Supplementary Material

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

(*Quantitative Economics*, Vol. 9, No. 2, July 2018, 643–658)

SAM SCHULHOFER-WOHL
Federal Reserve Bank of Chicago

This Appendix contains additional results on the estimation of structural parameters in life-cycle models in the face of the age-time-cohort problem. Section S.1 gives a more detailed proof of Proposition 1. Section S.2 further discusses the relationship between nonlinear least squares and the estimator in the paper. Section S.3 gives details on the derivation of the cross-sectional variance of consumption in the analytic example in the main paper. Section S.4 proves that Condition NL holds in the analytic example.

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

JEL CLASSIFICATION. C23, D91, J1.

## S.1. DETAILED PROOF OF PROPOSITION 1

Let $\mathbf{R} = \mathbf{I} - \mathbf{a}(\mathbf{a}'\mathbf{W}\mathbf{a})^{-1}\mathbf{a}'\mathbf{W}$ be the matrix that produces residuals from projecting any vector of length $A$ on a linear trend in $a - \bar{a}$ in a generalized least squares regression with weighting matrix $\mathbf{W}$. Then

$$\mathbf{q}(\boldsymbol{\theta}) = c_1(\boldsymbol{\theta})\mathbf{a} + \mathbf{R}\mathbf{q}(\boldsymbol{\theta}), \tag{A.1}$$

where $c_1(\boldsymbol{\theta})$ is the slope in the GLS regression of $q(a; \boldsymbol{\theta})$ on $a$ with weighting matrix $\mathbf{W}$ and $\mathbf{R}\mathbf{q}(\boldsymbol{\theta})$ is orthogonal to $\mathbf{a}$ under the weighting given by $\mathbf{W}$. Similarly, let $\hat{c}_2$ be the slope in the GLS regression of $\hat{\alpha}_a$ on $a$ and write

$$\hat{\boldsymbol{\alpha}} = \hat{c}_2\mathbf{a} + \mathbf{R}\hat{\boldsymbol{\alpha}}. \tag{A.2}$$

Then, for any $k$ and $\boldsymbol{\theta}$,

$$\mathbf{q}(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}} - k\mathbf{a} = \mathbf{R}\mathbf{q}(\boldsymbol{\theta}) - \mathbf{R}\hat{\boldsymbol{\alpha}} - \left[k - c_1(\boldsymbol{\theta}) + \hat{c}_2\right]\mathbf{a}. \tag{A.3}$$

Because $\mathbf{R}'\mathbf{W}\mathbf{a} = \mathbf{0}$, we then have

$$\begin{aligned} &\left[\mathbf{q}(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}} - k\mathbf{a}\right]'\mathbf{W}\left[\mathbf{q}(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}} - k\mathbf{a}\right] \\ &\quad = \left[\mathbf{R}\mathbf{q}(\boldsymbol{\theta}) - \mathbf{R}\hat{\boldsymbol{\alpha}}\right]'\mathbf{W}\left[\mathbf{R}\mathbf{q}(\boldsymbol{\theta}) - \mathbf{R}\hat{\boldsymbol{\alpha}}\right] + \left[k - c_1(\boldsymbol{\theta}) + \hat{c}_2\right]^2\mathbf{a}'\mathbf{W}\mathbf{a}. \end{aligned} \tag{A.4}$$

Sam Schulhofer-Wohl: samuel.schulhofer-wohl@chi.frb.org


Available at http://qeconomics.org. https://doi.org/10.3982/QE738

Hence the solution to the minimization problem in equation (3) of the main paper is

$$\hat{\boldsymbol{\theta}} \in \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{R}\hat{\boldsymbol{\alpha}}\big]'\mathbf{W}\big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{R}\hat{\boldsymbol{\alpha}}\big], \tag{A.5a}$$

$$\hat{k} = c_1(\hat{\boldsymbol{\theta}}) - \hat{c}_2. \tag{A.5b}$$

Now let $\mathbf{M}$ be the first $A$ rows of the Moore–Penrose pseudoinverse of the design matrix of the regression in step 1, so $\hat{\boldsymbol{\alpha}} = \mathbf{My}$. Then, as indicated in the main paper, the estimator of $\boldsymbol{\theta}^*$ can be expressed as

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

We now need to show that 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 objective function in (A.6) converges uniformly in probability to $Q_0$.

Let $g(\boldsymbol{\theta}, \mathbf{u})$ be the objective function in (A.6). Under Assumption 1,

$$\hat{\boldsymbol{\alpha}} = \mathbf{q}(\boldsymbol{\theta}^*) - k^*\mathbf{a} + \mathbf{Mu}, \tag{A.7}$$

where $k^*$ is an unknown real number determined by the normalization in step 1. Therefore,

$$\mathbf{RMy} = \mathbf{R}\hat{\boldsymbol{\alpha}} = \mathbf{R}\big[\mathbf{q}(\boldsymbol{\theta}^*) - k^*\mathbf{a} + \mathbf{Mu}\big] = \mathbf{Rq}(\boldsymbol{\theta}^*) + \mathbf{RMu}, \tag{A.8}$$

where we have used $\mathbf{Ra} = 0$. Hence

$$\begin{aligned} g(\boldsymbol{\theta}, \mathbf{u}) &= \big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*) - \mathbf{RMu}\big]'\mathbf{W}\big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*) - \mathbf{RMu}\big] \\ &= Q_0(\boldsymbol{\theta}) + [\mathbf{RMu}]'\mathbf{W}[\mathbf{RMu}] - 2\big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*)\big]'\mathbf{W}[\mathbf{RMu}], \end{aligned} \tag{A.9}$$

where

$$Q_0(\boldsymbol{\theta}) = \big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*)\big]'\mathbf{W}\big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*)\big]. \tag{A.10}$$

Because $\boldsymbol{\Theta}$ is compact and $\mathbf{q}$ is continuous by Assumption 3, $\mathbf{q}$ is bounded on $\boldsymbol{\Theta}$. Therefore, there exists a number $b < \infty$ such that, for all $a$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $|q(a;\boldsymbol{\theta}) - q(a;\boldsymbol{\theta}^*)| < b$. Let $\mathbf{b}$ be an $A \times 1$ column vector all of whose entries are $b$. Then

$$\begin{aligned} 0 &\le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|g(\boldsymbol{\theta}, \mathbf{u}) - Q_0(\boldsymbol{\theta})\big| \\ &= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|[\mathbf{RMu}]'\mathbf{W}[\mathbf{RMu}] - 2\big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*)\big]'\mathbf{W}[\mathbf{RMu}]\big| \\ &\le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|[\mathbf{RMu}]'\mathbf{W}[\mathbf{RMu}]\big| + 2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\big[\mathbf{Rq}(\boldsymbol{\theta}) - \mathbf{Rq}(\boldsymbol{\theta}^*)\big]'\mathbf{W}[\mathbf{RMu}]\big| \\ &= \big|\mathbf{u}'(\mathbf{M}'\mathbf{R}'\mathbf{WRM})\mathbf{u}\big| + 2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\big[\mathbf{q}(\boldsymbol{\theta}) - \mathbf{q}(\boldsymbol{\theta}^*)\big]'\mathbf{R}'\mathbf{WRMu}\big| \\ &\le \big|\mathbf{u}'(\mathbf{M}'\mathbf{R}'\mathbf{WRM})\mathbf{u}\big| + 2A\mathbf{b}'\big|\mathbf{R}'\mathbf{WRMu}\big|. \end{aligned} \tag{A.11}$$

By Assumption 2,

$$\big|\mathbf{u}'(\mathbf{M}'\mathbf{R}'\mathbf{WRM})\mathbf{u}\big| \xrightarrow{p} 0 \tag{A.12}$$

and

$$2A\mathbf{b}'\left|\mathbf{R}'\mathbf{W}\mathbf{R}\mathbf{M}\mathbf{u}\right| \xrightarrow{p} 0. \tag{A.13}$$

Therefore,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|g(\boldsymbol{\theta},\mathbf{u}) - Q_0(\boldsymbol{\theta})\right| \xrightarrow{p} 0 \tag{A.14}$$

and we have shown that $g(\boldsymbol{\theta},\mathbf{u})$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$, satisfying hypothesis (iv).

Hypotheses (ii) and (iii) are satisfied by Assumption 3. It remains to show that $\boldsymbol{\theta}^*$ uniquely minimizes $Q_0$ on $\boldsymbol{\Theta}$. Clearly, $Q_0(\boldsymbol{\theta}^*) = 0$ and $Q_0 \geq 0$ everywhere. So it is sufficient to show that $Q_0(\boldsymbol{\theta}) = 0$ only if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Because $\mathbf{W}$ is positive definite, $Q_0(\boldsymbol{\theta}) = 0$ only if $\mathbf{R}\mathbf{q}(\boldsymbol{\theta}) - \mathbf{R}\mathbf{q}(\boldsymbol{\theta}^*) = \mathbf{0}$, which in turn is true only if $\mathbf{q}(\boldsymbol{\theta}) - \mathbf{q}(\boldsymbol{\theta}^*)$ lies in the null space of $\mathbf{R}$. The null space of $\mathbf{R}$ is $\mathbf{a}$. Hence $Q_0(\boldsymbol{\theta}) = 0$ only if $\mathbf{q}(\boldsymbol{\theta}) - \mathbf{q}(\boldsymbol{\theta}^*)$ is proportional to $\mathbf{a}$. But under Condition NL, there is no $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ other than $\boldsymbol{\theta}^*$ for which $\mathbf{q}(\boldsymbol{\theta}) - \mathbf{q}(\boldsymbol{\theta}^*)$ is proportional to $\mathbf{a}$. Therefore, $\boldsymbol{\theta}^*$ uniquely minimizes $Q_0$ on $\boldsymbol{\Theta}$. Thus, the conditions of Newey and McFadden (1994, Theorem 2.1), are satisfied and $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*$. □

## S.2. Comparison with nonlinear least squares

As indicated in the main paper, an alternative approach would be to estimate $\boldsymbol{\theta}^*$ and the period and cohort effects simultaneously by nonlinear least squares (NLS) on:

$$\begin{aligned}&\left(\acute{\boldsymbol{\theta}} \in \boldsymbol{\Theta}, \acute{\xi}_0, \{\acute{\beta}_t\}, \{\acute{\gamma}_c\}\right) \in \arg\min_{\boldsymbol{\theta},\xi_0,\{\beta_t\},\{\gamma_c\}} \sum_{a,t}\left[y_{a,t} - \xi_0 - q(a;\boldsymbol{\theta}) - \beta_t - \gamma_c\right]^2 \\ &\text{s.t.} \quad \sum_t \beta_t = \sum_c \gamma_c = 0.\end{aligned} \tag{A.15}$$

Under Assumptions 1 and 2, the true parameters $\boldsymbol{\theta}^*$, $\xi_0^*$, $\beta_t^*$, $\gamma_c^*$ are one (asymptotic) solution to (A.15). The asymptotic objective function is zero at this solution. Because the asymptotic objective function is nonnegative, it must be zero—and all of the residuals must be zero— at any asymptotic solution. The solution therefore will be asymptotically unique in terms of $\acute{\boldsymbol{\theta}}$ if and only if there does not exist $\grave{\boldsymbol{\theta}} \neq \boldsymbol{\theta}^*$ such that the residuals are exactly the same at all $a$ and $t$ whether the objective function is evaluated at $\boldsymbol{\theta}^*$ or at $\grave{\boldsymbol{\theta}}$. In other words, the solution is asymptotically unique in terms of $\acute{\boldsymbol{\theta}}$ if and only if there do not exist $\grave{\boldsymbol{\theta}} \neq \boldsymbol{\theta}^*$ and $\grave{\xi}_0$, $\{\grave{\beta}_t\}$, $\{\grave{\gamma}_c\}$ such that

$$\forall a, t \quad q(a;\grave{\boldsymbol{\theta}}) - q\left(a;\boldsymbol{\theta}^*\right) = \xi_0^* - \grave{\xi}_0 + \beta_t^* - \grave{\beta}_t + \gamma_c^* - \grave{\gamma}_c. \tag{A.16}$$

Equation (A.16) implies there is some real number $\bar{k}$ such that

$$\forall c, t \quad \grave{\gamma}_c = \gamma_c^* + \bar{k}(c - \bar{c}), \qquad \grave{\beta}_t = \beta_t^* - \bar{k}(t - \bar{t}). \tag{A.17}$$

Also, because we have normalized $\sum_a q(a;\boldsymbol{\theta}) = \sum_t \beta_t = \sum_c \gamma_c$ for all parameter vectors, (A.16) requires $\xi_0^* = \grave{\xi}_0$. Hence the NLS solution is asymptotically unique in terms of $\acute{\boldsymbol{\theta}}$ if and only if there is no real number $\bar{k}$ such that

$$\forall a, t \quad q(a;\grave{\boldsymbol{\theta}}) - q\left(a;\boldsymbol{\theta}^*\right) = \bar{k}(a - \bar{a}). \tag{A.18}$$

This is exactly Condition NL. Therefore, NLS asymptotically identifies the parameters if and only if Condition NL holds, which is the same situation in which this paper's method asymptotically identifies the parameters.

### S.3. Derivation of the consumption function in the analytic example

This section presents additional details on the derivation of the consumption function and the cross-sectional variance of consumption in the paper's analytic example.

Krueger (2016, equation (5.46)) shows that

$$C_{i,a,c+a} = \theta_a^{-1} \frac{rW_{i,a,c+a}}{1+r}, \tag{A.19}$$

where

$$\theta_a = 1 - \frac{1}{(1+r)^{A-a+1}}, \tag{A.20a}$$

$$W_{i,a,c+a} = x_{i,a,c+a} + y_{i,a,c+a} + \mathrm{E}_{a,c+a} \sum_{s=1}^{A-a} \frac{y_{i,a+s,c+a+s}}{(1+r)^s}, \tag{A.20b}$$

$$x_{i,a,c+a} = (1+r)(x_{i,a-1,c+a-1} + y_{i,a-1,c+a-1} - C_{i,a-1,c+a-1}). \tag{A.20c}$$

Using the i.i.d. distribution of income and $(1+r) = \rho^{-1}$,

$$W_{i,a,c+a} = x_{i,a,c+a} + y_{i,a,c+a} + \mu \sum_{s=1}^{A-a} \rho^s = x_{i,a,c+a} + y_{i,a,c+a} + \mu\phi_a, \tag{A.21}$$

where

$$\phi_a = \sum_{s=1}^{A-a} \rho^s. \tag{A.22}$$

Observe that

$$1 + \phi_a = \sum_{s=0}^{A-a} \rho^s = \frac{1-\rho^{A-a+1}}{1-\rho} \tag{A.23}$$

and hence

$$C_{i,a,c+a} = (1+\phi_a)^{-1}(x_{i,a,c+a} + y_{i,a,c+a} + \mu\phi_a). \tag{A.24}$$

Substituting (A.24) and $(1+r) = \rho^{-1}$ into (A.20c), we have

$$x_{i,a+1,c+a+1} = \rho^{-1}\big[1-(1+\phi_a)^{-1}\big](x_{i,a,c+a} + y_{i,a,c+a}) - \rho^{-1}\phi_a(1+\phi_a)^{-1}\mu. \tag{A.25}$$

Observe that

$$1-(1+\phi_a)^{-1} = 1 - \frac{1-\rho}{1-\rho^{A-a+1}} = \frac{\rho-\rho^{A-a+1}}{1-\rho^{A-a+1}} = \rho\frac{1-\rho^{A-(a+1)+1}}{1-\rho^{A-a+1}} = \rho\frac{1+\phi_{a+1}}{1+\phi_a} \tag{A.26}$$

and that

$$\phi_a = \frac{1-\rho^{A-a+1}}{1-\rho} - 1 = \frac{\rho-\rho^{A-a+1}}{1-\rho} = \rho\frac{1-\rho^{A-(a+1)+1}}{1-\rho} = \rho(1+\phi_{a+1}). \tag{A.27}$$

Substituting (A.26) and (A.27) into (A.25) gives

$$x_{i,a+1,c+a+1} = \frac{1+\phi_{a+1}}{1+\phi_a}(x_{i,a,c+a} + y_{i,a,c+a} - \mu). \tag{A.28}$$

Working backwards from (A.28), we have

$$x_{i,a,c+a} = \frac{1+\phi_a}{1+\phi_0}x_{i,0,c} + \sum_{j=1}^{a}\frac{1+\phi_a}{1+\phi_{a-j}}(y_{i,a-j,c+a-j} - \mu). \tag{A.29}$$

The cross-sectional variance of consumption in cohort $c$ at age $a$ is therefore

$$\begin{aligned}
\operatorname{Var}[C_{i,a,c+a}|a,c] &= (1+\phi_a)^{-2}\big(\operatorname{Var}[x_{i,a,c+a}] + \sigma^2\big)\\
&= (1+\phi_a)^{-2}\left[\left(\frac{1+\phi_a}{1+\phi_0}\right)^2\operatorname{Var}[x_{i,0,c}] + \sum_{j=1}^{a}\left(\frac{1+\phi_a}{1+\phi_{a-j}}\right)^2\sigma^2 + \sigma^2\right]\\
&= (1+\phi_0)^{-2}\operatorname{Var}[x_{i,0,c}] + \sigma^2\sum_{j=0}^{a}[1+\phi_{a-j}]^{-2}\\
&= (1+\phi_0)^{-2}\operatorname{Var}[x_{i,0,c}] + \sigma^2\sum_{s=0}^{a}[1+\phi_s]^{-2}
\end{aligned} \tag{A.30}$$

as claimed in the main paper.

## S.4. Proof that Condition NL holds in the analytic example

Using $\phi_s = \rho(1-\rho^{A-s})/(1-\rho)$, Condition NL requires that the following equations have a unique solution $\hat\sigma^2 = \sigma^2$, $\hat\rho = \rho$, $k = 0$:

$$\sigma^2\sum_{s=0}^{a}\left(1+\rho\frac{1-\rho^{A-s}}{1-\rho}\right)^{-2} = ka + \hat\sigma^2\sum_{s=0}^{a}\left(1+\hat\rho\frac{1-\hat\rho^{A-s}}{1-\hat\rho}\right)^{-2}, \quad a = 0,\dots,A. \tag{A.31}$$

One obvious solution is $\hat\sigma^2 = \sigma^2$, $\hat\rho = \rho$, $k = 0$; we need to prove that there is no other. Specializing to $a = 0, 1, 2$, we have

$$\sigma^2\left(1+\rho\frac{1-\rho^{A}}{1-\rho}\right)^{-2} = \hat\sigma^2\left(1+\hat\rho\frac{1-\hat\rho^{A}}{1-\hat\rho}\right)^{-2}, \tag{A.32a}$$

$$\sigma^2\sum_{s=0}^{1}\left(1+\rho\frac{1-\rho^{A-s}}{1-\rho}\right)^{-2} = k + \hat\sigma^2\sum_{s=0}^{1}\left(1+\hat\rho\frac{1-\hat\rho^{A-s}}{1-\hat\rho}\right)^{-2}, \tag{A.32b}$$

$$\sigma^2\sum_{s=0}^{2}\left(1+\rho\frac{1-\rho^{A-s}}{1-\rho}\right)^{-2} = 2k + \hat\sigma^2\sum_{s=0}^{2}\left(1+\hat\rho\frac{1-\hat\rho^{A-s}}{1-\hat\rho}\right)^{-2}. \tag{A.32c}$$

Using (A.32a) to substitute for $\hat{\sigma}^2$ in (A.32b) and (A.32c), then using (A.32b) to eliminate $k$ and simplifying, we have

$$\left(\frac{1-\rho^{A-1}}{1-\rho^{A+1}}\right)^{-2} - \left(\frac{1-\rho^{A}}{1-\rho^{A+1}}\right)^{-2} = \left(\frac{1-\hat{\rho}^{A-1}}{1-\hat{\rho}^{A+1}}\right)^{-2} - \left(\frac{1-\hat{\rho}^{A}}{1-\hat{\rho}^{A+1}}\right)^{-2}. \tag{A.33}$$

For $\hat{\rho} \in (0, 1)$, the right-hand side of (A.33) is monotonically increasing in $\hat{\rho}$; therefore, (A.33) has a unique solution, $\hat{\rho} = \rho$. We then obtain $\hat{\sigma}^2 = \sigma^2$ from (A.32a) and $k = 0$ from (A.32b). Thus, the solution is unique, and Condition NL holds.

## References

Krueger, D. (2016), "An introduction to macroeconomics with household heterogeneity." Manuscript, University of Pennsylvania. [4]

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. [3]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 22 June, 2016; final version accepted 11 October, 2017; available online 5 December, 2017.