

# Disentangling Age, Cohort, and Time Effects in the Additive Model

David J. McKenzie\*

*Stanford University*<sup>†</sup>

July 2002

## 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. First differenced and level effects can then be consistently estimated with a small number of additional normalizing assumptions. Moreover, it is demonstrated that coefficients on additional exogenous regressors can be consistently estimated in this framework without the need for normalizing assumptions.

*JEL Classifications:* C23, C81, J11.

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

---

\*Thanks to Luigi Pistaferri, Edward Vytlačil, and participants at seminars at Stanford University and the 2002 North American Summer Meeting of the Econometric Society for useful comments.

<sup>†</sup>*Address:* Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072. *Email:* david.mckenzie@stanford.edu

# 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. The method is illustrated with pseudo-panel data from Mexico, and clearly shows the time effect of the 1995 peso crisis on consumption, and that a quadratic in age is not sufficient for explaining the age effect.

As a second example application, we consider 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 and statistically insignificant from zero for most ages,

implying that convexity is not a strong feature of the data.

While in some settings the focus is on the cohort, age, and time effects themselves, in other applications the researcher merely wishes to control for these effects when determining the influence of additional regressors. A second contribution of the paper lies in showing that one can consistently estimate the coefficients on these additional regressors, without requiring any normalizing assumptions on the age, cohort and time effects, provided that the additional regressors in question exhibit sufficient variation over age, cohorts and time. The same methods proposed here are equally applicable with genuine panel data when it is available, and we finish by showing how our methods can be applied to this form of data.

The existing economics literature<sup>1</sup> 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 Vytlačil (2001) find no support for the widely accepted practice of imposing

---

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

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 is attributed to age and cohort effects. This approach has rightly become quite popular, in part due to its flexibility in capturing the effects of interest. The present paper allows greater flexibility still, while enabling the researcher to explicitly test whether the popular quadratic or linear trends can adequately capture specific effects. The ability to estimate the effects of additional regressors whilst controlling generally for age, cohort and time effects should also be of interest to the applied researcher.

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 introduces additional regressors to the base model and Section 7 shows how the same method can be used with genuine panel data. Section 8 concludes and mathematical proofs are presented in Section 9.

## 2 Model

Observations are made on individuals of  $A$  age groups,  $a_1, \dots, a_A$ , over  $T$  time periods,  $t_1, \dots, t_T$ . 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. The cohort of individuals aged  $a_j$  in time period  $t_k$  is denoted as cohort  $c_{j-k+1}$ . For example, cohort  $c_1$  is aged  $a_1$  at time  $t_1$ , while cohort  $c_A$  is aged  $a_A$  at time  $t_1$ . 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<sup>2</sup>:

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

The error term  $\varepsilon_{i,c_{j-k+1},a_j,t_k}$  is assumed to satisfy either

### Assumption 1

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

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

or

### Assumption 2

$$\varepsilon_{i,c_{j-k+1},a_j,t_k} = \omega_{i,c_{j-k+1}} + \nu_{c_{j-k+1},a_j,t_k} + \eta_{i,c_{j-k+1},a_j,t_k} \quad (3)$$

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

Note that Assumption 2 reduces to Assumption 1 if  $\sigma_\nu^2 = 0$ . This is the case if the error term consists of only individual idiosyncratic components, and contains no cohort-level variation.<sup>3</sup>

With repeated cross-sections, individuals are only observed once, and so we proceed by taking means of equation (1) over cohorts at each time period:

$$\frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} y_{i(t_k),c_{j-k+1},a_j,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + \gamma_{t_k} + \frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} \varepsilon_{i(t_k),c_{j-k+1},a_j,t_k} \quad (4)$$

<sup>2</sup>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.

<sup>3</sup>The independence assumptions made here simplify the presentation of the results in this paper. Consistent estimation is possible in pseudo-panels under certain forms of weak temporal and spatial correlation, and is discussed in the context of dynamic models in McKenzie (2001a).

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. Letting  $\bar{y}_{c_{j-k+1},a_j,t_k} = \frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} y_{i(t_k),c_{j-k+1},a_j,t_k}$  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

$$\bar{y}_{c_{j-k+1},a_j,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + \gamma_{t_k} + \bar{\varepsilon}_{c_{j-k+1},a_j,t_k} . \quad (5)$$

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 (5) is singular. That is, without further assumptions, one cannot separately identify age, cohort and time effects. In fact, even linear trends in all three effects cannot be separately identified.

### 3 Identification with no parameter restrictions:

#### Second Derivatives

##### 3.1 Age Effects

Consider equation (5) for cohort  $c_1$  at time periods  $t_1$  and  $t_2$ :

$$\bar{y}_{c_1,a_1,t_1} = \alpha_{c_1} + \beta_{a_1} + \gamma_{t_1} + \bar{\varepsilon}_{c_1,a_1,t_1} \quad (6)$$

$$\bar{y}_{c_1,a_2,t_2} = \alpha_{c_1} + \beta_{a_2} + \gamma_{t_2} + \bar{\varepsilon}_{c_1,a_2,t_2} \quad (7)$$

Subtracting (6) from (7) eliminates the cohort effect and gives

$$\Delta_t \bar{y}_{c_1,a_2,t_2} = (\beta_{a_2} - \beta_{a_1}) + (\gamma_{t_2} - \gamma_{t_1}) + \Delta_t \bar{\varepsilon}_{c_1,a_2,t_2} . \quad (8)$$

where  $\Delta_t \bar{y}_{c_1,a_2,t_2} \equiv \bar{y}_{c_1,a_2,t_2} - \bar{y}_{c_1,a_1,t_1}$  denotes the first time difference of  $\bar{y}_{c_1,a_2,t_2}$ , and similarly  $\Delta_t \bar{\varepsilon}_{c_1,a_2,t_2}$  is the first time differenced error term. 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,  $\Delta_{c,t}\bar{y}_{c_2,a_3,t_2} = \bar{y}_{c_2,a_3,t_2} - \bar{y}_{c_1,a_1,t_1}$  and  $\Delta_{-t}\bar{y}_{c_2,a_3,t_2} = \bar{y}_{c_2,a_3,t_2} - \bar{y}_{c_2,a_4,t_3}$ . Similarly, time differencing the observations for cohort  $c_2$  between time periods  $t_1$  and  $t_2$  gives:

$$\Delta_t\bar{y}_{c_2,a_3,t_2} = (\beta_{a_3} - \beta_{a_2}) + (\gamma_{t_2} - \gamma_{t_1}) + \Delta_t\bar{\varepsilon}_{c_2,a_3,t_2} . \quad (9)$$

Now subtracting (8) from (9) eliminates the differenced time effect and yields:

$$\Delta_c\Delta_t\bar{y}_{c_2,a_3,t_2} = (\beta_{a_3} - \beta_{a_2}) - (\beta_{a_2} - \beta_{a_1}) + \Delta_c\Delta_t\bar{\varepsilon}_{c_2,a_3,t_2} \quad (10)$$

where  $\Delta_c\Delta_t\bar{y}_{c_2,a_3,t_2} \equiv \Delta_t\bar{y}_{c_2,a_3,t_2} - \Delta_t\bar{y}_{c_1,a_2,t_2}$  denotes the first cohort difference of  $\Delta_t\bar{y}_{c_2,a_3,t_2}$ .

Likewise, for cohort  $c_j$ ,  $j = 2, \dots, A - 1$  we have at time period  $t_2$ :

$$\Delta_c\Delta_t\bar{y}_{c_j,a_{j+1},t_2} = (\beta_{a_{j+1}} - \beta_{a_j}) - (\beta_{a_j} - \beta_{a_{j-1}}) + \Delta_c\Delta_t\bar{\varepsilon}_{c_j,a_{j+1},t_2} \quad (11)$$

More generally, taking time differences between periods  $t_k$  and  $t_{k-1}$  for cohort  $c_{j-k+2}$ , gives for  $j = 2, \dots, A - 1$ ;  $k = 2, \dots, T$ :

$$\Delta_c\Delta_t\bar{y}_{c_{j-k+2},a_{j+1},t_k} = (\beta_{a_{j+1}} - \beta_{a_j}) - (\beta_{a_j} - \beta_{a_{j-1}}) + \Delta_c\Delta_t\bar{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} . \quad (12)$$

Defining  $\tilde{\beta}_{a_{j+1}} = (\beta_{a_{j+1}} - \beta_{a_j}) - (\beta_{a_j} - \beta_{a_{j-1}})$ , we arrive at the following regression, for  $j = 2, \dots, A - 1$ ;  $k = 2, \dots, T$ :

$$\Delta_c\Delta_t\bar{y}_{c_{j-k+2},a_{j+1},t_k} = \tilde{\beta}_{a_{j+1}} + \Delta_c\Delta_t\bar{\varepsilon}_{c_{j-k+2},a_{j+1},t_k} . \quad (13)$$

Let  $\hat{\beta}_{a_{j+1}}$  denote the ordinary least squares estimator of  $\tilde{\beta}_{a_{j+1}}$  from equation (13). That is

$$\hat{\beta}_{a_{j+1}} = \frac{1}{T-1} \sum_{k=2}^T \Delta_c\Delta_t\bar{y}_{c_{j-k+2},a_{j+1},t_k} \quad (14)$$

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

The following theorem then applies:

**Theorem 1** For the data generating process given in (1),

(a) under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,

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

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

(b) under Assumption 2, as  $T \rightarrow \infty$ , for  $n_{c_s}$  fixed for all  $s = 2 - T, \dots, A$ ,

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

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

(c) (Corollary) Under Assumptions 1 and 3, (i)  $\widehat{\beta}_{a_{j+1}} \xrightarrow{p} \widetilde{\beta}_{a_{j+1}}$  in sequential limit as  $(n_{c_1}, T \rightarrow \infty)_{seq}$

and under Assumptions 2 and 3, (ii)  $\widehat{\beta}_{a_{j+1}} \xrightarrow{p} \widetilde{\beta}_{a_{j+1}}$  in sequential limit as  $(T, n_{c_1} \rightarrow \infty)_{seq}$  and

(iii)  $\widehat{\beta}_{a_{j+1}} \xrightarrow{p} \widetilde{\beta}_{a_{j+1}}$  in sequential limit as  $(n_{c_1}, T \rightarrow \infty)_{seq}$ .

(d) under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,

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

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

and letting  $\text{var}_c(y_{i,c_{j-k+2},a_{j+1},t_k}) \equiv \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \left( y_{i,c_{j-k+2},a_{j+1},t_k} - \bar{y}_{c_{j-k+2},a_{j+1},t_k} \right)^2$  denote



the cross-sectional sample variance across individuals in cohort  $c_{j-k+2}$  at time  $t_k$ , we have  $\text{var}_c (y_{i,c_{j-k+2},a_{j+1},t_k}) \xrightarrow{P} (\sigma_\omega^2 + \sigma_\eta^2)$ .

(e) under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,

$$s^2 = \frac{1}{T-2} \sum_{k=2}^T \left( \Delta_c \Delta_t \bar{y}_{c_{j-k+2},a_{j+1},t_k} - \widehat{\beta}_{a_{j+1}} \right)^2 \xrightarrow{P} 0 .$$

**Remarks:**

(a) It is thus possible to identify changes in the slopes of the age effects profile without any parameter restrictions. In essence, one obtains the second derivative of the age effect function. For example,  $\widetilde{\beta}_{a_3} = (\beta_{a_3} - \beta_{a_2}) - (\beta_{a_2} - \beta_{a_1})$  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  $a_1$ . If the age effects are linear, these terms should all be zero.

(b) In most practical settings, the number of individuals per cohort,  $n_c$ , will be large relative to the number of time periods, hence the use of  $n_c \rightarrow \infty$  asymptotics will be most appropriate. For fixed  $T$ , this requires that the error terms  $\Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2},a_{j+1},t_k}$  converge to zero as  $n_c \rightarrow \infty$ , which is only the case if  $\sigma_\nu^2 = 0$ , that is if there are only individual level components in the error term. Such a condition is a common restriction in pseudo-panel estimation with a fixed number of time periods. For example, Verbeek (1995) notes that such a condition is required for consistency of the pseudo-panel estimator of Deaton (1985) as  $n_{c_1} \rightarrow \infty$  for  $T$  fixed.

(c) If  $T$  is also large, then part (b) shows that one can allow for cohort-level error components and the estimator in (14) is still consistent. However, an implicit assumption in equation (1) is that the age effects are independent of time. This assumption is more credible over a small number of time periods, and as  $T$  gets large, one may wish to allow the age effects themselves to depend on time. If one wishes to do this generally, then again one is forced to rely on  $n_{c_1} \rightarrow \infty$  asymptotics and assume that  $\sigma_\nu^2 = 0$ .

(d) Part (d) 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

$$\widetilde{\sigma}_{j+1} = 2 \text{var}_c \left( y_{i, c_{j-k+2}, a_{j+1}, t_k} \right) \frac{1}{(T-1)^2} \sum_{k=2}^T \left( \frac{n_{c_1}}{n_{c_{j-k+2}}} + \frac{n_{c_1}}{n_{c_{j-k+1}}} \right). \quad (15)$$

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:

$$\widehat{\sigma}_{j+1} = 2 \frac{1}{(T-1)^2} \sum_{k=2}^T \left( \frac{n_{c_1}}{n_{c_{j-k+2}}} + \frac{n_{c_1}}{n_{c_{j-k+1}}} \right) \frac{1}{AT} \sum_{k=1}^T \sum_{j=1}^A \text{var}_c \left( y_{i, c_{j-k+2}, a_{j+1}, t_k} \right). \quad (16)$$

- (e) Note that although the estimator  $\widehat{\beta}_{a_{j+1}}$  is obtained by running OLS on (13), part (e) shows that the usual OLS standard errors will not be correct. Hence the need to use cross-sectional sample variances as in (16).

### 3.2 Time Effects

Time differencing the observations for cohort  $c_0$  between time periods  $t_2$  and  $t_3$  gives

$$\Delta_t \bar{y}_{c_0, a_2, t_3} = (\beta_{a_2} - \beta_{a_1}) + (\gamma_{t_3} - \gamma_{t_2}) + \Delta_t \bar{\varepsilon}_{c_0, a_2, t_3} \quad (17)$$

Subtracting equation (8) from (17) eliminates the age effects, giving

$$\Delta_{-c, t} \Delta_t \bar{y}_{c_0, a_2, t_3} = (\gamma_{t_3} - \gamma_{t_2}) - (\gamma_{t_2} - \gamma_{t_1}) + \Delta_{-c, t} \Delta_t \bar{\varepsilon}_{c_0, a_2, t_3}, \quad (18)$$

where  $\Delta_{-c, t} \Delta_t \bar{y}_{c_0, a_2, t_3} \equiv \Delta_t \bar{y}_{c_0, a_2, t_3} - \Delta_t \bar{y}_{c_1, a_2, t_2}$  and  $\Delta_{-c, t} \Delta_t \bar{\varepsilon}_{c_0, a_2, t_3} = \Delta_t \bar{\varepsilon}_{c_0, a_2, t_3} - \Delta_t \bar{\varepsilon}_{c_1, a_2, t_2}$ .

Likewise, for cohort  $c_{j-2}$ ,  $j = 2, \dots, A$ , in time period  $t_3$  we have

$$\Delta_{-c, t} \Delta_t \bar{y}_{c_{j-2}, a_j, t_3} = (\gamma_{t_3} - \gamma_{t_2}) - (\gamma_{t_2} - \gamma_{t_1}) + \Delta_{-c, t} \Delta_t \bar{\varepsilon}_{c_{j-2}, a_j, t_3} \quad (19)$$

More generally still, for cohort  $c_{j-k+1}$ ,  $j = 2, \dots, A$ , in time period  $t_k$ ,  $k = 3, \dots, T$ ,

$$\Delta_{-c,t} \Delta_t \bar{y}_{c_{j-k+1}, a_j, t_k} = \left( \gamma_{t_k} - \gamma_{t_{k-1}} \right) - \left( \gamma_{t_{k-1}} - \gamma_{t_{k-2}} \right) + \Delta_{-c,t} \Delta_t \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} \quad (20)$$

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

$$\Delta_{-c,t} \Delta_t \bar{y}_{c_{j-k+1}, a_j, t_k} = \tilde{\gamma}_{t_k} + \Delta_{-c,t} \Delta_t \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} . \quad (21)$$

Let  $\hat{\gamma}_{t_k}$  denote the ordinary least squares estimator of  $\tilde{\gamma}_{t_k}$  from equation (21). That is

$$\hat{\gamma}_{t_k} = \frac{1}{A-1} \sum_{j=2}^A \Delta_{-c,t} \Delta_t \bar{y}_{c_{j-k+1}, a_j, t_k} \quad (22)$$

The following theorem then shows that consistent estimation of changes in the slopes of the time effect profile is possible under appropriate assumptions.

**Theorem 2** *For the data generating process given in (1), for all  $k = 2, \dots, T$ ,*

(a) *under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,*

$$\hat{\gamma}_{t_k} \xrightarrow{P} \tilde{\gamma}_{t_k} = \left( \gamma_{t_k} - \gamma_{t_{k-1}} \right) - \left( \gamma_{t_{k-1}} - \gamma_{t_{k-2}} \right) .$$

(b) *under Assumption 2, as  $A \rightarrow \infty$ , for  $n_{c_s}$  fixed for all  $s = 2 - T, \dots, A$ ,  $\hat{\gamma}_{t_k} \xrightarrow{P} \tilde{\gamma}_{t_k}$ .*

(c) *(Corollary) Under Assumptions 1 and 3, (i)  $\hat{\gamma}_{t_k} \xrightarrow{P} \tilde{\gamma}_{t_k}$  in sequential limit as  $(n_{c_1}, A \rightarrow \infty)_{seq}$  and, under Assumptions 2 and 3, (ii)  $\hat{\gamma}_{t_k} \xrightarrow{P} \tilde{\gamma}_{t_k}$  in sequential limit as  $(A, n_{c_1} \rightarrow \infty)_{seq}$  and (iii)  $\hat{\gamma}_{t_k} \xrightarrow{P} \tilde{\gamma}_{t_k}$  in sequential limit as  $(n_{c_1}, A \rightarrow \infty)_{seq}$ .*

(d) *under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,*

$$\sqrt{n_{c_1}} (\hat{\gamma}_{t_k} - \tilde{\gamma}_{t_k}) \xrightarrow{d} N(0, \kappa_k) ,$$

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

**Remarks:**

- (a) While asymptotics as the number of age groups  $A \rightarrow \infty$  may seem strange, recall that the number of cohorts  $C = A + T - 1$ . Hence one can reinterpret parts (b) and (c) as giving asymptotic results as the number of cohorts,  $C$ , tends to infinity. Asymptotics as  $C \rightarrow \infty$  is reasonably common in pseudo-panel econometric work.<sup>4</sup> However, in many practical situations it is likely to be the case that the number of individuals per cohort,  $n_c$ , is larger than the number of cohorts,  $C$ , meaning that the asymptotics in either part (a) of Theorem 2, or in part (c) as  $(n_{c_1}, A \rightarrow \infty)_{seq}$  are likely to be more appropriate.
- (b) 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.
- (c) The variance  $\kappa_k$  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 changes in cohort effects, first consider (5) for cohort  $c_2$  at time period  $t_2$ :

$$\bar{y}_{c_2, a_3, t_2} = \alpha_{c_2} + \beta_{a_3} + \gamma_{t_2} + \bar{\varepsilon}_{c_2, a_3, t_2} . \tag{23}$$

Subtracting (7) from (23) eliminates the time effects and gives:

$$\Delta_c \bar{y}_{c_2, a_3, t_2} = (\alpha_{c_2} - \alpha_{c_1}) + (\beta_{a_3} - \beta_{a_2}) + \Delta_c \bar{\varepsilon}_{c_2, a_3, t_2} , \tag{24}$$

---

<sup>4</sup>For example, see Verbeek and Nijman (1993) and Collado (1997).

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

$$\Delta_c \bar{y}_{c_3, a_3, t_1} = (\alpha_{c_3} - \alpha_{c_2}) + (\beta_{a_3} - \beta_{a_2}) + \Delta_c \bar{\varepsilon}_{c_3, a_3, t_2} . \quad (25)$$

Subtracting (24) from (25) eliminates the age effects, giving

$$\Delta_{c,-t} \Delta_c \bar{y}_{c_3, a_3, t_1} = (\alpha_{c_3} - \alpha_{c_2}) - (\alpha_{c_2} - \alpha_{c_1}) + \Delta_{c,-t} \Delta_c \bar{\varepsilon}_{c_3, a_3, t_1} , \quad (26)$$

where  $\Delta_{c,-t} \Delta_c \bar{y}_{c_3, a_3, t_1} \equiv \Delta_c \bar{y}_{c_3, a_3, t_1} - \Delta_c \bar{y}_{c_2, a_3, t_2}$ . For  $k = 1, \dots, \min(A-2, T-1)$  this extends to

$$\Delta_{c,-t} \Delta_c \bar{y}_{c_3, a_{k+2}, t_k} = (\alpha_{c_3} - \alpha_{c_2}) - (\alpha_{c_2} - \alpha_{c_1}) + \Delta_{c,-t} \Delta_c \bar{\varepsilon}_{c_3, a_{k+2}, t_k} . \quad (27)$$

and for cohort  $j$ ,  $j = 4 - T, \dots, A$ ,  $k = 1, \dots, T - 1$ , such that  $2 \leq k + j - 1 \leq A$ ,

$$\Delta_{c,-t} \Delta_c \bar{y}_{c_j, a_{k+j-1}, t_k} = (\alpha_{c_j} - \alpha_{c_{j-1}}) - (\alpha_{c_{j-1}} - \alpha_{c_{j-2}}) + \Delta_{c,-t} \Delta_c \bar{\varepsilon}_{c_j, a_{k+j-1}, t_k} . \quad (28)$$

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

$$\Delta_{c,-t} \Delta_c \bar{y}_{c_j, a_{k+j-1}, t_k} = \tilde{\alpha}_{c_j} + \Delta_{c,-t} \Delta_c \bar{\varepsilon}_{c_j, a_{k+j-1}, t_k} . \quad (29)$$

Let  $\hat{\alpha}_{c_j}$  denote the ordinary least squares estimator of  $\tilde{\alpha}_{c_j}$  from equation (29). That is

$$\hat{\alpha}_{c_j} = \frac{1}{H_j} \sum_{k=\max(3-j, 1)}^{\min(A-j+1, T-1)} \Delta_{c,-t} \Delta_c \bar{y}_{c_j, a_{k+j-1}, t_k} , \quad (30)$$

where  $H_j = \min(A - j + 1, T - 1) - \max(3 - j, 1) + 1$  is the number of times  $\Delta_{c,-t} \Delta_c \bar{y}_{c_j, a_{k+j-1}, t_k}$  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** For the data generating process given in (1), for all  $j = 4 - T, \dots, A$ ,

(a) under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,

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

(b) under Assumption 2, as  $A \rightarrow \infty$  and  $T \rightarrow \infty$ , for  $n_{c_s}$  fixed for all  $s = 2 - T, \dots, A$ ,  $\widehat{\alpha}_{c_j} \xrightarrow{P} \widetilde{\alpha}_{c_j}$ .

(c) (Corollary) Under Assumptions 1 and 3, (i)  $\widehat{\alpha}_{c_j} \xrightarrow{P} \widetilde{\alpha}_{c_j}$  in sequential limit as  $(n_{c_1}, T \rightarrow \infty)_{seq}$

and, under Assumptions 2 and 3, (ii)  $\widehat{\alpha}_{c_j} \xrightarrow{P} \widetilde{\alpha}_{c_j}$  in sequential limit as  $(n_{c_1}, T \rightarrow \infty)_{seq}$  and

(iii)  $\widehat{\alpha}_{c_j} \xrightarrow{P} \widetilde{\alpha}_{c_j}$  in sequential limit as  $(A \rightarrow \infty$  and  $T \rightarrow \infty, n_{c_1} \rightarrow \infty)_{seq}$ .

(d) under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,

$$\begin{aligned} & \sqrt{n_{c_1}} (\widehat{\alpha}_{c_j} - \widetilde{\alpha}_{c_j}) \xrightarrow{d} N(0, \pi_j), \\ \text{where } \pi_j &= (\sigma_\omega^2 + \sigma_\eta^2) \frac{1}{H_j^2} \sum_{k=\max(3-j,1)}^{\min(A-j+1, T-1)} \left( \frac{1}{\delta_j} + \frac{2}{\delta_{j-1}} + \frac{1}{\delta_{j-2}} \right). \end{aligned}$$

**Remark:**

Part (b) shows that if the number of individuals per cohort is held fixed, both the number of time periods and the number of age groups need to pass to infinity for consistency. This arises as the age restriction causes the number of times a given cohort is observed to depend on both the number of age groups and the number of time periods.

### 3.4 Wald Tests of Age, Cohort and Time Effects

In addition to consistently estimating changes in the slopes of the age, cohort and time effect profiles, one can also 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:

$$\text{i) } H_{10} : \widetilde{\beta}_{a_3} = \widetilde{\beta}_{a_4} = \dots = \widetilde{\beta}_{a_A}$$

$$\Leftrightarrow (\beta_{a_3} - \beta_{a_2}) - (\beta_{a_2} - \beta_{a_1}) = (\beta_{a_4} - \beta_{a_3}) - (\beta_{a_3} - \beta_{a_2}) = \dots = (\beta_{a_A} - \beta_{a_{A-1}}) - (\beta_{a_{A-1}} - \beta_{a_{A-2}})$$

$$\text{ii) } H_{20} : \tilde{\beta}_{a_3} = \tilde{\beta}_{a_4} = \dots = \tilde{\beta}_{a_A} = 0 \Leftrightarrow (\beta_{a_2} - \beta_{a_1}) = (\beta_{a_3} - \beta_{a_2}) = \dots = (\beta_{a_A} - \beta_{a_{A-1}})$$

Testing  $H_{10}$  enables one to see whether the change in the slope of the age effect function itself 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,  $H_{20}$ , goes further, testing for linearity of the age effect. Failure to reject  $H_{20}$  means failing to reject that the age effects can be replaced by a linear age term. The corresponding tests of whether all of the  $\tilde{\alpha}_{c_j}$  or all of the  $\tilde{\gamma}_{t_k}$  are zero likewise enables one to determine whether cohort effects or time effects are linear in the data.

Stack the estimates  $\hat{\beta}_{a_{j+1}}$  from (14) to form the vector  $\hat{B} = (\hat{\beta}_{a_3}, \hat{\beta}_{a_4}, \dots, \hat{\beta}_{a_A})'$ , let  $B = (\tilde{\beta}_{a_3}, \tilde{\beta}_{a_4}, \dots, \tilde{\beta}_{a_A})'$ , and  $\Omega_a$  be the  $(A-2) \times (A-2)$  covariance matrix of the  $\sqrt{n_{c_1}} (\hat{\beta}_{a_{j+1}} - \tilde{\beta}_{a_{j+1}})$ 's, with elements to be given shortly. Let  $\hat{\Omega}_a$  be the consistent estimator of  $\Omega_a$  obtained by estimating  $(\sigma_\omega^2 + \sigma_\eta^2)$  by  $\frac{1}{AT} \sum_{k=1}^T \sum_{j=1}^A \text{var}_c (y_{i, c_{j-k+2}, a_{j+1}, t_k})$  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:

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

**Theorem 4** *under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,*

$$(a) \Omega_a \text{ has elements } \Omega_{j,h} = \text{cov} \left( \sqrt{n_{c_1}} (\hat{\beta}_{a_{j+1}} - \tilde{\beta}_{a_{j+1}}), \sqrt{n_{c_1}} (\hat{\beta}_{a_{h+1}} - \tilde{\beta}_{a_{h+1}}) \right),$$

for  $j, h = 2, \dots, A-1$  given by

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

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

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

## 4 Identification with Normalizations

### 4.1 One Normalization: First Derivatives

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 of the age, cohort, and time effect profiles to identifying the first derivatives. That is, one normalizing assumption will enable the actual slopes of the age, cohort, and time effect functions to be identified. Generalizing equation (8), we have for  $k = 1, \dots, T - 1$ , and  $j = 1, \dots, A - 1$ ,

$$\Delta_t \bar{y}_{c_{j-k+1}, a_{j+1}, t_{k+1}} = \left( \beta_{a_{j+1}} - \beta_{a_j} \right) + \left( \gamma_{t_{k+1}} - \gamma_{t_k} \right) + \Delta_t \bar{\varepsilon}_{c_{j-k+1}, a_{j+1}, t_{k+1}} . \quad (31)$$

and generalizing equation (24), for  $k = 1, \dots, T - 1$ , and  $j = 1, \dots, A - 1$ ,

$$\Delta_c \bar{y}_{c_{j-k+1}, a_{j+1}, t_{k+1}} = \left( \alpha_{c_{j-k+1}} - \alpha_{c_{j-k}} \right) + \left( \beta_{a_{j+1}} - \beta_{a_j} \right) + \Delta_c \bar{\varepsilon}_{c_{j-k+1}, a_{j+1}, t_{k+1}} . \quad (32)$$

From (31) and (32) 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. In many economic applications, the number of time periods is relatively small, and the time effects may be believed to vary greatly from period to period.<sup>5</sup> In contrast, with a reasonable number of age groups and cohorts, one may be more willing to defend an explicit normalization for one of these effects. 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:

---

<sup>5</sup>On the other hand, the researcher may wish to use information from other sources, such as macroeconomic statistics, to argue that time effects are relatively constant in a certain period. The procedure given here can be easily adapted for the case where the normalizing assumption is on the time effects.



**Assumption 4** (*Normalization of Age Effects*):  $(\beta_{a_{h+1}} - \beta_{a_h}) = \lambda$  for some constant  $\lambda$  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  $\widehat{b}_{a_j} \equiv (\widehat{\beta}_{a_j} - \widehat{\beta}_{a_{j-1}})$  denote the estimator of the slope of the age profile between age  $a_{j-1}$  and  $a_j$ , we recover:

$$\begin{aligned}\widehat{b}_{a_{h-s}} &= \lambda - \sum_{m=0}^s \widehat{\beta}_{a_{h-m+1}} \text{ for } s = 0, 1, \dots, j-2, \\ \widehat{b}_{a_{h+s}} &= \lambda + \sum_{m=2}^s \widehat{\beta}_{a_{h+m}} \text{ for } s = 2, 3, \dots, A-j.\end{aligned}\quad (33)$$

If the normalization made in Assumption 4 is true, then under the conditions for consistency of the  $\widehat{\beta}_{a_{j+1}}$  given in Theorem 1, we have that  $\widehat{b}_{a_j} \xrightarrow{p} (\beta_{a_j} - \beta_{a_{j-1}})$  as either  $n_{c_1} \rightarrow \infty$  or  $T \rightarrow \infty$  (or both). Recall that  $\widehat{B} = (\widehat{\beta}_{a_3}, \widehat{\beta}_{a_4}, \dots, \widehat{\beta}_{a_A})'$ , 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 (33). Then from the proof of the Wald test, we have  $\sqrt{n_{c_1}} (\widehat{B} - B) \xrightarrow{d} N(0, \Omega_a)$ , and hence

$$\sqrt{n_{c_1}} (\widehat{b}_{a_{j+s}} - (\beta_{a_{j+s}} - \beta_{a_{j+s-1}})) \xrightarrow{d} N(0, m'_{j+s} \Omega_a m_{j+s}). \quad (34)$$

Substituting  $(\beta_{a_{h+1}} - \beta_{a_h}) = \lambda$  into equations (31) and (32), one then obtains the following estimators of the slopes of the time and cohort effects for  $k = 1, \dots, T-1$ :

$$\begin{aligned}(\widetilde{\gamma}_{t_{k+1}} - \widetilde{\gamma}_{t_k}) &= \Delta_t \bar{y}_{c_{h-k+1}, a_{h+1}, t_{k+1}} - \lambda, \\ (\widetilde{\alpha}_{c_{h-k+1}} - \widetilde{\alpha}_{c_{h-k}}) &= \Delta_c \bar{y}_{c_{h-k+1}, a_{h+1}, t_{k+1}} - \lambda.\end{aligned}\quad (35)$$

Under assumptions 1, 3, and 4, as  $n_{c_1} \rightarrow \infty$  these estimators will be consistent estimators of the slopes of the time effect and cohort effect profiles since the error terms in (31) and (32) will both converge in probability to zero. However, more efficient estimators than those in (35) can be obtained by also using the estimated changes in age effects,  $\widehat{b}_{a_j}$ . Substituting the  $\widehat{b}_{a_j}$  into (31) and (32) and

rearranging gives for  $k = 1, \dots, T - 1$ , and  $j = 1, \dots, A - 1$ ,

$$\Delta_t \bar{y}_{c_{j-k+1}, a_{j+1}, t_{k+1}} - \widehat{b}_{a_{j+1}} = \left( \gamma_{t_{k+1}} - \gamma_{t_k} \right) + \Delta_t \bar{\varepsilon}_{c_{j-k+1}, a_{j+1}, t_{k+1}} \quad (36)$$

$$\Delta_c \bar{y}_{c_{j-k+1}, a_{j+1}, t_{k+1}} - \widehat{b}_{a_{j+1}} = \left( \alpha_{c_{j-k+1}} - \alpha_{c_{j-k}} \right) + \Delta_c \bar{\varepsilon}_{c_{j-k+1}, a_{j+1}, t_{k+1}} \quad (37)$$

Let  $\widehat{g}_{t_{k+1}} \equiv \left( \widehat{\gamma}_{t_{k+1}} - \widehat{\gamma}_{t_k} \right)$  and  $\widehat{a}_{c_j} = \left( \alpha_{c_j} - \alpha_{c_{j-1}} \right)$  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  $H_{1j} = \min(T, A - j + 1) - \max(1, 3 - j) + 1$  cohort-level observations respectively. This can be done for a particular  $k$ , with the remaining slopes being recovered from the  $\widehat{g}_{t_k}$  and  $\widehat{a}_{c_j}$  through the time and cohort effect versions of equation (33). Alternatively, (36) and (37) can be used to obtain each cohort and time slope.

**Theorem 5** *For the data generating process given in (1), under Assumption 4, and*

(a) *Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ ,  $\widehat{g}_{t_{k+1}} \xrightarrow{P} \left( \gamma_{t_{k+1}} - \gamma_{t_k} \right)$  and  $\widehat{a}_{c_j} \xrightarrow{P} \left( \alpha_{c_j} - \alpha_{c_{j-1}} \right)$ .*

(b) *Assumption 2, as  $A \rightarrow \infty$  and  $T \rightarrow \infty$  (i),  $\widehat{g}_{t_{k+1}} \xrightarrow{P} \left( \gamma_{t_{k+1}} - \gamma_{t_k} \right)$ ; and (ii)  $\widehat{a}_{c_j} \xrightarrow{P} \left( \alpha_{c_j} - \alpha_{c_{j-1}} \right)$ .*

**Remark:**

With the normalization in Assumption 4, one can identify relative age, time and cohort effects, and see whether such effects exist. For example, a Wald test of  $\widehat{b}_{a_{j+1}} = 0$  for all  $j$  would test whether the age effect is the same for each age group, while finding that  $\widehat{b}_{a_{j+1}} > 0$  would indicate that the age effect is greater for age group  $a_{j+1}$  than it is for age  $a_j$ .

## 4.2 Identification of Actual Effects

To identify the age, cohort, and time effects themselves, rather than just their changes, additional normalizations are required. From the basic specification in equation (5), it can be seen that normalizing two of the effects implicitly places a normalizing restriction on the third effect. We normalize by setting the first time effect and the first cohort effect both equal to zero, that is:

**Assumption 5** :  $\gamma_{t_1} = 0$  and  $\alpha_{c_1} = 0$ .

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  $\widehat{a}_{c_j}$ , namely:

$$\begin{aligned}\gamma_{t_k}^* &= \sum_{s=2}^k \widehat{g}_{t_s} \text{ for } k = 2, \dots, T, \\ \alpha_{c_j}^* &= \sum_{h=2}^j \widehat{a}_{c_h} \text{ for } j = 2, \dots, A, \\ \alpha_{c_{-j}}^* &= -\sum_{h=j}^0 \widehat{a}_{c_{h+1}} \text{ for } j = 2 - T, \dots, 0.\end{aligned}\tag{38}$$

Under the conditions of Theorem 5 and Assumption 5,  $\gamma_{t_k}^*$  will be a consistent estimator of  $\gamma_{t_k}$  and  $\alpha_{c_j}^*$  a consistent estimator of  $\alpha_{c_j}$ , where the asymptotic directions are as in Theorem 5. Substituting these effects into (5) and rearranging gives for  $k = 1, \dots, T$ , and  $j = 1, \dots, A$ ,

$$\bar{y}_{c_{j-k+1}, a_j, t_k} - \alpha_{c_{j-k+1}}^* - \gamma_{t_k}^* = \beta_{a_j} + \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k}.\tag{39}$$

Let  $\beta_{a_j}^*$  be the least squares estimator of  $\beta_{a_j}$  from (39), ie.

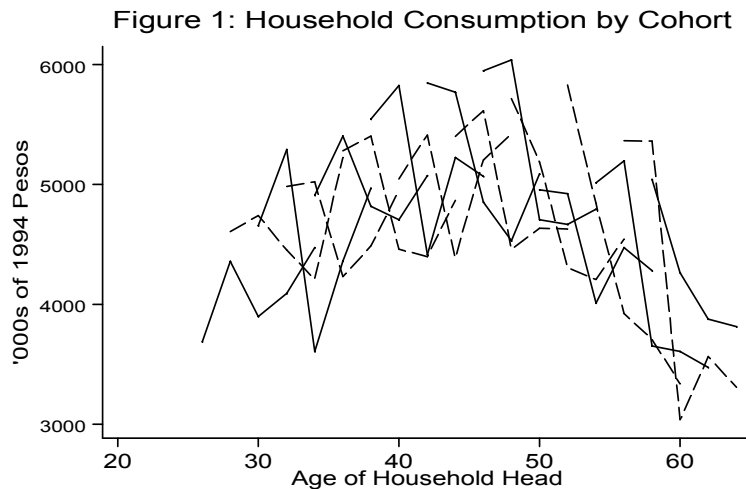
$$\beta_{a_j}^* = \frac{1}{T} \sum_{k=1}^T \left( \bar{y}_{c_{j-k+1}, a_j, t_k} - \alpha_{c_{j-k+1}}^* - \gamma_{t_k}^* \right)\tag{40}$$

Then  $\beta_{a_j}^* \xrightarrow{p} \beta_{a_j}$  as  $n_{c_1} \rightarrow \infty$  under Assumptions 1, 3, 4 and 5, and as  $A \rightarrow \infty$  and  $T \rightarrow \infty$ , under Assumptions 2, 3, 4 and 5.

Note that it is only meaningful to talk of positive or negative cohort, age, and time effects if the normalizations in Assumption 5 are correct. Otherwise, all comparisons are relative to the normalizations, and the differentials in these effects should be the primary concern. The focus in empirical work should therefore be on changes in effects, rather than the effects themselves, unless convincing normalizaations are available.

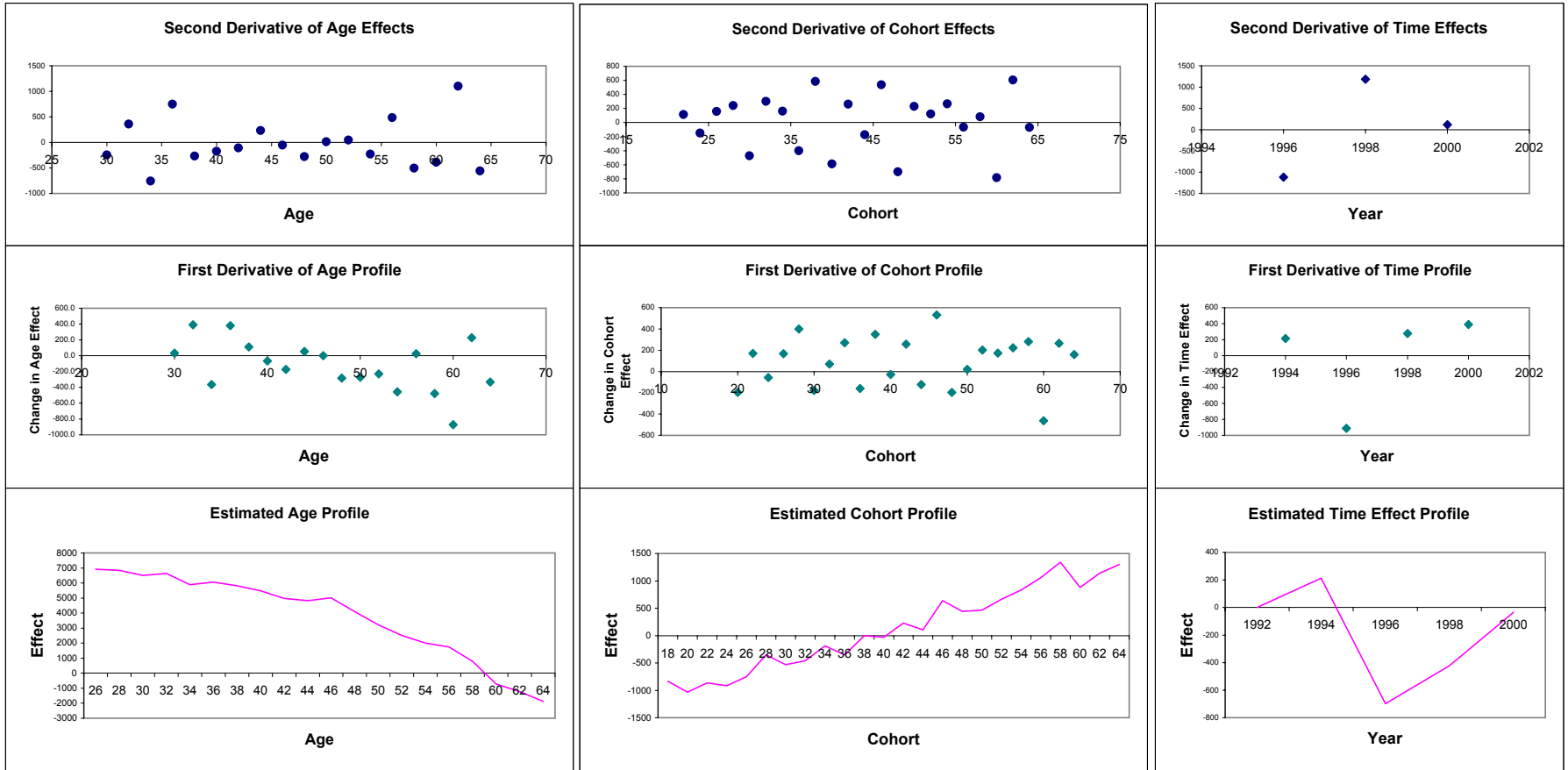
## 5 Empirical Examples

Figure 1 graphs Mexican household consumption by two-year birth cohort against two-year age group. 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 (2001b). 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

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



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.

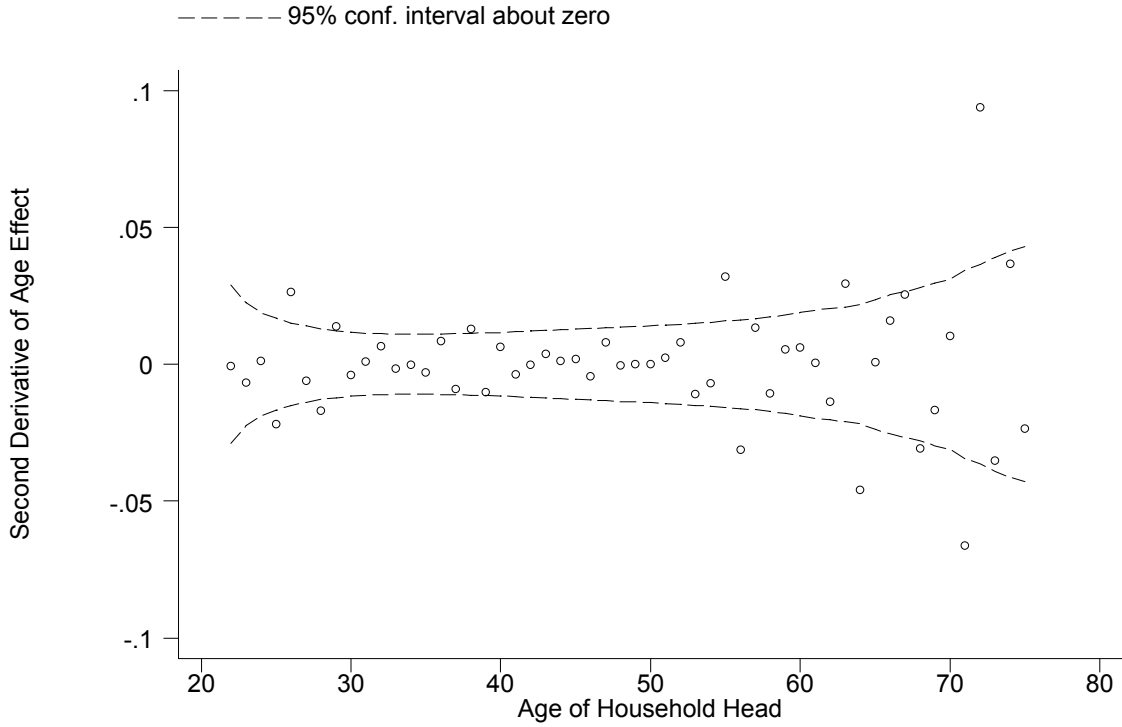
As a second example, we examine the age effects in the variance of log consumption in Taiwan, studied in Deaton and Paxson (1994). The data are annual data on real household consumption from 1976-96, taken from the Personal Income Distribution Surveys and described further in McKenzie (2001c). One year 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 (14) for each age group, together with an approximate 95 percent pointwise confidence interval about zero.<sup>6</sup> The effect of smaller cell sizes is seen in a widening of the confidence bands at very young and older age groups. Below age 55, the second differences in age effects appear relatively equal, suggesting at most a quadratic is needed to model age effects over this range. Furthermore, between ages 30 and 55, the second differenced effects are insignificant from zero, suggesting linearity of the age effect profile, rather than convexity or concavity, over this range. Above age 55, the differenced age effects appear more noisy, but we start to see some significant second differenced effects.

Figure 4 then shows the estimated age effects under different normalizing assumptions. Deaton and Paxson (1994) just use a set of age and cohort dummies, omitting the time effects, and normalize the age effect at age 38 to be the actual variance of log consumption. Normalizing the slope of the age effect function to be 0.005 at age 40 and using the second derivative estimates from Figure 3

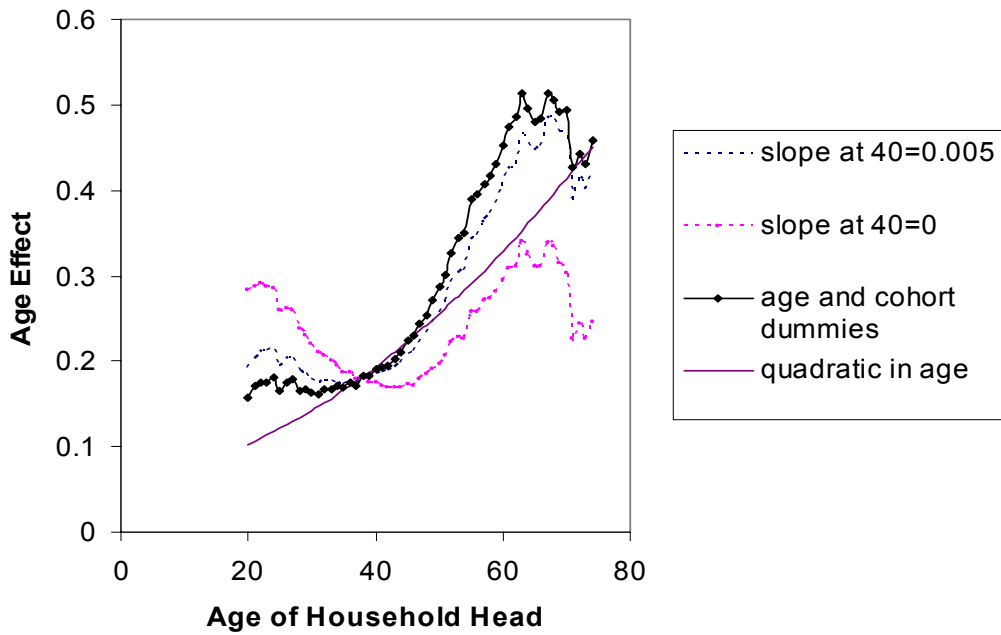
---

<sup>6</sup>The confidence interval assumes normality of log consumption in order to be able to calculate the standard error of the sample variances.

**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 Normalization Assumptions**



is seen to give very similar results. However, if we change this normalization slightly, setting the slope at age 40 to be zero, the estimated curve now appears much more U-shaped, with inequality at first decreasing and then increasing. Finally, fitting a quadratic in age along with a set of cohort dummies leads to a fairly linear upward sloping profile. The strong convexity present in Deaton and Paxson's estimates therefore seem to be mainly an artifact of the normalization used.

## 6 Adding Additional Regressors

While in some settings interest centres on the cohort, age, and time effects themselves, in other settings the researcher merely wishes to control for these effects when examining the relationship between  $y_{i,c_{j-k+1},a_j,t_k}$  and a  $r \times 1$  vector of other regressors,  $x_{i,c_{j-k+1},a_j,t_k}$ . That is, the question of interest is to estimate the vector of parameters  $\varphi$  in the model:

$$y_{i,c_{j-k+1},a_j,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + \gamma_{t_k} + x'_{i,c_{j-k+1},a_j,t_k} \varphi + \varepsilon_{i,c_{j-k+1},a_j,t_k} . \quad (41)$$

Taking cohort means in each time period, the pseudo-panel version is

$$\bar{y}_{c_{j-k+1},a_j,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + \gamma_{t_k} + \bar{x}'_{c_{j-k+1},a_j,t_k} \varphi + \bar{\varepsilon}_{c_{j-k+1},a_j,t_k} . \quad (42)$$

Proceeding as before, for  $j = 2, \dots, A - 1$ , and  $k = 2, \dots, T$ , (13) generalizes to

$$\Delta_c \Delta_t \bar{y}_{c_{j-k+1},a_j,t_k} = \tilde{\beta}_{a_{j+1}} + \Delta_c \Delta_t \bar{x}'_{c_{j-k+1},a_j,t_k} \varphi + \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+1},a_j,t_k} , \quad (43)$$

equation (21) generalizes for  $j = 2, \dots, A$ , and  $k = 3, \dots, T$ , to

$$\Delta_{-c,t} \Delta_t \bar{y}_{c_{j-k+1},a_j,t_k} = \tilde{\gamma}_{t_k} + \Delta_{-c,t} \Delta_t \bar{x}'_{c_{j-k+1},a_j,t_k} \varphi + \Delta_{-c,t} \Delta_t \bar{\varepsilon}_{c_{j-k+1},a_j,t_k} , \quad (44)$$



and for  $j = 4 - T, \dots, A$ ,  $k = 1, \dots, T - 1$ , and  $2 \leq k + j - 1 \leq A$ , (29) generalizes to

$$\Delta_{c,-t} \Delta_c \bar{y}_{c_{j-k+1}, a_j, t_k} = \tilde{\alpha}_{c_j} + \Delta_{c,-t} \Delta_c \bar{x}'_{c_{j-k+1}, a_j, t_k} \varphi + \Delta_{c,-t} \Delta_c \bar{e}_{c_{j-k+1}, a_j, t_k} . \quad (45)$$

Equations (43), (44) and (45) can now either be used separately, or together, to estimate  $\varphi$ , providing certain identifying conditions are met.

First consider estimating  $\varphi$  using just the set of equations in (43). There are  $(T - 1) \times (A - 2)$  equations to estimate the  $(A - 2) + r$  parameters  $\left\{ \tilde{\beta}_{a_{j+1}} \right\}_{j=2}^{A-1}$  and  $\varphi$ . Estimation of  $\varphi$  is only possible if  $r \leq (A - 2)(T - 2)$ , which requires that there be at least three time periods and age groups for any estimation to take place. The ordinary least squares estimator of  $\varphi$  from (43),  $\hat{\varphi}$ , is given by

$$\begin{aligned} \hat{\varphi} &= \varphi + \bar{F}^{-1} \bar{G} \\ \text{where } \bar{F} &= \sum_{j=2}^{A-1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t \bar{x}'_{c_{j-k+1}, a_j, t_k} \\ &\quad - \sum_{j=2}^{A-1} \left( \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \right) \left( \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{x}'_{c_{j-k+1}, a_j, t_k} \right) \\ \text{and } \bar{G} &= \sum_{j=2}^{A-1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{e}_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \\ &\quad - \sum_{j=2}^{A-1} \left( \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{e}_{c_{j-k+1}, a_j, t_k} \right) \left( \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \right) . \quad (46) \end{aligned}$$

Consistency of  $\hat{\varphi}$  requires the following assumption:

**Assumption 6** *Either*

- (a)  $x_{i, c_{j-k+1}, a_j, t_k} = x_{c_{j-k+1}, a_j, t_k} + \xi_{i, c_{j-k+1}, a_j, t_k}$ , where  $\xi_{i, c_{j-k+1}, a_j, t_k} \sim i.i.d. (0, \Sigma_\xi)$ ,  $x_{c_{j-k+1}, a_j, t_k}$ ,  $\xi_{i, c_{j-k+1}, a_j, t_k}$  and  $\eta_{i, c_{j-k+1}, a_j, t_k}$  are independent of one another, and  $F_1$  is of full rank  $r$ , where

$$\begin{aligned} F_1 &= \sum_{j=2}^{A-1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t x_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t x'_{c_{j-k+1}, a_j, t_k} \\ &\quad - \sum_{j=2}^{A-1} \left( \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t x_{c_{j-k+1}, a_j, t_k} \right) \left( \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t x'_{c_{j-k+1}, a_j, t_k} \right) , \quad (47) \end{aligned}$$

or

- (b)  $x_{i,c_{j-k+1},a_j,t_k} = x_{c_{j-k+1},a_j,t_k} + \zeta_{c_{j-k+1},a_j,t_k} + \xi_{i,c_{j-k+1},a_j,t_k}$  where  $\zeta_{c_{j-k+1},a_j,t_k} \sim i.i.d. (0, \Sigma_\zeta)$ ,  $\xi_{i,c_{j-k+1},a_j,t_k} \sim i.i.d. (0, \Sigma_\xi)$ ,  $x_{c_{j-k+1},a_j,t_k}$ ,  $\zeta_{c_{j-k+1},a_j,t_k}$ ,  $\xi_{i,c_{j-k+1},a_j,t_k}$ ,  $\nu_{c_{j-k+1},a_j,t_k}$  and  $\eta_{i,c_{j-k+1},a_j,t_k}$  are independent of one another; for all  $i, j, k$ ,  $E(x_{i,c_{j-k+1},a_j,t_k} \omega_{i,c_{j-k+1}}) < \infty$ ,

$$\lim_{T \rightarrow \infty} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t x_{c_{j-k+1},a_j,t_k} \equiv E_t(\Delta_c \Delta_t x_{c_{j-k+1},a_j,t_k}) < \infty, \text{ and}$$

$$\lim_{T \rightarrow \infty} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{x}_{c_{j-k+1},a_j,t_k} \Delta_c \Delta_t x'_{c_{j-k+1},a_j,t_k} \equiv E_t(\Delta_c \Delta_t x_{c_{j-k+1},a_j,t_k} \Delta_c \Delta_t x'_{c_{j-k+1},a_j,t_k}) < \infty ;$$

and

$$F_2 \equiv \sum_{j=2}^{A-1} \left( E_t(\Delta_c \Delta_t x_{c_{j-k+1},a_j,t_k} \Delta_c \Delta_t x'_{c_{j-k+1},a_j,t_k}) - E_t(\Delta_c \Delta_t x_{c_{j-k+1},a_j,t_k}) E_t(\Delta_c \Delta_t x'_{c_{j-k+1},a_j,t_k}) \right) + 4(A-2) \left( \Sigma_\zeta + \frac{\Sigma_\xi}{n_{c_1}} \right) \text{ is of full rank } r .$$

Note that we do not require that  $E(x_{i,c_{j-k+1},a_j,t_k} \omega_{i,c_{j-k+1}}) = 0$ , so that we allow for the possibility that individual-specific effects are correlated with some of the included regressor variables. We now have:

**Theorem 6** For the data generating process given in (41),

(a) under Assumptions 1, 3, and 6(a), as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,  $\hat{\varphi} \xrightarrow{P} \varphi$ .

(b) under Assumptions 2, 3, and 6(b), as  $T \rightarrow \infty$ , for  $n_{c_1}$  fixed,  $\hat{\varphi} \xrightarrow{P} \varphi$ .

**Remarks:**

- (a) A necessary condition to identify  $\varphi$  is that  $\Delta_c \Delta_t (x_{c_{j-k+1},a_j,t_k})_p \neq 0$  for some  $j$  and  $k$  and each  $p = 1, 2, \dots, r$ , where  $(x_{c_{j-k+1},a_j,t_k})_p$  denotes the  $p$ th column of  $x_{c_{j-k+1},a_j,t_k}$ . That is, each of the variables in  $x$  must exhibit variation both over cohorts and across time. Intuitively, if one wishes to ascertain the impact of  $x$  on  $y$ , controlling for cohort, age, and time effects,

then  $x$  must not be colinear with these effects. This rules out aggregate variables such as macroeconomic effects, and also time-invariant variables such as sex and race.<sup>7</sup> Note that such variables are implicitly being controlled for, but one can not consistently estimate their coefficients together with cohort and time effects.

- (b) Note that the changes in the slopes of the age effect function, the  $\tilde{\beta}_{a_{j+1}}$ 's, can also be consistently estimated along with  $\varphi$ . In a similar manner,  $\varphi$  could be also estimated using either (44) or (45) separately, enabling the changes in the slopes of the cohort effect and time effect functions to also be estimated. The parameter estimators  $\alpha_{c_{j-k+1}}^*$ ,  $\beta_{a_j}^*$ , and  $\gamma_{t_k}^*$  can then be recovered using the normalizing assumptions 4 and 5, as was done for the model without additional regressors.
- (c) Under these normalizing assumptions, it is possible to obtain consistent estimates of coefficients on time-invariant and cohort-invariant (aggregate) variables. For example, consider adding the time-invariant term  $Z'_{i,c_{j-k+1}}\rho$  to (41). Then  $\rho$  can be estimated using the regression:

$$\bar{y}_{c_{j-k+1},a_j,t_k} - \alpha_{c_{j-k+1}}^* - \beta_{a_j}^* - \gamma_{t_k}^* - \bar{x}'_{c_{j-k+1},a_j,t_k}\hat{\varphi} = \bar{Z}'_{c_{j-k+1}}\rho + v_{c_{j-k+1},a_j,t_k}. \quad (48)$$

The least squares estimator of  $\rho$  from (48) will be consistent as  $n_{c_1} \rightarrow \infty$  under the normalization assumptions. For fixed  $n_{c_1}$ , least squares will not be consistent as  $T \rightarrow \infty$  if  $E(Z_{i,c_{j-k+1}}\omega_{i,c_{j-k+1}}) \neq 0$ . The approach of Hausman and Taylor (1981) could then be used to obtain consistent estimates if there are sufficient columns of  $\bar{x}'_{c_{j-k+1},a_j,t_k}$  which are uncorrelated with  $\bar{\omega}_{c_{j-k+1}}$ , and can thus be used as an instrument for  $\bar{Z}'_{c_{j-k+1}}$ .

Stacking the equations in (43) with those in (44) and (45), one can potentially achieve greater efficiency using all equations together. Let  $\tilde{Y}$  be the vector consisting of stacked  $\Delta_c\Delta_t\bar{y}_{c_{j-k+1},a_j,t_k}$ 's,  $\Delta_{-c,t}\Delta_t\bar{y}_{c_{j-k+1},a_j,t_k}$ 's and  $\Delta_{c,-t}\Delta_c\bar{y}_{c_{j-k+1},a_j,t_k}$ 's;  $\Theta$  the vector of stacked parameters  $\left\{\tilde{\beta}_{a_{j+1}}\right\}_{j=2}^{A-1}$ ,

---

<sup>7</sup>Note that while the sex or race composition of a birth cohort may change over time, due to non-random mortality for example, such changes mean that we are not comparing the same group of individuals from one period to the next, and hence are generally a cause for concern rather than a means of identifying the effects of time-invariant individual level variables.

$\{\tilde{\gamma}_{t_k}\}_{k=3}^T$ ,  $\{\tilde{\alpha}_{c_j}\}_{j=4-T}^A$  and  $\varphi$ ; and  $\tilde{X}$  and  $\epsilon$  the corresponding matrix of stacked regressors and vector of stacked errors respectively. Letting  $\hat{\Theta} = \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}'\tilde{Y}$  be the least squares estimator from this stacked regression, it should be clear that under Assumptions 1 and 3, as  $n_{c_1} \rightarrow \infty$ , for  $T$  fixed,  $\hat{\Theta} \xrightarrow{p} \Theta$ , provided that the identification condition that  $\lim_{n_{c_1} \rightarrow \infty} \tilde{X}'\tilde{X}$  is of full rank  $2(A+T) + r - 8$  holds.

## 7 Genuine Panel Data

The methods presented in this paper can easily be applied with only a few modifications to the genuine panel data case. The panel data version of (13) for  $j = 2, \dots, A-1$ ,  $k = 2, \dots, T$ , and  $i = 1, \dots, n_{c_{j-k+2}}$ , is

$$\Delta_c \Delta_t y_{i, c_{j-k+2}, a_{j+1}, t_k} = \tilde{\beta}_{a_{j+1}} + \Delta_c \Delta_t \varepsilon_{i, c_{j-k+2}, a_{j+1}, t_k} , \quad (49)$$

where  $\Delta_c \Delta_t y_{i, c_{j-k+2}, a_{j+1}, t_k} = \Delta_t y_{i, c_{j-k+2}, a_{j+1}, t_k} - \Delta_t y_{i, c_{j-k+1}, a_j, t_k}$ . Letting  $N_j = \sum_{k=2}^T n_{c_{j-k+2}}$  be the total number of observations available for estimating  $\tilde{\beta}_{a_{j+1}}$ , it is straightforward to show that  $\hat{\beta}_{a_{j+1}} = \frac{1}{N} \sum_{k=2}^T \sum_{i=1}^{n_{c_{j-k+2}}} \Delta_c \Delta_t y_{i, c_{j-k+2}, a_{j+1}, t_k} \xrightarrow{p} \tilde{\beta}_{a_{j+1}}$  as  $N_j \rightarrow \infty$  under Assumption 1, and as  $T \rightarrow \infty$  under Assumption 2. Likewise, one can show that as  $N_j \rightarrow \infty$ , under Assumption 1,

$$\sqrt{N_j} \left( \hat{\beta}_{a_{j+1}} - \tilde{\beta}_{a_{j+1}} \right) \xrightarrow{d} N \left( 0, 4\sigma_\eta^2 \right) . \quad (50)$$

One can hence consistently estimate the second differences of the age effects with genuine panel data. Note that with genuine panel data,  $\Delta_c \Delta_t \omega_{i, c_{j-k+1}} = 0$ , and hence the individual effects are eliminated by the differencing process and therefore their variance does not appear in the limiting distribution of the standardized estimator. This contrasts with the pseudo-panel case, where  $\Delta_c \Delta_t \bar{\omega}_{c_{j-k+1}} \neq 0$  due to a different sample of individuals being taken each period from a given cohort. Wald tests can again be used on the  $\hat{\beta}_{a_{j+1}}$  to test whether they all equal (the age profile is quadratic) or all equal to zero (linear age profile).

Similarly, for the time effects, we have  $\hat{\gamma}_{t_k} = \frac{1}{D_k} \sum_{j=2}^A \sum_{i=1}^{n_{c_j-k+2}} \Delta_{-c,t} \Delta_t y_{i,c_j-k+1,a_j,t_k} \xrightarrow{p} \tilde{\gamma}_{t_k}$  as  $D_k = \sum_{j=2}^A n_{c_j-k+2} \rightarrow \infty$  under Assumption 1. That is identification of second differenced time effects requires the number of observations in a given time period to pass to infinity. For the second differenced cohort effects,  $\hat{\alpha}_{c_j} = \frac{1}{F_j} \sum_{k=\max(3-j,1)}^{\min(A-j+1,T-1)} \sum_{i=1}^{n_{c_j-k+2}} \Delta_{c,-t} \Delta_c y_{i,c_j,a_{k+j-1},t_k} \xrightarrow{p} \tilde{\alpha}_{c_j}$  as  $F_j = \sum_{k=\max(3-j,1)}^{\min(A-j+1,T-1)} n_{c_j-k+2} \rightarrow \infty$  under Assumption 1. Normalization assumptions can then be further imposed in the same way as was done for the pseudo-panel data. The more general model in (41) can also be estimated with panel data by applying the techniques given in the previous section.

## 8 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. We have also provided minimal normalizing assumptions, and shown that normalizing one of the effects imposes explicit normalization restrictions on the other effects. Estimation of more general models which include other regressors is possible without the need to make these normalizations. It is hoped that the methods provided will be of use in the wide variety of empirical applications in which age, cohort and time effects potentially figure.

## 9 Mathematical Proofs

### 9.1 Proof of Theorem 1

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

$$\begin{aligned}\bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} &= \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \varepsilon_{i(t_k), c_{j-k+2}, a_{j+1}, t_k} \\ &= \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \omega_{i(t_k), c_{j-k+2}} + \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \eta_{i(t_k), c_{j-k+2}, a_{j+1}, t_k}\end{aligned}$$

Now as  $n_{c_1} \rightarrow \infty$ , under Assumption 3  $n_{c_{j-k+2}} \rightarrow \infty$ , 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  $\frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \eta_{i(t_k), c_{j-k+2}, a_{j+1}, t_k} \xrightarrow{p} E(\eta_{i, c_{j-k+2}, a_{j+1}, t_k}) = 0$ . Hence as  $n_{c_{j-k+2}} \rightarrow \infty$ ,

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

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

To prove (b), note that under Assumption 2,

$$\begin{aligned}& \frac{1}{T-1} \sum_{k=2}^T \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} = \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \frac{1}{T-1} \sum_{k=2}^T \omega_{i(t_k), c_{j-k+2}} \\ & + \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \frac{1}{T-1} \sum_{k=2}^T \nu_{c_{j-k+2}, a_{j+1}, t_k} + \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \frac{1}{T-1} \sum_{k=2}^T \eta_{i(t_k), c_{j-k+2}, a_{j+1}, t_k} .\end{aligned}\quad (51)$$

As  $T \rightarrow \infty$ ,  $\frac{1}{T-1} \sum_{k=2}^T \nu_{c_{j-k+2}, a_{j+1}, t_k} \xrightarrow{p} E_t(\nu_{c_{j-k+2}, a_{j+1}, t_k}) = 0$  where  $E_t(\cdot)$  denotes an expectation over time, and likewise  $\frac{1}{T-1} \sum_{k=2}^T \eta_{i(t_k), c_{j-k+2}, a_{j+1}, t_k} \xrightarrow{p} 0$ , both by the WLLN under Assumption 2. Also, as different individuals are observed each period,  $\frac{1}{T-1} \sum_{k=2}^T \omega_{i(t_k), c_{j-k+2}} \xrightarrow{p} 0$  as  $T \rightarrow \infty$ . Hence from (51),  $\frac{1}{T-1} \sum_{k=2}^T \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \xrightarrow{p} 0$  as  $T \rightarrow \infty$  for all  $j$ , and thus

$$\begin{aligned}
\widehat{\beta}_{a_{j+1}} &= \widetilde{\beta}_{a_{j+1}} + \frac{1}{T-1} \sum_{k=2}^T \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} - \frac{1}{T-1} \sum_{k=2}^T \bar{\varepsilon}_{c_{j-k+2}, a_j, t_{k-1}} \\
&\quad - \frac{1}{T-1} \sum_{k=2}^T \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} + \frac{1}{T-1} \sum_{k=2}^T \bar{\varepsilon}_{c_{j-k+1}, a_{j-1}, t_{k-1}} \\
&\xrightarrow{p} \widetilde{\beta}_{a_{j+1}} + 0.
\end{aligned}$$

Parts (i) and (ii) of (c) follow immediately from (a) and (b). To show (iii), note that the first and third terms of (51) converge in probability to zero as  $n_{c_1} \rightarrow \infty$ , as per the proof of part (a). The middle term under Assumption 2 is

$$\frac{1}{T-1} \sum_{k=2}^T \frac{1}{n_{c_{j-k+2}}} \sum_{i=1}^{n_{c_{j-k+2}}} \nu_{c_{j-k+2}, a_{j+1}, t_k} = \frac{1}{T-1} \sum_{k=2}^T \nu_{c_{j-k+2}, a_{j+1}, t_k},$$

which does not converge in probability to zero as  $n_{c_1} \rightarrow \infty$  for  $T$  fixed. However, as  $(n_{c_1}, T \rightarrow \infty)_{seq}$ ,  $\frac{1}{T-1} \sum_{k=2}^T \nu_{c_{j-k+2}, a_{j+1}, t_k} \xrightarrow{p} E_t(\nu_{c_{j-k+2}, a_{j+1}, t_k}) = 0$ , as per part (b), and hence one obtains that  $\frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \xrightarrow{p} 0$  in sequential limit as  $(n_{c_1}, T \rightarrow \infty)_{seq}$ .

To show (d), 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.  $(0, \sigma_\omega^2 + \sigma_\eta^2)$ . By the Lindeberg-Levy Central Limit Theorem,

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

and similarly  $\sqrt{n_{c_1}} \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \xrightarrow{d} N\left(0, (\sigma_\omega^2 + \sigma_\eta^2) \left(\frac{2}{\delta_{j-k+2}} + \frac{2}{\delta_{j-k+1}}\right)\right)$ . The convergence of  $\frac{1}{T-1} \sum_{k=2}^T \sqrt{n_{c_1}} \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k}$  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  $var_c(y_{i, c_{j-k+1}, a_j, t_k}) = var_c(\varepsilon_{i, c_{j-k+1}, a_j, t_k}) = \sigma_\omega^2 + \sigma_\eta^2$ . 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.

Part (e) follows from (a), as  $\Delta_c \Delta_t \bar{y}_{c_{j-k+2}, a_{j+1}, t_k} - \widehat{\beta}_{a_{j+1}} \xrightarrow{p} 0$ , and hence

$$\frac{1}{T-2} \sum_{k=2}^T \left( \Delta_c \Delta_t \bar{y}_{c_{j-k+2}, a_{j+1}, t_k} - \widehat{\beta}_{a_{j+1}} \right)^2 \xrightarrow{p} 0.$$

## 9.2 Proof of Theorem 2

**Proof.** The proof of part (a) follows from the proof of part (a) of Theorem 1 as each  $\bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} \xrightarrow{p} 0$  as  $n_{c_1} \rightarrow \infty$ . To prove part (b), note that

$$\begin{aligned} & \frac{1}{A-1} \sum_{j=2}^A \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} = \frac{1}{A-1} \sum_{j=2}^A \frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} \omega_{i(t_k), c_{j-k+1}} \\ & + \frac{1}{A-1} \sum_{j=2}^A \frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} \nu_{c_{j-k+1}, a_j, t_k} + \frac{1}{A-1} \sum_{j=2}^A \frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} \eta_{i(t_k), c_{j-k+1}, a_j, t_k}. \end{aligned} \quad (52)$$

Now the  $\left\{ \frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} \omega_{i(t_k), c_{j-k+1}} \right\}_{j=2}^A$  are i.n.i.d. random variables with expectation zero and finite variance under Assumption 2. Hence by the WLLN for i.n.i.d. random variables, as  $A \rightarrow \infty$ ,  $\frac{1}{A-1} \sum_{j=2}^A \left( \frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} \omega_{i(t_k), c_{j-k+1}} \right) \xrightarrow{p} E \left( \frac{1}{n_{c_{j-k+1}}} \sum_{i=1}^{n_{c_{j-k+1}}} \omega_{i(t_k), c_{j-k+1}} \right) = 0$ . Likewise the second and third terms of (52) are also sample means of random variables with expectation zero, and hence converge in probability to zero as  $A \rightarrow \infty$ . Thus  $\frac{1}{A-1} \sum_{j=2}^A \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} \xrightarrow{p} 0$  for all  $k$ , giving the result. Part (c) follows immediately from parts (a) and (b), and from the proof of part (c) of Theorem 1. Part (d) follows the proof of part (d) of Theorem 1.

## 9.3 Proof of Theorem 3

**Proof.** Again the proof of (a) follows from the proof of part (a) of Theorem 1 as each  $\bar{\varepsilon}_{c_j, a_{k+j-1}, t_k} \xrightarrow{p} 0$  as  $n_{c_1} \rightarrow \infty$ . To prove part (b), write

$$\frac{1}{H_j} \sum_{k=\max(3-j, 1)}^{\min(A-j+1, T-1)} \bar{\varepsilon}_{c_j, a_{k+j-1}, t_k} = \frac{1}{H_j} \sum_{k=\max(3-j, 1)}^{\min(A-j+1, T-1)} \frac{1}{n_{c_j}} \sum_{i=1}^{n_{c_j}} \varepsilon_{i(t_k), c_j, a_{k+j-1}, t_k}, \quad (53)$$



and note that as  $A \rightarrow \infty$  and  $T \rightarrow \infty$ ,  $H_j \rightarrow \infty$ . Then following the proof of part (b) of Theorem 2, one can show that the right-hand side of (53) converges in probability to zero as  $H_j \rightarrow \infty$ , giving the required result. Parts (c) and (d) follow immediately from the proofs of part (c) and (d) of Theorem 1.

#### 9.4 Proof of Theorem 4

**Proof.** Theorem 1 gives that  $\widehat{\beta}_{a_{j+1}} \xrightarrow{p} \widetilde{\beta}_{a_{j+1}}$  as  $n_{c_1} \rightarrow \infty$  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  $\Omega_a$ . For  $h \neq j$ ,  $h = 2, \dots, A-1$  we therefore wish to evaluate

$$\begin{aligned} & E \left( \sqrt{n_{c_1}} \left( \widehat{\beta}_{a_{j+1}} - \widetilde{\beta}_{a_{j+1}} \right) \sqrt{n_{c_1}} \left( \widehat{\beta}_{a_{h+1}} - \widetilde{\beta}_{a_{h+1}} \right) \right) \\ &= E \left( n_{c_1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \frac{1}{T-1} \sum_{s=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} \right). \end{aligned} \quad (54)$$

Recall that

$$\Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} = \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} - \bar{\varepsilon}_{c_{j-k+2}, a_j, t_{k-1}} - \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} + \bar{\varepsilon}_{c_{j-k+1}, a_{j-1}, t_{k-1}}. \quad (55)$$

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

$$E \left( \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} \right) = 0 \text{ for all } (h, s) \neq (j, k). \quad (56)$$

From (55) we have that

$$\begin{aligned} & E \left( \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \Delta_c \Delta_t \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} \right) \\ &= E \left( \left( \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} - \bar{\varepsilon}_{c_{j-k+2}, a_j, t_{k-1}} - \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} + \bar{\varepsilon}_{c_{j-k+1}, a_{j-1}, t_{k-1}} \right) \Delta_c \Delta_t \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} \right) \end{aligned} \quad (57)$$

Expanding out the first element of (57) and using (56) gives

$$\begin{aligned}
& E \left( \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \Delta_c \Delta_t \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} \right) \\
&= E \left( \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \left( \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} - \bar{\varepsilon}_{c_{h-s+2}, a_h, t_{s-1}} - \bar{\varepsilon}_{c_{h-s+1}, a_h, t_s} + \bar{\varepsilon}_{c_{h-s+1}, a_{h-1}, t_{s-1}} \right) \right) \\
&= \begin{cases} -E \left( \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \bar{\varepsilon}_{c_{h-s+2}, a_h, t_{s-1}} \right) = \frac{-(\sigma_\omega^2 + \sigma_\eta^2)}{n_{c_{j-k+2}}} \text{ if } (h, s) = (j+1, k+1) \\ -E \left( \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \bar{\varepsilon}_{c_{h-s+1}, a_h, t_s} \right) = \frac{-(\sigma_\omega^2 + \sigma_\eta^2)}{n_{c_{j-k+2}}} \text{ if } (h, s) = (j+1, k) \\ E \left( \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \bar{\varepsilon}_{c_{h-s+1}, a_{h-1}, t_{s-1}} \right) = \frac{(\sigma_\omega^2 + \sigma_\eta^2)}{n_{c_{j-k+2}}} \text{ if } (h, s) = (j+2, k+1) \\ 0 \text{ otherwise} \end{cases} \quad (58)
\end{aligned}$$

Expanding out the other elements of (57) and combining their results with (58) then gives:

$$\begin{aligned}
& E \left( \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \Delta_c \Delta_t \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} \right) \\
&= \begin{cases} \frac{(\sigma_\omega^2 + \sigma_\eta^2)}{n_{c_{j-k+2}}} \text{ if } (h, s) = (j+2, k+1) \\ -(\sigma_\omega^2 + \sigma_\eta^2) \left( \frac{1}{n_{c_{j-k+2}}} + \frac{1}{n_{c_{j-k+1}}} \right) \text{ if } (h, s) = (j+1, k+1) \\ -(\sigma_\omega^2 + \sigma_\eta^2) \frac{2}{n_{c_{j-k+2}}} \text{ if } (h, s) = (j+1, k) \\ -(\sigma_\omega^2 + \sigma_\eta^2) \frac{2}{n_{c_{j-k+1}}} \text{ if } (h, s) = (j-1, k) \\ -(\sigma_\omega^2 + \sigma_\eta^2) \left( \frac{1}{n_{c_{j-k+2}}} + \frac{1}{n_{c_{j-k+1}}} \right) \text{ if } (h, s) = (j-1, k-1) \\ \frac{(\sigma_\omega^2 + \sigma_\eta^2)}{n_{c_{j-k+1}}} \text{ if } (h, s) = (j-2, k-1) \\ 0 \text{ otherwise} \end{cases} \quad (59)
\end{aligned}$$

From (58) it follows that

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

$$= \begin{cases} \frac{(\sigma_\omega^2 + \sigma_\eta^2)}{n_{c_{j-k+2}}} \text{ if } h = j + 2, k \leq T - 1 \\ -(\sigma_\omega^2 + \sigma_\eta^2) \left( \frac{3}{n_{c_{j-k+2}}} + \frac{1}{n_{c_{j-k+1}}} \right) \text{ if } h = j + 1, k \leq T - 1 \\ -(\sigma_\omega^2 + \sigma_\eta^2) \frac{2}{n_{c_{j-k+2}}} \text{ if } h = j + 1, k = T \\ -(\sigma_\omega^2 + \sigma_\eta^2) \frac{2}{n_{c_{j-k+1}}} \text{ if } h = j - 1, k = 2 \\ -(\sigma_\omega^2 + \sigma_\eta^2) \left( \frac{1}{n_{c_{j-k+2}}} + \frac{3}{n_{c_{j-k+1}}} \right) \text{ if } h = j - 1, k \geq 3 \\ \frac{(\sigma_\omega^2 + \sigma_\eta^2)}{n_{c_{j-k+1}}} \text{ if } h = j - 2, k \geq 3 \\ 0 \text{ otherwise} \end{cases} \quad (60)$$

and hence that

$$\begin{aligned} & E \left( \sum_{k=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \sum_{s=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} \right) \\ &= \begin{cases} \sum_{k=2}^{T-1} \frac{(\sigma_\omega^2 + \sigma_\eta^2)}{n_{c_{j-k+2}}} \text{ if } h = j + 2 \\ -(\sigma_\omega^2 + \sigma_\eta^2) \left[ \sum_{k=2}^{T-1} \left( \frac{3}{n_{c_{j-k+2}}} + \frac{1}{n_{c_{j-k+1}}} \right) + \frac{2}{n_{c_{j-T+2}}} \right] \text{ if } h = j + 1 \\ 0 \text{ otherwise} \end{cases} \quad (61) \end{aligned}$$

Using Assumption 3 and (61), we therefore have that as  $n_{c_1} \rightarrow \infty$ ,

$$\begin{aligned} & E \left( n_{c_1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+2}, a_{j+1}, t_k} \frac{1}{T-1} \sum_{s=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{h-s+2}, a_{h+1}, t_s} \right) \\ & \xrightarrow{P} \begin{cases} \frac{1}{(T-1)^2} \sum_{k=2}^{T-1} \frac{(\sigma_\omega^2 + \sigma_\eta^2)}{\delta_{c_{j-k+2}}} \text{ if } h = j + 2 \\ \frac{-(\sigma_\omega^2 + \sigma_\eta^2)}{(T-1)^2} \left[ \sum_{k=2}^{T-1} \left( \frac{3}{\delta_{c_{j-k+2}}} + \frac{1}{\delta_{c_{j-k+1}}} \right) + \frac{2}{\delta_{c_{j-T+2}}} \right] \text{ if } h = j + 1 \\ 0 \text{ otherwise} \end{cases} \quad (62) \end{aligned}$$

## 9.5 Proof of Theorem 5

**Proof.** Part (a) follows from the consistency of  $\hat{b}_{a_{j+1}}$  as  $n_{c_1} \rightarrow \infty$  and the fact that  $\bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} \xrightarrow{P} 0$  as  $n_{c_1} \rightarrow \infty$ . To prove (i) of part (b), note that:

$$\begin{aligned}
\widehat{g}_{t_{k+1}} &= \frac{1}{A-1} \sum_{j=1}^{A-1} \left( \Delta_t \bar{y}_{c_{j-k+1}, a_{j+1}, t_{k+1}} - \widehat{b}_{a_{j+1}} \right) \\
&= \left( \gamma_{t_{k+1}} - \gamma_{t_k} \right) + \frac{1}{A-1} \sum_{j=1}^{A-1} \Delta_t \bar{\varepsilon}_{c_{j-k+1}, a_{j+1}, t_{k+1}} + \frac{1}{A-1} \sum_{j=1}^{A-1} \left( b_{a_{j+1}} - \widehat{b}_{a_{j+1}} \right)
\end{aligned}$$

Convergence in probability to zero of the second term as  $A \rightarrow \infty$  is shown in the proof part (b) of Theorem 2. Also, as  $A \rightarrow \infty$ ,  $b_{a_{j+1}} \xrightarrow{p} \widehat{b}_{a_{j+1}}$  by part (b) of Theorem 1 and Assumption 4. Part (ii) of (b) likewise follows from Assumption 4, Theorem 1 and the proof of Theorem 3 as  $H_{1j} \rightarrow \infty$ .

## 9.6 Proof of Theorem 6

**Proof.** To prove (a), note that under the given Assumptions,  $\bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} \xrightarrow{p} 0$  as  $n_{c_1} \rightarrow \infty$  for  $T$  fixed as per the proof of Theorem 1, and similarly  $\bar{x}_{c_{j-k+1}, a_j, t_k} \xrightarrow{p} x_{c_{j-k+1}, a_j, t_k}$ . Hence  $\bar{G} \xrightarrow{p} 0$  as  $n_{c_1} \rightarrow \infty$  for  $T$  fixed. Also, as  $\bar{x}_{c_{j-k+1}, a_j, t_k} \xrightarrow{p} x_{c_{j-k+1}, a_j, t_k}$ ,  $\bar{F} \xrightarrow{p} F_1$ , a non-singular matrix. Thus  $\widehat{\varphi} \xrightarrow{p} \varphi$  as required.

For (b), note first that  $\frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} \xrightarrow{p} 0$  by the WLLN as  $T \rightarrow \infty$  for  $n_{c_1}$  fixed, and  $\frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \xrightarrow{p} E_t (\Delta_c \Delta_t x_{c_{j-k+1}, a_j, t_k}) < \infty$  by Assumption 6(b). Secondly, write

$$\begin{aligned}
& \sum_{j=2}^{A-1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\varepsilon}_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} = \sum_{j=2}^{A-1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\omega}_{c_{j-k+1}} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \\
& + \sum_{j=2}^{A-1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \nu_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} + \sum_{j=2}^{A-1} \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\eta}_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} .
\end{aligned} \tag{63}$$

The second and third terms converge in probability to zero as  $T \rightarrow \infty$  by the independence of

$\nu_{c_{j-k+1}, a_j, t_k}$  and  $\eta_{i, c_{j-k+1}, a_j, t_k}$  from  $x_{i, c_{j-k+1}, a_j, t_k}$ . Now write

$$\begin{aligned}
& \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\omega}_{c_{j-k+1}} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \\
&= \left( \frac{1}{T-1} \sum_{k=2}^T \bar{\omega}_{c_{j-k+1}} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} - \frac{1}{T-1} \sum_{k=2}^T \bar{\omega}_{c_{j-k+1}} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_{j-1}, t_{k-1}} \right) \\
&\quad - \left( \frac{1}{T-1} \sum_{k=2}^T \bar{\omega}_{c_{j-k}} \Delta_c \Delta_t \bar{x}_{c_{j-k}, a_{j-1}, t_k} - \frac{1}{T-1} \sum_{k=2}^T \bar{\omega}_{c_{j-k}} \Delta_c \Delta_t \bar{x}_{c_{j-k}, a_{j-2}, t_{k-1}} \right) \\
&\quad \xrightarrow{p} \left( E_t \left( \bar{\omega}_{c_{j-k+1}} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \right) - E_t \left( \bar{\omega}_{c_{j-k+1}} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_{j-1}, t_{k-1}} \right) \right) \\
&\quad - \left( E_t \left( \bar{\omega}_{c_{j-k}} \Delta_c \Delta_t \bar{x}_{c_{j-k}, a_{j-1}, t_k} \right) - E_t \left( \bar{\omega}_{c_{j-k}} \Delta_c \Delta_t \bar{x}_{c_{j-k}, a_{j-2}, t_{k-1}} \right) \right) = 0,
\end{aligned}$$

where  $E_t \left( \bar{\omega}_{c_{j-k+1}} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \right) = E_t \left( \bar{\omega}_{c_{j-k+1}} \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_{j-1}, t_{k-1}} \right) < \infty$ , since under Assumption 6 (b)  $E \left( x_{i, c_{j-k+1}, a_j, t_k} \omega_{i, c_{j-k+1}} \right) < \infty$ . Hence  $\bar{G} \xrightarrow{p} 0$  as  $T \rightarrow \infty$ . Also, under Assumption 6 (b), one can show that as  $T \rightarrow \infty$

$$\begin{aligned}
& \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{x}_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t \bar{x}'_{c_{j-k+1}, a_j, t_k} = \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t x_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t x'_{c_{j-k+1}, a_j, t_k} \\
&+ \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \zeta_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t \zeta'_{c_{j-k+1}, a_j, t_k} + \frac{1}{T-1} \sum_{k=2}^T \Delta_c \Delta_t \bar{\xi}_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t \bar{\xi}'_{c_{j-k+1}, a_j, t_k} + o_p(1) \\
&\quad \xrightarrow{p} E_t \left( \Delta_c \Delta_t x_{c_{j-k+1}, a_j, t_k} \Delta_c \Delta_t x'_{c_{j-k+1}, a_j, t_k} \right) + 4\Sigma_\zeta + \frac{4}{n_{c_1}} \Sigma_\xi.
\end{aligned}$$

We therefore have that  $\bar{F} \xrightarrow{p} F_2$ , a full rank matrix as  $T \rightarrow \infty$  under Assumption 6 (b), and hence

$$\hat{\varphi} \xrightarrow{p} \varphi.$$

## References

- Collado, MD. (1997) "Estimating dynamic models from time series of independent cross-sections", *Journal of Econometrics*, 82: 37-62.
- Deaton (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*, 30: 109-126.
- Deaton, A. and C. Paxson (1994). "Intertemporal Choice and Inequality", *Journal of Political Economy* 102(3): 437-467.
- Denton, FT., DC. Mountain and BG. Spencer (1999) "Age, Trend, and Cohort Effects in a Macro Model of Canadian Expenditure Patterns", *Journal of Business and Economic Statistics* 17(4): 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*, 44(2): 287-315.
- Hausman, J. and W. Taylor (1981). "Panel Data and Unobservable Individual Effects", *Econometrica* 49(6): 1377-1398.
- Heckman, J. and R. Robb (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 E. Vytlacil (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* 83(1): 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* 45(1): 57-75.
- Mason, WM. and SE. Fienberg (1985) *Cohort Analysis in Social Research: Beyond the Identification Problem*, Springer-Verlag: New York.
- Mckenzie, DJ. (2001a) "Asymptotic Theory for Heterogeneous Dynamic Pseudo-Panels", mimeo, Stanford University.
- Mckenzie, DJ. (2001b) "The Household Response to the Mexican Peso Crisis", Stanford Economics Department Working Paper 01-017.
- Mckenzie, DJ. (2001c) "Consumption Growth in a Booming Economy: Taiwan 1976-96", Yale University Economic Growth Center Discussion Paper No. 823.
- Verbeek, M. (1995) "Pseudo-panel Data", pp. 280-92 in Laszlo Matyas and Patrick Sevestre ed. *The Econometrics of Panel Data: A Handbook of Theory with Applications*, 2nd edition. Kluwer Academic Publishers: Boston.
- Verbeek, M. and T. Nijman (1993). "Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections", *Journal of Econometrics*, 59: 125-36.