The age-time-cohort problem and the identification of structural parameters in life-cycle models

SAM SCHULHOFER-WOHL
Federal Reserve Bank of Chicago

A standard approach to estimating structural parameters in life-cycle models imposes sufficient assumptions on the data to identify the "age profile" of outcomes, then chooses model parameters so that the model's age profile matches this empirical age profile. I show that this approach is both incorrect and unnecessary: incorrect, because it generally produces inconsistent estimators of the structural parameters, and unnecessary, because consistent estimators can be obtained under weaker assumptions. I derive an estimation method that avoids the problems of the standard approach. I illustrate the method's benefits analytically in a simple model of consumption inequality and numerically by reestimating the classic life-cycle consumption model of Gourinchas and Parker (2002).

KEYWORDS. Age-time-cohort identification problem, life-cycle models.

JEL CLASSIFICATION. C23, D91, J1.

1. INTRODUCTION

A well-known difficulty in investigating how economic choices change over the life cycle is that it is impossible to separately identify the effects of age, time, and birth cohort on the outcome of interest. This paper shows that a standard solution to this age-time-cohort identification problem will, in general, cause researchers to make incorrect inferences about the structural parameters of their economic models. I provide a simple alternative that allows accurate identification of the structural parameters, without having to first identify age, time, and cohort effects. The alternative method identifies structural parameters from the second and higher derivatives of the age effects; by comparison, the standard solution resorts to a normalization on the first derivative.

Consider an economic model that says outcome y depends on age a according to

$$y(a) = \xi_0 + q(a; \boldsymbol{\theta}^*), \tag{1}$$

where ξ_0 is an intercept, q is a known function, and $\boldsymbol{\theta}^*$ is a vector of structural parameters. A researcher who has data on the age profile $y(a)$ might estimate $\boldsymbol{\theta}^*$ by the vector of parameters that makes $q(a; \boldsymbol{\theta}^*)$ as close as possible to the observed age profile.

In the real world, outcomes y depend not only on age but also on other variables—in particular, time and birth cohort. For example, an investor's allocation to stocks may depend not only on her age but also on expected returns this year (time) and on whether she is averse to stocks because she grew up during the Great Depression (cohort). A researcher who wishes to confront a model of the form (1) with data therefore has two choices: enrich the model to describe time and cohort effects, or remove time and cohort effects from the data before confronting the model. In some applications, the theoretical source of time and cohort effects is clear, and it is straightforward to enrich the model to include them. But in other applications, a researcher may prefer a semistructural approach that models only age effects but not time and cohort effects—either to avoid reliance on assumptions about time and cohort effects that are not central to the issue being analyzed, or to make the model more tractable. A researcher who takes a semistructural approach will need to remove time and cohort effects from the data. This paper is concerned with how best to do so.

Let $y_{a,t}$ be the outcome for people who are age a at time t. One might hope to recover an age profile purged of time and cohort effects by regressing $y_{a,t}$ on a full set of age, time, and cohort dummy variables:

$$y_{a,t} = \xi_0 + \alpha_a + \beta_t + \gamma_c + u_{a,t}, \tag{2}$$

where $c = t - a$ is the birth cohort and $u_{a,t}$ is an unobservable error. (I impose throughout the innocuous normalization that $\sum_a q(a; \boldsymbol{\theta}) = \sum_a \alpha_a = \sum_t \beta_t = \sum_c \gamma_c = 0$.) The age coefficients α_a represent the age profile of y after controlling for period and cohort effects. However, the α_a's in (2) are not identified: If (2) holds, then for any real number k, so does

$$y_{a,t} = \xi_0 + (\alpha_a + ka - k\bar{a}) + (\beta_t - kt + k\bar{t}) + (\gamma_c + kc - k\bar{c}) + u_{a,t}, \tag{2'}$$

where $\bar{a}$, $\bar{t}$, and $\bar{c}$ are the means of a, t, and c.

The standard method for solving this identification problem is to impose a normalization on the age, period, or cohort effects to pin down k, so that the α_a's can be identified and the parameters $\boldsymbol{\theta}^*$ chosen to match them. But the estimator of $\boldsymbol{\theta}^*$ in the standard approach depends on the arbitrary normalization used to pin down k. If the normalization is incorrect, the estimator will be inconsistent. As a trivial example, suppose the model predicts that y increases with age if and only if a scalar parameter θ^* is positive. If the age effects estimated under the chosen normalization increase with age, it would be tempting to conclude that $\theta^* > 0$. But this conclusion would be incorrect. For k sufficiently negative, $\alpha_a + ka - k\bar{a}$ decreases with age, and if a restriction were chosen that corresponded to such a negative value of k, one would obtain age effect estimates that implied $\theta^* \leq 0$. Checking the estimator's robustness to a small number of possible normalizations, as is common in the literature, does not solve the problem for two reasons. First, even if the estimates do not vary much across the normalizations that are

tested, other normalizations might still have produced different results. Second, if one of the tested normalizations is correct and another is not, the estimates may vary across normalizations but the researcher will not know which one is correct.

The new method proposed in this paper exploits the fact that the age effects are identified up to a single constant k. The method obtains estimated age effects $\hat{\alpha}_a$ using any just-identified normalization on (2), then chooses k and $\boldsymbol{\theta}^*$ such that $\hat{\alpha}_a + ka - k\bar{a}$ is as close as possible to $q(a; \boldsymbol{\theta}^*)$. The logic is that the model should fit at least as well with a correct value for k as with an incorrect value. Thus, optimizing over k should produce a consistent estimator, a result formally proven below.

This paper's method amounts to removing a linear trend from both the model age profile $q(a, \boldsymbol{\theta}^*)$ and the empirical age profile $\hat{\alpha}_a$, then choosing $\boldsymbol{\theta}^*$ so that these de-trended age profiles match. Thus, this paper's method identifies $\boldsymbol{\theta}^*$ from the second and higher derivatives of the age profile, discarding all information about the first derivative. The method therefore uses strictly weaker assumptions than the standard method, which imposes a normalization on the first derivative. Hall (1968) shows that the second and higher derivatives of the age profile are identified even though the first derivative is not. McKenzie (2006) uses the second derivative to characterize the reduced-form relationship between a and y. The innovation here is that I show how to use the second and higher derivatives to identify structural parameters.

This procedure requires q to be sufficiently nonlinear, in a sense made precise below. The method therefore does not guarantee point identification of $\boldsymbol{\theta}^*$, but when q is nonlinear enough that a normalization on the first derivative of the age effects is not needed to identify $\boldsymbol{\theta}^*$, the method prevents this unneeded normalization from contaminating the estimates.

The paper proceeds as follows. Section 2 formally defines the new estimator, states conditions under which it identifies the structural parameters, and shows why the standard method generally fails to do so. Section 3 reviews literature using the standard method. Section 4 illustrates this paper's method analytically in a simple life-cycle model of consumption inequality, while Section 5 shows that this paper's method produces different empirical results in the life-cycle consumption model of Gourinchas and Parker (2002). Section 6 concludes.

2. The method

Assume that the observed ages run from 1 to A and that $\boldsymbol{\theta}^*$ is known to be in a set $\boldsymbol{\Theta}$. This paper's method for estimating $\boldsymbol{\theta}^*$ is:

1. Estimate the ordinary least squares regression (2), subject to $\sum_a \alpha_a = \sum_t \beta_t = \sum_c \gamma_c = 0$ and to any one additional linear restriction that identifies the parameters. The restriction does not matter so long as there is exactly one.

2. Let $\hat{\boldsymbol{\alpha}}$ be the vector of estimated age effects from step 1. Define the column vectors $\mathbf{a} = [1 - \bar{a}, \ldots, A - \bar{a}]'$ and $\mathbf{q}(\theta) = [q(1, \boldsymbol{\theta}), \ldots, q(A, \boldsymbol{\theta})]'$. Choose $\hat{\boldsymbol{\theta}}$ and $\hat{k}$ to solve

$$(\hat{\boldsymbol{\theta}}, \hat{k}) \in \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, k} \big[\mathbf{q}(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}} - k\mathbf{a}\big]' \mathbf{W}\big[\mathbf{q}(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}} - k\mathbf{a}\big], \tag{3}$$

where $\mathbf{W}$ is any $A \times A$ symmetric, positive definite weighting matrix.

The choice of identifying restriction in step 1 cannot affect the value of $\hat{\boldsymbol{\theta}}$ in step 2, because changing the restriction merely adds a linear trend to $\hat{\boldsymbol{\alpha}}$, which can be removed by changing the choice of k in (3). The standard method is identical to this paper's method but imposes $k = 0$ in (3). Thus, the identifying restriction in step 1 can affect the estimator under the standard method, and this paper's method relaxes the assumptions of the standard method.

2.1 Identification

Identification requires that age, time, and cohort effects are additively separable.

ASSUMPTION 1. *The observed data satisfy*

$$y_{a,t} = \xi_0^* + q\big(a; \boldsymbol{\theta}^*\big) + \beta_t^* + \gamma_c^* + u_{a,t} \tag{4}$$

for some intercept ξ_0^*, time effects β_t^*, cohort effects γ_c^*, and measurement errors $u_{a,t}$ satisfying

$$\mathrm{E}[u_{a,t}|a, t] = 0 \tag{5}$$

and the normalizations $\forall \boldsymbol{\theta} \sum_a q(a; \boldsymbol{\theta}) = 0$ and $\sum_t \beta_t^* = \sum_c \gamma_c^* = 0$.

Assumption 1 is a joint restriction on the sources of time and cohort effects, the functional form of the structural model, and the choice of variable y with which to estimate the model. At the level of generality considered here, I cannot provide primitive conditions under which this joint restriction holds. However, some examples are illustrative. As shown in Section 4, additively separable time and cohort effects can arise from measurement error that has a different distribution in different years or for different cohorts. Alternatively, if a model's objective function and constraint set are homothetic with respect to a shock, this shock will enter multiplicatively in policy functions, and logs of choice variables will satisfy Assumption 1. For example, a cake-eating problem satisfies this requirement if preferences are homothetic and the initial size of the cake varies across cohorts. But if the constraint set does not scale with the shock, then the shock will not be additively separable in logs of choice variables. For example, in a model with idiosyncratic income shocks and an exogenous borrowing limit, cohorts with different initial wealth will respond differently to identical income shocks at a given age, because they will be at different distances from the borrowing limit. However, even in such a model, some object other than the log of a choice variable might satisfy Assumption 1. Alternatively, Assumption 1 may be viewed as a first-order approximation to the way time and cohort effects enter the data when the model provides no guidance on this matter.

I assume the vector of measurement errors $\mathbf{u} = \{u_{a,t}\}$ is asymptotically normal, as will occur, for example, if the data $\mathbf{y} = \{y_{a,t}\}$ are moments of a random sample. I consider asymptotics in which the set of observed ages and dates is fixed but the sample size used to calculate these moments grows.

ASSUMPTION 2. $\sqrt{N}\mathbf{u} \xrightarrow{d} \mathrm{N}(0, \boldsymbol{\Sigma}_u)$ *as the sample size* $N \to \infty$.

I also assume that the model satisfies some standard regularity conditions.

ASSUMPTION 3. $\boldsymbol{\Theta}$ *is compact, and* $\mathbf{q}(\boldsymbol{\theta})$ *is continuous on* $\boldsymbol{\Theta}$.

Identification requires the structural model to be sufficiently nonlinear.

CONDITION NL. For all $\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \{\boldsymbol{\theta}^*\}$, there is no real number $\bar{k}$ such that $q(\boldsymbol{\theta}) - q(\boldsymbol{\theta}^*) = \bar{k}\mathbf{a}$.

Condition NL says there is no parameter vector $\boldsymbol{\theta}$ whose age profile $q(a; \boldsymbol{\theta})$ differs from the age profile under the true parameters by only a linear trend in age. If this condition failed, it would be impossible to identify the structural parameters from the age profile because the age profile itself is identified only up to an unknown linear trend. Because Condition NL is stated in terms of the unknown true parameters $\boldsymbol{\theta}^*$, it is not directly testable. Two testable conditions that imply Condition NL are:

- For all $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, there is no real number $\bar{k}$ such that $\mathbf{q}(\boldsymbol{\theta}_1) - \mathbf{q}(\boldsymbol{\theta}_2) = \bar{k}\mathbf{a}$.
- $\partial^3 q / \partial^2 a\, \partial\boldsymbol{\theta} \neq 0$ for some a and all $\boldsymbol{\theta}$.

We can now show that this paper's method identifies the structural parameters.

PROPOSITION 1. *Under Assumptions* 1, 2, *and* 3, *and Condition* NL, *in the limit as* N *goes to infinity, the solution* $\hat{\boldsymbol{\theta}}$ *to problem* (3) *converges in probability to* $\boldsymbol{\theta}^*$.

PROOF. I sketch the proof here and refer readers to the journal website http://qeconomics.org/supp/738/supplement.pdf for details. Let $\mathbf{R} = \mathbf{I} - \mathbf{a}(\mathbf{a}'\mathbf{Wa})^{-1}\mathbf{a}'\mathbf{W}$ be the matrix that produces residuals from projecting any vector of length A on $\mathbf{a}$ by generalized least squares (GLS) with weighting matrix $\mathbf{W}$. Let $\mathbf{M}$ be the first A rows of the Moore–Penrose pseudoinverse of the design matrix of the regression in step 1, so $\hat{\boldsymbol{\alpha}} = \mathbf{M}\mathbf{y}$. Minimizing out k in (3) shows that the solution to (3) is

$$\hat{\boldsymbol{\theta}} \in \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{RMy}\big]'\mathbf{W}\big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{RMy}\big], \tag{6a}$$

$$\hat{k} = c_1(\hat{\boldsymbol{\theta}}) - \hat{c}_2, \tag{6b}$$

where $c_1(\hat{\boldsymbol{\theta}})$ and $\hat{c}_2$ are, respectively, the slopes in GLS regressions of $q(a; \hat{\boldsymbol{\theta}})$ and $\hat{\alpha}_a$ on a. Equation (6a) expresses $\hat{\boldsymbol{\theta}}$ as a minimum distance estimator. We need only verify conditions for consistency of such estimators. Theorem 2.1 of Newey and McFadden (1994) shows that $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*$ if there is a function $Q_0(\boldsymbol{\theta})$ such that (i) Q_0 is uniquely minimized at $\boldsymbol{\theta}^*$, (ii) $\boldsymbol{\Theta}$ is compact, (iii) Q_0 is continuous, and (iv) the estimator's objective function converges uniformly in probability to Q_0. Define

$$Q_0(\boldsymbol{\theta}) = \big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*)\big]'\mathbf{W}\big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*)\big]. \tag{7}$$

Because $\mathbf{W}$ is positive definite, any minimizer of (7) satisfies $\mathbf{Rq}(\boldsymbol{\theta}) = \mathbf{Rq}(\boldsymbol{\theta}^*)$, that is, the residuals from projecting $\mathbf{q}(\boldsymbol{\theta})$ on $\mathbf{a}$ are the same as those from projecting $\mathbf{q}(\boldsymbol{\theta}^*)$ on $\mathbf{a}$. Condition NL then implies $\boldsymbol{\theta}^*$ is the unique minimizer, satisfying hypothesis (i). Hypotheses (ii) and (iii) hold by Assumption 3. Under Assumption 1, $\hat{\boldsymbol{\alpha}} = \mathbf{q}(\boldsymbol{\theta}^*) + k^*\mathbf{a} + \mathbf{Mu}$ for some k^* determined by the normalization in step 1. It follows that $\mathbf{RMy} = \mathbf{R\hat{\alpha}} = \mathbf{Rq}(\boldsymbol{\theta}^*) + \mathbf{RMu}$. Hence, under Assumptions 2 and 3, the objective function in (6a) converges uniformly in probability on $\boldsymbol{\Theta}$ to Q_0. Therefore, hypothesis (iv) also holds, and $\hat{\boldsymbol{\theta}} \overset{p}{\to} \boldsymbol{\theta}^*$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

2.2 Remarks

Inference about $\boldsymbol{\theta}^*$ If the conditions for identification of $\boldsymbol{\theta}^*$ hold, then the estimator of $\boldsymbol{\theta}^*$ is a standard minimum distance estimator and the usual inference techniques for such estimators apply, subject to appropriate regularity conditions such as differentiability of q.

Interpretation in terms of detrended age profiles   As equation (6a) shows, the new method chooses the structural parameters $\boldsymbol{\theta}^*$ so that the *detrended* age profile from the model matches, as closely as possible, the *detrended* age profile in the data.

Necessity of Condition NL   Condition NL is necessary for the asymptotic objective function (7) to have a unique minimum. However, even if Condition NL fails, the objective may have a unique minimum in finite samples. Therefore, Condition NL should be verified independently, without relying on finite sample behavior as a test of identification.

Age profiles of multiple variables   In many applications, researchers fit a model to the age profiles of two or more variables. In general, there is no reason to use the same normalization on the age, time, and cohort effects for all variables. Hence, a different slope k_j should be estimated for each variable j. (However, if theory suggests restrictions on the relationship between the slopes of different variables' age profiles, these restrictions could be imposed in estimation.) For example, suppose the model predicts age profiles of income i and consumption c:

$$i(a) = \xi_{0,i} + q_i(a; \boldsymbol{\theta}^*), \qquad c(a) = \xi_{0,c} + q_c(a; \boldsymbol{\theta}^*). \tag{8}$$

The structural parameters should be estimated as follows. First, estimate (2) separately for income and consumption, obtaining age profiles $\hat{\boldsymbol{\alpha}}_i$ and $\hat{\boldsymbol{\alpha}}_c$. Second, estimate $\boldsymbol{\theta}^*$ by solving

$$(\hat{\boldsymbol{\theta}}, \hat{k}_i, \hat{k}_c) \in \arg\min_{\boldsymbol{\theta}, k_i, k_c} \begin{bmatrix} \mathbf{q}_i(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}}_i - k_i\mathbf{a} \\ \mathbf{q}_c(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}}_c - k_c\mathbf{a} \end{bmatrix}' \mathbf{W} \begin{bmatrix} \mathbf{q}_i(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}}_i - k_i\mathbf{a} \\ \mathbf{q}_c(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}}_c - k_c\mathbf{a} \end{bmatrix}. \tag{9}$$

A special case: Estimation using second differences of age profiles   Let the weighting matrix be $\mathbf{W} = \tilde{\mathbf{W}}'\tilde{\mathbf{W}}$, where

$$\tilde{\mathbf{W}} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ & & & & \cdots & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}. \tag{10}$$

The objective function for this **W** is the sum of squared deviations between the second-differenced age profiles from the model and from the data. Because, as shown by Hall (1968), the second-differenced age profile from the data is invariant to the normalization used to estimate it, the objective function is invariant to the normalization and so is $\hat{\boldsymbol{\theta}}$.

Comparison with nonlinear least squares (NLS)   An alternative approach would be to estimate $\boldsymbol{\theta}^*$ and the period and cohort effects simultaneously by NLS on (4):

$$
(\hat{\boldsymbol{\theta}}, \acute{\xi}_0, \{\acute{\beta}_t\}, \{\acute{\gamma}_c\}) \in \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}, \xi_0, \{\beta_t\}, \{\gamma_c\}} \sum_{a,t} \big[y_{a,t} - \xi_0 - q(a; \boldsymbol{\theta}) - \beta_t - \gamma_c \big]^2
$$

$$
\text{s.t.} \quad \sum_t \beta_t = \sum_c \gamma_c = 0.
$$
(11)

Under Assumptions 1, 2, and 3, $(\boldsymbol{\theta}^*, \xi_0^*, \beta_t^*, \gamma_c^*)$ is one (asymptotic) solution to (11). I show in the online appendix that this is the unique asymptotic solution if and only if Condition NL holds. Thus, for both NLS and this paper's method, Condition NL is necessary and sufficient for the asymptotic objective function to identify $\boldsymbol{\theta}$. However, NLS is more computationally challenging. If there are P structural parameters and age profiles of L variables, this paper's method requires L linear regressions, followed by nonlinear optimization over $L + P$ parameters, while NLS requires nonlinear optimization over $(2T + A - 2)L + P$ parameters. For example, the quantitative exercise in Section 5 involves 40 ages, 14 time periods, four structural parameters, and three age profiles, so this paper's method requires optimization over seven parameters, while NLS would require optimization over 202 parameters. In addition, NLS requires raw data on $y_{a,t}$, while this paper's method requires only an estimated age profile $\hat{\boldsymbol{\alpha}}$, so this paper's method can be used whenever estimated age profiles have been published, without needing to reconstruct the raw data.

Partial identification of $\boldsymbol{\theta}^*$ If Condition NL fails, the parameter vector may still be partially identified. For example, suppose we can partition the parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ where $\mathbf{q}(\boldsymbol{\theta}) = \mathbf{q}_1(\boldsymbol{\theta}_1) + q_2(\boldsymbol{\theta}_2)\mathbf{a}$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and where $\mathbf{q}_1(\boldsymbol{\theta}_1)$ satisfies Condition NL. Then the same arguments as above show that the paper's method point identifies $\boldsymbol{\theta}_1^*$ using a minimum distance estimator for which standard inference techniques are available. In this sense, the paper's method may make it possible to learn something about $\boldsymbol{\theta}^*$ even if the full parameter vector is not point identified. If $\boldsymbol{\theta}^*$ cannot be partitioned in this way, estimation of and inference about identified subsets of the parameter space based on (6a) might still be possible, but methods such as those in Chernozhukov, Hong, and Tamer (2007) would be needed, and detailed investigation of the required assumptions is left for further research.

Incorrect results from the standard method   The standard method will generally produce incorrect results even when this paper's method produces correct results. Recall that the standard method is identical to this paper's method but imposes $k = 0$ in (3), and that for any normalization in step 1, we can find k^* such that $\hat{\boldsymbol{\alpha}} = \mathbf{q}(\boldsymbol{\theta}^*) + k^*\mathbf{a} + \mathbf{Mu}$. The standard method therefore estimates the structural parameters not by (6a) but by

$$
\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big[\mathbf{q}(\boldsymbol{\theta}) - \mathbf{q}(\boldsymbol{\theta}^*) - k^*\mathbf{a} - \mathbf{Mu}\big]' \mathbf{W} \big[\mathbf{q}(\boldsymbol{\theta}) - \mathbf{q}(\boldsymbol{\theta}^*) - k^*\mathbf{a} - \mathbf{Mu}\big].
$$
(12)

This objective function converges to one whose minimizer generally is not $\boldsymbol{\theta}^*$, unless either (i) $k^* = 0$ or (ii) $\mathbf{q}(\boldsymbol{\theta})$ is orthogonal to $\mathbf{a}$ for all $\boldsymbol{\theta}$. In case (i), the chosen normalization in step 1 is correct. In case (ii), the model describes a detrended age profile, so the choice of trend in the empirical age profile does not matter.

3. Research using the standard method

The literature employing the standard estimation method is large. This literature requires a normalization on the period or cohort effects. Two common normalizations are:

Cohort view: Secular trends appear only in cohort effects. Period effects are orthogonal to a time trend ($\sum_t \beta_t(t - \bar{t}) = 0$), are all zero ($\beta_t = 0$ for all t), or can be replaced with observables such as the unemployment rate that measure cyclical economic variation.

Period view: Secular trends appear only in period effects. Cohort effects are orthogonal to a time trend ($\sum_c \gamma_c(c - \bar{c}) = 0$) or are all zero ($\gamma_c = 0$ for all c).[1]

Some authors maintain one normalization throughout. Others investigate how their results depend on the choice between the cohort view and the period view.

A leading example, examined further in Section 5, is Gourinchas and Parker (2002). The authors model the mean consumption of households of age a as a function of the rate of time preference, coefficient of relative risk aversion, and other parameters. The model does not contain cohort or time effects. The authors estimate the empirical age profile of consumption by regressing log consumption y_{at} on age and cohort dummies and on the unemployment rate, which substitutes for time effects. Then the authors find the structural parameters that make the model's predicted age profile of consumption come as close as possible to the estimated coefficients on the age dummies in the first-stage regression.

Another example is Huggett, Ventura, and Yaron (2011), who model how learning and shocks to human capital produce inequality in earnings over the life cycle, as a function of parameters including the variance of shocks. The model does not contain cohort or time effects. The authors estimate moments of the earnings distribution in data on people who are age a in year t and regress these moments y_{at} on age, time, and cohort dummies. They produce two sets of age profiles $\hat{\alpha}_a$—one under the cohort view and one under the period view. Then they choose parameters to fit the model's predictions for the same moments of earnings as a function of age to each of the two sets of empirical age profiles.

Other papers employing this method include the studies of wealth accumulation by Cagetti (2003); of household investments by Wachter and Yogo (2010); of inequality in consumption, wages, and hours by Kaplan (2012); and of consumption over the life cycle by Aguiar and Hurst (2013). De Nardi, French, and Jones (2010) use a variant of the method to study health expenses and saving among the elderly; they model

---

[1]The all-zero normalization is overidentified because it imposes as many restrictions as there are periods or cohorts, not just the one restriction needed to identify the slope.

cohort effects structurally, as a function of lifetime income, but assume the time effects are all zero. Deaton and Paxson (1994a, 1994b), Ameriks and Zeldes (2004), and Heathcote, Storesletten, and Violante (2005) follow a similar but more qualitative procedure by comparing models' broad predictions to the observed relationship between y and a, after controlling for period and cohort effects and imposing a normalization to identify the age effects.

4. Analytic example: Consumption inequality over the life cycle

This section exhibits a simple analytic example in which the standard method does not identify the structural parameters of an economic model but this paper's method does. Agent i is born in year c with assets $x_{i,0,c} > 0$ and lives for $A + 1$ periods, receiving a stochastic income $y_{i,a,t}$ in each period. Income is independently and identically distributed across agents and dates with mean μ and variance σ^2. Let $C_{i,a,t}$ be i's consumption in year t, when he is age $a = t - c$. The agent's preferences are represented by

$$-\frac{1}{2}\mathrm{E}_c \sum_{a=0}^{A} \rho^a [\bar{C} - C_{i,a,t}]^2, \tag{13}$$

where ρ is the rate of time preference and $\bar{C}$ is a bliss level of consumption. The agent can borrow or save without limit at the gross interest rate $(1 + r) = \rho^{-1}$, but cannot borrow at age A. The agent maximizes (13) by choice of $\{C_{i,a,t}\}_{a=0}^{A}$, given $x_{i,0,c}$. The online appendix shows that the cross-sectional variance of consumption among agents in cohort c at age a is

$$\mathrm{Var}[c_{i,a,c+a}|a, c] = (1 + \phi_0)^{-2} \mathrm{Var}[x_{i,0,c}] + \sigma^2 \sum_{s=0}^{a}(1 + \phi_s)^{-2}, \quad \phi_a = \sum_{s=1}^{A-a} \rho^s. \tag{14}$$

Suppose that, as in Deaton and Paxson (1994a), an econometrician observes consumption in repeated cross sections of agents of various ages at various dates. Assume that observed consumption $\hat{C}_{i,a,t}$ is measured with an error that is independent of $C_{i,a,t}$, uncorrelated across agents, and has mean $\nu_{a,t}$ and variance η_t^2 at date t. (The bias $\nu_{a,t}$ and measurement error variance η_t^2 could change over time due to, e.g., changes in the survey instrument.) The econometrician can construct moments of consumption for each age and date. For simplicity, assume the sample is infinitely large so that sample moments equal population moments. The mean of observed consumption is uninformative because of the unknown bias $\nu_{a,t}$. The variance of observed consumption among people who are age a at date t is

$$\mathrm{Var}[\hat{C}_{i,a,t}|a, t] = \eta_t^2 + \mathrm{Var}[C_{i,a,t}|a, t] = \eta_t^2 + (1 + \phi_0)^{-2} \mathrm{Var}[x_{i,0,c}] + \sigma^2 \sum_{s=0}^{a}(1 + \phi_s)^{-2}. \tag{15}$$

Equation (15) is identical to (4) with $\boldsymbol{\theta}^* = (\sigma^2, \rho)$, $\xi_0^* = 0$, $q(a; \boldsymbol{\theta}^*) = \sigma^2 \sum_{s=0}^{a}(1 + \phi_s)^{-2}$, $\beta_t = \eta_t^2$, and $\gamma_c = (1 + \phi_0)^{-2} \mathrm{Var}[x_{i,0,c}]$. It follows that this paper's method identifies σ^2

and ρ as long as Condition NL holds, which in turn requires that the following equations have a unique solution $\hat{\sigma}^2 = \sigma^2$, $\hat{\rho} = \rho$, $k = 0$:

$$\sigma^2 \sum_{s=0}^{a} (1 + \phi_s)^{-2} = ka + \hat{\sigma}^2 \sum_{s=0}^{a} (1 + \hat{\phi}_s)^{-2}, \quad \phi_s = \sum_{s=1}^{A-a} \rho^s, \hat{\phi}_s = \sum_{s=1}^{A-a} \hat{\rho}^s, a = 0, \ldots, A. \quad (16)$$

The online appendix shows that $\hat{\sigma}^2 = \sigma^2$, $\hat{\rho} = \rho$, $k = 0$ is indeed the unique solution and hence that this paper's method identifies the structural parameters.

By contrast, the standard method obtains estimated age effects $\hat{\alpha}_a = k^*(a - \bar{a}) + \sigma^2 \sum_{s=0}^{a}(1 + \phi_s)^{-2}$ for some number k^* determined by the normalization on the regression (2), then chooses $\hat{\sigma}^2$ and $\hat{\rho}$ to solve (or best fit)

$$\hat{\sigma}^2 \sum_{s=0}^{a} (1 + \hat{\phi}_s)^{-2} = \hat{\alpha}_a, \quad a = 0, \ldots, A. \quad (17)$$

If $k^* \neq 0$, then for all $a \neq \bar{a}$, these equations do not hold when $\hat{\sigma}^2 = \sigma^2$ and $\hat{\rho} = \rho$, so the standard method must obtain $\hat{\sigma}^2 \neq \sigma^2$, $\hat{\rho} \neq \rho$, or both. Thus, if $k^* \neq 0$—that is, if the normalization is incorrect—the standard method fails to identify the structural parameters.

5. Quantitative example: Revisiting Gourinchas and Parker (2002)

Gourinchas and Parker (2002) use the standard method to estimate structural parameters of a life-cycle model in which households receive a stochastic income and decide how much to consume. This section tests how the results change with this paper's estimator.

5.1 *Model*

I briefly review the model here and refer readers to the original paper for details. Households work for $T = 40$ periods and then retire. Preferences while working are given by

$$\mathrm{E}\left[\sum_{t=1}^{T} \beta^t \frac{(C_t/Z_t)^{1-\rho}}{1-\rho} \right], \quad (18)$$

where β is the rate of time preference; ρ is the coefficient of relative risk aversion; Z_t is a deterministic family size adjustment, reflecting how changes in average family size with age affect the marginal utility of consumption; κ is a constant; and ζ_{T+1} is terminal liquid and illiquid wealth. Households choose consumption and savings at each age to maximize utility given an initial liquid wealth level W_1, the constraint that terminal liquid wealth W_{T+1} is nonnegative, and the budget constraint $W_{t+1} = R(W_t + Y_t - C_t)$. Income Y_t evolves according to a stochastic process with permanent and transitory shocks as well as an age-specific deterministic component. At retirement, the household follows a terminal consumption rule that is linear in liquid wealth normalized by permanent income P_{T+1},

$$(C_{T+1}/P_{T+1}) = \gamma_0 + \gamma_1 (W_{T+1} + Y_{T+1})/P_{T+1}. \quad (19)$$

5.2 *Original estimation procedure*

Gourinchas and Parker (2002) use external data to estimate the interest rate R, the variances of the income shocks, and the mean initial wealth level W_1. Next, they use repeated cross sections from the Consumer Expenditure Survey to estimate age profiles of log consumption, income, and family size. The age profile of log consumption is estimated by an equation analogous to (2), but dummy variables to control for within-age differences in family size are added, and the time effects are replaced by the unemployment rate to solve the identification problem. The age profile of the across-ages family size adjustment Z_t is calculated as the mean of the coefficients on the age dummies, weighted by the distribution of family sizes among households of age t; thus, there are assumed to be no period or cohort effects in the family size adjustment Z_t.[2] Income is normalized by the estimated family size adjustment, and the age profile of normalized income is estimated from an equation analogous to (2) but with time effects replaced by the unemployment rate.

The remaining parameters—β, ρ, γ_0, and γ_1—are chosen by the Method of Simulated Moments to fit the age profile of consumption. Given $\theta = (\beta, \rho, \gamma_0, \gamma_1)$ and first-stage parameters χ (which include age profiles of income and family size), Gourinchas and Parker (2002) calculate the household's consumption rule in the model and simulate the behavior of a large number of households. They then solve

$$\min_{\theta}\big[\overline{\ln C_t} - \ln \widehat{C_t(\theta, \chi)}\big]' \mathbf{W}\big[\overline{\ln C_t} - \ln \widehat{C_t(\theta, \chi)}\big], \tag{20}$$

where $\overline{\ln C_t}$ is the estimated age profile of log consumption in the data, $\ln \widehat{C_t(\theta, \chi)}$ is the mean of log consumption among simulated households of age t, and $\mathbf{W}$ is a weighting matrix.

5.3 *Replication*

Before implementing this paper's estimation method, I replicated the results of Gourinchas and Parker (2002) using their method. Jonathan Parker kindly shared with me the estimated age profiles and the GAUSS code used to estimate the parameters for the original paper. Because so much time has passed since the original code was written, I could not obtain a copy of the GAUSS software that could run the original code, so I wrote new code in C++.[3] My code follows as closely as possible all of the decisions made in the original code, such as the interpolation method used to approximate the consumption rule.

The parameters that minimize my implementation of (20) are close but not identical to the estimates published by Gourinchas and Parker (2002). The first column of Table 1 shows the estimates from Gourinchas and Parker (2002), while the second column shows the parameters that minimize my implementation of their objective function. Following Gourinchas and Parker (2002), I focus on results using a robust weighting

---

[2]Controlling for family size in the consumption regression removes within-age differences in family size, while including the family size age profile Z_t in the structural model accounts for how deterministic changes in average family size as an average family ages affect the marginal utility of consumption.

[3]I use the nonlinear optimization package of Johnson (2012) and utilities from Galassi et al. (2011).

TABLE 1. Comparison of estimation methods.

| | Standard Method | | New Method | | | |
| | Published | Replication | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Structural parameters:						
β	0.9598	0.9533	0.9570	0.9533	0.9520	0.9582
	(0.0179)	(0.0080)	(0.0180)	(0.0055)	(0.0061)	(0.3512)
ρ	0.5140	0.7440	1.7802	0.7452	0.8395	1.6527
	(0.1707)	(0.2516)	(0.2995)	(0.1922)	(0.2335)	(0.3747)
γ_0	0.0015	0.0002	0.0005	4.28×10^{-5}	4.29×10^{-5}	1.37×10^{-5}
	(3.85)	(0.4734)	(0.3404)	(0.2685)	(0.2221)	(0.0397)
γ_1	0.0710	0.0663	0.0546	0.0661	0.0629	0.0363
	(0.1244)	(0.0450)	(0.0228)	(0.0630)	(0.0199)	(0.0294)
Slope nuisance parameters:						
$k_{\text{consumption}}$		0	0.0150	0	0	0.0129
		–	(0.0081)	–	–	(0.2208)
$k_{\text{family size}}$		0	0	−0.0004	0	−0.0162
		–	–	(0.0516)	–	(0.2303)
k_{income}		0	0	0	−0.0018	-1.45×10^{-5}
		–	–	–	(0.0010)	(0.2214)
χ^2	174.10	149.40	109.12	147.15	136.15	108.43
d.f.	36	36	35	35	35	33

Note: Column 1 shows the parameter estimates and χ^2 statistic published by Gourinchas and Parker (2002). Column 2 shows the parameter estimates and χ^2 statistic produced in a replication exercise using the same method as Gourinchas and Parker (2002). Columns 3 through (6) show the parameter estimates produced using this paper's method. In column 3, the slope of the consumption age profile is allowed to vary freely, while the slopes of the family size and income age profiles are fixed at those estimated by Gourinchas and Parker (2002). In column 4, the slope of the family size age profile is allowed to vary freely, while the slopes of the consumption and income age profiles are fixed, and in column 5, only the slope of the income age profile is allowed to vary freely. Column 6 allows the slopes of all three age profiles to vary freely. Estimates using robust weighting matrix. Standard errors (in parentheses) and χ^2 statistics corrected for first-stage estimation. "d.f." indicates degrees of freedom.

matrix; results using the optimal weighting matrix proved to be unstable due to the need to numerically differentiate the objective function to estimate the optimal weights. The discrepancy between my results and those of Gourinchas and Parker (2002) for identical estimation procedures could be due to differences in the random number draws used for the simulations or differences in the numerical accuracy of the calculations. (For example, the standard errors are very sensitive to a tolerance used to calculate numerical gradients. I could not determine the value of this tolerance in GAUSS.) In all, though, the discrepancies in the point estimates are small and show that my replication essentially reproduces the published point estimates. If there are economically significant differences in the point estimates when I apply this paper's new estimation method, those differences must be due to the change in method—not to differences between my replication code and the original code.

5.4 *Estimation without normalizations on the age profiles*

Columns 3 through 6 of Table 1 show the results of applying this paper's estimation method. The estimates with this paper's method should be compared with my estimates from the standard method, shown in column 2, given that my code produces results slightly different from the published estimates even when applying the standard method.

Because there are three age profiles, I estimate up to three arbitrary slopes along with the structural parameters.[4] Column 3 allows an arbitrary trend only in the age profile of consumption; this change causes the estimated coefficient of relative risk aversion to more than double—to 1.78 from 0.74—and decreases γ_1, the marginal propensity to consume out of final wealth, by nearly 20 percent. Columns 4 and 5 instead allow trends in family size or income, with relatively little effect on the structural parameters. Finally, column 6 allows trends in all three age profiles; the coefficient of relative risk aversion is similar to that obtained using a consumption trend, and the marginal propensity to consume out of final wealth is even lower. Allowing arbitrary trends also improves the χ^2 statistic for model fit, though the overidentifying restrictions are still rejected.

Figure 1 shows how the new method identifies parameters from the curvature of the age profile. The figure shows detrended age profiles of log consumption, in the data and as predicted by the model at different parameter values. The new method brings the curvature closer to that in the data, compared with the standard method, by choosing structural parameters that make the age profile less curved during the first half of the life cycle.

Of course, changing the parameters also affects the first derivative of the age profile. Figure 2 plots the level of consumption as simulated with the parameters estimated by each method. The dashed dark line shows the age profile for parameters estimated by the standard method, while the dashed light line shows the age profile for parameters estimated by the new method. The new method estimates higher risk aversion. Hence, the precautionary motive is stronger, and households save more early in life, implying that consumption rises faster with age. Such a pattern would be grossly inconsistent with the age profile of consumption that Gourinchas and Parker (2002) estimate in the data using the normalization they chose, illustrated with the solid dark line. Therefore, the standard method strongly rejects a high coefficient of relative risk aversion. But the first derivative of the age profile is a function of the normalization, not of the data. This paper's method does not allow this unidentified first derivative to drive inferences about the structural parameters. Instead, the new method identifies the structural parameters by matching the curvature, as shown in Figure 1. Then the new method adds or subtracts

---

[4]Gourinchas and Parker (2002) impose more normalizations than are necessary to identify the age profiles: For consumption and income, they restrict the time effects to move in parallel with the unemployment rate, and for family size, they restrict the cohort and time effects to be zero. To maintain comparability, I do not relax the extra restrictions. Instead, I treat Gourinchas and Parker's (2002) estimated age profiles as if they were estimated using only the minimum required restrictions and then apply this paper's estimation method. The model also suffers from an unrelated identification problem: If $R\beta = 1$ and $\rho = 0$, the household is indifferent as to the timing of consumption, and *any* age profile of consumption that satisfies the budget constraint is consistent with the model. Therefore, I impose $R\beta < 1$ and $\rho > 0$.
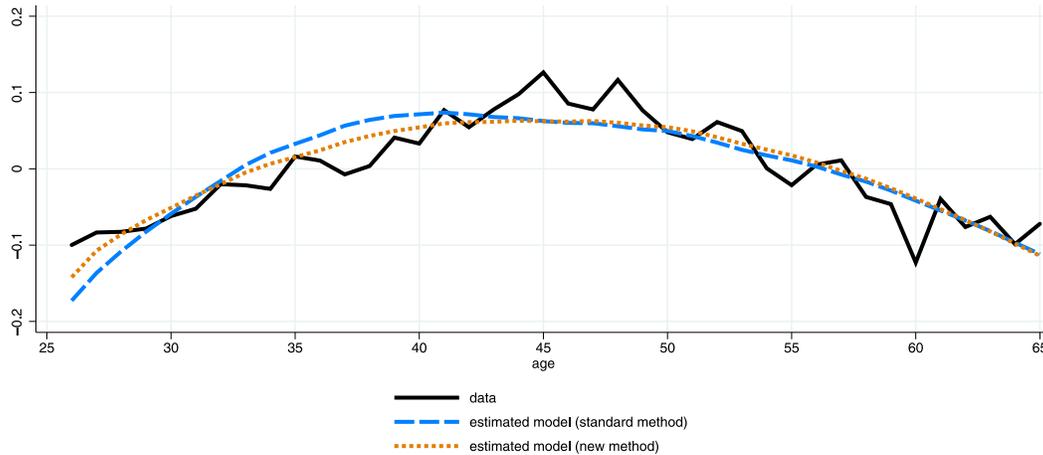
FIGURE 1. Detrended age profiles of ln(consumption). Graph shows residuals from regressing age profiles of the natural logarithm of consumption on a linear trend in age. Lines labeled "estimated model (standard method)" and "estimated model (new method)" are simulated from the model using parameter values in Table 1, columns 1 and 5, respectively.

a linear trend to rotate the empirical age profile so its slope is consistent with the age profile that the structural parameters predict. The solid light line in Figure 2 illustrates this rotated empirical age profile.

Table 1 shows that allowing arbitrary trends increases some standard errors but decreases others. In general, the parameter estimates remain relatively precise even after
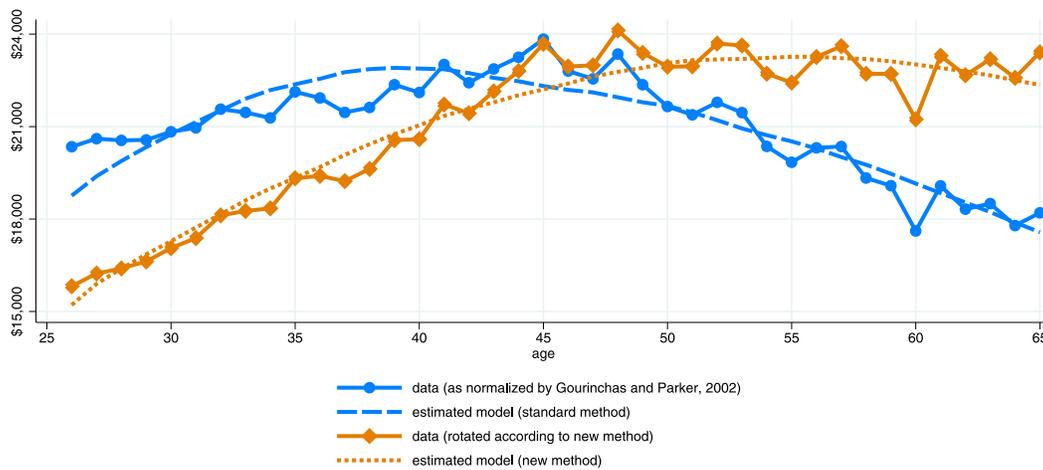

FIGURE 2. Age profiles of consumption. Lines labeled "estimated model (standard method)" and "estimated model (new method)" are simulated from the model using parameter values in Table 1, columns 1 and 5, respectively. The line labeled "data (rotated according to new method)" is the age profile in the data, rotated by the estimated consumption trend shown in Table 1, column 5.

allowing for arbitrary trends. Hence, in the model of Gourinchas and Parker (2002), the structural parameters remain well identified without having to resort to unneeded normalizations on age profiles, but removing those normalizations substantially changes one's conclusions about the true values of the parameters—significantly increasing the coefficient of relative risk aversion and reducing the marginal propensity to consume out of final wealth.

Gourinchas and Parker (2002) report a range of robustness checks with widely varying estimates of risk aversion. Depending on various choices, their estimate of the coefficient of relative risk aversion ranges from 0.1278 to 5.2823—a much larger difference than is produced by the change in normalization in this paper, and one with substantial economic consequences given that, for example, risk premia are generally proportional to risk aversion. Beyond the rejection of overidentifying restrictions, the sensitivity of the estimates to methodological choices may indicate that the model does not perfectly describe the data.

6. Conclusion

In estimating structural life-cycle models, an age-time-cohort identification problem arises when researchers project data that vary with both age and time onto a one-dimensional model that varies only with age. There are many ways to make such projections. A standard estimation strategy assumes a particular projection is correct, then estimates the structural parameters conditional on that assumption. This paper shows that such an assumption is unnecessary and, in general, leads to incorrect results. I provide an alternative approach that does not have this pitfall. The new method demonstrates that the structural parameters can be identified—and, in an empirical example, estimated relatively precisely—without having to first identify the age profile.

References

Aguiar, M. and E. Hurst (2013), "Deconstructing life cycle expenditure." *Journal of Political Economy*, 121 (3), 437–492. [650]

Ameriks, J. and S. P. Zeldes (2004), "How do household portfolio shares vary with age?" Manuscript, Columbia University. [651]

Cagetti, M. (2003), "Wealth accumulation over the life cycle and precautionary savings." *Journal of Business & Economic Statistics*, 21 (3), 339–353. [650]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75 (5), 1243–1284. [649]

De Nardi, M., E. French, and J. B. Jones (2010), "Why do the elderly save? The role of medical expenses." *Journal of Political Economy*, 118 (1), 39–75. [650]

Deaton, A. and C. Paxson (1994a), "Intertemporal choice and inequality." *Journal of Political Economy*, 102 (3), 437–467. [651]

Deaton, A. S. and C. H. Paxson (1994b), "Saving, growth, and aging in Taiwan." In *Studies in the Economics of Aging* (D. A. Wise, ed.), 331–357, University of Chicago Press, Chicago, IL. [651]

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi (2011), *GNU Scientific Library Reference Manual*, version 1.15. [653]

Gourinchas, P.-O. and J. A. Parker (2002), "Consumption over the life cycle." *Econometrica*, 70 (1), 47–89. [643, 645, 650, 652, 653, 654, 655, 657]

Hall, R. E. (1968), "Technical change and capital from the point of view of the dual." *Review of Economic Studies*, 35 (1), 35–46. [645, 649]

Heathcote, J., K. Storesletten, and G. L. Violante (2005), "Two views of inequality over the life cycle." *Journal of the European Economic Association*, 3 (2–3), 765–775. [651]

Huggett, M., G. Ventura, and A. Yaron (2011), "Sources of lifetime inequality." *American Economic Review*, 101 (7), 2923–2954. [650]

Johnson, S. G. (2012), "The NLopt nonlinear-optimization package," version 2.3. Available at https://nlopt.readthedocs.io/en/latest/. [653]

Kaplan, G. (2012), "Inequality and the life cycle." *Quantitative Economics*, 3 (3), 471–525. [650]

McKenzie, D. J. (2006), "Disentangling age, cohort and time effects in the additive model." *Oxford Bulletin of Economics and Statistics*, 68 (4), 473–495. [645]

Newey, W. K. and D. McFadden (1994), "Large sample estimation and hypothesis testing." In *Handbook of Econometrics*, Vol. IV (R. F. Engle and D. L. McFadden, eds.), 2111–2245, North-Holland, Amsterdam. [647]

Wachter, J. A. and M. Yogo (2010), "Why do household portfolio shares rise in wealth?" *Review of Financial Studies*, 23 (11), 3929–3965. [650]