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

Fedor Iskhakov†   Michael Keane‡

March 17, 2019

Abstract

We structurally estimate a life-cycle model of consumption, labor supply and retirement, using data from the Australian HILDA panel. We use the model to evaluate effects of Australia's Age Pension system and income tax policy on labor supply, consumption and retirement. Our model accounts for human capital, savings, uninsurable wage risk and credit constraints. We account for "bunching" of hours by assuming a discrete set of hours levels, and we investigate labor supply on both the 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 Australia's Age Pension system as currently designed is poorly targeted. Our simulations suggest that a doubling of taper rates, combined with a 5.9% reduction of income tax rates, would be budget neutral and Pareto improving.

Keywords: Labor supply, human capital, retirement, pensions, income taxes, dynamic structural model, superannuation, Age Pension, means tests

JEL codes: J22, J24, J26, C63

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 Age 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).[1]

The defined contribution aspect of the Australian system, known as "Superannuation," takes a large part of Social Security taxes and payments off the federal government budget.[2] However, in contrast to a completely privatized system, the Age Pension provides fairly generous insurance for those who accumulate inadequate assets to finance retirement. A measure of the success of the superannuation scheme is that government spending on the Age Pension is only 2.9% of GDP, less than half than OECD average (Chomik and MacLennan, 2014). Nevertheless, the Age Pension is a very large program in the Australian context. In 2014 it cost \$60 billion (Australian dollars), compared to total federal income taxes of \$180 billion.

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 Age 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 study the impact of the US Social Security (SS) system on life-cycle labor supply and asset accumulation.[4] A key difference is that, while SS is a universal system, the Age Pension is a means tested transfer (or "welfare") program where participation is

---

[1]The 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).

[2]Participation in superannuation is mandatory. Currently, employers contribute an amount equal to 9.5% of workers' salaries. Contributions are managed by a large number of private investment funds ("super funds"), and workers can take the accumulated balance as a lump sum or annuity at retirement.

[3]Keane 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 discrete levels.

[4]Keane 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. A 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).

endogenous in the sense that eligibility depends on earnings and assets. As Tran and Woodland (2014) point out, a means tested system affects behavior through more channels than a universal system. Indeed, a universal system is a special case of a means tested system in which the "taper rates" at which benefits are reduced with earnings and assets are set at zero.

We estimate our model via the method of simulated moments (MSM) using a sample of males aged 19 to 89. Our data source is the "Household Income and Labor Dynamics in Australia Survey" (HILDA), which contains socio-economic data on roughly 20,000 households collected annually since 2001. The moments we match include life-cycle labor supply profiles (i.e., proportions at each of the six hours levels at each age), profiles of wages and wage dispersion, wealth profiles, superannuation balances, and annual transition rates between employment states.

The continuous choice of consumption combined with the discrete choice of hours renders the agents' problem non-convex. There are kinks in the value functions, and the optimal consumption policy has discontinuities.[5] 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).[6] For instance, the Frisch elasticity is only 0.30 for college workers at age 30, but it grows to roughly 2.2 at age 65. Similarly, if we look at the effect of unanticipated permanent wage changes, Marshallian elasticities are very small prior to age 45, but grow to about 0.80 at age 60, and 1.75 at age 65.

In one policy experiment we consider complete elimination of the Age Pension. This allows a lowering of income tax rates by 37% in a budget neutral scenario. Our welfare calculations indicate that, ex ante, the large majority (92%) of agents in the model would prefer to be born into a world without the Age Pension (and with the lower tax rate). We take this to indicate that the Age Pension is not well targeted. Indeed, income and asset taper rates of the program are low enough that about 75% of Australians over 65 receive some benefits.

Hence, we also consider an experiment where we double the taper rates on income and assets. This increases labor supply of age 65+ college graduates, reducing their dependence on the Age Pension. The reverse pattern holds for dropouts. Thus, the program becomes better targeted.

---

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

[6]Imai 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.

When combined with a 5.9% income tax cut to maintain budget neutrality, this policy is Pareto improving. A phase in of higher taper rates is consistent with NCOA (2014) recommendations.

The outline of the paper is as follows: In Section 2 we describe the Australian social security system. Sections 3 and 4 present our model and the DC-EGM solution method. Sections 5 and 6 describe the data and the estimation method. Section 7 present the estimates. 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

Australia's retirement income system consists of two parts. The first is the "superannuation" system, which involves mandatory employer contributions paid into individual pension accounts, managed by private firms known as "super funds." The mandatory contribution rate was 9% of earnings from July 1, 2002 through June 30, 2013, which covers the bulk of our sample period.[7]

The second component is the means-tested *Age Pension*, funded from general revenue. The pension is paid from the age of 65, and benefits are calculated according to the formula:

$$
\begin{aligned}
\text{pension} &= \max\big\{0, \text{full benefit} - \max\{\text{income test}, \text{asset test}\}\big\}, \\
\text{income test} &= \max\big\{0, t^I \cdot (\text{income} - \text{income threshold})\}\big\}, \\
\text{asset test} &= \max\big\{0, t^A \cdot (\text{wealth} - \text{asset threshold})\}\big\}.
\end{aligned}
$$

Thus, the full benefit is reduced by income and assets above the thresholds,[8] based on the income and asset taper rates, t^I and t^A, respectively. Taper rates are rather low, so about 70% to 75% of workers aged 65+ received some Age Pension benefit in 2012 (FaHCSIA, 2012).

At age 65, the private super accounts can be accessed. About half of workers take the funds in a lump sum, while most others transfer them to phased withdrawal products with no longevity protection and no upper limit on spending (Australian Prudential Regulation Authority, 2014).

Age Pension benefits are indexed to the maximum of price and wage growth twice each year, and the means-test thresholds are adjusted annually to changes in the CPI or the Pensioner and Beneficiaries Cost of Living Index (PBCI), whichever is higher.

---

[7]A 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. Employers and employees may sometimes make additional voluntary contributions (encouraged by tax concessions), but we ignore this.

[8]Different 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. Different assets are also treated differently by 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.

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

(a) Maximum pension benefit, $1000

(b) Tapers of income means testing
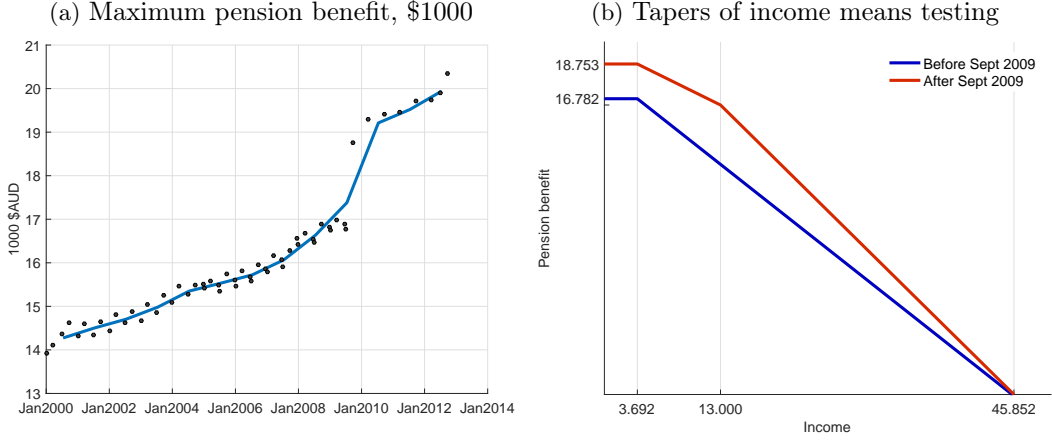
Notes: The left panel shows the real maximum Age Pension benefit from 2000 to 2012. The dots are the actual levels, and the line shows annual weighted averages. The right panel shows changes in taper rates and thresholds for the income test introduced in Sept 2009.

In September 2009 the Australian government introduced several changes in Age Pension rules: (a) the full annual benefit was increased by $1689, (b) the income taper rate was increased from 40% to 50%, (c) the first $13,000 in annual earnings was subject to a lower taper rate of 25% (Commonwealth of Australia, 2009). The policy changes are illustrated in Figure 1.

3 The model

In our model an individual, who has completed schooling, makes annual decisions on consumption and labor supply. 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. More precisely:

In each year starting at age $t = t_0 \geq 0$, 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

$$h_t \in H = \{h^{(0)}, ..., h^{(5)}\} \tag{1}$$

where $H = \{0, 1000, 2000, 2250, 2500, 3000\}$ hours per year.[9]

---

[9]This corresponds to $H = \{0, 20, 40, 45, 50, 60\}$ hours per week. We find 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).

The hourly offer wage rate is the product of human capital K_t, the rental price of a unit of human capital R_t and a log-normally distributed idiosyncratic wage shock $\varepsilon_t^{wage} \sim lnN(0, \sigma_t^{wage})$:

$$wage_t = K_t \cdot R_t \cdot \varepsilon_t^{wage}. \tag{2}$$

The shock ε_t^{wage} is revealed after the labor supply decision is made. We let the variance of wage shocks change over the life cycle, $\sigma_t^{wage} = \varsigma_0 + \varsigma_1 t$. The market for human capital is perfect, so all workers face the same rental price R_t. It is constant over time and normalized to one.

Human capital K_t is a function of age and work experience $\mathcal{E}_t$. It also depends on a worker's education level, indexed by $\tau_{edu} \in \{1, \ldots, J^{edu}\}$ and, as in Keane and Wolpin (1997), their unobserved skill endowment $\tau_{uh} \in \{1, \ldots, J^{uh}\}$. To simplify notation, denote *type* as $\tau = (\tau_{edu}, \tau_{uh})$. We allow for $J^{uh} = 2$ unobserved skill types ("high" and "low") and $J^{edu} = 3$ levels of education: high school dropouts ("dr"), high school graduates ("hs") and college graduates ("cg").

We define $\mathcal{E}_t \in [0, 1]$ as the fraction of total time budget $t \cdot h^{(5)}$ devoted to work up to and including period $t - 1$.[10] The law of motion for $\mathcal{E}_t$ is given by the recursive expression:

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

$\mathcal{E}_t \in [0, 1]$ spans the unit interval in every time period, so it far more convenient to use it, rather than human capital, as a state variable. The human capital production function is:

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

where $t \cdot \mathcal{E}_t$ is total work experience. The η are free parameters. The initial level of human capital is determined by the sum of the education specific intercept $\eta_0(\tau_{edu}) \in \{\eta_{0,cg}, \eta_{0,hs}, \eta_{0,dr}\}$, and the unobserved type specific intercept $\eta_0(\tau_{uh}) \in \{0, \eta_{0h}\}$. The mapping from experience to human capital, governed by $\eta_1(\tau_{edu})$ and $\eta_2(\tau_{edu})$, varies by education level.

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

$$M_{t+1} = (M_t - c_t)(1 + r) + h_t \cdot wage_{t+1} - Tax(h_t \cdot wage_{t+1})$$
$$+ pens_{t+1} \cdot \mathbb{1}\{t + 1 \geq 65\} + super_{t+1} \cdot \mathbb{1}\{t + 1 = 65\} \tag{5}$$
$$+ tr \cdot \mathbb{1}\{t + 1 \leq 23\}$$

The term $(M_t - c_t)$ is assets carried over from period t, at interest at rate r. The next two terms

---

[10]Here we use t to denote the period index ranging from 0 to $T - t_0$, rather than age ranging from t_0 to T.

are labor earnings and the tax on those earnings. The last three terms are non-labor income.

In writing (5) we adopt the timing assumption that the wage shock ε_{t+1}^{wage} is realized *after* time t labor supply is chosen. Thus the labor supply decision is based of the expected wage rate. Labor income is received at the end of the period and becomes part of M_{t+1}.[11]

Both the tax function (Tax) and the Age Pension benefit rule $(pens_{t+1})$ in equation(5) are estimated from the data in a first stage (see Section 7). This enables us to circumvent modelling details of the tax and benefit rules, such as how deductions and exemptions are determined.

The term $super_t$ in equation (5) denotes the one-time withdrawal of the accumulated balance in the super account at age 65, the earliest age for withdrawal without tax penalties.[12] To avoid modelling details of voluntary contributions (which few workers make), we assume the super account balance is proportional to the age 65 level of human capital, which in turn is a deterministic function of the initial skill endowment and total labor supply $(t \cdot \mathcal{E}_t)$, consistent with the fact that mandatory contributions are proportional to earnings.[13] Specifically, we assume:

$$super_t = \rho_1(\tau_{edu}) \cdot K_{65} \tag{6}$$

where the parameters $\rho_1(\tau_{edu}) \in \{\rho_{cg}, \rho_{hs}, \rho_{dr}\}$ are structurally estimated. Our simple treatment of the superannuation system is internally consistent, as agents take the formula in equation (6) into account when making labor supply and consumption choices throughout the life cycle.[14]

The last term in equation (5) is a transfer from parents, tr, that workers receive in each period from t_0 to age 23. We include this because Keane and Wolpin (2001) show that in a model with liquidity constraints parental transfers help to explain the relative low employment rates of young workers. We assume the parameter tr is the same across types.

Finally, agents in our model are endowed with an initial asset level M_0. This is given by $M_0 = tr + \tilde{M}_0$ which is the sum of the t_0 transfer from parents and a log-normal random variable $\tilde{M}_0 \sim lnN(0, \varsigma_{w_0})$. The M_0 can be interpreted as a one time bequest from parents received at

---

[11]As we will see, this timing assumption lets us put the model in Rust (1987)'s framework, where the only unobservables seen by agents and not the econometrician at the time choices are additive extreme value errors, giving multinomial logit choice probabilities. This contrasts with Keane and Wolpin (1997), where wage draws are known by agents when choices are made, so choice probabilities involve integrals over those draws.

[12]Our assumption of lump sum withdrawal is not too restrictive, as the majority of the Australian retirees either take out lump sums, or obtain full control to draw down their accumulated pension wealth. Other withdrawal options, such as annuities, are much less common, and modelling them is beyond the scope of this paper.

[13]Thus our model abstracts from voluntary contributions. The assumption that the super balance is a determinstic function of human capital means we do not need to introduce additional state and choice variables to model the evolution of the super balance. Feng et al. (2014, p. 14) report that only about a quarter of employees made voluntary contributions 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.

[14]Because we abstract from 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 are invariant to the experiment.

$t = t_0$. But it is observationally equivalent if M_0 and tr are received from a different source.

Having completely specified technology and constraints, we now turn to the specification of preferences. 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. We assume a CRRA form with parameter $\zeta > 0$ for $u(c_t)$:

$$u(c_t) = \frac{c_t^{1-\zeta} - 1}{1 - \zeta}, \tag{7}$$

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, $\gamma = (\gamma^{(1)}, \ldots, \gamma^{(5)})$. These coefficients are allowed to vary with age and the unobserved heterogeneity type. We have:

$$v_t(h_t) = \mathbb{1}\{h_t > 0\} \cdot \kappa_{type}(\tau_{uh}) \cdot \kappa_{age}(t) \cdot \gamma(h_t), \tag{8}$$

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

$$\kappa_{type}(\tau_{uh}) = 1 + \kappa_1 \cdot \mathbb{1}\{\tau_{uh} = \text{low}\}, \tag{9}$$

$$\kappa_{age}(t) = 1 + \kappa_2 \cdot (t - 40)^2 \cdot \mathbb{1}\{t > 40\} + \kappa_3 \cdot (t - 25) \cdot \mathbb{1}\{t < 25\}. \tag{10}$$

As we abstract from job availability and involuntary job loss, estimates of the parameters $\gamma^{(1)}, \ldots, \gamma^{(5)}$ may 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 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.

Note that out model includes unobserved heterogeneity, captured by two discrete types in the intercepts $\eta_0(\tau_{uh})$ of the human capital production function (4) and the disutility of work parameter κ_1 in (9). We call the types "high" and "low." We let the "high" type proportions $(p_{\text{cg}}, p_{\text{hs}}, p_{\text{dr}})$, which are estimated, differ by education level.

If an agent dies before the start of period $t + 1$ he leaves a bequest $b_t = M_t - c_t$. The utility from bequests is given by a CRRA function with parameter $\xi > 0$:

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

We let utility from bequests depend on $b_t + a_0$, rather than b_t itself, as this is guaranteed to be non-negative.[15] This translation implies that agents do not get infinite disutility from leaving zero bequests, which allows us to fit zero or small negative bequests in the data.

---

[15] $b_t + a_0 \geq 0$ because the credit constraint $c_t \leq M_t + a_0$ holds in every period.

We assume preferences are additively separable within period and over time. Let $\beta(\tau_{edu})$ denote the discount factor, which is education level specific, and let δ_t denote the annual survival probability. Let $V_t(X_t)$ denote the maximum of expected discounted utility over the remaining life cycle, given the state $X_t = (M_t, \mathcal{E}_t, \tau)$ of the decision maker at period t. Then the intertemporal decision problem is characterized by the *Bellman equation*:

$$V_t(X_t) = \max_{\substack{0 \leq c_t \leq M_t + a_0, \\ h_t \in H_t(\tau)}} \left\{ u(c_t) - v_t(h_t, \tau_{uh}) + \right.$$
$$\left. + \delta_t \beta(\tau_{edu}) E\left[V_{t+1}(X_{t+1}) \,|\, X_t, c_t, h_t\right] + (1 - \delta_t) \cdot B(M_t - c_t) \right\}, \tag{12}$$

where the expectation is over the ε_{t+1}^{wage}. The solution of (12) is given by the decision rules

$$c_t^\star(X_t): \; X_t \rightarrow [0, M_t + a_0],$$
$$h_t^\star(X_t): \; X_t \rightarrow H_t(\tau) \subset H, \tag{13}$$

which map the state X_t into optimal choices of consumption and labor supply.[16] These choices are deterministic in this model. In Section 4 we introduce taste shocks that generate multinomial logit choice probabilities for labor supply.

Finally, we assume agents cannot work after age 85, so $H_t(\tau) = \{h^{(0)}\}$ for $t \geq 85$. We also set $t_0 = 19$ for high school graduates and dropouts, and $t_0 = 23$ for college graduates.

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 and additional state variables. As in the original EGM, the main idea of DC-EGM is to start with an exogenous grid over the action space, which is then mapped point-by-point into the *endogenous* grid over the state space using the (analytically) inverted Euler equation. This contrasts with methods that approximate value functions over an exogenous grid in the state space. The EGM method is appealing in our case as it avoids costly root finding operations that would be required to find optimal consumption at each point in an exogenous state space grid.

For concave maximization problems the first order conditions of the Euler equation are sufficient. But problems containing mixed discrete and continuous choices (like ours) are in general non-concave, and solution methods cannot rely on first order conditions alone. Fella (2014) applies EGM to non-concave problems by first identifying the regions where first order

---

[16]We disallow the choice of zero hours if it results in optimal consumption below \$100.

conditions are not sufficient, and using traditional solution methods in those regions. Instead, Iskhakov et al. (2017) develop an algorithmic approach which automatically detects regions of non-concavity by studying the shape of the choice-specific value functions.

Iskhakov et al. (2017) show that in *deterministic* discrete-continuous dynamic choice problems (like that in Section 3) non-concavities accumulate in the backward induction, leading to numerous kinks in value functions and discontinuities in the policy functions. The accumulation of these kinks and discontinuities complicates the numerical solution. They also show that the introduction of discrete choice specific random utility components in the tradition of McFadden (1973) and Rust (1987) alleviates this problem. Yet, as models with discrete-continuous choices are not necessarily fully concave even with this 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, which ensures its speed and accuracy.[17]

As in Iskhakov et al. (2017), we assume discrete hours choices are affected by additional state variables (i.e., "taste shocks") known by agents but not by the econometrician. These taste shocks are added to the deterministic utilities of the hours levels. Denote the 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, $\varepsilon(h_t) = \varepsilon_t^{(i)} \Leftrightarrow h_t = h^{(i)}$, $i \in \{0, \dots, 5\}$. The vector ε_t is independent multivariate extreme value distributed with scale parameter λ. It is independent of wage shocks ε_t^{wage} and over time.

By modifying the Bellman equation in (12) to include these taste shocks we obtain:

$$V_t(X_t) = \max_{h_t \in H_t(\tau)} \left\{ W_t(X_t, h_t) + \lambda\varepsilon(h_t) \right\}, \tag{14}$$

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

$$
\begin{aligned}
W_t(X_t, h_t) &= \max_{0 \le c_t \le M_t + a_0} \Big\{ u(c_t) - v_t(h_t, \tau_{uh}) + (1 - \delta_t)B(M_t - c_t) \\
&\quad + \delta_t\beta(\tau_{edu})E\Big[\max_{h_{t+1} \in H_{t+1}(\tau)} \{ W_{t+1}(X_{t+1}, h_{t+1}) + \lambda\varepsilon_{t+1}(h_{t+1}) \} | X_t, c_t, h_t \Big] \Big\} = \\
&= \max_{0 \le c_t \le M_t + a_0} \Big\{ u(c_t) - v_t(h_t, \tau_{uh}) + (1 - \delta_t)B(M_t - c_t) \\
&\quad + \delta_t\beta(\tau_{edu})E\Big[\mathcal{LS}_\lambda_{h_{t+1} \in H_{t+1}(\tau)} \big(W_{t+1}(X_{t+1}, h_{t+1}) \big) \Big| X_t, c_t, h_t \Big] \Big\},
\end{aligned}
\tag{15}
$$

and $\mathcal{LS}_\lambda(x_1, \dots, x_k)$ is the logsum function.

---

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

The solution to (15)-(14) gives the optimal decision rules

$$
\begin{aligned}
c_t^\star(X_t): \ & X_t \to [0, M_t + a_0], \\
h_t^\star(X_t): \ & X_t \to \Big(P_t(h^{(0)}), \ldots, P_t(h^{(6)}) \Big), P_t(h_t) = 0 \ \forall h_t \notin H_t(\tau),
\end{aligned}
\tag{16}
$$

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

$$
P_t(h^{(i)}) = \frac{\exp \big(W_t(X_t, h^{(i)})/\lambda \big)}{\sum_{h_t' \in H_t(\tau)} \exp \big(W_t(X_t, h_t')/\lambda \big)}.
\tag{17}
$$

Contrast the probabilistic solution for hours in (16)-(17) with the deterministic solution in (13).

The DC-EGM algorithm relies on the Euler equation, which is a necessary condition despite the presence of discrete choices (Clausen and Strub, 2016). To derive the Euler equation, note that, *conditional* on the discrete hours choice h_t, optimal consumption choices are characterized by the Bellman equation (15). Provided that the credit constraint is not binding, the first order conditions for the consumption problem are given by:

$$
\begin{aligned}
u'(c_t) \ = \ & (1 - \delta_t)B'(M_t - c_t) \\
& + \delta_t \beta(\tau_{edu}) E \Big[\sum_{h_{t+1} \in H_{t+1}(\tau)} P_{t+1}(h_{t+1}) \frac{\partial W_{t+1}(X_{t+1}, h_{t+1})}{\partial M_{t+1}} \big[1 + r - R^{pens} \big] \Big| X_t, c_t, h_t \Big],
\end{aligned}
\tag{18}
$$

where the term $R^{pens} = \frac{\partial pens_{t+1}}{\partial M_{t+1}'} \cdot \mathbb{1}\{t+1 \geq 65\}$ arises because the age pension depends on previous period consumption due to means testing. The choice probabilities in (18) $P_{t+1}(h_{t+1})$ are given by (17). 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 (18) evaluated at the optimal level of consumption c_t. Therefore, it holds

$$
\partial W_t(X_t, h_t)/\partial M_t = u'(c_t)
\tag{19}
$$

in every period, in particular in period $t + 1$. Plugging (19) into (18) yields the Euler equation:

$$
\begin{aligned}
u'(c_t) \ = \ & (1 - \delta_t)B'(M_t - c_t) \\
& + \delta_t \beta(\tau_{edu}) E \Big[\sum_{h_{t+1} \in H_{t+1}(\tau)} P_{t+1}(h_{t+1}) u'(c_{t+1}) \big[1 + r - R^{pens} \big] \Big| X_t, c_t, h_t \Big].
\end{aligned}
\tag{20}
$$

Given our CRRA specification of utility in (7), equation (20) can be inverted analytically to yield the level of current peiod consumption c_t for every level of *end-of-period* assets $A_t = M_t - c_t$.

This completely characterizes the asset level at period $t+1$. Analytical invertibility of the Euler equation ensures the speed of our numerical solution.

The additional state variables in X_t, namely work experience $\mathcal{E}_t$ and agent type τ, do not hinder the application of DC-EGM as long as their transitions do not depend on the consumption choice.[18] In our model, the type is time invariant, and the deterministic evolution of $\mathcal{E}_t$ only depends on the labor supply. Thus, similar to Bellman equation (12), the expectation in (20) is taken over wage shocks. Therefore, the only additional step required in the computation of the right hand side of the Euler equation is interpolation over the next period work experience $\mathcal{E}_{t+1}$. The fact that $\mathcal{E}_t$ is bounded to the unit interval in every t eliminates the need for extrapolation. We discretize $\mathcal{E}_t$ with simple linear grid which includes 0 and 1, and use linear interpolation to approximate the values between the grid points.

We now describe in detail how we solve the model using the DC-EGM algorithm:

In the terminal period T, optimal consumption $c_T(X_T, h_T)$ and the choice specific value functions $W_T(X_T, h_T)$ for every value of τ, every point of the grid over $\mathcal{E}_T$, and conditional on h_T, are given by the solution to the static problem:

$$
c_T(X_T, h_T) = \underset{0 \le c \le M_T + a_0}{\arg\max} \Big\{ u(c) + B(M_T - c) \Big\}, \tag{21}
$$

$$
W_T(X_T, h_T) = u\big(c_T(X_T, h_T)\big) + B\big(M_T - c_T(X_T, h_T)\big). \tag{22}
$$

In all preceding periods $t < T$ and for all $h(t) \in H_t(\tau)$, τ and points of the grid over $\mathcal{E}_T$, we take as given the (already computed) solutions for $c_{t+1}(X_{t+1}, h_{t+1})$ and $W_{t+1}(X_{t+1}, h_{t+1})$, and we perform the following steps:

1. Let $\{A_1, \ldots, A_J\}$ denote the exogenously fixed monotonic grid over "savings," defined as $A_t = M_t - c_t$, with $A_1 = -a_0$ and $A_j < A_{j+1}$.

2. For each point A_j we calculate the right hand side (RHS) of equation (20) by integrating over the next period wage shocks ε_{t+1}^{wage} using Gaussian quadrature. For each quadrature point representing a possible value of ε_{t+1}^{wage} we compute:

   (a) Next period work experience $\mathcal{E}_{t+1}$ using (3) and current period labor supply $h(t)$;

   (b) Next period human capital K_{t+1} using (4);

   (c) Next period wage $wage_{t+1}$ using (2) and the given value for the wage shock;

   (d) Next period wealth M_{t+1} using (5) (Note that M_{t+1} only depends on M_t and c_t through $A_t = M_t - c_t$, which serves as a 'sufficient statistic' for period t);

---

[18]See Iskhakov (2015) for general conditions.

(e) Finally, optimal levels of consumption c_{t+1} using the policy function $c_{t+1}(X_{t+1}, h_{t+1})$ computed on the previous iteration (we are using two-dimensional linear interpolation over M_{t+1} and $\mathcal{E}_{t+1}$).

3. Once the RHS of the Euler equation is computed for given X_t, h_t and A_j, compute the optimal level of current consumption as $c_t(X_t, h_t) = (u')^{-1}(RHS)$ and plug it into (15) to also compute the current period $W_t(X_t, h_t)$.

4. Form the endogenous choice specific grids over current period wealth $\{M_1, \ldots, M_J\}$ as $M_j = A_j + c_t(X_t, h_t)$.

5. If the endogenous grid $\{M_1, \ldots, M_J\}$ is not monotone in savings $\{A_1, \ldots, A_J\}$ use the *upper envelope* algorithm from (Iskhakov et al., 2017) to eliminate the sub-optimal points corresponding to the Euler equation solutions which do not correspond to the highest values of $W_t(X_t, h_t)$ for given wealth.

6. Now $c_t(X_t, h_t)$ and $W_t(X_t, h_t)$ can be computed by interpolation on the (subset) of endogenous grid points $\{M_1, \ldots, M_J\}$ that form the upper envelope.

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 involve solving an optimization problem or finding roots of an equation (as the derivative of the CRRA utility function is analytically invertible).

5 Data

We estimate the model using the first 16 waves of the HILDA survey. The survey collects annual data on household income and labor supply, and it also includes several reoccurring modules on wealth and retirement, among other topics (Wooden and Watson, 2007). The first wave of HILDA was administered in 2001 to a large national probability sample of households that contained in total 19,914 individuals. 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 screens to limit the sample to male household heads with completed schooling[19] who are part of the labor force. We also restrict the data to ages 19 to 89. The resulting unbalanced

---

[19]We treat respondents as attending school if they are below their reported "age of leaving school," or if their reported fraction of calendar time in school is above 50% in a given year. Also, we treat respondents under age 25 as in school in any year where school status in both the previous year and following year is positive or unknown. This is to account for gap years popular in Australia.

Table 1: Description of the structural estimation sample.

	College graduates	Highschool graduates	Highschool dropouts	All
Number of individuals	2,391	5,254	2,488	10,133
Number of data points	20,207	41,965	19,025	81,197
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 *wage*	75.03%	65.77%	45.71%	63.38%
Non-missing *wealth*	22.68%	22.59%	22.14%	22.51%
Non-missing *super*	20.01%	17.66%	12.94%	17.14%

panel contains 10,133 male household heads born between 1912 and 1997 and observed between 2001 and 2016. There are 81,197 person-year observations.

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.[20] Table 1 contains summary of the estimation sample, with detailed description of data preparation available in the Online Appendix.

To construct our measure of annual hours we use the HILDA question on hours worked in a "typical" week along with the extensive record of time use in the previous year. The survey collects information on employment, unemployment and labor force status in the early, middle and late part of each month, for the period July through June. We calculate annual hours as "typical" weekly hours times 50 weeks times the fraction of the year the person reports being in the employed state (i.e., not unemployed or out of the labor force).[21]

To our knowledge this is the first study in the dynamic labor supply literature to accommodate the bunching of hours at discrete levels, which is evident in the data. We first categorize those working less than 500 hours in a year as "unemployed." Then we use a K-median clustering algorithm to partition the remaining workers into discrete hours levels. We decided to use $K = 5$ partitions as this provided an accurate description of the data. Table 2 presents the partitioning results. As we see, the medians of the five category partition that give the best fit to the data are 24, 40, 45, 50 and 60 hours per week. The largest partition with positive hours

---

[20]Due to occasional missing data on key variables (e.g., hours, consumption) there can be gaps in the observed life cycles. But only 17.83% of life cycles included such gaps.

[21]We impose a maximum of 4200 hours, which corresponds to 12 hours per day, 7 days per week for 50 weeks.

Table 2: Discrete levels of hours of labor supply.

	Hours in 6 levels per year	per week	N obs	Median per year	Median per week	Min	Max
0	0	0	26,411	0	0	0	500.00
1	1000	20	6,711	1200	24	500.03	1600.00
2	2000	40	23,387	2000	40	1600.02	2124.90
3	2250	45	7,622	2250	45	2133.36	2368.03
4	2500	50	12,115	2500	50	2375.10	2750.00
5	3000	60	8,368	3000	60	2750.10	4200.00

Notes: The table presents results of the k-median clustering algorithm for k=5. The discrete levels of hours in the model are based on the medians found by the cluster analysis.

Table 3: HILDA classifications within hours levels.

	Hours in 6 discrete levels					
Labor force status in HILDA	0	1	2	3	4	5
Employed FT	353	1,303	23,212	7,622	12,115	8,368
Employed PT	1,877	5,408	175	0	0	0
Unemployed, looking for FT	1,959	0	0	0	0	0
Unemployed looking for PT	257	0	0	0	0	0
OLF, marginally attached	2,888	0	0	0	0	0
OLF, not marginally attached	19,072	0	0	0	0	0
Total	26,411	6,711	23,387	7,622	12,115	8,368

Notes: The table reports how people in six labor market states are allocated to the six discrete hours levels. HILDA defines full-time work as 35 hours or more per week (usual hours).

is $k = 2$ the where hours range from 1600 to 2125 in the year, and the median is 2000.[22] Table 3 shows how our six hours levels correspond to labor market states reported in HILDA. The zero hours observations include both the unemployed and those who report being out-of-labor-force (OLF) states. It also includes some employed workers who report low hours of work.

We compute the hourly wage rate by taking the HILDA variable *last financial year gross wages and salaries* and dividing by our (continuous) measure of annual hours. We set the wage to missing if the calculated value falls outside of the interval from $5 to $150.

We calculate wealth using the HILDA wealth model, which is based on 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. Our measure of wealth is financial wealth plus non-financial wealth minus household debt minus combined household super, where the components are defined as in (Summerfield et al., 2013, pp. 71-75). We exclude the super balance because it is a separate variable in our model. Wages and wealth are converted to 2012 Australian dollars

---

[22]Although the median of the $k = 1$ partition is 24 hours per week, we decided to use 20 instead because, unlike other cells where the median and mode coincide, the mode of the k=1 cell was 20 hours per week. But the choice between 20 and 24 made little difference to our results.

using the CPI index from the Australian Bureau of Statistics.

The superannuation balance *profiles* are not used in the estimation. As 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.

6   Estimation Method and Moment Conditions

We estimate the structural parameters of our model using the method of simulated moments (MSM) (McFadden, 1989). We construct moments based on age profiles of labor supply, wages and assets and their simulated counterparts. Specifically, we match age profiles of the following variables: 1) the fraction of the population employed, 2) mean hours conditional on working, 3) the mean hourly wage conditional on working, 4) the variance of the wage conditional on working, 5) the skewness of log earnings conditional on working, 6)-9) the fractions of the population who work 20, 40, 45 or 50 hours per week, 10) mean wealth, 11) the fraction of the population who work in periods t and $t-1$, 12) the fraction of population who do not work in periods t and $t-1$, and, finally, 13) the superannuation balance at exactly age 65.

We aggregate data on wealth, variance of wages and skewness of earnings into 5-year intervals, and we disregard moments that are *conditional* on working after age 70 (due to small cell sizes). Moreover, we disregard age-education cells that contain less than 10 observations. Table 4 lists the number of moments provided by each of the thirteen variables listed above, for each education group. The three education groups are combined in estimation, giving 1592 moments, although a subset of parameters (human capital production function, discount rate, superannuation function and type proportions) are allowed to differ by education group.

Let θ deonte the complete vector of model parameters, and let $\mu(\theta)$ denote the vector of 1592 simulated moment conditions, composed of the stacked vectors of simulated moments $\mu_i^e(\theta)$ for all education levels $e \in \{\text{hs}, \text{dr}, \text{cg}\}$ and each of the $i \in \{1, \ldots, 13\}$ statistics listed above. The MSM estimator $\hat{\theta}$ minimizes the criterion $\mu(\theta)'W\mu(\theta)$ where W is the moment weighting matrix. Under standard assumptions, the MSM estimator $\hat{\theta}$ is consistent and asymptotically normal, i.e. $\sqrt{N}(\hat{\theta} - \theta_0) \sim N(0, \Sigma)$. The variance-covariance matrix Σ can be approximated by:

$$\hat{\Sigma} = (1 + N/N_{sim})(D'WD)^{-1}D'WSWD(D'WD)^{-1}, \tag{23}$$

where N is the number of observed individual life-cycles, N_{sim} is the number of simulated life cycles, 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 a consistent estimate of θ. Thus,

Table 4: Number of empirical moments.

Moments	High school Ages	N	Dropouts Ages	N	College Ages	N
work status	19 - 86	67	19 - 88	70	23 - 89	64
hours conditional on working	19 - 70	48	19 - 70	48	23 - 70	44
wage conditional on working	19 - 70	48	19 - 70	48	23 - 70	44
variance of wage ⋆	19 - 70	10	19 - 70	10	23 - 70	10
skewness of log-earnings ⋆	19 - 85	13	19 - 85	13	23 - 85	13
hours = 20	19 - 86	67	19 - 86	68	23 - 89	64
hours = 40	19 - 82	61	19 - 84	64	23 - 79	57
hours = 45	19 - 77	55	19 - 83	56	23 - 76	51
hours = 50	19 - 76	58	19 - 88	66	23 - 77	53
wealth ⋆	25 - 85	13	25 - 85	13	25 - 85	13
work to work	19 - 70	48	19 - 70	48	23 - 70	44
nowork to nowork	19 - 70	48	19 - 70	48	23 - 70	44
super	65	1	65	1	65	1
Total		537		553		502

Notes: The model is fit to life-cycle profiles of 13 variables. The table reports (by education level) the age range for which each variable is observed, and the number of moments generated. A ⋆ indicates the variable is aggregated to 5 year intervals.

simulation error inflates the standard method of moments (MOM) covariance matrix by the factor $1 + 1/\tau$, where $\tau = N_{sim}/N$ is the ratio of simulated to observed life-cycle histories.

We implement the MSM estimation procedure as follows: Given a trial value of parameter vector θ, we solve the dynamic model as described in Section 4, and obtain the decision rules $c_t^\star(X_t)$ and $h_t^\star(X_t)$. We then generate N_{sim} simulated individual life-cycle histories from the model, using independent draws for the stochastic terms. Then compute the moment conditions, i.e., the distance between statistics calculated from the N life-cycle histories in the data and those calculated from the N_{sim} simulated histories. We then construct the objective function by computing $\mu(\theta)'W\mu(\theta)$. Next we require a search algorithm to update the trial value of parameter vector θ so as to improve the objective. We stop when we are satisfied that the objective is minimized, and the result is the MSM estimate $\hat\theta$.

We use the POUNDerS optimization algorithm (Wild, 2015) to update θ as we seek to minimize the MSM objective.[23] POUNDerS is tailored to problems of minimizing a (weighted) sum of squared residuals. It works by approximating each moment condition as a quadratic function of the model parameters. It is derivative free, which is useful because computing derivatives is computationally expensive. But it relies on existence of derivatives for the quality of the approximation. The algorithm terminates if it succeeds in setting the (approximate) derivatives of the moments with respect to the parameters to zero (within tolerance).

---

[23]The acronym stands for practical optimization using no derivatives for sums of squares.

In our model labor supply choices are are discrete, which causes the simulated choice and hence the simulated moment conditions (and hence the MSM objective) to be piece-wise flat in the parameters. To ensure existence of derivatives, and to prevent the maximization loop from "stalling" in the flat areas, when we simulate data from the model we replace the maximum operators that produce simulated discrete choices with logsum (soft max) functions that produce "smoothed choices" that can be represented by near integer numbers.[24] Because the utilities of discrete choices calculated when the dynamic programming problem is solved are smooth functions of structural parameters, with our smooth simulator the MSM objective function is smooth as well. Our smooth simulation approach is closely related to Bruins et al. (2018) with the 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. The Appendix provides a complete description of our approach.

The optimal choice of the MSM weighting matrix is $W = S^{-1}$ in which case (23) 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.

To construct W, we let d_t and $s_t(\theta)$ denote empirical and simulated statistics in period t, and let $\mu_t(\theta) = d_t - s_t(\theta)$ denote the simulated moment. We restrict $Var(s_t(\theta))$ to be age invariant. Assuming orthogonality between sampling and simulation errors, the variance of the moment condition can be decomposed as $Var(\mu_t) = Var(d_t) + Var(s_t(\theta))$. We then have:

$$Var(s_t(\theta)) = \frac{1}{T}\sum_{\tau=t_0}^{T} Var(\mu_\tau(\theta)) - \frac{1}{T}\sum_{\tau=1}^{T} Var(d_\tau), \ \forall t = t_0, \ldots, T. \tag{24}$$

Substituting this for $Var(s_t(\theta))$ gives the following formula for diagonal element W_{μ_t} of the weighting matrix W corresponding to the moment condition $\mu_t(\theta)$:

$$W_{\mu_t} = Var(d_t) + \frac{1}{T}\sum_{\tau=1}^{T} Var(\mu_\tau(\theta)) - \frac{1}{T}\sum_{\tau=1}^{T} Var(d_\tau). \tag{25}$$

We estimate the variances on the right hand side of (25) using the data and the simulated moments at a consistent estimate of θ.

---

[24] All quantities dependent on the chosen options are then interpolated using these smoothed values (that sum to one by construction) as weights.

7 Estimation results

7.1 First stage estimates

The Age Pension rule, the tax function and the survival curve can be consistently estimated separately from the full structure, and treated as known within the structural estimation.

First, we use HILDA data on Age Pension receipt to estimate a model of pension payments conditional on covariates that appear in our structural model. Our preferred specification is:

$$
\begin{aligned}
pens_t = \mathcal{M}_\nu \Big(& \underset{(183.96)}{10,759.73} + \underset{(173.52)}{1846.92} \cdot \mathbb{1}\{year \geq 2010\} \\
& - \mathcal{M}_\nu \big[0, \mathcal{M}_\nu \big(\underset{(0.020)}{0.27749} \cdot wage_t, \underset{(0.0004)}{0.00499} \cdot (M_t' - \underset{(46,895.27)}{117,082.60})) \big], 0 \Big),
\end{aligned}
\tag{26}
$$

where $\mathcal{M}_\nu(x, y) = \nu \cdot log\big(\exp(x/\nu) + \exp(y/\nu) \big)$ is a soft max with $\nu = 0.1$, $\mathcal{M}_\nu(x, y) \to max(x, y)$ as $\nu \to 0$, and M_t' is wealth counted in the asset test, $M_t' = M_t - pens_t$, $t \geq 65$.[25]

The coefficient of 1846 on the indicator for $t \geq 2010$ captures the increase in the benefit level in 2010. In the structural estimation we treat the shift as unanticipated, and solve the model separately for the pre and post-2010 regimes. When an agent reaches the 2010 calendar year the simulator switches from using the optimal decision rules for the pre-2010 policy regime to those for the post-2010 regime. Obviously different people are at different ages when the shift occurs.

Next, we estimate the income tax function using tax and income data in HILDA.[26] We adopt this approach to avoid modelling the full complexity of the tax rules. Fortunately, the Australian income tax structure was very stable over our sample period, both in terms of rates and the bracket structure. Using non-linear least squares we obtained the following three bracket rule:

$$
\text{tax} = \begin{cases}
0, & \text{if income} < \text{thld}_1 = \underset{(0.10016)}{17.39184}, \\[2mm]
\underset{(0.00096)}{0.29907} \cdot (\text{income} - \text{thld}_1), & \text{if thld}_1 \leq \text{income} < \text{thld}_2, \\[2mm]
\underset{(0.00556)}{0.37930} \cdot (\text{income} - \text{thld}_2) + \underset{(0.00096)}{0.29907} \cdot \text{thld}_1, & \text{if income} \geq \text{thld}_2 = \underset{(0.59292)}{73.17661},
\end{cases}
\tag{27}
$$

Finally, we estimated survival probabilities from life tables, Australian Government Actuary (2009). We assume the probability of death is (perceived to be) zero up until age 40, and fit the following parametric model to the survival function for males between ages 40 and 90:

$$
\delta_t = 1 - \underset{(0.0000249)}{0.0006569} \left[\exp \left(\underset{(0.0008145)}{0.1078507}(age_t - 40) \right) - 1 \right] \text{ if } age_t \geq 40,
\tag{28}
$$

---

[25]Here and below, standard errors are in parenthesis. Smoothing results in a maximum error of $log(2)\nu = \$69$.

[26]To focus on the tax function for labor earnings, we used a sub-sample of individuals with small non-labor earnings (less than $500) and positive wage earnings (not exceeding $200,000).

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 multiplicator of the utility of bequest	0.68659	1.42044
β_{cg}	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

7.2 Structural estimates

The estimates of the structural parameters are reported in Tables 5 to 7. Table 5 reports the preference parameters. The CRRA parameter for consumption, ζ in (7), is estimated as 0.79. So utility is slightly less concave than logarithmic in consumption. In simple versions of the life-cycle model this would imply the substitution effect of wages on labor supply slightly dominates 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 (8) imply disutility is increasing in hours. But the increase is not monotonic, as the estimated disutility 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, but also the extent to which jobs with each level of hours are available.

The estimate of κ_1 in (9) implies the disutility of work is roughly 50% greater for the "low" type. The estimate of κ_2 in (10) 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 b_{scale} and ξ in (11) are imprecisely estimated, similar to what Imai and Keane (2004) found. This is due to the 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, behavior changes greatly if bequests are ignored entirely, suggesting

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

Parameter	Description	Estimate	Std.Err.
$\eta_{0,cg}$	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
$\eta_{1,cg}$	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
$\eta_{2,cg}$	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

they are important, but that the precise shape of the bequest function is difficult to pin down.

The education-specific discount factors are precisely estimated, ranging from .968 to .970. Finally, we report our estimate of the scale λ of the extreme value taste shocks in (14). This parameter is important in distributing workers across the 6 discrete hours levels.[27]

We report estimates of the human capital production function in Table 6. The offer wage function intercepts for the three education types differ as expected - e.g., the offer wage premium for college vs. high school, evaluated at zero work experience, is 0.22 log points, implying 25% higher initial offer wages. And the intercept shift for the "high" latent skill type, $\eta_{0,high}$ is .393, implying $\exp(.393) = 48\%$ higher offer wages, *ceteris paribus*.

Surprisingly, the offer wage difference between latent types is very imprecisely estimated, which is unusual for these type of models.[28] We explored this phenomenon carefully and came to the following conclusion: As we show later, the income distribution has fat left and right tails. Our model has some trouble capturing both tails with two types. We can vary $\eta_{0,high}$ over a fairly wide range (from about 0.30 to 0.50) with little impact on overall model fit, except that for lower values of $\eta_{0,high}$ we obtain a better fit to the left tail, while for higher values we obtain 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.[29]

---

[27]Note that as λ approaches infinity workers are more equally distributed across hours levels .

[28]See Keane and Wolpin (1997, 2001, 2010), and other related mixture models, where offer wage differences between latent types are very precisely estimated. A key difference is their models are fit to individual life-cycle wage and employment histories, while our model is fit to aggregate summary statistics.

[29]We are not too concerned with this phenomenon because (i) the fit to the overall wage and income distri-

Table 7: Estimates of Miscellaneous Structural 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_{cg}	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	1.48960	6.69399
p_{cg}	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

The next six rows of Table 6 report quadratics in work experience, which capture learning-by-doing. The effects of experience are allowed to differ for the three education groups. Consistent with results in Imai and Keane (2004) and Keane and Wasi (2016), 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 high school type. Thus, endogenous human capital accumulation over the life-cycle is more important for more educated workers.

The last two parameters in Table 6 are a quadratic in age, which captures *exogenous* variation in offer wages over the life-cycle. It is highly significant and quantitatively important, implying offer wages follow an inverted U-shaped path in age (just as in experience). Initially the age effect is positive 2.67% per year, but it 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 not picked up by work experience alone. For example, early in 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 so the model is not forced to explain all life-cycle variation in wages by work experience alone, thus potentially exaggerating returns to experience (which could, in turn, distort estimates of labor supply elasticities).

The remaining structural parameters are reported in Table 7. The parameters ς that determine how the standard deviation of wage shocks varies with age imply a U-shaped path. This is similar to the reduced form pattern found by Geweke and Keane (2000). The estimated transfer from parents, which we assume individuals receive annually between t_0 and 23, is \$5.5k. The

---

butions are 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.

superannuation process parameters from equation (6) imply, as expected, that the accumulated super balance is an increasing function of the level of accumulated human capital.

Table 7 also reports our estimate of ς_{w0}, the standard deviation of (log) initial wealth. This is \$1.5k but very imprecisely estimated. ς_{w0} 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, "excess" heterogeneity that is hard to explain given observed earnings).

Finally, Table 7 contains estimates of latent type proportions conditional on education. Interestingly, the probability a person is the "high" type is about 90% among the college educated, 80% among the high school 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

Here we describe the fit of the model to several key dimensions of the data:

Figure 2 presents model vs. data paths for average annual hours over the life-cycle. This average includes both working and non-working individuals. Notice the model provides a good fit to the life-cycle path of hours for high school workers. 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. 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 3 and 4 decompose total hours into the extensive margin (employment) and the intensive margin (hours conditional on employment). The model captures this decomposition quite well for all three groups. Figure 4 reveals that hours conditional on employment are fairly flat over the life-cycle. For example, for high school workers, average hours are 2250 per year at age 40, and this only drops to 1800 at age 70. As we see in Figure 3 it is the sharp drop in employment after age 50 that accounts for most of the decline in total hours we see in Figure 2.

Recall that agents can work 20, 40, 45, 50 or 60 hours per week in our model. Figure 5 shows that the model provides a very good fit to the distribution of workers across hours levels (by age) with a couple notable exceptions. In the data there is a jump in the number of people who work part-time (i.e., 20 hour per week) in their early to mid 60s. This pattern exists for all education groups, but is most pronounced for 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

Figure 2: Fitted hours of labor supply by age/education

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts
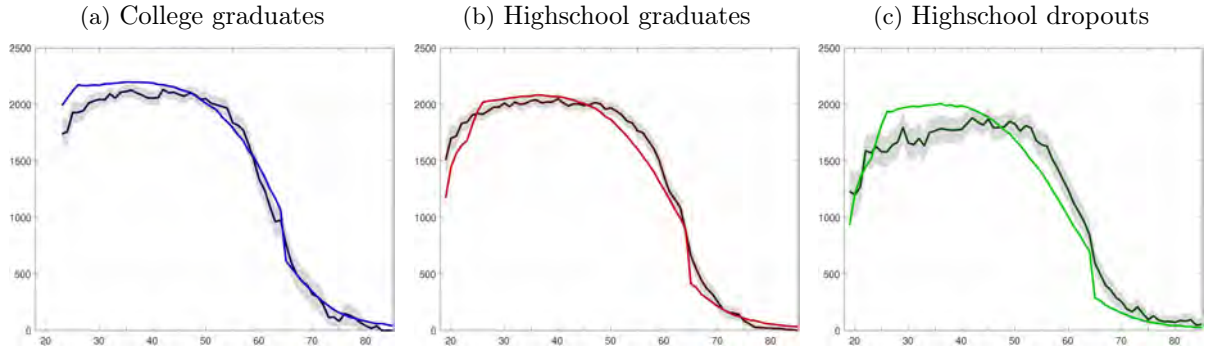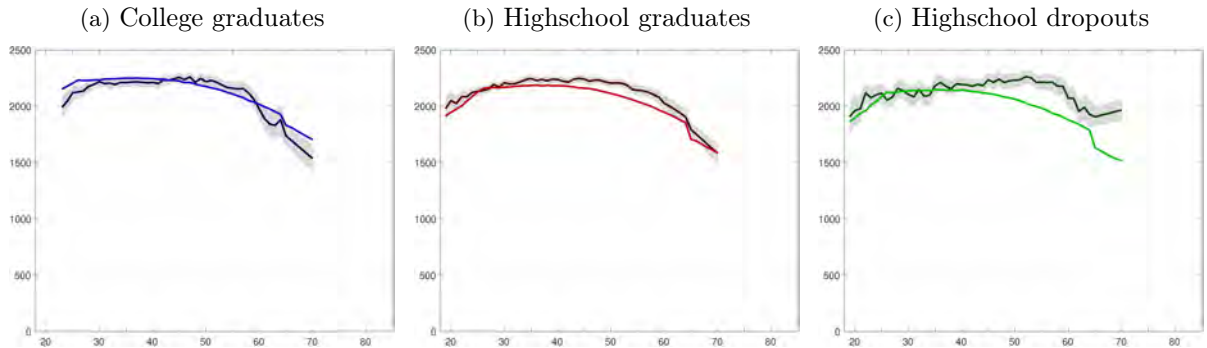

Notes: The plots show simulated life cycle profiles of hours of labor supply for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 3: Fitted employment rates by age/education

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts
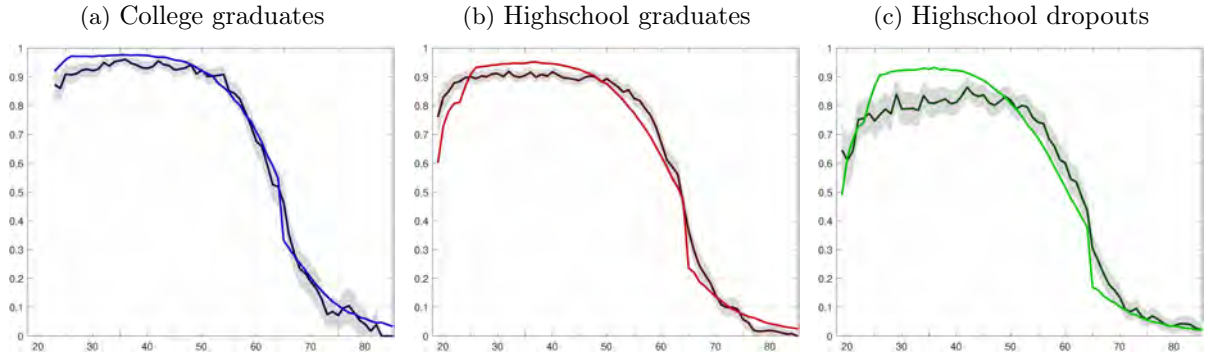
Notes: The plots show simulated life cycle profiles of fraction of working people for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 4: Fitted hours conditional on working.

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts

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.

pattern. However, we did not feel this shortcoming was serious enough to complicate the model by adding parameters designed specifically to capture this data feature. So we leave this as a topic for future research. The other shortcoming of the model is that it seriously understates the fraction of high school dropouts who work 60 hours per week in their 50s and 60s.

Figure 6 reports the fit of the model to the mean, variance and skewness of wages of employed workers, broken down by education and age. In our view, if one wants to credibly evaluate the impact of social insurance programs like the Age 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 6 reports results for average wage profiles. Clearly there are details of the life-cycle wage path for each education group 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 slightly from workers' first entrance into the labor market until their life cycle peak. But for college workers wages roughly double. High school workers are in between. The model does a very good job of capturing these broad differences. Obviously this is primarily because the estimates imply greater returns to experience for more educated workers.
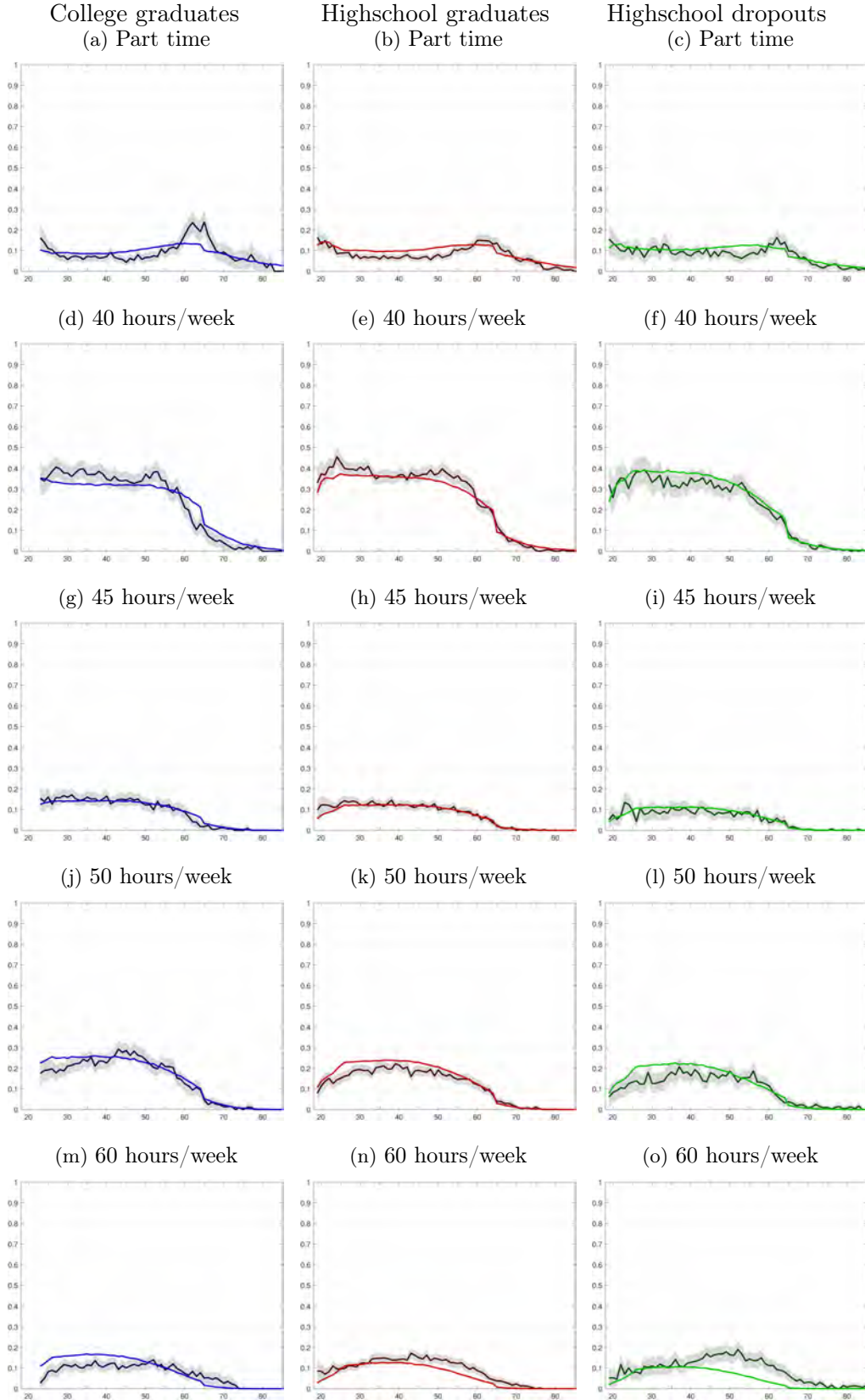
On the other hand, it is striking that in the HILDA data all three education groups experience rapid wage growth at young ages, but this flattens out dramatically after about age 35. Thus it seems that quadratics in age and experience do not capture wage paths in Australia as well as they do in the US or UK. We leave the source of this difference as a subject for future research.

The second row of Figure 6 reports results for the variance of observed wages. Again, the big picture is that the life-cycle variance paths are very different for the three education groups. Variance rises much more sharply with age for the more educated. The model successfully captures that difference, as well as the level differences across groups.

Finally, the third row of Figure 6 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 high school workers in their 20s and 30s.

Figure 7 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 high school 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 show how labor market returns to experience are much greater for more educated workers.

Figure 5: Fitted fractions of the population working particular levels of hours

College graduates	Highschool graduates	Highschool dropouts
(a) Part time	(b) Part time	(c) Part time
(d) 40 hours/week	(e) 40 hours/week	(f) 40 hours/week
(g) 45 hours/week	(h) 45 hours/week	(i) 45 hours/week
(j) 50 hours/week	(k) 50 hours/week	(l) 50 hours/week
(m) 60 hours/week	(n) 60 hours/week	(o) 60 hours/week

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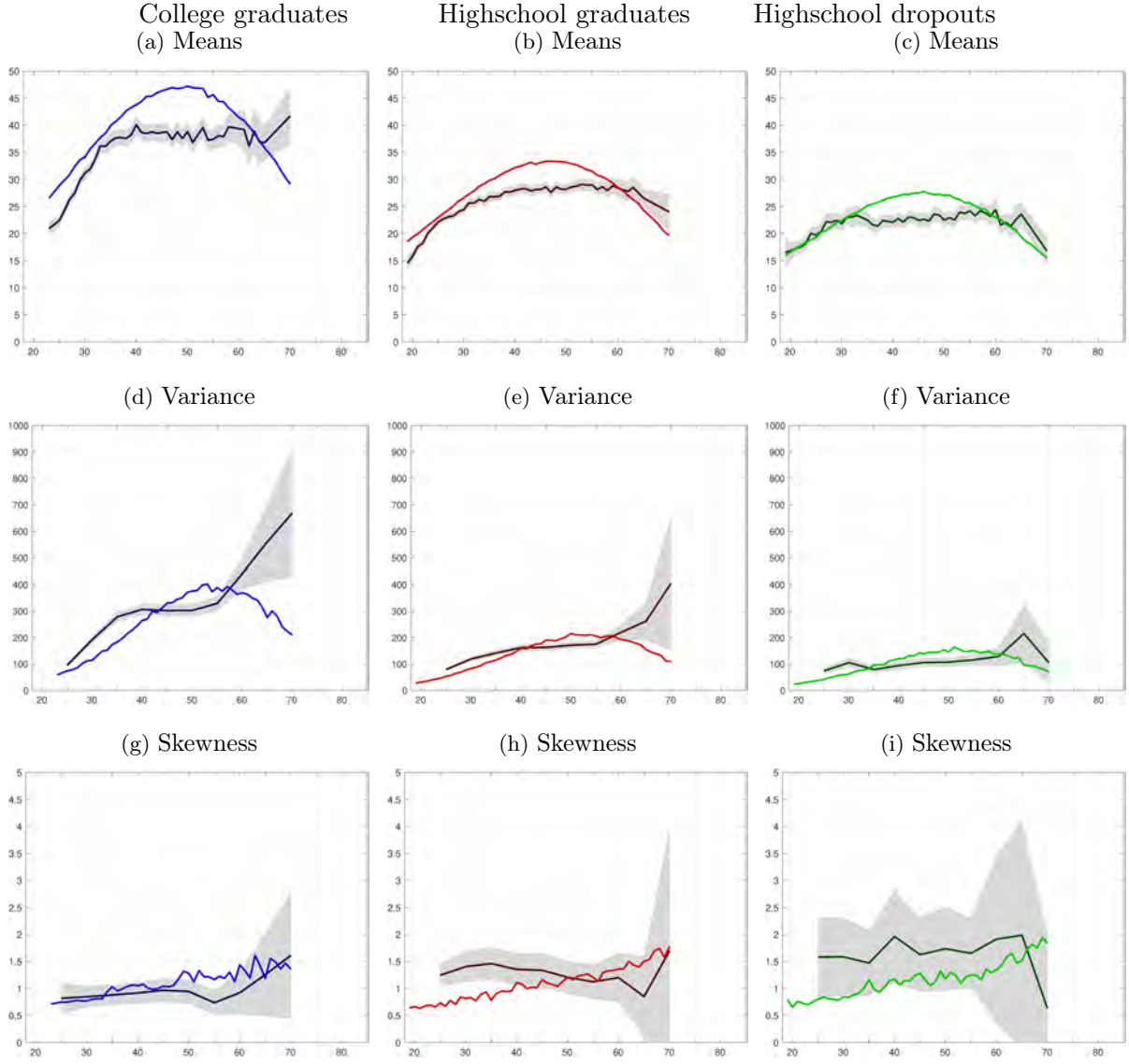
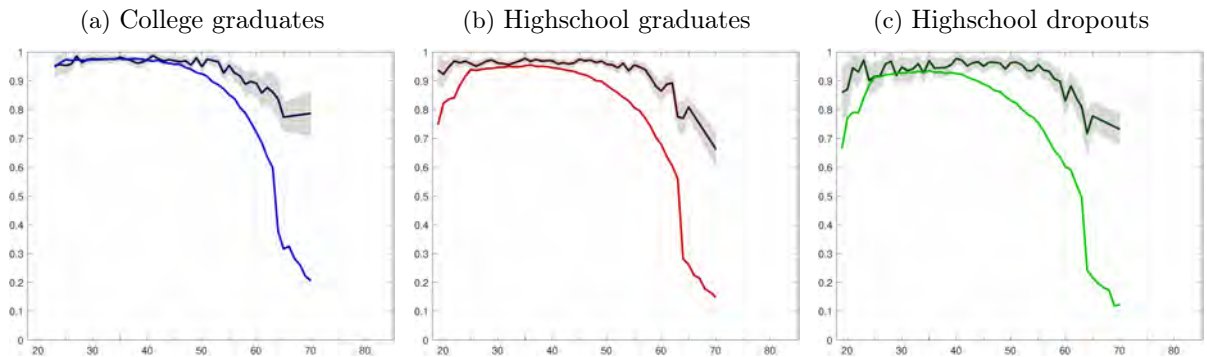
Notes: The plots show simulated life cycle profiles of wages for the three education groups against the profiles observed in the data. Top row shows mean, second row shows variance, and the third row skewness of simulated wages. The profiles observed in the data are shown with the bands of 1.96 standard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

Figure 7: Fitted labor earnings by age/education

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts
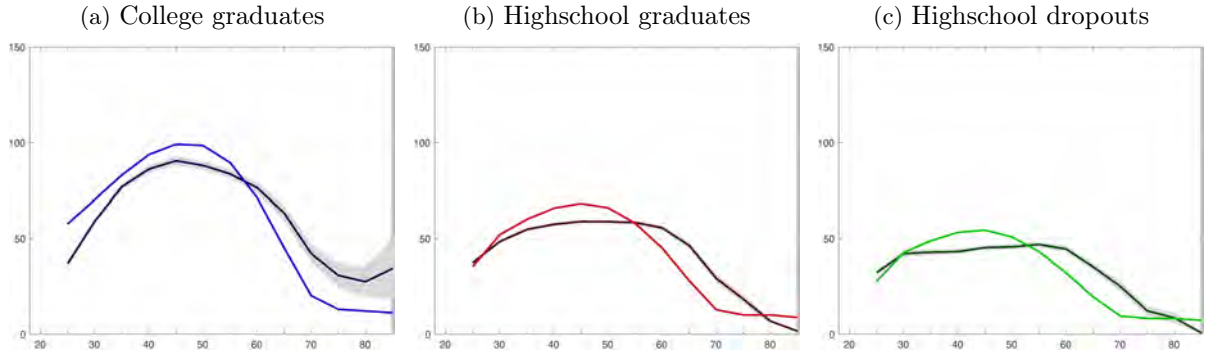
Notes: The plots show simulated life cycle profiles of wage earnings for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 8: Fitted wealth profiles

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts
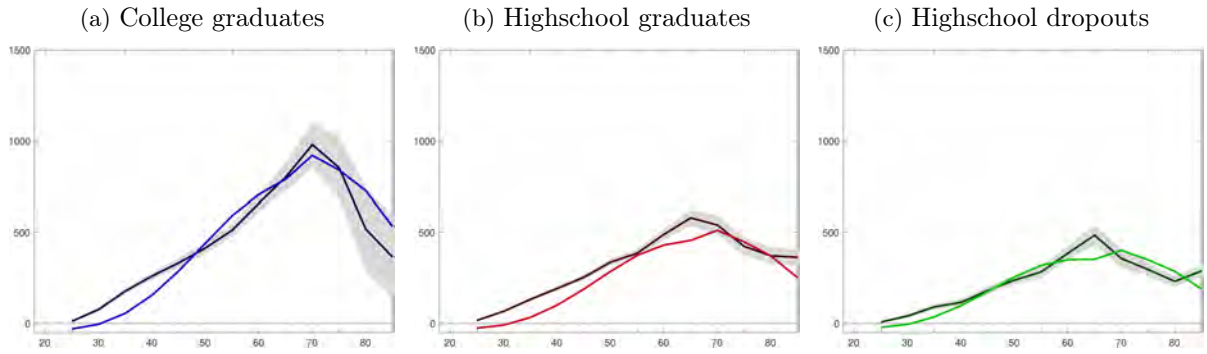
Notes: The plots show simulated life cycle profiles of assets for the three education groups against the profiles observed in the data. The profiles observed in the data are shown with the bands of 1.96 standard deviations around the mean.

Figure 9: Fitted work to work transition rates

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts

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 10: Observed and simulated income distribution

(a) Wage earnings, $1000   (b) Logarithm of wage earnings
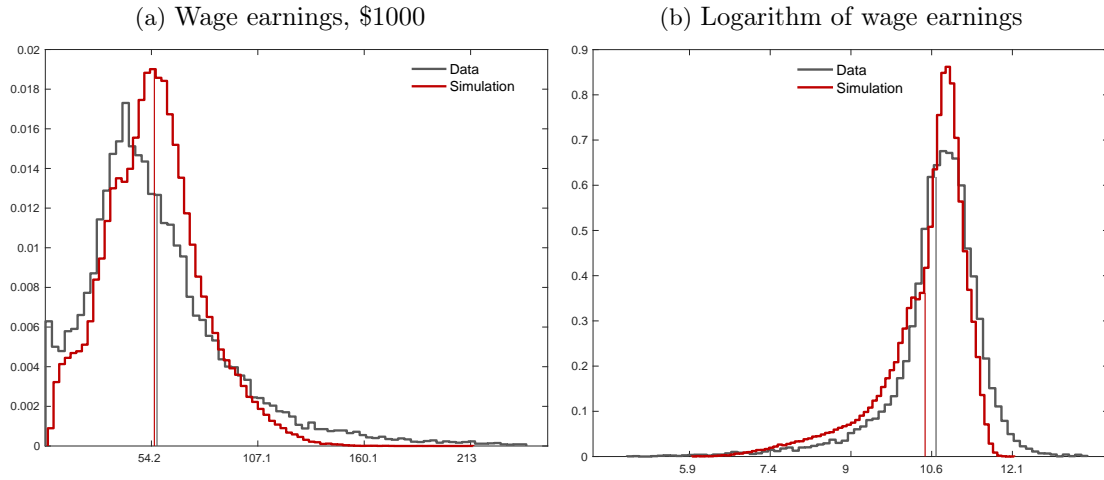
Notes: Histograms for observed vs. simulated labor earnings in levels (panel a) and logs (panel b). The distribution in panel a is truncated at $240,000.

Figure 8 reports on the evolution of wealth over the wealth cycle. The model provides a very good fit to 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, 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 Age Pension on older workers.

Figure 9 describes how the model fits work-to-work transition rates. The model matches the high level of persistence in the data at most ages, but it understates persistence at ages 50+. 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 more 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. 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, the effect of changing Age Pension rules on worker behavior may differ greatly depending on levels of assets and work experience that workers have at age 65, but we doubt the persistence in their employment history is an important factor. Thus, we chose not to complicate the model in an attempt to generate even more persistence at older ages.

Finally, Figure 10 shows the 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 fairly good approximation to the earnings distribution - viewed either way. This is important: In evaluating the impact of a safety net program like the Age Pension, face validity requires a model that generates an earnings distribution similar to that in the data.

Figure 10 reveals that the actual earnings distribution has a complex shape, with fat left and right tails. This is hard to fit with just two types; as we noted in Section 7.2, our estimate of the skill difference between latent types is a compromise that attempts to fill out both tails. In the levels graph we see how the model fails to generate the small mass of people with extremely low earnings (i.e., a couple thousand dollars a year or less). But in the log 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 is the first dynamic labor supply model to account for bunching of hours, so its implications for labor supply elasticities are of interest. Then, in Section 9.2, we turn to our main policy simulations that assess how the Age Pension affects labor supply, savings and human capital over the life cycle.

9.1   Labor Supply Simulations

In Figure 11, 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 11 reports effects of fully anticipated transitory wage changes, so we obtain pure Frisch elasticities. As we see, our model predicts Frisch elasticities increase with age, which is consistent with the pattern predicted by the simpler life-cycle model in Imai and Keane (2004).[30] For instance, the Frisch elasticity is about 0.30 for college types at age 30, but it grows to roughly 1.5 at age 60.

Figure 11 also reveals how the increase in Frisch elasticities with age differs by human capital level, as predicted in Keane (2016) and found empirically in Keane and Wasi (2016). For dropouts, who have relatively less human capital accumulation over the life cycle, the Frisch

---

[30]Theirs' 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 benefits.

Figure 11: Responses of hours to transitory 10% wage reduction

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts
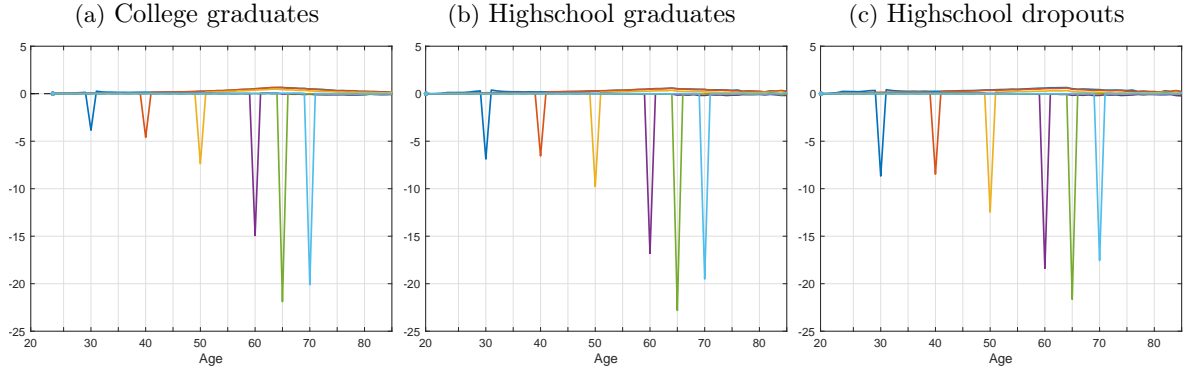
Note: Percent change in work hours induced by anticipated temporary 10% wage reductions at different ages.

Figure 12: Responses of hours to permanent unanticipated 10% wage reductions
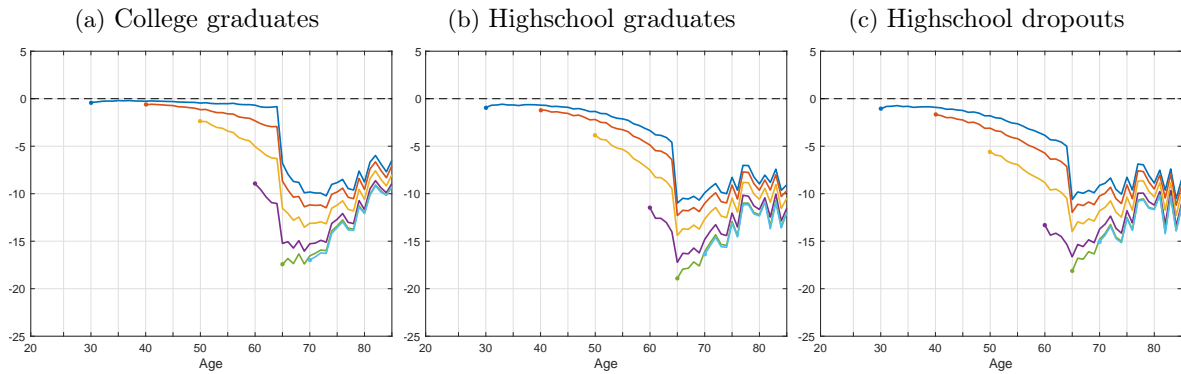
(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts
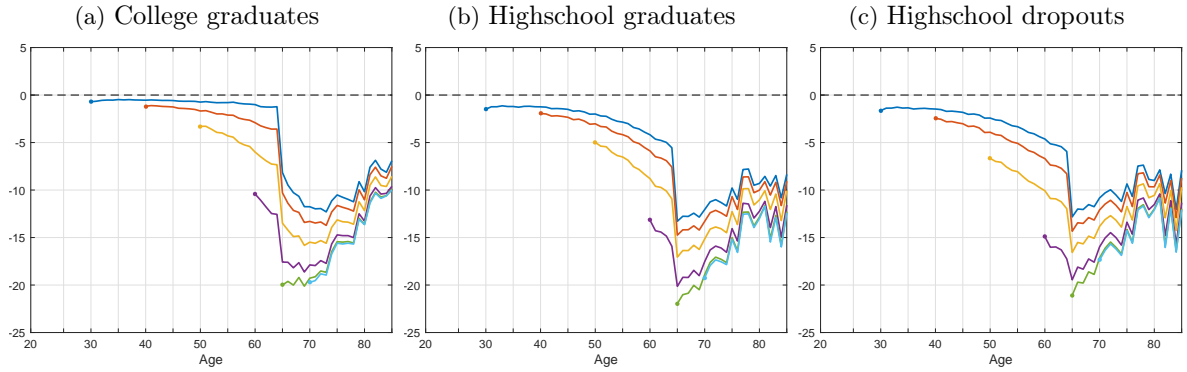
Note: Percent changes in hours induced by permanent unanticipated 10% wage reductions at different ages.

elasticity is about 0.80 at ages 30 and grows to about 1.8 at age 60. This is substantial, but much less than the 5-fold increase we see for college types.

We also decomposed the Frisch elasticity into changes at the extensive (i.e., participation) vs. the intensive margin (i.e., hours conditional on work). The Frisch elasticity for hours (given employment) is roughly 0.15 to 0.25 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 about 0.15 to 0.25 smaller than that for total hours, and it has about the same patterns with education/age discussed earlier. These results are consistent with the empirical work that looked only at employed men and found very small Frisch elasticities (see, e.g., MaCurdy (1981)).

Next we simulate *permanent* 10% wage reductions. These are uncompensated, so we obtain Marshallian elasticities. One can think of these wage reductions as induced by unanticipated permanent 10 percentage point increases in the tax rate on labor earnings, going into effect at the

Figure 13: Response of employment to permanent unanticipated 10% wage reduction

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts

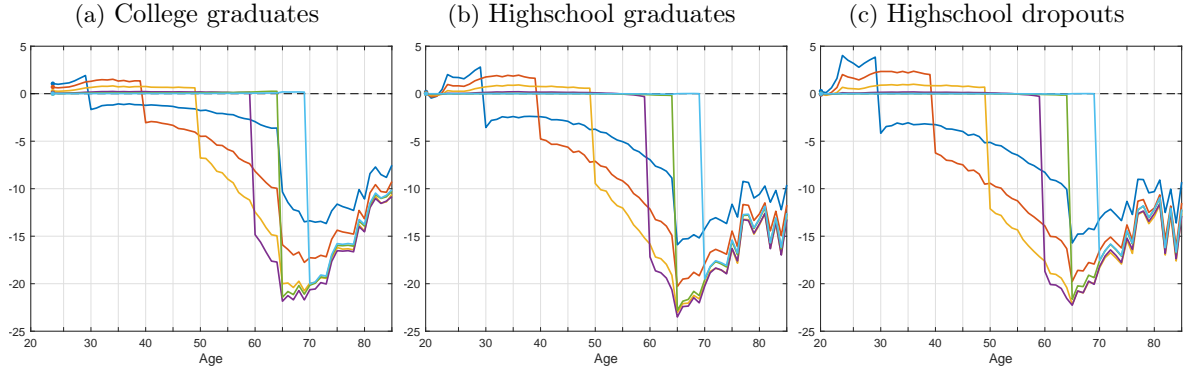Note: Percent change in employment induced by permanent unanticipated 10% wage reduction at different ages.

indicated age, and with the revenue discarded. Figure 12 plots the life-cycle impact of 10% wage reductions that occur at ages 30, 40, 50, 60, 65 or 70 and persist for the rest of the life. For all three education groups the initial impact on labor supply is negative (i.e., the substitution effect dominates the income effect). These initial impacts are greater for older workers. For example, for a high school graduate who experiences a 10% wage reduction at age 40, the decline in hours is roughly 2%, implying a Marshallian elasticity of only 0.20. But if the wage reduction occurs at age 50 the Marshallian elasticity is 0.50, and if it occurs at age 60 it is 1.3.

The changing impact of wage reductions with age is interesting. For example, for a high school graduate who experiences a 10% wage reduction at age 40, the decline in hours is roughly 2% at age 40 and grows slowly to 7% at age 60. But at age 65 the drop in hours is much greater (15%). This is because at age 65 workers become eligible for retirement benefits (both superannuation and the Age Pension). Retirement benefits amplify the negative effect of the 10% wage reduction substantially at age 65, as more workers retire.

Figure 13 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 12), implying again that most of the action is on the participation margin. Second, Figure 13 clarifies the point that most of the large drops in hours at ages 65+ that we saw in Figure 12 are due to additional retirement, rather than reductions in hours.

Finally we consider permanent and fully anticipated 10% wage reductions. These may arise from anticipated permanent tax changes. As we see in Figure 14, anticipation of a wage reduction causes hours to increase up until the age when the tax increase occurs. 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

31

Figure 14: Response of hours to permanent anticipated 10% wage reductions

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts

Notes: Percent change in hours induced by permanent anticipated 10% wage reductions at different ages.

Figure 15: Response of employment to permanent anticipated 10% wage reduction
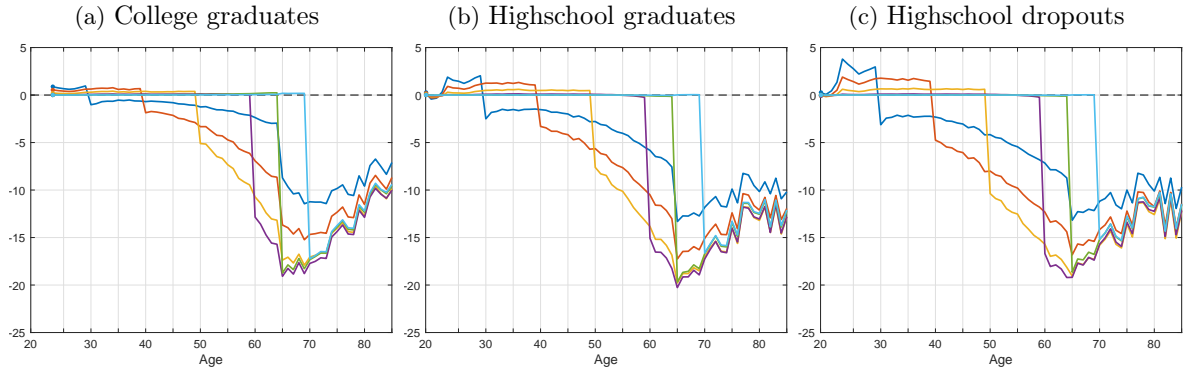
(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts

Notes: Percent change in employment induced by permanent anticipated 10% wage reduction at different ages.

increase goes into effect at age 30, the "immediate impact" at age 30 is for labor supply to drop by 4% for dropouts, 3% for high school graduates, and only 2% for college graduates.

Figure 14 reveals that the immediate impact of anticipated tax increases grows with age. For example, for high school graduates the Marshallian elasticities in the year of a tax increase are 0.30, 0.50, 1.0, and 1.7 and 2.3, if the tax is increased at ages 30, 40, 50, 60 and 65, respectively.

Figure 14 also reveals that the Marshallian elasticity increases with age after a tax change is implemented. It is notable how it jumps up at ages 60 to 65. For example, for a high school graduate, if a 10% permanent tax increase is implemented at age 30, the immediate impact corresponds to an elasticity of 0.30, but this quickly falls to about 0.25. Then it starts gradually increasing, until it reaches 0.7 at age 60. Then at age 65 the Marshallian elasticity jumps to 1.6. This means the tax increase causes workers to reallocate labor from older ages to younger ages (at least in relative terms), as labor supply falls much more at ages 60+.

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 Marshallian elasticity is only about 0.10 at ages 30-45 but shoots up to about 1.0 at age 65 for high school and college types and 0.70 for dropouts. It is comforting that our model generates fairly similar patterns, despite the substantial differences in model specification and national context. This gives us more confidence that this is a robust finding.

Finally, Figure 15 reports the extensive margin responses to a permanent anticipated 10% tax increase implemented at different ages. Figures 14 and 15 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.[31]

9.2 Age Pension Experiments

We now turn to our policy simulations designed to assess how the Age Pension affects labor supply, savings and human capital over the life cycle. We first considered changes in pension rules. For example, focusing on high school types, we predict a 25% increase in the pension grant amount would reduce labor supply of high school type workers by 1% at age 55, and by less at earlier ages. At ages 65+ labor supply drops by about 13%. Taper rates have very little effect on labor supply prior to age 65. But a doubling of the taper rate on assets leads to 17% increase in hours from 65 onward (due to the income effect), and a doubling the taper rate on income reduces hours from 65 onward by at least 12% (due to the substitution effect).

Our model predicts the significant change in the Age 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.

Given that modest program changes had modest simulated effects, we decided to focus our attention on experiments where we completely eliminate the Age Pension. Admittedly this is unrealistic, as 75% of the 65+ population currently relies on the program (and planned for retirement based on its existence). But such a large policy change creates a very clear 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 to keep the policy budget neutral. As we noted in the introduction, the Age Pension is a

---

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

large program, with total costs equal to 1/3 of income tax revenue. Thus, given no behavioral response, income tax rates could be cut by roughly one-third if the Age Pension was eliminated. In fact, as we show below, eliminating the Age Pension has a strong positive labor supply effect. As a result, the income tax rate can be cut by 37% in our budget neutral simulations.[32]

Even our revenue neutral simulations should be viewed as partial equilibrium. Policy experiments that increase labor supply may reduce equilibrium wages, thus dampening the labor supply increase. We do not account for this in our simulations. Eliminating the Age Pension might lead to other important changes in equilibrium as well (e.g., the possible creation of new insurance and annuity products). Nevertheless, our partial equilibrium experiments are useful to gain insights into the mechanisms through which the Age 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 16 to 18 report the effects of eliminating the Age 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, high school = blue, dropouts = red).
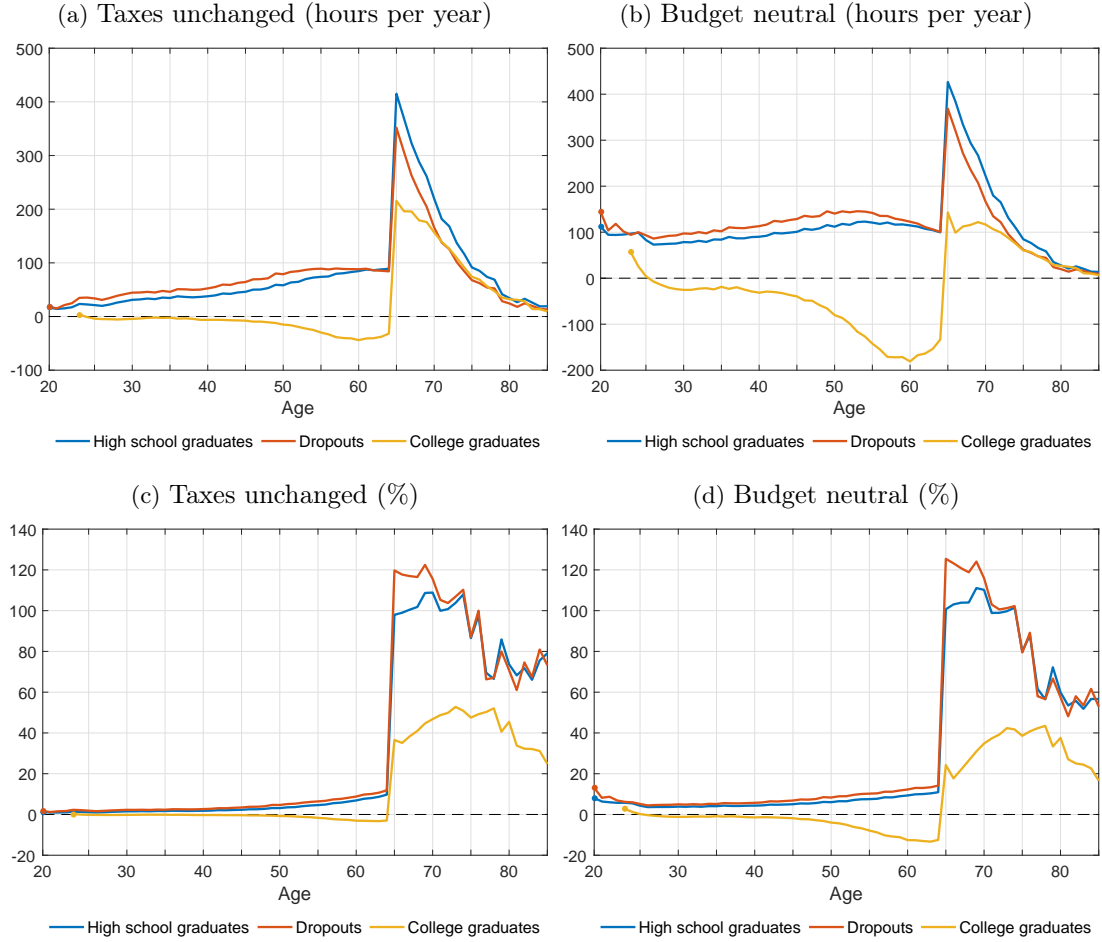
Consider the fixed tax rate simulation first. In Figure 16 we see that labor supply increases prior to age 65 for high school and dropout workers. This is due to the income effect of the elimination of the pension benefit. Forward-looking agents know throughout their working lives that the pension will not be available to them when they retire, so they work more (Figure 16), consume less (Figure 17) and accumulate more assets (Figure 18). By age 64 the increases in labor supply reach about 10%.

In contrast, for college workers there is a smaller income effect from eliminating the Age Pension, because they only expect to receive modest benefits from the program. But college types do work more *after* age 65. This is because the elimination of the Age Pension raises their after-tax wage rate at age 65+, due to the elimination of the earnings test. This pivot in the life-cycle wage profile also induces a small intertemporal substitution effect that causes them to work slightly *less* before age 65.

As we see in Figure 17, all three groups reduce consumption prior to age 65. In levels it is the college group that reduces consumption most, by about $1.5k to $2.5k per year in the 30 to 64 age range, while the dropout and high school graduate groups only decrease consumption by

---

[32]More precisely, the tax rates in both brackets can be cut by 37%, thus still keeping the tax system progressive. Given that required revenue falls by 33% in the budget neutral scenario, the 37% tax reduction is made possible by a 5.8% increase in aggregate labor supply, induced by the combination of eliminating the age pension and the tax rate cut itself (we solve for a fixed point to find the 37% rate reduction).

Figure 16: Response of hours of labor supply to eliminating the Aged Pension

(a) Taxes unchanged (hours per year)

(b) Budget neutral (hours per year)

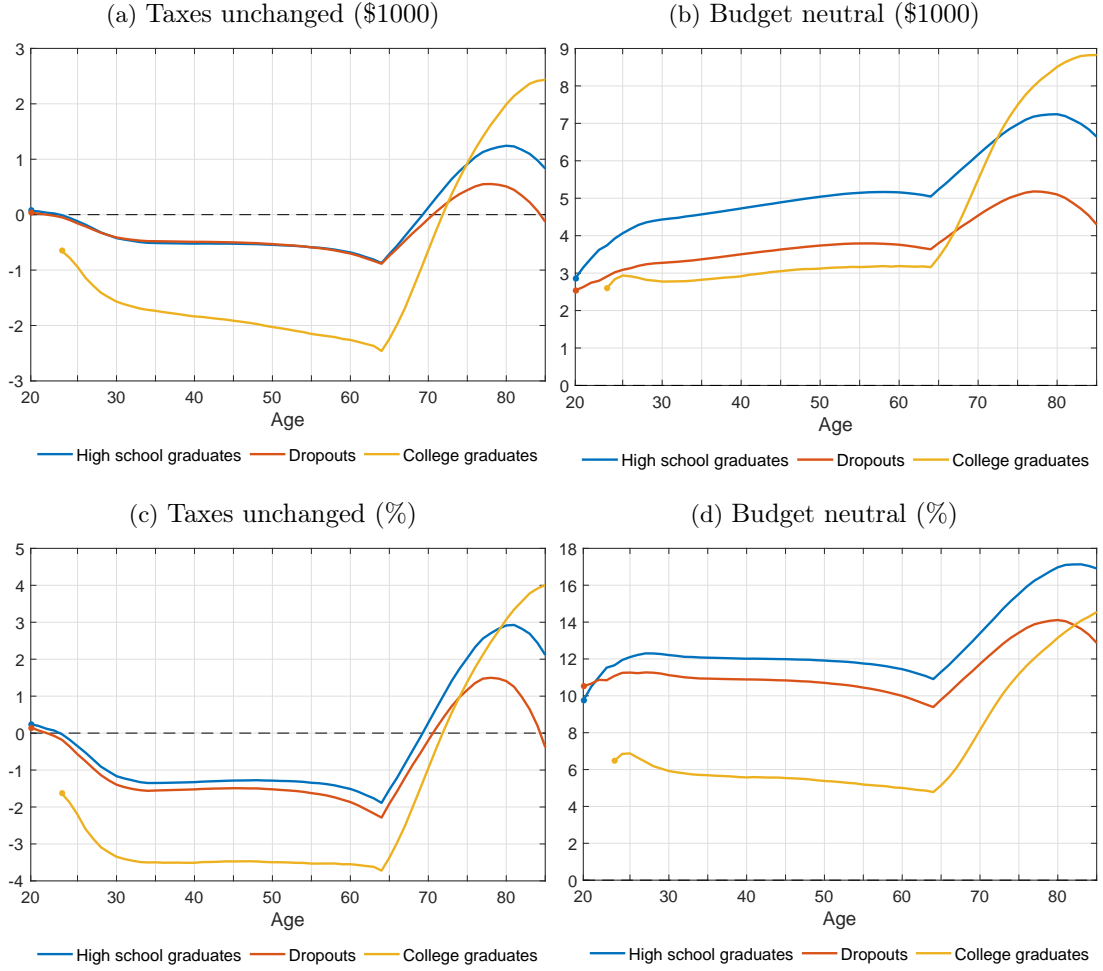(c) Taxes unchanged (%)

(d) Budget neutral (%)

Notes: Plots show simulated changes in hours of work induced by elimination of the Aged Pension, with the same income tax (panels a and c) and with tax rates lowered by 37% to keep government budget unchanged (panels b and d). Panels a and b show absolute changes, while panels c and d show relative (%) changes.

about $500. In percentage terms the college types reduce consumption by about 3.5% per year in the 30 to 65 age range, compared to about 1.5% to 2% for the high school and dropout types.

The net result of these changes is that all three groups accumulate more assets by age 64. As we see in Figure 18, the college graduates accumulate $125k more on average, while the high school graduates and dropouts accumulate about $135k and $145k more, respectively. It is to be expected that the additional wealth accumulation by the college graduates is somewhat less, as they are less likely to utilize the Age Pension in the baseline economy. But in percentage terms (see panel c) the differences between education types is more substantial. The dropouts accumulate 38% more assets by age 64, while the college graduates accumulate 15% more.

Nevertheless, even the effects on the college educated are substantial. There are two main reasons that college workers respond strongly to elimination of the Age Pension. First, the Age

Figure 17: Response of consumption to eliminating the Aged Pension

(a) Taxes unchanged ($1000)

(b) Budget neutral ($1000)

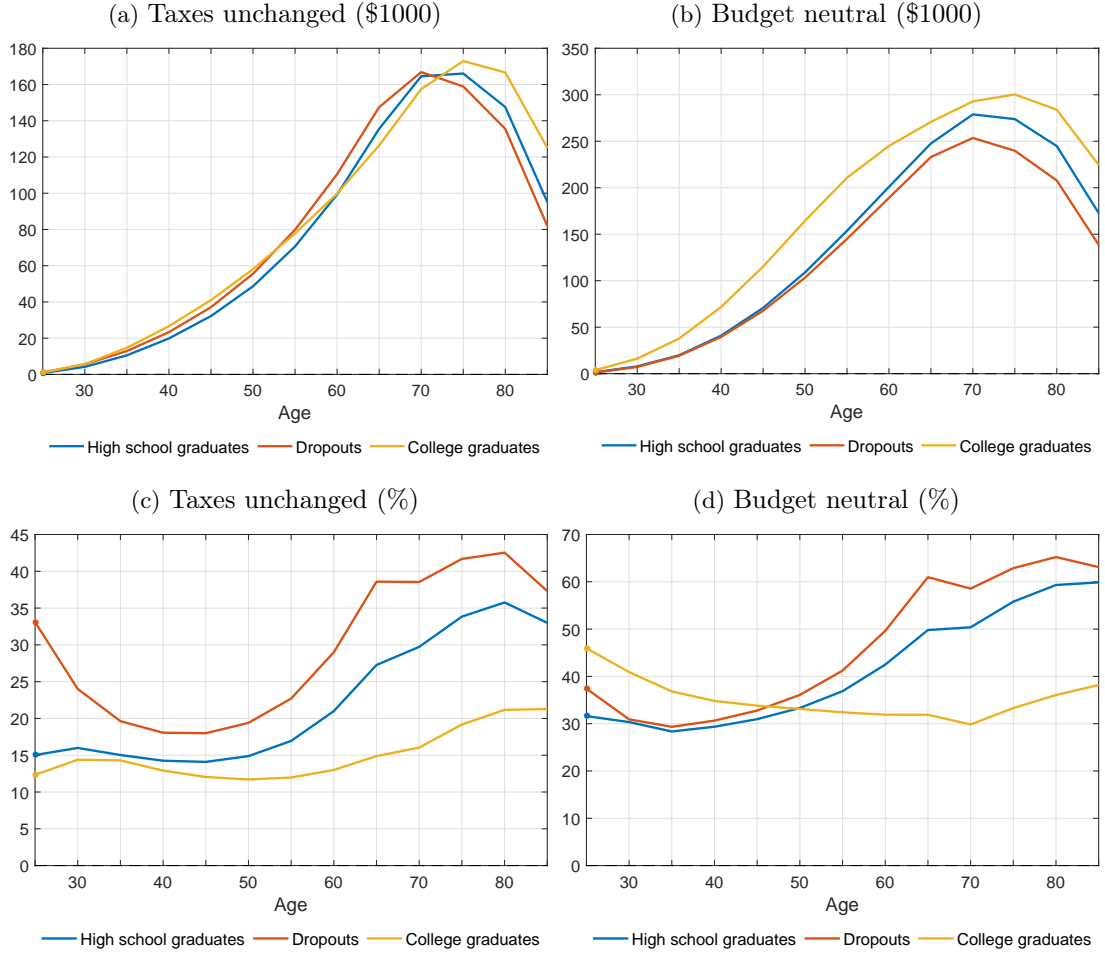
(c) Taxes unchanged (%)

(d) Budget neutral (%)

Notes: Plots show simulated changes in consumption induced by elimination of the Aged Pension, with the same income tax (panels a and c) and with tax rates lowered by 37% to keep government budget unchanged (panels b and d). Panels a and b show absolute changes, while panels c and d show relative (%) changes.

Pension is a consumption insurance program, that mitigates 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 Age 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. In Figure 16, we see that annual hours of work at age 65 increase by roughly 350 hours (120%) for the dropouts, 420 hours (98%) for the high school graduates, and 210 hours (37%) for the college graduates. People also work more at ages beyond 65, but the absolute increment to labor supply diminishes with age. Figure 19 reports 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 16. Thus, the increase

Figure 18: Response of wealth to eliminating the Aged Pension

(a) Taxes unchanged ($1000)

(b) Budget neutral ($1000)

(c) Taxes unchanged (%)
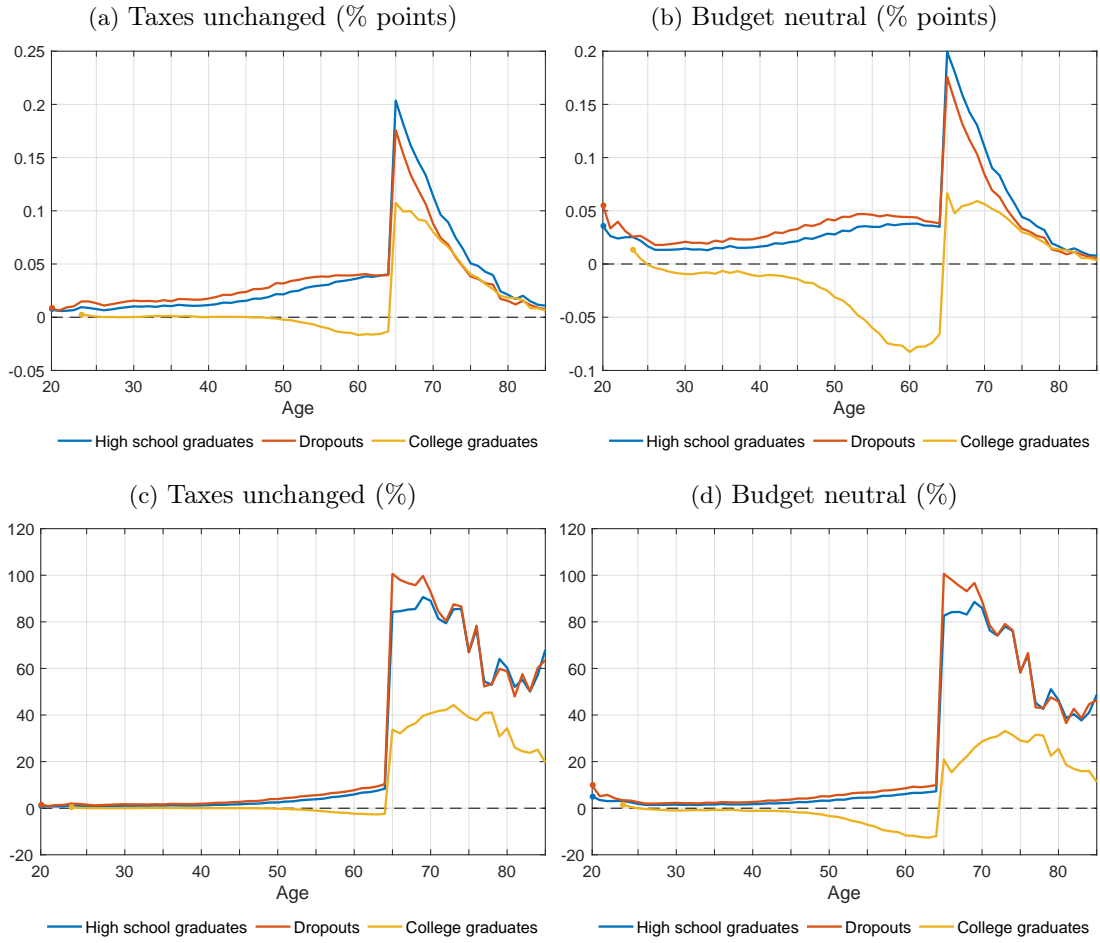
(d) Budget neutral (%)

Notes: Plots show simulated changes in wealth induced by elimination of the Aged Pension, with the same income tax (panels a and c) and with tax rates lowered by 37% to keep government budget unchanged (panels b and d). Panels a and b show absolute changes, while panels c and d show relative (%) changes.

in labor supply from eliminating the Age Pension is mostly due to increased employment, not increased hours of work.[33] Interestingly, this is true both before and after age 65.

Consider next the revenue neutral simulations reported in the right panels of Figures 16-19. As we see in Figure 16, both high school and dropout workers now have labor supply increases of about 5% at younger ages (i.e., 20 to 45) due to the tax reduction. At ages 65+, the change in labor supply of the high school graduate and dropout types is very similar to the no tax change case. Because the tax rate is reduced at the same time the Age 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

---

[33]That is, in Figure 19 panel c we see increases in participation at age 65 of about 30% for high school 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 16 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 19: Response of employment rates to eliminating the Aged Pension

(a) Taxes unchanged (% points)

(b) Budget neutral (% points)

(c) Taxes unchanged (%)

(d) Budget neutral (%)

Notes: Plots show simulated changes in employment rates induced by elimination of the Aged Pension, with the same income tax (panels a and c) and with tax rates lowered by 37% to keep government budget unchanged (panels b and d). Panels a and b show changes in percentage points, while panels c and d show (%) changes.

so the net positive effect of eliminating the Age 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 17, the lower tax rate enables the dropout and high school graduate types to consume about $3k to $5k more per year during their working lives, despite the knowledge that the Age Pension will not be available to them in retirement. And as we see in Figure 18, the lower tax rate enables the dropout and high school 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 Age Pension and higher taxes). In fact, as we see in the right panel of Figure 17, the dropout and high school graduate types are, on average, able to consume significantly more at ages 65+ in the world with no Age Pension but lower taxes.

Figure 20: Distribution of hours responses to eliminating the Aged Pension

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts
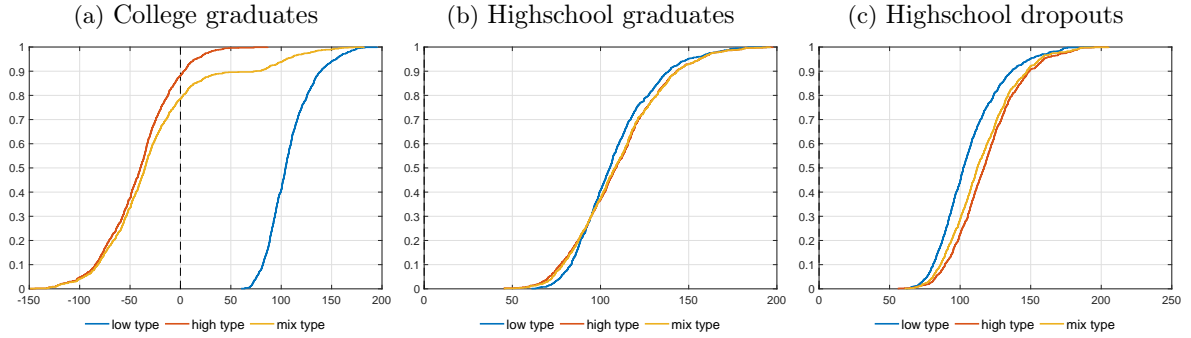
Notes: Plots show cumulative distribution functions for change in hours of work over the life-cycle due to elimination of the Aged Pension, with income tax rates lowered by 37% to keep government budget unchanged. In each panel the data is split by latent type τ_{uh} (high and low), with the overall distribution in yellow.

The results for the college graduates are very different. As we see in the right panel of Figure 16, in a budget neutral simulation of a world with no Age Pension and lower tax rates, college graduates work quite a bit less in the years close to retirement. For example, they work about 80 fewer hours at age 50 (or 4% less), 140 fewer hours at age 55 (8% less), and 175 fewer hours at age 60 (13% less). Two factors explain why the college graduates are so differently affected by the tax cut: First, the college types pay more taxes in the baseline, so the income effect of the tax cut is greater for them. Second, as we saw in Section 9.1, they also experience an inter-temporal substitution effect that induces them to shift labor supply to 65+ periods.

As we see in Figure 17, in the budget neutral simulation college graduates consume about $3k (6%) more per year at ages prior to 65, and at 65+ their consumption grows even more. And in Figure 18 we see that college graduates accumulate (on average) an extra $270k (32%) in assets by age 64 compared to the baseline (and $145k more than in the constant tax rate simulation). The 37% tax cut enables the college type to achieve this higher level of asset accumulation while working a bit less and consuming more.

9.2.2 Heterogeneity and Welfare

Here we analyze heterogeneity in the impact of eliminating the Age Pension. Our model has three sources of ex ante heterogeneity: education level, latent skill/taste for work types, and initial wealth endowments. Within each education level there are "high" and "low" types, where the "high" type has higher skill and lower disutility of work. We call these high/low skill types to be concise. Within 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

Figure 21: Distributions of consumption responses to eliminating the Aged Pension

(a) College graduates   (b) Highschool graduates   (c) Highschool dropouts
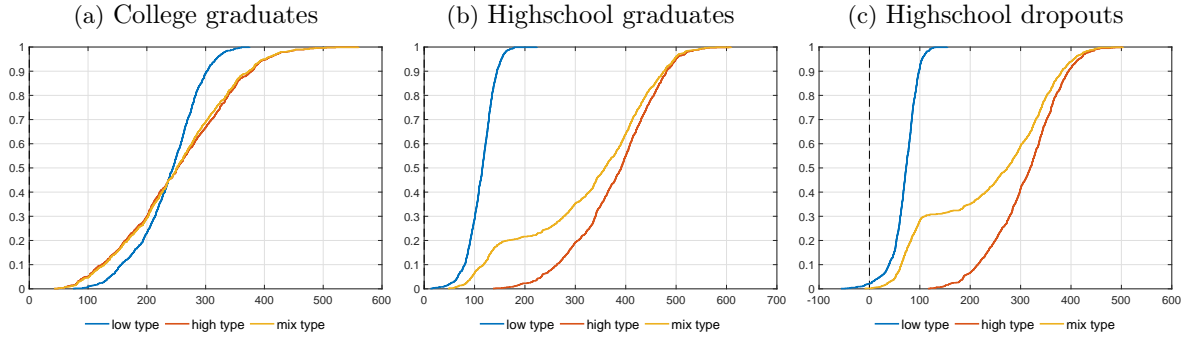
Notes: Plots show cumulative distribution functions for change in the non-discounted sum of lifetime consumption (in $1000s) induced by elimination of the Aged Pension, with tax rates lowered by 37% to keep government budget unchanged. In each panel results are split by type τ_{uh} (high and low), with the overall distribution in yellow.

otherwise identical agents to make different choices and have different earnings ("luck").

Figure 20 reports education/skill type-specific cumulative distribution functions (CDFs) for changes in mean annual hours of work (over the entire life). We focus on the budget neutral scenario. 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 proportions which are education specific in our model.

For the high school and dropout types the entire CDF lies to the right of zero. So every member of these groups works higher average hours as a result of eliminating the Age Pension and reducing taxes. Indeed, nearly everyone increases hours by at least 60 per year, and there is non-negligible mass up to about 160 hours per year. The median individual increases hours by about 110 per year. The differences between the high and low skill type are not large.

The CDFs for college graduates look very different. In Figure 20 we see that for the high type 90% have a decrease in hours, and the median decrease is about 40 hours per year. But for the low skill type all individuals have an increase in hours, and the median increase is about 105 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 homogeneous across the high/low types. However, as Keane and Rogerson (2015) and Keane and Wasi (2016) discuss, in more complex models with features like human capital, an extensive margin and non-linear budget sets, the situation is much more complex. Labor supply elasticities become functions of the entire economic environment.

Figure 22: Distributions of ex-ante welfare effects of eliminating the Aged Pension

(a) College graduates (b) Highschool graduates (c) Highschool dropouts
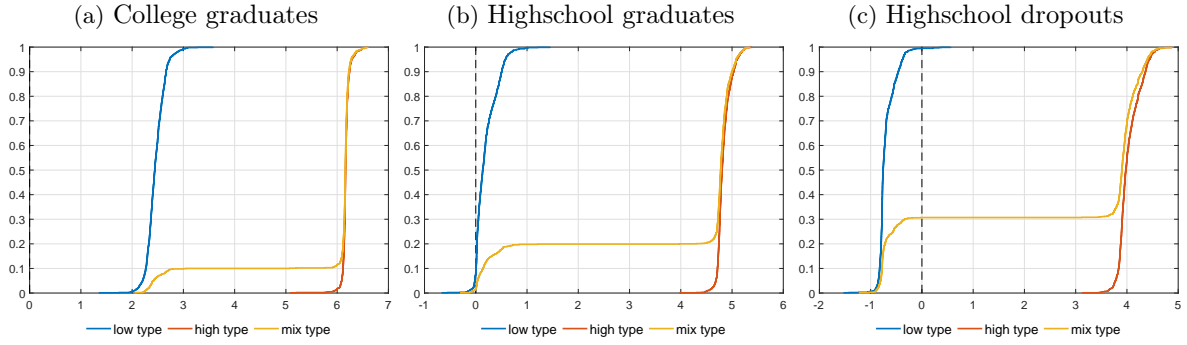
Notes: Plots show the cumulative distribution functions for changes in expected discounted lifetime utility at age 19 (23 for college graduates) induced by elimination of the Aged Pension, with income tax rates lowered by 37% to keep government budget unchanged. In each panel results are split by type τ_{uh} (high and low), with the overall distribution shown in yellow.

The point that preference parameters are "insufficient" to characterize labor supply responses is well illustrated here. The reactions of high and low skill college types to elimination of the Age Pension differ not only in magnitude but even in sign (despite having identical curvature parameters on consumption/leisure). In fact, the low-skill college types behave quite similarly to high school types. There are two reasons for this: First, the low skill types are much more likely to use the Age 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.

Figure 21 reports CDFs of the change in cumulative consumption over the life cycle (not discounted). Almost all workers consume more when the Age Pension is eliminated (and taxes are reduced). The median increase is about $250k for college graduates, $360k for high school graduates, and $280k for dropouts. The only group with a drop in consumption is about 4% of the low skill drop out type. But this represents a very small fraction of the overall population (as only about 25% are dropouts, and only 31% of dropouts are the low type).

Figure 22 reports the distribution of changes in expected present value of lifetime utility given elimination of the Age Pension (and reduced taxes). Every member of the college group is ex ante better off.[34] Note that, in these graphs, as we are taking expectations in the initial period, so "luck" plays 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. Thus, among the

---

[34]The median increase in expected present value of lifetime utility at age 23 is about 6.2 for the high type and 2.4 for the low type.

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 too surprising that the college types are better off with less social insurance and lower taxes. What is surprising is that the story is little different for high school graduates. The high types are all better off, while the large majority of the low type are better off (although most are close to indifferent).[35] The story is more mixed for dropouts. The high type are all better off, but the low type are almost all worse off.[36] However, the low type dropouts are only $(.31)(.25) = 8\%$ of the population. It would only take modest lump sum transfers from gainers to losers to also make them better off, and render the policy Pareto improving.

Finally, we translate the utility changes into consumption equivalent values. The mean increase in the present value of lifetime utility evaluated in the first period is 2.48 for dropouts, 3.82 for high school graduates and 5.69 for college graduates. These translate into consumption equivalent values (CEVs) of $7.4k, $16.8k and $43.5k, respectively.[37]

9.2.3 A Better Targeted Age Pension

In order to improve the targeting of the Age Pension, we next consider an experiment where we double the taper rates on both income and assets. In equation (26), the income tamper rate is increased from 27.7% to 55.5%, while that on assets is increased from 0.005 to 0.10. The effects on labor supply are shown in Figure 23, left panels.

The increase in taper rates has very little effect on labor supply of any education group prior to age 65. But at age 65 the labor supply of college graduates increases by 20%, while the labor supply of dropouts falls by 8%. This implies the college types rely less on the Age Pension while the dropouts rely more - which in turn suggests the program is better targeted.

When we calculate welfare changes we find that, not surprisingly, all six education/skill types are *ex ante* worse off as a result of higher taper rates. However, we observed that the welfare losses are fairly small, and that the increase in taper rates raises some revenue. This raised the possibility that if we factored in a small tax cut we might achieve a Pareto improvement.

Thus, we ran a budget neutral simulation. We found that in a world with higher taper rates it is possible to cut income tax rates by 5.9%.[38] In this world, we simulate that all workers are *ex ante* better off. The mean CEVs for low-skill types are $1.4k for dropouts, $1.5k for high school graduates and $1.7k for college graduates. The gains for the high types are $1.7k, $1.9k

---

[35] For high types the median utility gain is about 4.8 but for low type it is only 0.1.

[36] For high types the median utility gain is about 4.0 and for the low type the median utility loss is about 0.8.

[37] For the low types, the CEVs are -$500 for dropouts, $1.3k for high school graduates and $7.2k for college graduates. The gains for the high types are much greater (i.e., $17.1k, $27.0k and $51.3k, respectively).

[38] This means that in equation (27) the top rate is reduced from 37.9% to 35.7% and the middle bracket rate is reduced from 29.9% to 28.1%.

Figure 23: Response of labor supply to doubling the Age Pension income and asset taper rates

(a) Taxes unchanged (hours per year)

(b) Budget neutral (hours per year)

(c) Taxes unchanged (%)
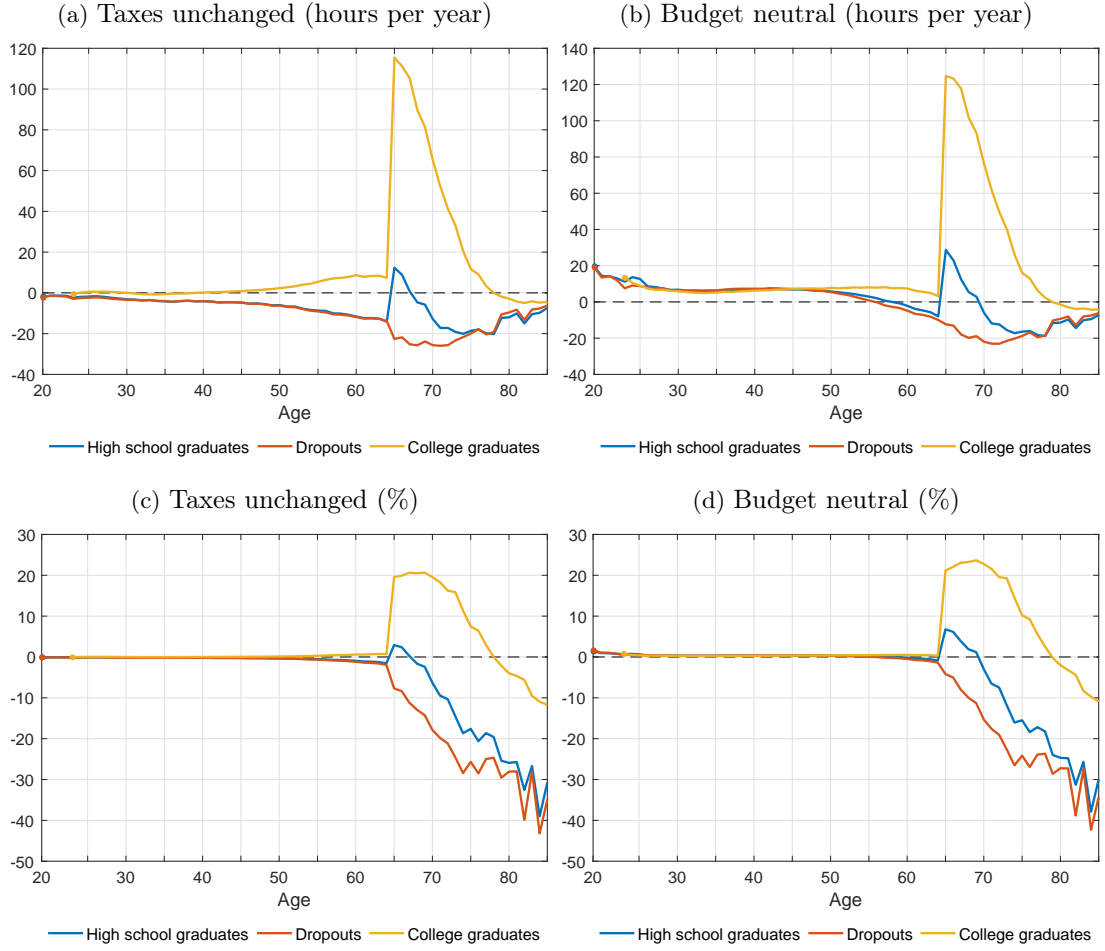
(d) Budget neutral (%)

Notes: Plots show simulated changes in hours of work induced by doubling the income and asset test taper rates. Income tax rates are held fixed in panels a and c. Tax rates are lowered by 5.9% in panels b and d to maintain budget neutrality. Panels a and b show absolute changes, while panels c and d show relative (%) changes.

and $2.2k, respectively. Even within types, the entire distribution of gains is positive. Thus, this policy change is Pareto improving. Interestingly, NCOA (2014) has recommended that 50% higher taper rates be phased in by 2028. Our simulations support this recommendation, provided it is paired with income tax redictions.[39, 40]

Finally, the labor supply effects of the budget neutral policy are shown in the right panels of Figure 23. The policy leads to a small increase in labor supply for all education groups prior

---

[39]The NCOA argue that Age Pension rule changes should be grandfathered in, so that workers born prior to 1960 are not affected, as they would have made retirement plans under current rules.

[40]In related work, Tran and Woodland (2014) analyzed optimal design of the Age 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 and face liquidity constraints. To make OLG modelling feasible they abstract from human capital investment and specify a simpler model of individual decision making. They find, as we do, that elimination of the Age Pension, along with the tax reductions this permits, would increase welfare. However, in contrast to our results, they find the optimal taper rate on earnings is roughly 30% to 40%, which is similar to the existing rate.

to age 65, and a strong increase in labor supply of college workers at ages 65+.

10   Conclusion

In this paper we have developed and estimated a life-cycle model of labor supply and consumption behavior and used it to evaluate the Age Pension program in Australia. Our results imply that *ex ante* 92% of agents would prefer to live in an economy without the Age Pension, with income tax rates reduced by 37% to maintain budget neutrality. The exception is the lower skilled 31% of high school dropouts, who make up roughly 8% of the population.

Three key factors drive this result: First, the Age 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. Third, while the Age Pension is meant as a social insurance program against the risk of low consumption in old age, it is poorly targeted. Although the program is means tested, taper rates are very low, so that benefits extend well up into the income/wealth distribution.

Our policy simulations imply that an improved version of the Age Pension, with a doubling of income and asset taper rates used to finance a 5.9% reduction in income tax rates, would generate a Pareto improvement. Notably, the NCOA (2014) has proposed that substantial increases in taper rates be phased in by 2028. Our results are strongly supportive of this policy recommendation, so long as it is paired with income tax rate reductions.

Our results should be viewed in historical context: The superannuation guarantee was only created in 1992 and contributions were not ramped up to 8% of earnings until 2000. Hence, current retirees were not able to contribute to super for their whole working life. Thus, the current Age Pension program needs to be rather generous because we are still in a transition period where super balances have not reached their full level. However, starting in roughly 2030-2035 we can expect that new retirees will have made a full 40 to 45 years of super contributions. Thus, our policy advice is that Age Pension taper rates should be gradually increased over the next 15 years until the program is better targeted as a safety net for the low income population.

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, and the bunching of hours of work at several discrete levels. Of course we also include the superannuation and Age Pension rules, as well as retirement decisions. There is 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 on 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. And we propose a new method of modelling the human capital state variable that has important computational advantages.

An obvious extension of our model would be 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 Security 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 Age 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. One might also examine the transition path to full phase in of superannuation. Of course, embedding dynamic structural models 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 Age 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 Age Pension.

References

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 Government Actuary. Australian life tables. Technical report, Department of Treasury, Canberra, ACT, 2009.

Australian Prudential Regulation Authority. Insight, Issue one, 2014. URL http://www.apra.gov.au/Insight/Documents/Insight_Issue1_2014_FINAL.pdf.

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.

Andrew Clausen and Carlo Strub. A General and Intuitive Envelope Theorem. ESE Discussion Papers 274, Edinburgh School of Economics, University of Edinburgh, April 2016. URL https://ideas.repec.org/p/edn/esedps/274.html.

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 multidimensional 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.

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. Multidimensional endogenous gridpoint method: Solving triangular dynamic stochastic optimization problems without root-finding operations. *Economics Letters*, 135:72–76, October 2015. doi: 10.1016/j.econlet.2015.07.033. URL `http://dx.doi.org/10.1016/j.econlet.2015.07.033`.

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. 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 effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4):1051–1103, 2001.

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 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`.

NCOA. The Report of the National Commission of Audit. Technical report, National Commission of Audit, Canberra, ACT, February 2014.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, pages 999–1033, 1987.

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.

Steven M. Wild. Solving derivative free nonlinear least squares problems with 'pounders'. Technical Report ANL/MCS-P5120-0414, Mathematics and Computer Science Division, Argonne National Laboratory, 2015. https://www.mcs.anl.gov/papers/P5120-0414.pdf.

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: 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 $\mathbf{1}\{v_1 \geq v_2\}$ as $\phi \to 0$
3. $\mathcal{S}(v_1, v_2, \phi)$ is invariant to scale of the values, i.e. $\mathcal{S}(\alpha v, \alpha v_2, \phi) = \mathcal{S}(v_1, v_2, \phi)$, $\alpha > 0$.

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

$$\mathcal{S}^{(1)}(v_1, v_2, \phi) = \frac{\exp(\frac{v_1}{\phi v_m})}{\exp(\frac{v_1}{\phi v_m}) + \exp(\frac{v_2}{\phi v_m})} = \frac{1}{1 + \exp(\frac{v_2 - v_1}{\phi v_m})}. \tag{29}$$

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

$$\mathcal{S}^{(2)}(v_1, v_2, \phi) = \frac{\exp(\frac{v_1}{\phi|v_m|})}{\exp(\frac{v_1}{\phi|v_m|}) + \exp(\frac{v_2}{\phi|v_m|})} = \frac{1}{1 + \exp(\frac{v_2 - v_1}{\phi|v_m|})} = \frac{1}{1 + \exp(\text{sign}(v_m)\frac{v_2 - v_1}{\phi v_m})}, \tag{30}$$

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 $(v_1, \ldots, v_5)$ denote the values of 5 choices of positive hours, and v_0 denote the value of not working. Further, let $v_{m5} = \max\{v_1, \ldots, v_5\}$ and $v_m = \max\{v_0, v_{m5}\}$. Then

$$\mathcal{S}_0(v_0, \ldots, v_5, \phi_1, \phi_2) = \frac{\exp(\frac{v_0}{\phi_1|v_m|})}{\exp(\frac{v_0}{\phi_1|v_m|}) + \exp(\frac{v_{m5}}{\phi_1|v_m|})} = \frac{1}{1 + \exp(\frac{v_{m5} - v_0}{\phi_1|v_m|})}, \tag{31}$$

$$\mathcal{S}_i(v_0, \ldots, v_5, \phi_1, \phi_2) = \frac{\exp(\frac{v_{m5}}{\phi_1|v_m|})}{\exp(\frac{v_0}{\phi_1|v_m|}) + \exp(\frac{v_{m5}}{\phi_1|v_m|})} \cdot \frac{\exp(\frac{v_i}{\phi_2|v_{m5}|})}{\sum_{j=1}^{5} \exp(\frac{v_j}{\phi_2|v_{m5}|})}, \quad i = 1, \ldots, 5, \tag{32}$$

where $\mathcal{S}_0$ is the *smoothed indicator* for not working and $\mathcal{S}_i$, $i = 1, \ldots, 5$ 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) $\mathcal{S}_i(v_0, \ldots v_5, \phi_1, \phi_2)$ is a smooth function of $v_0, \ldots, v_5$ as long as $v_{m5} \neq 0$ and $v_m \neq 0$, (2) $\mathcal{S}_i(v_0, \ldots v_5, \phi_1, \phi_2)$ converges to the appropriate indicators when $\phi_1 \to 0$ and $\phi_2 \to 0$, and (3) $\mathcal{S}_i(v_0, \ldots v_5, \phi_1, \phi_2)$ is invariant to scale of $v_0, \ldots, v_5$.