Article:

Bell, A.J. and Jones, K. (2015) Bayesian Informative Priors with Yang and Land's Hierarchical Age-Period-Cohort model. Quality and Quantity, 49 (1). 255 - 266. ISSN 1573-7845

https://doi.org/10.1007/s11135-013-9985-3

Bayesian Informative Priors with Yang and Land's Hierarchical Age-Period-Cohort model

Andrew Bell and Kelvyn Jones

School of Geographical Sciences and Centre for Multilevel Modelling, University of Bristol

School of Geographical Sciences
University of Bristol
University Road
Bristol
BS8 1SS

Contact: Andrew.bell@bristol.ac.uk

Last updated: 5th December 2013

Abstract

Previous work (Bell and Jones 2013a, c; Luo and Hodges 2013) has shown that, when there are trends in either the period or cohort residuals of Yang and Land's Hierarchical Age-Period-Cohort (APC) model (Yang and Land 2006; Yang and Land 2013), the model can incorrectly estimate those trends, because of the well-known APC identification problem. Here we consider modelling possibilities when the age effect is known, allowing any period or cohort trends to be estimated. In particular, we suggest the application of informative priors, in a Bayesian framework, to the age trend, and we use a variety of simulated but realistic datasets to explicate this. Similarly, an informative prior could be applied to an estimated period or cohort trend, allowing the other two APC trends to be estimated. We show that a very strong informative prior is required for this purpose. As such, models of this kind can be fitted but are only useful when very strong evidence of the age trend (for example physiological evidence regarding health). Alternatively, a variety of strong priors can be tested and the most plausible solution argued for on the basis of theory.

Keywords

Age-period-cohort models, MCMC, collinearity, informative priors

1. Introduction

Age-period-cohort modelling has recently been reinvigorated by a number of recent methodological innovations (Yang and Land 2013; Yang and Land 2006; Yang et al. 2008; Tu et al. 2011) which, it is claimed, overcome the identification problem caused by the collinearity between age, period and cohort. However, it has long been argued that, for identification of such a model to occur, some quite strong assumptions must be made (Glenn 1976, 2005). Indeed recent work on such models has shown that the recent wave of innovations are no different in making assumptions which, if unjustified, will lead to extremely misleading inference (Luo 2013; Luo and Hodges 2013; Bell and Jones 2013a, b, c, d).

Despite this, there are circumstances where such models are of use. Yang and Land's Hierarchical Age-Period-Cohort (HAPC) model is one of the more conceptually appealing of these models. Whilst it is unable automatically to tell apart trends in age, periods or cohorts, and solve the identification problem as its authors have claimed, it is capable of analyzing period and cohort random fluctuation at the same time as modelling an age trend, so long as there are no trends in the periods and cohorts. However, if we wish to analyse trends in periods and cohorts, we must make big assumptions about at least one the effects of of age, period and cohort – effectively we must constrain one of the three, usually to zero.

Here we consider the possibility of using Bayesian informative priors to implement this constraint. In doing so, we are able to consider the functioning of the MCMC estimator in running such a model, and the reasons why the HAPC model with diffuse priors produces the results that it does.

The paper proceeds as follows. We first outline Yang and Land's HAPC model and the problems with it. We then outline the simulations that we undertake to assess the affect of various informative priors on the results that are found. Finally, the implications of these results, in terms of practical uses with real data, are discussed.

2. The HAPC model

Yang and Land's method to identify an age-period-cohort model is to use a cross-classified multilevel model, treating age as a fixed effect and period and cohort groups as random effects (and as such contexts in which individuals are situated). Where the dependent variable is continuous, the model can be specified as:

$$y_{i(j_1j_2)} = \beta_{0j_1j_2} + \beta_1 Age_{i(j_1j_2)} + \beta_2 Age^2_{i(j_1j_2)} + e_{i(j_1j_2)}$$

$$\beta_{0j_1j_2} = \beta_0 + u_{1j_1} + u_{2j_2}$$

$$e_{i(j_1j_2)} \sim N(0, \sigma^2_e), \qquad u_{1j_1} \sim N(0, \sigma^2_{u1}), \qquad u_{2j_2} \sim N(0, \sigma^2_{u2})$$

(1)

Where $y_{i(j_1j_2)}$ is the dependent variable, for individual i in period j_1 and cohort j_2. β_1 and β_2 are the coefficients for the linear and quadratic age effects respectively, $\beta_{0j_1j_2}$ is the constant which varies across both periods and cohorts around a non-varying mean constant β_0, and $e_{i(j_1j_2)}$ is the level 1 residual error term. Periods, cohorts and level-1 residuals are all assumed to be Normally distributed, and their variances are estimated.

Yang and Land argue that modelling in this way, with age modelled differently from periods and cohorts and age specified as a quadratic polynomial, negates the problem of model identification encountered by APC analysts (Yang and Land 2006; Yang and Land 2013). However, a number of recent critiques have shown that this is not the case. Where there are trends in periods and/or cohorts, the HAPC model will arbitrarily assign this trend between age, period and cohort, leading to results that can be extremely misleading (Bell and Jones 2013a, c; Luo and Hodges 2013). However, the model does function relatively well when there are no trends in the periods or cohorts, or when one of those trends is assumed to be zero by including the other as a linear effect in the fixed part of the model (Bell and Jones 2013a).

3. HAPC models in a Bayesian framework

Even leaving aside the identification problem, researchers wishing to use the HAPC model face a number of methodological questions. For example it is often the case that there are not enough periods or cohort groups "to satisfy the large-sample criteria required" for maximum likelihood (ML) or restricted ML estimation (Yang and Land 2013, p. 263). As a superior alternative, Yang (2006) suggests using Bayesian estimation with MCMC estimation, since such models are not biased to the same extent when higher level sample sizes are small (Browne and Draper 2006; Stegmueller 2013).

Bayesian modelling additionally allows the possibility of using informative priors in modelling. This allows you to impose, with an appropriate distribution of uncertainty, your prior beliefs about a parameter, and these beliefs are combined with the likelihood of the data to produce a posterior distribution which produces what is an overall result, given the data and the prior distribution.

It is most common that, when using Bayesian methods, diffuse priors are used, which impose as little prior information as possible upon the result, with the intent that only the data influence the estimates. In contrast, when the researcher is very sure about the result of a particular parameter, they could impose a very strong prior which has almost no uncertainty around it. Where the researcher has some reason to believe a parameter takes a certain value (or wishes to combine the results of prior research into the model) they could use a less strong informative prior, which expresses a belief but with a level of certainty that reflects the uncertainty with which they hold their belief.

This could be of particular interest in the case of the HAPC model: where a number of possible APC combinations fit the data equally well, an informative prior could be used to draw the posterior distribution towards a particular solution, justified by a priori reasoning. For example, consider the case of obesity, modelled by Reither et al. (2009) and later critiqued by us (Bell and Jones 2013c). In that case, the model arbitrarily assigned the increasing rates of obesity to periods, when cohorts could just as easily have generated the same data. However, in the case of obesity, there are physiological and cultural reasons to believe that we know what the lifecourse (age) patterns of obesity are likely to be – i.e. that the probability of obesity will increase as individuals age, and then decline slightly in old age (Visscher et al. 2010). If this can be quantified in parameter values for β_1 in equation 1, then informative priors could be used to 'nudge' β_1 towards the correct solution. Once the age trend is constrained in this way, any trends in periods and cohorts will be identified, and will be estimated to fit with the age trend being modelled. Meanwhile, any random fluctuations in periods and cohorts can also be reliably modelled and interpreted.

Similar techniques could be used to constrain period or cohort linear trends. Our previous simulation exercises (Bell and Jones 2013a) have shown that including a cohort trend in the fixed part of the HAPC model allows both an age and cohort trend to be identified, but also assumes that there is no period trend since all degrees of freedom have been consumed by the age and cohort trends. Such a model is specified as follows:

$$y_{i(j_1 j_2)} = \beta_{0 j_1 j_2} + \beta_1 Age_{i(j_1 j_2)} + \beta_2 Age^2_{i(j_1 j_2)} + e_{i(j_1 j_2)}$$

$$\beta_{0 j_1 j_2} = \beta_0 + \beta_1 Cohort_{j_2} + u_{1 j_1} + u_{2 j_2}$$

(2)

In this case, an informative prior could be used to constrain the cohort trend to a certain value; any period and age trends would then be estimated given that constraint, in the period level residuals and the age linear effect estimates respectively.

Some words of warning at this point. This is not an automatic solution to the identification problem in any way, shape or form. The informative priors being imposed on the Bayesian model cannot be based on the data, since a multitude of possible APC solutions are equally statistically plausible given the data. The priors can only be based on theory, and therefore the results of that model are only as reliable as that theory. There are many situations where theory is not well established, or where the research question is itself aiming to find out the extent and nature of the age effect. In such cases, this model will not help the researcher find what they require.

The purpose of this article is to assess the practicalities of using such informative priors – in particular, how informative the priors need to be to function in the way required.

4. Simulations

Two data generating processes (DGPs) were used to generate data for use in these simulations:

$$y = 1 + (0.1 * Age) - (0.005 * Age^2) + (0.1 * Cohort) + u_c + u_p + e_{L1} \tag{3}$$

and

$$y = 1 + (0.1 * Age) - (0.005 * Age^2) + (0.1 * Cohort) - (0.005 * Cohort^2) + (0.1 * Year) - (0.005 * Year^2) + u_c + u_p + e_{L1} \tag{4}$$

Where $u_c \sim N(0,1)$ for cohorts, $u_p \sim N(0,1)$ for periods, and $e_{L1} \sim N(0,4)$ for individual, level 1 residuals, for data 'collected' between 1990 and 2010, for individuals aged 20-60. These data were fitted to the models in equation (1) and equation (2), with MCMC estimation (Browne 2009) using a range of priors: a diffuse (non-informative) prior, and two informative prior distributions of mean 0.1 (matching the DGP) with standard deviations of 0.01 and 0.001 (that is with moderate and high degrees of certainty respectively, described as weak and strong informative priors henceforth) on the age term for equation (1) and the cohort linear term for equation (2). Models were tested using both grouped (by 5-year intervals) and ungrouped (1-year intervals) cohort groups. For each simulation scenario, 100 datasets were generated and modelled. The models were run in MLwiN version 2.28 (Rasbash et al. 2013) using the runmlwin command (Leckie and Charlton 2013) in Stata[1]. The true parameter values from the DGP were used as MCMC starting values. A burn in of 500 iterations was followed by a monitored chain of 50,000 iterations. All models were checked for convergence and appropriate chain length by visual inspection of the chains for each parameter estimated, as well as with statistics such as the Effective Sample Size (ESS – see Table 1). None of our scenarios showed any signs of poor mixing or trending in the chains.

The results are shown in Figure 1 for the ungrouped scenario with model (1) and DGP (3), whilst the results from a single model run under this scenario are presented in Table 1. As can be seen, with diffuse priors, and in line with results found previously (Bell and Jones 2013a), the model finds an apparent trend in the periods that was not present in the DGP. For similar reasons, the cohort trend in the DGP is not found, whilst only the quadratic trend is found for age (the linear trend is estimated as 0 rather than 0.1). In short, when no informative prior is used, the HAPC model produces an answer that is incorrect given the known DGP.

[Table 1 and Figure 1 about here]

In contrast, including informative priors improves the results; however a very strong informative prior is required to push the posterior distribution away from the incorrect solution found by the model with diffuse priors, to the true value. A prior with a standard deviation of 0.01 (that is quite a

[1] Do-files for replicating these simulations can be found in the online appendix.

bit of certainty – our prior evidence means that we are 95% sure that the true parameter value lies between 0.08 and 0.12) produced a result approximately halfway between the truth (0.1) and the results with diffuse priors (zero). It takes a much stronger prior (here with a SD of just 0.001) to nudge the model results to the truth (see Figure 2) although incrementally lower SDs produce incrementally more accurate parameter estimates (see Figure 3). This is despite the Deviance information criteria - DIC (Spiegelhalter et al. 2002) – suggesting that the three models presented fit the data approximately as well as each other (although the DIC is lower in the diffuse prior case, the differences would not normally be considered significant enough to choose one model over the other). In other words, despite the various solutions presented in figure 1 apparently fitting the data equally well, the model still requires strong evidence to shift the results found towards the truth.

[Figures 2 and 3 about here]

That such strong priors are required is related to properties of the MCMC estimator under conditions of high collinearity. Browne et al. (2001, p. 113) show that, in the presence of very high degrees of collinearity in the fixed or the random effects, MCMC methods will identify this, by poor mixing being evident in the parameter chains. At slightly lower levels of collinearity, however, Wheeler and Calder (2006) show that MCMC models function in a similar way to ridge regression models, thus dealing well with quite high levels of collinearity. The HAPC model, by splitting the APC terms between the fixed and random parts, effectively tricks the model into behaving as it would under inexact rather than exact collinearity, producing apparently well behaved chains which require strong priors to be overcome, much like the ridge estimators produce relatively well behaved estimates. The difference is that there is exact collinearity between the variables in the DGP, and as such the apparent certainty of the parameter estimates is unwarranted.

The other scenarios correspond to the understanding articulated here and elsewhere. Grouping of the data does not affect the average effect, but does make the result more variable across the 100 simulations (see figure 4). This variation reduces as increasingly strong informative priors are used.

When modelling data where there are all of age, period and cohort linear effects in the DGP (as in equation 4), the model incorrectly produces only a large period trend, until informative priors are applied (see figure 5[2]). Finally, including a linear cohort term into the model, and imposing an informative prior on this linear effect, allows us to estimate the effects of period and age effectively (see figure 6), assuming, as ever, that our prior knowledge of the cohort trend is correct.

[Figures 4, 5 and 6 about here]

5. Discussion

The results presented here provide a possible option for those wishing to distinguish period and cohort effects when the age effect is known, or when any one of the APC trends are known and the effects of the other two are required to be estimated. However it also presents significant cautions to those wishing to do so. It is surprising how strong a prior is required to fix the APC effect appropriately, given that the HAPC model applies a given APC combination apparently arbitrarily. As such, a vague idea about the nature of the age trend is not enough to apply this method rigorously – the researcher must be *almost certain* that the age trend is as they believe and effectively constrain it to that trend. Unfortunately, this is only likely to occur when the researcher already has a very clear understanding of the processes involved, and completely trusts that the data is unbiased. Similarly, a researcher could use such a process to get a particular result that they wanted, which would obviously not be good science.

An alternative could be for researchers to model a number of different possible APC combinations (using a range of strong priors), thus presenting a range of possible outcomes which could explain the patterns in the data. This is in line with the idea of a "community of priors" (Spiegelhalter 2004, p. 159; Lunn et al. 2013, p. 97) expressing different possible prior beliefs that the researcher might

[2] It seems that, when the DGP with all of APC linear trends is used, a weaker prior is required to pull the result towards the truth. However this is not particularly valuable for applied research where the DGP is unknown.

have[3]. They could then use theoretical arguments to justify which scenario they consider to be the most plausible. It must be emphasized, however, that this justification cannot be based on the data but only on the theoretical beliefs of the researcher, and they should be presented as such. The APC identification problem cannot be solved by this method or any other technical legerdemain, and any research that claims to do so should be treated with an appropriate degree of skepticism.

Acknowledgements

Thanks to Bill Browne for his help and advice, and the anonymous reviewer for his/her suggestions – neither are responsible for what we have written.

References

Bell, A., Jones, K.: Another 'futile quest'? A simulation study of Yang and Land's Hierarchical Age-Period-Cohort model. Demographic Research **forthcoming** (2013a)

Bell, A., Jones, K.: Current practice in the modelling of Age, Period and Cohort effects with panel data: a commentary on Tawfik et al (2012), Clarke et al (2009), and McCulloch (2012). Quality and Quantity **in press** (2013b). doi:10.1007/s11135-013-9881-x

Bell, A., Jones, K.: 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 **in press** (2013c). doi:10.1016/j.socscimed.2013.09.004

Bell, A., Jones, K.: The impossibility of separating age, period and cohort effects. Social Science & Medicine **93**, 163-165 (2013d)

Browne, W.J.: MCMC estimation in MLwiN, Version 2.25. Centre for Multilevel Modelling, University of Bristol (2009)

Browne, W.J., Draper, D.: A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Anal **1**(3), 473-513 (2006)

[3] Kass and Greenhouse (1989) argue that "the beliefs we specify need not be our own, nor need they be the beliefs of any actual person we happen to know, nor derived in some way from any group of "experts"." As such, the purpose here is in understanding a range of possible beliefs and the sensitivity to these of the resultant posterior distribution, rather than imposing a prior distribution given the researchers' actual beliefs.

Browne, W.J., Goldstein, H., Rasbash, J.: Multiple membership multiple classification (MMMC) models. Statistical Modelling **1**, 103-124 (2001)

Glenn, N.D.: Cohort Analysts Futile Quest - Statistical Attempts to Separate Age, Period and Cohort Effects. Am Sociol Rev **41**(5), 900-904 (1976)

Glenn, N.D.: Cohort Analysis, 2nd ed ed. Sage, London (2005)

Kass, R.E., Greenhouse, J.B.: Comment: A Bayesian perspective. Stat Sci **4**, 310-317 (1989)

Leckie, G., Charlton, C.: runmlwin: A program to run the MLwiN multilevel modelling software from within Stata. J Stat Softw **52**(11) (2013)

Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D.: The BUGS Book: A Practical Introduction to Bayesian Analysis. CRC Press, Boca Raton, FL (2013)

Luo, L.: Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem. Demography **in press** (2013)

Luo, L., Hodges, J.: The cross-classified age-period-cohort model as a constrained estimator. Under review (2013)

Rasbash, J., Charlton, C., Browne, W.J., Healy, M., Cameron, B.: MLwiN version 2.28. In. Centre for Multilevel Modelling, University of Bristol, (2013)

Reither, E.N., Hauser, R.M., Yang, Y.: Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. Social Science & Medicine **69**(10), 1439-1448 (2009)

Spiegelhalter, D.J.: Incorporating Bayesian ideas into health-care evaluation. Stat Sci **19**(1), 156-174 (2004)

Spiegelhalter, D.J., Best, N.G., Carlin, B.R., van der Linde, A.: Bayesian measures of model complexity and fit. J Roy Stat Soc B **64**, 583-616 (2002)

Stegmueller, D.: How many countries do you need for multilevel modeling? A comparison of frequentist and Bayesian approaches. Am J Polit Sci **57**(3), 748-761 (2013)

Tu, Y.K., Smith, G.D., Gilthorpe, M.S.: A New Approach to Age-Period-Cohort Analysis Using Partial Least Squares Regression: The Trend in Blood Pressure in the Glasgow Alumni Cohort. Plos One **6**(4) (2011)

Visscher, T.L.S., Snijder, M.B., Seidell, J.C.: Epidemiology: definition and classification of obesity. In: Kopelman, P.G., Caterson, I.D., Dietz, W.H. (eds.) Clinical Obesity in Adults and Children. pp. 3-14. Blackwell, Oxford (2010)

Wheeler, D.C., Calder, C.A.: Bayesian spatially varying coefficient models in the presence of collinearity. ASA Section on Bayesian Statistical Science (2006)

Yang, Y.: Bayesian inference for hierarchical age-period-cohort models of repeated cross-section survey data. Sociological Methodology **36**, 39-74 (2006)

Yang, Y., Land, K.C.: 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 (2006)

Yang, Y., Land, K.C.: Age-Period-Cohort Analysis: New models, methods, and empirical applications. CRC Press, Boca Raton, FL (2013)

Yang, Y., Schulhofer-Wohl, S., Fu, W.J.J., Land, K.C.: The intrinsic estimator for age-period-cohort analysis: What it is and how to use it. Am J Sociol **113**(6), 1697-1736 (2008)

Figure 1: Results for age, period and cohort effects (rows 1, 2 and 3 respectively) found from HAPC models using a diffuse prior, a 'weak' informative prior, and a 'strong' informative prior, when fitted with the model in equation (1), to 100 simulated datasets generated by the DGP in equation (3). Cohorts are assumed to be ungrouped in the model. The dashed lines represent the true trends.

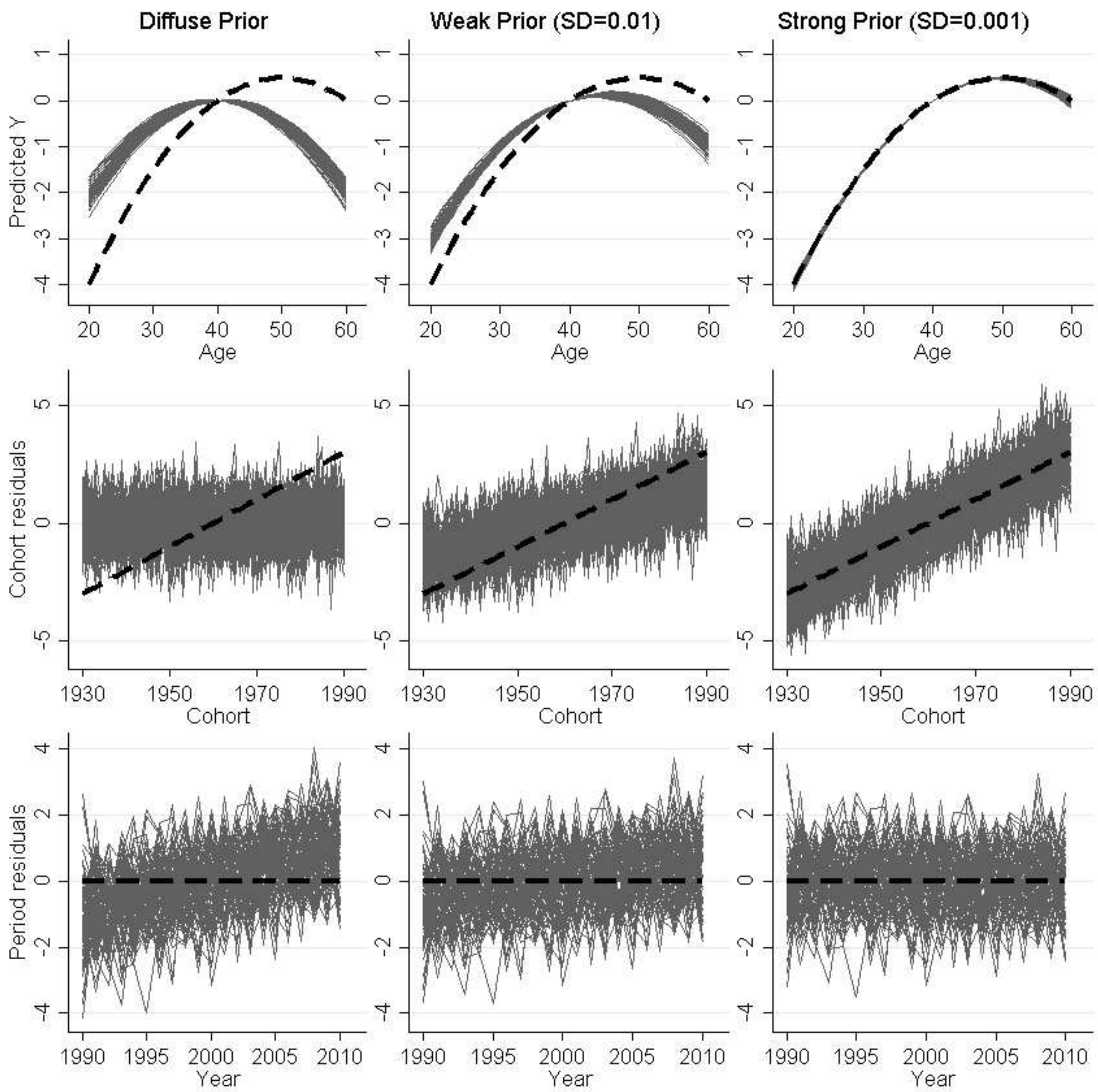

Figure 2: Prior (dotted line) and posterior (solid line) distributions for the age parameter, where age is specified with (a) a diffuse prior, (b) a 'weak' informative prior, and (c) a 'strong' informative prior, for a single dataset, fit with the model in equation (1), with DGP as in equation (3).

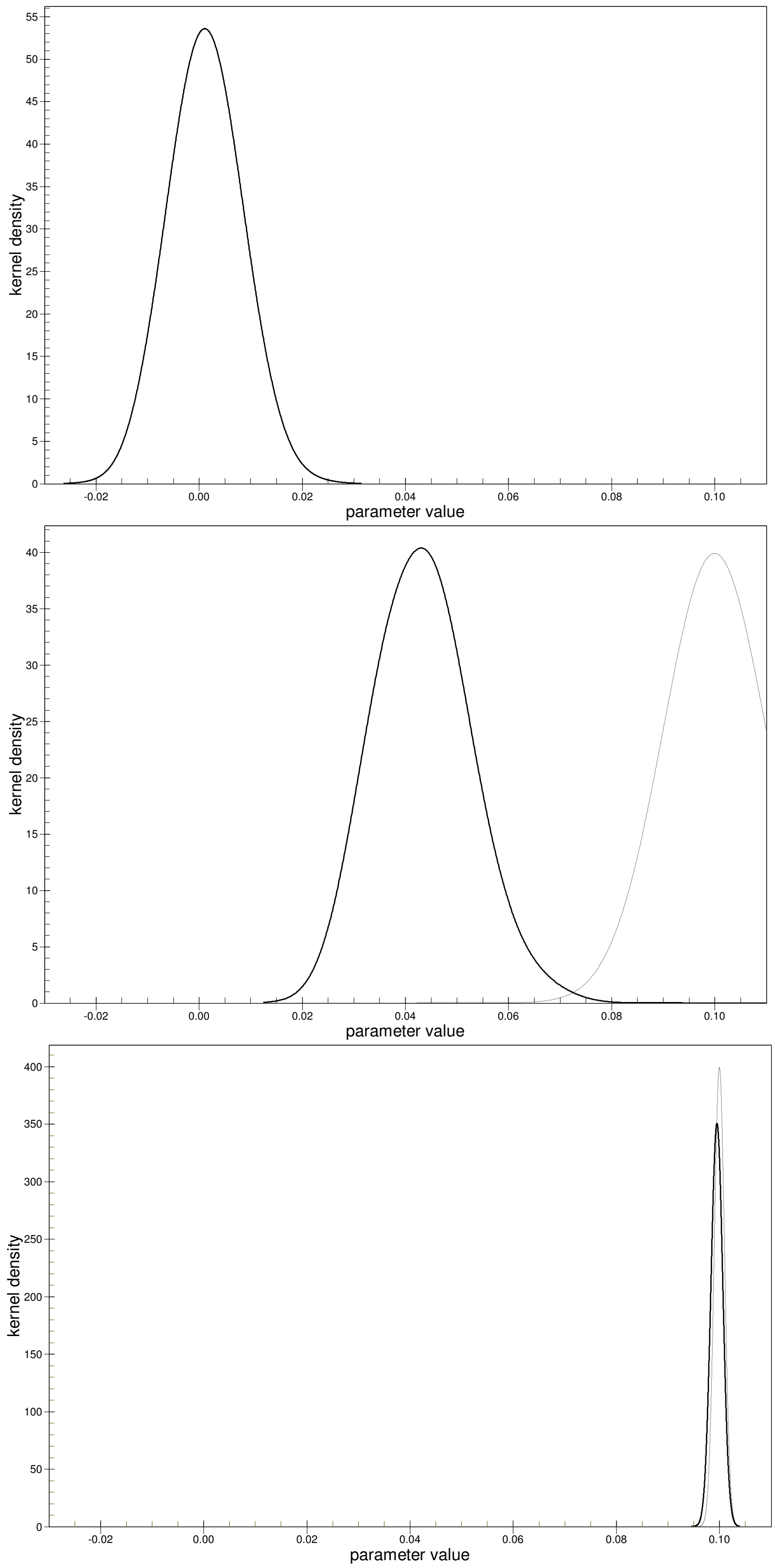

Figure 3: The effect of the SD of the prior distribution on the parameter estimate, for a single dataset, fit with the model in equation (1), with DGP as in equation (3). The true parameter value is 0.1.

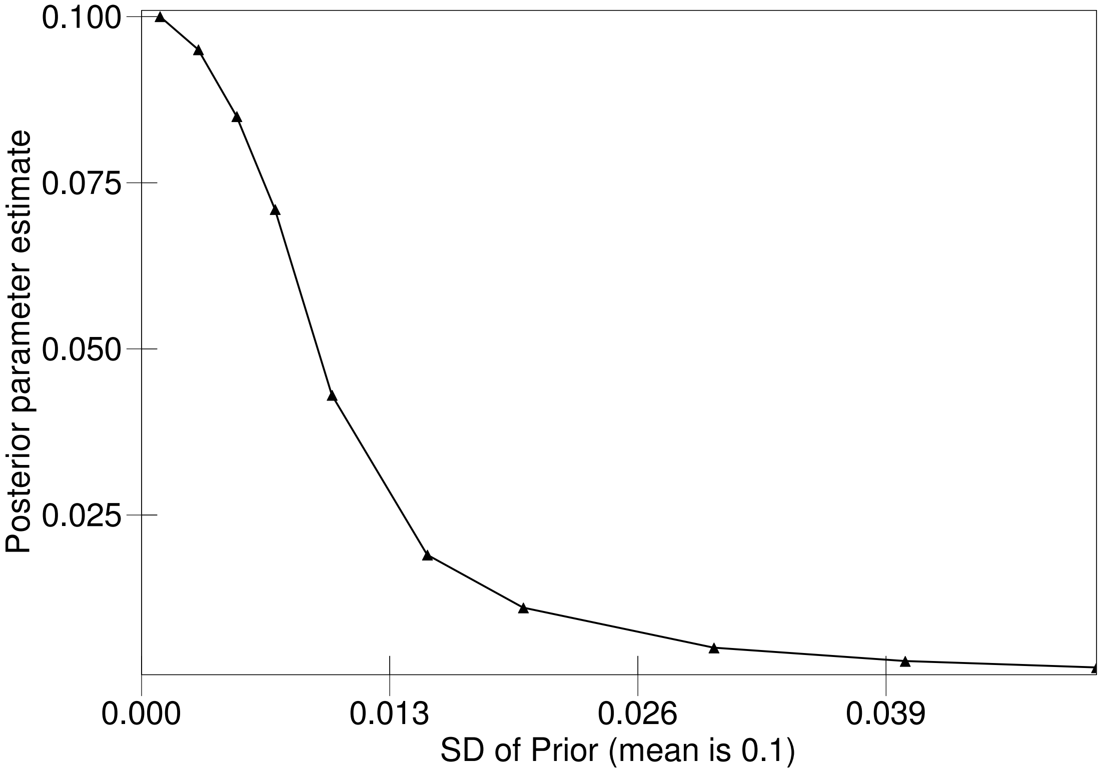

Figure 4: Results for age, period and cohort effects (rows 1, 2 and 3 respectively) found from HAPC models using a diffuse prior, a ‘weak’ informative prior, and a ‘strong’ informative prior, when fitted with the model in equation (1), to 100 simulated datasets generated by the DGP in equation (3). Cohorts are assumed to be grouped in 5-year intervals in the model. The dashed lines represent the true trends.

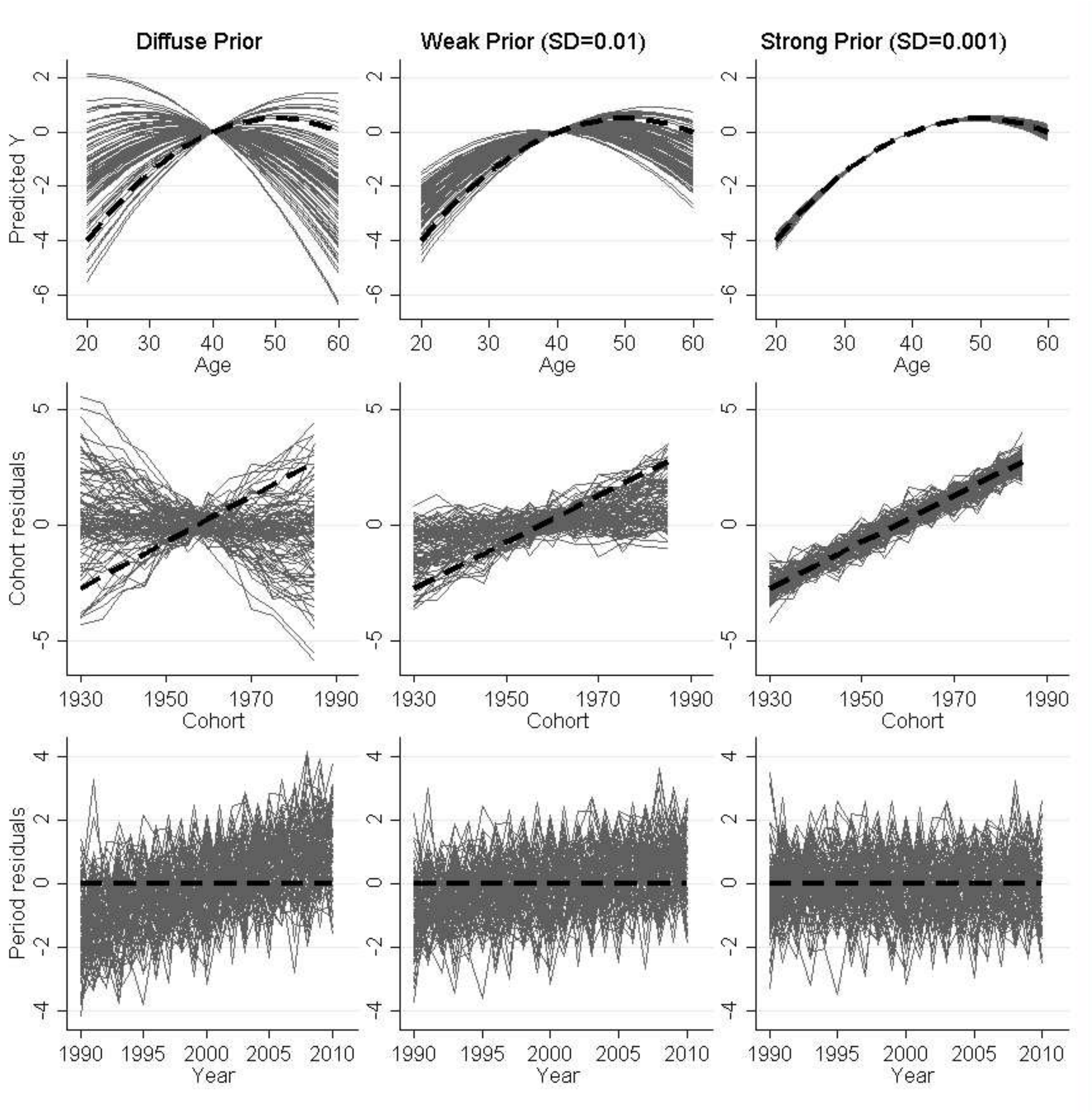

Figure 5: Results for age, period and cohort effects (rows 1, 2 and 3 respectively) found from HAPC models using a diffuse prior, a 'weak' informative prior, and a 'strong' informative prior, when fitted with the model in equation (1), to 100 simulated datasets generated by the DGP in equation (4). Cohorts are assumed to be ungrouped in the model. The dashed lines represent the true trends.

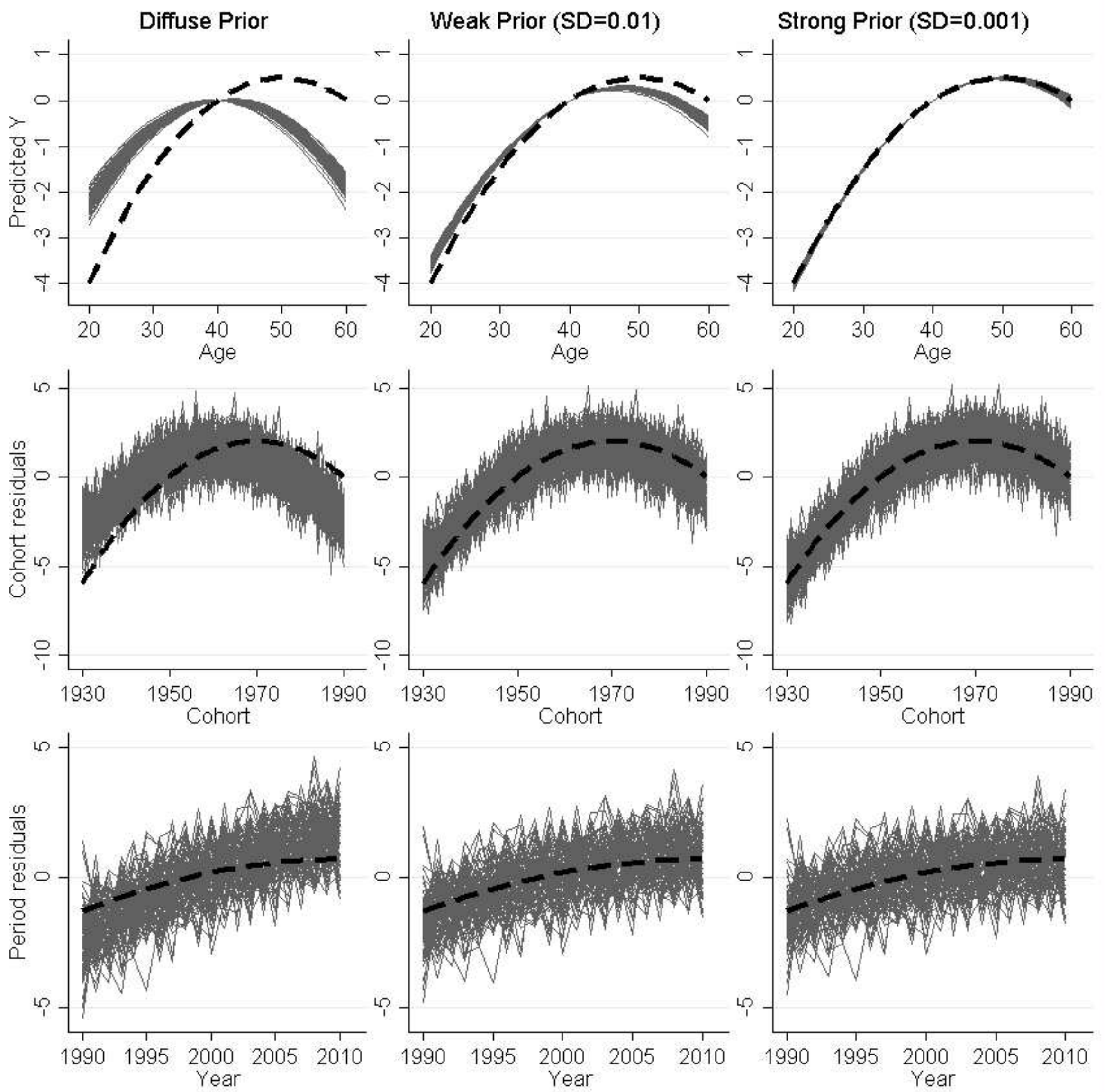

Figure 6: Results for age, period and cohort effects (rows 1, 2 and 3 respectively) found from HAPC models using a diffuse prior, a 'weak' informative prior, and a 'strong' informative prior, when fitted with the model in equation (2), to 100 simulated datasets generated by the DGP in equation (4). Cohorts are assumed to be ungrouped in the model. The dashed lines represent the true trends.

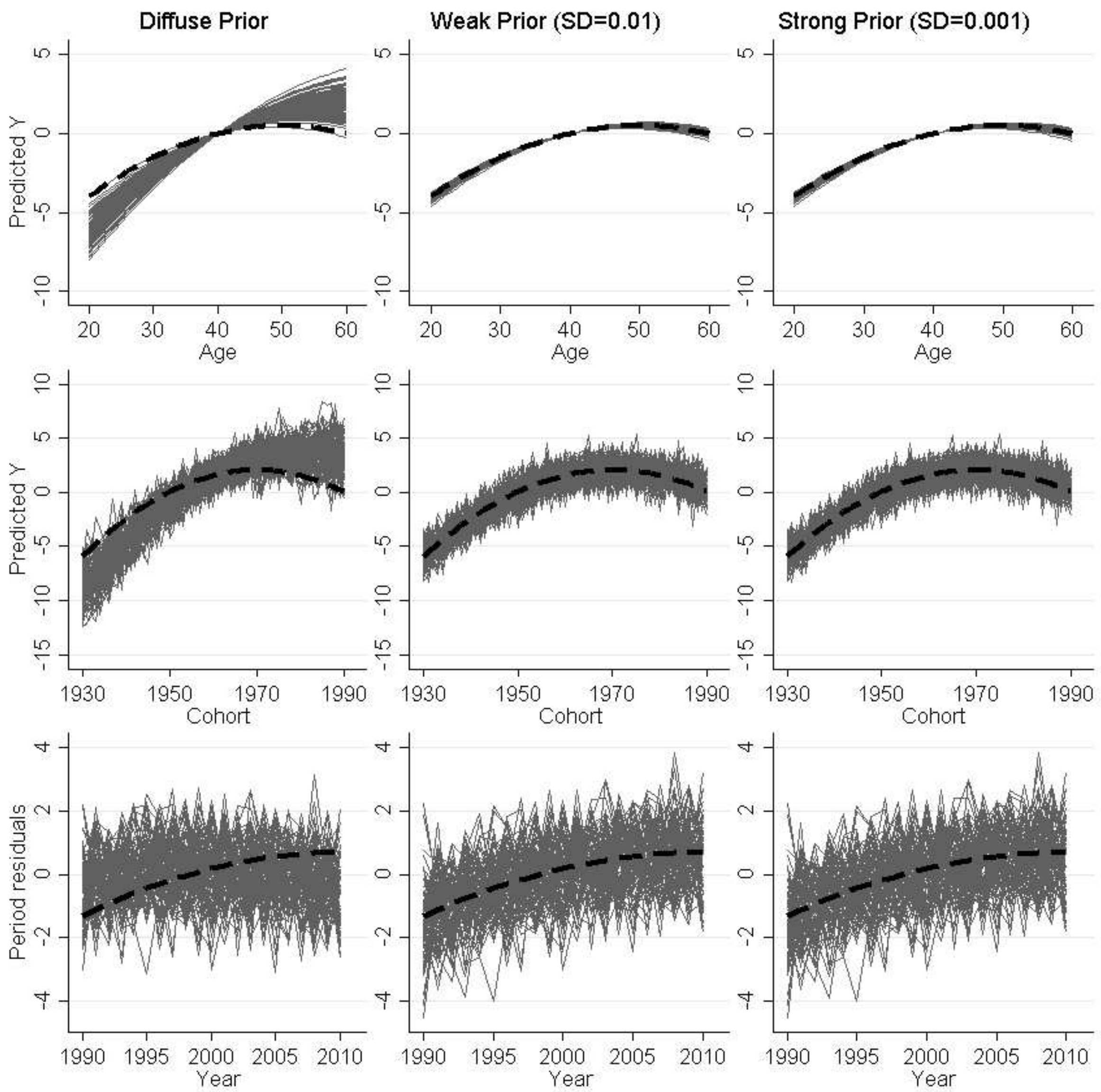

Table 1: Parameter estimates from HAPC models using a diffuse prior, a weak informative prior, and a strong informative prior, when fitted to a single simulated dataset. Deviance, effective number of parameters (pD) and DIC (which effectively combines the two) are also shown.

	Diffuse priors					Weak informative prior (SD=0.01)					Strong informative prior (SD=0.001)				
	β	S.E.	CIs		ESS	β	S.E.	CIs		ESS	β	S.E.	CIs		ESS
Fixed Part															
cons	1.326	0.343	0.635	2.044	55	1.321	0.349	0.614	2.041	57	1.322	0.397	0.511	2.121	50
(age-gm)^1	0.001	0.007	-0.011	0.015	740	0.043	0.009	0.027	0.062	390	0.100	0.001	0.098	0.101	18258
(age-gm)^2	-0.005	0.000	-0.005	-0.005	19247	-0.005	0.000	-0.005	-0.005	19089	-0.005	0.000	-0.005	-0.005	18957
Random Part															
Cohort Level															
Cons	0.785	0.153	0.540	1.142	35207	1.388	0.378	0.824	2.291	877	3.97	0.761	2.746	5.717	39758
Period Level															
Cons	2.075	0.741	1.090	3.896	29902	1.845	0.666	0.969	3.474	27496	1.748	0.626	0.92	3.275	29814
Level 1															
Cons	4.014	0.040	3.935	4.094	48181	4.014	0.040	3.936	4.094	48160	4.014	0.04	3.936	4.094	48153
DIC:			84631.432					84633.154					84633.7		
pD:			81.332					82.010					82.695		
Deviance			84550.10					84551.14					84551.00		