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

Zoë Fannon, Christiaan Monden, Bent Nielsen

06 October 2019

We develop a model to study non-linearities in age, period, and cohort for repeated cross-section data with covariates. The effects of age, period, and cohort are of interest in many applications in economics, such as life-cycle saving and growth, consumption, migration, and inequality (Krueger & Pischke, 1992; Deaton & Paxson, 2000; Browning et al., 2016; Beenstock et al., 2010; Kalwij & Alessie, 2007). The age of a person, the period in which they are observed, and their cohort (i.e. the generation to which they belong) may all separately influence their behaviour in any of these contexts, and so are all of interest.

We focus on non-linearities in each of age, period, and cohort because the linear parts are not identifiable. They are not identifiable because of the linear relationship between the three variables of interest: age = period - cohort. Applied researchers have long been aware of this identification problem, and of the fact that the non-linear parts can be identified (Heckman & Robb, 1985; Deaton & Paxson, 2000). While this awareness exists in principle, in practice applied researchers often use methods which do not fully separate linear and non-linear effects, and thus end up drawing untestable conclusions about linear effects (Allman-Farinelli et al., 2008; Keyes et al., 2010).

In this paper, we employ a model that uses a parametrization of the age, period, and cohort effects which completely separates the linear and non-linear parts of these effects. This parametrization was developed by Kuang et al. (2008) for aggregate data, but to date has not been used for individual level data. The parametrization has two parts. The first part is the non-linear effects in each of age, period, and cohort. For each of the three, a series of "double-differences" is estimated. These double-differences constitute a non-parametric model for the non-linear effects of age, period, and cohort. The second part of the parametrization is a linear plane, which combines the linear parts of age, period, and cohort and which should not be directly interpreted.

Our first contribution is to demonstrate that this parametrization can be adapted for use with repeated cross-sectional data and covariates. We embed both the age-period-cohort parametrization and the covariates in the linear predictor of a generalized linear model (McCullagh & Nelder, 1989). We use the generalized linear modelling framework because it allows us to consider both continuous and binary outcomes. Asymptotic inference is straightforward in this framework, and this enables us to test restrictions on the model, such as that non-linearities in one or more of age, period, and cohort are insignificant.

Our second contribution is to provide a test of this model of age, period, and cohort effects against a more general model of time effects, which is suitable for repeated cross-sectional data. The model of age, period, and cohort effects we use assumes that there are no interaction effects between the three, a point that is recognised in the literature (Kupper et al., 1985). Such interaction effects would capture the impact of, for instance, time- and age-limited government programs like healthy school meals. The assumption that there are no interaction effects is therefore restrictive, and must be tested.

The test we use compares our model of age, period, and cohort effects to a model with an indicator for each age-by-period interaction that is present in the data, which we call the "time-saturated" model. This time-saturated model nests our age-period-cohort model, and therefore the two can be compared using likelihood ratio tests. This test bears similarity to the deviance test of the age-period-cohort model used in the context of aggregate data (Martínez Miranda et al., 2015). The time-saturated model is similar to the fully-saturated model used in the deviance test.

The high dimension of the time-saturated model poses computational challenges, so we develop a custom algorithm to estimate it. The computational challenge is that estimation requires inversion of a high-dimensional matrix. The algorithm to overcome this challenge relies on two elements: the fact that the indicators in the time-saturated model are mutually exclusive, and the orthogonalization of the indicators from the other covariates. We separate the time-saturating indicators from the covariates in the linear predictor of the generalized linear model. The part relating to the mutually-exclusive indicators then involves a diagonal matrix, which is easy to invert. The part involving the covariates is small, and so also easy to invert.

We demonstrate the use of our age-period-cohort model with covariates, and our test of this model against the time-saturated model, in a study of obesity in England from 2001-2014 . Obesity is a growing public health concern. It is linked to type II diabetes, reduced life expectancy, higher rates of sickness, and negative psychological effects (Moody, 2016; Hruby et al., 2016). 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). Understanding the relationship between obesity and age, period, and cohort can help to identify the most at-risk populations. Future research on the phenomenon, as well as interventions to address it, can then be targeted towards those populations.

For our study of obesity, we use data from fourteen waves of the Health Survey for England, between 2001 and 2014. We consider individuals aged between 28 and 80, with age measured in one-year increments. Obesity is constructed as having body mass index in excess of 30, where body mass index is calculated from weight and height, which are directly measured by a nurse. Following the literature, we analyse men and women separately; our sample includes observations for 38,316 men and 43,077 women. We include ethnicity, education, social class, alcohol consumption, and smoking as covariates.

Considering obesity among women, we find that the age-period-cohort model passes the test against the time-saturated model, and that the only significant non-linearity is concavity in age. The concavity may be consistent with general metabolic effects or selection effects towards the end of life, as obesity reduces life expectancy (Hruby et al., 2016). Child-bearing and child-rearing may also be a factor. The included covariates have signs and significance consistent with the existing literature on obesity (An & Xiang, 2016; McPherson et al., 2007; Akbartabartoori et al., 2005; Traversy & Chaput, 2015).

Considering obesity among men, we find that the age-period-cohort model passes the test against the time-saturated model, and that the only significant non-linearity is concavity in cohort. The concavity may be related to generational shifts in the nature of employment, in particular to the shift towards sedentary employment. Again, the included covariates have signs and significance consistent with the existing literature on obesity.

The substantive implications of our findings on obesity, then, are that men and women must be considered separately, and that future research into the drivers of obesity should focus on explanations consistent with concavity in age for women and cohort for men. Similar conclusions have been reached by previous studies, but those studies used methods which failed to adequately separate non-linear and linear effects (Howel, 2011; Lean et al., 2013). The present paper adds to the literature on obesity by showing that those results are robust to using a method which fully separates the non-linear and linear effects of age, period, and cohort. It also provides evidence that a model of age, period, and cohort effects alone, with no interaction effects, is sufficient for a model of obesity.

The age-period-cohort model and the test against the time-saturated model can be applied to other settings where age, period, and cohort are of interest to economists. Examples include life-cycle saving and growth, consumption, migration, and inequality. In each case, it will be necessary to express questions and hypotheses in terms of the identifiable non-linearities, rather than the unidentifiable linearities, of age, period, and cohort. This will require substantial future research. Further research will also be required to extend the theoretical framework to consider additional settings, such as panel data and gaps in the sequence of age, period, or cohort.

References

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

Allman-Farinelli, M. A., Chey, T., Bauman, A. E., Gill, T., & James, W. P. T. (2008). Age, period and birth cohort effects on prevalence of overweight and obesity in Australian adults from 1990 to 2000. *European Journal of Clinical Nutrition*, *62*, 898–907.

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

Beenstock, M., Chiswick, B. R., & Paltiel, A. (2010). Testing the immigrant assimilation hypothesis with longitudinal data. *Review of Economics of the Household*, *8*, 7–7.

Browning, M., Crossley, T. F., & Luhrmann, M. (2016). Durable purchases over the later life cycle. *Oxford Bulletin of Economics and Statistics*, *78*, 145–169.

Deaton, A. & Paxson, C. (2000). Growth, demographic structure, and national saving in taiwan. *Population and Development Review*, *26*, 141–173.

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

Heckman, J. & Robb, R. (1985). *Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations*, (pp. 137–150). New York, NY: Springer New York.

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*(1), 27–33.

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.

Kalwij, A. S. & Alessie, R. (2007). Permanent and transitory wages of british men, 1975-2001: year, age, and cohort effects. *Journal of Applied Econometrics*, *22*, 1063–1093.

Keyes, K. M., Utz, R. L., Robinson, W., & Li, G. (2010). What is a cohort effect? comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971–2006. *Social science & medicine*, *70*(7), 1100–1108.

Krueger, A. B. & Pischke, J. (1992). The effect of social security on labor supply: a cohort analysis of the notch generation. *Journal of Labor Economics*, *102*, 412–437.

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.

Kupper, L. L., Janis, J. M., Karmous, A., & Greenberg, B. G. (1985). Statistical age-period-cohort analysis: A review and critique. *Journal of Chronic Diseases*, *38*, 811–830.

Lean, M. E. J., Katsarou, C., McLoone, P., & 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.

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.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models*. Springer.

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.

Traversy, G. & Chaput, J. (2015). Alcohol consumption and obesity: an update. *Current obesity reports*, *4*, 122–130.