Age-period-cohort modelling and covariates, with an application to obesity in England 2001-2014

Zoë Fannon, Christiaan Monden, Bent Nielsen

27 November 2018

Abstract

We develop an age-period-cohort model for repeated cross-section data with individual covariates. This is done for both continuous and binary dependent variables. The age-period-cohort identification problem is addressed by use of the canonical parametrization which has freely varying parameters. We develop specification tests against a time saturated model where there is a dummy for each age-cohort combination. The method is applied to an analysis of the obesity epidemic in England using survey data.

1 Introduction

We use repeated cross-section data to disentangle the socio-demographic determinants of the rise in obesity rates in England. We examine models for both continuous and binary measurements of obesity. The determinants are individuals' age and birth cohort, the period of observation, and other individual characteristics such as sex, race, education, and socio-economic status. There is a well-known identification problem when working with age, period and cohort. Given that we are interested in these time effects we use a recently suggested parametrization of the age, period, and cohort (APC) effects that is both freely varying and invariant to the identification problem. We develop a specification test for this parametrization in the context of repeated cross-sections. This resembles a deviance test in comparing against a full saturation of the time effects and applies for both discrete and continuous outcomes. When applying the methods to English obesity data we find that for both men and women the data can be parsimoniously described using an age-cohort model for women and an age-drift model for men.

Adult obesity almost tripled in the UK in the years between 1980 and 2011, with over a quarter of adults estimated to be obese by 2016 (Department of Health, 2011; Moody, 2016). An individual is classified as obese if their body mass index (BMI, defined in equation 1) exceeds 30. Excess weight is linked to numerous immediate and long term health risks, the most well-known being type II diabetes. Being obese can also have negative psychological consequences (Moody, 2016; Department of Health, 2011; Hruby et al., 2016). The Department of Health estimated that the total cost of obesity and overweight to the UK was about £16 billion in 2011, including both direct healthcare costs, and lost earnings due to sickness and premature mortality (Department of Health, 2011). Reducing obesity has therefore been a policy goal for many years, with specific government directives issued in 2007, 2011, and 2016.

In this paper we investigate the socio-demographic determinants of the rise in obesity. We use data from the 2001 through 2014 waves of the Health Survey for England. Our dependent variable is either the continuous measure of log BMI or the obesity indicator. The explanatory variables include age and the period of observation, from which we construct cohort through *cohort* = *period* - *age*, as well as other socio-demographic variables including education and smoking behaviour.

We employ a generalized linear model with APC time effects and other variables. The three distinct time effects may reflect different underlying factors that are difficult to measure directly. For example, period effects may reflect environmental conditions, while cohort effects might capture habits formed by generation-specific experiences. A preliminary analysis can guide subsequent research into these factors. Estimates of these time effects can also be used to produce forecasts.

The APC time effects in the model are not fully identified, which is a well-known problem, see Holford (1983), Clayton & Schifflers (1987), Glenn (2005), and Carstensen (2007). We reparametrize the model in terms of freely varying parameters as suggested by Kuang et al. (2008), henceforth KNN. The generalized linear models then become regular exponential families where the freely varying parameters are canonical. Two important features of the canonical parametrization are as follows. First, it is easy to impose restrictions on the time effects and count the associated degrees freedom. Second, it is simple to incorporate extensions beyond the time horizon of the sample. For instance, we may want to conduct recursive analysis where the number of waves change, or forecast beyond the last sample period. The canonical parametrization is invariant to such changes.

When working with the time effects we are effectively thinking of the data as a two-way array in age and period with lots of individual information in each cell of the array. In the data we have 53 age groups, 14 period groups, and 56 cohorts. In the asymptotic analysis we keep the dimension of the age-period array fixed and exploit the individual level information for inference. In light of the canonical parametrization the statistical analysis is then fairly simple. This asymptotic approach resembles earlier work for aggregate data by Martínez Miranda et al. (2015), who developed a Poisson model for counts of cancer data in which the count in each cell increases corresponding to an increase in the number of individuals in the cell. Recently, Harnau & Nielsen (2017) presented a similar model for over-dispersed Poisson data. Other papers working with aggregate data include Fu (2016) in which the author studies a class of constrained estimators where the dimension of the array increases. That approach would be inappropriate for our data given its small period range. There is also a Bayesian approach to aggregate data presented by Smith & Wakefield (2016).

The Bayesian approach has been used in models with individual data. A prominent model is the hierarchical age-period-cohort model by Yang & Land (2006), which is further generalized in the cross-classified random effects model by Yang (2008). The latter model has been used to study obesity by Reither et al. (2009) and An & Xiang (2016). These models impose a quadratic age structure, which is a testable restriction in our model. More importantly, the models do not fully address the identification problem since priors are imposed on both identified and non-identified parameters. It is well-known that the likelihood cannot update all of these parameters. In particular, the conditional prior for the non-identified parameters given the identified parameters is not updated, see Poirier (1998) and Nielsen & Nielsen (2014).

We also present a specification test for the age-period-cohort structure. Our focus on the two-way array yields a natural alternative specification where each age-cohort cell has its own parameter. We call this a time saturated (TS) model. The proposed test resembles a deviance-type test. This works both for discrete and continuous dependent variables. Inference is standard, but there are some numerical challenges which we address.

When applying these methods to the data from the Health Survey for England, 2001-2014, we find that an age-drift model fits the data on women while an age-cohort model fits the data on men. These models are consistent with the idea that obesity rates and mean BMI are both increasing over time for the aggregate population. The age-drift model for women includes an increasing linear plane and a deviation from linearity in the age dimension which takes the form of acceleration to age 50 and deceleration thereafter. For men, the non-linearity also involves an acceleration effect for the 1960's cohorts. The effects of covariates are broadly consistent with existing literature.

The paper is outlined as follows: §2 introduces the elements needed to understand the approach, including the data used in the application, the notation employed, and a more detailed summary of both the APC identification problem and the ideas in KNN. §3 and §4 contain the main theoretical contributions of this paper. First for the normal and then for the logit, conditions for standard inference are discussed, a new test is proposed for assessing mis-specification of the time effects, and an algorithm for this test is developed. In §5 the situation in which the time effects are nuisance parameters rather than direct objects of interest is considered. §6 contains the application of the methods to the question of obesity trends in England, while §7 concludes.

2 Data and statistical model

We describe the data, which are repeated cross-section with an age-period-cohort structure. The statistical model is defined, in which the classical APC problem is addressed using a reparametrization.

2.1 Obesity Data

The data used is drawn from the Health Survey for England[1] (HSE) and analysed using the R-package apc, version 1.3.3; see Nielsen (2015). The data is a repeated cross-section of a representative sample of the English population. We use waves from 2001-2014 as these include the National Statistics Socio-Economic Classification, which is one of our explanatory variables.

We observe 81,393 individuals of which 43,077 are women and 38,316 are men. We analyse women and men separately, and index the observations by $h = 1, \dots, H$, reserving the letter i for the age index. The data is repeated cross-section, so each individual is observed only once. For each individual we have information on weight and height directly measured by a registered nurse, so we do not worry about self-reporting bias in the outcome variable. From these we compute body mass index as

$$BMI = (\text{weight in kg})/(\text{height in metres})^2. \quad (1)$$

A small number of observations have BMI outside the range 12 to 60. These were presumed to be subject to measurement error and were excluded.

In addition to BMI we have data on age and the period in which the individual is observed. We consider as covariates ethnicity, level of education, NSSEC at three and at eight levels of specificity, smoking history, and alcohol consumption. Descriptive statistics are reported in tables 12 and 13 in Appendix C.

We consider two choices of dependent variable: either log BMI or an indicator for obesity defined as $BMI \geq 30$. For each individual h then Y_h is the dependent variable, i_h is the individual's age, j_h indicates the period in which the individual is observed, and k_h is the cohort of the individual which is constructed from i_h and j_h. Finally, Z_h is the d_z-length vector of covariates.

In this dataset age and period vary in a rectangular array, where age is between 28 and 80 inclusive and period is between 2001 and 2014. We therefore have $I = 53$ age groups and $J = 14$ period groups. Cohort therefore varies between 1921 and 1986. However, we exclude the first and last five cohorts because these cohorts are sparsely observed. This leaves $K = 56$ cohort groups. The range of the data, as an age-period array, is shown in Figure 1. The shading in that figure gives an indication of the variation in survey size with the period.

[1] https://discover.ukdataservice.ac.uk/series/?sn=2000021

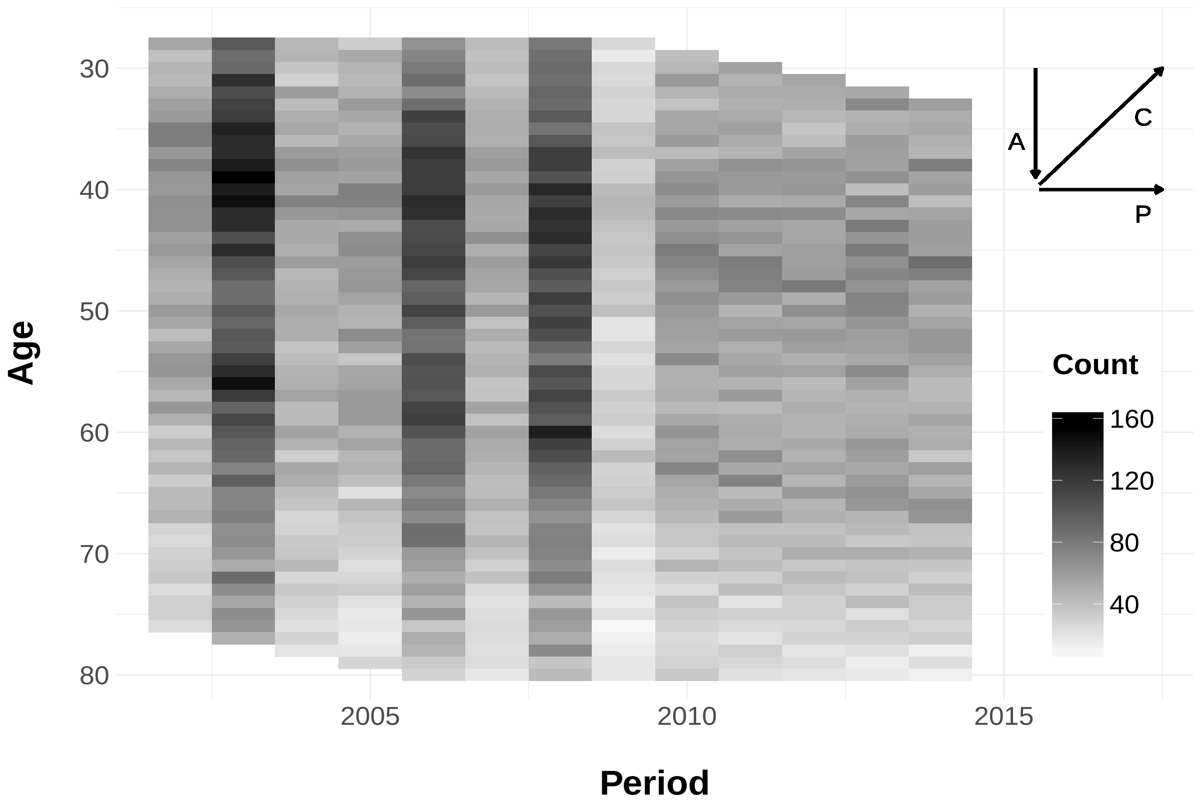


Figure 1: Within-cell observation counts women

The data is an example of a generalized trapezoid in the sense of KNN. We find it is easier to switch from an age-period coordinate system to an age-cohort coordinate system, because of the age-cohort symmetry in the relation $age + cohort = period$. Thus, throughout the paper we consider an age-cohort array, where $i = 1, \ldots, I$ is the age index and $k = 1, \ldots, K$ is the cohort index. We define the period index through $j = i + k - 1$ and get an index set of the form

$$1 \leq i \leq I, \qquad 1 \leq k \leq K, \qquad L + 1 \leq j \leq L + J. \tag{2}$$

Here L is the necessary offset in the period index due to beginning the age and cohort indices at 1. With the present data, we then have that $I = 53$, $J = 14$, $K = 56$, $L = 48$ so that $age = 28$, $per = 2001$, $coh = 1921$ correspond to $i = 1$, $j = L + 1$, $k = 1$.

2.2 Generalized Linear Model

We use a generalized linear model for the dependent variable, Y_h, where the linear predictor η_h is a function of the covariates and the age-period-cohort structure as described below. In particular, for continuous dependent variables a normal model is employed which has the form

$$Y_h = \eta_h + \varepsilon_h \qquad \text{for } h = 1, \ldots, H. \tag{3}$$

The errors ε_h are independent over individuals and normally distributed conditional on the linear predictor: $\varepsilon_h \sim \mathsf{N}(0, \sigma^2)$. For dichotomous dependent variables, a logistic model is employed with

$$\log \frac{\mathsf{P}(Y_h = 1)}{\mathsf{P}(Y_h = 0)} = \eta_h \qquad \text{for } h = 1, \ldots, H. \tag{4}$$

The linear predictor η_h is individual-specific and has the form

$$\eta_h = Z_h' \zeta + \mu_{i_h k_h}. \tag{5}$$

Here, ζ is a d_z-length vector of parameters while $\mu_{i_h k_h}$ describes the age-period-cohort structure as follows: an individual h with age i_h and cohort k_h observed in period $j_h = i_h + k_h - 1$ will have linear predictor $\mu_{i_h k_h}$ where

$$\mu_{ik} = \alpha_i + \beta_j + \gamma_k + \delta. \tag{6}$$

In the above α_i, β_j, γ_k are fixed effects for age i, period j, and cohort k respectively; we refer to these as time effects. The full set of such time effects is of dimension q, where $q = I + J + K + 1$. Collecting the time effects as

$$\theta = (\alpha_1, ..., \alpha_I, \beta_{L+1}, ..., \beta_{L+J}, \gamma_1, ..., \gamma_K, \delta)' \tag{7}$$

we can write $\mu_{i_h k_h} = D_h' \theta$ where D_h is a q-dimensional design vector of age, period, and cohort indicators and an intercept. It is well known that the vector θ is not fully identified. We address this in the following.

2.3 Identification Problem

It is not possible to identify the vector of time effects θ from the likelihood, since for any constants $a, b, c, d \in \mathbb{R}$ the predictor μ_{ik} in (6) satisfies

$$\mu_{ik} = \{\alpha_i + a + (i-1)d\} + \{\beta_j + b - (j-1)d\} \\ + \{\gamma_k + c + (k-1)d\} + \{\delta - a - b - c\}, \tag{8}$$

see for instance Carstensen (2007). The fact that this holds for any set of arbitrary constants makes it impossible to identify the linear parts of the time effects.

There are two ways to address the identification problem. First, we can work with an identified version of the original parameter vector θ, by imposing four constraints, if we keep track of the consequences for interpretation, count of degrees of freedom, plotting, and forecasting (see Nielsen & Nielsen (2014) and §5 of this paper). Second, we can reparametrize the model in terms of a freely varying parameter, which is invariant to the class of transformations in (8), and which is of a lower dimension, $p = q - 4$, than the original time effect. We follow the second approach.

The reparametrization is expressed in terms of the p-dimensional parameter vector ξ and a design vector X_h following KNN. The parameter vector is

$$\xi = (\upsilon_o, \upsilon_a, \upsilon_c, \Delta^2\alpha_3, \dots, \Delta^2\alpha_I, \Delta^2\beta_{L+3}, \dots, \Delta^2\beta_{L+J}, \Delta^2\gamma_3, \dots, \Delta^2\gamma_K)'. \tag{9}$$

Here, $\upsilon_o, \upsilon_a, \upsilon_c$ parametrize a linear plane while the double differences, e.g. $\Delta^2\alpha_i = (\alpha_i - \alpha_{i-1}) - (\alpha_{i-1} - \alpha_{i-2})$, measure deviation of the time effects from that plane. The p-dimensional design vector X_h combines the available information on an individual's age and cohort such that $\mu_{i_h k_h} = X_h' \xi$. The reparametrization can be expressed in terms of a $p \times q$ transformation matrix A' which satisfies $D = XA'$ and $A'\theta = \xi$. Further details are given in Appendix A.

This reparametrization confers the following advantages. First, the vector ξ is invariant to the transformations in (8). For instance, the unidentified age effects $\alpha_i + a + (i-1)d$, $\alpha_{i-1} + a + (i-2)d$, etc. yield double difference $\Delta^2\alpha_i$ regardless of the values of a, d. Moreover, the linear plane parameters $\upsilon_o, \upsilon_a, \upsilon_c$ can be chosen to be invariant to the transformation in (8), see Appendix A. Second, the design matrix X, formed from stacking X_h, has full column rank, whereas the design matrix D, formed from D_h, has reduced column rank. Third, there is a unique ξ that can satisfy the relation between X_h and the linear predictor for all $h = 1, ..., H$, so that $\forall\ \xi^\dagger \neq \xi$ it holds that $\mu(\xi^\dagger) \neq \mu(\xi)$. Finally, for linear exponential family models such as the normal and logit models used here, ξ is the canonical parameter.

2.4 Sub-models

We can examine whether all parts of the age-period-cohort structure are necessary. This is achieved in the same way as was done for aggregate data (Nielsen, 2014). A variety of restrictions are of interest, which are explored empirically in §6.

First, we can test for the absence of non-linearities in one of the time effects. For instance, we test period non-linearities by imposing $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+J} = 0$. In terms of the unidentified original parametrization, this is written as $\beta_{L+1} = \ldots\beta_{L+J} = 0$. This gives an age-cohort (AC) model. The two formulations of the hypothesis are in fact equivalent, see Nielsen & Nielsen (2014) for a formal analysis. The latter formulation gives the misleading impression that we simultaneously test for the absence of both non-linear and linear period effects, which is not the case.

Second, we can test for the absence of non-linearities in two components in a similar way. To test the period and cohort non-linearities we impose $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+J} = 0$ and $\Delta^2\gamma_3 = \cdots = \Delta^2\gamma_K = 0$, while leaving the linear plane unrestricted. Clayton & Schifflers (1987) refer to this as the age-drift (Ad) model.

Third, a model (A) with only age effect or, equivalently, no non-linearities or linearities in the period and cohort effects arises by imposing $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+J} = 0$ and $\Delta^2\gamma_3 = \cdots = \Delta^2\gamma_K = 0$, as well as $\upsilon_c = 0$. This leaves a model of the type $\mu_{i,k} = \alpha_i$.

Finally, we will be interested in a linear plane model (t), where there are no non-linearities of any kind. Then $\Delta^2\alpha_3 = \cdots = \Delta^2\alpha_I = 0$, and $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+J} = 0$ and $\Delta^2\gamma_3 = \cdots = \Delta^2\gamma_K = 0$. This model corresponds to chosing the time effects as linear functions so that $\alpha_i = \alpha_0 + \alpha_1 i$ and $\beta_j = \beta_0 + \beta_1 j$ and $\gamma_k = \gamma_0 + \gamma_1 k$.

3 Statistical Analysis of the normal model

We have described the data and the model we want to use, and have shown how identification of that model is achieved through reparametrization. We proceed to discuss estimation of the normal reparametrized model (5).

3.1 Estimation

For the purposes of discussing estimation, it is convenient to stack observations. We define Y as the $H \times 1$ vector of individual observations on the dependent variable, $Y = [Y_1, ..., Y_H]'$. Correspondingly let η, μ, X, and Z represent stacked individual information. We then have

$$\eta = Z\zeta + \mu, \qquad \mu = X\xi. \tag{10}$$

We embed this in the normal model (3) which can be estimated by least squares regression of Y on (X, Z), noting that ζ and ξ are freely varying parameters.

3.2 Inference

Following estimation, we conduct inference on the estimated model. We investigate whether some elements of the full APC structure are unnecessary. Achieving a more parsimonious representation is desirable for forecasting purposes and for ease of interpretation.

Exact inference on the OLS parameter estimates $(\hat{\zeta}, \hat{\xi})$ can be performed by appealing to the classical results for analysis of variance in the linear model. Since it is desirable to relax the strict assumption of conditional normality of ε, we also consider asymptotic inference. For this we must be clear about the repetitive structure. We treat the dimensions of the generalized trapezoid in age-cohort array as fixed: the numbers of ages, periods, and cohorts,

I, J, and K, does not change. It is the number of individual observations H that is assumed to increase.

To justify asymptotic inference we make the following assumptions:

1. The triplets Y_h, X_h, Z_h are independent and identically distributed across individuals h.
2. The regressors X_h, Z_h jointly have a positive definite covariance matrix.
3. The errors ε_h have zero conditional mean and finite variance: $\mathsf{E}(\varepsilon_h|X_h, Z_h) = 0$ and $\mathsf{Var}(\varepsilon_h|X_h, Z_h) = \sigma^2$.

It is a consequence of these assumptions that as we increase the sample size H the relative frequency of individuals at all age-cohort combinations should remain constant. This could be considered problematic given that the size of the HSE survey varies from year to year in a way that does not reflect changes to the underlying UK population. However, from Wooldridge (2010, §19.4) we know that maximum likelihood estimators will remain consistent and asymptotically normal, even in the presence of sample selection on exogenous variables. This requires that the distribution of the outcome variable conditional on a selection indicator and the exogenous variables is equal to the distribution conditional on the exogenous variables alone. That is, the distribution of log BMI among those who were selected for the HSE should be the same, conditional on X and Z, as the distribution of log BMI among those who were not selected. This means that the fact that a particular year happened to be over-represented in this survey must not be related to log BMI. Since the variation in sample size across years is due to financial constraints of the surveying body this seems reasonable in our setting. In a similar vein, these assumptions imply that selection into the HSE is independent of the covariates Z_h. This is plausible as the HSE is a representative sample. The distribution of Z_h may vary between age-cohort cells.

Under these assumptions, inference can be performed in the usual way. Under exact normality, t- and F-tests can be used. Under asymptotic inference, likelihood ratio tests are asymptotically χ^2. Such tests can be used to investigate the sub-models outlined in §2.4.

3.3 Misspecification Testing

We describe a new misspecification test for the normal APC model and explain how computational issues encountered in developing this test were overcome.

3.3.1 Normal model misspecification test

Once a model has been estimated on data, it is important to evaluate how well that model describes the variation present in the data. In a discrete data context, this is often done using a deviance test, which can be thought of as a comparison between the model of interest and a “fully-saturated” model. The fully-saturated model has degrees of freedom exactly equal to the number of observations, so all variation in the data is described by the model. A comparison between this fully-saturated model and the more parsimonious model of interest indicates how well the assumptions of the more parsimonious specification match the data.

Building on the ideas behind the deviance test, we develop a new strategy for testing the specification of the age-period-cohort component of a model. This is achieved by saturating the age-cohort array resulting in a model that we refer to as the time-saturated model. We replace the design matrix X with a matrix T of dimension $H \times n$, where n is the number of cells in the age-cohort array. Each row of the matrix T is a unit vector, indicating the age-cohort cell to which that individual belongs. Consequently, $T'T$ is diagonal. All of the earlier statistical analysis carries through, so it is possible to compare a model of the form $Y = \eta + \varepsilon$ where $\eta = Z\zeta + X\xi$ with the more general model where

$$\eta = Z\zeta + T\kappa. \tag{11}$$

F-tests are used under classical normality assumptions, which become χ^2 likelihood ratio tests under asymptotic assumptions. However, the dimension of the general time-saturated model is sufficiently large that it poses numerical issues, which we now discuss.

3.3.2 Computational challenges

The overall dimension of the combined design matrix $M = (Z, T)$ in the above general model is $H \times (d_z + n)$. In the data example, $d_z = 15$ with $n = 684$. Consequently it is challenging to evaluate and to invert $M'M$ using a computer due to memory allocation. We can address this problem by orthogonalizing the regressors and exploiting the unique structure of the design matrix T. Instead of estimating equation (11) directly, we evaluate the partitioned regression

$$Y = \left[Z - T(T'T)^{-1}T'Z\right]\zeta + T\rho + \varepsilon. \tag{12}$$

Here $\left[Z - T(T'T)^{-1}T'Z\right] = \tilde{Z}$ is the residual of a first-stage regression of Z on T.

Since $T'T$ is diagonal by virtue of the dummy structure of T, we do not need to store the entire matrix $T'T$; we need only store the vector of elements of the main diagonal. We can thus avoid the memory allocation problem associated with $M'M$. Because $T'T$ is diagonal the inverse is found by taking the reciprocal of the diagonal elements. It is therefore easy to calculate $\tilde{Z}$. Since $\tilde{Z}$ and T are orthogonal by construction, ζ and ρ can easily be estimated by regression of Y on $\tilde{Z}$ and T, respectively. This poses no computational challenge since $\tilde{Z}$ is of dimension $H \times d_z$ and d_z is small.

We can retrieve κ from $\hat{\kappa} = \hat{\rho} - (T'T)^{-1}T'Z\hat{\zeta}$. Note that the equations (11) and (12) give equivalent models with the same fit and the same residual variance. As a consequence we are normally not interested in the value of $\hat{\kappa}$.

We can test the APC model against the saturated model using

$$F = \frac{(RSS_X - RSS_T)/(d_T - d_X)}{RSS_T/(H - d_T)}, \tag{13}$$

where d is the number of parameters in a model and H is the number of individuals. RSS is the residual sum of squares from a model. Subscripts indicate the model in question: X refers to the APC model, because it has design matrix X for the time component, while T refers to the TS model. RSS_T is equal to the residual sum of squares from the model (12). This F-statistic is asymptotically χ^2, or F-distributed under exact inference.

4 Statistical Analysis of the logit model

We discuss analysis of a logit model of form (4), with η specified as in (10).

4.1 Estimation

The logit log-likelihood is

$$\ell(\zeta, \xi) = \sum_{h=1}^{H} \eta_h Y_h - \sum_{h=1}^{H} \ln\left(1 + \exp \eta_h\right). \tag{14}$$

There is no closed-form expression for the ζ, ξ that maximise this log-likelihood. However, the log-likelihood is strictly concave when the design matrix has full rank so the maximum likelihood estimator is unique (Wedderburn, 1976). It is finite in the absence of separation or quasi-separation (Agresti, 2013, §6.5). Under these conditions the maximum likelihood estimator can be found by Newton iteration.

4.2 Inference

The asymptotic theory of the estimator is outlined by Fahrmeir & Kaufmann (1986). Their Theorem 2 shows consistency and asymptotic normality under the following assumptions:

1. the triplets $\{Y_h, X_h, Z_h\}$ are independent, identically distributed;
2. the regressors X_h, Z_h have a positive definite covariance matrix.

The asymptotic variance-covariance matrix of this estimator is given by $J = -\ddot{\ell}$, for $\ddot{\ell}$ the second derivative of the log-likelihood. Theorem 3 of Fahrmeir & Kaufmann (1986) shows that likelihood ratio test statistics are asymptotically χ^2. These can be used to examine whether any of the covariates in Z are redundant, as was done for the normal model. They can also be used to test the necessity of different parts of the APC structure as outlined in §3.2.

4.3 Misspecification Testing

The time-saturated (TS) model introduced in §3.3 can also be used to test the specification of the logit APC model, again by comparing the fit of the two models. The linear predictor is exactly as it appears in equation (11). However it is now embedded in a logit model of the form in equation (4).

We face similar numerical problems in seeking to estimate the logit TS model as we did in the normal case, because of the size of the design matrix. We address these by exploiting the structure of the derivatives of the log-likelihood in conjunction with the dummy structure of T.

The time-saturated model here is a logit model with mean $\eta = Z\zeta + T\kappa$ given by equation (11). Thus, the score is

$$\dot{\ell} = \begin{pmatrix} Z' \\ T' \end{pmatrix} (Y - \Pi), \tag{15}$$

where Π is a H-length vector of logistic probabilities π_h that depend on individual values of Z_h and X_h through (11). The matrix J as defined in §4.2 is

$$J = -\ddot{\ell} = \begin{pmatrix} Z' \\ T' \end{pmatrix} W \begin{pmatrix} Z & T \end{pmatrix} = \begin{pmatrix} J_{ZZ} & J_{ZT} \\ J_{TZ} & J_{TT} \end{pmatrix}, \tag{16}$$

where W is a diagonal matrix of Bernoulli variances, $\pi_h(1 - \pi_h)$. Using partitioned inversion we find, with $J_{ZZ \cdot T} = J_{ZZ} - J_{ZT} J_{TT}^{-1} J_{TZ}$, that

$$J^{-1} = \begin{pmatrix} J_{ZZ\cdot T}^{-1} & -J_{ZZ\cdot T}^{-1} J_{ZT} J_{TT}^{-1} \\ -J_{TT}^{-1} J_{TZ} J_{ZZ\cdot T}^{-1} & J_{TT}^{-1} + J_{TT}^{-1} J_{TZ} J_{ZZ\cdot T}^{-1} J_{ZT} J_{TT}^{-1} \end{pmatrix}. \tag{17}$$

Here we make use of the dummy structure of T. In the normal model, we used the fact that $T'T$ is diagonal. Here, we exploit the fact that the large-dimensional matrix $J_{TT} = T'WT$ is diagonal, and so we need only deal with the main diagonal as a vector. This greatly reduces the computational cost of calculating J^{-1}.

To estimate the parameters of the time-saturated logit model we again use a Newton iterative procedure. We initialize the parameters at zero, and update using the score and inverse observed information calculated by the above formulas, noting that the diagonal structure is preserved in each step.

Given the estimated values of the parameters attained by this procedure, we can calculate the log-likelihood of the time-saturated model for the data in question. This can then be compared to the log-likelihood of the APC model by a likelihood ratio test. Under the assumptions outlined in §4.2 the likelihood ratio test statistic will be asymptotically χ^2.

5 Ad hoc identification

When a researcher is directly interested in the effects of age, period, and cohort the choice of identication is important and the canonical parametrization of equation (9) is preferred for the reasons outlined in §2.3. This is true of our obesity analysis. However, in other situations researchers are primarily interested in the effect of a covariate such as education, but need to control for age, period, and cohort effects. An example is found in Ejrnæs & Hochguertel (2013) where the authors are interested in the effect of insurance status on unemployment but must isolate this from the effects of age and cohort. In that situation it does not matter how the time effects are identified.

Recall the model for the linear predictor as outlined in equation (10):

$$\eta = Z\zeta + \mu, \qquad \mu = X\xi.$$

This can be viewed as the linear predictor in the normal model, the logit model or any other generalized linear model. The maximum likelihood estimators are denoted $\hat{\xi}, \hat{\zeta}$. In the preceding sections $\hat{\xi}$ was of interest; we now consider a situation where only the estimator $\hat{\zeta}$ is of interest. Now, suppose that we identify the age-period-cohort structure differently so that

$$\eta = Z\zeta + \mu, \qquad \mu = XQ\phi,$$

where Q is a known, invertible $q \times q$-matrix and $\phi = Q^{-1}\xi$. We suppose that the researcher has arrived at this formulation not via the canonical parametrization $X\xi$ but from some other identification strategy. Maximum likelihood gives the estimators $\hat{\phi}, \hat{\zeta}_\phi$, say. The two sets of parameters are linked through

$$\begin{pmatrix} \xi \\ \zeta \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} \phi \\ \zeta \end{pmatrix}.$$

The mapping is one-one since Q is invertible. Due to the equivariance of maximum likelihood estimators (Davidson & MacKinnon, 1993, §8.3) we have in the same way that $(\hat{\xi}, \hat{\zeta}) = (Q\hat{\phi}, \hat{\zeta}_\phi)$ and in particular $\hat{\zeta} = \hat{\zeta}_\phi$. Thus, the estimator for ζ is invariant to the choice of Q.

We note that the reparametrization

$$\xi = Q\phi \tag{18}$$

covers a range of ad hoc identification schemes appearing in the age-period-cohort literature. By ad hoc we mean that the identification is not invariant to the group of transformations described in (8).

An example of ad hoc identification is the constraint, from Mason et al. (1973),

$$\alpha_1 = \alpha_2 = \beta_J = \gamma_K = 0. \tag{19}$$

To conduct estimation given this constraint, the columns of the design D corresponding to $\alpha_1, \alpha_2, \beta_J, \gamma_K$ are dropped. This gives a design matrix D_λ of dimension $n \times p$ which has full column rank. Thus, this approach replaces $D\theta = X\xi$ with $D_\lambda\psi$, for a p-vector ψ, which makes regression computationally feasible. By combining ψ with the four constraints in equation (19) a q-vector is formed. In Appendix B we show that the model implied by (19) is indeed of the form $\xi = Q\phi$ by finding Q and ϕ. Note that this ad hoc identication is not invariant to (8), since replacing α_i by $\alpha_i + a$ for some arbitrary constant a does not respect the constraint $\alpha_1 = 0$. Thus the estimated age, period, and cohort effects are with reference to these constraints.

The analysis of Ejrnæs & Hochguertel (2013) actually imposes a quadratic constraint in addition to ad hoc identification and so cannot be represented in the form $\xi = Q\phi$. Specifically they assume

$$\mu_{ik} = \alpha_{(1)}i + \alpha_{(2)}i^2 + \beta_j + \gamma_{(1)}k + \gamma_{(2)}k^2 + \delta \qquad \text{with} \qquad \beta_1 = \beta_2 = 0.$$

Here, the level and the slope of the period effect β_j are not identified, hence the need for the ad hoc identification by $\beta_1 = \beta_2 = 0$. The constraint imposed by this model can be expressed in terms of a testable linear restriction on the canonical parameter. Since the age and cohort effects are quadratic their double differences are constant, see Nielsen & Nielsen (2014, §5.4.5). This linear restriction can be tested against the unrestricted age-period-cohort model as well as the time saturated model using the tests outlined above. Subject to imposing this linear restriction on the canonical parameter the estimate for ζ will be the same whether one uses this restricted canonical parametrization or the ad hoc identication.

6 Empirical Application

We apply the methods outlined above to examine trends in obesity in England using the data described in §2.1. The object of the analysis is to establish whether a better model of this epidemic can be achieved by decomposing the aggregate trend in terms of the reparametrized time effects. We assess whether the APC model or any of its sub-models is sufficient to describe the trends in the data.

6.1 Preliminary Data Analysis

We begin by visually inspecting the data. The heatmap in Figure 1 of §2.1 displays the number of women observed in the data, disaggregated by age-cohort cell. The observation counts for men are similar. The demographic bulge in cohorts from the mid-1940s to the mid-1970s is evident from the slightly darker shading across the centre of Figure 1. In certain years there were substantially more observations than in other years; this is a consequence of budget constraints affecting the HSE. As discussed in §3.2, this contradicts the assumption that the data is IID. To ensure this did not affect our analysis we ran robustness checks described in §6.3.

Figures 2 and 3 show the mean values of BMI in each age-cohort cell for men and women respectively. Overall women have lower BMI than men. The broad pattern of changes over age, cohort, and period is similar between the sexes. BMI means are lowest in the top-right corner of the graph, indicating low BMI either among later cohorts or at younger ages. The prevalence of darker shades towards the right sides of figures 2 and 3 suggests a period effect. The lighter shades in the bottom-left of figure 3 indicates that earlier cohorts of men may have lower mean BMI. Our analysis will help to disentangle the contributions of these time effects.

6.2 Covariates

The covariates were adapted to facilitate regression analysis. Descriptive statistics are reported in tables 12, 13 in Appendix C.

For education, those who left school after attaining a GCSE or equivalent qualification were taken as the reference group, and we included three dummies: education below GCSE level, holders of a university degree, and those with education beyond GCSE but below degree level. More detail on the harmonization of different qualifications is available in the HSE documentation.

Smoking behaviour is captured by two dummies. One records whether an individual currently smokes, while the other captures former regular smokers.

For alcohol consumption, the casual drinking population (those drinking one to four times a week) was taken to be the reference and dummies were introduced classifying individuals as not drinking at all, drinking rarely (less than once a week), and drinking frequently (five or more times a week).

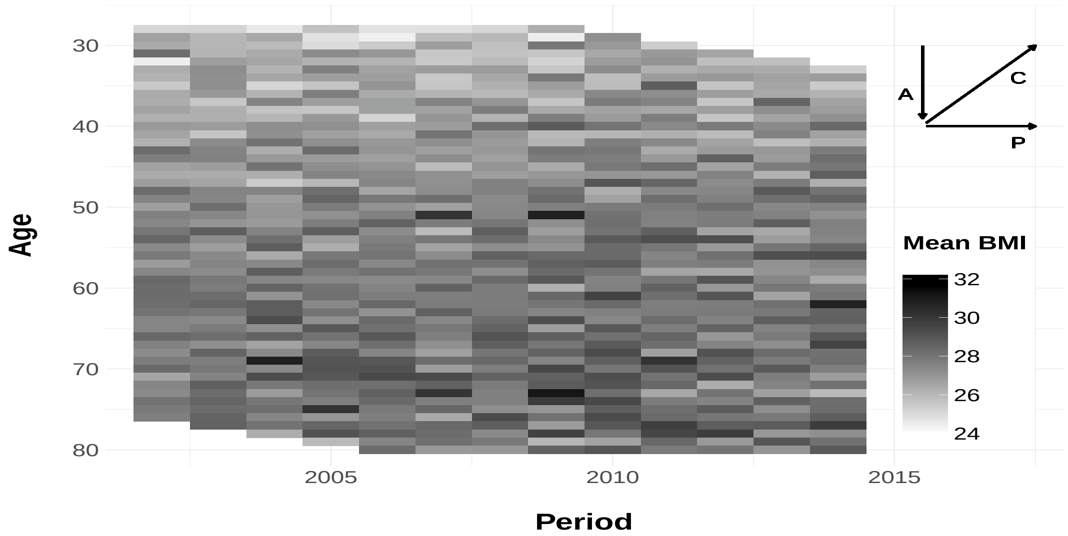


Figure 2: Within-cell BMI means women

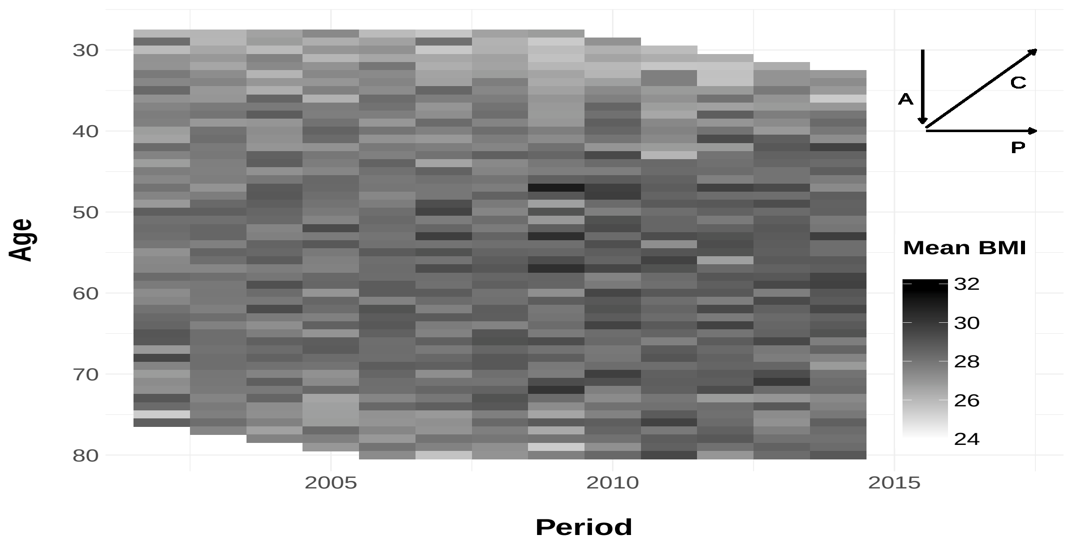


Figure 3: Within-cell BMI means men

The reference ethnicity was taken to be white, with dummies for whether an individual identified as black, Asian, of mixed ethnicity, or of "other" ethnicity (including e.g. Arab).

The National Statistics Socio-economic Classification (NSSEC) is used at three levels of specificity. The reference category is "Routine and Manual" occupations. Indicators are included for "Intermediate", "Managerial and Professional", and "Other" occupation groups. The "Other" group includes the students, those permanently outside the labour force, the long-term unemployed, and anyone whose employment could not be satisfactorily classified.

6.3 Normal Model

The techniques described in §3 are used to estimate and conduct inference on a model with log BMI as the dependent variable. The explanatory variables are the covariates described in the preceding section and the reparametrized APC structure. The model is fitted separately for men and women. The analysis begins with fitting the TS model, the APC model, and all sub-models, and comparing them using F-statistics and the Akaike Information Criterion (AIC). A preferred model is selected, and results for that model are presented and discussed.

6.3.1 Women

Table 1: **Model comparisons, log BMI, women**

	Against TS			Against APC				
	F	df	p	F	df	p	AIC	ℓ
TS							-22747.33	12101.67
APC	1.02	592	0.36				-23321.91	11796.95
AP	0.99	646	0.53	0.73	54	0.93	-23390.28	11777.14
AC	1.03	604	0.32	1.38	12	0.17	-23329.33	11788.67
PC	1.11	643	0.03	2.16	51	0.00	-23313.70	11741.85
Ad	1.00	658	0.47	0.85	66	0.80	-23397.46	11768.73
Pd	1.22	697	0.00	2.35	105	0.00	-23284.82	11673.41
Cd	1.11	655	0.02	2.00	63	0.00	-23321.56	11733.78
A	1.05	659	0.20	1.29	67	0.06	-23369.27	11753.64
P	1.51	698	0.00	4.25	106	0.00	-23084.91	11572.46
C	1.28	656	0.00	3.74	64	0.00	-23210.61	11677.31
t	1.22	709	0.00	2.25	117	0.00	-23292.97	11665.49

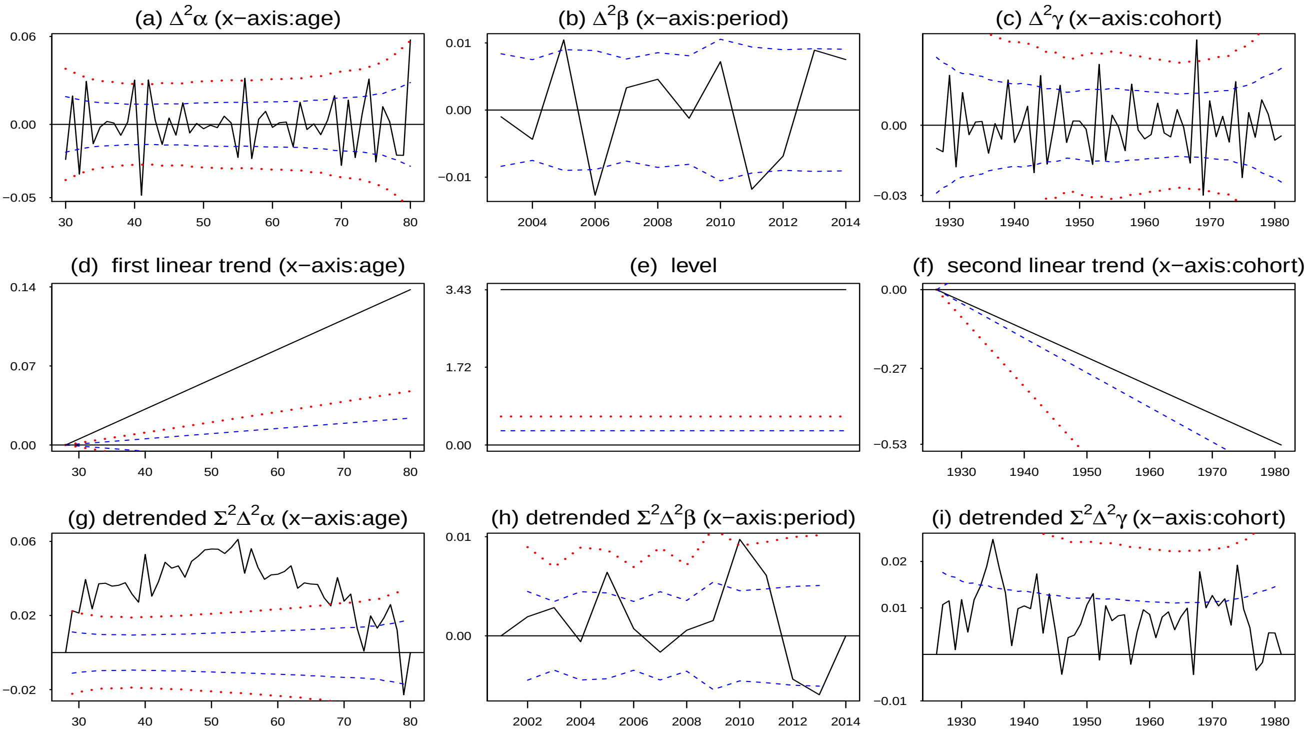
Table 2: **Model comparisons, log BMI, women**

Models compared	Ad vs AC	Ad vs AP	A vs Ad
p	0.926	0.158	0.000

For women, table 1 displays statistics that facilitate comparison between models. Each row represents a model. The results from F-tests against the TS and APC models are shown, along with the AIC and log-likelihood. We conduct model selection by first ranking models in terms of the AIC; the most preferred model is that with the smallest AIC. We then check the F-tests of these models, and select the model with the highest ranking in terms of AIC that is also not rejected by the likelihood ratio tests comparing it to larger models.

Looking at table 1, the Ad model has the smallest AIC, followed by AP and A. The F-tests comparing the AP and Ad models to the APC model are not rejected. In table 2 we conduct

Figure 4: **Time effects, APC model of log BMI, women**



solid line = estimate; blue (red) dotted line = 1 (2) standard deviation

direct F-tests between the AP, Ad, and A models. The reduction to Ad from the larger models is supported, so the Ad model is selected according to our criteria above. The A model is not supported as a reduction either from the APC or the Ad model.

We plot the estimated reparametrized time effects for the full APC model in figure 4. The middle row of panels contains the linear plane. The slopes are always estimated along the age and cohort dimensions, but they combine the linear parts of all three time effects and so are not attributable to age and cohort. The top row of panels contains the series of estimated double differences in age, period, and cohort. The cumulative effect of double differences at each age, period, and cohort is shown in the bottom row of panels. In the panels shown here, these cumulative effects have been "de-trended" post-estimation so that they begin and end at zero; the removed trends have been added on to the linear plane in the middle panels. The detrended time effects can be interpreted individually due to the two zero constraints. They show the non-linear development in the time effects over and above the unidentified linear trends. The concave shape in age is often found in epidemiological studies, see for instance Nielsen (2015).

In figure 4(h,i) we can see that neither the period nor the cohort non-linearities exceed the red dotted line that marks two standard deviations. The age non-linearity does, however. This is visual support for the earlier conclusion that the Ad model would be sufficient to describe the data.

Figure 4 could be repeated for the Ad model. That model excludes the double differences in period and cohort, so the plots in panels (b,c,h,i) fall away. Since the Ad model cannot be rejected against the APC model and since the canonical parametrization is chosen invariantly, we find that the age double differences are nearly identical in the Ad and the APC models. Thus, the corresponding figure for the Ad model has nearly the exact same panels (a,g) and we therefore omit it. This type of stability has commonly been found for APC models for aggregate data. The linear plane represented by panels (d,e,f) does however change when the double differences in period and cohort are eliminated. The second slope is now slightly positive and significiant; this explains why the direct F-test between the A and Ad models rejected the reduction to the A model as this would have restricted the second slope to be zero.

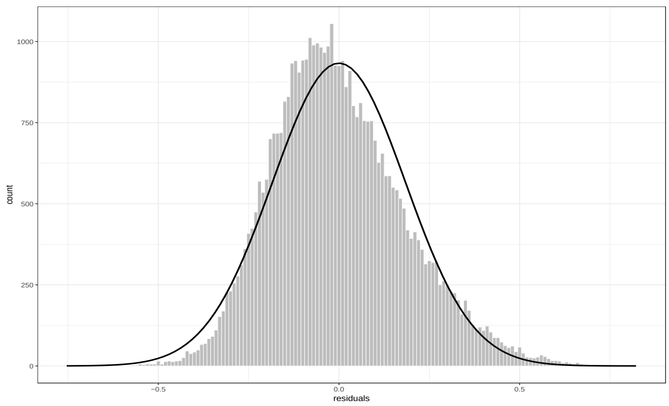
The coefficients on the covariates of the Ad model are seen in table 6. Interpretation of these is deferred to §6.3.3, where they are discussed in conjunction with the estimated effects for men.

Formal misspecification tests for the Ad model are reported in table 3 in situations with and without log transformation of the dependent variable. The tests include a cumulant based test for normality of residuals and tests for functional form misspecification and heteroskedasticity (Ramsey, 1969; White, 1980). The log transformation clearly improves the specification. Yet, given the large sample size, $n = 43,077$, it is difficult to avoid very small p-values. The histogram of the residuals in figure 5 suggests that the non-normality is not too severe. Nonetheless, as a precaution we therefore conducted various robustness checks reported in §6.3.3. Those checks indicate that the mis-specification is not detrimental for inference.

Table 3: **Ad model specification tests, women**

	BMI			Log BMI			
Test	value	statistic	p	value	statistic	p	distribution
Skewness	1.02	7488.99	0.00	0.45	1445.07	0.00	$\chi^2(1)$
Excess kurtosis	1.59	4524.01	0.00	0.24	102.92	0.00	$\chi^2(1)$
Normality test		12012.99	0.00		1547.99	0.00	$\chi^2(2)$
RESET test		23.07	0.00		18.93	0.00	F(2, 43006)
hetero test		5.20	0.00		4.94	0.00	F(120, 42956)

Figure 5: **Residuals from Ad model of log BMI, women**



solid line = normal distribution with mean and standard deviation from the data

6.3.2 Men

For the men, a similar approach is followed. First, the table comparing all candidate models is constructed, see table 4. We see that the AIC is minimized by the Cd model. However, the F-test comparing the Cd model to the APC model rejects, suggesting that there is important information lost in moving from the APC to the Cd model. Looking at the remaining models, a case could be made for either the AC model (on the basis of the F-test) or the PC model (based on the AIC). To aid selection we conducted direct tests comparing the AC, PC, and Cd models, seen in table 5. These tests suggest that age non-linearities are important, but period non-linearities are not. Thus an age-cohort model appeared optimal.

The estimated reparametrized time effects from the APC model are seen in figure 6. There is some curvature in each of age and cohort, while the period non-linearity is driven by the anomalous spike in 2010. While there was no evidence in the HSE documentation of a sampling or other technical reason for the spike, we could think of no good meaningful explanation for it, and so chose to ignore it. This lent support to our decision to exclude the PC and focus on the AC model.

The double differences, linear plane, and detrended time effects for the AC model were very similar to those from the APC model apart from the omission of panels b and h. The plot of the AC model is therefore omitted. The estimated coefficients on the covariates are seen in table 6. Misspecification tests on the residuals are similar to those for women reported in table 3 and therefore omitted.

Table 4: **Model comparisons, log BMI, men**

	Against TS			Against APC				
	F	df	p	F	df	p	AIC	ℓ
TS							-36449.64	18952.82
APC	0.98	592	0.59				-37043.86	18657.93
AP	1.06	646	0.15	1.85	54	0.00	-37051.87	18607.93
AC	0.99	604	0.56	1.22	12	0.26	-37053.16	18650.58
PC	1.03	643	0.29	1.57	51	0.01	-37065.67	18617.83
Ad	1.06	658	0.14	1.73	66	0.00	-37061.31	18600.65
Pd	1.56	697	0.00	4.80	105	0.00	-36750.82	18406.41
Cd	1.03	655	0.27	1.50	63	0.01	-37075.39	18610.70
A	1.15	659	0.00	2.63	67	0.00	-37001.23	18569.62
P	1.60	698	0.00	5.05	106	0.00	-36722.71	18391.35
C	1.21	656	0.00	3.33	64	0.00	-36958.31	18551.16
t	1.55	709	0.00	4.42	117	0.00	-36762.34	18400.17

Table 5: **Model comparisons, log BMI, men**

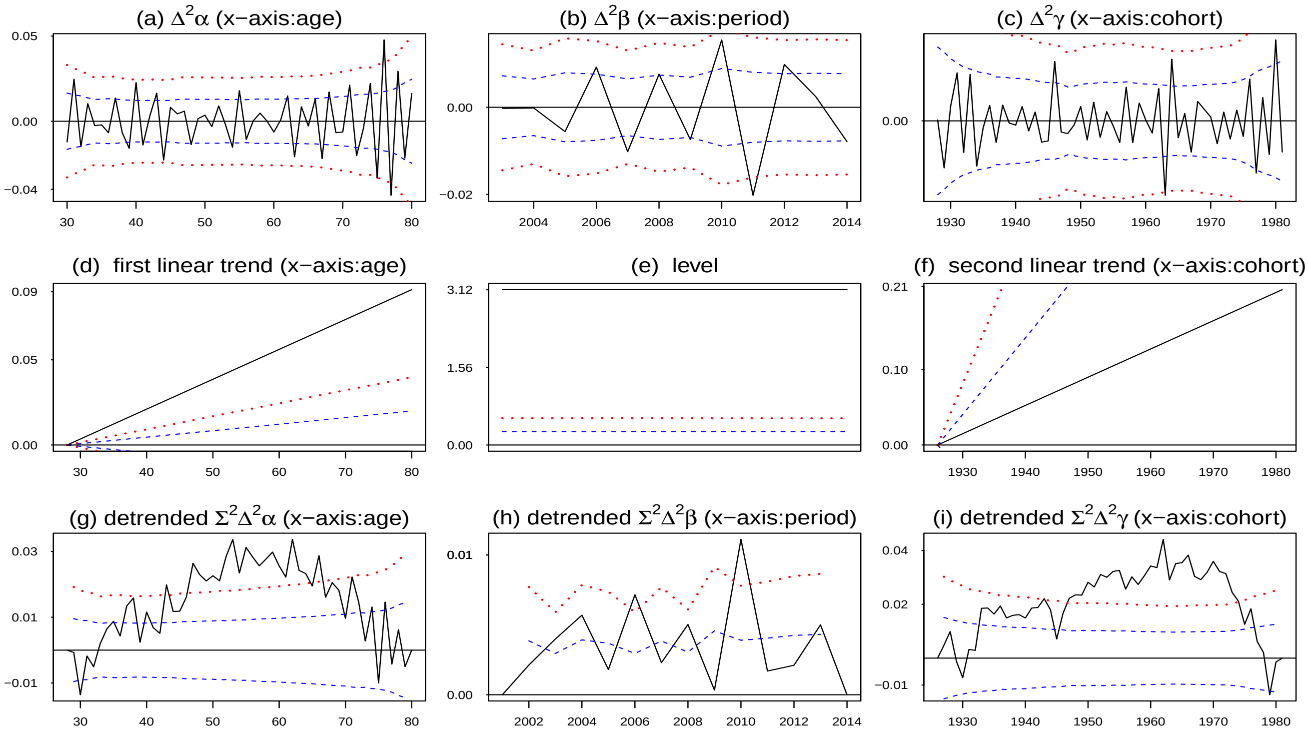
Models compared	Cd vs AC	Cd vs PC
p	0.006	0.285

6.3.3 Interpretation

Recall that for women we selected the Ad model and for men we selected the AC model. In both cases the estimated intercept is within the plausible range for BMI (about 3 on the log scale, corresponding to a BMI of 20). One slope is positive and significant in both models. The second slope is only significant in the Ad model for women, where it is positive. This general plane shape is consistent with the increase in mean BMI over time for the aggregate population, although the slope is not steep. Due to the identification problem it is impossible to say whether that is a period effect, or a result of the aging population combined with an age or cohort effect.

For both men and women, there are significant deviations from linearity. For women there is some acceleration in log BMI up to age 50, although the slope is far from smooth, and then a

Figure 6: **Time effects, APC model of log BMI, men**



solid line = estimate; blue (red) dotted line = 1 (2) standard deviation

substantial deceleration in log BMI thereafter. This may be consistent with general metabolic effects or selection effects towards the end of life, as those with higher BMI die sooner (Hruby et al., 2016). Children may also be a factor, both due to the biological effect of child-bearing on weight and the impact of child-rearing on free time for personal healthcare. For men, there is curvature in both the age and cohort dimensions. The age non-linearity is not as significant as that for women, and it begins later, suggesting that child-bearing may be an important factor among women. The significance of cohort among men is more difficult to explain, but may be related to generational shifts in the nature of employment. We hypothesize that men from the central cohorts may have similar dietary habits to men of earlier cohorts, but have a more sedentary lifestyle and do less physical labour; whereas more recent cohorts eat a more varied diet with less heavy, traditional British fare. Such factors could affect men more than women due to the long-standing social pressure on women to moderate their diets to "keep their figure". Further targeted research would be required to validate any of these hypotheses.

There is little in the way of period non-linearities; the only point at which the period effect attains significance is in 2010, where there is an unusual and so far unexplained spike in log BMI. As discussed earlier, we judge this spike to be non-informative about the evolution of BMI.

The effects of the covariates are largely as one would expect. Where they are significant, the signs of the coefficients on the ethnicity indicators are consistent with previous literature (Ogden et al., 2015; An & Xiang, 2016), as is the negative correlation between BMI and social class (McPherson et al., 2007). Those with more education have lower BMI on average, again consistent with the literature (Baum II & Ruhm, 2009; An & Xiang, 2016). Somewhat more interesting are the correlations with other negative health behaviours, alcohol and smoking. Those who currently smoke have lower BMI on average, while those with a history of smoking have higher BMI on average, than those who have never smoked. Non-drinkers and rare drinkers have higher BMI than casual drinkers (the reference group), who in turn have higher BMI than frequent drinkers. This pattern might be explained by a story of substitution between smoking, drinking, and sugar consumption, although again further research would be required to confirm this.

There are some sex differences in the covariates, primarily relating to significance. Black women and women of mixed ethnicity have significantly higher and lower BMI, respectively, than white women; whereas for men these ethnicities are not significant. Non-drinking men do not differ significantly from casual drinkers, and the effects of social class are not significant for men.

Table 6: **Covariate effects, log BMI models**

	Women, Ad			Men, AC		
	$\hat{\zeta}$	*se*	*p*	$\hat{\zeta}$	*se*	*p*
Ethnicity indicators (excl. white)						
Black	0.068	0.007	0.000	-0.008	0.006	0.208
Asian	-0.045	0.007	0.000	-0.039	0.005	0.000
Mixed ethnicity	-0.023	0.011	0.035	-0.008	0.011	0.428
Other ethnicity	-0.069	0.012	0.000	-0.033	0.011	0.004
Behaviour indicators (excl. never smoked, occasionally drink alcohol)						
Former smoker	0.024	0.002	0.000	0.026	0.002	0.000
Current smoker	-0.030	0.002	0.000	-0.045	0.002	0.000
Never drink alcohol	0.043	0.008	0.000	0.001	0.010	0.921
Rarely drink alcohol	0.037	0.002	0.000	0.014	0.002	0.000
Frequently drink alcohol	-0.032	0.003	0.000	-0.017	0.002	0.000
Education level indicators (excl. GCSE)						
Below GCSE	0.016	0.003	0.000	0.010	0.002	0.000
Some higher education	-0.012	0.003	0.000	0.002	0.002	0.393
University degree	-0.048	0.003	0.000	-0.026	0.003	0.000
3 level NSSEC indicators (excl. routine/manual)						
Intermediate occupations	-0.021	0.002	0.000	-0.001	0.002	0.696
Managerial/Professional	-0.009	0.003	0.003	-0.001	0.002	0.485
Other occupations	-0.013	0.007	0.064	-0.020	0.011	0.063

6.3.4 Robustness checks

A range of alternative specifications of the normal model were examined as robustness checks. Using the same data, we replaced the three-level NSSEC with the eight-level classification. We considered a model with log weight as the dependent variable and log height as the explanatory variable; a model with log BMI as the dependent variable implicitly imposes a coefficient of 2 in this regression, and we wanted to evaluate whether this was restrictive. These models did not change our substantive findings.

We also considered different subsets of the original HSE data. To examine whether income yielded different results to the NSSEC, we tried a specification which replaced the NSSEC with inflation-adjusted household income (quadratic in logs) using two samples: first with all observations where income information was available, then for only observations where both income and NSSEC information was available. The only substantive change to our results was that the second slope for men became borderline significant (it was positive). Given the apparent insensitivity of the estimated covariate coefficients to the APC specification, we decided that the time and covariate effects were largely orthogonal and tested a model which excluded the covariates. This gave us a much larger sample size due to less missing information. The substantive results were unchanged. Finally, to check whether the differences in sample size across years affected our results we randomly selected 2000 observations from each year and ran the original analysis on this smaller sample, using three different random seeds. The plane and age non-linearities were robust to this check for both men and women.

In our final set of robustness checks we tested extensions of the age-cohort space. We considered the original model (with NSSEC) but with the age range extended to be from 20-80, and the cohort range extended accordingly. This incorporated some cells in which perfect separation was present, but that should not be a problem in the normal model. The main consequence of this was a strengthening of the significance of age non-linearities for men, with rapid acceleration of log BMI in the early twenties. The NSSEC was not recorded prior to 2001, but we have income information back to 1997, so we were able to consider the model with income over a longer period horizon. We were also able to evaluate a no-covariates model with data back to 1992. The estimated age effects remained similar to the original models throughout. With an extended period range, the period non-linearities become significant and exhibit curvature, suggesting that there was acceleration in log BMI in the 1990s which has now ceased.

In addition to the robustness checks above, we have the misspecification tests (normality, functional form, heteroskedasticity) on the estimated models. While our mis-specification tests show imperfections in our models, they do not invalidate our results. Fat tails mean that our standard errors may be incorrect, but the estimators will still be consistent. The functional form and heteroskedasticity results might be resolved with a more careful choice of covariates. We also intend to consider heteroskedasticity arising from the APC structure in a future paper. The lack of variation in the main substantive findings across all robustness checks is encouraging.

6.4 Logit Model

We defined a binary outcome variable, `obese`, that takes the value 1 for $BMI \geq 30$ and 0 otherwise. This was analysed using a logit model as described in §4. The covariates used are the same as those for the model of log BMI, above. Again, the sexes are considered separately. The first step in the analysis is to produce a table of model comparison statistics, and to estimate and visually inspect the full APC model, to determine which APC sub-model is appropriate for the data.

For the women, the model comparison statistics are presented in table 7. Direct tests comparing plausible candidate models are seen in table 8. Both sets of tests favour the Ad model. Visual inspection of the APC model, not presented here, also supported the Ad model. The estimated Ad model is seen in figure 7 with the covariate coefficients in table 11.

For the men, the model comparison statistics are presented in table 9. In this case models A, P, C, and t are clearly rejected and thus omitted. Direct tests are shown in table 10. From these, and from the estimated full APC model (not shown), it is clear that the Cd model is favoured. This contrasts with the study of log BMI, where the AC model was chosen over the Cd model. The estimated Cd model effects are seen in figure 8 with the covariate coefficients in table 11.

The substantive patterns seen when using obesity as the outcome variable are broadly similar to those seen when using log BMI as the outcome variable. For women, the plane includes two significant positive slopes. The age deviations are similar, but in the logit model the initial acceleration is more rapid. The probability of being obese jumps to a high level at age 30 and does not deviate from linearity linearity until deceleration begins at age 50. In the normal model, there is gradual acceleration to age 50.. This implies that between the ages of 30 and 50, an increasing number of women gain weight but few pass the obesity threshold.

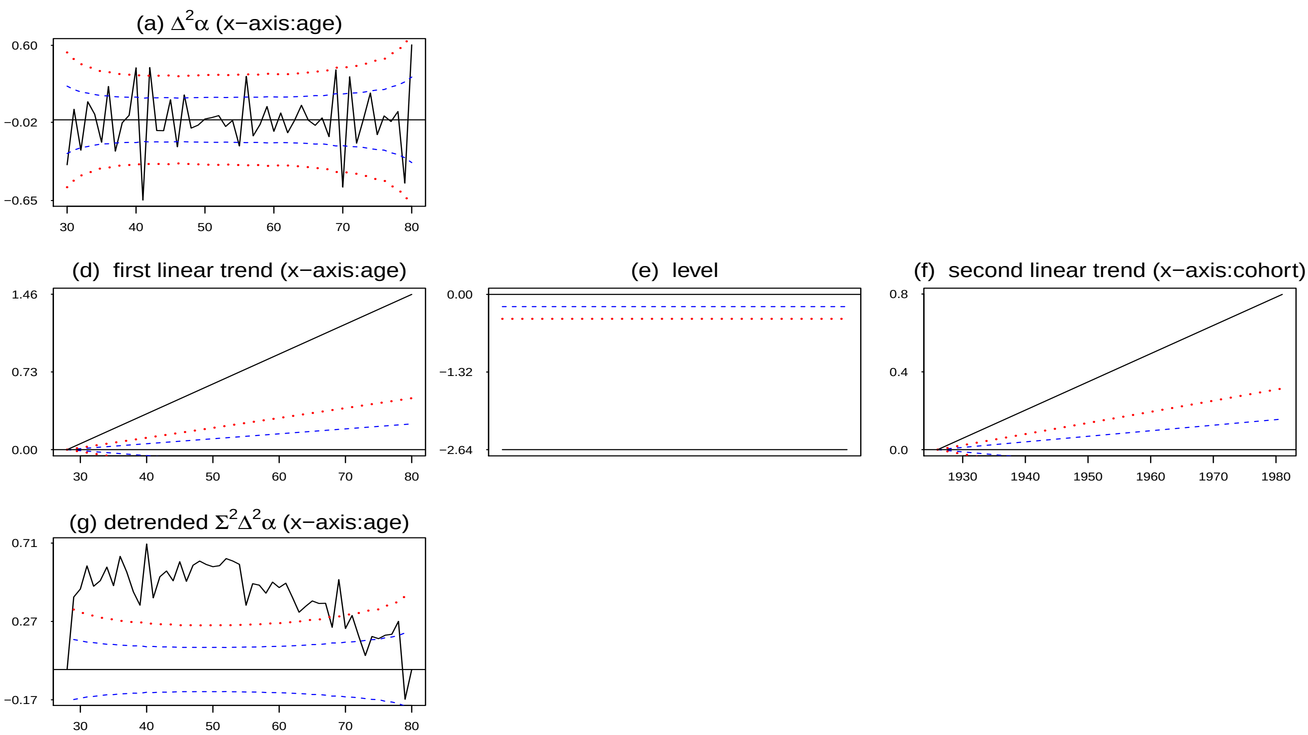
Table 7: **Model comparisons, obesity indicator, women**

	Against TS			Against APC				
	LR	df	p	LR	df	p	AIC	ℓ
TS							48708.92	-23627.46
APC	598.10	592	0.42				48123.02	-23926.51
AP	632.60	646	0.64	34.49	54	0.98	48049.51	-23943.76
AC	609.29	604	0.43	11.19	12	0.51	48110.21	-23932.10
PC	671.85	643	0.21	73.75	51	0.02	48094.76	-23963.38
Ad	643.83	658	0.65	45.72	66	0.97	48036.74	-23949.37
Pd	752.79	697	0.07	154.69	105	0.00	48067.71	-24003.85
Cd	683.41	655	0.21	85.31	63	0.03	48082.32	-23969.16
A	669.59	659	0.38	71.48	67	0.33	48060.50	-23962.25
P	809.36	698	0.00	211.26	106	0.00	48122.28	-24032.14
C	745.72	656	0.01	147.62	64	0.00	48142.63	-24000.32
t	763.97	709	0.07	165.86	117	0.00	48054.88	-24009.44

Table 8: **Model comparisons, obesity indicator, women**

Models compared	Ad vs AC	Ad vs AP	A vs Ad	t vs Ad	t vs A
p	0.98	0.51	0.00	0.00	0.00

Figure 7: **Time effects, Ad model of obesity indicator, women**



solid line = estimate; blue (red) dotted line = 1 (2) standard deviation

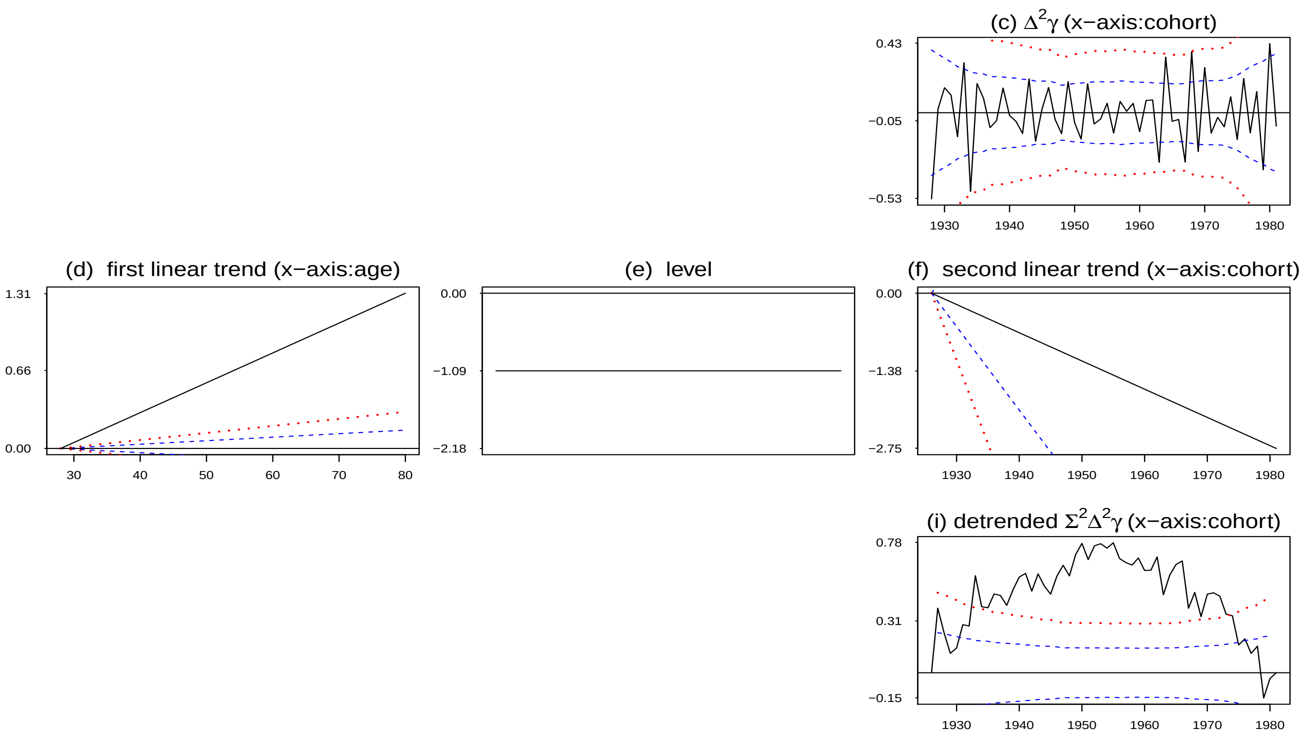
Table 9: **Model comparisons, obesity indicator, men**

	Against TS			Against APC				
	LR	df	p	LR	df	p	AIC	ℓ
TS							44563.50	-21554.75
APC	556.26	592	0.85				43935.76	-21832.88
AP	630.12	646	0.66	73.86	54	0.04	43901.62	-21869.81
AC	577.07	604	0.78	20.81	12	0.05	43932.57	-21843.29
PC	615.23	643	0.78	58.97	51	0.21	43892.74	-21862.37
Ad	651.07	658	0.57	94.81	66	0.01	43898.57	-21880.28
Pd	861.50	697	0.00	305.24	105	0.00	44031.00	-21985.50
Cd	635.40	655	0.70	79.14	63	0.08	43888.90	-21872.45

Table 10: **Model comparisons, obesity indicator, men**

Models compared	Cd vs AC	Cd vs PC	Ad vs AC
p	0.224	0.064	0.037

Figure 8: **Time effects, Cd model of obesity indicator, men**



solid line = estimate; blue (red) dotted line = 1 (2) standard deviation

One candidate explanation is that the psychological effect of being classified "obese" causes women to avoid moving into the category, but this does not affect weight loss.

For the men, there is one significant positive trend and one more ambiguous trend, which we also saw with log BMI. In the APC model for obesity and in the AC model for log BMI, there is something like curvature in each of age and cohort. Upon reduction to the Cd model, it appears that the two are combined to result in a larger cohort curve with an earlier peak. The way to think about this is as follows: the group of men aged approximately 40-60 in 2001-2014 have higher mean BMI than those of other ages. Because of the limited period range of this dataset, however, the observation of 40-60 year olds overlaps with the observation of cohorts born in 1940-1980; we do not observe middle-aged men from other cohorts, and we do not observe these cohorts at anything other than middle age. It is therefore impossible, even with good APC techniques, to separate between the cohort and age influences for this group. A longer period range is needed.

Table 11: **Covariate effects, obesity models**

	Women, Ad			Men, Cd		
	$\hat{\zeta}$	*se*	*p*	$\hat{\zeta}$	*se*	*p*
Ethnicity indicators (excl. white)						
Black	0.658	0.077	0.000	-0.034	0.095	0.721
Asian	-0.593	0.105	0.000	-0.588	0.089	0.000
Mixed ethnicity	-0.190	0.149	0.203	-0.170	0.172	0.326
Other ethnicity	-0.628	0.178	0.000	-0.343	0.188	0.068
Behaviour indicators (excl. never smoked, occasionally drink alcohol)						
Former smoker	0.201	0.027	0.000	0.310	0.027	0.000
Current smoker	-0.216	0.030	0.000	-0.356	0.033	0.000
Never drink alcohol	0.507	0.097	0.000	0.347	0.143	0.015
Rarely drink alcohol	0.428	0.025	0.000	0.201	0.029	0.000
Frequently drink alcohol	-0.332	0.035	0.000	-0.158	0.029	0.000
Education level indicators (excl. GCSE)						
Below GCSE	0.194	0.031	0.000	0.160	0.035	0.000
Some higher education	-0.109	0.033	0.001	-0.008	0.034	0.811
University degree	-0.452	0.040	0.000	-0.280	0.040	0.000
3 level NSSEC indicators (excl. routine/manual)						
Intermediate occupations	-0.239	0.029	0.000	-0.036	0.033	0.267
Managerial/Professional	-0.084	0.032	0.009	-0.117	0.031	0.000
Other occupations	-0.078	0.086	0.363	0.015	0.162	0.924

7 Conclusion

Our analysis of data from the Health Survey for England demonstrated clear time effects and covariate effects that were robust to a range of specifications. The already well-documented increasing aggregate trend is picked up in the linear plane for both men and women. The deviations around the plane differ between the sexes. For women, the only significant deviation from linearity is curvature in the age dimension, with an increase in log BMI up to middle age and a decline thereafter. We suggest metabolic changes, child-bearing, and child-rearing as potential reasons for this. For men, there is significant curvature in the cohort dimension, and, for log BMI only, in the age dimension. For both genders the impact of the covariates is largely consistent with existing literature, although more covariates are significant in the models for women. The only surprising result relates to the correlations with alcohol consumption: those who consume alcohol frequently have lower BMI than casual drinkers, while rare drinkers have higher BMI on average.

To estimate these effects, we employed the canonical parametrization of Kuang et al. (2008) for age, period, and cohort identification. We used a generalized linear modelling framework to introduce this reparametrization at the individual level via normal and logit models. To assess the goodness-of-fit of the classical APC model and its assumptions, in particular that of no interaction effects between time effects, we introduced the idea of testing against a "time-saturated" model. We developed an algorithm for estimation of the time-saturated model that exploits its mutually exclusive dummy variable structure. Additionally, we showed in §5 that minimally-constraining ad hoc identification yields the same covariate coefficient estimates as the reparametrized model. This strategy is suitable for inference when only the covariates are of interest and the time effects are considered nuisance parameters. Overall, this is an important methodological contribution to the study of age, period, and cohort effects. We extended the range of options available for analysis of individual-level data to include a method that does not impose constraints on the time effects, and have highlighted that the choice of method depends on the object of interest.

That said, there is plenty of work yet to do. The existing framework must be expanded to allow for mixture models, interaction terms between time effects and covariates, and heteroskedastic errors. This would enable us to address some of the misspecification concerns with the log BMI analysis. While standard methods for these situations exist, it is not clear how they would be affected by the collinearity in time effects and so care is needed in adding them to APC models. A clear limitation is that the implications of missing age-cohort cells within the generalized trapezoid have not been thought through; in the present application, this meant we had to drop all ages below 28 due to perfect separation in one age-cohort cell at age 27. It would also be of interest to test parametric models for the deviations from linearity, for example a quadratic model for the women's age deviations.

Acknowledgements

Funding was received from ESRC grant ES/J500112/1 (Fannon), ERC grants 681546, FAMSIZEMATTERS (Monden), and 694262, DisCont (Nielsen).

A Details of Reparametrization

The classical APC model we wish to use to represent the data is given in equation (6) of §2.2, that is $\mu_{ik} = \alpha_i + \beta_j + \gamma_k + \delta$.

The reparametrization of this model is produced by introducing telescopic sums of the following form into the model equation:

$$\alpha_i = \alpha_1 + \sum_{t=2}^{i} \Delta\alpha_t \qquad \Delta\alpha_t = \Delta\alpha_2 + \sum_{s=3}^{t} \Delta^2\alpha_s, \tag{20}$$

where $\Delta\alpha_t = \alpha_t - \alpha_{t-1}$ and $\Delta^2\alpha_s = \Delta\alpha_s - \Delta\alpha_{s-1}$. This results in an equation of the general form

$$\mu_{ik} = \upsilon_o + (i-U)\upsilon_a + (k-U)\upsilon_c + A_i + B_j + C_k \tag{21}$$

$$A_i = 1_{(i<U)} \sum_{t=i+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2\alpha_s + 1_{(i>U+1)} \sum_{t=U+2}^{i} \sum_{s=U+2}^{t} \Delta^2\alpha_s$$

$$B_j = 1_{(L\,odd\,\&\,j=2U-2)} \Delta^2\beta_{2U} + 1_{(j>2U)} \sum_{t=2U+1}^{j} \sum_{s=2U+1}^{t} \Delta^2\beta_s$$

$$C_k = 1_{(k<U)} \sum_{t=k+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2\alpha_s + 1_{(k>U+1)} \sum_{t=U+2}^{k} \sum_{s=U+2}^{t} \Delta^2\alpha_s$$

Here L is the offset mentioned in (2), §2.1, U is a central reference point calculated as the integer value of $(L+3)/2$, υ_o is an intercept that combines structural level effects, and υ_a and υ_c are slope parameters in age and cohort directions respectively that combine single-differences of all three time effects. Note also the symmetry in the definitions of A_i and C_k.

Upon inspection it can be seen that the above may be written as $X_h'\xi$ for ξ given in equation (9), §2.3. X_i will contain an intercept, the two slopes, and cumulations of each double-difference. Further details of this reparametrization including insight regarding the choice of L and U can be found in Nielsen (2014).

B Properties of ad hoc identification schemes

We show that the model $D_\lambda\psi$, estimated under the ad hoc identification scheme $\alpha_1 = \alpha_2 = \beta_J = \gamma_K = 0$ in equation (19) of §5 can be expressed in the form $\xi = Q\phi$ in (18) by finding Q and ϕ. The proofs appeal to the analysis of Nielsen & Nielsen (2014), henceforth NN14.

Some notation from linear algebra is needed, specifically the orthogonal complement. A matrix m has full column rank if $m'm$ is invertible. In this case the orthogonal complement $m_\perp$ is a matrix so $m_\perp' m = 0$ and $(m, m_\perp)$ is invertible.

Write the constraint in equation (19) as $L'\theta_\lambda = 0$, where L is the $(q \times 4)$-matrix that selects the coordinates in (19) and the subscript λ indicates that θ_λ is a constrained version of θ. Note that $L'L = I_4$.

The design matrix for the ad hoc identified model is D_λ. This is found by dropping the columns DL from D. That is $D_\lambda = DL_\perp$, where $L_\perp$ is the selection matrix for the remaining columns with the properties that $L_\perp' L_\perp = I_p$ and $L_\perp' L = 0$. Thus, the ad hoc identification gives $\mu = DL_\perp\psi$ where $\psi = L_\perp'\theta$ are the remaining elements of θ. Since $\mu = DL_\perp\psi$ as well as $D = XA'$ we get $\mu = XA'L_\perp\psi$. At the same time we have $\mu = X\xi$ implying $\xi = A'L_\perp\psi$. Here $Q = A'L_\perp$ and if it is invertible we can choose $\phi = \psi = Q^{-1}\xi$.

To verify that $(A'L_\perp)$ is invertible, recall from §2.3 and NN14 that the canonical parametrization can be expressed in terms of a $q \times p$ matrix A so $D = XA'$ and $A'\theta = \xi$. We note that the constraint satisfies the property that (A, L) is invertible. This is proved by writing out A which is implicitly defined in §2.3 and then finding an orthogonal complement $A_\perp$. An explicit expression for $A_\perp$ is given in NN14, equation 58. Next, it is checked that $A'_\perp L$ is invertible. This implies that (A, L) is invertible, see NN14, Lemma A.1. Lemma A.1 also shows that $A'L_\perp$ is invertible.

The difference between the two approaches is that with the canonical parametrization we focus on the estimable p-vector ξ whereas the ad hoc identification considers the q-vector $\theta_\lambda = L_\perp \psi$, but only the p-vector $L_\perp \theta_\lambda = \psi$ can be determined from data.

C Data and Design

The alcohol categories are: rare = drinks less than once a week, casual = drinks one to four times per week, frequent = drinks five or more times per week. Note this does not account for quantity of drinks per drinking event.

Table 12: **Descriptive statistics, women** ($N = 43077$)

Continuous variables	*Minimum*	*Mean*	*Median*	*Maximum*	
Age	28	51	50	80	
Period	2001	2007	2006	2014	
Cohort	1926	1956	1957	1981	
BMI	13.2	27.37	26.35	58.94	
Height (cm)	123.6	161.67	161.6	202	
Weight (kg)	28.4	71.49	69	164	
Categorical variables					
Ethnicity	*Black*	*White*	*Asian*	*Mixed*	*Other*
	762	41071	702	288	254
NSSEC (3 level)	*Routine*	*Intermediate*	*Professional*	*Other*	
	4585	11721	14302	748	
Education level	*Below GCSE*	*GCSE*	*Some higher*	*Degree*	
	13138	11847	9644	8448	
Alcohol	*Never*	*Rare*	*Casual*	*Frequent*	
	497	15703	19534	7343	
Smoking	*Never*	*Former*	*Current*		
	23037	10724	9316		

Table 13: **Descriptive statistics, men** ($N = 38316$)

Continuous variables	*Minimum*	*Mean*	*Median*	*Maximum*	
Age	28	52	51	80	
Period	2001	2007	2006	2014	
Cohort	1926	1955	1956	1981	
BMI	13.63	27.93	27.44	59.45	
Height (cm)	138.2	174.9	174.8	203.1	
Weight (kg)	34.2	85.52	84	203.4	
Categorical variables					
Ethnicity	*Black*	*White*	*Asian*	*Mixed*	*Other*
	600	36363	972	201	180
NSSEC (3 level)	*Routine*	*Intermediate*	*Professional*	*Other*	
	6972	7390	16363	201	
Education level	*Below GCSE*	*GCSE*	*Some higher*	*Degree*	
	10927	7865	10456	9068	
Alcohol	*Never*	*Rare*	*Casual*	*Frequent*	
	228	8330	19489	10269	
Smoking	*Never*	*Former*	*Current*		
	16692	13084	8540		

References

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

An, R. & Xiang, X. (2016). Age-period-cohort analyses of obesity prevalence in US adults. *Public Health*, *141*, 163–169.

Baum II, C. L. & Ruhm, C. J. (2009). Age, socioeconomic status and obesity growth. *Journal of health economics*, *28*, 635–648.

Carstensen, B. (2007). Age-period-cohort models for the Lexis diagram. *26*, 3018–3045.

Clayton, D. & Schifflers, E. (1987). Models for tempral variation in cancer rates. II Age-period-cohort models. *6*, 469–481.

Davidson, R. & MacKinnon, J. (1993). *Estimation and Inference in Econometrics.* Oxford: Oxford University Press.

Department of Health (2011). Healthy lives, healthy people: A call to action on obesity in England. Technical Report 16166, HM Government.

Ejrnæs, M. & Hochguertel, S. (2013). Is business failure due to lack of effort? Empirical evidence from a large administrative sample. *Economic Journal*, *123*, 791–830.

Fahrmeir, L. & Kaufmann, H. (1986). Asymptotic inference in discrete response models. *Statistical Papers*, *27*, 179–205.

Fu, W. (2016). Constrained estimators and consistency of a regression model on a Lexis diagram. *111*, 180–199.

Glenn, N. D. (2005). *Cohort Analysis* (2nd ed.)., volume 5 of *Quantitative Applications in the Social Sciences.* SAGE Publications, Inc.

Harnau, J. & Nielsen, B. (2017). Over-dispersed age-period-cohort models. *Journal of the American Statistical Association*, *to appear.*

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, *39*, 311–324.

Hruby, A., Manson, J. E., Qi, L., Malik, V. S., Rimm, E. B., Sun, Q., Willet, W. C., & Hu, F. B. (2016). Determinants and consequences of obesity. *AJPH Special Section: Nurses' Health Study Contributions*, *106*, 1656–1662.

Kuang, D., Nielsen, B., & Nielsen, J. P. (2008). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika*, *95*, 979–986.

Martínez Miranda, M. D., Nielsen, B., & Nielsen, J. P. (2015). Inference and forecasting in the age-period-cohort model with unknown exposure with an application to mesothelioma mortality. *Journal of the Royal Statistical Society, Series A*, *178*, 29–55.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, K. (1973). Some methodological issues in cohort analysis of archival data. *American Sociological Review*, *38*, 242–258.

McPherson, K., Marsh, T., & Brown, M. (2007). Tackling obesities: Future choices - modelling future trends in obesity and the impact on health. Technical report, Government Office for Science.

Moody, A. (2016). Health Survey for England 2015 adult overweight and obesity. Technical report, Health and Social Care Information Centre.

NatCen Social Research, University College London. Department of Epidemiology and Public Health (2016). Health Survey for England 2014 [data collection]. Data retrieved from https://discover.ukdataservice.ac.uk/series/?sn=2000021.

Nielsen, B. (2014). Deviance analysis of age-period-cohort models. Discussion paper, Nuffield College.

Nielsen, B. (2015). apc: An R package for age-period-cohort analysis. *The R Journal*, *7*, 52–64.

Nielsen, B. & Nielsen, J. P. (2014). Identification and forecasting in mortality models. *The Scientific World Journal*, *2014*, Article ID 347043, 24 pages.

Ogden, C. L., Carroll, M. D., Fryar, C. D., & Flegal, K. M. (2015). Prevalence of obesity among adults and youth: United States, 2011-2014. Technical Report 219, National Center for Health Statistics, Hyattsville, MD.

Poirier, D. J. (1998). Revising beliefs in nonidentified model. *Econometric Theory*, *14*, 483–509.

Ramsey, J. B. (1969). Test for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, Series B (Methodological)*, *31*, 350–371.

Reither, E. N., Hauser, R. M., & Yang, Y. (2009). Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. *Social Science & Medicine*, *69*, 1439–1448.

Smith, T. R. & Wakefield, J. (2016). A review and comparison of age-period-cohort models for cancer incidence. *Statistical Science*, *31*, 591–610.

Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika*, *63*, 27–32.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817–838.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: Massachusetts Institute of Technology.

Yang, Y. (2008). Social inequalities in happiness in the United States, 1972 to 2004: An age-period-cohort analysis. *American Sociological Review*, *73*, 204–226.

Yang, Y. & Land, K. C. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological Methodology*, *36*, 75–97.