Disentangling age, cohort, and time effects in the additive model

DAVID J. MCKENZIE*

Development Research Group, The World Bank,

1818 H Street N.W., Washington, D.C. 20433, U.S.A.

(e-mail: dmckenzie@worldbank.org)

August 4, 2005

Abstract

This paper presents a new approach to the old problem of linear dependency of age, cohort and time effects. It is shown that second differences of the effects can be estimated without any normalization restrictions, providing information on the shape of the age, cohort and time effect profiles, and enabling identification of structural breaks. A Wald test is provided to test the popular linear and quadratic specifications against a very general alternative. The method is illustrated through examples which show its ability to detect structural breaks in time effects as a result of the Mexican peso crisis, and to determine whether the age effects profile in the variance of Taiwanese log consumption is concave or convex.

JEL Classifications: C23, C81, J11.

Keywords: Cohort effect, Age effect, Time effect, Pseudo-Panel.

*Thanks to Luigi Pistaferri, Edward Vytlacil, a co-editor (Adriaan Kalwij), an anonymous referee, and participants at seminars at Stanford University, NEUDC, and the 2002 North American Summer Meeting of the Econometric Society for useful comments.

1 Introduction

Many economic and social phenomena are modelled as a confluence of age, birth cohort, and time effects. General linear models which attempt to capture all three effects are faced with one of the most well-known identification problems in economics: a person's age added to their birth year gives the current year, so that there is an exact linear relationship between the age, cohort, and time effects. Identifying the level effects of these three factors therefore requires additional normalization or exclusion assumptions, which is the current practice in the literature.

In this paper we provide a different approach to the identification of these effects in the additively separable model. Using pseudo-panel methods, we show that although the level effects can not be identified without further restrictions, one can identify economically meaningful linear combinations of these effects. In particular, with no normalizing assumptions, one can identify second differences of the effects. These effectively provide the second derivatives of the age, cohort, and time effect profiles, providing valuable information on changes in growth rates, the convexity or concavity of the effects, and enabling structural breaks to be seen. Furthermore, we show that Wald tests on the estimated second differenced effects can be used to test whether either a quadratic or a linear term can adequately capture a given effect. Next, we show that with only one normalizing assumption on the slopes, the first differences of all three effects can be identified, with identification of the level effects requiring two further assumptions.

We illustrate the methods in the paper by way of two empirical examples. The first example shows the ability of this approach to identify structural breaks in the time effect profile through examination of Mexican household consumption data. Our method clearly shows the time effect of the 1995 peso crisis on consumption, and in addition shows that a quadratic in age is not sufficient for explaining the age effect. Our second example shows how the method can be used to make inferences about the shape of the age effect profile for the variance of log consumption in Taiwan. Deaton and Paxson (1994) show that the permanent income hypothesis (PIH) implies that consumption inequality increases with age. However, whether the age effect profile is concave or convex depends on the degree of persistence in shocks to earnings. If the PIH is correct, then a convex age effects profile implies that individual earnings must contain a large stationary component. Applying the methods of this paper, it is found that the second derivatives of the age effect profile are in fact fairly equal, implying at most a quadratic in age is needed. The second differences are statistically insignificant from zero for most ages, and we can not reject linearity of the age effects profile for ages 30 to 46. However, over a wider range linearity is rejected in favor of a quadratic specification, which does turn out to be convex.

The existing economics literature[1] suggests a variety of restrictions for overcoming the identification problem. One view is that the age, cohort, and time effects are proxy variables for underlying unobserved variables which are not themselves linearly dependent (Heckman and Robb, 1985). In some situations, the business cycle may capture the period effect and cohort size the cohort effect. Foster (1990) notes that, in demography, parametric schedules which take advantage of strong regularities in the age patterns of vital rates in human populations can be used to capture demographic events. When multiple proxies are available, Heckman and Robb (1985) propose a latent variable approach, enabling estimation of a multiple-cause-multiple-indicator model. However, often the underlying variable(s) for a given effect may be unclear, and so the researcher prefers to agnostically model all three effects simultaneously.

A second approach is then to model the age and cohort effects with small-order polynomials, in some cases just linear effects. For example, Japelli (1999) uses fifth order polynomials in age and cohort, while Denton, Mountain and Spencer (1999) use a quadratic for cohort and time effects, and a cubic spline for age effects with knots at 17 and 57 to capture variations associated with lifecycle transitions. These models may be reasonable, but there is often an ad-hoc nature to their specification, and they can struggle to adequately capture trend breaks. Using non-parametric methods, Heckman and Vytlacil (2001) find no support for the widely accepted practice of imposing linear effects of time and age. Deaton (1997) argues that when data are plentiful, it is better to allow dummy variables for all three sets of effects, allowing the data to choose the profiles. He provides a normalization which makes the year effects orthogonal to a time trend, so that all growth

[1] See the volume edited by Mason and Fienberg (1985) for a discussion of the identification approaches used in other social sciences.

is attributed to age and cohort effects.

In a working paper, Attanasio (1993) showed that by using information on the stock of financial assets, one can identify the shape of the age profile for changes in financial assets, which represent saving.[2] His approach to identification is very similar to that set out in this paper to identify the second derivative of the age profile: where he uses time differences in a polynomial in age, cohort, and time effects, I use a more general dummy variable specification. This paper builds on Attanasio's approach by showing that similar methods can be used to identify the shapes of cohort and time effect profiles, by allowing for the presence of individual effects, and by providing for a distribution theory which allows one to test hypotheses about the shapes of the age, cohort and time effect profile.

The remainder of this paper is organized as follows. Section 2 presents the basic additive model in age, cohort, and time effects to be used in this paper. Section 3 shows that one can consistently estimate second derivatives without further parameter restrictions, and provides a Wald test for testing whether the effects can be captured by quadratic or linear terms. In Section 4 it is shown that the introduction of one normalization assumption can allow identification of first derivatives, while two more normalizations allow identification of actual effects. The method is applied to Mexican household consumption data and Taiwanese inequality data in Section 5. Section 6 concludes and mathematical proofs of the Theorems are presented in Section 7.

2 Model

Observations are made on individuals of A age groups, $a_1, ..., a_A$, over T time periods, $t_1, ..., t_T$. The population is divided into $C = A + T - 1$ cohorts, with some cohorts observed in more time periods than others due to the restriction on ages. For example, both our applications involve consumption, which is only measured at the household level. Cohorts are then defined by the age of the household head, and so the standard practice is to focus on an age range over which one can

[2] I thank a referee for alerting me to this previous work. In the published version of this working paper, Attanasio (1998) does not use this method but instead employs the same normalization assumption as Deaton (1997), assuming that all time effects are orthogonal to a linear trend and average to zero.

assume no systematic changes in who is the household head. Thus one might consider consumption patterns of 20-65 year olds between 1980 and 2005. Then the cohort aged 63 in 1980 would only be followed for two more years before they age out of the sample.

The cohort of individuals aged a_j in time period t_k is denoted as cohort c_{j-k+1}. Cohort c_1 is thus aged a_1 at time t_1, while cohort c_A is aged a_A at time t_1, and c_0 is aged a_1 at time t_2. We assume that selection of the number of age groups and cohorts is predetermined, leaving issues of optimal cohort selection for further research. The number of individuals sampled from cohort c_j is n_{c_j}, which can vary from cohort to cohort, but for notational purposes is assumed to be the same in each time period the cohort is sampled.

For individual i in cohort c_{j-k+1}, of age a_j in time period t_k, the variable of interest y_{i,c_{j-k+1},a_j,t_k} is modelled as the sum of a cohort effect, an age effect, a time effect, and an individual error term[3]:

$$y_{i,c_{j-k+1},a_j,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + \gamma_{t_k} + \varepsilon_{i,c_{j-k+1},a_j,t_k} \tag{1}$$

Our underlying assumption will be that the error term $\varepsilon_{i,c_{j-k+1},a_j,t_k}$ contains an individual fixed effect and a time-varying individual level component.[4] That is:

Assumption 1

$$\varepsilon_{i,c_{j-k+1},a_j,t_k} = \omega_{i,c_{j-k+1}} + \eta_{i,c_{j-k+1},a_j,t_k} \tag{2}$$

where $\omega_{i,c_{j-k+1}} \sim i.i.d.\left(0,\sigma_\omega^2\right)$, $\eta_{i,c_{j-k+1},a_j,t_k} \sim i.i.d.\left(0,\sigma_\eta^2\right)$, $\omega_{i,c_{j-k+1}}$ *and* $\eta_{i,c_{j-k+1},a_j,t_k}$ *are independent, and* $E\left(\varepsilon^4_{i,c_{j-k+1},a_j,t_k}\right) < \infty$.

In many empirical applications, including those in this paper, we do not observe the same individuals for multiple periods. Instead, following Deaton (1985), one can form a pseudo-panel, whereby cohorts are followed over time. Each individual is only observed once, and so we proceed by taking

[3] This is the additively separable model common in some of the literature. It could be further motivated in practice by estimating a saturated model of cohort, age, and interaction effects, and testing that the interaction terms are zero. Additional variables can also be included in the model (see the working paper, McKenzie 2002).

[4] In certain applications one may wish to allow for a time-varying cohort-level component to the error term. Consistency then requires using asymptotics in which either the number of cohorts or the number of time periods goes to infinity. McKenzie (2004) discusses pseudo-panel estimation under these alternative asymptotics.

means of equation (1) over cohorts at each time period:

$$\frac{1}{n_{c_{j-k+1}}}\sum_{i=1}^{n_{c_{j-k+1}}} y_{i(t_k),c_{j-k+1},a_j,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + \gamma_{t_k} + \frac{1}{n_{c_{j-k+1}}}\sum_{i=1}^{n_{c_{j-k+1}}} \varepsilon_{i(t_k),c_{j-k+1},a_j,t_k} \ . \quad (3)$$

Here the $i\,(t_k)$ subscript is used to make explicit the fact that different individuals are observed in each time period in the pseudo-panel.

The following assumption is made on the relative cohort sizes, which assures that one continues to obtain new observations from each cohort as the total sample grows:

Assumption 2 *: For all $s = 2 - T, ..., A$, there exists $0 < \delta_s < \infty$, such that $n_{c_s}/n_{c_1} \to \delta_s$ as $n_{c_1} \to \infty$.*

Letting $\overline{y}_{c_{j-k+1},a_j,t_k} = \frac{1}{n_{c_{j-k+1}}}\sum_{i=1}^{n_{c_{j-k+1}}} y_{i(t_k),c_{j-k+1},a_j,t_k}$ denote the cohort sample mean of the variable of interest for cohort c_{j-k+1} in time period t_k, the pseudo-panel version of equation (1) is then

$$\overline{y}_{c_{j-k+1},a_j,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + \gamma_{t_k} + \overline{\varepsilon}_{c_{j-k+1},a_j,t_k} \ . \quad (4)$$

The identification problem which arises is that an individual's birth year added to their age gives the current year, so that the regressor matrix of equation (4) is singular. That is, without further assumptions, one cannot separately identify age, cohort and time effects.

3 Identification of second derivatives

3.1 Age effects

Consider equation (4) for cohort c_1 at time periods t_1 and t_2:

$$\overline{y}_{c_1,a_1,t_1} = \alpha_{c_1} + \beta_{a_1} + \gamma_{t_1} + \overline{\varepsilon}_{c_1,a_1,t_1} \quad (5)$$

$$\overline{y}_{c_1,a_2,t_2} = \alpha_{c_1} + \beta_{a_2} + \gamma_{t_2} + \overline{\varepsilon}_{c_1,a_2,t_2} \quad (6)$$

Subtracting (5) from (6) eliminates the cohort effect and gives

$$\Delta_t \overline{y}_{c_1,a_2,t_2} = \left(\beta_{a_2} - \beta_{a_1}\right) + \left(\gamma_{t_2} - \gamma_{t_1}\right) + \Delta_t \overline{\varepsilon}_{c_1,a_2,t_2} \ . \tag{7}$$

where $\Delta_t \overline{y}_{c_1,a_2,t_2} \equiv \overline{y}_{c_1,a_2,t_2} - \overline{y}_{c_1,a_1,t_1}$ denotes the first time difference of $\overline{y}_{c_1,a_2,t_2}$, and similarly $\Delta_t \overline{\varepsilon}_{c_1,a_2,t_2}$ is the first time differenced error term. More generally, time differencing for cohort c_{j-k+1} between periods t_{k-1} and t_k gives:

$$\Delta_t \overline{y}_{c_{j-k+1},a_j,t_k} = \left(\beta_{a_j} - \beta_{a_{j-1}}\right) + \left(\gamma_{t_k} - \gamma_{t_{k-1}}\right) + \Delta_t \overline{\varepsilon}_{c_{j-k+1},a_j,t_k} \ . \tag{8}$$

This equation is an extension of equation (4) in Attanasio (1993) from the case of polynomials in age, cohort, and time effects to dummy variables for all three effects. Attanasio suggests estimating this equation by OLS in order to identify the shape of the age profile for changes in the variable of interest, but does not provide conditions for consistency of this estimator, or details of the standard errors of such estimates.

Our method proceeds by noting that from (7), time differencing the observations for cohort c_2 between time periods t_1 and t_2 gives:

$$\Delta_t \overline{y}_{c_2,a_3,t_2} = \left(\beta_{a_3} - \beta_{a_2}\right) + \left(\gamma_{t_2} - \gamma_{t_1}\right) + \Delta_t \overline{\varepsilon}_{c_2,a_3,t_2} \ . \tag{9}$$

Now subtracting (7) from (9) eliminates the differenced time effect and yields:

$$\Delta_c \Delta_t \overline{y}_{c_2,a_3,t_2} = \left(\beta_{a_3} - \beta_{a_2}\right) - \left(\beta_{a_2} - \beta_{a_1}\right) + \Delta_c \Delta_t \overline{\varepsilon}_{c_2,a_3,t_2} \tag{10}$$

where $\Delta_c \Delta_t \overline{y}_{c_2,a_3,t_2} \equiv \Delta_t \overline{y}_{c_2,a_3,t_2} - \Delta_t \overline{y}_{c_1,a_2,t_2}$ denotes the first cohort difference of $\Delta_t \overline{y}_{c_2,a_3,t_2}$. In terms of notation, we define differences over cohorts and time periods, which will also implicitly define age differences. A c and/or t subscript indicates that the difference is taken over cohorts and/or time periods, while the absence of a subscript for cohort (or time) indicates that the difference is for the same cohort (or time). Negative signs indicate forward differences. For example, $\Delta_{-t} \overline{y}_{c_2,a_3,t_2} =$

$\overline{y}_{c_2,a_3,t_2} - \overline{y}_{c_2,a_4,t_3}$.

Generalizing equation (10) to cohort c_{j-k+2}, gives for $j = 2, ..., A-1$; $k = 2, ..., T$:

$$\Delta_c \Delta_t \overline{y}_{c_{j-k+2},a_{j+1},t_k} = \left(\beta_{a_{j+1}} - \beta_{a_j}\right) - \left(\beta_{a_j} - \beta_{a_{j-1}}\right) + \Delta_c \Delta_t \overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}. \tag{11}$$

Define $\widetilde{\beta}_{a_{j+1}} = \left(\beta_{a_{j+1}} - \beta_{a_j}\right) - \left(\beta_{a_j} - \beta_{a_{j-1}}\right)$. This parameter $\widetilde{\beta}_{a_{j+1}}$ then measures the change in the slope of the age effect profile for y. For example, $\widetilde{\beta}_{a_3} = \left(\beta_{a_3} - \beta_{a_2}\right) - \left(\beta_{a_2} - \beta_{a_1}\right)$ gives the difference in the slope of the age profile between ages a_2 and a_3 from the slope between ages a_2 and a_1. From equation (11) we arrive at the following regression, for

$j = 2, ..., A-1$; $k = 2, ..., T$:

$$\Delta_c \Delta_t \overline{y}_{c_{j-k+2},a_{j+1},t_k} = \widetilde{\beta}_{a_{j+1}} + \Delta_c \Delta_t \overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} \ . \tag{12}$$

Let $\widehat{\beta}_{a_{j+1}}$ denote the ordinary least squares estimator of $\widetilde{\beta}_{a_{j+1}}$ from equation (12). That is

$$\widehat{\beta}_{a_{j+1}} = \frac{1}{T-1} \sum_{k=2}^{T} \Delta_c \Delta_t \overline{y}_{c_{j-k+2},a_{j+1},t_k} \tag{13}$$

The following theorem then applies:

Theorem 1 *Under Assumptions 1 and 2 and the data generating process given in equation (1), as* $n_{c_1} \to \infty$, *for* T *fixed, we have:*

(a) Consistency of the Second-Differenced Age Effects:

$$\widehat{\beta}_{a_{j+1}} \xrightarrow{p} \widetilde{\beta}_{a_{j+1}} = \left(\beta_{a_{j+1}} - \beta_{a_j}\right) - \left(\beta_{a_j} - \beta_{a_{j-1}}\right)$$

for all $j = 2, ..., A-1$.

(b) Asymptotic Normality:

$$\sqrt{n_{c_1}}\left(\widehat{\beta}_{a_{j+1}} - \widetilde{\beta}_{a_{j+1}}\right) \xrightarrow{d} N\left(0,\ \sigma_{j+1}\right) \ ,$$

$$\textit{where } \sigma_{j+1} = 2\left(\sigma_\omega^2 + \sigma_\eta^2\right) \frac{1}{(T-1)^2} \sum_{k=2}^{T} \left(\frac{1}{\delta_{j-k+2}} + \frac{1}{\delta_{j-k+1}}\right),$$

and letting $var_c\left(y_{i,c_{j-k+2},a_{j+1},t_k}\right) \equiv \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \left(y_{i,c_{j-k+2},a_{j+1},t_k} - \overline{y}_{c_{j-k+2},a_{j+1},t_k}\right)^2$ *denote the cross-sectional sample variance across individuals in cohort* c_{j-k+2} *at time* t_k*, we have* $var_c\left(y_{i,c_{j-k+2},a_{j+1},t_k}\right) \xrightarrow{p} \left(\sigma_\omega^2 + \sigma_\eta^2\right)$.

Remarks:

(a) Theorem 1 shows that it is thus possible to identify changes in the slopes of the age effects profile without any parameter restrictions. If the age effects are linear, these terms should all be zero.

(b) Part (b) shows that the variance of $\widehat{\beta}_{a_{j+1}}$ can be estimated using the cross-sectional sample variance and the relative sample sizes of the different cohorts. That is, a consistent estimator of the variance of $\sqrt{n_{c_1}}\left(\widehat{\beta}_{a_{j+1}} - \widetilde{\beta}_{a_{j+1}}\right)$ is

$$\widetilde{\sigma}_{j+1} = 2var_c\left(y_{i,c_{j-k+2},a_{j+1},t_k}\right) \frac{1}{(T-1)^2} \sum_{k=2}^{T} \left(\frac{n_{c_1}}{n_{c_{j-k+2}}} + \frac{n_{c_1}}{n_{c_{j-k+1}}}\right) \ . \tag{14}$$

The efficiency of this estimator can be improved by averaging the cross-sectional sample variances for all cohorts and all time periods, to arrive at the estimator:

$$\widehat{\sigma}_{j+1} = 2\frac{1}{(T-1)^2} \sum_{k=2}^{T} \left(\frac{n_{c_1}}{n_{c_{j-k+2}}} + \frac{n_{c_1}}{n_{c_{j-k+1}}}\right) \frac{1}{AT} \sum_{k=1}^{T} \sum_{j=1}^{A} var_c\left(y_{i,c_{j-k+2},a_{j+1},t_k}\right) \ . \tag{15}$$

(c) Note that although the estimator $\widehat{\beta}_{a_{j+1}}$ is obtained by running OLS on (12), the usual OLS standard errors will not be correct. Hence the need to use cross-sectional sample variances as in (15).

3.2 Time effects

We can use similar methods to those above to identify the second differences of time effects and cohort effects. Time differencing the observations for cohort c_0 between time periods t_2 and t_3 gives

$$\Delta_t \overline{y}_{c_0,a_2,t_3} = \left(\beta_{a_2} - \beta_{a_1}\right) + \left(\gamma_{t_3} - \gamma_{t_2}\right) + \Delta_t \overline{\varepsilon}_{c_0,a_2,t_3} \tag{16}$$

Subtracting equation (7) from (16) eliminates the age effects, giving

$$\Delta_{-c,t}\Delta_t \overline{y}_{c_0,a_2,t_3} = \left(\gamma_{t_3} - \gamma_{t_2}\right) - \left(\gamma_{t_2} - \gamma_{t_1}\right) + \Delta_{-c,t}\Delta_t \overline{\varepsilon}_{c_0,a_2,t_3}\ , \tag{17}$$

where $\Delta_{-c,t}\Delta_t \overline{y}_{c_0,a_2,t_3} \equiv \Delta_t \overline{y}_{c_0,a_2,t_3} - \Delta_t \overline{y}_{c_1,a_2,t_2}$. In general, for cohort c_{j-k+1}, $j = 2, ..., A$, in time period t_k, $k = 3, ..., T$,

$$\Delta_{-c,t}\Delta_t \overline{y}_{c_{j-k+1},a_j,t_k} = \left(\gamma_{t_k} - \gamma_{t_{k-1}}\right) - \left(\gamma_{t_{k-1}} - \gamma_{t_{k-2}}\right) + \Delta_{-c,t}\Delta_t \overline{\varepsilon}_{c_{j-k+1},a_j,t_k} \tag{18}$$

Defining $\widetilde{\gamma}_{t_k} = \left(\gamma_{t_k} - \gamma_{t_{k-1}}\right) - \left(\gamma_{t_{k-1}} - \gamma_{t_{k-2}}\right)$, we arrive at the following regression, for $j = 2, ..., A$; $k = 2, ..., T$:

$$\Delta_{-c,t}\Delta_t \overline{y}_{c_{j-k+1},a_j,t_k} = \widetilde{\gamma}_{t_k} + \Delta_{-c,t}\Delta_t \overline{\varepsilon}_{c_{j-k+1},a_j,t_k}\ . \tag{19}$$

The $\widetilde{\gamma}_{t_k}$ are the changes in the slopes of the time effects profile. In a number of applications one might be interested in determining whether there was a structural break or shock in a particular period. Estimation of $\widetilde{\gamma}_{t_k}$ will be informative for this purpose. For example, if y measures consumption, then $\widetilde{\gamma}_{t_k} < 0$ implies that $\left(\gamma_{t_k} - \gamma_{t_{k-1}}\right) < \left(\gamma_{t_{k-1}} - \gamma_{t_{k-2}}\right)$. That is, the rate of change of the time effect has fallen, meaning that consumption growth has fallen in this period.

Let $\widehat{\gamma}_{t_k}$ denote the ordinary least squares estimator of $\widetilde{\gamma}_{t_k}$ from equation (19). That is

$$\widehat{\gamma}_{t_k} = \frac{1}{A-1}\sum_{j=2}^{A} \Delta_{-c,t}\Delta_t \overline{y}_{c_{j-k+1},a_j,t_k} \tag{20}$$

The following theorem then shows $\widehat{\gamma}_{t_k}$ is consistent and asymptotically normally distributed after appropriate scaling:

Theorem 2 *Under Assumptions 1 and 2 and the data generating process given in (1), for all* $k = 2, ..., T,$ *as* $n_{c_1} \to \infty$*, for* T *fixed,*

(a) Consistency of the Second-Differenced Time Effects:

$$\widehat{\gamma}_{t_k} \xrightarrow{p} \widetilde{\gamma}_{t_k} = \left(\gamma_{t_k} - \gamma_{t_{k-1}}\right) - \left(\gamma_{t_{k-1}} - \gamma_{t_{k-2}}\right) \ .$$

(b) Asymptotic Normality:

$$\sqrt{n_{c_1}} \left(\widehat{\gamma}_{t_k} - \widetilde{\gamma}_{t_k}\right) \xrightarrow{d} N\left(0, \kappa_k\right) \ ,$$

$$\text{where } \kappa_k \quad = \quad 2\left(\sigma_\omega^2 + \sigma_\eta^2\right) \frac{1}{\left(A-1\right)^2} \sum_{j=2}^{A} \left(\frac{1}{\delta_{j-k+2}} + \frac{1}{\delta_{j-k+1}}\right) .$$

Remarks:

(a) Note that Theorem 2 requires at least three time periods for changes in relative time effects to be identified, whereas Theorem 1 shows that changes in the slope of the age profile can be identified with only two time periods, which may be all that is available for some data sets.

(b) The variance κ_k can be estimated using cross-sectional sample variances and relative cohort sample sizes as was done for the age effects.

3.3 Cohort effects

Finally, to estimate second differences in the cohort effects, consider (4) for cohort c_2 at time period t_2:

$$\overline{y}_{c_2,a_3,t_2} = \alpha_{c_2} + \beta_{a_3} + \gamma_{t_2} + \overline{\varepsilon}_{c_2,a_3,t_2} \ . \tag{21}$$

Subtracting (6) from (21) eliminates the time effects and gives:

$$\Delta_c \overline{y}_{c_2,a_3,t_2} = \left(\alpha_{c_2} - \alpha_{c_1}\right) + \left(\beta_{a_3} - \beta_{a_2}\right) + \Delta_c \overline{\varepsilon}_{c_2,a_3,t_2} \ , \tag{22}$$

where $\Delta_c \overline{y}_{c_2,a_3,t_2} = \overline{y}_{c_2,a_3,t_2} - \overline{y}_{c_1,a_2,t_2}$ denotes the first cohort difference of $\overline{y}_{c_2,a_3,t_2}$. Taking first cohort differences of $\overline{y}_{c_3,a_3,t_1}$ likewise gives:

$$\Delta_c \overline{y}_{c_3,a_3,t_1} = (\alpha_{c_3} - \alpha_{c_2}) + \left(\beta_{a_3} - \beta_{a_2}\right) + \Delta_c \overline{\varepsilon}_{c_3,a_3,t_2} \ . \tag{23}$$

Subtracting (22) from (23) eliminates the age effects, giving

$$\Delta_{c,-t} \Delta_c \overline{y}_{c_3,a_3,t_1} = (\alpha_{c_3} - \alpha_{c_2}) - (\alpha_{c_2} - \alpha_{c_1}) + \Delta_{c,-t} \Delta_c \overline{\varepsilon}_{c_3,a_3,t_1} \ , \tag{24}$$

where $\Delta_{c,-t} \Delta_c \overline{y}_{c_3,a_3,t_1} \equiv \Delta_c \overline{y}_{c_3,a_3,t_1} - \Delta_c \overline{y}_{c_2,a_3,t_2}$. For cohort j, $j = 4 - T, ..., A$, $k = 1, ..., T - 1$, such that $2 \leq k + j - 1 \leq A$, this extends to:

$$\Delta_{c,-t} \Delta_c \overline{y}_{c_j,a_{k+j-1},t_k} = \left(\alpha_{c_j} - \alpha_{c_{j-1}}\right) - \left(\alpha_{c_{j-1}} - \alpha_{c_{j-2}}\right) + \Delta_{c,-t} \Delta_c \overline{\varepsilon}_{c_j,a_{k+j-1},t_k} \ . \tag{25}$$

Cohort j is only observed when its members are aged $a_1, ..., a_A$. Defining $\widetilde{\alpha}_{c_j} = \left(\alpha_{c_j} - \alpha_{c_{j-1}}\right) - \left(\alpha_{c_{j-1}} - \alpha_{c_{j-2}}\right)$, we arrive at the following regression for $j = 4 - T, ..., A$, $k = 1, ..., T - 1$ such that $2 \leq k + j - 1 \leq A$,

$$\Delta_{c,-t} \Delta_c \overline{y}_{c_j,a_{k+j-1},t_k} = \widetilde{\alpha}_{c_j} + \Delta_{c,-t} \Delta_c \overline{\varepsilon}_{c_j,a_{k+j-1},t_k} \ . \tag{26}$$

Let $\widehat{\alpha}_{c_j}$ denote the ordinary least squares estimator of $\widetilde{\alpha}_{c_j}$ from equation (26). That is

$$\widehat{\alpha}_{c_j} = \frac{1}{H_j} \sum_{k=\max(3-j,1)}^{\min(A-j+1,T-1)} \Delta_{c,-t} \Delta_c \overline{y}_{c_j,a_{k+j-1},t_k} \ , \tag{27}$$

where $H_j = \min(A - j + 1, T - 1) - \max(3 - j, 1) + 1$ is the number of times $\Delta_{c,-t} \Delta_c \overline{y}_{c_j,a_{k+j-1},t_k}$ is observed for a given c_j .The following theorem provides for consistent estimation of the change in slopes of the cohort effect profile.

Theorem 3 *Under Assumptions 1 and 2 and the data generating process given in (1), for all* $j = 4 - T, ..., A$, *as* $n_{c_1} \rightarrow \infty$, *for* T *fixed,* *(a) Consistency:*

$$\widehat{\alpha}_{c_j} \xrightarrow{p} \widetilde{\alpha}_{c_j} = \left(\alpha_{c_j} - \alpha_{c_{j-1}}\right) - \left(\alpha_{c_{j-1}} - \alpha_{c_{j-2}}\right) \ .$$

(b) Asymptotic Normality:

$$\sqrt{n_{c_1}}\left(\widehat{\alpha}_{c_j} - \widetilde{\alpha}_{c_j}\right) \xrightarrow{d} N\left(0, \pi_j\right),$$

$$\text{where } \pi_j = \left(\sigma_\omega^2 + \sigma_\eta^2\right) \frac{1}{H_j^2} \sum_{k=\max(3-j,1)}^{\min(A-j+1,T-1)} \left(\frac{1}{\delta_j} + \frac{2}{\delta_{j-1}} + \frac{1}{\delta_{j-2}}\right).$$

3.4 Wald tests of age, cohort and time effects

Now that we have consistent estimates of the changes in the slopes of the age, cohort, and time effect profiles, one can carry out Wald tests to test specific hypotheses about the shapes of these profiles. In particular, one may wish to test one of the following hypotheses:

i) $H_1 : \widetilde{\beta}_{a_3} = \widetilde{\beta}_{a_4} = ... = \widetilde{\beta}_{a_A}$

$\Leftrightarrow \left(\beta_{a_3} - \beta_{a_2}\right) - \left(\beta_{a_2} - \beta_{a_1}\right) = \left(\beta_{a_4} - \beta_{a3}\right) - \left(\beta_{a_3} - \beta_{a_2}\right) = ... = \left(\beta_{a_A} - \beta_{a_{A-1}}\right) - \left(\beta_{a_{A-1}} - \beta_{a_{A-2}}\right)$

ii) $H_2 : \widetilde{\beta}_{a_3} = \widetilde{\beta}_{a_4} = ... = \widetilde{\beta}_{a_A} = 0 \Leftrightarrow \left(\beta_{a_2} - \beta_{a_1}\right) = \left(\beta_{a_3} - \beta_{a_2}\right) = ... = \left(\beta_{a_A} - \beta_{a_{A-1}}\right)$

Testing H_1 enables one to see whether the change in the slope of the age effect function is constant or whether it changes over the range of ages considered, which can be considered a test of whether the age effect function is quadratic. The corresponding test applied to time effects will also enable one to determine whether there are any trend breaks in the time effect function. The second hypothesis, H_2, goes further, testing for linearity of the age effect. Failure to reject H_2 means failing to reject that the age effects can be replaced by a linear age term. The corresponding tests of whether all of the $\widetilde{\alpha}_{c_j}$ or all of the $\widetilde{\gamma}_{t_k}$ are zero likewise enables one to determine whether cohort effects or time effects are linear in the data.

In order to carry out this Wald test, stack the estimates $\widehat{\beta}_{a_{j+1}}$ from (13) to form the vector $\widehat{B} = \left(\widehat{\beta}_{a_3}, \widehat{\beta}_{a_4}, ..., \widehat{\beta}_{a_A}\right)'$, let $B = \left(\widetilde{\beta}_{a_3}, \widetilde{\beta}_{a_4}, ..., \widetilde{\beta}_{a_A}\right)'$, and let Ω_a be the $(A-2) \times (A-2)$ covariance matrix of the $\sqrt{n_{c_1}}\left(\widehat{\beta}_{a_{j+1}} - \widetilde{\beta}_{a_{j+1}}\right)$'s, with elements to be given shortly. Let $\widehat{\Omega}_a$ be the consistent estimator of Ω_a obtained by estimating $\left(\sigma_\omega^2+\sigma_\eta^2\right)$ by $\frac{1}{AT}\sum_{k=1}^{T}\sum_{j=1}^{A} var_c\left(y_{i,c_{j-k+2},a_{j+1},t_k}\right)$ and $\delta_{c_{j-k+2}}$ by $\frac{n_{c_{j-k+2}}}{n_{c_1}}$. The Wald test of the null hypothesis $H_0 : RB = r$, for a known $d \times (A-2)$ matrix R and $d \times 1$ vector r, is then given by its standard form:

$$W_{n_{c_1}} = n_{c_1}\left(R\widehat{B}-r\right)'\left(R\widehat{\Omega}_a R'\right)^{-1}\left(R\widehat{B}-r\right)$$

Theorem 4 *under Assumptions 1 and 2, as $n_{c_1}\to\infty$, for T fixed,*

(a) Ω_a *has elements* $\Omega_{j,h} = cov\left(\sqrt{n_{c_1}}\left(\widehat{\beta}_{a_{j+1}}-\widetilde{\beta}_{a_{j+1}}\right), \sqrt{n_{c_1}}\left(\widehat{\beta}_{a_{h+1}}-\widetilde{\beta}_{a_{h+1}}\right)\right)$,

for $j,h = 2,...,A-1$ *given by*

$$\Omega_{j,h} = \begin{cases} \frac{\left(\sigma_\omega^2+\sigma_\eta^2\right)}{(T-1)^2}\sum_{k=2}^{T}\left(\frac{2}{\delta_{j-k+2}}+\frac{2}{\delta_{j-k+1}}\right) \ \textit{if } h=j \\ \frac{-\left(\sigma_\omega^2+\sigma_\eta^2\right)}{(T-1)^2}\left[\sum_{k=2}^{T-1}\left(\frac{3}{\delta_{c_{j-k+2}}}+\frac{1}{\delta_{c_{j-k+1}}}\right)+\frac{2}{\delta_{c_{j-T+2}}}\right] \ \textit{if } h=j+1 \\ \frac{\left(\sigma_\omega^2+\sigma_\eta^2\right)}{(T-1)^2}\sum_{k=2}^{T-1}\frac{1}{\delta_{c_{j-k+2}}} \ \textit{if } h=j+2 \\ 0 \ \textit{otherwise} \end{cases}$$

(b) under $H_0 : RB = r$, $W_{n_{c_1}} \xrightarrow{d} \chi_d^2$.

Wald tests on the cohort effects and time effects can similarly be formulated in the standard way, and will also have the usual χ^2 distribution under the null hypothesis.

4 Identification with normalizations

The preceding section showed that with no restrictions on parameters, one can identify changes in the slopes of the age, cohort, and time effect functions. Next we show that with a normalization on one of the slopes of these effects, one can move from identifying the second derivatives to identifying the first derivatives or slopes of the age, cohort, and time effect functions. Recall from equation (8), we have for $k = 1,...,T-1$, and $j = 1,...,A-1$,

$$\Delta_t\overline{y}_{c_{j-k+1},a_{j+1},t_{k+1}} = \left(\beta_{a_{j+1}}-\beta_{a_j}\right)+\left(\gamma_{t_{k+1}}-\gamma_{t_k}\right)+\Delta_t\overline{\varepsilon}_{c_{j-k+1},a_{j+1},t_{k+1}} \ . \tag{28}$$

and generalizing equation (22), for $k = 1,...,T-1$, and $j = 1,...,A-1$,

$$\Delta_c \overline{y}_{c_{j-k+1},a_{j+1},t_{k+1}} = \left(\alpha_{c_{j-k+1}} - \alpha_{c_{j-k}}\right) + \left(\beta_{a_{j+1}} - \beta_{a_j}\right) + \Delta_c \overline{\varepsilon}_{c_{j-k+1},a_{j+1},t_{k+1}} \ . \tag{29}$$

From (28) and (29) it becomes clear that by normalizing any one particular slope for one effect, a normalization can then be recovered for the slopes of the remaining two effects. The Wald tests given above may be used to guide the normalization choice, and together with the sample context will determine which explicit normalization the researcher has the most confidence in making. For example, one could make the normalization $\beta_{a_2} = \beta_{a_1}$. With a large number of age groups, the difference in age effects between two successive ages may be relatively small, making this assumption more credible. More generally, we make the following normalization assumption:

Assumption 3 *(Normalization of Age Effects):* $\left(\beta_{a_{h+1}} - \beta_{a_h}\right) = \lambda$ *for some constant* λ *and a given* $h \in (1, A-1)$*.*

With this assumption, one can then recover all remaining slopes of the age effect profile using the estimated $\widehat{\beta}_{a_{j+1}}$'s. Letting $\widehat{b}_{a_j} \equiv \left(\widehat{\beta_{a_j} - \beta_{a_{j-1}}}\right)$ denote the estimator of the slope of the age profile between age a_{j-1} and a_j, we recover:

$$\begin{aligned} \widehat{b}_{a_{h-s}} &= \lambda - \sum_{m=0}^{s} \widehat{\beta}_{a_{h-m+1}} \text{ for } s = 0, 1, ..., j-2 \ , \\ \widehat{b}_{a_{h+s}} &= \lambda + \sum_{m=2}^{s} \widehat{\beta}_{a_{h+m}} \text{ for } s = 2, 3, ..., A-j \ . \end{aligned} \tag{30}$$

If the normalization made in Assumption 3 is true, then by Theorem 1, we have that $\widehat{b}_{a_j} \xrightarrow{p} \left(\beta_{a_j} - \beta_{a_{j-1}}\right)$ as $n_{c_1} \to \infty$. Recall that $\widehat{B} = \left(\widehat{\beta}_{a_3}, \widehat{\beta}_{a_4}, ..., \widehat{\beta}_{a_A}\right)'$, and so we can write $\widehat{b}_{a_j} = \lambda + m_j' \widehat{B}$, where m_j is a $(A-2) \times 1$ vector with elements of zeros, ones, and negative ones, as given in (30). Then from the proof of the Wald test, we have $\sqrt{n_{c_1}} \left(\widehat{B} - B\right) \xrightarrow{d} N\left(0, \Omega_a\right)$, and hence

$$\sqrt{n_{c_1}} \left(\widehat{b}_{a_{j+s}} - \left(\beta_{a_{j+s}} - \beta_{a_{j+s-1}}\right)\right) \xrightarrow{d} N\left(0, m_{j+s}' \Omega_a m_{j+s}\right) . \tag{31}$$

Substituting $\left(\beta_{a_{h+1}} - \beta_{a_h}\right) = \lambda$ into equations (28) and (29), one then obtains the following estimators of the slopes of the time and cohort effects for $k = 1, .., T-1$:

$$\begin{aligned} \widetilde{\left(\gamma_{t_{k+1}} - \gamma_{t_k}\right)} &= \Delta_t \overline{y}_{c_{h-k+1}, a_{h+1}, t_{k+1}} - \lambda \ , \\ \widetilde{\left(\alpha_{c_{h-k+1}} - \alpha_{c_{h-k}}\right)} &= \Delta_c \overline{y}_{c_{h-k+1}, a_{h+1}, t_{k+1}} - \lambda \ . \end{aligned} \tag{32}$$

However, more efficient estimators than those in (32) can be obtained by also using the estimated changes in age effects, $\widehat{b}_{a_j}$. Substituting the $\widehat{b}_{a_j}$ into (28) and (29) and rearranging gives for $k = 1, ..., T-1$, and $j = 1, ..., A-1$,

$$\Delta_t \overline{y}_{c_{j-k+1}, a_{j+1}, t_{k+1}} - \widehat{b}_{a_{j+1}} = \left(\gamma_{t_{k+1}} - \gamma_{t_k}\right) + \Delta_t \overline{\varepsilon}_{c_{j-k+1}, a_{j+1}, t_{k+1}} \tag{33}$$

$$\Delta_c \overline{y}_{c_{j-k+1}, a_{j+1}, t_{k+1}} - \widehat{b}_{a_{j+1}} = \left(\alpha_{c_{j-k+1}} - \alpha_{c_{j-k}}\right) + \Delta_c \overline{\varepsilon}_{c_{j-k+1}, a_{j+1}, t_{k+1}} \tag{34}$$

Let $\widehat{g}_{t_{k+1}} \equiv \widehat{\left(\gamma_{t_{k+1}} - \gamma_{t_k}\right)}$ and $\widehat{a}_{c_j} = \widehat{\left(\alpha_{c_j} - \alpha_{c_{j-1}}\right)}$ denote the least squares estimators of $\left(\gamma_{t_{k+1}} - \gamma_{t_k}\right)$ and $\left(\alpha_{c_j} - \alpha_{c_{j-1}}\right)$, based on $A-1$ and $H_{1j} = \min\left(T, A-j+1\right) - \max\left(1, 3-j\right) + 1$ cohort-level observations respectively. This can be done for a particular k, with the remaining slopes being recovered from the $\widehat{\gamma}_{t_k}$ and $\widehat{\alpha}_{c_j}$ through the time and cohort effect versions of equation (30). Alternatively, (33) and (34) can be used to obtain each cohort and time slope.

To identify the age, cohort, and time effects themselves, rather than just their changes, additional normalizations are required. From the basic specification in equation (4), it can be seen that normalizing two of the effects implicitly places a normalizing restriction on the third effect. For example, one could set the first time effect and the first cohort effect both equal to zero. Under these normalizations, one can recover estimates of the other time effects and cohort effects from the slope estimators $\widehat{g}_{t_{k+1}}$ and $\widehat{a}_{c_j}$. Letting $\alpha^*_{c_{j-k+1}}$ and $\gamma^*_{t_k}$ denote these estimates, one can then estimate the age effects from the following regression:

$$\overline{y}_{c_{j-k+1}, a_j, t_k} - \alpha^*_{c_{j-k+1}} - \gamma^*_{t_k} = \beta_{a_j} + \overline{\varepsilon}_{c_{j-k+1}, a_j, t_k} \ . \tag{35}$$

5 Empirical examples

5.1 Example 1: The effect of a crisis

We provide two applications to illustrate potential uses of this method. The first application is to determining whether there was a significant effect of the 1995 Mexican peso crisis on consumption. Figure 1 graphs Mexican household consumption by two-year birth cohort against two-year age group for households with heads aged 25-64. Cohort sample sizes range from 214 to 754, with a mean cohort sample size of 450 observations. The data are taken from the 1992, 1994, 1996, 1998 and 2000 Mexican ENIGH household surveys of income and expenditure, and are described in more detail in McKenzie (2003). The 1995 peso crisis resulted in large drops in income and consumption between 1994 and 1996. The standard hump-shaped age pattern can be seen, although there is a lot of noise around this, and it appears that at a given age, younger cohorts generally consume less than older cohorts did.

The top row of Figure 2 plots the estimated changes in the slopes of the age, cohort, and time effect profiles estimated using the methods developed in this paper. There are noticeable variations in the slopes of the age and cohort effect functions, suggesting that a quadratic in age and in cohort would not adequately capture these effects. The Wald test statistic for testing the hypothesis that the changes in the slope of the age profile are constant is 593.5, with 17 degrees of freedom. One thus overwhelmingly rejects that a quadratic captures the age effects present in the data. The time effects are easier to interpret. The slope of the time effect function has a growth effect on consumption, and hence the second derivative essentially captures a change in growth rate. The effects of the peso crisis are seen in a slowdown in the growth rate between 1994 and 1996, as compared to between 1992 and 1994, and are captured by the negative value of the second derivative in 1996. The subsequent recovery is seen in the positive coefficients on the second derivative in 1998 and 2000.

The second derivative of the age effect function is relatively constant around age 46, and coupled with life-cycle theory which suggests that this is a relatively stable period of life, we normalize by setting $\beta_{a_{46}} - \beta_{a_{44}} = 0$. With this one normalization, we arrive at the first derivatives shown in the second row of Figure 2. Making the further normalizations that the cohort effect at age 38 is zero, and the time effect in 2000 is zero, we arrive at the estimated profiles in the bottom row of Figure 2. The age and cohort effects are seen to offset each other to a degree, while the large time effects reflect well Mexico's macroeconomic performance over the period of study.

5.2 Example 2: Is the age effect profile convex or concave?

As a second example, we examine the shape of the age effects in the variance of log consumption in Taiwan, studied in Deaton and Paxson (1994). They show that an implication of the permanent income hypothesis is that consumption inequality should increase on average with age. However, they derive that the shape of the age-inequality profile should depend on the degree of persistence in shocks to earnings. In particular, under the permanent income hypothesis, they show that the age-inequality profile will be concave unless individual earnings contain a large stationary component. Using data from Taiwan, they estimate the following equation for the variance of log consumption, σ

$$\sigma_{c_{j-k+1},a_k,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + u_{c_{j-k+1},a_j,t_k} \tag{36}$$

Note that they omit year dummies in this model.[5] They estimate this model and after normalizing so that the fitted age effect at age 38 equals the average variance of log consumption for 38-year olds over all cohorts, plot the fitted age effects. The fitted profile of age effects looks convex, which is in contrast to their prior that the profile would be concave.

The methods outlined in this paper allow us to test formally whether the age profile is indeed convex or concave. A globally convex age profile would imply that the slope of the age profile is increasing with age, in which case we should find $\widetilde{\beta}_{a_j} > 0$ for all j. Estimation of the $\widetilde{\beta}_{a_j}$ allows us to test this hypothesis, and also test whether the age profile is quadratic or linear over certain ranges by means of Wald tests.

[5] Deaton and Paxson (1994) refer to an earlier working paper in which they normalize the time dummies to be orthogonal to a linear trend and to sum to zero, and say they obtained very similar results under this normalization.

We employ the same data as Deaton and Paxson (1994), consisting of annual data on real household consumption from 1976-96, taken from the Taiwanese Personal Income Distribution Surveys. One year cohorts are formed based on the birth year of the household head, and all heads aged 20-75 are considered. Approximately 300-400 heads of each age group are observed each year, although less than 100 observations are made on average for heads aged under 22 or above 66. Figure 3 plots the second differences in age effects estimated from equation (13) for each age group, together with an approximate 95 percent pointwise confidence interval about zero.[6] The effect of smaller cell sizes is seen in a widening of the confidence bands at very young and older age groups.

The point estimates in Figure 3 show approximately an equal number of positive and negative coefficients. However, the majority of these points lie within a 95 percent pointwise confidence interval about zero, with second differences in age effects taking on larger values for ages above 55, where the sample sizes start getting small. For household heads aged between 30 to 55, the second differences in age effects appear relatively equal. Testing formally for equality of the $\widetilde{\beta}_{a_j}$ for this age range using the Wald Test gives a p-value of 0.895, so that we can not reject the null of equality of second differences. As a result, a quadratic in age is at most all that is needed to model age effects over this range. We can test further to determine whether the $\widetilde{\beta}_{a_j}$ are jointly equal to zero in this age range. We can not reject this hypothesis for the age range 30-46 (Wald test p-value of 0.134), but reject it strongly once we extend to age range 30-55 (Wald test p-value of 0.001). Therefore for the age range of 30-46, a linear term in age is sufficient, and the age profile is neither strongly convex or concave over this range.

Figure 4 then shows the estimated age effects profile based on this analysis for age range 30 to 55. The estimated age dummies from equation (36) are plotted under the same normalization as Deaton and Paxson (1994) and show a convex age profile over this range. Our Wald tests reveal a quadratic in age to be sufficient over this age range, and so we reestimate equation (36) with a quadratic in age, again normalizing so that the fitted value at age 38 is the average of the variance of log consumption at this age. The quadratic curve closely fits the curve based on age dummies, as

[6] The confidence interval assumes normality of log consumption in order to be able to calculate the standard error of the sample variances.

the Wald tests would suggest. Finally, we fit equation (36) with only a linear term in age, estimated over the age range 30-46. Visually the curve based on age dummies closely tracks this fitted line over this age range, in accordance with our inability to reject the $\widetilde{\beta}_{a_j}$ being jointly equal to zero over this range.

This example therefore illustrates the ability of the methods in this paper to determine whether a general dummy variable specification for age effects can be replaced by a quadratic or linear term in age, and to determine whether or not the age profile is convex. Overall our results show that while the age profile has a convex quadratic shape over a wide age range, it can be well approximated by a linear age effect over a 16 year age range.

Note that the data generating process in equation (36) is given at the cohort level, since the dependent variable is the variance of log consumption for a given cohort in a given time period. As such, the estimates of $\widetilde{\beta}_{a_j}$ based on OLS estimation of this equation will be also be consistent. However, the method set out in this paper in Section 3 has at least two advantages over OLS estimation of equation (36) for estimating $\widetilde{\beta}_{a_j}$. This first, seen in this example, is that it automatically provides standard errors needed for testing hypotheses about the shape of the age profile. Secondly, the method provided by this paper will continue to be consistent when the data generating process is at the individual level and contains individual fixed effects, whereas equation (36) does not allow for fixed effects.

6 Conclusions

Our analysis shows that the linear dependence of age, cohort and time effects does not prevent the estimation of meaningful linear combinations of these effects without the need to impose further normalizing restrictions. Estimation of what is effectively the second derivative of each of the age, cohort, and time effect profiles provides some information about their shapes, clearly identifies trend breaks, and enables testing of quadratic and linear specifications to be done with respect to a very general alternative. This was illustrated by means of two examples. The first example showed the ability of the method to detect structural breaks in the time effect profile arising from the Mexican peso crisis. The second example showed how the method can be used to test for the convexity of an age effect profile, and to determine whether a quadratic specification in age is sufficient.

7 Mathematical proofs

7.1 Proof of Theorem 1

Proof. To prove part (a), recall that under Assumption 1,

$$\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} = \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \omega_{i(t_k),c_{j-k+2}} + \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \eta_{i(t_k),c_{j-k+2},a_{j+1},t_k}$$

Now as $n_{c_1} \to \infty$, under Assumption 2 $n_{c_{j-k+2}} \to \infty$, and then the Weak Law of Large Numbers (WLLN) for i.i.d. random variables gives that $\frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \omega_{i(t_k),c_{j-k+2}} \xrightarrow{p} E(\omega_i) = 0$ and $\frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \eta_{i(t_k),c_{j-k+2},a_{j+1},t_k} \xrightarrow{p} E\left(\eta_{i,c_{j-k+2},a_{j+1},t_k}\right) = 0$. Hence as $n_{c_{j-k+2}} \to \infty$,

$$\Delta_c \Delta_t \overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} = \overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} - \overline{\varepsilon}_{c_{j-k+2},a_j,t_{k-1}} - \overline{\varepsilon}_{c_{j-k+1},a_j,t_k} + \overline{\varepsilon}_{c_{j-k+1},a_{j-1},t_{k-1}} \xrightarrow{p} 0$$

and thus $\widehat{\beta}_{a_{j+1}} = \widetilde{\beta}_{a_{j+1}} + \frac{1}{T-1} \sum_{k=2}^{T} \Delta_c \Delta_t \overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} \xrightarrow{p} \widetilde{\beta}_{a_{j+1}}$.

To show (b), note that as different individuals are sampled each period, $\omega_{i(t_k),c_{j-k+2}}$ and $\omega_{i(t_s),c_{j-k+2}}$ are independent under Assumption 1 for all $s \neq k$, and so each $\varepsilon_{i(t_k),c_{j-k+2},a_{j+1},t_k}$ is i.i.d.$\left(0, \sigma_\omega^2 + \sigma_\eta^2\right)$. By the Lindeberg-Levy Central Limit Theorem,

$$\sqrt{n_{c_1}}\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} = \sqrt{\frac{n_{c_1}}{n_{c_{j-k+2}}}} \frac{1}{\sqrt{n_{c_{j-k+2}}}} \sum_{i=1}^{n_{c_{j-k+2}}} \varepsilon_{i(t_k),c_{j-k+2},a_{j+1},t_k} \xrightarrow{d} \sqrt{\frac{1}{\delta_{j-k+2}}} N\left(0, \sigma_\omega^2 + \sigma_\eta^2\right),$$

and similarly $\sqrt{n_{c_1}} \Delta_c \Delta_t \overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} \xrightarrow{d} N\left(0, \left(\sigma_\omega^2 + \sigma_\eta^2\right)\left(\frac{2}{\delta_{j-k+2}} + \frac{2}{\delta_{j-k+1}}\right)\right)$. The convergence of $\frac{1}{T-1} \sum_{k=2}^{T} \sqrt{n_{c_1}} \Delta_c \Delta_t \overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}$ follows under the independence assumptions. All age, cohort and time effects in (1) are constant when we consider only individuals in the same cohort in a given time period, and so taking cross-sectional variances, we have $var_c\left(y_{i,c_{j-k+1},a_j,t_k}\right) = var_c\left(\varepsilon_{i,c_{j-k+1},a_j,t_k}\right) = \sigma_\omega^2 + \sigma_\eta^2$. Convergence of the sample cross-sectional variance to the population cross-sectional variance follows from the assumption of existence of fourth moments made in Assumption 1.

7.2 Proof of Theorem 2

Proof. The proof of part (a) follows from the proof of part (a) of Theorem 1 as each $\overline{\varepsilon}_{c_{j-k+1},a_j,t_k} \xrightarrow{p} 0$ as $n_{c_1} \to \infty$. Part (b) follows the proof of part (b) of Theorem 1.

7.3 Proof of Theorem 3

Proof. The proof follows closely the proof of Theorem 1.

7.4 Proof of Theorem 4

Proof. Theorem 1 gives that $\widehat{\beta}_{a_{j+1}} \xrightarrow{p} \widetilde{\beta}_{a_{j+1}}$ as $n_{c_1} \to \infty$ for all j and thus that $\widehat{B} \xrightarrow{p} B$, $\sqrt{n_{c_1}}\left(\widehat{\beta}_{a_{j+1}} - \widetilde{\beta}_{a_{j+1}}\right) \xrightarrow{d} N(0,\ \sigma_{j+1})$, and $\widehat{\sigma}_{j+1} \xrightarrow{p} \sigma_{j+1}$. To determine the limiting covariance matrix, we need to evaluate the off-diagonal entries of Ω_a. For $h \neq j$, $h = 2, ..., A-1$ we therefore wish to evaluate

$$\begin{aligned} & E\left(\sqrt{n_{c_1}}\left(\widehat{\beta}_{a_{j+1}} - \widetilde{\beta}_{a_{j+1}}\right)\sqrt{n_{c_1}}\left(\widehat{\beta}_{a_{h+1}} - \widetilde{\beta}_{a_{h+1}}\right)\right) \\ = \quad & E\left(n_{c_1}\frac{1}{T-1}\sum_{k=2}^{T}\Delta_c\Delta_t\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\frac{1}{T-1}\sum_{s=2}^{T}\Delta_c\Delta_t\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right). \end{aligned} \tag{37}$$

Recall that

$$\Delta_c\Delta_t\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} = \overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} - \overline{\varepsilon}_{c_{j-k+2},a_j,t_{k-1}} - \overline{\varepsilon}_{c_{j-k+1},a_j,t_k} + \overline{\varepsilon}_{c_{j-k+1},a_{j-1},t_{k-1}}. \tag{38}$$

Under the i.i.d. assumptions made in Assumption 1, it is easily seen that

$$E\left(\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right) = 0 \text{ for all } (h,s) \neq (j,k)\ . \tag{39}$$

From (38) we have that

$$
\begin{aligned}
&E\left(\Delta_c\Delta_t\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\Delta_c\Delta_t\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right) = \\
&E\left(\left(\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} - \overline{\varepsilon}_{c_{j-k+2},a_j,t_{k-1}} - \overline{\varepsilon}_{c_{j-k+1},a_j,t_k} + \overline{\varepsilon}_{c_{j-k+1},a_{j-1},t_{k-1}}\right)\Delta_c\Delta_t\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right)
\end{aligned} \tag{40}
$$

Expanding out the first element of (40) and using (39) gives

$$
\begin{aligned}
& E\left(\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\Delta_c\Delta_t\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right) \\
= \; & E\left(\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\left(\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s} - \overline{\varepsilon}_{c_{h-s+2},a_h,t_{s-1}} - \overline{\varepsilon}_{c_{h-s+1},a_h,t_s} + \overline{\varepsilon}_{c_{h-s+1},a_{h-1},t_{s-1}}\right)\right) \\
= \; & \begin{cases}
-E\left(\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\overline{\varepsilon}_{c_{h-s+2},a_h,t_{s-1}}\right) = \frac{-\left(\sigma_\omega^2+\sigma_\eta^2\right)}{n_{c_{j-k+2}}} \text{ if } (h,s)=(j+1,k+1) \\
-E\left(\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\overline{\varepsilon}_{c_{h-s+1},a_h,t_s}\right) = \frac{-\left(\sigma_\omega^2+\sigma_\eta^2\right)}{n_{c_{j-k+2}}} \text{ if } (h,s)=(j+1,k) \\
E\left(\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\overline{\varepsilon}_{c_{h-s+1},a_{h-1},t_{s-1}}\right) = \frac{\left(\sigma_\omega^2+\sigma_\eta^2\right)}{n_{c_{j-k+2}}} \text{ if } (h,s)=(j+2,k+1) \\
0 \text{ otherwise}
\end{cases}
\end{aligned} \tag{41}
$$

Expanding out the other elements of (40) and combining their results with (41) then gives:

$$
\begin{aligned}
& E\left(\Delta_c\Delta_t\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\Delta_c\Delta_t\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right) \\
= & \begin{cases}
\frac{\left(\sigma_\omega^2+\sigma_\eta^2\right)}{n_{c_{j-k+2}}} \text{ if } (h,s)=(j+2,k+1) \\
-\left(\sigma_\omega^2+\sigma_\eta^2\right)\left(\frac{1}{n_{c_{j-k+2}}}+\frac{1}{n_{c_{j-k+1}}}\right) \text{ if } (h,s)=(j+1,k+1) \\
-\left(\sigma_\omega^2+\sigma_\eta^2\right)\frac{2}{n_{c_{j-k+2}}} \text{ if } (h,s)=(j+1,k) \\
-\left(\sigma_\omega^2+\sigma_\eta^2\right)\frac{2}{n_{c_{j-k+1}}} \text{ if } (h,s)=(j-1,k) \\
-\left(\sigma_\omega^2+\sigma_\eta^2\right)\left(\frac{1}{n_{c_{j-k+2}}}+\frac{1}{n_{c_{j-k+1}}}\right) \text{ if } (h,s)=(j-1,k-1) \\
\frac{\left(\sigma_\omega^2+\sigma_\eta^2\right)}{n_{c_{j-k+1}}} \text{ if } (h,s)=(j-2,k-1) \\
0 \text{ otherwise}
\end{cases}
\end{aligned} \tag{42}
$$

From (41) it follows that

$$
E\left(\Delta_c\Delta_t\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\sum_{s=2}^{T}\Delta_c\Delta_t\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right)
$$

$$= \begin{cases} \frac{(\sigma_\omega^2+\sigma_\eta^2)}{n_{c_{j-k+2}}} \text{ if } h=j+2,\ k\le T-1 \\ -\left(\sigma_\omega^2+\sigma_\eta^2\right)\left(\frac{3}{n_{c_{j-k+2}}}+\frac{1}{n_{c_{j-k+1}}}\right) \text{ if } h=j+1,\ k\le T-1 \\ -\left(\sigma_\omega^2+\sigma_\eta^2\right)\frac{2}{n_{c_{j-k+2}}} \text{ if } h=j+1,\ k=T \\ -\left(\sigma_\omega^2+\sigma_\eta^2\right)\frac{2}{n_{c_{j-k+1}}} \text{ if } h=j-1,\ k=2 \\ -\left(\sigma_\omega^2+\sigma_\eta^2\right)\left(\frac{1}{n_{c_{j-k+2}}}+\frac{3}{n_{c_{j-k+1}}}\right) \text{ if } h=j-1,\ k\ge 3 \\ \frac{(\sigma_\omega^2+\sigma_\eta^2)}{n_{c_{j-k+1}}} \text{ if } h=j-2,\ k\ge 3 \\ 0 \text{ otherwise} \end{cases} \tag{43}$$

and hence that

$$E\left(\sum_{k=2}^{T}\Delta_c\Delta_t\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\sum_{s=2}^{T}\Delta_c\Delta_t\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right)$$

$$= \begin{cases} \sum_{k=2}^{T-1}\frac{(\sigma_\omega^2+\sigma_\eta^2)}{n_{c_{j-k+2}}} \text{ if } h=j+2 \\ -\left(\sigma_\omega^2+\sigma_\eta^2\right)\left[\sum_{k=2}^{T-1}\left(\frac{3}{n_{c_{j-k+2}}}+\frac{1}{n_{c_{j-k+1}}}\right)+\frac{2}{n_{c_{j-T+2}}}\right] \text{ if } h=j+1 \\ 0 \text{ otherwise} \end{cases} \tag{44}$$

Using Assumption 3 and (44), we therefore have that as $n_{c_1}\to\infty$,

$$E\left(n_{c_1}\frac{1}{T-1}\sum_{k=2}^{T}\Delta_c\Delta_t\overline{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}\frac{1}{T-1}\sum_{s=2}^{T}\Delta_c\Delta_t\overline{\varepsilon}_{c_{h-s+2},a_{h+1},t_s}\right)$$

$$\xrightarrow{p} \begin{cases} \frac{1}{(T-1)^2}\sum_{k=2}^{T-1}\frac{(\sigma_\omega^2+\sigma_\eta^2)}{\delta_{c_{j-k+2}}} \text{ if } h=j+2 \\ \frac{-(\sigma_\omega^2+\sigma_\eta^2)}{(T-1)^2}\left[\sum_{k=2}^{T-1}\left(\frac{3}{\delta_{c_{j-k+2}}}+\frac{1}{\delta_{c_{j-k+1}}}\right)+\frac{2}{\delta_{c_{j-T+2}}}\right] \text{ if } h=j+1 \\ 0 \text{ otherwise} \end{cases} \tag{45}$$

References

Attanasio, O. (1993) 'A Cohort Analysis of Saving Behavior by U.S. Households', NBER Working Paper No. 4454.

Attanasio, O. (1998) 'Cohort Analysis of Saving Behavior by U.S. Households', *Journal of Human Resources* Vol 33, pp. 575-609.

Deaton, A. (1997) *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, The Johns Hopkins University Press: Baltimore.

Deaton, A. (1985) 'Panel Data from Time Series of Cross-sections', *Journal of Econometrics* Vol. 30, pp. 109-126.

Deaton, A. and Paxson, C. (1994). 'Intertemporal Choice and Inequality', *Journal of Political Economy* Vol. 102, pp. 437-467.

Denton, F., Mountain, D. and Spencer, B. (1999) 'Age, Trend, and Cohort Effects in a Macro Model of Canadian Expenditure Patterns', *Journal of Business and Economic Statistics* Vol. 17, pp. 430-443.

Foster, A. (1990) 'Cohort Analysis and Demographic Translation: A Comparative Study of Recent Trends in Age Specific Fertility Rates from Europe and North America', *Population Studies*, Vol. 44, pp. 287-315.

Heckman, J. and Robb, R. (1985) 'Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations', pp. 137-150 in WM. Mason and SE. Fienberg (eds.) *Cohort Analysis in Social Research: Beyond the Identification Problem*, Springer-Verlag: New York.

Heckman, J. and Vytlacil, E. (2001) 'Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return to Schooling', *The Review of Economics and Statistics* Vol. 83, pp. 1-12.

Japelli, T. (1999) 'The Age-Wealth Profile and the Life-Cycle Hypothesis: A Cohort Analysis with a Time Series of Cross-Sections of Italian Households', *Review of Income and Wealth* Vol 45, pp. 57-75.

Mason, W. and Fienberg, S. (1985) *Cohort Analysis in Social Research: Beyond the Identification Problem*, Springer-Verlag: New York.

McKenzie, D. (2004) 'Asymptotic theory for Heterogeneous Dynamic Pseudo-Panels', *Journal of Econometrics* Vol. 120, pp. 235-262.

McKenzie, D. (2003) 'How do Households Cope with Aggregate Shocks? Evidence from the Mexican Peso Crisis', *World Development* Vol. 31, pp. 1179-1199.

McKenzie, D. (2002) 'Disentangling Age, Cohort, and Time Effects in the Additive Model', *Stanford University Economics Department Working Paper* 02-009.

Figure 1: Household Consumption in Mexico 1992-2000 by Two-Year Cohort

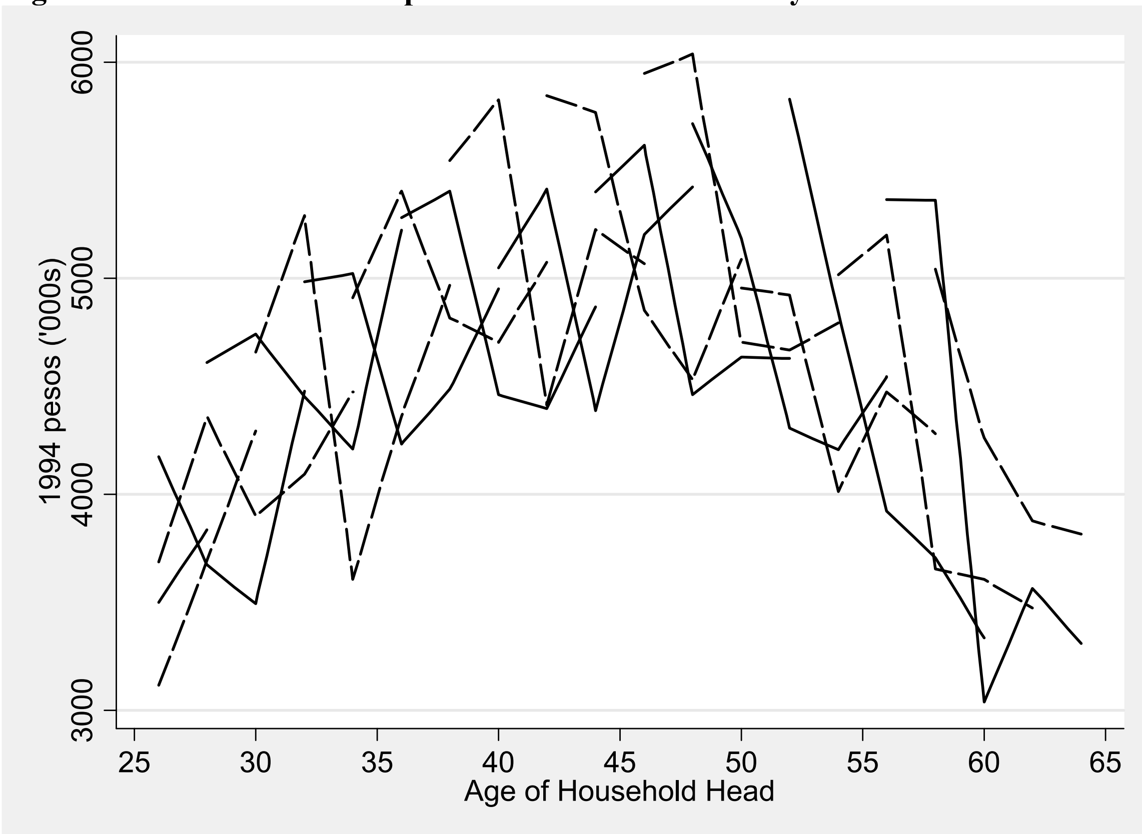

Figure 2: Estimated Age, Cohort and Time Effects for Mexican Household Consumption

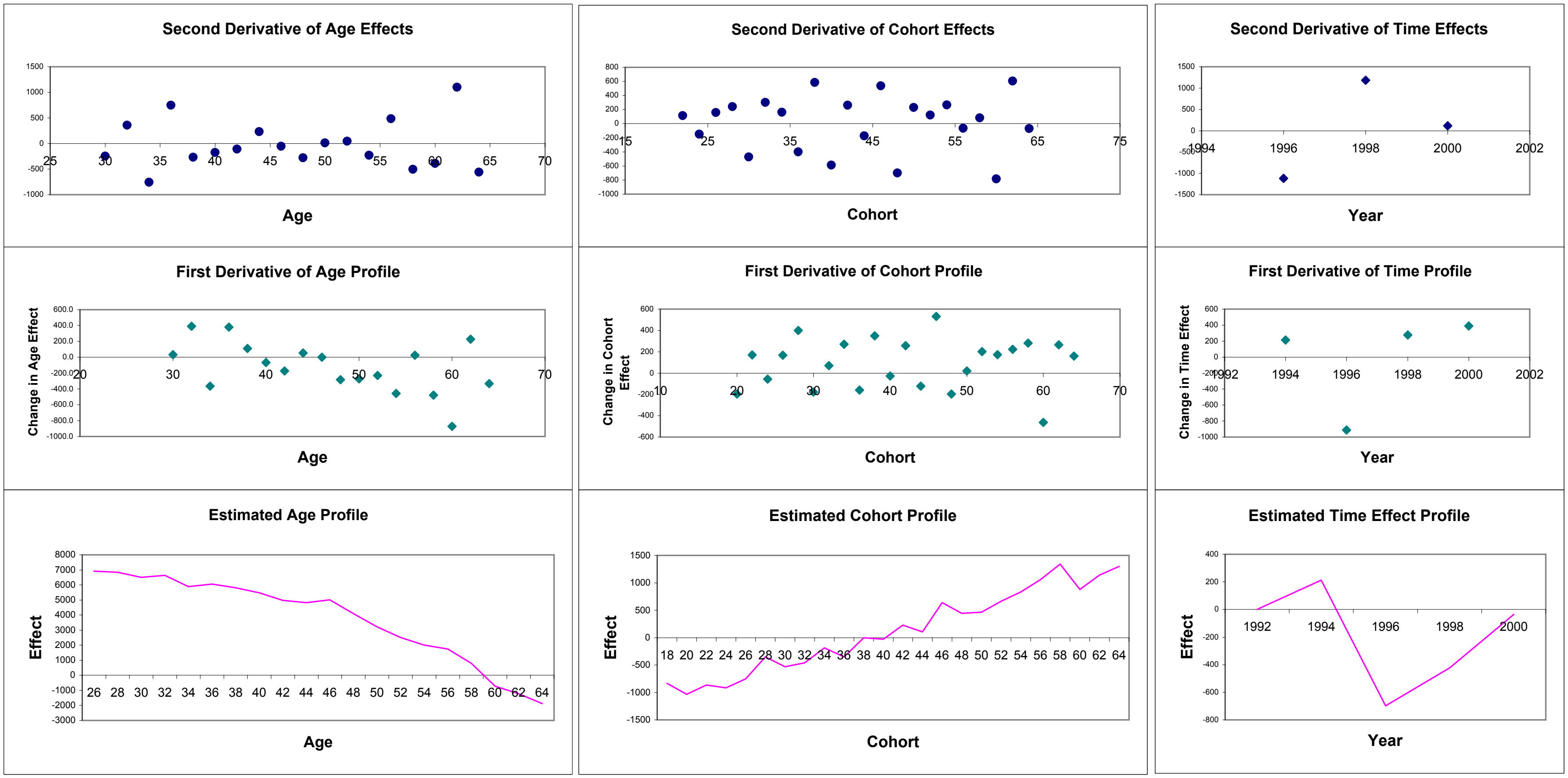

Figure 3: Second Differenced Age Effects for the Variance of Taiwanese Log Consumption

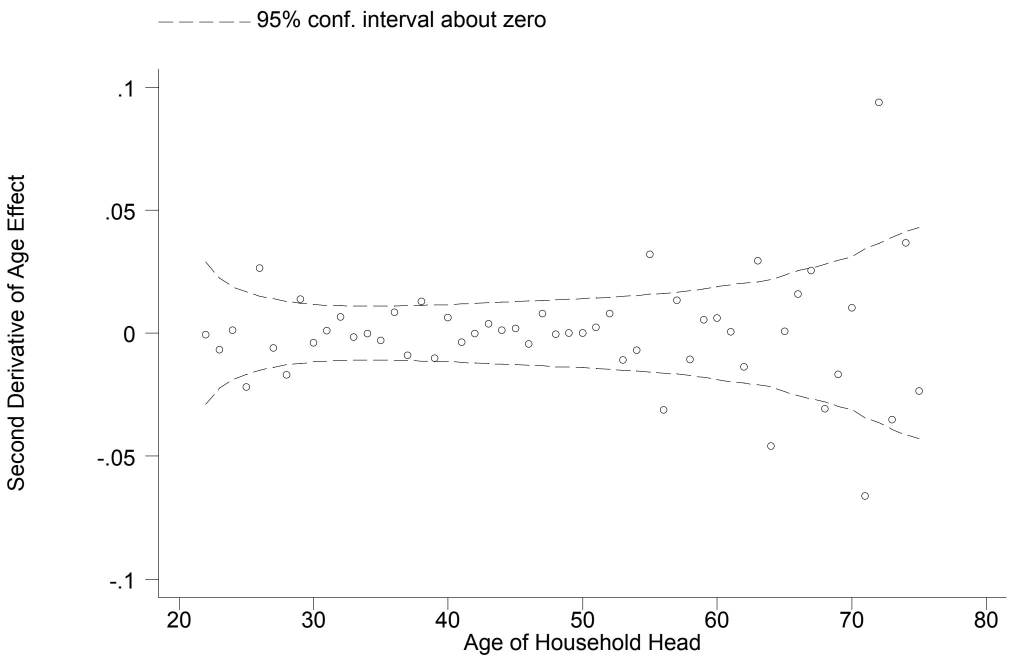

Figure 4: Estimated Age Effects in the Variance of Taiwanese Log Consumption under Different Functional Forms