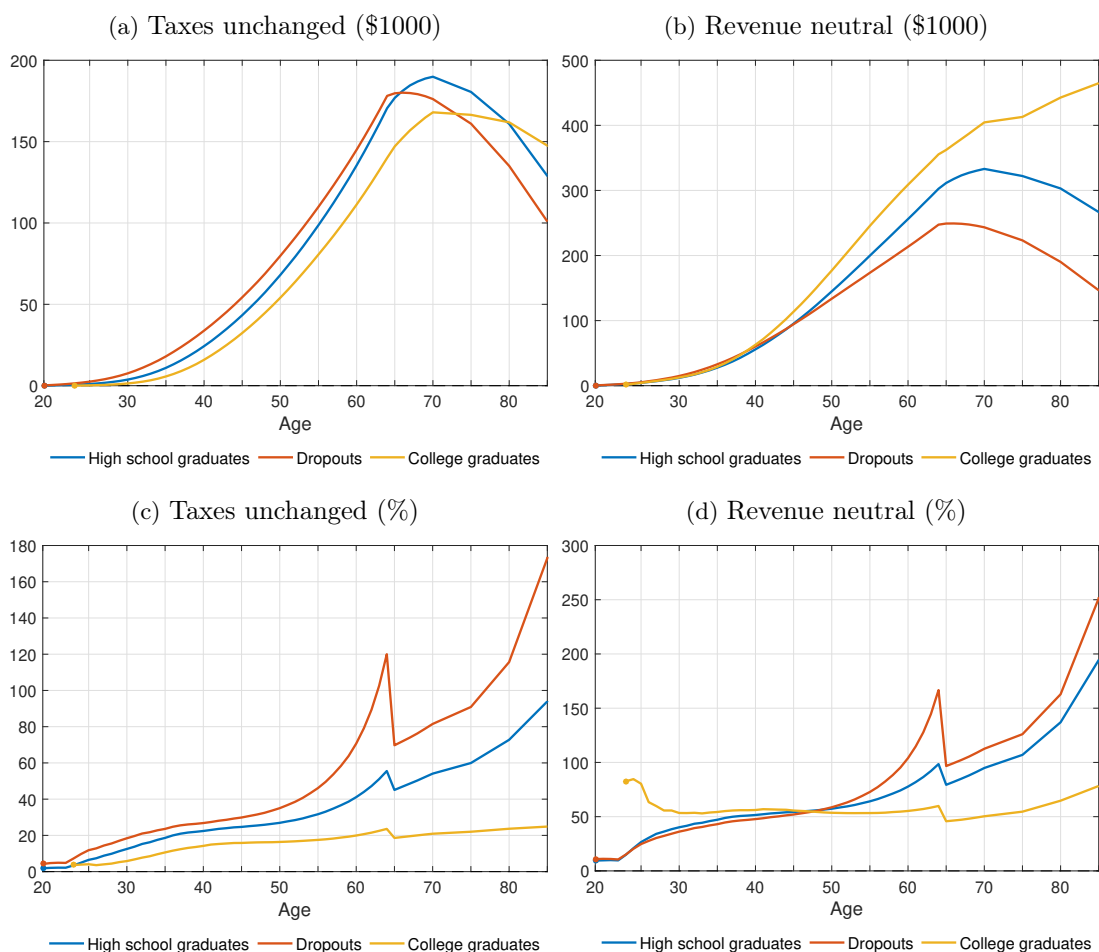


Notes: The plots show simulated absolute change in consumption induced by complete elimination of Age Pension with the same income tax (panels a and c) and with income tax rates lowered by 42% to keep the governmental revenue unchanged (panels b and d). Top row (panels a and b) show absolute changes, bottom row (panels c and d) show relative (%) changes.

Nevertheless, we see that even the effects on the college educated are substantial. There are two main reasons that college workers respond strongly to elimination of the Aged Pension. First, the Aged Pension is a consumption insurance program, that mitigates the risk of poverty in old age. Risk averse agents will save less prior to age 65 if they have such consumption insurance available, even if the probability of using it is rather small. Second, as we discussed in Section 2, the rate at which Aged Pension benefits are taxed away with income and assets is very low. Thus, even fairly high income individuals will in fact be eligible for some benefits under the program.

Now consider what happens at age 65. Returning to Figure 17, we see that annual hours of work at age 65 increase by roughly 140 hours (35%) for the HS dropouts, 220 hours (35%)

Figure 19: Responses of wealth to elimination of Age Pension.



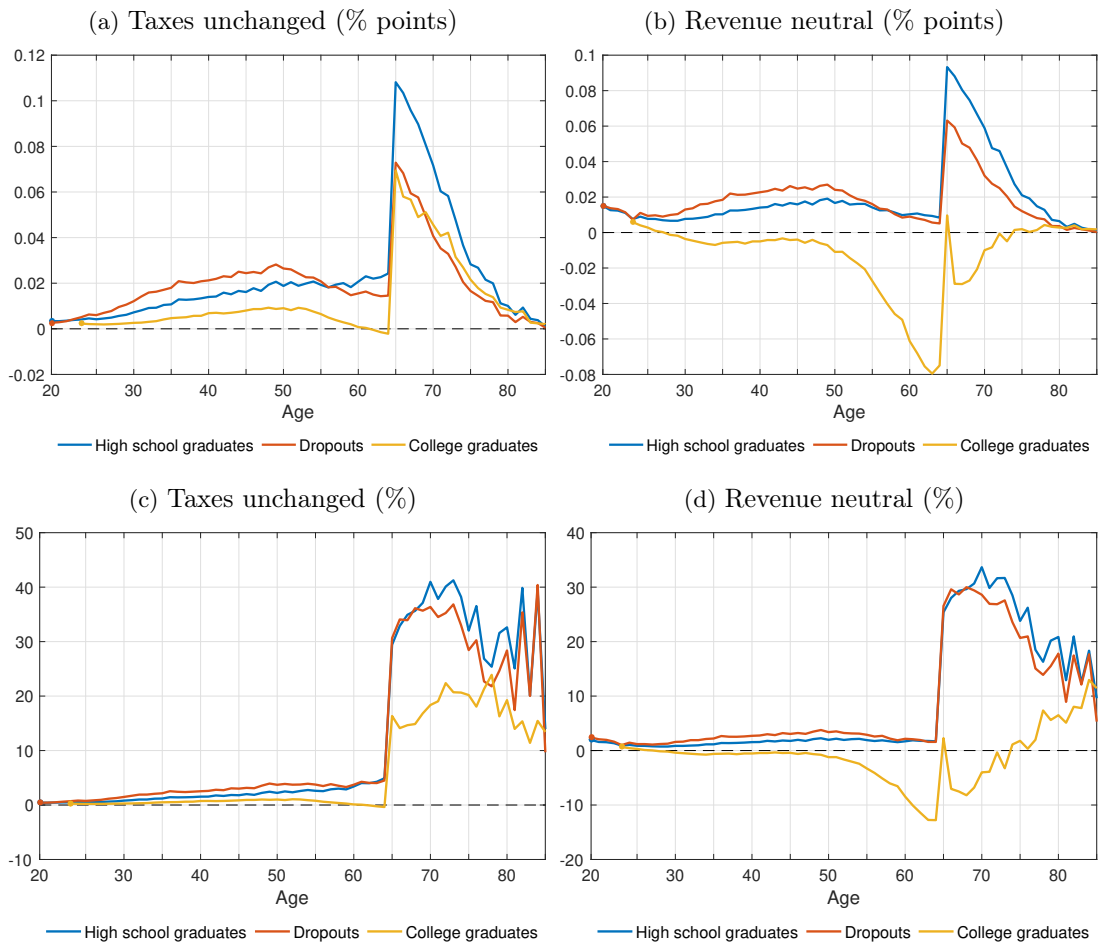
Notes: The plots show simulated absolute change in wealth induced by complete elimination of Age Pension with the same income tax (panels a and c) and with income tax rates lowered by 42% to keep the governmental revenue unchanged (panels b and d). Top row (panels a and b) show absolute changes, bottom row (panels c and d) show relative (%) changes.

for the HS graduates, and 130 hours (18%) for the college graduates. Workers also continue to work more at ages beyond age 65, but the absolute increment to labor supply slowly diminishes with age. Figure 20 reports the results for labor force participation (employment). This figure (panel c) shows percentage increases in participation almost as great as the increases in hours we saw in Figure 17.²⁷ Thus, we see that the increase in labor supply from eliminating the aged pension is mostly due to increased employment, not increased hours of work. Interestingly, this is true both before and after age 65.

Consider next the revenue neutral simulations reported in the right panels of Figures 17-20.

²⁷That is, in Figure 20 panel c we see increases in participation at age 65 of about 30% for HS and dropout workers, and 15% for college workers. This compares to overall hours increases of about 35% and 18%, respectively, that we see in Figure 17 panel c. So for all education groups about 85% of the hours increase at age 65 is due to higher participation (or delayed retirement).

Figure 20: Responses of participation rate to elimination of Age Pension.



Notes: The plots show simulated increase in participation in percentage points (top panels a and b), and the percentage increase in participation (bottom panels c and d) induced by complete elimination of Age Pension with the same income tax (panels a and c) and with income tax rates lowered by 42% to keep the governmental revenue unchanged (panels b and d). Top row (panels a and b) show absolute changes in percentage points, bottom row (panels c and d) show relative (%) changes in percentages.

As we see in Figure 17, the impact on the labor supply behavior of the HS graduate and HS dropout types is very similar to the no tax change case. Because the tax rate is reduced at the same time the aged pension is eliminated, the positive income effect on labor supply is slightly lessened, but on the other hand the tax cut now creates a positive substitution effect. Apparently these differences roughly cancel, and so the net positive effect of eliminating the aged pension on labor supply for the less educated groups is very similar regardless of whether taxes are reduced.

The same cannot be said of consumption and assets however. As we see in Figure 18, the lower tax rate enables the HS dropout and HS graduate types to consume about \$2k to \$2.5k more per year during their working lives, despite the knowledge that the aged pension will not be available to them in retirement. And as we see in Figure 19, the lower tax rate enables the HS

dropout and HS graduate types to accumulate an extra \$75k or \$125k, respectively, by age 65 compared to the fixed tax case, or an extra \$250k to \$300k compared to the baseline economy (that has both the aged pension and higher taxes). In fact, as we see in the right panel of Figure 18, the HS dropout and HS graduate types are, on average, able to consume significantly more at ages 65+ in the world that has no aged pension but lower taxes.

The results for the college graduates are very different. As we see in the right panel of Figure 17, the labor supply behavior of the college graduates is dramatically different between the no tax change and the revenue neutral simulations. In fact, in the revenue neutral simulation, the college graduates work quite a bit *less* in the world with no aged pension and lower tax rates. This reduction in labor supply is especially focused in the years close to retirement. For example, they work about 30 fewer hours at age 50 (or 1.5% less), 75 fewer hours at age 55 (4% less), and 140 fewer hours age age 60 (10% less). Two factors drive this result, and account for why the college type is so differently affected by the tax cut than the lower educated types: First, the college types simply pay much more taxes in the baseline, so the income effect of the tax cut is much greater for them. Second, as we saw in Section 9.1, the income effect dominates the substitution effect for college workers.

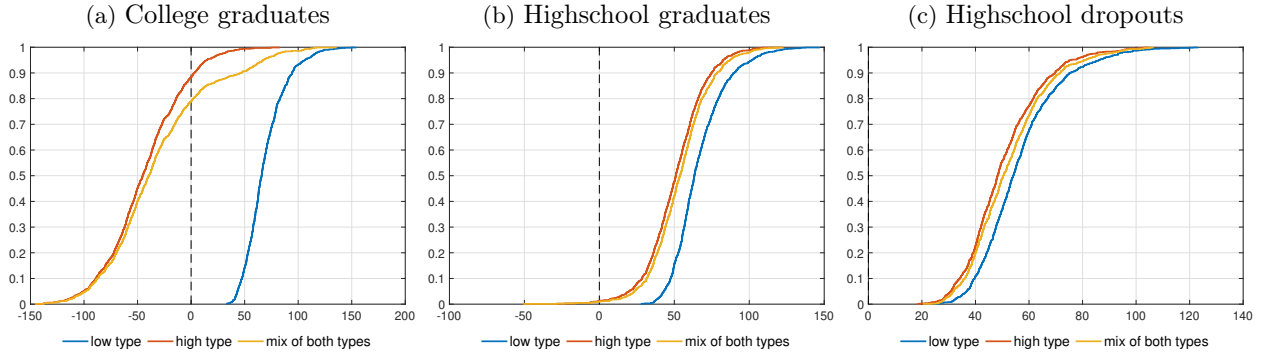
As we see in Figure 18, the college types consume about \$2k to \$3k more per year in their 30s in the revenue neutral simulation, but after that they choose to consume about the same as they did in the baseline economy. In Figure 19 we see that in the revenue neutral simulation the college type is able to accumulate on average an extra \$350k (or 60%) in assets by age 64 compared to the baseline (and \$200k extra compared to the no tax change simulation). The 42% tax cut enables the college type to achieve this very high level of asset accumulation while working less and with little change in consumption. Indeed, our results show that their extra accumulated wealth does not go into extra consumption in retirement (see Figure 18) but rather into more leisure and higher bequests.

9.2.2 Heterogeneity and Welfare

In this section we analyze the heterogenous impacts of eliminating the aged pension on different types of agents in the economy. We focus on the revenue neutral simulation, and look at effects on labor supply, consumption and welfare.

Recall that *ex ante* heterogeneity in our model comes from three sources: education level, latent skill and taste for work types, and initial wealth endowments (inheritance). Within each education level there are “high” and “low” types, where the high type has both higher skill and lower disutility of work. We will refer to these as high/low skill types to be concise. Within

Figure 21: Distributions of responses of hours to elimination of Age Pension.



Notes: The plots show the empirical cumulative distribution functions for the mean change of hours of labor supply over the life-cycle induced by complete elimination of Age Pension with income tax lowered by 42% to keep the governmental revenue unchanged. In each panel the data is split by unobserved heterogeneity type τ_{uh} (high and low), with the overall distribution shown in light color.

each education/skill type there is a continuous distribution of initial wealth. Heterogeneity in outcomes (ex post) also arises from transitory wage/taste shocks that cause otherwise identical agents to make different choices and have different earnings (“luck”).

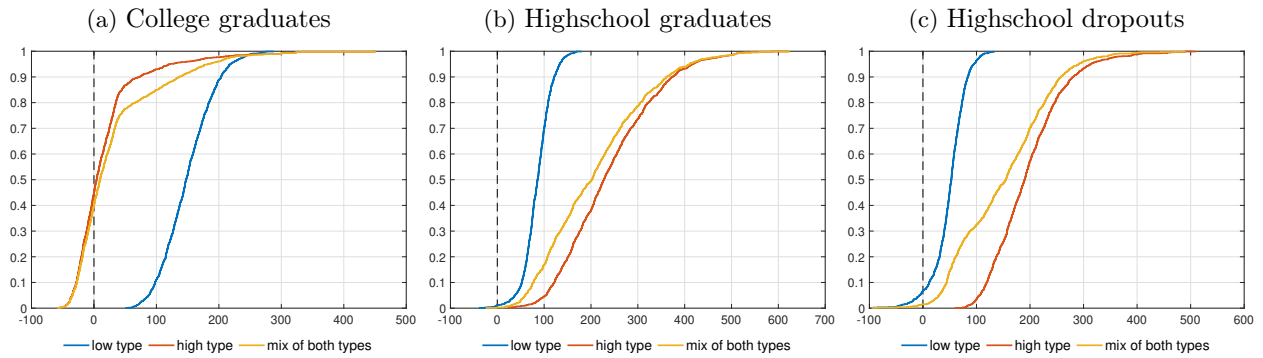
Figure 21 reports education/skill type-specific cumulative distribution functions (CDFs) for changes in mean annual hours of work (over the entire life). In each graph, the red curve represents the high-skill type, the blue curve represents the low-skill type, and the yellow curve represents the distribution for the whole education group, accounting for the type membership probabilities which are education specific in our model.

For example, for the HS dropout group we see that the entire CDF lies to the right of zero. Thus, every member of this group works higher average hours as a result of eliminating the aged pension (and reducing taxes). Indeed, nearly everyone has an hours increase of at least 20 hours per year, and there is non-negligible mass up to about 90 hours per year. The median individual increases hours by roughly 50 per year. The differences between the high and low skill type are not large. For instance, the median hours increase for the high skill type is about 48 hours per year, while that for the low skill type is about 55 hours per year.

The effects on the HS graduate group are very similar overall. The median increase in hours is about 55 per year, about 50 for the high type and 65 for the low type. The only slight difference from the dropouts is that the distribution of hours changes are a bit more spread out at the tails. A small fraction of the high skilled HS graduate type have little to no hours increase, and about 10% of the low type have hours increases over 90+ per year.

However, the CDFs for college graduates look very different. In Figure 21 we see that for the

Figure 22: Distributions of responses of consumption to elimination of Age Pension.

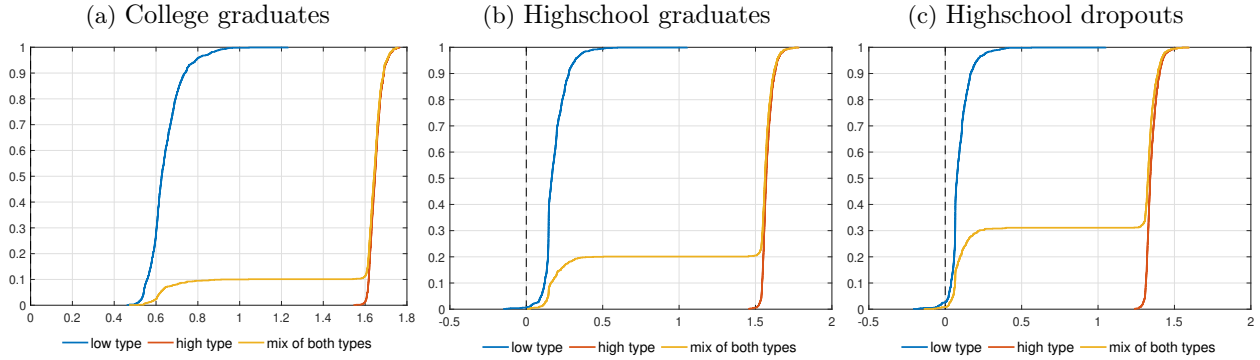


Notes: The plots show the empirical cumulative distribution functions for the non-discounted sum of simulated absolute changes of consumption (in \$1000) over the whole lifecycle induced by complete elimination of Age Pension with income tax lowered by 42% to keep the governmental revenue unchanged. In each panel the data is split by unobserved heterogeneity type τ_{uh} (high and low), with the overall distribution shown in light color.

high type almost 90% have a decrease in hours, and the median decrease is about 48 hours per year. But for the low skill type all individuals have an increase in hours, and the median increase is about 65 per year. What accounts for the large difference in behavior? In simple life-cycle models like MaCurdy (1981) with a linear budget constraint, interior solutions for continuous hours, and savings as the only source of dynamics (i.e., no human capital), elasticities of labor supply depend only on the parameters that describe curvature of utility in consumption and hours. And these utility parameters are homogenous across the high and low types. However, as Keane and Rogerson (2015) and Keane and Wasi (2016) discuss, in more complex models with features like human capital, and operative participation margin and non-linear budget sets, the situation is much more complex. Labor supply elasticities become functions of the entire economic environment.

The point that preference parameters are “insufficient” to characterize labor supply responses is well illustrated here. As we see that the high and low skill types have reactions to the elimination of the aged pension that differ not only in magnitude but even in sign. As we see, the low-skill college types behave quite similarly to high school types. Intuitively, there are two main reasons for this: First, the low skill college types are much more likely to use the aged pension than the high skill types, so the positive income effect of eliminating it is greater. Second, the low-skilled college type pay less taxes than the high-skilled type, so the negative income effect from lowering the tax rate is smaller. So, on net, the income effect on labor supply is positive for the low-skill type but negative for the high skill type (despite the fact that both types have the same curvature parameters on consumption and leisure).

Figure 23: Distributions of ex-ante welfare effects of elimination of Age Pension.



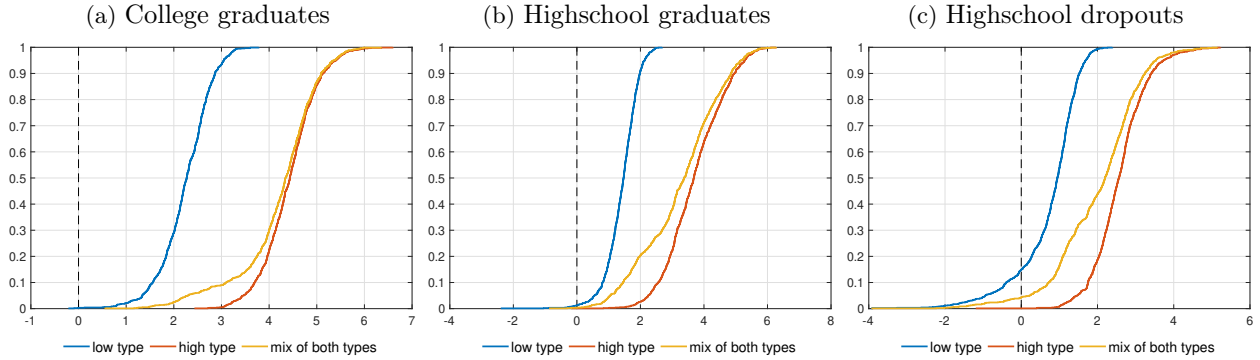
Notes: The plots show the empirical cumulative distribution functions for the discounted expected maximum utility at age 19 (23 for college graduates) induced by complete elimination of Age Pension with income tax lowered by 42% to keep the governmental revenue unchanged. In each panel the data is split by unobserved heterogeneity type τ_{uh} (high and low), with the overall distribution shown in light color.

Figure 22 reports CDFs of the change in cumulative consumption over the life cycle (not discounted). Note that almost all HS educated workers consume more when the aged pension is eliminated (and taxes are reduced). The median increase is about \$200k. It is not surprising that the median increase is more than twice as great for the high type as for the low type, as consumption for the high type is much greater under the baseline. The results for dropouts are very similar, with a median increase of roughly \$150k. The only notable difference is that about 6% of the low type experience a drop in consumption under the experiment. However, this represents a very small fraction of the overall population (as only 25% are HS dropouts, and only 20% of the dropouts are the low type).

Again, the results for college graduates look very different. The median change in lifetime consumption for the high type is almost zero. As we saw in Figure 21, these people tend to work less, and consume about the same, as under the baseline. But there is a great deal of heterogeneity in response within this group. Roughly 10% of the high skill college types both work more and consume quite a bit more (i.e., an additional \$100k to \$250k) under the new policy. On the other hand, the low skill college types all work more and consume more under the new policy. Their median increase in consumption is a substantial \$150k.

Figure 23 reports the distribution of the change in expected present value of lifetime utility expected under the policy experiment of eliminating the age pension (and reducing taxes). We see that every single member of the college group is ex ante better off. The median increase in expected PV of lifetime utility at age 23 is roughly 1.65 for the high type and 0.62 for the low type. Note that in these graphs, as we are taking expectations in the initial period, “luck” plays

Figure 24: Distributions of ex-post welfare effects of elimination of Age Pension.



Notes: The plots show the empirical cumulative distribution functions for the discounted expected maximum utility at age 65 induced by complete elimination of Age Pension with income tax lowered by 42% to keep the governmental revenue unchanged. In each panel the data is split by unobserved heterogeneity type τ_{uh} (high and low), with the integrated line shown in light color.

no role (i.e., wage and taste shocks are integrated out). Thus all the within education/skill heterogeneity is due to heterogeneity in initial assets (inheritance). Thus, among the college educated, even the lowest skilled with the lowest initial wealth would prefer (ex ante) to live in a world with the age pension abolished and tax rates lowered.

It is not surprising that the college educated are better off with less social insurance and lower taxes. What is surprising is that the story is little different for the HS graduates and even the HS dropouts. For the high types, the median utility gains are about 1.3 and 1.6 for the HS and dropout types, respectively. Even for the low skill HS graduates the utility change is almost universally positive, with a median increase of about 0.2. Even for low skill HS dropouts only about 3% are slightly worse off ex ante. It would only require trivial transfers from other groups to render this group better off as well, and render the elimination of the aged pension Pareto improving ex ante.

We note that the mean increase in the present value of lifetime utility evaluated in the first period is 0.95 for the dropouts, 1.27 for the HS graduates and 1.52 for the college graduates. These translate into consumption equivalent values (CEVs) of \$2386, \$3104 and \$3753, respectively. These are the amounts of additional assets we would have to give a typical person of each education level at t_0 to make them indifferent between living in the baseline economy vs. the economy with no aged pension and the lower tax rate. If we look only at the low types, the CEVs are \$1212 for dropouts, \$1219 for HS graduates and \$1818 for college graduates, respectively. The gains for the high types are greater (i.e., \$3242, \$3827 and \$4048, respectively).

Finally, in order to look at the insurance aspect of the program, it is interesting to look at

the distribution of the value function at age 65. Although almost everyone prefers to abolish the aged pension *ex ante*, are there a significant number of people who, at age 65, due to bad luck (i.e., poor wage draws or poor taste for work draws) find themselves in a position where they wish *ex post* that the program existed? This is reported in Figure 24. Strikingly, nearly everyone is better off even *ex post*. The exception is about 5% of the HS dropouts, which corresponds to only about 1% of the total population.

10 Discussion and conclusions

In this paper we have developed and estimated a life-cycle model of labor supply and consumption behavior and used it to evaluate the aged pension program in Australia. Our results imply that *ex ante* almost all agents would prefer to live in a world where the aged pension is abolished, and where taxes are reduced to maintain revenue neutrality. Three key factors drive this result. First, the aged pension is a large program whose cost is equal to roughly one third of income tax revenues. Second, the program has negative labor supply effects. As a result, eliminating the program would allow the tax rate on earning to be reduced by roughly 40% in a revenue neutral simulation. Third, while the aged pension is meant as a social insurance program against the risk of low consumption in old age, it is very poorly targeted. Although the program is means tested, taper rates are very low, so that benefits extend well up into the income/wealth distribution. Hence, many fairly wealthy people receive transfers that give them very minor increments to utility in retirement. The poor targeting makes the insurance program expensive for everyone.

The dynamic life-cycle model estimated in this paper is quite rich, in that it incorporates several key features that have not previously been brought together in a single model: asset accumulation, liquidity constraints, human capital accumulation via learning by doing, and a discrete choice of hours involving six possible levels. Of course we also include the superannuation and Aged Pension rules, as well as retirement decisions. There is of course a large literature on estimating the effects of means tested transfer programs (Moffitt (2016)), but most prior work analyzes these programs in a static framework, ignoring their effects on asset and human capital accumulation. A notable exception is the literature that has developed in response to the US welfare reform of the mid-1990s. This introduced time limits into means tested transfer programs, motivating a number of researchers to analyze these programs using dynamic models. See, for example, Chan (2013), Keane and Wolpin (2010).

Our work also extends prior literature on structural estimation of dynamic programming models. In our model the continuous choice of consumption combined with discrete choice of hours renders the problem non-convex. There are kinks in the value functions, and the optimal

consumption policy has discontinuities. To deal with this problem, we use the discrete-continuous generalization of the endogenous grid point method (DC-EGM) developed in Iskhakov et al. (2017). This paper is the first application of DC-EGM to an empirical model as complex as ours. The paper also represents the first application of the smooth simulation algorithm of Bruins et al. (2018) to a dynamic structural model.

An obvious extension of the present paper would be to use the structural model to experiment with changes in design of the Aged Pension parameters that would improve the targeting of the program and perhaps improve ex ante welfare. In related work, Tran and Woodland (2014) have analyzed optimal design of the Australian Aged Pension using a very different framework from ours. Specifically, they look an overlapping generations (OLG) equilibrium model that includes labor supply and savings decisions, in an environment where consumers are (exogenously) heterogeneous in their human capital liquidity and face liquidity constraints. But in order to make OLG modelling feasible they abstract from human capital investment and specify a simpler model of individual decision making. In their setting, they find, as we do, that elimination of the aged pension, along with the tax reductions this permits, would increase welfare. Furthermore, they find that conditional on the program existing, the optimal taper rate for benefits as a function of labor earnings is roughly 30% to 40%. That is very close to the existing rate. This suggests that finding improved program parameters (that mitigate reductions in labor supply and savings) may be quite difficult. It will likely involve more than "tinkering" with taper rates, and require more creative ways of enhancing work incentives in the system. One approach that seems useful in the design of means tested transfers in the static setting is the combination of means tested benefits with work subsidies designed to undo some of the work disincentives created by means testing (see Keane (1995)). This is worth considering in a dynamic context.

Another obvious extension of our model would be to try to embed it in an equilibrium OLG model. These type of models have frequently been used to analyze universal pension schemes like the US Social System (see Krueger and Kubler (2006), Imrohoroglu et al. (1995)), and they have been adapted by Tran and Woodland (2014) to study the means tested Aged Pension program. The tradeoff here is that in a partial equilibrium setting, like that adopted here or in van der Klaauw and Wolpin (2008), it is possible to implement much richer models of individual behavior. On the other hand, the obvious advantage of the OLG equilibrium framework is that it allows one to examine effects of the program on both wages and asset returns, as well as to analyze inter-generational risk sharing if there are aggregate shocks to wages and returns on savings. Of course, embedding dynamic structural models of individual behavior within equilibrium models is a challenging enterprise that has rarely been accomplished. One of the

few examples is Lee and Wolpin (2010).

Finally, an important extension of our model would be to incorporate changes in health, as in van der Klaauw and Wolpin (2008), DeNardi et al. (2010), French and Jones (2011). Our model abstracts from work limitations (or disability) due to poor health, and thus implicitly assumes that all workers can increase labor supply at older ages if the Aged Pension were abolished. But for some workers this response may not be possible. Thus, by ignoring health shocks we may understate the insurance value of the Aged Pension.

References

- Australian Government Actuary. Australian life tables. Technical report, Department of Treasury, Canberra, ACT, 2009.
- Joseph G Altonji and Lewis M Segal. Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366, 1996.
- Australian Prudential Regulation Authority. Insight, Issue one, 2014. URL http://www.apra.gov.au/Insight/Documents/Insight_Issue1_2014_FINAL.pdf.
- Satish Balay, Shirang Abhyankar, Mark F. Adams, Jed Brown, Peter Brune, Kris Buschelman, Lisandro Dalcin, Victor Eijkhout, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Dave A. May, Lois Curfman McInnes, Richard Tran Mills, Todd Munson, Karl Rupp, Patrick Sanan, Barry F. Smith, Stefano Zampini, Hong Zhang, and Hong Zhang. PETSc users manual. Technical Report ANL-95/11–Revision 3.9, Argonne National Laboratory, 2018. URL <http://www.mcs.anl.gov/petsc>.
- Marianne Bruins, James A Duffy, Michael P Keane, and Anthony A Smith Jr. Generalized indirect inference for discrete choice models. *arXiv preprint arXiv:1507.06115*, 2018.
- Christopher D Carroll. The method of endogenous gridpoints for solving dynamic stochastic optimization problems. *Economics letters*, 91(3):312–320, 2006.
- Mark K Chan. A dynamic model of welfare reform. *Econometrica*, 81(3):941–1001, 2013.
- Rafal Chomik and Mary MacLennan. Aged care in australia: Part I – Policy, demand and funding. Technical Report 01, ARC Centre of Excellence in Population Ageing Research, 2014.
- Commonwealth of Australia. Budget strategy and outlook 2009–2010, budget paper no. 1. Technical report, Commonwealth of Australia, Canberra, 2009.
- Mariacristina DeNardi, Eric French, and John Bailey Jones. Why do the elderly save? the role of medical expenses. *Journal of Political Economy*, 118(1):39–75, 2010.
- John Dixon. A global ranking of national social security systems. *International Social Security Review*, 53(1):109–122, 2000.
- Jeppe Druedahl and Thomas Høgholm Jørgensen. A general endogenous grid method for multi-dimensional models with non-convexities and constraints. *Journal of Economic Dynamics and Control*, 74:87–107, 2017.

- FaHCSIA. Annual Report 2012-13. Technical report, Department of Families, Housing, Community Services and Indigenous Affairs, Canberra, ACT, 2012.
- Giulio Fella. A generalized endogenous grid method for non-smooth and non-concave problems. *Review of Economic Dynamics*, 17(2):329–344, 2014.
- Jun Feng, Paul Gerrans, and Gordon Clark. Understanding superannuation contribution decisions: Theory and evidence. Technical report, CSIRO–Monash Superannuation Research Cluster, 2014.
- Eric French and John Bailey Jones. The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, 79(3):693–732, 2011.
- John Geweke and Michael Keane. An empirical analysis of earnings dynamics among men in the psid: 1968-1989. *Journal of Econometrics*, 96(2):293–356, 2000.
- Andrew Herscovitch and David Stanton. History of social security in australia. Technical report, Australian Institute of Family Studies, 2008.
- Susumu Imai and Michael P Keane. Intertemporal labor supply and human capital accumulation. *International Economic Review*, 45(2):601–641, 2004.
- Ayse Imrohoroglu, Selahattin Imrohoroglu, and Douglas H. Joines. A life cycle analysis of social security. *Economic Theory*, 6(1):83–114, 1995.
- Fedor Iskhakov, Thomas H. Jørgensen, John Rust, and Bertel Schjerning. The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics*, 8(2):317–365, 2017. doi: 10.3982/QE643.
- Michael P. Keane. A new idea for welfare reform. *Quarterly Review, Federal Reserve Bank of Minneapolis*, pages 2–28, 1995.
- Michael P Keane. Life-cycle labour supply with human capital: Econometric and behavioural implications. *The Economic Journal*, 126(592):546–577, 2016.
- Michael P Keane and Richard Rogerson. Reconciling micro and macro labor supply elasticities: A structural perspective. *Annual Review of Economics*, 7(6):89–117, 2015.
- Michael P Keane and Nada Wasi. Labour supply: the roles of human capital and the extensive margin. *The Economic Journal*, 126(592):578–617, 2016.
- Michael P Keane and Kenneth I Wolpin. The career decisions of young men. *Journal of political Economy*, 105(3):473–522, 1997.
- Michael P Keane and Kenneth I Wolpin. The role of labor and marriage markets, preference heterogeneity and the welfare system in the life-cycle decisions of black, hispanic and white women. *International Economic Review*, 51(3):851–892, 2010.
- Dirk Krueger and Felix Kubler. Pareto-improving social security reform when financial markets are incomplete!? *The American Economic Review*, 96(3):737–755, 2006.
- Donghoon Lee and Kenneth I. Wolpin. Accounting for wage and employment changes in the us from 1968-2000: A dynamic model of labor market equilibrium. *Journal of Econometrics*, 156(1):68–85, 2010. Structural Models of Optimization Behavior in Labor, Aging, and Health.
- Thomas E MaCurdy. An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy*, 89(6):1059–1085, 1981.

- Daniel McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in econometrics*. Academic Press, New York, 1973.
- Daniel McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, pages 995–1026, 1989.
- Mercer. Melbourne mercer’s global pension index. Technical report, The Australian Centre for Financial Studies (ACFS), October 2015.
- R.A. Moffitt. *Economics of Means-Tested Transfer Programs in the United States*. Number v. 1 in *Economics of Means-tested Transfer Programs in the United States*. University of Chicago Press, 2016. ISBN 9780226370477. URL <https://books.google.co.uk/books?id=LKhSDQAAQBAJ>.
- Todd Munson, Jason Sarich, Stefan Wild, Steven Benson, and Lois Curfman McInnes. Tao 2.0 users manual. Technical Report ANL/MCS-TM-322, Mathematics and Computer Science Division, Argonne National Laboratory, 2012. <http://www.mcs.anl.gov/tao>.
- Ariel Pakes and David Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica*, pages 1027–1057, 1989.
- John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, pages 999–1033, 1987.
- Social Security Administration. Social security programs throughout the world: Asia and the pacific, 2014. Technical Report SSA Publication No. 13-11802, Social Security Administration, 500 E Street, SW, 8th Floor Washington, DC 20254, March 2015.
- Carlo Strub and Andrew Clausen. A general and intuitive envelope theorem. In *2014 Meeting Papers*, number 235. Society for Economic Dynamics, 2014.
- M. Summerfield, S. Freidin, M. Hahn, P. Ittak, N. Li, N. Macalalad, N. Watson, R. Wilkins, and M. Wooden. *HILDA User Manual – Release 12*. Melbourne Institute of Applied Economic and Social Research, University of Melbourne, December 2013.
- Chung Tran and Alan Woodland. Trade-offs in means tested pension design. *Journal of Economic Dynamics and Control*, 47:72–93, 2014.
- Wilbert van der Klaauw and Kenneth I. Wolpin. Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, 145(1):21–42, 2008.
- Mark Wooden and Nicole Watson. The HILDA survey and its contribution to economic and social research (so far). *Economic Record*, 83(261):208–231, 2007.

Appendix

A Complete list of model parameters

Table 8: Parameters of the model.

	Symbol	Description
Preference parameters		
1	ζ	CRRA coefficient in consumption
2	γ_1	Disutility of working 1000 hours (20 per week)
3	γ_2	Disutility of working 2000 hours (40 per week)
4	γ_3	Disutility of working 2250 hours (45 per week)
5	γ_4	Disutility of working 2500 hours (50 per week)
6	γ_5	Disutility of working 3000 hours (60 per week)
7	κ_1	Correction coefficient for low type with disutility of work
8	κ_2	Quadratic coefficient on age for older workers
9	κ_3	Linear coefficient on age for young workers
10	ξ	CRRA coefficient in utility of bequest
11	b_{scale}	Scale multiplier of the utility of bequest
12	β_{clg}	Discount factor, college
13	β_{hs}	Discount factor, highschool
14	β_{dr}	Discount factor, dropouts
15	λ	Scale of EV taste shocks
Parameters of state transition (beliefs)		
16	$\eta_{0,high}$	Human capital: constant for high type
17	$\eta_{0,clg}$	Human capital: constant for college
18	$\eta_{0,hs}$	Human capital: constant for high school
19	$\eta_{0,dr}$	Human capital: constant for dropouts
20	$\eta_{1,clg}$	Human capital: work experience for college
21	$\eta_{1,hs}$	Human capital: work experience for high school
22	$\eta_{1,dr}$	Human capital: work experience for dropout
23	$\eta_{2,clg}$	Human capital: work experience square for college
24	$\eta_{2,hs}$	Human capital: work experience square for high school
25	$\eta_{2,dr}$	Human capital: work experience square for dropout
26	η_3	Human capital: ag
27	η_4	Human capital: age squar
28	ς_0	St.dev. in shock distribution: constant
29	ς_1	St.dev. in shock distribution: ag
30	ρ_{clg}	Superannuation lumpsum: human capital for college
31	ρ_{hs}	Superannuation lumpsum: human capital for high school
32	ρ_{dr}	Superannuation lumpsum: human capital for dropouts
33	tr	Transfer from parents between t_0 and age 23
Initial conditions and smooth simulator parameters		
34	ς_{w_0}	Sigma parameter in log-normal distribution of initial asset
35	p_{clg}	High type proportion college
36	p_{hs}	High type proportion high school
37	p_{dr}	High type proportion dropout
Fixed and calibrated parameters		
	20	Credit constraint (in \$1000)
	4%	Interest rate on savings, r
	85	Compulsory retirement age

Table 8: (continued)

Value	Description
0.1	Minimum consumption (in \$1000) for not working, c_{min}
23	Minimum age of college student working career beginning
23	Maximum age to receive money transfer from parents
65	Minimum age to receive pension benefits
65	The age when lump sum payment of superannuation is received
0.1	Smoothing parameter in smooth maximum function in pension equation, ν
25	Young age threshold in disutility of work moderation multiplier
40	Old age threshold in disutility of work moderation multiplier
0.002	Smooth simulator: scale invariant smoothing parameter for $h^{(0)}$
0.005	Smooth simulator: scale invariant smoothing parameter for $h^{(1)}, \dots, h^{(5)}$

Notes: The table lists all estimated, calibrated and fixed parameters of the dynamic programming model, excluding parameters estimated outside of the model during the first stage of estimation. For the fixed parameters the values are also shown. The index is displayed only for the 38 structurally estimated parameters.

B Derivation of the Euler equation for the model

Conditional on the discrete choice h_t , the optimal consumption choices are characterized by the Bellman equation (16). Provided that the credit constraint is not binding, the first order conditions for this problem are given by

$$u'(c_t) = (1 - \delta_t)w'(M_t - c_t) + \delta_t \beta (\tau_{edu}) E \left[\sum_{h_{t+1} \in H_{t+1}(\tau)} P_{t+1}(h_{t+1}) \frac{\partial W_{t+1}(X_{t+1}, h_{t+1})}{\partial M_{t+1}} [1 + r - R^{pens}] \middle| X_t, c_t, h_t \right], \quad (31)$$

where $R^{pens} = \frac{\partial pens_{t+1}}{\partial M'_{t+1}} \cdot \mathbb{1}\{t+1 \geq 65\}$ is the effect of age pensions that depend on previous period consumption due to means testing. The choice probabilities $P_{t+1}(h_{t+1})$ are given by (18). Even though the choice specific value function $W_{t+1}(X_{t+1}, h_{t+1})$ may not be differentiable everywhere, as Iskhakov et al. (2017) show, the measure of the set of non-differentiable points is zero. By envelope theorem $\partial W_t(X_t, h_t)/\partial M_t$ is identical to the right hand side of (31) evaluated at the optimal level of consumption c_t . Therefore, it holds

$$\partial W_t(X_t, h_t)/\partial M_t = u'(c_t) \quad (32)$$

in every period, in particular in period $t+1$. Plugging (32) into (31) yields the Euler equation (19).

C Smoothed simulator of discrete choices

Consider first a simple example of simulating binomial discrete choice between two alternatives which have values v_1 and v_2 . We devise a function $\mathcal{S}(v_1, v_2, \phi)$ such that for any v_1, v_2 and the *smoothing parameter* ϕ it holds:

1. $\mathcal{S}(v_1, v_2, \phi)$ is a smooth function of v_1 and v_2 .
2. $\mathcal{S}(v_1, v_2, \phi)$ converges to the indicator function $\mathbb{1}\{v_1 \geq v_2\}$ as $\phi \rightarrow 0$
3. $\mathcal{S}(v_1, v_2, \phi)$ is invariant to scale of the values, i.e. $\mathcal{S}(\alpha v_1, \alpha v_2, \phi) = \mathcal{S}(v_1, v_2, \phi)$, $\alpha > 0$.

Denote $v_m = \max(v_1, v_2)$. One candidate smoothing function \mathcal{S} is

$$\mathcal{S}^{(1)}(v_1, v_2, \phi) = \frac{\exp(\frac{v_1}{\phi v_m})}{\exp(\frac{v_1}{\phi v_m}) + \exp(\frac{v_2}{\phi v_m})} = \frac{1}{1 + \exp(\frac{v_2 - v_1}{\phi v_m})}. \quad (33)$$

It is easy to verify that desired properties are satisfied as long as $v_m > 0$. The latter can be corrected for with a slightly different specification of the smoothing function, namely

$$\mathcal{S}^{(2)}(v_1, v_2, \phi) = \frac{\exp(\frac{v_1}{\phi|v_m|})}{\exp(\frac{v_1}{\phi|v_m|}) + \exp(\frac{v_2}{\phi|v_m|})} = \frac{1}{1 + \exp(\frac{v_2 - v_1}{\phi|v_m|})} = \frac{1}{1 + \exp(\text{sign}(v_m) \frac{v_2 - v_1}{\phi v_m})}, \quad (34)$$

where $|\cdot|$ denotes absolute value. Then all three desired properties are satisfied as long as $v_m \neq 0$.

Smoothing functions $\mathcal{S}^{(1)}(v_1, v_2, \phi)$ and $\mathcal{S}^{(2)}(v_1, v_2, \phi)$ are straightforward to generalize to multinomial choice. It is worth noting though that *uniform* smoothing with a single smoothing parameter ϕ may not be desirable in particular applications.

When applied to multinomial choice of hours in our model, uniform smoothing inflates the proportion of working because it is composed on multiple categories (several positive levels of hours). To correct for that, we devise the following nested smoothing function. Let (v_1, \dots, v_5) denote the values of 5 choices of positive hours, and v_0 denote the value of not working. Further, let $v_{m5} = \max\{v_1, \dots, v_5\}$ and $v_m = \max\{v_0, v_{m5}\}$. Then

$$\mathcal{S}_0(v_0, \dots, v_5, \phi_1, \phi_2) = \frac{\exp(\frac{v_0}{\phi_1|v_m|})}{\exp(\frac{v_0}{\phi_1|v_m|}) + \exp(\frac{v_{m5}}{\phi_1|v_m|})} = \frac{1}{1 + \exp(\frac{v_{m5} - v_0}{\phi_1|v_m|})}, \quad (35)$$

$$\mathcal{S}_i(v_0, \dots, v_5, \phi_1, \phi_2) = \frac{\exp(\frac{v_{m5}}{\phi_1|v_m|})}{\exp(\frac{v_0}{\phi_1|v_m|}) + \exp(\frac{v_{m5}}{\phi_1|v_m|})} \cdot \frac{\exp(\frac{v_i}{\phi_2|v_{m5}|})}{\sum_{j=1}^5 \exp(\frac{v_j}{\phi_2|v_{m5}|})}, \quad i = 1, \dots, 5, \quad (36)$$

where \mathcal{S}_0 is the *smoothed indicator* for not working and \mathcal{S}_i , $i = 1, \dots, 5$ are *smoothed indicators* for the 5 levels of hours. We use two different smoothing parameters ϕ_1 and ϕ_2 to allow for flexible adjustment of the amount of smoothing to the number of alternatives. It is not hard to show that (1) $\mathcal{S}_i(v_0, \dots, v_5, \phi_1, \phi_2)$ is a smooth function of v_0, \dots, v_5 as long as $v_{m5} \neq 0$ and $v_m \neq 0$, (2) $\mathcal{S}_i(v_0, \dots, v_5, \phi_1, \phi_2)$ converges to the appropriate indicators when $\phi_1 \rightarrow 0$ and $\phi_2 \rightarrow 0$, and (3) $\mathcal{S}_i(v_0, \dots, v_5, \phi_1, \phi_2)$ is invariant to scale of v_0, \dots, v_5 .