DISCUSSION PAPER SERIES

DP15719

Identification of Dynamic Discrete-Continuous Choice Models, with an Application to Consumption-Savings-Retirement

Pasquale Schiraldi and Matthew Levy

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

Identification of Dynamic Discrete-Continuous Choice Models, with an Application to Consumption-Savings-Retirement

Pasquale Schiraldi and Matthew Levy

Discussion Paper DP15719
Published 25 January 2021
Submitted 15 December 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Identification of Dynamic Discrete-Continuous Choice Models, with an Application to Consumption-Savings-Retirement

Abstract

This paper studies the non-parametric identification of the discount factor and utility function in the class of dynamic discrete-continuous choice (DDCC) models. In contrast to the discrete-only model we show the discount factor is identified. Our results further highlight why Euler equation estimation approaches that ignore agents' discrete choices are inconsistent. We estimate utility and discount factors for a consumption- savings-retirement choice problem using the Panel Study of Income Dynamics (PSID). We show that the relative risk aversion parameter and the intertemporal elasticity of substitution are separately identified, and that the latter varies across agents due to the wealth-dependence of the surplus from the discrete choice. This surplus also implies that the value function may be locally convex in wealth, and we find that a simulated Universal Basic Income (UBI) policy counterintuitively benefits wealthier working households more than poorer ones due to this effect.

JEL Classification: N/A

Keywords: N/A

Pasquale Schiraldi - p.schiraldi@lse.ac.uk
LSE and CEPR

Matthew Levy - m.r.levy@lse.ac.uk
LSE

Identification of Dynamic Discrete-Continuous Choice Models, with an Application to Consumption-Savings-Retirement

Matthew R. Levy* and Pasquale Schiraldi†

January 2021

Abstract

This paper studies the non-parametric identification of the dynamic discrete-continuous choice models. In contrast to the discrete-only model we show the discount factor is identified. Our results further highlight why Euler equation estimation approaches that ignore agents discrete choices are inconsistent. We estimate utility and discount factors for a consumption-savings-retirement choice problem using the PSID. We show that the relative risk aversion parameter and the intertemporal elasticity of substitution are separately identified, and that the latter varies across agents due to the presence of the discrete choice. The value function in this setup may be locally convex in wealth, and we find that a simulated Universal Basic Income policy counterintuitively benefits wealthier working households due to this effect.

*London School of Economics, m.r.levy@lse.ac.uk

†London School of Economics and CEPR, p.schiraldi@lse.ac.uk

1 Introduction

It is well known that the optimal decision rule of an economic agent who chooses a continuous quantity of a good so as to maximize their inter-temporal utility can be characterized by the first-order condition Euler equation. This approach has been used in numerous literatures such as consumption and saving, risk-sharing, asset pricing, labor supply, investment, and more. Importantly, however, the assumptions that justify these models include agents being able to freely and continuously adjust the optimal consumption of the goods in any given time. In reality, agents make complex decisions which involve not only continuous but also discrete decisions: either because of the nature of the goods, notably durables; or because consumption decisions require adjustment or transaction costs and therefore are infrequently and discretely adjusted. In general, the inter-temporal decision problem of an economic agent requires a combination of (possibly multiple) discrete and continuous decisions: consumers choose which retail stores to visit and how much to buy, which investment products to buy and how much to invest, which tasks to perform and how much effort to exert, which retirement products to chose and how much to consume, and so on.[1] As noted by Chetty and Szeidl (2007), "the canonical expected utility model of risk preferences...assumes that agents consume a single composite commodity... When some goods cannot be costlessly adjusted, a composite commodity does not exist, and the standard expected utility model cannot be applied".

In this paper, we study the non-parametric identification of the utility function as well as the discount factor of dynamic discrete-continuous choice (DDCC) models, and apply the results to estimate these preferences in the context of a cosumption-saving-retirement choice problem. On the one hand, adding the continuous choice to a discrete-only problem allows us to achieve identification of the discount factor which would not typically be identified otherwise. On the other hand, adding the discrete choice to a continuous-only model allows us to handle the frictions that would otherwise lead to an inconsistent estimate. On this subject we make three main contributions: first, we establish a new set of identification results of the utility function and the discount factor in this class of model; second, we highlight the source of inconsistency that characterizes the standard model of intertemporal decision based on the simple Euler equation; and third, we estimate a consumption-saving-retirement model using a

[1] For example, the optimal consumption and investment with multiple risky assets likely involves fixed/variable and observed/unobserved transaction costs as well as individual observed and unobserved heterogeneity. This problem has proven quite difficult to analyze even with parametric restrictions on the form of the utility function (see for example Liu, 2004; Lynch and Tan, 2011; Gârleanu and Pederson, 2016).

simple two-step method approach which can be generally used to estimate this class of model. We show that in this application, failing to account for the discrete choice yields a substantial bias in the key estimates.

Specifically, we apply our non-parametric identification results to estimate a lifecycle problem with continuous consumption and binary retirement decisions using the Panel Survey of Income Dynamics (PSID). A naive Euler equation approach which ignores the discrete choice yields an estimated coefficient of relative risk aversion equal to 1.71 for working households, which is thirty-eight percent larger than the estimate of 1.24 generated by the full model. This is because the naive Euler equation approach mis-attributes the option value of the future discrete choice — i.e., being able to retire in the next period — as risk-aversion in the instantaneous utility function. Such a degree of bias has important implications for welfare and policy. Equally important, in the full model we estimate an annual discount factor of 0.96 on average while the naive Euler equation approach yields a much lower value of 0.94, or roughly twice the discount rate. Moreover, we find that the discrete choice is directly important to understanding welfare, as both the level and the degree of risk-aversion are typically much higher under retirement than when working — and increasingly so with age.

We use the model to estimate a counterfactual policy exercise of a Universal Basic Income (UBI), in which a \$750/month cash transfer is financed by an across-the-board 15 percentage point increase in income taxes. Such a policy has clear implications for both the continuous consumption choice as well as the discrete labor force participation choice. Having identified not only the utility functions of working and retired households, but also the discount factor, we are able to compute the change in lifetime expected discounted utility from this policy. The effect is strongly progressive among retired households, benefitting the poorest households the most. However, we find that among working households the benefits are actually higher for wealthier households. Working households retire on average 1.2 years earlier under the policy reform, but this is concentrated among wealthier households. They therefore benefit from a greater option value from the discrete choice than do poorer households, which tend only to benefit from the increased consumption conditional on remaining working. The presence of the discrete choice thus overturns the welfare conclusions which would arise in a continuous-only model.

Models of economic agents' dynamic optimization problems based on the estimation of Euler Equations using micro data are workhorses of modern macroeconomics, public finance, consumer finance, and many other fields. However, this literature (see for example Alan et al.

(2019), Attanasio et al. (1999), Bond and Meghir (1994), Mulligan (2004) among others) has mostly ignored the presence of discrete goods (such as durables) or costly consumption adjustment. When costly adjustment or a discrete choice affects agents utility the Euler equation can and must be augmented by two additional considerations: first, the alternative chosen tomorrow may have a marginal utility of consumption different from the current alternative; and second, the continuous choice may affect the probability with which future alternatives are chosen. Both channels influence the optimal continuous choice.

To understand the first channel, consider a simple setting with two alternatives, one with a relatively high marginal utility of consumption and the other relatively low. Taking for now the discrete choice probabilities as exogenously given, an agent who has chosen the high marginal utility good today knows that with some chance she will choose the low marginal utility good tomorrow. The expected marginal utility of consumption tomorrow is therefore lower than would result from considering only the chosen alternative, and the agent should optimally consume relatively more today. Ignoring the possibility of alternatives with differing marginal utilities would be similar to ignoring the presence of states that affect the marginal utility, and the model would be misspecified in both cases.

In general, however, the discrete choice probabilities will themselves be affected by the continuous state, and this introduces a second effect. Even if the marginal utilities from the continuous choice are identical across all alternatives, the agent must take into account the expected surplus from the next-period discrete choice when deciding her continuous choice. The agent's marginal utility from consumption may be unaffected by her discrete choice, but both her level of utility and the possible transition of her wealth may be. As changes in the agent's wealth may cause her to switch from one discrete choice to another, her value function may in fact be locally convex in wealth. Consumption today therefore affects the expected surplus from the discrete choice tomorrow, and the agent will therefore consume differently than if the choice probabilities were fixed. Ignoring the presence of the discrete choice would mis-attribute this response to the option value from the discrete choice as pertaining to the marginal utility of consumption and potentially yield substantially biased results.

Literature Review The limitations of identification of the dynamic discrete choice model have been well studied. In general, the dynamic discrete choice model is not identified without imposing additional restrictions, usually on the utility or value functions. Rust (1994), Magnac and Thesmar (2002), and Levy and Schiraldi (2020) study the problem of identification of the discount factor and utility function and Komarova et al. (2018) and Abbring and Daljord (2020) discuss the identification of the discount factor, respectively in a fully-parametric and non-parametric framework.

In the context of the Euler-equation models identification has been studied by Chen and Ludvigson (2009), Chen et al. (2014) and Escanciano et al. (2017) among others. Specifically, Escanciano et al. (2017) show that the discount factor and the agents' marginal utility is non-parametrically set-identified under mild restrictions and point identified with the additional assumption that the utility is increasing in the continuous-choice variable. The result is based on the result that the Euler Equation can be interpreted as a Fredholm integral equation of the second kind, and that the discount factor and the marginal utility are the eigenvalue-eigenfunction pair that solve such an equation. Unfortunately, these results are not robust to the presence of a discrete choice or adjustment costs which modify the Euler equation such that the problem is no longer nested in the Type-II Fredholm class, and the results therefore cannot be applied to a DDCC setting.

In the context of dynamic discrete-continuous model, Chetty and Szeidl (2007) points out the importance of accounting for costly adjustment and infrequent consumption decisions problem in evaluating risk aversion with respect to moderate-stake shocks. However, they do not study the identification or the estimation of the model, instead analyzing a stylized setting where all uncertainty is resolved in the first period and there is no unobserved heterogeneity. Beffy et al. (2019) consider a consumption-savings problem with a discrete set of labor supply choices, which introduces similar optimization frictions but imposes a common utility function across alternatives. Iskhakov et al. (2017) and Gayle (2017) study estimation methods for DDCC models. Iskhakov et al. (2017) provides a computational method to fully solve and estimate these models, Gayle (2017) proposes a two-step estimation procedure in the class of DDCC models that are characterized by finite dependence[2]. These papers do not study the identification of these models and set the discount factor at a given known level, moreover we also allow for time varying individual heterogeneity that affect both the discrete and continuous choices. More closely related to our result is Blevins (2014) who extends the results of non-parametric identification of the utility function in the dynamic discrete choice model to the discrete-continuous models, but in the spirit of that literature he assumes the discount factor is known and imposes a strong form of normalization of utility function in order to achieve identification. In providing identification of the time preferences, we also are able to relax

[2]See Arcidiacono and Miller (2011)

some assumptions about the normalization of the utility function. Finally, Schiraldi ⓡ Levy ⓡ Bracke (2019) study a dynamic discrete-continuous model where individuals are present-biased.

2 Set-up and Assumptions

We consider a single-agent discrete-continuous dynamic optimization model with an infinite time horizon indexed by $t = 1, 2, ...$ and a stationary environment. The agent's period-t flow utility depends on their choices and the period-t state variables $(\bar{s}_t, \zeta_t, \varepsilon_t)$. We assume that $\bar{s}_t$ is observed by the researcher, while $\varepsilon_t \equiv (\varepsilon_{i0t}, \ldots, \varepsilon_{Jt}) \in \mathbb{R}^{J+1}$ and $\zeta_t \in \mathbb{R}$ are unobserved. The observed state may be decomposed as $\bar{s}_t \equiv (s_t, L_t, z_t)$, where $L_t \in \mathbb{R}$ is the stock variable in the agent's intertemporal budget constraint (usually "wealth"), while $s_t \in \mathcal{S}$ captures any remaining payoff-relevant states that are observed by the researcher. We separately write z_t as we will impose some additional constraints relative to s_t later to address selection. As typical in the discrete-choice literature, we assume that $\mathcal{S}$ is finite for simplicity. Given that q_t is continuous, however, we assume that L_t is also continuous. Unobservable preferences shocks are associated with both the discrete and continuous choices, and thus ε_t represents the vector of individual idiosyncratic random preference shocks for each of the discrete alternatives, while ζ_t is an individual-level shock which affects the marginal utility of consumption. The joint state $(\bar{s}_t, \zeta_t, \varepsilon_t)$ evolves according to a time-homogeneous process that we describe below.[3]

At the beginning of the period, the agent observes the realization of $(\bar{s}_t, \zeta_t, \varepsilon_t)$ and simultaneously makes a discrete and a continuous choice $(d_t, q_{jt}) \in \mathcal{J} \times \mathbb{R}$ which is observed by the researcher. Specifically, the agent chooses d_t from a set $\mathcal{J} = \{0, 1, \ldots, J\}$ of discrete, mutually exclusive, and exhaustive alternatives. The agent also chooses the continuous quantity $q_t \in \mathbb{R}$. They then receive an instantaneous payoff $\bar{u}_{d_t}(q_t, s_t, z_t, \zeta_t, \varepsilon_t)$. Note that L_t is omitted from the utility function as we are controlling for q_t and thus considering a direct utility function. L_t is considered to affect choices only through its effect on the intertemporal budget constraint.[4]

[3] Our analysis can be extended to allow for the presence of an unobservable market-level shock, ξ_{jt}, and for the possibility that it is correlated with some observable characteristics of the product, e.g. the price, as in the spirit of the discrete choice models in the I.O. literature. To guarantee that ξ_{jt} is non-parametrically identified, we may assume that it enters additively with the individual unobservable shock and replace in the text below ζ_t with ζ_{jt} where $\zeta_{jt} \equiv \zeta_t + \xi_{jt}$.

[4] Note that in the applied discrete choice literature, the problem is often specified in terms of an indirect utility function and therefore wealth (or income) appears in this function.

Assumptions. We make the following additional assumptions, most of which are standard assumptions used in the dynamic discrete choice literature (Rust, 1994; Aguirregabiria and Mira, 2010, see,), with slight modifications in some cases to allow for the additional continuous choice and the presence of an intertemporal budget constraint. Assumptions 1, 2 and 3 have been widely used in dynamic discrete choice models since Rust (1987) demonstrated their role in generating empirically tractable structural models of dynamic discrete choice (see Rust (1994), Aguirregabiria and Mira (2010)).

Assumption 1. *(Additive Separability of Discrete Shock) The instantaneous utilities are given by, for each $j \in \mathcal{J}$,*

$$\bar{u}_j(q_t, s_t, z_t, \zeta_t, \varepsilon_t) = u_j(q_t, s_t, z_t, \zeta_t) + \varepsilon_{jt}$$

where $u_j(\cdot)$ is continuously differentiable for all j

Assumption 1 is an additive separability condition of the sort used in both static and dynamic discrete choice analysis (e.g. McFadden (1974), Rust (1994)). Note that in our context, although it requires ε_t to affect payoffs additively, ζ_t may still affect payoffs in a nonseparable manner.

Assumption 2. *(Independent Discrete Shock) The unobservable state variable ε_t is iid distributed over time and across agents with support $\mathbb{R}^J$ and CDF $F(\varepsilon_t)$ which has finite first moments and is continuous and twice differentiable in ε_t.*[5]

We proceed assuming that $F(\varepsilon_t)$ is known. As discussed by Rust (1994) and Magnac and Thesmar (2002), establishing non-parametric identification when the distribution of the errors is unknown in the infinite-horizon case presents additional challenges that we want to abstract away from.[6]

Future states are uncertain, and the agent's actions and states today affect the distribution. The evolution of the states is summarized by a Markov transition law $\Gamma(\bar{s}_{t+1}, \zeta_{t+1}, \varepsilon_{t+1} | \bar{s}_t, \zeta_t, \varepsilon_t; d_t, q_t)$. We make the following additional assumptions about the transition of the state variables and the distribution of the shocks:[7]

[5] Rust(1994) presents a weaker version of this assumption where the second and higher moments of ε_t may depend on the observable states. However, this weaker version of the assumption is hardly ever used in practice.

[6] See for example Blevins (2014) who provides new identification results for $F(\varepsilon_t)$ in a similar context to ours under additional exclusion restrictions.

[7] See Blevins (2014) and Hong and Shum (2010) for a similar extension of the conditional independent assumption to the discrete-choice environment.

Assumption 3. *(Conditional Independence): The transition distribution of the states has the following factorization:*

$$\Gamma(\bar{s}_{t+1}, \zeta_{t+1}, \varepsilon_{t+1} | \bar{s}_t, \zeta_t, \varepsilon_t, d_t, q_t) = \\ \lambda(L_{t+1} | \bar{s}_t, s_{t+1}, d_t, q_t)\pi(s_{t+1} | \bar{s}_t, d_t, q_t)W(z_{t+1} | s_t, L_t, d_t)F(\varepsilon_{t+1})G(\zeta_{t+1})$$

where $\lambda(\cdot)$, $W(\cdot)$, $F(\cdot)$, and $G(\cdot)$ have finite first moment and additionally $F(\cdot)$, and $G(\cdot)$ have support $\mathbb{R}$ and are twice differentiable.

Assumption 3 is the conditional independence assumption of Rust (1987, 1994), which limits the serial dependence of the unobservables. We allow s_{t+1} to depend arbitrarily on the period-t state variables and choices, and allow the same for L_{t+1} but also permit dependence on s_{t+1} to allow for stochastic returns. As the unobserved shocks are not the main focus of this paper, we assume that ε_{jt} and ζ_t are independently distributed over time.[8] Similarly, Assumption 3 may in fact allow for time-varying distributions but we suppress such variation and, for example, write $\pi(\cdot)$ rather than $\pi_t(\cdot)$. We assume that z_{t+1} may depend on s_t, L_t, and d_t, but is conditionally independent of q_t and z_t, which is motivated by the next assumption:

Assumption 4. *(Unlimited encouragement) There exists k such that for all $j \in \mathcal{J} \setminus \{k\}$:*

1. *The marginal utility $\partial u_j / \partial q$ is independent of z for all $(q, \bar{s}, \zeta)$.*

2. *For all (s, L) there exists a sequence $\{z_{j,n}\}$ such that $\lim_{n\to\infty} Pr(d_t = j | s, L, z_{j,n}) = 1$*

Assumption 4 is made to address the selection on unobservables — particularly ζ_t — which may affect the conditional continuous choice, and resembles the assumptions made in both the static structural and reduced-form literatures (see for example Dubin and McFadden (1984), Hewitt and Hanemann (1995) and Heckman (1979)).[9] We impose two conditions, which may fail to be satisfied by at most one alternative. The first condition excludes z_t from the marginal utility, and this combined with the conditional independence from Assumption 3 will guarantee

[8]We note that ζ_{t+1} may be permitted to depend on s_t and ζ_t. In that case, Lemma 1 recovers not ζ_t but a polynomial in the lagged state and conditional percentiles of consumption, the coefficients of which must then be estimated along with the rest of the model. As our focus here is on the effect of the discrete choice, we instead impose independence to keep the model parsimonious.

[9]In a static model Assumption 4 part 1 is sufficient to deal with the selection as (for example) discussed in Dubin and McFadden (1984), when coupled with additive separability of the policy function in ζ_t. This separability, however, is generically not satisfied in a dynamic setting even if the marginal utility of consumption is additively separable in ζ_t due to the expected continuation value.

that the optimal continuous choice is independent of z_t for most alternatives. In contrast, the optimal discrete choice may be affected by z_t and the second part of Assumption 4 guarantees that this is the case. Specifically, it states that for all but one alternative, there must be a sequence of z_t such that the probability of choosing alternative j approaches one.

Without the second part of Assumption 4, unknown selection on ζ would render the continuous policy function unrecoverable without imposing difficult-to-interpret and unverifiable assumptions on it.[10] Instead, our Assumption is both clear and verifiable, and applies in many common settings. It may be viewed as a strengthening of the common encouragement design of experiments with non-random treatment assignment. For example, suppose a consumer is choosing between a local bodega in Manhattan and a Costco in New Jersey (which differ in the marginal utility of their goods) but that the econometrician has data on the traffic in the Holland Tunnel. We require that traffic may be sufficiently bad that the consumer chooses the Manhattan store with arbitrarily high probability — but note that we do not require traffic conditions that force them to New Jersey (and moreover the marginal utility in New Jersey may always depend on traffic).

Two common features of decision environments will also satisfy this assumption. In our application to consumption-savings-retirement, we treat retirement as an absorbing choice. We may therefore use the lagged discrete choice as z_t, which clearly satisfies both the conditional independence and part 1 of Assumption 4. Moreover, conditional on starting the period retired, the probability of remaining in this absorbing choice is one. Thus the second part is satisfied as well.[11] We also note that when there is variation in the choice set, as studied by Levy and Schiraldi (2020), the realized choice set will clearly affect the choice probabilities without affecting the conditional continuous choices, and if there is the possibility that the choice set may be a singleton then it may be applied here.

Our next assumption relies on standard economic modeling of the intertemporal budget constraint. We assume that the continuous choice q_t and the continuous state variable L_t enter the problem in a manner similar to how consumption and assets are usually treated. Treating L_t as a general state variable requires some additional notation, but follows the standard budget

[10] In general, equation (12) may have an infinite number of solutions beyond the true policy functions. Unlimited encouragement is a sufficient condition for uniqueness. There are other conditions on the policy functions which are sufficient for uniqueness (e.g. linearity in ζ), but unlike 4, they would not be testable in the data.

[11] A second option in our application would be to use the head of household's age, as empirically the probability of retirement for sufficiently young households approaches zero. Instead, we choose to allow the marginal utility of consumption to be age-dependent (violating part 1 of the assumption) and make use of the absorbing retirement choice.

constraint formulation. First, we assume that L_{t+1} is a known function of $\bar{s}_t, s_{t+1}, L_t$, and the period-t choices. Unlike the the state variables in s_t, economic theory can often provide guidance on how L_t is modeled. Indeed, as in our application, it is often merely an accounting identity. We will therefore make use of this economic structure rather than attempting to recover it from the data. Second, we place restrictions on the form this function may take. Specifically, that the next-period value is a known linear function of the end-of-period state: $f_{d_t}(s_{t+1}, s_t, L_t, q_t)\,(L_t - q_t - \phi_{jt}\,(s_t, L_t))$, where $\phi_{jt}(s_t, L_t)$ accounts for any fixed costs of the chosen alternative. For example, if part of the state is the market interest rate (i.e. $s_t^n = r_t$) then $f_{d_t}(s_{t+1}, s_t, L_t, q_t) = (1 + r_{t+1})$ yields the standard intertemporal asset accumulation equation. That L_t and q_{jt} enter in this form corresponds to an assumption that the agent is a price-taker, as the price does not depend on the quantity chosen.[12] We allow that the returns on the decision-maker's assets are stochastic and depend on s_{t+1}.

Assumption 5. *(Intertemporal budget constraint evolution)*

$$\lambda(L_{t+1}|s_{t+1}, \bar{s}_t, d_t, q_t) = \begin{cases} 0, & L_{t+1} < B(\bar{s}_t, s_{t+1}, d_t, q_t) \\ 1 & L_{t+1} \geq B(\bar{s}_t, s_{t+1}, d_t, q_t) \end{cases}$$

where $B(\bar{s}_t, s_{t+1}, d_t, q_t) \equiv f_{d_t}(s_{t+1}, s_t, L_t, q_t)(L_t - q_t - \phi_{d_t}(s_t, L_t))$ *for known functions* $f_{d_t}(\cdot) \neq 0$ *and* $\phi_j(\bar{s}_t)$ *which are nondecreasing in* L_t*.*

In our setting, Assumption 5 can be interpreted as the usual intertemporal budget constraint. The consumption-independent term $\phi_j(s_t, L_t)$ reflects any other expenditures or fixed costs, and in our application is net income (negative, given that $\phi_j(s_t, L_t)$ enters negatively). In other settings it may reflect transaction costs, switching costs, and so forth. We also note that $\phi_j(\bar{s}_t)$ will only play an important role in Theorem 2, and may be arbitrary or omitted entirely in Theorem 1, which forms the basis of our empirical analysis.

We conclude with the regularity conditions that guarantee an interior solution for the continuous choice.

Assumption 6. *(Regularity Conditions) For all* j, s, L, *and* ζ*:* $u_j(\cdot)$ *is continuous, differentiable, and* $\lim_{q\to\infty} \frac{\partial u_j}{\partial q} = 0$

Finally, we make a non-triviality assumption regarding the transitions of s:

[12] In some settings this will be a restrictive assumption. It may be relaxed, but doing so will require correspondingly stronger assumptions to provide identification.

Assumption 7. *(Non-trivial dependence of transitions) There exists j such that if $q \neq q'$ or $L \neq L'$ then $f_j(\cdot, L, q)\pi(\cdot|s, L, z, q, j) \neq f_j(\cdot, L', q')\pi(\cdot|s, L', z, q', j)$*

Assumption 7 requires that the continuous part of the decision-maker's problem affects the states in some way. One economically meaningful interpretation is that wealth or consumption affect either the return or the distribution of returns the decision-maker faces. Specifically, while they may not affect market returns per se, they may affect their own distribution of returns for example by changing the marginal rate at which the returns are taxed as in our application below. In other contexts, the continuous choice may directly affect the next-period consumption utility, for example through habit formation. As we have assumed a stationary utility function, it is natural to operationalize such spillovers through the state variables.

Finally, we must make an assumption regarding how the unobservable shock ζ enters the utility function. Given that ζ is unobserved, some assumptions such as monotonicity may be without loss of generality as they are a matter of labelling. However, we will require that the policy function for the continuous choice is invertible in ζ, and so we make the common assumption that the utility function is supermodular in ζ and q:[13]

Assumption 8. *(Supermodularity). For all $j \in \mathcal{J}$, the instantaneous marginal utility $\partial u_j(q, s, z, \zeta)/\partial q$ is (weakly) supermodular in q and ζ*

3 Decision process and Identification

In the rest of this paper, for notational convenience, we consider that the utility function does not depend on z_t.

3.1 Preliminary results

We assume individuals discount the future at rate δ in maximizing the present discounted value of their lifetime utilities. Under the assumptions above, the value function from the perspective of the beginning of the period can be expressed recursively as

$$V(\bar{s}_t, \zeta_t, \varepsilon_t) = \max_{d_t, q_t} \{u_{d_t}(q_t, s_t, \zeta_t) + \varepsilon_{d_t t} + \delta E[V(\bar{s}_{t+1}, \zeta_{t+1}, \varepsilon_{t+1})|d_t, q_t, \bar{s}_t]\} \tag{1}$$

[13]Similar monotonicity conditions have been widely used both in empirical work and in identification analysis in related models, including, but not limited to, Matzkin (2003), Blevins (2014), Bajari et al. (2007), Chesher (2003), Hong and Shum (2010), and Srisuma (2015).

Note that the presence of the max operator in equation (1) means that the unconditional value function is not guaranteed to be concave. Because the decision-maker is able to take the upper envelope of the optimal continuous choices across all $|\mathcal{J}|$ discrete choices, there are likely to be important convexities as the agent shifts their discrete choice. These may be smoothed out to a degree by considering the ex-ante value function (or integrated value function), $\bar{V}(\bar{s}_t, \zeta_t)$, defined as the continuation value of being in state $(\bar{s}_t, \zeta_t)$ and integrating $V(\bar{s}_t, \zeta_t, \varepsilon_t)$ over ε_t (Rust, 1987, see,) :

$$\bar{V}(\bar{s}_t, \zeta_t) = \int V(\bar{s}_t, \zeta_t, \varepsilon_t) dF(\varepsilon_t) \tag{2}$$

We now define the conditional value function $v_j(\bar{s}_t)$ as the present discounted value (net of ε_t only) of choosing alternative j and the conditionally optimal quantity q^*_{jt}, and behaving optimally from period t + 1 on:

$$v_j(\bar{s}_t, \zeta_t) \equiv u_j(q^*_{jt}, s_t, \zeta_t) + \delta E\left[\bar{V}(\bar{s}_{t+1}, \zeta_{t+1}) | j_t, \bar{s}_t, q^*_{jt}\right] \tag{3}$$

Turning to the discrete choice, we may write the discrete choice probabilities in terms of the discrete-choice-specific value function as:

$$\Pr(d_t = j_t | \bar{s}_t, \zeta_t) = E\left[1 \cdot (j_t \in \arg\max_d v_d(\bar{s}_t, \zeta_t) + \varepsilon_{dt}) | \bar{s}_t, \zeta_t\right] \tag{4}$$

Under Assumptions 1 and 2, there exists a one-to-one mapping from the conditional choice probabilities to differences in the choice-specific value function in given the vector of states:

$$(\Delta v_1(\bar{s}, \zeta), ..., \Delta v_J(\bar{s}, \zeta)) = \Psi(Pr(d = 1 | \bar{s}, \zeta), ..., Pr(d = J | \bar{s}, \zeta)) \tag{5}$$

where $\Delta v_j(\bar{s}, \zeta) = v_j(\bar{s}, \zeta) - v_0(\bar{s}, \zeta)$ for any $j \neq 0$. Moreover, the ex-ante value function has the additivity property (Rust (1994, Theorem 3.1)):

$$\bar{V}(\bar{s}, \zeta) = \Phi_k(\bar{s}, \zeta) + v_k(\bar{s}, \zeta) \tag{6}$$

for any alternative k. Notice that (5) implies that Φ_k is a unique function of the choice probabilities, which are in turn a function of the state, and so we write $\Phi_k(\bar{s}, \zeta)$ for a more compact notation. We may therefore re-write equation (3) as:

$$v_j(\bar{s}_t, \zeta_t) = u_j(q^*_{jt}, s_t, \zeta_{jt}) + \delta \mathbb{E}\left[\Phi_0(\bar{s}_{t+1}, \zeta_{t+1}) + v_0(\bar{s}_{t+1}, \zeta_{t+1}) | j_t, \bar{s}_t, q^*_{jt}\right] \tag{7}$$

Under the regularity conditions assumed, the agent's continuous choice will be characterized by the first-order condition of (7). Given that the agent is choosing optimally, an envelope condition holds and we have $\partial v_j(\bar{s}_t, \zeta_t)/\partial L = \partial u_j(q^*_j, s, \zeta)/\partial q_j$ (see Appendix A.1). We may therefore write the first-order condition of (7) as:

$$\frac{\partial u_j(q^*_{jt}, s_t, \zeta_t)}{\partial q} = \delta \mathbb{E}\left[f_j(s_{t+1}, s_t, L_t, q_t)\left(\frac{\partial \Phi_j(\bar{s}_{t+1}, \zeta_{t+1})}{\partial L_{t+1}} + \frac{\partial u_j(q^*_{jt+1}, \bar{s}_{t+1}, \zeta_{t+1})}{\partial q}\right) \big| j_t, \bar{s}_t, q^*_{jt}\right] \tag{8}$$

Equation (8) generalizes the familiar Euler equation to our discrete-continuous setting. Relative to the problem with no discrete choice, the right-hand side accounts for the marginal effect on the surplus from choice tomorrow from today's consumption. As the choice surplus in period $t+1$ depends on the period $t+1$ wealth, the agent's period-t choice must account not only for the expected future marginal utility but also this additional effect.

It is possible to write (8) for choice j in terms of any choice and therefore in terms of choice 0. Notice that

$$\mathbb{E}\left[\frac{\partial \Phi_j(\bar{s}_{t+1}, \zeta_{t+1})}{\partial L_{t+1}} + \frac{\partial u_j(q^*_{jt+1}, s_{t+1}, \zeta_{t+1})}{\partial q} \big| j_t, \bar{s}_t, q^*_{jt}\right] = \mathbb{E}\left[\frac{\partial \Phi_0(\bar{s}_{t+1}, \zeta_{t+1})}{\partial L_{t+1}} + \frac{\partial u_0(q^*_{0t+1}, s_{t+1}, \zeta_{t+1})}{\partial q} \big| j_t, \bar{s}_t, q^*_{jt}\right] \tag{9}$$

Moreover, from (5), we know that $v_j(\bar{s}, \zeta) - v_0(\bar{s}, \zeta) = \Psi_j(Pr(d = 1|\bar{s}, \zeta), ..., Pr(d = J|\bar{s}, \zeta))$. As the choice probabilities are in turn functions of the state variables, we will write $\Psi_j(\bar{s}, \zeta)$ for a more compact notation. We differentiate this and again apply the envelope condition from the continuous choice in order to rewrite the marginal utility from any alternative j in terms of alternative 0:

$$\frac{\partial u_j}{\partial q_j} = \frac{\partial u_0}{\partial q_0} + \frac{\partial}{\partial L}\Psi_j(\bar{s}, \zeta) \tag{10}$$

Substituting in to equation (8), we obtain:

$$\frac{\partial u_0(q^*_{0t}, s_t, \zeta_{jt})}{\partial q^*_{0t}} + \frac{\partial \Psi_{jt}(\bar{s}_t, \zeta_t)}{\partial L_t} = \delta \mathbb{E}\left[f_j(s_{t+1}, s_t, q_t, L_t) \left(\frac{\partial \Phi_0(\bar{s}_{t+1}, \zeta_{t+1})}{\partial L_{t+1}} + \frac{\partial u_0(q^*_{0t+1}, s_{t+1}, \zeta_{t+1})}{\partial q} \right) | j_t, s_t, q^*_{jt} \right] \tag{11}$$

We will refer to equation (11) as the generalized Euler equation, and it will form the basis of our identification results.

3.2 Identification

In this section, we provide three approaches to identification based on the variation in the data. The first approach requires the least structure on the data, but requires a more stringent assumption on utility. This is relaxed in the second approach, which in return demands the presence of alternative-specific fixed costs. The third approach establishes identification in the case where the state transition probabilities are choice-independent (i.e. exogenous to the decision-maker). Prior to these results, however, we first establish that in all three cases the continuous policy function is recoverable from the data:

Lemma 1. *Under Assumptions 1–6 and 8, ζ and $\{q^*_j(\bar{s}, z, \zeta)\}_{j \in \mathcal{J}}$ are identified*

Proof of Lemma 1. Under Assumptions 3–6 and 8, $q^*_j(\bar{s}, \zeta_t)$ is invertible in ζ_t for all j. Moreover, Assumption 3 implies the continuation value (given d_t and $\bar{s}_t$) is independent of z_t, while Assumption 4 implies the marginal flow utility is independent of z_t for all but one alternative (without loss, let $k = 0$) and thus the maximizer of v_j must be independent of z_t as well. Therefore for $j > 0$, q^*_j is not a function of z_t.

We thus have $G(q^{*-1}_j(q, s, L)|\bar{s}, j) = Q(q^*_j(s, L, \zeta)|\bar{s}, j)$, where $Q(\cdot)$ is the (conditional) empirical distribution of q.

Using the law of total probability, we have for all $(\bar{s}, z, \zeta)$:

$$\begin{aligned} G(\zeta) &= \sum_{j \in \mathcal{J}} Pr(d_t = j|\bar{s}) G(\zeta|\bar{s}, d_t = j) \\ &= Pr(d_t = 0|\bar{s}) Q(q^*_0(\bar{s}, \zeta)|\bar{s}, 0) + \sum_{k=1}^{J} Pr(d_t = k|\bar{s}) Q(q^*_k(s, L, \zeta)|\bar{s}, k) \end{aligned} \tag{12}$$

Thus, by part 2 of Assumption 4, we have for any $j > 0$ and any q:

$$\lim_{n\to\infty} Q(q_j^*(s, L, \zeta)|s, L, z_{jn}, j) = G(\zeta)$$

where the unconditional $G(\cdot)$ results from the fact that $\lim_{n\to\infty} Pr(j|s, L, z_{jn}) = 1$ and the continuity of (12) in the choice probabilities. As $Q(\cdot)$ is a CDF of a continuous density and therefore strictly monotone and continuous, we may obtain $q_j^*(s, L, \zeta) = \lim_{n\to\infty} Q^{-1}(G(\zeta)|s, L, z_{jn}, j))$. As q_j^* is invertible, for any $\bar{s} = (s, L, z)$ we also obtain $\zeta = q_j^{*-1}(q, \bar{s})$ and the conditional distribution $G(\zeta|s, L, z, j) = Q(q_j^*(\bar{s}, \zeta)|\bar{s}, j)$.

Finally, consider the remaining alternative 0 and any z such that $Pr(d_t = 0|\bar{s}, z) \neq 0$. We obtain the conditional distribution of ζ by:

$$G(\zeta|\bar{s}, 0, z) = (Pr(0|\bar{s}, 0))^{-1} \left(G(\zeta) - \sum_{k=1}^{J} Pr(k|\bar{s}, z) G(\zeta|\bar{s}, k, z) \right)$$

and then given $(\bar{s}, q)$ we obtain $\zeta = G^{-1}(Q(q|\bar{s}, 0, z)|\bar{s}, 0)$ if $d_t = 0$, while $q_0^*(\bar{s}, \zeta) = Q^{-1}(G(\zeta|\bar{s}, 0)|\bar{s}, 0)$. □

We note that the result is robust to the inclusion of a market-level shock to the marginal utility, as we show in Appendix A.5.[14] Moreover, if Assumption 3 is relaxed such that ζ_{t+1} is correlated with $\bar{s}_t$ or ζ_t, the lemma serves only to recover ζ_t as an unknown function of $\bar{s}_{t-1}$ and the conditional percentiles of period-t consumption. This must therefore be estimated as part of the continuous policy function. As the unobserved shocks are not the focus of this paper, we take the simple version of Lemma 1 as stated and proceed to establish our main identification results.

3.2.1 Theorem 1

Our first identification result, which will underpin our empirical estimation, makes a standard, though substantive, assumption on the marginal utility. We strengthen Assumption 8 to assume that the marginal utility is separable in ζ.

Assumption 8′. *(Separable marginal utility)*

[14] That is, the marginal utility of alternative j receives both an individual shock ζ_{it} and aggregate shock ξ_{jt}.

For all $j \in \mathcal{J}$, marginal utility is given by:

$$T\left(\frac{\partial u_j(q,s,\zeta)}{\partial q}\right) = T\left(\frac{\partial u_j(q,s)}{\partial q}\right) + T(\zeta)$$

for some strictly monotone transform T. Moreover, $\frac{\partial u_j(q,s)}{\partial q}$ is strictly monotone in q.

We state Assumption 8′ generally, but note that it nests two commonly used variations. When $T(x) = x$, this reduces to additive separability in ζ. Similarly, if $T(x) = \ln(x)$ then Assumption 8′ implies multiplicative separability, i.e. $\partial u_j(q,s,\zeta)/\partial q = \partial u_j(q,s)/\partial q \cdot \zeta$. The latter is commonly used in the macroeconomics literature, and we will impose this case in our empirical application. Finally, the last part of Assumption 8′ ensures that the marginal utility is invertible in q.

Theorem 1. *Suppose Assumptions 1–7 and 8′. Then $(\partial u(\cdot)/\partial q, \delta(\cdot))$ is point identified.*

The full proof appears in Appendix A.2. Under Assumption 8′, the separability of the continuous shock means that (L_t, ζ_t) pairs which would lead to the same quantity conditional on choosing alternative 0 differ in their current-period marginal utility only through ζ_t. Because it is possible to recover ζ_t by Lemma 1, this difference is known to the econometrician. The proof proceeds by establishing that there exist also states which lead to the same end-of-period wealth, and thus have the same support for period-$(t+1)$ wealth. The distribution of s_{t+1} differs, however, and thus it is possible to construct a system of linear equations which is invertible after considering multiple starting wealth levels. One may first recover the expected marginal continuation value, then the marginal utility, and finally $\delta(s_{t+1})$.

3.2.2 Theorem 2

In some settings the separability of Assumption 8′ is too strong. Our next result relax this assumption by using the possibility that there is variation in the fixed costs of the various alternatives in the data.

Suppose that $\bar{s}_t = (s_t, L_t, z_t, \nu_t)$, where we now introduce $\nu_t \equiv \{\nu_{0t}, \ldots, \nu_{Jt}\}$ to represent possible shocks to the fixed costs of choosing each alternative. We assume that ν_t is independent over time, and maintain the rest of the factorization of Assumption 3.[15]. As ν_t represents cost

[15]That is, we now write:
$\Gamma(\bar{s}_{t+1}, z_{t+1}, \zeta_{t+1}, \varepsilon_{t+1}|\bar{s}_t, z_t, \zeta_t, \varepsilon_t, d_t, q_t) = \lambda(L_{t+1}|\bar{s}_t, s_{t+1}, d_t, q_t)\pi(s_{t+1}|\bar{s}_t, d_t, q_t W(z_{t+1}|\bar{s}_t, d_t))F(\varepsilon_{t+1})G(\zeta_{t+1})H(\nu_{t+1})$
where $H(\cdot)$ is treated as known as ν_t is observed.

variation, we make the following assumption:

Assumption 9. *(Cost variation) For all j, s, L: $\phi_j(s, L, \nu_j)$ is continuous in ν_j; $\lim_{\nu_j \to -\infty} \phi_j(s, L, z, \nu_j) = -\infty$; $\lim_{\nu_j \to \infty} \phi_j(s, L, \nu_j) = \infty$; and $E_\nu[\phi_j(s, L, \nu_j)]$ is finite. Furthermore, the marginal utility does not depend on ν.*

Assumption (9) implies that arbitrary variation in the fixed costs $\phi_j(s, L, \nu_j)$ will be available in the data. In practice, only a finite amount of variation will be available and indeed the proof of Theorem 2 will only require finite variation. In particular, there must be sufficient variation in the costs to match continuous choices or end-of-period wealth levels across states or discrete choices. The statement of assumption 9 is sufficient to guarantee this variation exists, but in practice the particular variation needed in any application may be checked ex-post. The assumption moreover states that the marginal utility of consumption is not affected by these cost shocks, which is standard in most settings as they represent a purely financial outcome.

Theorem 2. *Suppose Assumptions 1–9. Then $(\partial u(\cdot)/\partial q, \delta(\cdot))$ is point identified.*

The proof of Theorem 2 proceeds analogously to Theorem 1, but makes use of the observable cost shocks of Assumption 9 (along with the current wealth) to avoid increasing the number of unknowns in the period-t continuous choices and period-$(t+1)$ value function where Theorem 1 relied on simultaneous variation in L_t and ζ_t.

3.2.3 Theorem 3

In some applications, the state transitions may be truly independent of the decision-maker's choices. Such applications treat the decision-maker as a price-taker, and assume that only market-level variables affect utility. In such cases, Assumption 7 does not hold, and the rank condition in equation (38) will fail. We now show that an alternative assumption can suffice.

Assumption 7′. *(Choice-independent transitions and returns) There exist j, k such that for all s, s', L and q: $f_j(s) = \alpha f_k(s)$ where $\alpha \neq 1$ and $\pi(s'|\bar{s}, j, q) = \pi(s'|\bar{s}, k, q)$.*

Assumption (7′) places two requirements on the data. First, that there exist two discrete alternatives which generate the same conditional transition probabilities. Although this is a strong assumption to make in general, we note that the purpose of this section is to provide an identification approach in exactly those applications where the decision-maker is faced with market-level variables that are plausibly exogenous to their choices. The substantive requirement, therefore, is that the gross returns for these two alternatives differ by a common ratio across states. This will be necessary to keep the number of unknown values from proliferating when comparing the continuation values when choosing j and k, and while certainly not without loss of generality, it is nonetheless a condition that may be verified in the data.

Theorem 3. *Suppose Assumptions 1–7′ and either 8′ or 8–9. Then $(\partial u(\cdot)/\partial q, \delta(\cdot))$ is point identified.*

3.2.4 Theorem 4: Utility

We next turn to the identification of the levels of utility from the marginal utilities. As is well-known, some degree of normalization is required given that only differences in utilities affect the decision-maker's choices. We impose that the utility of consuming $q_0 = 0$ for the reference alternative is normalized to zero across states. Note that this is substantially weaker than what is often imposed. For example, Blevins (2014) makes a similar, though considerably stronger, assumption to identify the level of the utility function — namely, that u_0 is known for all s, L, ζ, *and* q.

Assumption 10 (Normalization of utility)**.** $u_0(0, s, \zeta) = 0$ *for all* $\bar{s}$ *and* ζ

Adding this normalization to our existing results allows us to recover the levels of utility:

Theorem 4. *Suppose Assumptions 1–8′, or 1–9, or 1–7′ and either 8′ or 8–9; and suppose Assumption 10. Then $(u(\cdot), \delta(\cdot))$ is point identified.*

Proof. Theorems 1–3 imply that the discount factor and marginal utilities are identified. We proceed by showing that given these and the additional normalization of Assumption 10, the levels of utility are also identified by the discrete choice.

Assumption 6 normalizes the level of utility when consuming $q_0 = 0$, and therefore the by the fundamental theorem of calculus u_0 is pointwise identified . Given u_0 and δ, we next observe that v_0 is also uniquely determined. Letting $j = 0$ in equation (7), v_0 may be defined recursively. Letting E_0 be the expectation operator defined in (7) conditioned to $j = 0$, and given $\delta < 1$, it follows that δE_0 is a contraction. This implies that $(I - \delta E_0)^{-1}$ exists, and we can thus write $v_0 = (I - \delta E_0)^{-1}(u_0 + \delta E_0 \Phi_0)$.

Finally, we identify the utilities from any alternative $k \neq 0$. By equation (10), the marginal utilities follow directly from the known marginals of alternative 0, and thus alternative k is known up to a constant. Given the marginal utility of alternative j as well as v_0, equation (7) is also known up to this same constant. We may therefore difference equation (6) for alternative k and alternative 0 to obtain a single equation which is linear in this constant and otherwise known, which completes the proof. □

4 Application to a Consumption-Savings-Retirement Setting

We now apply our identification results based on Theorem 1 to estimate a dynamic discrete-continuous choice model over consumption and working vs. retirement. The consumption-savings part of the problem is part of a large literature in macroeconomics,[16] but the discrete side is largely ignored or treated separately.[17] We show that it in fact has important consequences for the estimates and policy consequences in this setting.

4.1 Description and Data

We estimate the model described above using data from the Panel Survey of Income Dynamics (PSID). The PSID is a longitudinal study of individuals and families that began in 1968 which contains annual information about the income, employment, and demographic characteristics of individual households. A module of questions assessing wealth was first introduced in 1984 and systematically included in every wave beginning in 1999. Starting also with the 1999 wave, the PSID began collection of information on a larger number of consumption components. Our sample therefore covers the period 1999 to 2017. We eliminate households and observations with missing values, and treat new households formed by the offspring of panel members as separate entities. The final sample contains 15,864 households and 73,832 observations.

We follow Blundell et al. (2016) in constructing measures of wealth, income, and consumption from the data recorded in the PSID. Wealth and income variables are provided at a high degree of granularity. Household income comprises: wage income, farming/market gardening, rent from roomers or boarders, other rental income (net), dividends, interest, trust funds or royalties, alimony, all of the preceding accruing to the reference person's spouse, and total taxable income of any other household members. Assets comprise: real estate (net of mort-

[16] See among many others Attanasio and Weber (1995), Attanasio and Low (2004), Blundell et al. (2008)

[17] See for example Aguirregabiria (2010) where only the choice of retirement is analysed.

gages), vehicles, businesses, stocks, checking and savings accounts, retirement accounts (including defined-contribution, defined-benefit, and hybrid employer programs), cash value in life insurance policies, valuables and collectables, and any rights to a trust or estate. Counterfactual incomes for working newly retired heads are estimated by replacing the reference person's wage income or social security benefits with estimated social security benefits or wages in the previous wave, respectively. State and Federal income taxes and marginal tax rates were estimated using the TAXSIM27 program developed by the National Bureau of Economic Research (Feenberg and Coutts, 1993).

In contrast to the high degree of precision with which wealth and income are recorded, a known challenge to using the PSID is the relative paucity of expenditure variables. Prior to the 1999 expansion of the survey, only housing and food-related expenditures were recorded in the PSID, leading many researchers to use food expenditure as a proxy for total expenditure in order to make use of the full time-series. We make use of the expanded variables in order to construct a measure of consumption which includes: food at home, food away from home, rent (or imputed rent for homeowners), home insurance and utilities, travel (car insurance, repair, fuel, bus and taxi fares, and other transport), education (tuition, other school expenses), childcare, and health (insurance, hospital and doctor charges, and prescriptions). Although not exhaustive, reviews of these expenditure variables have found that match closely the values recorded in the more comprehensive Consumer Expenditure Survey where both are recorded, and constitute the great majority of total household expenditures (Andreski et al., 2014). Nevertheless, the PSID data may omit some of the variation in household expenditures.[18]

In order to maintain the stationarity of the state variables, all prices and returns are converted to real values using the annual GDP deflator provided by the US Bureau of Economic Analysis. The risk-free rate of return is assumed to be the effective federal funds rate. Both series were retrieved from the Federal Reserve Bank of St. Louis.

Finally, although a large literature studies intra-household bargaining over consumption and labour supply, we treat households as a single decision-maker. New households which enter the panel through offspring leaving the household or separation of adult partners are treated as independent households, and we treat household size itself as a state variable. The discrete choice of working or retirement is based on the labor force participation status of the head of

[18]An alternative approach would be to attempt to back out consumption from the change in assets across waves of the panel. This approach would require household-level data on returns, however, and would otherwise mis-interpret high returns as low consumption — a highly undesirable form of error to introduce in to the Euler condition estimation.

household, as the labor force attachment of non-primary earners is often poorly characterized by a binary status, and longer but non-permanent spells of voluntary unemployment are more common (e.g. for parental leave). Labor force participation is based primarily on self-reported status in order to separate retirement from unemployment. Finally, in order to avoid counting "soft retirement", reference persons with wage income are characterized as working regardless of their self-report.

Table 1: Summary Statistics

	All	Working	Retired	Difference
	(1)	(2)	(3)	(4)
Age	45.25	39.82	63.53	-23.71***
	(16.47)	(11.94)	(16.43)	
Household size	2.64	2.79	2.16	0.63***
	(1.47)	(1.47)	(1.36)	
Years of education	13.65	13.82	12.66	0.99***
	(2.79)	(2.68)	(3.03)	
Not in good health	0.162	0.106	0.355	-0.249***
	(0.369)	(0.308)	(0.478)	
Total income	65,421	77,937	23,351	54,585***
	(106,570)	(114,649)	(55,257)	
Total wealth	585,734	609,012	507,484	101,529***
	(1,568,329)	(1,533,979)	(1,676,334)	
Consumption	34,022	35,748	28,220	7,528***
	(25,251)	(25,778)	(22,445)	
Observations	73,832	56,904	16,928	

Notes: Income, wealth, and consumption are in 2017 dollars. "Not in good health" is self-rated health status "poor" or "fair" rather than "good", "very good", or "excellent". Household size is number of persons including reference person.

Table 1 reports summary statistics for the overall sample, and separately for the working and retired subsamples. Unsurprisingly, heads of retired households are significantly older than working households, The distributions of age of head, however, do have a substantial overlap in their support, which is consistent with our model of retirement choice. Working households are also larger on average, as children are more likely to be present in younger households, Retired households are more likely to not report being in good health. Working households also report a significantly mean household income than do retired households.[19] They also report higher total wealth than retired households, as many of the latter have already spent much of their

[19] While both means are of course heavily influenced by outliers, the difference in median income also differs significantly across the two groups.

retirement savings. This is also reflected in the relative consumption of the two subsamples, as retired households' mean annual consumption of $28,220 is higher than their mean income but lower than that of working households.

4.2 Empirical Specification

While we have formally established conditions for nonparametric identification and therefore the sources of variation in the data that identify the model, we now specify a utility function in order to apply a more conventional parametric estimation. Additionally, as it is known since Rust (1994) that a closed-form solution for the choice probabilities is available under the stronger assumption that ε_{ijt} are drawn from a common extreme value distribution, for convenience we will make use of this stronger alternative assumption instead of Assumption 2:

Assumption 2′. *(Extreme Value)* ε_{ijt} *is distributed iid* $EV(0, \sigma_\varepsilon)$.

and in the same spirit we operationalize Assumption 3 with

Assumption 3′. *(Normal)* ζ_{it} *is distributed iid* $N(0, 1)$.

4.2.1 Utility

We treat households as unitary actors, abstracting from considerations of intra-household bargaining. Household i maximises their lifetime expected discounted utility, given available resources, by choosing whether to retire or to work, i.e. $d_{it} \in \{R, W\}$,[20] and how much to consume, q_{ijt}. Utility is assumed to be inter-temporally separable, and to match the theoretical setup the time horizon is infinite and future utility flows are discounted geometrically at a rate δ_i (potentially a function of individual characteristics). The flow utility function for each alternative has two components, namely: $u_{ijt}(q_{it}, s_{it}; \theta_j) = u_j^q(q_{it}, s_{it}, \zeta_{it}; \theta_j^q) + u_j^d(s_{it}; \theta^d)$. $u_j^q(q_{it}, s_{it}, \zeta_{it}; \theta_j^q)$ is a function of the continuous quantity chosen q_{it}, and $u_j^d(s_{it})$ captures any utility that consumers obtain from working or retirement which is not related to consumption. The structural parameters $\theta_j = \{\theta_j^c, \theta^d\}$ along with the discount factor are to be estimated.

We assume that the consumption-dependent components are of the Constant Relative Risk Aversion (CRRA) form and, similarly to Attanasio and Weber (1995), assume that utility in

[20]We focus on the labour supply decision of the head of household. Including the spousal decision would create a state-dependent choice set, as two-adult households would have choices involving secondary-earner labour supply, which would not be available to single-adult households.

consumption is shifted by household characteristics, X_{ijt}, such as family size, head of household age, and the unobservable component ζ_{it}. That is, $u_j^q(q_{it}, s_{it}, \zeta_{it}) \equiv \frac{q_{it}^{1-\gamma_j}}{1-\gamma_j} \exp(X_{ijt}\beta_j + \sigma_\zeta \zeta_{it})$, where $\theta_j^q = \{\gamma_j, \beta_j, \sigma_\zeta\}$ are the parameters to estimate and may differ across the discrete choices.

The second component, $u_j^d(s_{it}) \equiv M_{ijt}\theta^d + \varepsilon_{ijt}$, is a function of consumer characteristics, M_{ijt}, which for the working choice includes: age, year of education, log of net real income (i.e. wage plus other sources of income), health status plus the logit error shock ε_{iWt}. For the retirement choice, it includes only the log of net real retirement income (which generically differs from working income) and the logit error shock ε_{iWt}.[21]

Households are able to move resources over time by saving or borrowing, and we abstract from credit constraints. We denote with L_{it+1} the wealth/stock of assets in period $t+1$ with real risk-free gross interest rate of r_{t+1} between periods t and $t+1$. The net return, r_{ijt+1}, depends from the individual decision of retirement as the marginal tax rate is potentially different as the income is different in these two different alternatives. We assume that a non-retired head of household may freely choose to retire and the quantity to consume, but once retired can choose the continuous consumption level only. Given these assumptions, the problem for a working household is given by:

$$\max_{\{(q_{it}, d_{it})\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \delta_i^t \left(\frac{q_{it}^{1-\gamma_{d_{it}}}}{1-\gamma_{d_{it}}} \exp(\beta'_{d_{it}} X_{id_{it}t} + \sigma_\zeta \zeta_{it}) + M_{id_{it}t}\theta^d + \varepsilon_{ijt} \right) |s_{it}, L_{it} \right]$$

subject to the inter-temporal budget constraint

$$L_{it+1} = (1 + r_{id_tt+1}) \left(L_{it} + Y_{ijt} - q^*_{id_{it}t} \right) \tag{13}$$

where $Y_{id_{it}t}$ is the net income at period t which includes labor income if the household works and pension and social security benefits otherwise net of the income tax paid. Income in either case is after-tax, and thus taxes enter the problem as ν_{it}. All income and consumption variables are real.

If the head of household is already retired then they solve the following problem which

[21] While the utility of consumption is recovered from the Euler equation, the second component is identified up to a level normalization one one alternative, therefore the constant is normalized to zero in M_{iRt}.

reduces to the standard one in the consumption-savings literature:

$$\max_{\{q_{it}\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \delta_i^t \left(\frac{q_{iRt}^{1-\gamma_R}}{1-\gamma_R} \exp(\beta_R' X_{iRt} + \sigma_\zeta \zeta_{it})\right) | s_{it}, L_{it}, R_i\right]$$

subject to the inter-temporal budget constraint

$$L_{it+1} = (1 + r_{iRt+1})\,(L_{it} + Y_{iRt} - q_{iRt}^*) \tag{14}$$

The utility function specified above satisfies Assumption $8'$, i.e. marginal utility is multiplicative separable in ζ_{it} and therefore the continuous utility preferences along with the discount factor are identified under Theorem 1. Finally, note that in this context $\phi(\bar{s}_{ijt})$ accounts for the gross income as well as taxes (on income and returns) conditional on the working/retirement choice, thus that variation in $\phi(\bar{s}_{ijt})$ is also useful in identifying the preferences and discount factor as described in Theorem 2.

4.3 Estimation

We propose a two-stage estimation procedure in the spirit of Hotz and Miller (1993), Hotz et al. (1994) and Bajari et al. (2007) among others. In the first stage, we estimate the policy functions and recover ζ_{it}, and the second stage, we estimate the structural parameters.

First stage. In the first stage we estimate the continuous and discrete policy functions and to retrieve ζ_{it}. We first estimate the unconditional probability of individual i choosing alternative j, which can be found by integrating the multivariate logit probability over the distribution of the unobserved individual characteristics ζ_{it}: $\overline{Pr}_{ijt}(\bar{s}_{it}; \lambda^d) = \int Pr_{ijt}(\bar{s}_{it}, \zeta; \lambda^d) g(\zeta) d\zeta$.[22] The probabilities are functions of state variables, with λ_d being the parameters to estimate. Therefore probability of observing individual i choosing to work or retire is given by the following expression:

[22] Simulated choice probabilities are computed averaging the results from 50 random draws taken for every observation from a standard-normal distribution. We use Halton draws to further reduce the sampling variance. Results do not change substantively when we use more draws.

$$LL = \sum_{i,j,t} d_{ijt} \log \overline{Pr}_{ijt}(\bar{s}_{it}; \lambda^d) \tag{15}$$

The unconditional probability $\overline{Pr}_{ijt}$ is sufficient to estimate the continuous policy function as discussed below and retrieve ζ_{it}. We will, however, require the discrete policy conditional on $\hat{\zeta}_{it}$, i.e. $\widehat{Pr}_{ijt}(\bar{s}_{it}, \hat{\zeta}; \hat{\lambda}^d)$, in the structural estimation step and the policy counterfactual.

We then proceed to recover ζ_{it} along with continuous policy functions for working and retired households respectively. Notice that in our application choosing to retire is an absorbing action, which satisfies Assumption 4 as given that the probability of remaining in this absorbing choice is one. It also satisfies the conditional independence assumption as the current choice is a sufficient statistic for the future value of the lagged choice. We thus may estimate the continuous policy function for retired households by MLE. Specifically, we specify $\ln q^*_{iRt} = \mu(\bar{s}^{\mu}_{iRt}; \lambda^{c1}_R) + \sigma(\bar{s}^{\sigma}_{iRt}; \lambda^{c2}_R) \cdot \zeta_{it}$, where $\mu(\bar{s}^{\mu}_{iRt}, j_{it}; \lambda^{c1}_R)$ and $\ln \sigma(\bar{s}^{\sigma}_{iRt}, j_{it}; \lambda^{c2}_R)$ are polynomials in the state variables with parameters $\lambda^c_R = \{\lambda^{c1}_R, \lambda^{c2}_R\}$ to estimate. As the policy is invertible, we can retrieve the unobserved ζ for all households that newly choose to retire or who are already retired, i.e. $\hat{\zeta}_{it}(q_{iRt}|d_t = R) = q^{*-1}_{iRt}(q_{iRt}; \hat{\lambda}^c_R)$ where q_{iRt} is the observed quantity consumed.

We must then retrieve the unobserved ζ_{it} for those households who choose to work, and in doing so also cover the continuous policy function conditional on working. We parametrize this similarly to before with $\ln q^*_{iWt} = \mu(\bar{s}^{\mu}_{iWt}; \lambda^{c1}_W) + \sigma(\bar{s}^{\sigma}_{iWt}; \lambda^{c2}_W) \cdot \zeta_{it}$. To estimate the unknown parameters λ^{c2}_W, we use a GMM estimator where the set of moment conditions are based on the fact we can write ζ_{it} as function of the unknown parameters entering the working policy function:

$$\zeta_{it} = q^{*-1}_{iRt}(q_{iRt}; \hat{\lambda}^c_R) \cdot 1_{j=R} + q^{*-1}_{iWt}(q_{iWt}; \lambda^c_W) \cdot 1_{j=W} \tag{16}$$

Assuming that ζ_{it} is drawn from a standard normal distribution, we compute the following empirical moments:

$$g_n(\lambda_W^{c2}) = \begin{bmatrix} \left(\sum_{i,t} \tilde{\zeta}_{it}\right)/(N \cdot T) \\ \left(\sum_{i,t} \bar{s}^{\mu}_{ijt}\overline{\Pr}_{ijt}(\bar{s}_{it}, \hat{\lambda}_d) * \tilde{\zeta}_{it}\right)/(N \cdot T) \\ \sum_{i,t} \left(\tilde{\zeta}_{it} - \left(\sum_{i,t} \tilde{\zeta}_{it}\right)/(N \cdot T)\right)^2/(N \cdot T) - 1 \\ \left(\bar{s}^{\sigma}_{iWt}{}' \, \bar{s}^{\sigma}_{iWt}\right)^{-1} \bar{s}^{\sigma}_{iWt}\tilde{\zeta}^2_{it} \end{bmatrix}$$

The first two moments are based on the condition that $E[\zeta_{it}] = 0$ and $E[\zeta_{it}|\bar{s}^{\mu}_{ijt}\bar{\Pr}_{ijt}] = 0$,[23] the third moment use the fact that $var(\zeta_{it}) = 1$. The last moments will guarantee that $\tilde{\zeta}_{it}$ is uncorrelated with observables in X_{it} and follow the spirit of the Breush-Pagan test.[24]

Second stage. Next, we construct a set of moments from the continuous choice problem as well as from the discrete one. We focus on the continuous part first. The optimal consumption quantity is determined given the discrete choice j by the first-order condition of the value function:

$$\left(\frac{\partial v_{ijt}}{\partial q_{ijt}}\right) = 0$$

As discussed in section 3.1 and by using the specification in (4.2.1), we can re-write the equation above as follows:

$$q_{ijt}^{-\gamma_j} \exp(X_{ijt}\beta_j + \sigma_\zeta \zeta_{it}) = E\left[(1 + r_{ikt+1})\delta_i \left((q_{ikt+1}^{-\gamma_k} \exp(X_{ikt+1}\beta_k + \sigma_\zeta \zeta_{it+1})) - \frac{\partial \log Pr_{ikt+1}}{\partial L_{it+1}}\right) |\bar{s}_{it}, j_{it}, q_{ijt}\right] \tag{17}$$

We then remove the expectation by using the realized choices and states in period $t+1$ and add the short run error term η^c_{it}. This will avoid making arbitrary assumptions on evolution of the states, instead relying only on the validity of the expectations operator in a random sample. By taking a log transformation and manipulating the equation above we obtain:

$$\Delta \log q_{ikt+1} = \frac{1}{\gamma_k}\Big(\log(1 + r_{kt+1}) + \log(\delta_i) + \beta_k' X_{ikt+1} - \beta_j' X_{ijt} + \sigma_\zeta \Delta\hat{\zeta}_{it+1} + \Upsilon_{ikt+1} - (\gamma_k - \gamma_j)\log(q_{ijt}) + \mu + \tilde{\eta}^c_{it}\Big) \tag{18}$$

[23] We condition on $\bar{s}^{\mu}_{iWt}\bar{\Pr}_{iWt}$ as in Dubin and McFadden (1984).

[24] We have tested the validity of these moments in a Monte Carlo simulation (available upon request).

where $\Delta\hat{\zeta}_{it} \equiv \hat{\zeta}_{it+1} - \hat{\zeta}_{it}$ and $\Upsilon_{ikt+1} \equiv \log\left(1 - \frac{\partial \log \hat{P}r_{kt+1}}{\partial L_{it+1}} \frac{q_{ikt+1}^{\gamma_k}}{\exp(\beta_k' X_{ikt+1} + \sigma_\zeta \hat{\zeta}_{it+1})}\right)$. Notice that while the expectation of the expectational error from equation (17) is zero (i.e. $E(\eta_{it}^c) = 0$), the non-linear transformations required to obtain equation (18) leave the term $\log(1 + \eta_{it}^c)$ which is not mean-zero. We thus define $\mu \equiv E(\log(1 + \eta_{it}^c))$ and treat it as a nuisance parameter, leaving the de-meaned $\tilde{\eta}_{it}^c \equiv \log(1 + \eta_{it}^c) - \mu$ as the error term.

Notice that the first line of equation (18) is similar to the one traditionally estimated in the consumption-savings literature (see for example Attanasio and Weber (1995)), but the discrete choice creates some key differences: (1) the expected future surplus generates the term Υ_{ijt+1}; (2) if the agent consumes different alternatives in periods t and $t+1$ then the change in γ must be directly accounted for; (3) the discount factor (along with $u_j^q(q_{ijt}, s_{it}, \zeta_{it}; \theta_j^q)$) is identified by Theorem 1 and may depend on (possibly stochastic) individual characteristics; (4) the elasticity of intertemporal substitution is no longer equal to $\frac{1}{\gamma_k}$ but to $\frac{1}{\gamma_k}(1 + \frac{\partial \Upsilon_{ijt+1}}{\partial \log(1+r_{ikt+1})})$ for the working households.

Define Z_{it}^c as the set of IVs[25] to obtain the following moments:

$$m_c = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{n_t} (Z_{it}^c)' \, \tilde{\eta}_{it}^c \tag{19}$$

where NT is the total number of observations. Moreover to separately identify the discount factor from $\mu_{\tilde{\eta}_i^c}$ we also use the following set of moments:[26]

$$m_{c2} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{n_t} \eta_{it}^c \tag{20}$$

For the discrete part, we start by taking the log of the probability ratio

$$\log\left(\frac{\Pr_{iWt}}{\Pr_{iRt}}\right) = v_{iWt} - v_{iRt} \tag{21}$$

First notice that left-hand side is estimated in the first step. Moreover, $v_{iRt} \equiv u_R(q_{iRt}^*, \bar{s}_{it}, \zeta_{it}) +$

[25]In all models estimated in Table 2, we use as instruments along with the constant term, the current real net return, lagged family size, lagged income, lagged consumption, and lagged log ζ , in addition to log age, log education, health status and in columns 2 and 3 we also use retirement status interacted with the previous IVs. Finally, we also use the probability of working interacted with the previous IVs in the full structural model in column 3.

[26]For the full structural model we also interact η_{it}^c with the probability of retiring.

$\mathbb{E}\left[\sum_{\tau=1}^{\infty} \delta_i^\tau u_R(q^*_{iR\tau}, \bar{s}_{i\tau}, \zeta_{i\tau}) | \bar{s}_{it}, R\right]$ and as discussed in section 3.1 we can write $v_{iWt} = u_W(q^*_{iWt}, \bar{s}_{it}, \zeta_{it}) + \mathbb{E}\left[v_{iRt+1} - \log(\Pr_{iRt+1}) | \bar{s}_{it}, W\right]$. As before we replace the future term with the realization of the future consumption and state variables as observed in the data. Similar to before we remove the expectation and we add the short run prediction error.[27] The realized v_{iRt} is computed using a forward simulation[28] for 30 periods (60 years), matching individuals on all contemporaneous state variables and using the realized transitions to avoid making parametric assumptions. We condition this process not only on $\bar{s}_t$ but on the choices d_t and q_t to obtain $\hat{v}_{iRt+1|\bar{s}_t,q_t,d_t} = \sum_{\tau=1}^{30} \delta_i^\tau \frac{(\hat{q}^*_{iRt+\tau})^{1-\gamma_R}}{1-\gamma_R} \exp(\beta_R' \hat{X}_{iRt+\tau} + \sigma_\zeta \ln \hat{\zeta}_{it+\tau})$. Note that the removal of the expectation and the matching each introduce a prediction error. We can then write equation (21) as:

$$\ln\left(\frac{\widehat{\Pr}_{iWt}}{\widehat{\Pr}_{iRt}}\right) = \frac{{q^*_{iWt}}^{1-\gamma_W}}{1-\gamma_W} \exp(\beta_W' X_{iWt} + \sigma_\zeta \hat{\zeta}_{it}) + \theta' M_{iWt} + \delta_i \left(\hat{v}_{iRt+1|\bar{s}_t,W,q^*_{iWt}} - \log(\widehat{\Pr}_{iRt+1})\right)$$
$$- \left(\frac{{q^*_{iRt}}^{1-\gamma_R}}{1-\gamma_R} \exp(\beta_R' X_{iRt} + \sigma_\zeta \hat{\zeta}_{it}) + \delta_i \hat{v}_{iRt+1|\bar{s}_t,R,q^*_{iRt}}\right) + \eta^d_{it} \tag{22}$$

where η^d_t is the (mean-zero) difference in prediction error. We define a new set of IVs Z^d_{it},[29] and use them to obtain the a new set of moments:

$$m_d = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{n_t} (Z^d_{it})' \eta^d_{it} \tag{23}$$

We then stacks the set of moments from (19), (20) and (23) and proceed to estimate the parameters using a two-step efficient GMM procedure.

4.4 Results

Table 2 reports the main estimates of the full model which includes both the continuous choice as well as the discrete choice and compares them with the Euler Equation estimations obtained

[27] See Kalouptsidi et al. (2020) for an overview of similar constructions used in the dynamic discrete choice literature.

[28] See Hotz et al. (1994).

[29] We use log age, log education, log income for working and retirement and health status along with constant term as instruments

ignoring the discrete part as well when pooling working and retired households.

Panel A of Table 2 presents the parameters entering the continuous part of the utility function, $u_j^q(q_{ijt}, s_{it}, \zeta_{it})$. In column 1, we present the results from the standard Euler Equation estimation approach. In a typical application, the main coefficient of interest is that attached to the net real return which is interpreted as $1/\gamma$. In the absence of a discrete choice, this coefficient is also the elasticity of intertemporal substitution and equal to 0.44 ($= 1/\gamma$ with γ equal to 2.29). In Column 2 we allow γ and the effect of family size to differ by working status as in the main model. We find a substantially lower value of γ_R, 1.71, for the working subsample compared to the retired households, $\gamma_W = 2.31$. This implies a higher intertemporal elasticity of substitution for the working than for retired households. As in the CRRA specification γ is the coefficient of relative risk aversion, this difference implies the intuitively appealing result that retired households exhibit a significantly greater degree of risk-aversion in their instantaneous utility than working households.

The estimated coefficients of relative risk aversion in the full model (column 3) for working and retired households are lower than the previous model, respectively 1.24 and 1.45. While the implied elasticity of intertemporal substitution is unique for retired household and equal to the reciprocal of the coefficient of relative risk aversion, i.e. 0.69, the elasticity of intertemporal substitution for working is not constant across households. The average value of 0.86 adjusts $1/\gamma$ to account for the marginal discrete choice surplus, i.e. $\frac{1}{\gamma_{d_{t+1}}}(1 + \frac{\partial \Upsilon_{id_{t+1}t+1}}{\partial \log(1+r_{id_{t+1}t+1})})$. The EIS thus depends on the realized state variables including individual characteristics, generating heterogeneity not only between households but *within* households over time.

Figure 1 shows the implications of our results in terms of the intertemporal elasticity of substitution. It is not obvious from inspection how large an effect the $\frac{\partial \Upsilon_{id_{t+1}t+1}}{\partial \log(1+r_{id_{t+1}t+1})}$ term will have on this particular elasticity, and we therefore plot the distribution of elasticities in our working sample by splitting it between over- and under-60s. The dashed orange line in Figure 1 indicates the EIS for retired households for reference. The under-60 working distribution in red shows a substantial degree of variation around its mean, confirming that heterogeneity in the EIS is important for this group despite the relatively low probability of retiring. Far more striking, however, is the over-60 working sample, whose bimodal distribution reflects the greater probability of retirement — and hence greater impact of the discrete choice — for this group. The additional peak for this group is clearly centered around the EIS for retired households, as a significant fraction of older working households anticipate imminent retirement and consume accordingly. In applications with a greater number of discrete alternatives, or in general where

Table 2: Structural Estimates

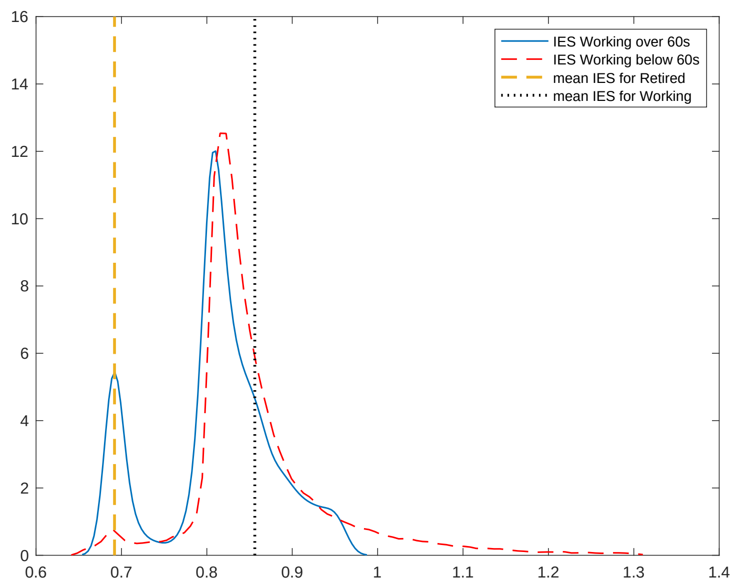
Panel A: $u_j^q(q_{ijt}^*, s_{it}, \zeta_{it})$	Euler Equation		Full Model
γ	2.2945	–	–
	(0.4671)		
γ_W	–	1.7098	1.2428
		(0.3905)	(0.1235)
γ_R	–	2.3104	1.4453
		(0.5165)	(0.1456)
Family Size	0.1589	–	–
	(0.0352)		
Family Size_{W}	–	0.0872	0.0692
	–	(0.0261)	(0.0117)
Family Size_{R}	–	-0.0086	0.0523
		(0.0667)	(0.0218)
Log(Age)	2.4170	2.4892	1.6748
	(0.4993)	(0.5640)	(0.2226)
σ_ζ	0.6507	0.4447	0.3092
	(0.1397)	(0.1101)	(0.0370)
Constant	–	–	-9.3160
			(1.4236)
$\mu_{\tilde{\eta}_i^c}$	-0.0808	0.0208	0.0136
	(0.0092)	(0.093)	(0.0040)

Panel B: $u_j^d(s_{it})$	Full Model
Health Status	-1.0593
(moderate/bad)	(0.0457)
Log(Educ)	-0.3152
	(0.0595)
Log(Net Income_j)	1.7448
	(0.0443)
Log(Age)	-1.9268
	(0.1705)
Constant	7.5114
	(0.5410)

Panel C: $\delta = exp(\beta' X)$	Euler Equation		Full Model
Constant	-0.0582	-0.1890	-0.1193
	(0.0260)	(0.0447)	(0.0176)
Log(Educ)	0.0233	0.0265	0.0178
	(0.0070)	(0.0078)	(0.0043)
Health Status	-0.0308	-0.0342	-0.0217
(moderate/bad)	(0.0075)	(0.0094)	(0.0038)
Implied mean δ_i	0.9885	0.9403	0.9633
	(0.0155)	(0.0227)	(0.0101)
Min δ_i	0.9757	0.9293	0.9562
Max δ_i	1.0031	0.9453	0.9666

there is a greater amount of variation in the marginal discrete choice surplus, one would expect to find an even higher degree of dispersion in this measure.

Figure 1: Distribution of intertemporal elasticity of substitution



In Panel B of Table 2, we present the parameters entering the discrete (consumption-independent) component of utility, $u_W^d(s_{it})$. Recall from the previous section that this is normalized to 0 for the retirement choice. We find that the coefficient on health status in the utility of working is negative and statistically significant, consistent with ill-health being an important driver of retirement. There is also a large negative effect of ln(Age) which indicates that retirement becomes relatively more desirable over time.[30]. Finally, we find that log income enters with a positive coefficient and the income associated with the working choice is typically larger than the retirement income.

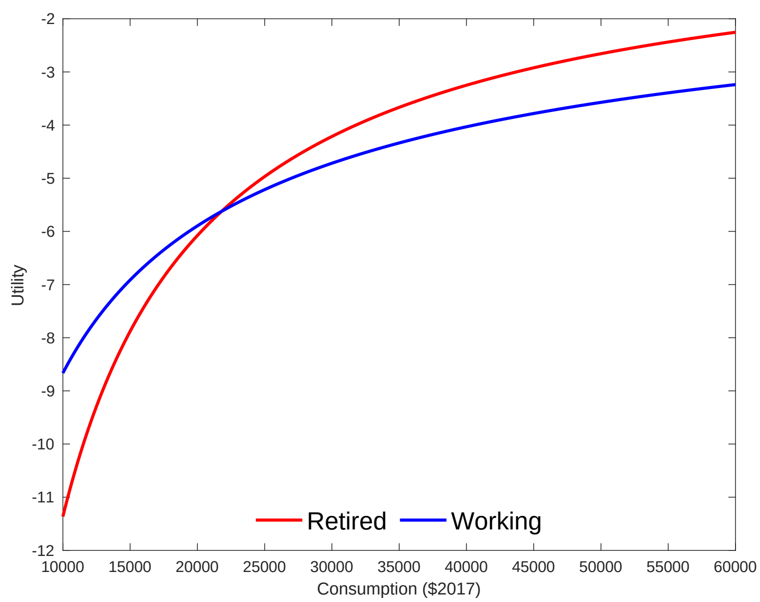
Finally, Panel C of Table 2 presents the estimates of the discount factor δ from the three approaches. The bias in the Euler equation approach is substantial: at the means of the covariates, we estimate a value of δ of 0.99 and 0.94 in columns 1 and 2, respectively, as compared to a mean of 0.963 for the full model, or roughly half the discount rate of column 2. We also estimate a substantial degree of heterogeneity in the discount factor depending on years of education and health status, with additional years of education being associated with

[30]Note this does not follow directly from the fact that retirement is increasing in age, as so too are retirement savings and retirement benefits and therefore retirement consumption opportunities.

a higher discount factor and poor health being associated with a lower discount factor.[31]

To better interpret the coefficients in Table 2, we plot in Figure 2 the utility functions from work and retirement at the means of the covariates. The x-axis indicates the level of consumption and the y-axis the level of utility. The figure indicates that starting from moderate levels of annual consumption of around $20,000, the level of utility is higher for retired households than for working households. Only at very low levels of consumption is the level of utility higher when working, though this may partially be an artefact of the constant relative risk aversion utility specification. As the retired utility crosses the working from below, the effect of age estimated in Panel 2 of Table 2 will serve to push the crossing point further to the left as the head of household ages. As previously noted, prior estimates of retirement vs. working utility based on a discrete-only framework do not identify the discount factor, and therefore could not distinguish the flow utility from the value function. In contrast, Figure 2 shows clearly that for most households that choose to work, it is primarily due to the continuation value (i.e. the ability to finance future consumption, including in future retirement) rather than the contemporaneous utility.

Figure 2: Estimated utility functions at covariate means



[31] Although we treat these covariates as exogenous in our model, the results here to not imply a causal relationship.

4.5 Universal Basic Income

In this section, we apply our results to estimate the welfare effects of a counterfactual policy implementation of a Universal Basic Income (UBI) program. As its name suggests, UBI is a form of negative income tax in which all residents of a country receive a regular unconditional cash transfer from the government, often in lieu of any other income assistance programs. The political debate surrounding UBI has grown dramatically in recent years, but given the substantial costs involved only a limited number of pilot experiments have been conducted (e.g. Kangas et al., 2019). Moreover, one of the primary economic concerns regarding a UBI is the unknown effect on labor force participation rates.

Our results are uniquely suited to estimating not just the magnitude of any labor force participation rate response, but the consequences on lifetime utility of any such response. The utility functions estimated in Section 4.4 indicate that at moderate levels of consumption, the level of utility achieved from the retirement discrete choice is significantly greater than that from the working discrete choice. As the baseline level of income is lower in retirement, a UBI could allow higher levels of non-working consumption, and allow households to change their discrete choice and obtain higher levels of utility. It is unclear *a priori* which households will benefit the most from this option value. The higher taxes needed to finance the program offset this wealth effect and may reduce lifetime wealth for higher-income households, but these households may also be more likely to supplement their UBI income with their existing savings and retire early.

In order to simulate this policy counterfactual it is first necessary to define the policy more precisely, as a wide range of options have been proposed under the heading of UBI. The two main dimensions of heterogeneity are the size of the cash transfer and the adjustments to the tax code required to fund the program. Given the unprecedented scope of such an intervention, it is impossible to state with any certainty what plausible parameters may be (Mogstad and Kearney, 2019, see, e.g.). We therefore consider a variant somewhere in the middle of what has been proposed: a \$750 per month per household transfer, financed by a 15pp increase across all marginal income tax rates (including both labor and capital income). We do not consider any other changes to the tax code, such as to the Earned Income Tax Credit (EITC), which would likely feature in a full UBI policy. The predictions we obtain follow from this particular policy experiment and may differ in important ways if the features of the UBI reform differ.

In order to keep the model tractable, we assume that the policy functions from the previous section remain valid. This requires the further restriction that there is no response on the intensive margin of labor supply, either through changes to wages or to hours worked conditional on employment. This is of course a strong restriction, but to estimate the general equilibrium effects of a UBI is well beyond the scope of this paper. We therefore focus on the effects on consumption and the extensive margin of labor supply only, leaving the rest to future research.

To evaluate the effects of the policy, we perform a Monte Carlo simulation for 30 periods (60 years) with 100 independent replications per household to compute expected utilities and retirement ages. State variables evolve according to an (estimated) AR(1) process, and the unobserved shocks ζ and ε are drawn according to the estimated distributions from the previous section. Simulated choices are then given by the estimated utilities and value functions. Standard errors are obtained by bootstrapping both the structural parameters and the policy functions and re-simulating the policy on the bootstrap sample. Given the computational burden, we choose a random 10% subsample of households for the Monte Carlo (though we resample from the full dataset when bootstrapping the parameter estimates).

We find that the UBI experiment overall leads to a negligible change in average lifetime utility and a modest change in retirement ages. Overall lifetime expected utility increases by approximately 1.3% compared to the baseline level, although this is not statistically distinguishable from zero. This average effect, however, masks substantial heterogeneity, both between working and retired households and within each group. We investigate this heterogeneity in Table 3 by regressing the post-UBI change in lifetime utility on initial age, income, and wealth. We de-mean the regressors within each specification, so that the constant is more usefully interpreted as the mean effect. Among retired households, which tend to be both older and wealthier than working households, the mean change in utility is slightly negative though not statistically significantly so. However, as there is only the continuous consumption choice for these households to make, we find that the policy is significantly better for poorer retired households. We find that wealth is associated with a statistically significant coefficient of -0.081. This conforms to the straightforward intuition that the $750/month benefit is constant across households, but both diminishing marginal utility and the increased tax burden reduce or eliminate the benefit for wealthier households.

In contrast to the progressivity of UBI among retired households, we find that the benefits are actually higher for better-off working households. In column (2) of Table 3, we find that the mean change in lifetime utility for working households is positive though not significant at 0.119 — equivalent to one-off increase in consumption of approximately $1800 at the covariate means. However, the effect is higher at higher wealth levels, and to a lesser extent at higher

Table 3: Welfare Effects of UBI

	Change in Lifetime Utility		Change in Retirement Age
	(1)	(2)	(3)
Constant	-0.414	0.119	-1.207***
	(1.0185)	(1.0299)	(0.0881)
log(Age)	-0.561***	0.057*	0.837***
	(0.1155)	(0.0354)	(0.018)
Education	-0.002	0.012	-0.015***
	(0.0088))	(0.0130)	(0.0024)
Log Wealth	–0.081***	0.062***	-0.261***
	(0.0190)	(0.0090)	(0.0089)
Observations	1519	4628	4628
Sample	Retired	Working	Working

Notes: Dependent variable in columns (1) and (2) is change in expected discounted lifetime utility with UBI policy reform relative to baseline. Dependent variable in column (3) is change in retirement age (in years). All covariates are de-meaned by sample. Bootstrapped standard errors in parentheses.

ages as well. At first glance this may appear counterintuitive, as the relative increase in taxes should diminish the benefits for these households as seen among retired households. However, we find that the discrete choice plays a key role here. As seen in section 3.1, the unconditional value function may be locally convex in ranges where increases in wealth enable the household to switch to the retired discrete choice and obtain the higher consumption utility thereof. If this option value is higher for wealthier households, then it can dominate the higher tax burden. In column (3), we show that the UBI policy does in fact have a larger effect on the discrete choices of wealthier households. While the overall effect is a significant 1.2 year reduction in the average retirement age, each doubling of wealth further magnifies the effect on average retirement age by 3 months.[32] For poorer households, the \$750/month from the policy may simply not be sufficient to enable retirement. It is important to note that these findings thus depend on the specific policy proposal, and may be overturned for different monthly payments or tax reforms. Nevertheless, they highlight the importance of considering the discrete labor supply choice in addition to the continuous consumption choice when evaluating any such reforms.

[32]Note that in addition to the higher retirement consumption available to higher-income households, the 15pp tax increase reduces the gap between retirement and working consumption.

5 Conclusion

In this paper, we study the non-parametric identification of the utility function as well as the discount factor of dynamic discrete-continuous choice models. We prove that these objects are identified under one of the two key identifying assumptions: either the unobservable continuous shock enters additively/multiplicatively in the marginal utility or there is an observable (product specific) fixed costs of choosing a discrete alternative that affects the intertemporal budget constraint but not the utility. We then apply the results to estimate these preferences in the context of a consumption-saving-retirement choice problem. Using the PSID data, we estimate a fully discrete-continuous model to study the consumption saving along with retirement decision. We show that ignoring the discrete choice will bias the estimate. Specifically, estimating the Euler Equation with a CRRA utility function while ignoring the discrete choice will constrain the relative risk aversion parameter to coincide with the intertemporal elasticity of substitution. We show that these two objects are generically separately identified. Moreover, the presence of the discrete alternative implies that agents must take into account the expected surplus from the next-period discrete choice when deciding her continuous choice and therefore the elasticity of intertemporal substitution will varies across agents depending on their expectation about the expected surplus which in turns depends on their current states. Therefore in a model where there are lumpy adjustments/decisions we should expect heterogeneous elasticity of intertemporal substitution a feature that is missing in a simple Euler Equation specification where the lumpy decisions are not fully accounted for. Finally, we use our estimated model to measure the effects of the UBI program. We consider a transfer of $750 per month per household, financed by a 15pp increase across all marginal income tax rates. We find that the UBI experiment leads to a minor increase in lifetime expected utility on average, but while the benefit is decreasing in income among retired households, the presence of the discrete choice reverses this finding among working households where higher-income households respond to the policy by retiring relatively sooner than poorer households. Accounting for discrete as well as continuous choices is therefore important not only for correctly estimating structural parameters, but also directly in conducting counterfactual policy analyses.

References

Abbring, Jaap H. and Øystein Daljord (2020) "Identifying the Discount Factor in Dynamic Discrete Choice Models," *Quantitative Economics*, 11 (2), 471–501.

Aguirregabiria, Victor (2010) "Another Look at the Identification of Dynamic Discrete Decision Processes: An Application to Retirement Behavior," *Journal of Business & Economic Statistics*, 28 (2), 201–218.

Aguirregabiria, Victor and Pedro Mira (2010) "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156 (1), 38–67.

Alan, Sule, Kadir Atalay, and Thomas F. Crossley (2019) "Euler Equation Estimation on Micro Data," *Macroeconomic Dynamics*, 23 (8), 3267–3292.

Andreski, Patricia, Geng Li, Mehmet Zahid Samancioglu, and Robert Schoeni (2014) "Estimates of Annual Consumption Expenditures and Its Major Components in the PSID in Comparison to the CE," *The American Economic Review: Papers and Proceedings*, 104 (5), 132–135.

Arcidiacono, Peter and Robert A. Miller (2011) "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79 (6), 1823–1867.

Attanasio, Orazio, James Banks, Costas Meghir, and Guglielmo Weber (1999) "Humps and Bumps in Lifetime Consumption," *Journal of Business & Economic Statistics*, 17 (1), 22–35.

Attanasio, Orazio and Hamish Low (2004) "Estimating Euler Equations," *Review of Economic Dynamics*, 7 (2), 405–435.

Attanasio, Orazio and Guglielmo Weber (1995) "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, 103 (6), 1121–57.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007) "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75 (5), 1331–1370.

Beffy, Magali, Richard Blundell, Antoine Bozio, Guy Laroque, and Maxime Tô (2019) "Labour supply and taxation with restricted choices," *Journal of Econometrics*, 211 (1), 16–46.

Blevins, Jason R. (2014) "Nonparametric Identification of Dynamic Decision Processes with Discrete and Continuous Choices," *Quantitative Economics*, 5, 531–554.

Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008) "Consumption Inequality and Partial Insurance," *American Economic Review*, 98 (5), 1887–1921.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten (2016) "Consumption Inequality and Family Labor Supply," *The American Economic Review*, 106 (2), 387–435.

Bond, Stephen and Costas Meghir (1994) "Dynamic Investment Models and the Firm's Financial Policy," *The Review of Economic Studies*, 61 (2), 197–222.

Chen, Xiaohong, Victor Chernozhukov, Sokbae Lee, and Whitney K. Newey (2014) "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, 82 (2), 785–809.

Chen, Xiaohong and Sydney C. Ludvigson (2009) "Land of addicts? an empirical investigation of habit-based asset pricing models," *Journal of Applied Econometrics*, 24 (7), 1057–1093.

Chesher, Andrew (2003) "Identification in Nonseparable Models," *Econometrica*, 71 (5), 1405–1441.

Chetty, Raj and Adam Szeidl (2007) "Consumption Commitments and Risk Preferences*," *The Quarterly Journal of Economics*, 122 (2), 831–877.

Dubin, Jeffrey A and Daniel McFadden (1984) "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52 (2), 345–62.

Escanciano, Juan Carlos, Stefan Hoderlein, Arthur Lewbel, Oliver Linton, and Sorawoot Srisuma (2017) "Nonparametric Euler Equation Identification and Estimation," February, Working Paper.

Feenberg, Daniel and Elisabeth Coutts (1993) "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management*, 12 (1).

Gârleanu, Nicolae and Lasse Heje Pederson (2016) "Dynamic portfolio choice with frictions," *Journal of Economic Theory*, 165, 487–516.

Gayle, Wayne-Roy (2017) "CCP Estimation of Dynamic Discrete/Continuous Choice Models with Generalized Finite Dependence and Correlated Unobserved Heterogeneity," August, Working Paper.

Heckman, James J. (1979) "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1), 153–161.

Hewitt, Julie A. and W. Michael Hanemann (1995) "A Discrete/Continuous Choice Approach to Residential Water Demand under Block Rate Pricing," *Land Economics*, 71 (2), 173–192.

Hong, Han and Matthew Shum (2010) "Pairwise-Difference Estimation of a Dynamic Optimization Model," *The Review of Economic Studies*, 77 (1), 273–304.

Hotz, V. Joseph and Robert A. Miller (1993) "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60 (3), 497–529.

Hotz, V. Joseph, Robert A. Miller, Seth Sanders, and Jeffrey Smith (1994) "A Simulation Estimator for Dynamic Models of Discrete Choice," *The Review of Economic Studies*, 61 (2), 265–289.

Iskhakov, Fedor, Thomas H. Jørgensen, John Rust, and Bertel Schjerning (2017) "The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks," *Quantitative Economics*, 8 (2), 317–365.

Kalouptsidi, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues (2020) "Linear IV regression estimators for structural dynamic discrete choice models," *forthcoming, Journal of Econometrics.*

Kangas, Olli, Signe Jauhiainen, Miska Simanainen, and Minna Ylikännö (2019) "The Basic Income Experiment 2017–2018 in Finland. Preliminary results,"Technical Report 9, Reports and Memorandums of the Ministry of Social Affairs and Health.

Komarova, Tatiana, Fabio Sanches, Daniel Silva Junior, and Sorawoot Srisuma (2018) "Joint analysis of the discount factor and payoff parameters in dynamic discrete choice models," *Quantitative Economics*, 9 (3), 1153–1194.

Levy, Matthew and Pasquale Schiraldi (2020) "Identification of Dynamic Discrete Choice Models with Time-Varying Choice Sets," February, Working Paper.

Liu, Hong (2004) "Optimal Consumption and Investment with Transaction Costs and Multiple Risky Assets," *The Journal of Finance*, 59 (1), 289–338.

Lynch, Anthony W. and Sinan Tan (2011) "Explaining the Magnitude of Liquidity Premia: The Roles of Return Predictability, Wealth Shocks, and State-Dependent Transaction Costs," *The Journal of Finance*, 66 (4), 1329–1368.

Magnac, Thierry and David Thesmar (2002) "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70 (2), 801–816.

Matzkin, Rosa L. (2003) "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71, 1339–1375.

McFadden, Daniel (1974) "Conditional logit analysis of qualitative choice analysis.," in Zarembka, P. ed. *Frontiers in Econometrics*, 105–142: Academic Press, New York.

Mogstad, Magne and Melissa Kearney (2019) "Universal Basic Income (UBI) as a Policy Response to Current Challenges," Aspen Institute Brief, November.

Mulligan, Casey B. (2004) "What Do Aggregate Consumption Euler Equations Say about the Capital-Income Tax Burden?" *The American Economic Review*, 94 (2), 166–170.

Rust, John (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), 999–1033.

——— (1994) "Structural estimation of markov decision processes," in Engle, R. F. and D. McFadden eds. *Handbook of Econometrics*, 4 of Handbook of Econometrics, Chap. 51, 3081–3143: Elsevier.

Schiraldi, Pasquale ⓡ Matthew Levy ⓡ Philippe Bracke (2019) "Heterogeneous time preferences: Evidence from the UK Mortgage Market," August, Working Paper.

Srisuma, Sorawoot (2015) "Identification in Discrete Markov Decision Models," *Econometric Theory*, 31 (3), 521–538.

Appendix A

A.1 Envelope Condition

In this section, we demonstrate how the envelope condition on the agent's continuous choice extends to the discrete-continuous model.

As in the continuous-only model, we begin by noting that the first-order condition of equation (3) must hold for any alternative chosen with positive probability:

$$\partial u_j(q, s_t, \zeta_t)/\partial q + \delta E\left[\frac{\partial \bar{V}(s_{t+1}, L_{t+1}, \nu_{t+1}, \zeta_{t+1})}{\partial L_{t+1}}|j_t, q, \bar{s}_t\right]\frac{\partial L_{t+1}}{\partial q} = 0 \tag{24}$$

where we apply the Dominated Convergence Theorem to reverse the order of the expectation and derivative operators in the second term.

Next, define $q_j(\bar{s}_t, \zeta_t)$ as the maximizer of the conditional value function given in (3), and also define:

$$\hat{v}_j(q, \bar{s}_t, \zeta_t) = u_j(q, s_t, \zeta_t) + \delta E\left[\bar{V}(\bar{s}_{t+1}, \zeta_{t+1})|j_t, q, \bar{s}_t\right] \tag{25}$$

Note that the following hold:

$$\partial \hat{v}_j/\partial q = \partial u_j/\partial q + \delta E\left[\frac{\partial \bar{V}(s_{t+1}, L_{t+1}, \nu_{t+1}, \zeta_{t+1})}{\partial L_{t+1}}|j_t, q, \bar{s}_t\right]\frac{\partial L_{t+1}}{\partial q} \tag{26}$$

and

$$\partial \hat{v}_j/\partial L_t = \delta E\left[\frac{\partial \bar{V}(s_{t+1}, L_{t+1}, \nu_{t+1}, \zeta_{t+1})}{\partial L_{t+1}}|j_t, q, \bar{s}_t\right]\frac{\partial L_{t+1}}{\partial L_t} \tag{27}$$

Finally, because $v_j(\bar{s}_t, \zeta_t) \equiv \hat{v}_j(q(\bar{s}_t, \zeta_t), \bar{s}_t, \zeta_t)$, we have:

$$\begin{aligned}
\frac{\partial v_j(s_t, L_t, \nu_t, \zeta_t)}{\partial L_t} &= \frac{d\hat{v}_j(q(\bar{s}_t, \zeta_t), \bar{s}_t, \zeta_t)}{dL_t} \\
&= \frac{\partial \hat{v}_j}{\partial q}\frac{\partial q}{\partial L_t} + \frac{\partial \hat{v}_j}{\partial L_t} \\
&= \delta E\left[\frac{\partial \bar{V}(s_{t+1}, L_{t+1}, \nu_{t+1}, \zeta_{t+1})}{\partial L_t}|j_t, q, \bar{s}_t\right]\frac{\partial L_{t+1}}{\partial L_t} = \frac{\partial u_j}{\partial q}
\end{aligned} \tag{28}$$

where the third equality follows since $\frac{\partial \hat{v}_j}{\partial q}\frac{\partial q}{\partial L} = 0$ by optimality of $q(\bar{s}, \zeta)$ and the fourth equality follows since $\partial L_{t+1}/\partial L_t = -\partial L_{t+1}/\partial q$.

A.2 Proof of Theorem 1

We prove pointwise identification. For notational convenience, we suppress z_t, we consider the additive separability under Assumption 8′,[33] and finally we use $f_j(s_{t+1}, s, L, q) = f_j(\cdot, L)$. Given q and s there exists a locus of (L, ζ) pairs such that $q_0^*(\bar{s}, \zeta) = q$. Choose an arbitrary wealth level $L^0(s)$, and consider $\zeta^0(s)$ such that $q = q_0^*(s, L^0(s), \zeta^0(s))$.

From equation (11) we have:

$$\frac{\partial u_0(q,s)}{\partial q} + \zeta^0(s) = \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) f_0(\cdot, L^0(s)) \mathcal{V}(s_{t+1}, B_0(s, L^0(s), s_{t+1})) \pi(s_{t+1}|s, 0, L^0(s), q) \quad (29)$$

where $\mathcal{V}(s_{t+1}, B) = \mathbb{E}_{\zeta_{t+1}}\left[\frac{\partial \Phi_0(s_{t+1},B,\zeta_{t+1})}{\partial B} + \frac{\partial u_0(q^*_{t+1},s_{t+1},B,\zeta_{t+1})}{\partial q}\right]$ is the expected marginal continuation value, and $B_0(s, L, s_{t+1}) = B(s, s_{t+1}, L, 0, q)$ is next-period wealth conditional on a realization of s_{t+1} given the period-t choices.

Next, consider an arbitrary $L^j(s)$. By assumption 8′ there exists some new $\zeta^j(s)$ such that $q_0^*(s, L^j(s), \zeta^j(s)) = q$ as before. For an alternative $j \neq 0$, in general $q_j^*(s, L^j(s), \zeta^j(s)) = q^j(s) \neq q$. Equation (11) implies, however:

$$\begin{aligned}\frac{\partial u_0(q,s)}{\partial q} + \zeta^j(s) + \frac{\partial \Psi_j(s, L^j(s), \zeta^j(s))}{\partial L} = \quad & \quad (30)\\ \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) f_0(\cdot, L^j(s)) \mathcal{V}(s_{t+1}, B_j(s, L^j(s), s_{t+1})) \pi(s_{t+1}|s, j, L^j(s), q^j(s)) &\end{aligned}$$

where $B_j(s, L^j(s), s_{t+1}) = B(s, s_{t+1}, L^j(s), j, q^j(s))$ similarly to B_0. Generically these will differ from the wealth levels induced by equation (29).

Subtracting equation (29) from (30) yields:

$$\begin{aligned}\frac{\partial \Psi_j(s, L^j(s), \zeta^j(s))}{\partial L} + \zeta^j(s) - \zeta^0(s) = \quad & \quad (31)\\ \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) \Big[f_j(\cdot, L^j(s)) \mathcal{V}(s_{t+1}, B_j(s, L^j(s), s_{t+1})) \pi(s_{t+1}|s, j, L^j(s), q^j(s)) & \\ - f_0(\cdot, L^0(s)) \mathcal{V}(s_{t+1}, B_0(s, L^0(s), s_{t+1})) \pi(s_{t+1}|s, 0, L^0(s), q) \Big] &\end{aligned}$$

Note that equation (31) has $2S$ unknowns: $\delta(s_{t+1})\mathcal{V}(s_{t+1}, B_k(s_{t+1}))$ evaluated at each of S

[33] Note the general case will follow by substituting $T^{-1}(T(\zeta^j(s_n) - T(\zeta^0(s_n)))$ in equation (31) and in the definition of $\nabla\boldsymbol{\Psi}$ in equations (36) and (38).

possible states and for $k \in \{0, j\}$.

Next, choose some other state s'. By assumption 6, the marginal propensity to consume out of wealth is less than one at sufficiently high wealth levels; i.e. $\partial q_0/\partial L < 1$. Given $L(s)$ and $\zeta^0(s)$, for any arbitrarily chosen $\zeta^0(s')$, there therefore exists some $L^0(s')$ which equalizes end-of-period wealth when choosing alternative 0, i.e.:

$$L^0(s') - q_0^*(s', L^0(s'), \zeta^0(s') - \phi_0(s') = L^0(s) - q_0^*(s, L^0(s), \zeta^0(s)) - \phi_0(s) \tag{32}$$

The end-of-period wealth is thus the same when choosing alternative 0 in state $(s, L^0(s), \zeta^0(s))$ as when choosing it in state $(s', L^0(s'), \zeta^0(s'))$, and thus so too will be the support of start-of-period-$(t+1)$ wealth. That is, $B_0(s', L^0(s'), s_{t+1}) = B_0(s, L^0(s), s_{t+1})$ for all s_{t+1}. The transition probabilities may of course differ. So too may the continuous choice, which we denote by $q(s') \equiv q_0^*(s', L^0(s'), \zeta^0(s'))$. Equation (11) then implies:

$$\frac{\partial u_0(q(s'), s')}{\partial q} + \zeta^0(s') = \sum_{s_{t+1} \in \mathcal{S}} \delta(s_{t+1}) f_0(\cdot, L^0(s')) \mathcal{V}(s_{t+1}, B_0(s, L^0(s), s_{t+1})) \pi(s_{t+1}|s', 0, L^0(s'), q(s')) \tag{33}$$

By the same argument as above, for any arbitrary $\zeta^j(s')$ there exists a $L^j(s')$ that induces the same end-of-period wealth when choosing j in state $(s', L^j(s'), \zeta^j(s'))$ as in state $(s, L^j(s), \zeta^j(s))$. That is, $B_j(s', L^j(s'), s_{t+1}) = B_j(s, L^j(s), s_{t+1})$ for all s_{t+1}. Furthermore, by Assumption 8′, there exists a $\zeta_j(s')$ such that $q_0^*(s', L^j(s'), \zeta^j(s') = q_0^*(s', L^0(s'), \zeta(s')) \equiv q^0(s')$. We thus obtain:

$$\begin{aligned} \frac{\partial u_0(q(s'), s')}{\partial q} + \zeta^j(s') + \frac{\partial \Psi_j(s', L^j(s'), \zeta^j(s'))}{\partial L} &= \\ \sum_{s_{t+1} \in \mathcal{S}} \delta(s_{t+1}) f_j(\cdot, L^j(s')) \mathcal{V}(s_{t+1}, B_j(s, L^j(s), s_{t+1})) \pi(s_{t+1}|s', j, L^j(s'), q^j(s')) \end{aligned} \tag{34}$$

where $q^j(s') \equiv q_j^*(s', L^j(s'), \zeta^j(s')$. Differencing (33) and (34) yields:

$$\begin{aligned} \frac{\partial \Psi_j(s', L^j(s'), \zeta^j(s'))}{\partial L} + \zeta^j(s') - \zeta^0(s') &= \\ \sum_{s_{t+1} \in \mathcal{S}} \delta(s_{t+1}) \big[f_j(\cdot, L^j(s')) \mathcal{V}(s_{t+1}, B_j(s, L^j(s), s_{t+1})) \pi(s_{t+1}|s', j, L^j(s'), q^j(s')) \\ - f_0(\cdot, L^0(s')) \mathcal{V}(s_{t+1}, B_0(s, L^0(s), s_{t+1})) \pi(s_{t+1}|s', 0, L^0(s'), q^0(s')) \big] \end{aligned} \tag{35}$$

Note that (35) introduces no new unknowns relative to equation (31): as the end-of-period wealth is the same as before, the support of the distribution of next-period wealth will be the same, though the probabilities will differ as the current state is different. We may repeat this process at all $s \in \mathcal{S}$ to generate a system of S linear equations in $2S$ unknowns. WLOG, we assume that $s = s_1$, and let $L^0 = L^0(s)$. Let $\boldsymbol{\nabla\Psi}(L^0)$ be the $S \times 1$ vector whose n^{th} element is $\partial\Psi_j(s_n, L^j(s_n), \zeta^j(s_n))/\partial L + \zeta^j(s_n) - \zeta^0(s_n)$ for each of the S states. Let $\boldsymbol{\pi}_k(L^0)$ be the $S \times S$ matrix whose $(n,m)^{th}$ element is the product $f_k(s_m, s_n, L^k(s_n), q^k(s_n))\pi(s_m|s_n, k, L^k(s_n), q^k(s_n))$. Finally, let $\boldsymbol{W}_k(L^0)$ be the $S \times 1$ vector whose n^{th} element is $\delta(s_n)\mathcal{V}(s_n, B_k(s_1, L^k(s_1), s_n))$. We thus have:

$$\boldsymbol{\nabla\Psi}(L^0) = \begin{bmatrix} \boldsymbol{\pi}_j(L^0) & -\boldsymbol{\pi}_0(L^0) \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_j(L^0) \\ \boldsymbol{W}_0(L^0) \end{bmatrix} \tag{36}$$

Now consider some new wealth level $\tilde{L}^0$. Similar to equation (32), there exists some $\tilde{\zeta}^0(s)$ such that:

$$\tilde{L}^0(s) - q_0^*(s, \tilde{L}^0(s), \tilde{\zeta}^0(s)) = L^0(s) - q_0^*(s, L^0(s), \zeta^0(s)) \tag{37}$$

Since the end-of-period wealth induced by choosing alternative 0 at this wealth level and marginal utility shock is the same as that induced at the original level choice in (29), so too will be the continuation wealth level state-by-state. We may therefore repeat the steps leading to equation (36) starting from $\tilde{L}^0$ rather than L^0, to generate an additional S linear equations in the same unknowns as before. We may add these to (45) to obtain:

$$\begin{bmatrix} \boldsymbol{\nabla\Psi}(L^0) \\ \boldsymbol{\nabla\Psi}(\tilde{L}^0) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi}_j(L^0) & -\boldsymbol{\pi}_0(L^0) \\ \boldsymbol{\pi}_j(\tilde{L}^0) & -\boldsymbol{\pi}_0(\tilde{L}^0) \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_j(L^0) \\ \boldsymbol{W}_0(L^0) \end{bmatrix} \tag{38}$$

Note that in the first matrix on the right-hand side of (38), $\boldsymbol{\pi}_j$ combines the state transition probabilities with the state-dependent returns f_j. Assumption 7 guarantees that we can choose $\tilde{L}^0$ such that it will be full-rank. We therefore recover $\delta(s_{t+1})\mathcal{V}(s_{t+1}, B_0(s_{t+1}))$, and indeed also $\delta(s_{t+1})\mathcal{V}(s_{t+1}, B_j(s_{t+1}))$ by inverting this linear system. Returning to (29), however, the marginal utility $\partial u_0(q,s)/\partial q$ for any state s is also identified as it is now written in terms of these parameters multiplied by data. Because q was arbitrary, the function $\partial u_0(\cdot,\cdot,s,\cdot)/\partial q$ is pointwise identified.

Once the marginal utility $\partial u_0/\partial q$ is known, $\mathcal{V}(s,B)$ may be recovered on its own by evaluating $E\left[\partial\Phi_0/\partial B + \partial u_0/\partial q\right]$. Note that these may be recovered at any wealth states, including those induced by choosing alternative j. The system of equations in (36) may therefore be written as a linear system with S unknowns only: the values of $\delta(s_{t+1})$. If there are any states not in the support of $\pi(s_{t+1}|s,0,L,q)$, their values may be obtained by repeating this process at some initial state where the transition probability is strictly positive. Thus $\delta(s)$ is identified.

A.3 Proof of Theorem 2

For notational convenience, we suppress z_t and set $f_j(s_{t+1},s,L,q) = f_j(\cdot,L)$. Consider an arbitrary (q,s,L,ζ). Under Assumption 6, $q_0^*(s,L,\zeta,\nu_0)$ is strictly monotone and unbounded in L and, given that $\phi_0(s,L,\nu_0)$ is strictly monotone in ν_0, it will be strictly monotone in ν_0 as well. There therefore exists ν_0 such that $q_0^*(s,\zeta,\nu_0) = q$.[34]

From equation (11) we have:

$$\frac{\partial u_0(q,s,\zeta)}{\partial q} = \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) f_0(\cdot,L)\mathcal{V}(s_{t+1},B_0(s_{t+1}))\pi(s_{t+1}|s,0,L,q) \tag{39}$$

where $\mathcal{V}(s_{t+1},B_0) = \mathbb{E}_{\zeta_{t+1},\nu_{t+1}}\left[\frac{\partial\Phi_0(s_{t+1},B_0,\zeta_{t+1},\nu_{t+1})}{\partial B} + \frac{\partial u_0(q_{t+1}^*,s_{t+1},\zeta_{t+1})}{\partial q}\right]$[35] and $B_0(s_{t+1}) = B(s,s_{t+1},L,\nu,0,q)$ suppressing arguments for notational ease.

Similarly, for some alternative j we have:

$$\frac{\partial u_0(q,s,\zeta)}{\partial q} + \frac{\partial\Psi_j}{\partial L} = \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) f_j(\cdot,L)\mathcal{V}(s_{t+1},B_j(s_{t+1}))\pi(s_{t+1}|s,j,L,q_j^*) \tag{40}$$

where $B_j(s_{t+1}) = B(s,s_{t+1},L,\nu,j,q_j^*(s,L,\zeta,\nu_j))$ as above. Generically these will differ from the wealth levels induced by equation (39).

Subtracting equation (39) from (40) yields:

$$\begin{aligned}\frac{\partial\Psi_j(L,s,\zeta,\nu)}{\partial L} = \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) \big[& f_j(\cdot,L)\mathcal{V}(s_{t+1},B_j(s_{t+1}))\pi(s_{t+1}|s,j,L,q_j^*) \\ & -f_0(\cdot,L)\mathcal{V}(s_{t+1},B_0(s_{t+1}))\pi(s_{t+1}|s,0,L,q)\big]\end{aligned} \tag{41}$$

[34] Strict monotonicity is not sufficient to imply that there is such a ν_0. If there is not, however, then this implies that the decision-maker will never consume q_0 in state (s,L,ζ), and therefore this particular q is irrelevant both in terms of observed behavior and in terms of calculating the value function. We therefore ignore such values, and assume that ν_0 does exist.

[35] We use the modified envelope to al

Note that equation (41) has $2S$ unknowns: $\delta(s_{t+1})\mathcal{V}(s_{t+1}, B_k(s_{t+1}))$ evaluated at each of S possible states and for $k \in \{0, j\}$.

Next, choose some state $\tilde{s} \neq s$. By assumption 9, there exists $\tilde{\nu}_0$ such that $q_0^*(\tilde{s}, L, \zeta, \tilde{\nu}_0) + \phi_0(\tilde{s}, L, \tilde{\nu}_0) = q + \phi_0(s, L, \nu_0)$.[36] Thus the end-of-period wealth is the same as in equation (39) and therefore so too is the support of period-$(t+1)$ wealth. The period-t continuous choice will generically differ, however, and we denote its new value as $\tilde{q}_0$:

$$\frac{\partial u_0(\tilde{q}_0, \tilde{s}, \zeta)}{\partial q} = \sum_{s_{t+1} \in \mathcal{S}} \delta(s_{t+1}) f_0(\cdot, L)\mathcal{V}(s_{t+1}, B_0(s_{t+1}))\pi(s_{t+1}|s, 0, L, \tilde{q}_0) \tag{42}$$

We next find values to match both the continuous choice and end-of-period wealth when choosing j. By the same argument as above, there exists a $\tilde{\nu}_j$ such that $q_j^*(\tilde{s}, \tilde{L}, \tilde{\zeta}, \tilde{\nu}_j) + \phi_j(\tilde{s}, \tilde{L}, \tilde{\nu}_j) = q_j^*(s, L, \zeta, \nu_j) - \phi_j(s, L, \nu_j)$ for any given $\tilde{L}$. The induced end-of-period wealth is therefore the same when choosing alternative j in state (s, L, ζ, ν) and in state $(\tilde{s}, \tilde{L}, \zeta, \tilde{\nu})$. Moreover, we choose $\tilde{L}_j(L)$ to guarantee that $q_0^*(\tilde{s}, \tilde{L}_j(L), \zeta, \tilde{\nu}_0) = q_0^*(\tilde{s}, L, \zeta, \tilde{\nu}_0)$ We therefore have:

$$\frac{\partial u_0(\tilde{q}_0, \tilde{s}, \zeta)}{\partial \tilde{q}_0} + \frac{\partial \Psi_j}{\partial L} = \sum_{s_{t+1} \in \mathcal{S}} \delta(s_{t+1}) f_j(\cdot, \tilde{L}_j)\mathcal{V}(s_{t+1}, B_j(s_{t+1}))\pi(s_{t+1}|\tilde{s}, j, \tilde{L}_j, \tilde{q}_j) \tag{43}$$

and differencing (42) and (43) yields:

$$\begin{aligned}\frac{\partial \Psi_j(\tilde{L}_j, \tilde{s}, \zeta, \tilde{\nu}_j)}{\partial L} = \sum_{s_{t+1} \in \mathcal{S}} \delta(s_{t+1}) \Big[& f_j(\cdot, \tilde{L}_j)\mathcal{V}(s_{t+1}, B_j(s_{t+1}))\pi(s_{t+1}|\tilde{s}, j, \tilde{L}_j, \tilde{q}_j) \\ & - f_0(\cdot, L)\mathcal{V}(s_{t+1}, B_0(s_{t+1}))\pi(s_{t+1}|\tilde{s}, 0, L, \tilde{q}_0)\Big]\end{aligned} \tag{44}$$

Note that (44) introduces no new unknowns relative to equation (41): as the end-of-period wealth is the same as before, the support of the distribution of next-period wealth will be the same, though the probabilities will differ as the current state is different. We may repeat this

[36] To see this, note that $B(s_t, s_{t+1}, L, \nu, j, q_j^*)$ is unbounded above as either $L_t \to \infty$ or $\nu_t \to -\infty$. If not, then there exists a $\bar{B}$ as an upper bound. If the agent has at most $\bar{B}$ resources, then conditional on any j_{t+1}, the expected quantity $q_{j,t+1}$ is finite, and therefore $E[u'(q_j)] > 0$. With a finite set of alternatives, the minimum expected marginal utility also exists and is larger than zero. Since the agent could spend any additional resources on consumption in period $t+1$, the optimal consumption path must yield at least this utility and therefore $\partial V/\partial L$ is also strictly positive. If $\bar{B}$ is an upper bound, however, then as $L_t \to \infty$ it must be that $q_t \to \infty$ as well. By assumption (6), this means the marginal utility of consumption in period t goes to zero, which leads to a contradiction of the Euler equation. The same contradiction arises when $\nu \to -\infty$. Because f_j is finite, the end-of-period wealth must be unbounded as a function of ν.

process at all $s \in \mathcal{S}$ to generate a system of S linear equations in $2S$ unknowns. Let $\boldsymbol{\nabla\Psi}(L)$ be the $S \times 1$ vector whose n^{th} element is $\partial \Psi_j(L, s_n, \zeta_n, \nu_n)/\partial L$, where ζ_n and ν_n are chosen as in equation (44). Let $\boldsymbol{\pi}_k(L)$ be the $S \times S$ matrix whose $(n,m)^{th}$ element is the product $f_k(s_m, s_n, \tilde{L}_k(s_n), q_k(s_n))\pi(s_m|s_n, k, \tilde{L}_k(s_n), q_k(s_n))$ — noting that $\tilde{L}_0(s) = L$ for all s. Finally, let $\boldsymbol{W}_k(L)$ be the $S \times 1$ vector whose n^{th} element is $\delta(s_n)\mathcal{V}(s_n, B_k(s_n))$. We thus have:

$$\boldsymbol{\nabla\Psi}(L) = \begin{bmatrix} \boldsymbol{\pi}_j(L) & -\boldsymbol{\pi}_0(L) \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_j(L) \\ \boldsymbol{W}_0(L) \end{bmatrix} \tag{45}$$

Notice that our starting point in Equation (39) is an arbitrary level of L (and ζ). We can therefore consider a new starting level L' to construct a similar equation to (45) without introducing any new unknowns by considering a value ζ' which equates end-of-period wealth in state (s, L', ζ') and (s, L, ζ) conditional on choosing alternative 0, and following the same construction for alternative j or states $\tilde{s}$. Thus, we may stack these new equations to (45) to obtain:

$$\begin{bmatrix} \boldsymbol{\nabla\Psi}(L) \\ \boldsymbol{\nabla\Psi}(L') \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi}_j(L) & -\boldsymbol{\pi}_0(L) \\ \boldsymbol{\pi}_j(L') & -\boldsymbol{\pi}_0(L') \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_j(L) \\ \boldsymbol{W}_0(L) \end{bmatrix} \tag{46}$$

Note that in the first matrix on the right-hand side of (46), $\boldsymbol{\pi}_j$ combines the state transition probabilities with the state-dependent returns f_j. Assumption 7 guarantees that we can choose L' such that it will be full-rank. We can therefore recover $\delta(s)$ and $\partial u_j(\cdot, \cdot, s, \cdot)/\partial q$ for every j pointwise following the same argument as the proof of Theorem 1.

A.4 Proof of Theorem 3

For notational convenience, we suppress z_t. Furthermore, as this Theorem addresses limited dependence of the state transitions on the decision-maker's choices we highlight this by suppressing all dependence of f and π on q and L. We begin by assuming that 9 is satisfied. Consider an arbitrary (q, s, L, ζ). Under Assumption 6, $q_0^*(s, L, \zeta, \nu_0)$ is strictly monotone and unbounded in L and, given that $\phi_0(s, L, \nu)$ is strictly monotone in ν_0, it will be strictly monotone in ν_0 as well. There therefore exists ν_0 such that $q_0^*(s, L, \zeta, \nu_0) = q$. Finally, assume WLOG that assumption $7'$ relates to alternatives 0 and j.

From equation (11) we have:

$$\frac{\partial u_0(q,s,\zeta)}{\partial q} = \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) f_0(s_{t+1}) \mathcal{V}(s_{t+1}, B_0(s_{t+1}))\pi(s_{t+1}|s,0) \tag{47}$$

where $\mathcal{V}(s_{t+1}, B) = \mathbb{E}_{\zeta_{t+1},\nu_{t+1}}\left[\frac{\partial \Phi_0(s_{t+1},B,\zeta_{t+1})}{\partial B} + \frac{\partial u_0(q^*_{t+1},s_{t+1},B,\zeta_{t+1})}{\partial q}\right]$ and $B_0(s_{t+1}) = B(s, s_{t+1}, L, \nu, 0, q)$ suppressing arguments for notational ease.

Similarly, for some alternative j we have:

$$\frac{\partial u_0(q,s,\zeta)}{\partial q} + \frac{\partial \Psi_j}{\partial L} = \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) f_j(s_{t+1}) \mathcal{V}(s_{t+1}, B_j(s_{t+1}))\pi(s_{t+1}|s,j) \tag{48}$$

where $B_j(s_{t+1}) = B(s, s_{t+1}, L, \nu, j, q^*_j(s, L, \zeta, \nu_j))$ as above. Generically these will differ from the wealth levels induced by equation (47). However, under assumption 9 there exists a ν_j such that:

$$q^*_j(s, L, \nu_j) = \alpha^{-1}\left[(\alpha-1)L + \alpha q^*_0(s, L, \nu_0) + \alpha\phi_0(s, L, \nu_0) - \phi_j(s, L, \nu_j)\right] \tag{49}$$

Equation (49) guarantees that the end-of-period wealth induced by choosing j is a proportion $1/\alpha$ of that induced by choosing alternative 0, and therefore state-by-state the start-of-period-$(t+1)$ wealth will also coincide.

Subtracting equation (47) from (40) therefore yields:

$$\frac{\partial \Psi_j(L,s,\zeta,\nu)}{\partial L} = \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1})\left[f_0(s_{t+1})(\alpha-1)\mathcal{V}(s_{t+1}, B_j(s_{t+1}))\pi(s_{t+1}|s,j)\right] \tag{50}$$

Note that equation (50) has S unknowns: $\delta(s_{t+1})\mathcal{V}(s_{t+1}, B_k(s_{t+1}))$ evaluated at each of S possible states and for $k \in \{0, j\}$.

Next, choose some state $\tilde{s} \neq s$. By assumption 9, there exists $\tilde{\nu}_0$ such that $q^*_0(\tilde{s}, L, \zeta, \tilde{\nu}_0) + \phi_0(\tilde{s}, L, \tilde{\nu}_0) = q + \phi_0(s, L, \nu_0)$. The period-t continuous will generically differ, however, and we denote its new value as $\tilde{q}_0$:

$$\frac{\partial u_0(\tilde{q}_0,\tilde{s},\zeta)}{\partial q} = \sum_{s_{t+1}\in\mathcal{S}} \delta(s_{t+1}) f_0(s_{t+1}) \mathcal{V}(s_{t+1}, B_0(s_{t+1}))\pi(s_{t+1}|s,0,L,\tilde{q}_0) \tag{51}$$

By the same argument as above, there exists a $\tilde{\nu}_j$ that will induce a $1/\gamma$ fraction of the end-of-period wealth induced by choosing alternative 0. We may again difference the Euler equations as in (50) to yield:

$$\frac{\partial \Psi_j(L, \tilde{s}, \zeta, \tilde{\nu})}{\partial L} = \sum_{s_{t+1} \in \mathcal{S}} \delta(s_{t+1}) \left[f_j(s_{t+1})(\alpha - 1)\mathcal{V}(s_{t+1}, B_j(s_{t+1}))\pi(s_{t+1}|s, j) \right] \tag{52}$$

Note that (52) introduces no new unknowns relative to equation (50): as the end-of-period wealth is the same as before, the support of the distribution of next-period wealth will be the same, though the probabilities will differ as the current state is different. We may repeat this process at all $s \in \mathcal{S}$ to generate a system of S linear equations in S unknowns. Let $\boldsymbol{\nabla\Psi}(L)$ be the $S \times 1$ vector whose n^{th} element is $\partial \Psi_j(L, s_n, \zeta_n, \nu_n)/\partial L$, where ζ_n and ν_n are chosen as in equation (52). Let $\boldsymbol{\pi}(L)$ be the $S \times S$ matrix whose $(n,m)^{th}$ element is the product $f_0(s_m)(1-\alpha)\pi(s_m|s_n, 0)$. Finally, let $\boldsymbol{W}(L)$ be the $S \times 1$ vector whose n^{th} element is $\delta(s_n)\mathcal{V}(s_n, B_0(s_n))$. We thus have:

$$\boldsymbol{\nabla\Psi}(L) = \boldsymbol{\pi}(L)\boldsymbol{W}(L) \tag{53}$$

Equation (53) is a linear system of S equations in S unknowns. Moreover, unlike equations (36) and (45), the Markov matrix $\boldsymbol{\pi}(L)$ will generically be full rank and therefore the system may be inverted. We therefore recover $\delta(s_{t+1})\mathcal{V}(s_{t+1}, B_0(s_{t+1}))$, and the rest of the proof proceeds as before.

Finally, we note that when Assumption $8'$ is holds, equation (49) may be satisfied by an appropriate choice of L and ζ as in Theorem 1. The remainder of the proof proceeds in the same way, with the substitution of $\boldsymbol{\nabla\Psi}(L)_n = \partial \Psi_j(L_{j,n}, s_n, \zeta_{j,n})/\partial L + \zeta_{j,n} - \zeta_{0,n}$ to accommodate the use of ζ rather than ϕ.

A.5 Market-level Shocks

To we allow for the presence of a product-level aggregate shock, we assume:

Assumption 11. *The continuous shock ζ_{ijt} is additive in the market-level and individual components: $\zeta_{ijt} = \xi_{jt} + \zeta_{it}$*

We then have the following corollary:

Corollary 1. *Lemma 1 and assumption 11 guarantee that ζ_{it} and ξ_{jt} are identified.*

Proof of Corollary 1. Notice that $\zeta_{ijt} = \zeta_{it} + \xi_{jt}$ and $E[\tilde{\zeta}_t | d_t = j] = 0$ which will immediately implies the results. $\square$