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

Zoë Fannon, Christiaan Monden, Bent Nielsen

2 March 2020

Abstract

We develop an age-period-cohort model for repeated cross-sectional data with individual covariates. This is done for both continuous and binary dependent variables. The age, period, and cohort effects in the model are represented by a parametrization with freely varying parameters that allows separation of the identified non-linear effects and the unidentifiable linear effects. We develop a test of the parametrization against a more general "time-saturated" model. The method is applied to an analysis of the obesity epidemic in England using survey data. We find that the main deviations from linearity present in English obesity data are age-related among women and cohort-related among men.

1 Introduction

We use repeated cross-sectional data to examine the socio-demographic determinants of obesity rates in England, using both continuous and binary measures of obesity. The explanatory variables of interest are an individual's age, birth cohort, and period of observation, as well as other individual characteristics such as sex, race, education, and socio-economic status. For the analysis, we develop new models for repeated cross-sectional data with fixed effects for age, period and cohort. The well-known age-period-cohort identification problem is addressed through the canonical parametrization, which is both freely varying and invariant to the identification problem. We develop a test of this parametrization against a more general model. This resembles a deviance test, but can be used for both continuous and binary outcomes. Applying the methods to English obesity data, we find that the data can be parsimoniously described using an age-cohort or cohort-drift model for men and an age-drift model for women.

Adult obesity is a major public health concern. UK obesity rates almost tripled 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 obese if their body mass index (BMI) exceeds 30. Obesity is linked to immediate and long term health risks, such as type II diabetes. The direct healthcare costs of obesity in 2006-07 were estimated to be £5.1 billion, 6% of the National Health Service budget (Scarborough et al., 2011). Reducing obesity has thus been a policy goal for many years, with specific government directives issued in 2007, 2011, and 2016. Policies to reduce obesity will have the greatest effect if they target the most at-risk subpopulations. The analysis in this paper helps to identify such subpopulations by showing how obesity evolves with age and cohort, independent of population level period effects.

We use data from the 2001 through 2014 waves of the Health Survey for England. Our dependent variable is either the continuous BMI measure or a binary obesity indicator. The explanatory variables include age, period of observation, and cohort, constructed through *cohort* = *period* - *age*, and socio-demographic variables including education and smoking behaviour.

The proposed models are generalized linear models (GLMs) for repeated cross-sectional data with age, period, and cohort (APC) effects and individual covariates. The models have three new features. First, several types of dependent variables are accommodated by GLMs, including continuous and binary variables. Second, the well-known APC identification problem, which is discussed at some length below, is addressed using the parametrization of Kuang et al. (2008), henceforth KNN. This parametrization had previously only been used with aggregate data, and avoids some of the issues associated with the identification problem. Third, the combination of the KNN parametrization with GLMs gives simple likelihood functions. It is then easy to formulate hypotheses as well as more general models and to pose associated statistical tests. In this spirit, we propose to test the APC models against a more general 'time saturated' model, with a fixed effect for each age-cohort combination.

The APC identification problem is well-known and discussed at length for models for aggregate data (Holford, 1983; Clayton and Schifflers, 1987; Glenn, 2005; Carstensen, 2007; O'Brien, 2011; Luo, 2013; Fannon and Nielsen, 2019). In APC models, the predictor is a linear combination of three APC effects. Knowing the predictor only allows partial identification of the three APC effects. The issue is often addressed by constraining the APC effects, but this gives difficulties in separating what is learned about the APC effects from the data and what is learned from the constraints, which in turn generates challenges with respect to interpretation, inference and recursive analysis; see Nielsen and Nielsen (2014) for a formal analysis. A less controversial idea is the estimable functions approach, see Holford (1983), Clayton and Schifflers (1987). There, constraints are used in the estimation, but only estimates of functions of the APC effects that are invariant to the identification problem are interpreted.

Here, we approach the identification issue with a focus on non-linear APC effects, which are identified, rather than the unidentified linear effects. We adopt the KNN parametrization of the predictor, see also Martínez Miranda et al. (2015) for a Poisson count model and Smith and Wakefield (2016) for a Bayesian implementation. The parametrization has two features. First, the predictor is parametrized as a linear function of a freely varying parameter, which is therefore canonical in a generalized linear model context. This eases imposition and interpretation of hypotheses as well as counting the associated degrees of freedom. Second, as a function of the APC effects, the KNN parametrization is invariant to the identification problem. It is also invariant to incorporating additional data waves and other variations of the APC time horizons. Specifically, the predictor consists of: deviations from linearity attributed to each APC effect, and a linear plane combining the inseparable linear APC effects.

The asymptotic analysis of our model can be understood by thinking of the data as a two-way age-cohort array with individual information accumulating in each cell. 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-cohort array fixed and exploit the individual level information for inference. With the canonical parametrization, the statistical analysis is then fairly simple. This asymptotic approach resembles earlier work on two-way arrays of aggregate data where each cell entry is large, including a Poisson model for counts of cancer deaths (Martínez Miranda et al., 2015) and an over-dispersed Poisson model for insurance claims (Harnau and Nielsen, 2018). Asymptotics for aggregate data with a large period dimension have been considered by Fu (2016). That approach would be inappropriate for our data given its small period range.

We develop a test of the reparametrized age-period-cohort model against a more general model, where each cell in the age-cohort array has its own parameter. We call this a time saturated (TS) model. The TS model nests the APC model, while permitting interactions between any two of the APC effects. Such interactions could capture, for instance, the effect on obesity of a time- and age-limited government programme promoting healthy eating in schools. The absence of such effects in the APC model is testable. The proposed tests of the APC models against TS models resemble deviance tests, but apply for both continuous and binary outcomes. Inference is standard, but we address some computational challenges.

Broadly speaking, the literature has used APC models for cross-sectional data in two contexts. First, in some studies researchers are primarily interested in the effect of individual level covariates, but include APC effects as controls. In that case it does not matter how the APC effects are identified, as we show in an appendix. Ejrnæs and Hochguertel (2013) provide an application to unemployment insurance. Second, in this paper and in other studies the APC effects are of primary interest. The hierarchical APC model, developed in Yang and Land (2006) and Yang (2008), is often used for that purpose. In those models the age effect is restricted to be quadratic, while the period and cohort effects have zero mean random effects. Since the period-cohort effects have zero mean, those models are essentially Age (A) models and the APC identification problem does not apply. The hypotheses of a quadratic age effect and absence of period-cohort effects are testable in the model presented in this paper.

Some previous studies of obesity use APC models. Reither et al. (2009) use the hierarchical APC model reviewed above. Their estimates of the zero mean period-cohort random effects deviate systematically from zero. This indicates model misspecification, which complicates interpretation of parameters and standard errors (Snijders and Bosker, 2012, §11). It also introduces an APC identification problem (Bell and Jones, 2014b). An and Xiang (2016) use a fixed-effects version of the HAPC model. This suffers from an APC identification problem, as the cohort effects are constrained in bands and this affects the estimated APC slopes.

When applying our 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. The age-drift model for women includes two parts. The first is an increasing linear plane, or "drift", that combines the inseparable linear APC effects. The second is a concave deviation from linearity in the age dimension which has a kink at age 50. This concavity is consistent with previous research (Lean et al., 2013; Wang et al., 2011; Howel, 2011), in which APC effects are considered more informally than here. For men, the age-cohort model includes a combined linear plane; a concave deviation from linearity in age, similar to that for women; and a concave deviation from linearity in cohort, which is particularly pronounced from the 1960s cohort onwards. The deviation from linearity in cohort was not detected by previous studies, which used different age and period ranges and often imposed constraints on the cohort effects. The effects of covariates are broadly consistent with existing literature.

The paper is outlined as follows: §2 introduces the data and its APC structure as well as the APC identification problem and the KNN parametrization. §3 and §4 contain the main theoretical contributions of this paper. In §3, we discuss the conditions for standard inference in normal and logit models (continuous and binary outcomes, respectively) using the KNN parametrization of APC effects and repeated cross-sectional data. In §4 a new test is proposed which compares the APC model to a more general model, and an algorithm for this test is developed. In §A.3 the situation in which the APC effects are nuisance parameters rather than direct objects of interest is considered. §5 contains the application of the methods to analyse obesity dynamics in England, while §6 concludes.

2 Preliminaries: data and APC parametrization

In this section we introduce the obesity data and the APC indices. We proceed to give an overview of APC models, the APC identification problem, the KNN parametrization, and APC sub-models as used for aggregated data.

2.1 Obesity data

The data is a repeated cross-section of representative samples of the English population taken from the Health Survey for England (HSE). We use waves from 2001-2014. The waves prior to 2001 do not include the National Statistics Socio-Economic Classification (NSSEC), one of

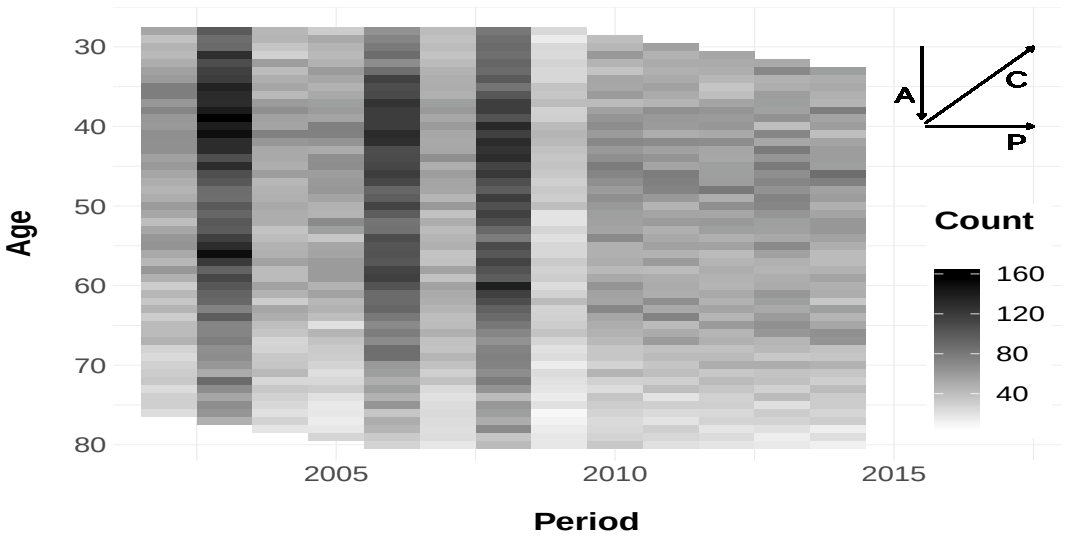


Figure 1: Within-cell observation counts women

our explanatory variables, while age is only recorded in five-year bands after 2014. For the analysis, we used the R-package apc, version 1.3.4 (R Core Team, 2019; Nielsen, 2015).

We separately analyse data for 43,077 women and 38,316 men. The sample size is denoted H and individuals are indexed $h = 1, \ldots, H$, reserving the letter i for the age index.

For each individual we have information on weight and height, measured by the interviewer. The body mass index (BMI) is defined as the weight in kg divided by the square of the height in metres. A small number of observations with BMI outside the range 12 to 60 were presumed to be subject to measurement error and excluded. In addition to BMI, age, and period of observation, we observe the following covariates: ethnicity, level of education, NSSEC, smoking history, and alcohol consumption. Tables 3 and 4 in Appendix B report descriptive statistics.

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 constructed through $i_h + k_h = j_h + 1$.

In this dataset age and period vary in a rectangular age-period array, where age is between 28 and 80 and period is between 2001 and 2014 (see Figure 1). We therefore have $I = 53$ age groups and $J = 14$ period groups. Cohort then varies between 1921 and 1986. However, we exclude the first and last five cohorts as they are sparsely observed. This leaves $K = 56$ cohort groups. The final data, as an age-period array, is shown in Figure 1. The shading in that figure reflects the variation in survey size across waves.

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

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

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 APC parametrization

The APC component in our model is treated the same way as it was in previous work with aggregate data (Martínez Miranda et al., 2015). For an individual h with age i and cohort k,

observed in period $j = i + k - 1$, the linear predictor will include the APC component

$$\mu_{ik} = \alpha_i + \beta_j + \gamma_k + \delta. \tag{2}$$

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

$$\theta = (\alpha_1, \ldots, \alpha_I, \beta_{L+1}, \ldots, \beta_{L+J}, \gamma_1, \ldots, \gamma_K, \delta)', \tag{3}$$

we can write (2) as $\mu_{ik} = D'_{ik}\theta$. Here, D_{ik} is a q-dimensional design vector taking the value one for the entries corresponding to α_i, β_j, γ_k, and δ and zero otherwise. Thus, we have a linear function from θ to μ, the collection of μ_{ik} for $i, k \in \mathcal{I}$. As θ is of dimension q the variation of μ_{ik} is at most of dimension q.

The APC identification problem is that the vectors D_{ik} have four linear dependencies. The problem can be characterized by a group of transformations (Carstensen, 2007), as follows. For any constants $a, b, c, d \in \mathbb{R}$ the predictor μ_{ik} in (2) satisfies

$$\begin{aligned}\mu_{ik} = \{\alpha_i + a + (i-1)d\} + \{\beta_j + b - (j-1)d\} \\ + \{\gamma_k + c + (k-1)d\} + \{\delta - a - b - c\}.\end{aligned} \tag{4}$$

KNN show that there are not more than these four dependencies. Equation (4) therefore summarizes the $p = q - 4$ dimensional variation of the linear function from θ to μ. In combination, the constants a, b, c, d cancel on the right hand side of (4), so that μ_{ik} does not depend on a, b, c, d. We say that μ_{ik} is invariant with respect to the transformations in (4). At the same time the individual APC effects, α_i, β_j, γ_k, are only identified up to linear trends. For instance, the age effect α_i is observationally equivalent to $\alpha_i + a + (i-1)d$ for any a, d.

As an example, let $\mu_{ik} = -2 + 0.1i - 0.001i^2 + 0.04k$ as in Bell and Jones (2014b). This could arise from θ defined by $\alpha_i = -2 + 0.1i - 0.001i^2$, $\gamma_k = 0.04k$, $\beta_j = \delta = 0$; or equally from θ^* defined by $\alpha^*_i = -1.96 + 0.06i - 0.001i^2$, $\beta^*_j = 0.04j$, $\gamma^*_k = \delta^* = 0$. Note, that θ^* arises from θ when choosing $a = 0$, $b = -c = -d = 0.04$ in (4). Thus, the vector θ is not identified.

A common response to the APC problem is the constraints approach. This imposes four constraints on θ, such as $\alpha_1 = \beta_{L+1} = \beta_{L+2} = \gamma_1 = 0$, giving a constrained vector θ_c. The vector θ_c has p-dimensional variation and is identifiable. However, the approach causes confusion if the constrained vector θ_c is mistaken for θ, see Nielsen and Nielsen (2014). The constraints approach does not exploit the invariance of μ with respect to the transformations in (4) and lacks clarity about which parts of θ are identifiable and which are not.

Our approach is to reparametrize the predictor μ in terms of a parameter ξ that has two desirable properties. First, ξ is invariant to the transformations in (4). Second, we want μ to be exactly identifiable from ξ, so that different values of ξ give different values of μ. Thus, ξ must be a p-vector giving a dimension reduction from q to p as for the constraints approach. By requiring invariance, we focus on the identifiable part of θ and simply ignore the part of θ that cannot be identified. While not exactly solving the APC identification problem, this approach makes the problem ignorable.

KNN construct a parameter ξ of the desired form from the following two observations. First, the predictors μ_{ik} are invariant. Second, double differences of the APC effects, such as $\Delta^2\alpha_i = (\alpha_i - \alpha_{i-1}) - (\alpha_{i-1} - \alpha_{i-2})$ are invariant, since $\Delta^2\alpha_i = \Delta^2\{\alpha_i + a + (i-1)d\}$, see also Clayton and Schifflers (1987). Returning to the Bell and Jones (2014a) example, We see that the quadratic term is identified with $\Delta^2\alpha_i = \Delta^2\alpha^*_i = -0.002$.

The parameter ξ proposed by KNN, which has the desired properties, is the p-vector

$$\xi = (v_o, v_a, v_c, \Delta^2\alpha_3, \ldots, \Delta^2\alpha_I, \Delta^2\beta_{L+3}, \ldots, \Delta^2\beta_{L+J}, \Delta^2\gamma_3, \ldots, \Delta^2\gamma_K)'. \tag{5}$$

Here, $\upsilon_o, \upsilon_a, \upsilon_c$ are the level and two slopes of a linear plane. These can be chosen in various ways. For computational implementations, it is useful to anchor the plane symmetrically in age and cohort. Thus, let U be the integer value of $(L+3)/2$ with L defined in (1) and let

$$\upsilon_o = \mu_{UU}, \qquad \upsilon_a = \mu_{U+1,U} - \mu_{UU}, \qquad \upsilon_c = \mu_{U,U+1} - \mu_{UU}, \tag{6}$$

The linear slopes of the APC effects remain entangled. Indeed, inserting the APC expression for the predictor in (2) shows that

$$\upsilon_o = \alpha_U + \beta_{2U-1} + \gamma_U + \delta, \qquad \upsilon_a = \Delta\alpha_{U+1} + \Delta\beta_{2U}, \qquad \upsilon_c = \Delta\gamma_{U+1} + \Delta\beta_{2U}. \tag{7}$$

In addition, the double differences in ξ represent the non-linear parts of the APC effects. Corollary 2 in KNN shows that ξ is a maximal invariant function of θ with respect to (4) and therefore identifiable from the predictor μ. Moreover, a design vector X_{ik} exists so that $\mu_{ik} = X'_{ik}\xi$. Nielsen (2015) gives the precise form of X_{ik}, see also Appendix A.1.

The KNN parametrization confers a number of advantages. The vector ξ is identifiable from μ through linearly independent design vectors X_{ik}. It is an invariant function of θ with respect to (4) and its formulation is independent of the index set $\mathcal{I}$. When working with models of the generalized linear model type, such as the normal and logit models used in this paper, ξ is the canonical parameter.

In summary, the proposal is to replace the APC effect parametrization $\mu_{ik} = D'_{ik}\theta$ and all its problems, with the parametrization $\mu_{ik} = X'_{ik}\xi$, where the X_{ik} vectors are linearly independent and ξ is identifiable and invariant. Here, ξ represents what can be learned in an APC model, while the transformations (4) describe what cannot be learned. Thus, the parametrization allows us to focus on using standard statistical tools to build statistical APC models for different types of data and safely ignore the APC identification problem.

2.3 APC sub-models

For aggregate data it is common to consider sub-models of the APC model, where one or more components are dropped (Nielsen, 2015; Oh and Holford, 2015; Fannon and Nielsen, 2019). We review these sub-models. Five categories of restriction are of interest.

First, we can test for the absence of deviations from linearity in one of the APC effects. For instance, we test period non-linearities by imposing $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+J} = 0$. This gives an Age-Cohort (AC) model. In terms of the unidentified original parametrization θ, this is written as $\beta_{L+1} = \ldots \beta_{L+J} = 0$. The two formulations of the hypothesis are in fact equivalent (Nielsen and Nielsen, 2014). The latter formulation obscures the degrees of freedom and hides that the hypothesis does not constrain the linear period effects. Period-Cohort (PC) and Age-Period (AP) models are analogous to AC models.

Second, we can test for absence of non-linearities in two components, such as period and cohort, by imposing $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+J} = 0$ and $\Delta^2\gamma_3 = \cdots = \Delta^2\gamma_K = 0$, while leaving the linear plane unrestricted. This is the Age-drift (Ad) model (Clayton and Schifflers, 1987). Analogous models are Cohort-drift (Cd) and Period-drift (Pd).

Third, we can test a model with only an age effect, by imposing $\upsilon_c = 0$ on the Ad model. This is the Age (A) model. The linear slope of the age effect remains unidentifiable in view of (7). The constraint is that $\upsilon_c = \Delta\gamma_{U+1} + \Delta\beta_{2U} = 0$. In this case, $\upsilon_a = \Delta\alpha_{U+1} + \Delta\beta_{2U}$ identifies a combined age-period slope. Analogous models are Cohort (C) and Period (P).

Fourth, we will be interested in a linear plane model, where all double differences $\Delta^2\alpha_i$, $\Delta^2\beta_i$, $\Delta^2\gamma_i$ are set to zero. In this model, the APC effects are linear trends, $\alpha_i = \alpha_c + \alpha_\ell i$ and $\beta_j = \beta_c + \beta_\ell j$ and $\gamma_k = \gamma_c + \gamma_\ell k$. We call this a trend (t) model.

Finally, functional form restrictions could be of interest. For instance, quadratic age effect $\alpha_i = \alpha_c + \alpha_\ell i + \alpha_q i^2$ arises when $\Delta^2\alpha_3 = \cdots = \Delta^2\alpha_I = 2\alpha_q$ is imposed (Fannon and Nielsen, 2019). The hierarchical APC model of Yang and Land (2006) is an A model, where α_i is restricted to be quadratic, but the period and cohort effects have zero mean random effects.

3 The APC models for repeated cross-sectional data

We have described the data and the APC parametrization. We proceed to introduce generalized linear models with APC effects for individual level data. In particular, we consider a normal model for continuous outcomes and a logistic model for binary outcomes. We discuss estimation and inference for both cases.

3.1 The generalized linear model

We use generalized linear models for the dependent variable, Y_h. The linear predictor η_h is a function of the covariates and the APC structure through

$$\eta_h = Z_h'\zeta + \mu_{i_h k_h}, \qquad \mu_{i_h k_h} = X'_{i_h k_h}\xi. \tag{8}$$

Here, ζ is a d_z-vector of parameters and Z_h is a d_z-vector of covariates. These covariates can include regressors varying at the individual level or APC interaction effects that cannot be expressed in terms of the linear APC effects in X_{i_h,k_k}. The vectors (Z_h, X_{i_h,k_k}) are linearly independent. The term $\mu_{i_h k_h}$ describes the APC structure for an individual h with age i_h and cohort k_h observed in period $j_h = i_h + k_h - 1$, where $\mu_{ik} = \alpha_i + \beta_j + \gamma_k + \delta = X'_{ik}\xi$ as in (2), with ξ and X_{ij} given in (5) and (20). In matrix notation, we stack η_h, $\mu_{i_h k_h}$, Z_h, X_h over individuals to get η, μ, X, and Z. The linear predictor can then be written as

$$\eta = Z\zeta + \mu, \qquad \mu = X\xi. \tag{9}$$

The matrix X will have full column rank as long as $I, J, K \geq 2$. We will require that the combined design matrix (Z, X) also has full column rank.

Yang and Land (2006) and Yang (2008) have proposed related models for repeated cross sections. Their models are essentially of the same structure as here, but with two differences. First, the APC component is constrained to follow an A model, with the age effect constrained to be quadratic. This is a testable constraint in the present framework. Second, zero mean random effects are added to the period and cohort components.

When the APC effects are of interest we recommend the parametrization (5) for the reasons outlined in §2.2. However, if the primary interest lies in the covariates Z_h and APC effects are just included as controls it does not matter how the time effects are identified, see Appendix A.3. Ejrnæs and Hochguertel (2013) give an empirical example, where Z_h includes both individual level regressors and a non-linear interaction effect.

3.2 The normal model

For continuous dependent variables a normal model is employed of the form

$$Y_h = \eta_h + \varepsilon_h \qquad \text{for } h = 1, \ldots, H. \tag{10}$$

Conditional on the linear predictor, the errors ε_h are assumed independently $\mathsf{N}(0, \sigma^2)$ distributed. The model is estimated by ordinary least squares.

When it comes to inference, some of APC parameters may be unnecessary as described in §2.3. Achieving a more parsimonious APC structure is desirable for interpretation and for forecasting purposes. Likewise, some elements of covariate vector Z_h may be redundant.

Exact inference on the parameters ζ, ξ can be performed using t- and F-tests by appealing to the classical results for analysis of variance under normality. Standard asymptotic inference can be conducted under weaker assumptions. In particular, the likelihood ratio test statistic will be asymptotically χ^2. For this we must be clear about the repetitive structure. We treat the dimensions I, J, K of the age-cohort array as fixed, but assume the number of individual observations H is large. To justify asymptotic inference we assume

1. The triplets Y_h, X_h, Z_h are i.i.d. across individuals h.
2. The covariance matrix of (Z_h, X_h) is positive definite .
3. The errors ε_h satisfy $\mathsf{E}(\varepsilon_h|X_h, Z_h) = 0$ and $\mathsf{Var}(\varepsilon_h|X_h, Z_h) = \sigma^2$.

These assumptions imply that the relative frequencies of different age-cohort combinations are independent of the sample size H. In our data, the annual variation in sample size is due to financial constraints of the HSE and thus unrelated to the distribution of BMI. Similarly, it is plausible that selection into the survey is independent of the covariates Z_h. It is therefore quite likely that inferences can be extrapolated beyond the sample (Wooldridge, 2010, §19.4).

3.3 The logit model

For binary dependent variables, a logistic model is employed with

$$\log\{\mathsf{P}(Y_h = 1)/\mathsf{P}(Y_h = 0)\} = \eta_h \qquad \text{for } h = 1, \dots, H. \tag{11}$$

The corresponding logit log-likelihood is

$$\ell(\zeta, \xi) = \sum_{h=1}^{H} \eta_h Y_h - \sum_{h=1}^{H} \ln(1 + \exp \eta_h). \tag{12}$$

It 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.

The asymptotic theory of the logit estimator is outlined by Fahrmeir and Kaufmann (1986). Their Theorem 2 shows consistency and asymptotic normality under the assumptions 1 and 2 listed for the normal model in §3.2. The asymptotic variance-covariance matrix of the logit estimator is given by $J = -\ddot{\ell}$, where $\ddot{\ell}$ is the second derivative of the log-likelihood. Theorem 3 of Fahrmeir and Kaufmann (1986) shows that likelihood ratio test statistics on the covariate parameter ζ and the APC parameter ξ are asymptotically χ^2.

4 Testing the APC model against a general model

In practice, the linear APC model may be too parsimonious. APC effects could be multiplicative, as in the Lee and Carter (1992) model, or they could change after interventions. To address this, we provide a mis-specification test of the APC model against a time-saturated (TS) model, where the linear predictor μ_{ik} is unconstrained and thus catches the above alternatives. We use a likelihood ratio test of the APC model against the TS model. For aggregate data, where there is only one observation per age-cohort cell, the test of APC against TS is a deviance test for the logistic model and not feasible for a normal model. However, with cross-sectional information the test applies for both the logistic and the normal model.

In the time-saturated model we replace the APC design $\mu_{ik} = \alpha_i + \beta_j + \gamma_k + \delta$ in (2) with a complete unstructured specification of μ_{ik}, using an indicator for each age-cohort combination. We replace the p-dimensional design vector X_h with an n-dimensional unit-vector T_h indicating the age-cohort combination of individual h, where $p \ll n$. Thus, we compare the APC predictor $\eta_h = Z_h'\zeta + X_{i_h k_h}'\xi$ in (8) with the TS predictor

$$\eta_h = Z_h\zeta + T_h\kappa. \tag{13}$$

We stack X_h and T_h in design matrices X and T, where X belongs to the linear span of T.

Under classical normality assumptions, F-tests can be used to compare the models. Under asymptotic assumptions, likelihood ratio tests will be asymptotically χ^2 both for normal and logit models. In the following we address the computational issues arising when n is large. Indeed, in the obesity data, $n = 53 \times 14 - 5 * 6 = 712$, while $p = 53 + 14 + 56 - 3 = 120$.

4.1 Estimation of the normal time-saturated model

The combined design matrix $M = (Z, T)$ has $d_z + n$ columns. In the obesity example, $d_z = 15$ and $n = 712$. Consequently, it is computationally challenging to invert $M'M$ directly. We can address this problem by orthogonalizing the regressors.

In the implementation, we exploit that $T'T$ is diagonal since each row of T is a unit vector. Instead of regressing Y directly on M directly, we evaluate the partitioned regression

$$Y = \{Z - T(T'T)^{-1}T'Z\}\zeta + T\rho + \varepsilon. \tag{14}$$

Here $\{Z - T(T'T)^{-1}T'Z\} = \tilde{Z}$ is the residual of a first-stage regression of Z on T. Since $T'T$ is diagonal it can be inverted by inverting the diagonal elements, avoiding general matrix inversion routines. It is therefore easy to compute $\tilde{Z}$. Since $\tilde{Z}$ and T are orthogonal by construction, ζ and ρ are estimated by regressing Y on $\tilde{Z}$ and T, respectively. This poses no computational challenge since $\tilde{Z}$ has d_z columns, and d_z is small. If Z includes non-linear APC interactions, these can be expressed as functions of T. Then $\tilde{Z}$ has reduced rank, so degrees of freedom calculations will be affected.

We can retrieve κ from $\hat{\kappa} = \hat{\rho} - (T'T)^{-1}T'Z\hat{\zeta}$. Note that (14) and the equation in terms of M, ζ, and κ give equivalent models with the same fit and the same residual variance. As a consequence we are normally not interested in the value of $\hat{\kappa}$.

4.2 Estimation of the logit time-saturated model

In the logit model as in the normal model, the many regressors in the TS model can cause computational problems in inverting the information matrix. The solution is similar to that used for the normal model.

The time-saturated model is a logit model with linear predictor $\eta_h = Z_h\zeta + T_h\kappa$ as in (13). The corresponding success probability is $\pi_h = \exp(\eta_h)/\{1 + \exp(\eta_h)\}$. Thus, the score is

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

where Π is a H-length vector of probabilities π_h. The information matrix J 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 then find the inverse information as

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

Here, the matrix $J_{TT} = T'WT$ is large, but diagonal, since the rows of T are unit vectors and W is diagonal. Thus, as before, the inverse of J_{TT} can be found simply by inverting the diagonal elements. Further, the matrices J_{ZZ} and $J_{ZZ\cdot T} = J_{ZZ} - J_{ZT}J_{TT}^{-1}J_{TZ}$ have low dimension and can be inverted by standard matrix inversion algorithms. The logit model can thus be estimated by Newton iteration using the above calculation of the inverse information.

5 Empirical application to obesity in England

We use the proposed methods to examine the APC dynamics of obesity in England. Using log BMI as a continuous dependent variable, we find an age-drift model for women and an age-cohort model for men. Using an obesity indicator as dependent variable, we find an age-drift model for women and a cohort-drift model for men. Previous studies had detected non-linearities in age but had not detected the non-linearity in cohort among men.

5.1 Preliminary data analysis

We begin by visually inspecting the data, which was described in §2.1. The heatmap in Figure 2 shows the mean values of BMI in each age-cohort cell for women. The heatmap for men is similar, and therefore is not shown. We see a pattern of darker shading concentrated towards the right and centre of the graph and lighter shading concentrated towards the top, left, and perhaps bottom-left. This suggests that there are relationships between log BMI and one or more of age, period, and cohort. Our statistical analysis cannot identify the linear components of such relationships, but it can identify non-linear components. For example, the area of the array around ages 30-45 and periods 2005-2010 seems lighter in colour than the central area below it (ages 45-60, periods 2005-2010). However, the central area is similar in colour to the area below it (ages 60-80, periods 2005-2010). This suggests curvature, i.e. non-linearity, in the relationship between age and log BMI.

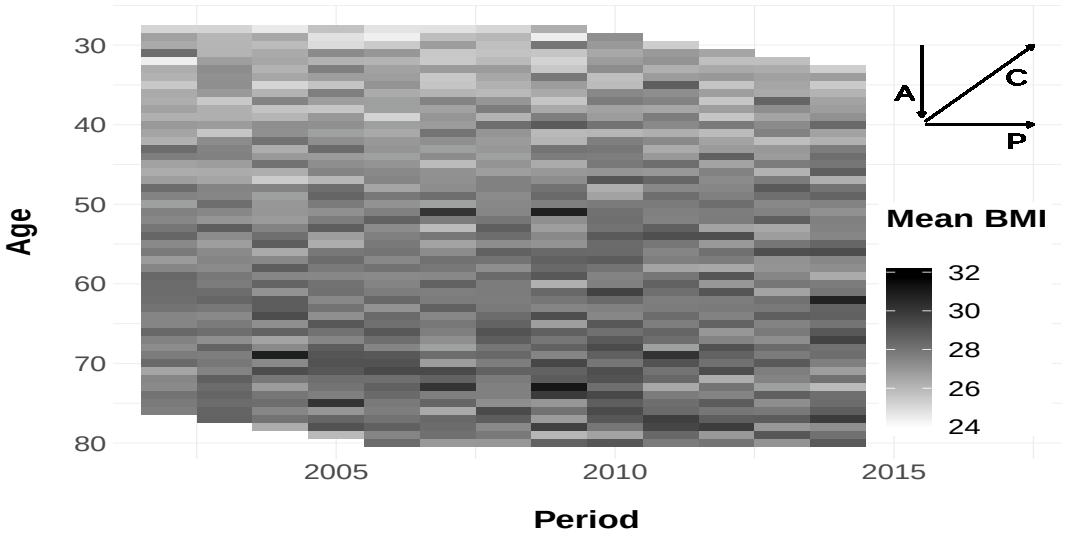


Figure 2: Within-cell BMI means women

5.2 Covariates

The covariates are ethnicity, level of education, NSSEC, smoking history, and alcohol consumption as described below. Tables 3 and 4 in Appendix B report descriptive statistics. The reference ethnicity was taken to be white, with indicators for self-identification as black, Asian, of mixed ethnicity, or of "other" ethnicity (including e.g. Arab).

For education, the reference group are those who left school after attaining a GCSE, the minimum school-leaving qualification obtained around age 16, or equivalent qualification. We include three indicators: education below GCSE level, holders of a university degree, and education beyond GCSE but below degree level.

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

Smoking behaviour is captured by two indicators. 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 indicators 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).

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

		Women				Men			
	df	F	p	AIC	ℓ	F	p	AIC	ℓ
TS				-22747.33	12101.67			-36449.64	18952.82
APC				-23321.91	11796.95			-37043.86	18657.93
AP	54	0.73	0.93	-23390.28	11777.14	1.85	0.00	-37051.87	18607.93
AC	12	1.38	0.17	-23329.33	11788.67	1.22	0.26	-37053.16	18650.58
PC	51	2.16	0.00	-23313.70	11741.85	1.57	0.01	-37065.67	18617.83
Ad	66	0.85	0.80	*-23397.46*	11768.73	1.73	0.00	-37061.31	18600.65
Pd	105	2.35	0.00	-23284.82	11673.41	4.80	0.00	-36750.82	18406.41
Cd	63	2.00	0.00	-23321.56	11733.78	1.50	0.01	*-37075.39*	18610.70
A	67	1.29	0.06	-23369.27	11753.64	2.63	0.00	-37001.23	18569.62
P	106	4.25	0.00	-23084.91	11572.46	5.05	0.00	-36722.71	18391.35
C	64	3.74	0.00	-23210.61	11677.31	3.33	0.00	-36958.31	18551.16

Degrees of freedom (df), F statistics and p-values are for tests against APC model. The sub-models are defined in §2.3. Italics indicate the minimum AIC values.

5.3 Model for a continuous outcome variable

We apply the normal APC model in §3.2 to analyse log BMI using the covariates described in §5.2. The analysis is performed separately for men and women.

5.3.1 Women

We begin by analysing the women. Table 1 provides an analysis of variance of the TS model, the APC model and the sub-models described in §2.3. Recall that the TS model has a indicator for each age-cohort combination; the AP, AC, and PC models each drop one non-linear effect; the Ad, Pd, and Cd models each drop two non-linear effects; and the A, P, C models drop one linear effect in addition to two non-linear effects. Columns 2-4 and 5-7 report F tests against, respectively, TS and APC models. Column 8 is the Asymptotic Information Criteria. Column 9 is the log-likelihood.

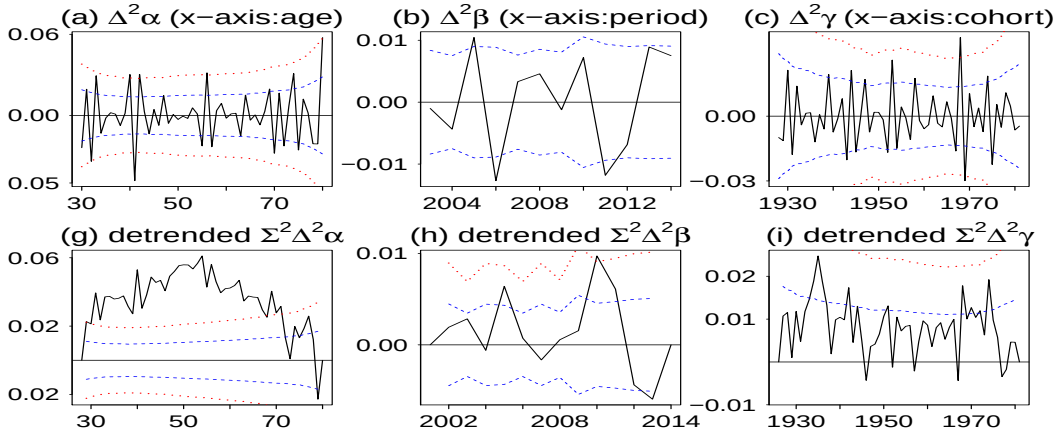
We start by checking the APC specification against the TS model. The APC model is preferred by the AIC and by an F-test. The $F(592, 42942)$ statistic 1.02 has a p-value of 0.36.

The APC model can be reduced further to an Ad model, which we recall from §2.3 is the age-drift model with a linear plane and a non-linear age effect. The reduction proceeds by the following logic. Minimizing the AIC points to the Ad model. The Ad model is rejected by neither the F test against the TS model nor that against the APC model. A more careful analysis follows the model reduction paths TS–APC–AP–Ad and TS–APC–AC–Ad. Neither path leads to rejection, noting that the p-values for F-tests of Ad against AP and AC are 0.93 and 0.16, respectively. There is no support for a further reduction to the A model. The p-value for testing A against Ad is negligible at 0.000. Since the A model is rejected we rule out the HAPC model of Yang and Land (2006), which has a quadratic age effect and zero mean period-cohort random effects.

Figure 3 displays the estimates for the APC model. The estimates for the preferred Ad model are very similar. Recall that the KNN parameter vector contains level υ_o and slopes υ_a, υ_c of a linear plane and three sets of double differences, $\Delta^2\alpha_i$, $\Delta^2\beta_j$, and $\Delta^2\gamma_k$. The double differences are plotted in the top three panels of Figure 3, marked (a), (b), (c).

The double differences can be interpreted as accelerations or differences-in-differences. For instance, $\Delta^2\alpha_i = \mu_{ik} - \mu_{i-1,k} - \mu_{i-1,k+1} + \mu_{i-2,k+1}$ for any value of the cohort k. Thus, $\Delta^2\alpha_i$

Figure 3: **Detrended time effects, APC model of log BMI, women**



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

captures the difference between the change of average log BMI from age $i-1$ to age i, for cohort k, and the change of average log BMI from age $i-2$ to age $i-1$, for cohort $k+1$. Both of these changes happen in period $j = i+k-1$. The interpretation of the double differences is not dependent on the index array $\mathcal{I}$ defined in (1). If, for instance, those over 70 were truncated from the sample, the remaining double differences would have the same appearance apart from sampling error.

In each of the panels (a)-(c), the reported pointwise confidence bands are not that informative about the joint significance, although there is a slight tendency to more variation in (a) than in (b), (c), in the line with the above reduction to the Ad model.

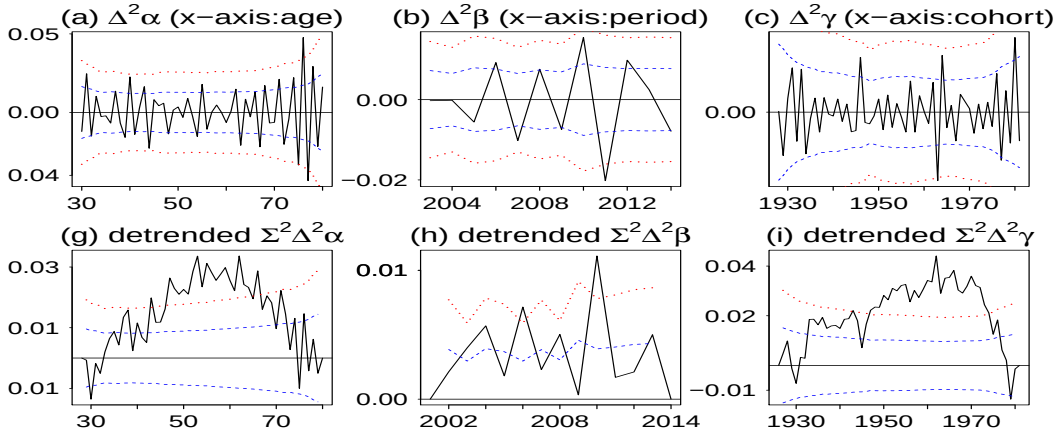
To ease interpretation of the double differences, the bottom panels marked (g), (h), and (i) show double cumulated double differences, detrended to start and end in 0. This cumulation gives us $\tilde{\alpha}_i = \sum_{t=3}^{i}\sum_{s=3}^{s}\Delta^2\alpha_s$ which is the same as the original age effect α_i in (2) *apart* from an unidentifiable linear trend. By detrending $\tilde{\alpha}_i$ to start and end in zero we visualize how α_i deviates from a linear trend, i.e. how it accelerates or decelerates. The age effect in panel (g) has a concave appearance in line with many epidemiological studies (Nielsen, 2015). The period and cohort effects in panels (h) and (i) show no particular signal, in line with their formal insignificance as found by the reduction to the Ad model. Recall that there are $I-2+K-2+J-2$ double-differences in ξ. There are $I+K+J$ cumulated double-differences, of which six are set to zero (two each in age, period, and cohort). Therefore the cumulated double differences have the same degrees of freedom as the double differences. Due to the cumulation, the pointwise confidence bands in (g)-(i) give a better indication of the significance of the age non-linearity in (g) and the insignificance of the period and cohort non-linearities in (h),(i), in line with the Ad model. A feature of these plots is that two constraints are imposed on each plot. This gives six constraints in total and implies that the detrended effects are unrelated to one another. This contrasts with an identification by constraints approach, where for instance the age effect cannot be interpreted in isolation from the period and cohort effect.

Figure 3 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. The corresponding figure for the Ad model has nearly the exact same panels (a,g). This is unsurprising given that the Ad model is not rejected against the APC model.

When detrending the removed trends have to be added to linear plane, see Nielsen (2015) for details. In the Ad model, the linear plane estimated this way is

$$\underset{(0.02)}{3.16} + \underset{(0.0004)}{0.0025}(age-28) - \underset{(0.0002)}{0.0013}(cohort-1921).$$

Figure 4: **Detrended time effects, APC model of log BMI, men**



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

We see that cohort coefficient is significant. This matches the earlier finding that the Ad model cannot be reduced to an A model. The cohort coefficient is interpreted as the change in linear plane when increasing the cohort by one while keeping the age fixed. This of course means that period is also increased by one. Therefore, the interpretation of the cohort coefficient in terms of the APC effect formulation in (2) is that it is the sum of the cohort and the period slopes. Similarly, the age coefficient is interpreted as the sum of the age and period slopes.

The coefficients on the covariates of the Ad model are seen in Table 2. Interpretation of these is deferred to §5.3.3, where they are discussed in conjunction with the model for men.

In appendix B, we report further mis-specification tests along with some comments on multiple testing issues and choice of significance level. Some further robustness checks are reported. We conclude that the overall outcomes are stable across changes to the specification.

5.3.2 Men

For the men, a similar approach is followed. Table 1 provides an analysis of variance. As before, both the AIC and an F-test favour the APC model over the TS model. The $F(592, 37589)$ statistic comparing the two is 0.98 with a p-value of 0.59. The APC model can be reduced further to an AC model. Minimizing AIC points to the Cd model. However, the F-test for the Cd model against the APC model rejects. Likewise, the p-value for testing the Cd against AC model is 0.006.

Figure 4 shows the estimates for the APC model. The estimates for the AC model are similar apart from omission of panels (b) and (h). There is some curvature in each of age and cohort, while the period non-linearity is driven by an anomalous spike in 2010. This lent support to our decision to exclude the PC model and focus on the AC model.

The estimated coefficients on the covariates are seen in Table 2. Mis-specification tests on the residuals are similar to those for women reported in Table 5 and therefore are not shown.

5.3.3 Interpretation

We selected the Ad model for women and the AC model for men. In neither case does the linear plane decompose uniquely into APC components. In both cases, there are deviations from linearity. For women there is concavity in log BMI. The concavity may be consistent with general metabolic effects or selection effects towards the end of life, as those with higher BMI may 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.

Table 2 shows the estimated effects of the covariates. These are largely consistent with the literature. Black individuals have higher BMI than white individuals, on average, while those of other ethnicities have lower BMI (Ogden et al., 2015; An and Xiang, 2016). BMI and social class are negatively correlated (McPherson et al., 2007). Those with more education have lower BMI on average (Baum II and Ruhm, 2009; An and Xiang, 2016). 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 (Akbartabartoori et al., 2005). Non-drinkers and rare drinkers have higher BMI than casual drinkers (the reference group), who in turn have higher BMI than frequent drinkers. This may be explained by the mismatch between frequency and quantity of consumption (O'Donovan et al., 2018).

There are some sex differences, primarily relating to significance. Black and mixed ethnicity women have significantly higher and lower BMI, respectively, than white women; whereas for men these ethnicities are not significant. (Sproston and Mindell, 2006; Agyemang et al., 2015). Non-drinking men do not differ significantly from occasional drinkers. Social class is not significant for men and effects are larger in women than in men (Devaux and Sassi, 2011).

5.4 Models for binary outcome variables

We apply the logit APC model in §3.3 to analyse an obesity indicator taking the value 1 for $BMI \geq 30$ and 0 otherwise. We use the same covariates as before and analyse women and men separately. We summarize the findings here, leaving details to §B.2 in the appendix.

For women we choose the Ad model, just as in the model for log BMI. The detrended age effects are shown in figure 6(a). The plot is similar to figure 3(g), with a change in curvature in the mid 50s. However, this is perhaps less pronounced in the obesity model, indicating that an increasing number of women gain weight, but few pass the obesity threshold.

For men we choose the Cd model, noting that for log BMI we favoured the AC model with weak support for the Cd model. Figure 6(b) shows the detrended cohort effects, which appear to be a mix of the separate detrended age and cohort effects for log BMI, but with an earlier peak. This could be explained as follows: the group of men aged 40-60 in 2001-2014 have higher mean BMI than those of other ages. These men belong to cohorts 1940-1980. Because of the limited period range of this dataset, 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 to separate the cohort and age influences for this group with this data.

Table 2 reports coefficients to the covariates. For women, significance (at a 5% level) and signs are exactly the same as for log BMI. Moreover, the logit coefficients are all about 10 times the least squares cofficients. For men, the logit coefficients are broadly speaking in line with those reported for log BMI.

Table 2: **Covariate effects**

	normal models for log BMI				logit models for obesity			
	Women, Ad		Men, AC		Women, Ad		Men, Cd	
	$\hat{\zeta}$	*se*	$\hat{\zeta}$	*se*	$\hat{\zeta}$	*se*	$\hat{\zeta}$	*se*
Ethnicity								
Black	0.068*	0.007	-0.008	0.006	0.658*	0.077	-0.034	0.095
Asian	-0.045*	0.007	-0.039*	0.005	-0.593*	0.105	-0.588*	0.089
Mixed	-0.023‡	0.011	-0.008	0.011	-0.190	0.149	-0.170	0.172
Other	-0.069*	0.012	-0.033*	0.011	-0.628*	0.178	-0.343†	0.188
Smoker								
Former	0.024*	0.002	0.026*	0.002	0.201*	0.027	0.310*	0.027
Current	-0.030*	0.002	-0.045*	0.002	-0.216*	0.030	-0.356*	0.033
Drinking frequency								
Never	0.043*	0.008	0.001	0.010	0.507*	0.097	0.347‡	0.143
Rarely	0.037*	0.002	0.014*	0.002	0.428*	0.025	0.201*	0.029
Frequently	-0.032*	0.003	-0.017*	0.002	-0.332*	0.035	-0.158*	0.029
Education								
Below GCSE	0.016*	0.003	0.010*	0.002	0.194*	0.031	0.160*	0.035
Some higher	-0.012*	0.003	0.002	0.002	-0.109*	0.033	-0.008	0.034
University	-0.048*	0.003	-0.026*	0.003	-0.452*	0.040	-0.280*	0.040
3 level NSSEC								
Intermediate	-0.021*	0.002	-0.001	0.002	-0.239*	0.029	-0.036	0.033
Managerial	-0.009*	0.003	-0.001	0.002	-0.084*	0.032	-0.117*	0.031
Other	-0.013†	0.007	-0.020†	0.011	-0.078	0.086	0.015	0.162

All variables are indicators with the following reference catagories: white ethnicity; never smoker; occasionally drink alcohol; GCSE education level; routine work for NSSEC. p-values: * $p \leq 0.01$, ‡ $0.01 < p \leq 0.05$, † $0.05 < p \leq 0.10$.

6 Conclusion

Our analysis of the obesity data for England demonstrated clear non-linear age and cohort effects as well as covariate effects that were robust to a range of specifications. For women, the only significant deviation from linearity in the APC effects is concavity in the age dimension. We suggest metabolic changes, child-bearing, and child-rearing as potential reasons for this. For men, there is significant concavity in the cohort dimension, and, for log BMI only, in the age dimension. For both sexes, the impact of covariates is largely consistent with existing literature, although more covariates are significant in the models for women.

To estimate these effects, we employed the canonical parametrization of Kuang et al. (2008). This parametrization facilitated estimation, inference, and identification of the non-linear APC effects, while avoiding issues relating to the unidentifiable linear trends. We used the parametrization in normal and logit models for individual level data. To assess the adequacy of the APC model, we presented tests against a time-saturated model. Additionally, we showed in §A.3 that minimally-constraining ad hoc identification yields the same covariate coefficient estimates as the KNN parametrization. This strategy is suitable for inference when only the covariates are of interest and the age, period, and cohort effects are considered nuisance parameters.

The existing framework could be expanded to allow for mixture models, interaction terms between APC effects and covariates, and heteroskedastic errors. This would enable us to

address some of the mis-specification concerns in the log BMI analysis. It would be of interest to analyse the consequence of missing age-cohort cells. Indeed, in the logit application we dropped all observations with ages below 28 to avoid perfect separation in a single age-cohort cell; it would be preferable to avoid this.

Acknowledgements

Suggestions from Jurgen Doornik are gratefully acknowledged. Funding was received from ESRC grant ES/J500112/1 (Fannon), ERC grants 681546, FAM-SIZEMATTERS (Monden), and 694262, DisCont (Fannon, Nielsen).

A Details on APC parametrizations

A.1 The canonical parametrization

The classical APC model is $\mu_{ik} = \alpha_i + \beta_j + \gamma_k + \delta$ as stated in (2) of §2.2. The KNN reparametrization $\mu_{ik} = X_{ik}'\xi$, with ξ defined in (5), arises as follows.

Define $\Delta\alpha_t = \alpha_t - \alpha_{t-1}$ and $\Delta^2\alpha_s = \Delta\alpha_s - \Delta\alpha_{s-1}$ as well as the telescopic sum

$$\alpha_i = \alpha_1 + \sum_{t=2}^{i} \Delta\alpha_t \qquad\qquad \Delta\alpha_t = \Delta\alpha_2 + \sum_{s=3}^{t} \Delta^2\alpha_s. \tag{18}$$

Insert this and similar expressions for β_j and γ_k in the equation $\mu_{ik} = \alpha_i + \beta_j + \gamma_k + \delta$ to get

$$\mu_{ik} \;=\; \upsilon_o + (i-U)\upsilon_a + (k-U)\upsilon_c + A_i + B_j + C_k, \tag{19}$$

where the reference point U is the integer value of $(L+3)/2$ and L is the offset in (1) of §2.1. Moreover, $\upsilon_o = \mu_{UU}$, $\upsilon_a = \mu_{U+1,U} - \mu_{UU}$ and $\upsilon_c = \mu_{U,U+1} - \mu_{UU}$, as in (6), while

$$\begin{aligned}
A_i &= 1_{(i<U)} \sum_{t=i+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2\alpha_s + 1_{(i>U+1)} \sum_{t=U+2}^{i} \sum_{s=U+2}^{t} \Delta^2\alpha_s, \\
B_j &= 1_{(L\,odd\,\&\,j=2U-2)} \Delta^2\beta_{2U} + 1_{(j>2U)} \sum_{t=2U+1}^{j} \sum_{s=2U+1}^{t} \Delta^2\beta_s, \\
C_k &= 1_{(k<U)} \sum_{t=k+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2\alpha_s + 1_{(k>U+1)} \sum_{t=U+2}^{k} \sum_{s=U+2}^{t} \Delta^2\alpha_s.
\end{aligned}$$

The component $\upsilon_o + (i-U)\upsilon_a + (k-U)\upsilon_c$ parametrizes a linear plane in age and cohort coordinates, but it does not identify the linear APC slopes, see (7). The present formulation gives symmetry in age and cohort and has a simple expression when $L = 0$ as in KNN. Equation (19) can be written as $\mu_{ik} = X_{ik}'\xi$ with ξ as in (5) and design vector

$$X_{ik}' = (1, i-U, k-U, x_\alpha, x_\beta, x_\gamma) \tag{20}$$

where we define the function $m(t,s) = \max(t - x + 1, 0)$ and the vectors

$$\begin{aligned}
x_\alpha &= \{m(1,i), \ldots m(U-1,i), m(i,U+2), \ldots m(i,I)\}, \\
x_\gamma &= \{m(1,k), \ldots m(U-1,k), m(k,U+2), \ldots m(k,K)\} \\
x_\beta &= \begin{cases} \{1_{j=L+1}, m(j,L+4), \ldots, m(j,L+J)\} & \text{for odd } L, \\ \{m(j,L+3), \ldots, m(j,L+J)\} & \text{for even } L. \end{cases}
\end{aligned}$$

See Fannon and Nielsen (2019) for further motivation and details.

A.2 Properties of ad hoc identification schemes

Suppose θ is ad hoc identified by the constraint $\alpha_1 = \alpha_2 = \beta_J = \gamma_K = 0$ as in Mason et al. (1973). Another example is the analysis of Ejrnæs and Hochguertel (2013), where one age, one cohort, and two period indicators are dropped. These are examples of constraints of the type $L'\theta = 0$. for L an appropriate selection matrix We show that the APC predictor thus constrained has the form $\mu = XQ\phi$ for some invertible $p \times p$ matrix Q and a parameter vector ϕ. We describe how the constrained APC effects are found from ϕ.

Suppose, that $L'L$ is invertible. Then L has orthogonal complement $L_\perp$ of dimension $q \times p$, so that $L'_\perp L = 0$ and $(L, L_\perp)$ is invertible. For simplicity, suppose $L'L = I_4$ and $L'_\perp L_\perp = I_p$. Then, the orthogonal projection identity is $I_q = LL' + L_\perp L'_\perp$.

The identification problem in (4) is that μ as a function of θ is invariant to the transformations $\theta \mapsto \theta + A_\perp v$, where $v = (a, b, c, d)'$ and $A_\perp$ is a $q \times 4$ matrix (Nielsen and Nielsen, 2014). Thus, μ only depends on θ through $\xi = A'\theta$ where A is a $q \times p$ matrix so that $A'A_\perp = 0$. Further, the collinearity of the D_{ik} vectors is given by $D_{ik} = X_{ik}A'$.

The original design matrix D has reduced column rank. Since $I_q = LL' + L_\perp L'_\perp$ we get $D\theta = DLL'\theta + DL_\perp L'_\perp\theta$. The constraint $L'\theta = 0$ implies $D\theta = DL_\perp L'_\perp\theta$. Here, $DL_\perp$ is a $q \times p$ matrix and it must have full column rank. The reason is that Corollary 2 in KNN implies that D has rank p. In other words, the constraint $L'\theta = 0$ is identifying when $DL_\perp$ has full column rank and the constrained θ is estimated by dropping the columns DL from D.

The identity $D = XA'$ shows $D\theta = XA'L_\perp L'_\perp\theta$. Let $Q = A'L_\perp$ and $\phi = L'_\perp\theta$, so that $D\theta = XQ\phi$. From above, $DL_\perp = XA'L_\perp = XQ$ has full column rank, so that Q is invertible.

To find the constrained parameter, satisfying $L'\theta = 0$ say, use $I_q = LL' + L_\perp L'_\perp$ to get $\theta = LL'\theta + L_\perp L'_\perp\theta$. Since $L'\theta = 0$ and $L'_\perp\theta = \phi$ then $\theta = L_\perp\phi$.

We note that in general ϕ is not invariant. Indeed, applying the transformation $\theta \mapsto \theta + A_\perp v$ to $\phi = L'_\perp\theta$ gives $L'_\perp\theta + L'_\perp A_\perp v$. In general, this depends on v, unless we choose $L_\perp = A$ so that $L'_\perp A_\perp v = 0$. With the choice $L_\perp = A$ we get $\phi = \xi$.

A.3 Covariate parameters and identification of APC effects

Recall the model $\eta = Z\zeta + \mu$ for the linear predictor in equation (9). We show that when only the covariate effects ζ are of interest it does not matter whether the APC effects in μ are parametrized through the parametrization $\mu = X\xi$ as in (9), or by ad hoc identification.

Suppose that we identify the APC structure by a constraint such as $\alpha_1 = \alpha_2 = \beta_J = \gamma_K = 0$ or some other constraint on the form $L'\theta = 0$ as analyzed in appendix A.2. Following appendix A.2, we can just as well suppose the predictor satisfies $\mu = XQ\phi$, where Q is a known, invertible $p \times p$-matrix and $\phi = Q^{-1}\xi$. Thus, we have two parametrizations: $\eta = Z\zeta + X\xi$ and $\eta = Z\zeta + XQ\phi$. The mapping between the two parametrizations is one-one, since Q is invertible. Due to the equivariance of maximum likelihood estimators (Cox and Hinkley, 1974, §1) the maximum likelihood estimators of ζ are the same under the two parametrizations.

B Further data analysis

The data is available through the UK Data Service[1]. Tables 3 and 4 give descriptive statistics.

Note that the alcohol categories do not account for quantity of drinks per drinking event.

B.1 Robustness checks for normal models

Formal mis-specification tests for the Ad model are reported in Table 5 in situations with and without log transformation of the dependent variable. The tests include a cumulant based test

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

Table 3: **Descriptive statistics: Continuous variables**

	Women				Men			
	min	mean	median	max	min	mean	median	max
Age	28	51	50	80	28	52	51	80
Period	2001	2007	2006	2014	2001	2007	2006	2014
Cohort	1926	1956	1957	1981	1926	1955	1956	1981
BMI	13.2	27.4	26.4	58.9	13.6	27.9	27.4	59.5
Height (cm)	124	162	162	202	138	175	175	203
Weight (kg)	28.4	71.5	69	164	34.2	85.5	84	203

Table 4: **Descriptive statistics: Indicators**

	Women, 43077 observations					Men, 38316 observations				
	1	2	3	4	5	1	2	3	4	5
Ethnicity	762	41071	702	288	254	600	36363	972	201	180
Class	4585	11721	14302	748		6972	7390	16363	201	
Education	13138	11847	9644	8448		10927	7865	10456	9068	
Alcohol	497	15703	19534	7343		228	8330	19489	10269	
Smoker	23037	10724	9316			16692	13084	8540		

Classification. Ethnicity: 1 black, 2 white, 3 Asian, 4 mixed, 5 other.
Social class: 1 routine & manual, 2 intermediate, 3 managerial & professional, 4 other.
Education level: 1 below GCSE, 2 GCSE, 3 some higher, 4 university degree.
Alcohol drinking (events per week): 1 never, 2 rare (<1), 3 occasional (1-4), 4 frequent ($\geq$ 5).
Smoker: 1 never, 2 former, 3 current.

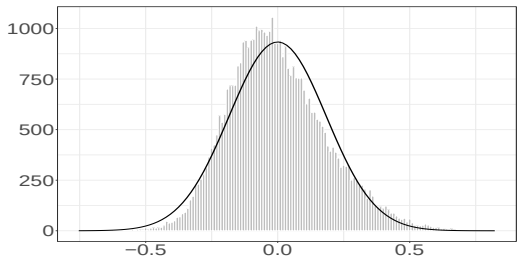
for normality of residuals and tests for functional form mis-specification and heteroskedasticity (Ramsey, 1969; White, 1980). The log transformation clearly improves the specification. Yet, given the large sample size, $H = 43,077$, it is difficult to avoid very small p-values. The histogram of the residuals in figure 5 suggests a systematic, if modest skewness in the residuals that could be addressed by dropping the least squares approach in favour of a regression based on a non-normal distribution.

In the above analysis, we have conducted many tests, yet taken an informal approach to the choice of significance level. This is guided by the following ideas. The APC model is the central model, which we test using a handful of mis-specification tests: the test against the TS model, and those reported in Table 5. If each of these tests is conducted at a 1% level

Table 5: **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

the overall level for mis-specification tests is about 5%. We then reduce the APC model by testing a number of nested sub-models. If each sub-model is compared both with the model immediately above and with the APC model at a 5% level the overall level for the reduction is likely to be in the neighbourhood of 5%. The reduction tests and the mis-specification tests are likely to be independent, which gives an overall size of about 10%. We refrain from a detailed analysis here, noting that it is more important to record marginal decisions with p-values in the range from 1% to 10% than to get the size calculation exactly right.

A range of alternative specifications of the normal model were examined as robustness checks. We considered a model replacing the three-class NSSEC with the eight-class version. We considered a using BMI instead of log BMI as the dependent variable. We also considered a model with log weight as the dependent variable and log height as an explanatory variable; a model with log BMI as the dependent variable implicitly imposes a coefficient of two in this regression, and we wanted to evaluate whether that was restrictive. We also re-estimated the log BMI and obesity models using the interview weights provided by the HSE; these weights account for differences in participation by age, sex, region, household type, and social class. 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. There was no substantial change to the non-linearities or the covariates. 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 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 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 is not a problem in the normal model. The main consequence of this was a strengthening of the significance of age non-linearities for men, with curvature in the early twenties that could be explained by particularly rapid growth in log BMI over those ages. 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 concavity, which could be explained

Table 6: **Model comparisons, obesity indicator, women and men**

		Women				Men			
	df	LR	p	AIC	ℓ	LR	p	AIC	ℓ
TS				48708.92	-23627.46			44563.50	-21554.75
APC				48123.02	-23926.51			43935.76	-21832.88
AP	54	34.49	0.98	48049.51	-23943.76	73.86	0.04	43901.62	-21869.81
AC	12	11.19	0.51	48110.21	-23932.10	20.81	0.05	43932.57	-21843.29
PC	51	73.75	0.02	48094.76	-23963.38	58.97	0.21	43892.74	-21862.37
Ad	66	45.72	0.97	*48036.74*	-23949.37	94.81	0.01	43898.57	-21880.28
Pd	105	154.69	0.00	48067.71	-24003.85	305.24	0.00	44031.00	-21985.50
Cd	63	85.31	0.03	48082.32	-23969.16	79.14	0.08	*43888.90*	-21872.45
A	67	71.48	0.33	48060.50	-23962.25	153.20	0.00	43954.96	-21909.48
P	106	211.26	0.00	48122.28	-24032.14	307.12	0.00	44030.88	-21986.44
C	64	147.62	0.00	48142.63	-24000.32	151.02	0.00	43958.78	-21908.39

Degrees of freedom (df), LR statistics and p-values are for tests against APC model.

by a reduction in the rate of growth in log BMI after the 1990s.

In addition to the robustness checks above, we have the mis-specification 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.

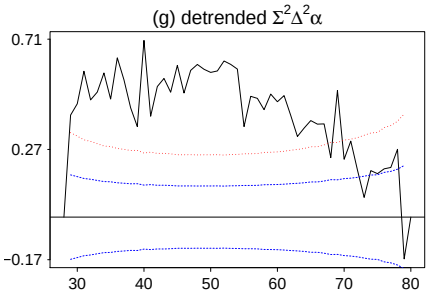
B.2 Details of binary analysis

The main findings of the binary analysis are summarized in §5.4. Here, we present details of the model reduction. This follows the approach for continuous outcomes in §5.3.

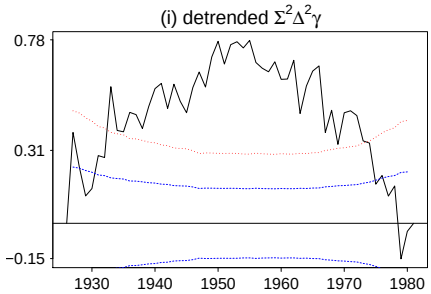
Model comparison statistics are presented in Table 6. For the women, the APC model is preferred over the TS specification according to the AIC and the LR test. The LR statistic is 598.10 with $p = 0.42$ when compared with a $\chi^2(592)$ distribution. Moving on to the model reduction, the preferred model is the Ad model for the following reasons: the AIC favours the Ad model; The Ad model is not rejected by an F-test against the APC model; it is not rejected by F-tests against the AP and the AP models, with p-values of 0.98 and 0.51, respectively. We note that the A model is not rejected against the APC model, but it has a p-value of 0.00 against the Ad model and so is rejected. Figure 6(a) shows the detrended age effect.

For the men, the APC model is preferred over the TS specification according to the AIC and the LR test. The LR statistic is 556.26 with $p = 0.85$ by a $\chi^2(592)$ distribution. As a mis-specification test, we see that the APC model is not rejected against the TS model. Moving on to the model reduction, the preferred model is the Cd model for the following reasons: the AIC favours the Cd model; the Cd model is not rejected by an F-test against the APC model; it is also not rejected by F-tests against the AC and the PC models with p-values of 0.22 and 0.064, respectively, noting that the decision against the PC model is marginal. There is no support for reduction to the smaller C model. Figure 6(b) shows the detrended cohort effect.

Figure 6: Detrended time effects, models of obesity indicator



(a) women, Ad model, age effect



(b) men, Cd model, cohort effect

References

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

Agyemang, C., Kunst, A., Bhopal, R., Zaninotto, P., Nazroo, J., M., N., Unwin, N., van Valkengoed, I., Redekop, K., and Stronks, K. (2015). Dutch versus English advantage in the epidemic of central and generalised obesity is not shared by ethnic minority groups: comparative secondary analysis of cross-sectional data. *International Journal of Obesity*, 35:1334–1346.

Akbartabartoori, M., Lean, M. E. J., and Hankey, C. R. (2005). Relationships between cigarette smoking and body shape. *International Journal of Obesity*, 29:236–243.

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

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

Bell, A. and Jones, K. (2014a). Don't birth cohorts matter? a commentary and simulation exercise on reither, hauser, and yang's (2009) age–period–cohort study of obesity. *Social Science & Medicine*, 101:176–180.

Bell, A. and Jones, K. (2014b). Don't birth cohorts matter? A commentary and simulation exercise on Reither, Hauser, and Yang's (2009) age-period-cohort study of obesity. *Social Science & Medicine*, 101:176–180.

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

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

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

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

Devaux, M. and Sassi, F. (2011). Social inequalities in obesity and overweight in 11 OECD countries. *European Journal of Public Health*, 23:464–469.

Ejrnæs, M. and 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. and Kaufmann, H. (1986). Asymptotic inference in discrete response models. *Statistical Papers*, 27:179–205.

Fannon, Z. and Nielsen, B. (2019). Age-period-cohort models. *Oxford Research Encyclopedia, Economics and Finance*. DOI: 10.1093/acrefore/9780190625979.013.495.

Fu, W. (2016). Constrained estimators and consistency of a regression model on a Lexis diagram. *Journal of the American Statistical Association*, 111:180–199.

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

Harnau, J. and Nielsen, B. (2018). Over-dispersed age-period-cohort models. *Journal of the American Statistical Association*, 113:1722–1732.

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

Howel, D. (2011). Trends in the prevalence of obesity and overweight in English adults by age and birth cohort, 1991-2006. *Public Health Nutrition*, 14:27–33.

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

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

Lean, M. E. J., Katsarou, C., McLoone, P., and Morrison, D. S. (2013). Changes in BMI and waist circumference in Scottish adults: use of repeated cross-sectional surveys to explore mulitple age groups and birth-cohorts. *International Journal of Obesity*, 37:800–808.

Lee, R. D. and Carter, L. R. (1992). Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87:659–671.

Luo, L. (2013). Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. *Demography*, 50:1945–1967.

Martínez Miranda, M. D., Nielsen, B., and 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., and Poole, K. (1973). Some methodological issues in cohort analysis of archival data. *American Sociological Review*, 38:242–258.

McPherson, K., Marsh, T., and 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.

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

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

O'Brien, R. M. (2011). Constrained estimators and age-period-cohort models (with discussion). *Sociological Methods & Research*, 40:419–470.

O'Donovan, G., Stamatakis, E., and Hamer, M. (2018). Associations between alcohol and obesity in more than 100 000 adults in England and Scotland. *British Journal of Nutrition*, 119:222–227.

Ogden, C. L., Carroll, M. D., Fryar, C. D., and 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.

Oh, C. and Holford, T. R. (2015). Age-period-cohort approaches to back-calculation of cancer incidence rate. *Statistics in Medicine*, 34:1953–1964.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

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., and 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.

Scarborough, P., Bhatnagar, P., Wickramasinghe, K. K., Allender, S., Foster, C., and Rayner, M. (2011). The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. *Journal of Public Health*, 33:527–535.

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

Snijders, T. A. B. and Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advance Multilevel Modeling*. Sage, London, 2nd edition.

Sproston, K. and Mindell, J. (2006). *Health survey for England 2004. Volume 1. The Health of minority ethnic groups*. The Information Centre, London.

Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L., and Brown, M. (2011). Health and economic burden of the projected obesity trends in the usa and the uk. *Lancet*, 378:815–825.

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*. Massachusetts Institute of Technology, Cambridge, MA.

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. and 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.