Flexible Latent Variable Methods for Panel Data

Igor Balnozan

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Economics

UNSW Business School

Supervisors:

Scientia Professor Robert Kohn

Professor Denzil G. Fiebig

Professor Scott A. Sisson

February 2022

Contents

Acknowledgements

Thank you Robert, Denzil and Scott, for your guidance over these four years. You provided the support I needed to keep progressing despite the challenges that came.

Thanks also to Anthony Asher, for your feedback and comments on the superannuation applications.

Thanks too, in no particular order, to David Gunawan, Matias Quiroz, Rob Salomone, Jay Lee and Tim Neal, for helpful discussions about methodology.

I am grateful for the support provided by the Commonwealth Government of Australia through the provision of an Australian Government Research Training Program Scholarship, and to State Super for provision of the State Super Academic Scholarship.

I was partially supported by the Australian Research Council through the Australian Centre of Excellence in Mathematical and Statistical Frontiers (ACEMS; CE140100049).

I thank Plan For Life, Actuaries & Researchers, who collected, cleaned and allowed me to analyse the superannuation data used in this research; this data capture forms one part of a broader survey into retirement incomes, commissioned by the Institute of Actuaries of Australia.

This research includes computations using the computational cluster Katana supported by Research Technology Services at UNSW Sydney. Katana DOI: 10.26190/669x-a286

This thesis uses unit record data from Household, Income and Labour Dynamics in Australia Survey [HILDA] conducted by the Australian Government Department of Social Services (DSS). The findings and views reported in this thesis, however, are mine and should not be attributed to the Australian Government, DSS, or any of DSS' contractors or partners. HILDA DOI: 10.26193/IYBXHM

List of Tables

List of Figures

Chapter 1

Introduction

This thesis develops and applies methods for estimating linear panel data models that control for both time-invariant and time-varying heterogeneity. The models considered are motivated from an empirical microeconomics perspective, where a researcher or practitioner is interested in the behavioural effects on individuals of unobservable preferences or external shocks, and where the effects are allowed to vary across individuals. Specifically, the methods are tailored for panel datasets that track a large number of observational units over a short to moderate number of successive response occasions.

Time-varying heterogeneity in preferences across individuals can be an important determinant of measured outcomes in microeconomic panels; however, the source of this heterogeneity may typically be unobserved in data. Latent preference heterogeneity may also be an obstacle to precise estimation of the partial effects of observed covariates, as failure to control for this heterogeneity can create an endogeneity problem in microeconomic models. Addressing both of these concerns by including a time-varying heterogeneity component in microeconomic panel data models can reveal economic insights derived from interpretations of the estimated heterogeneity parameters, while improving inference on covariate effects of interest.

In this thesis, practical estimation of the time-varying heterogeneity component of the models depends on classifying observational units into distinct groups, where group membership is unobserved and incorporated into the models using latent variables. The

most general model considered allows for separate group structures in both the covariate effects and the time-varying heterogeneity. The other models studied are special cases of this general specification, obtained by constraining the groupings in the covariate effects and time profiles to be the same, or by enforcing homogeneity in the covariate effects across all units in the sample. This latter model is the one considered by Bonhomme and Manresa (2015) for the Grouped Fixed-Effects (GFE) estimator; further assuming that all units follow the same set of time profiles recovers a simpler, textbook panel data model, the two-way fixed effects model, defined in Chapter 2.

The three main methodological contributions of this thesis are: a) proposing extensions to the GFE estimator that tailor it for use in analysing typical observational microeconomic datasets; b) proposing nonparametric Bayesian estimation procedures for models with grouped time-varying heterogeneity and individual-specific intercepts, and developing corresponding inference methods for the group-specific parameters; c) using simulation studies to validate the proposed approaches and to directly compare them on the same synthetic data with known generating processes.

The main contribution to the applied microeconomics literature is a novel analysis of retirement decumulation behaviours in Australian superannuation funds using all methods considered in this thesis. Another applied contribution is a preliminary analysis of different smoking behaviours in Australia during a period of significant government intervention through smoking policy.

Chapter 2 provides the methodological background and a review of closely related econometrics literature. Chapter 3 demonstrates how an existing classical method, the GFE estimator from Bonhomme and Manresa (2015), can, with some extensions, be repurposed for a variety of microeconomics applications. Chapters 4 and 5 extend an existing Bayesian analog to the GFE estimator, adding the ability to control for individual fixed effects, and group-specific covariate effects, respectively. Chapter 6 concludes with a summary and discussion of the key implications of the results in this thesis. Appendices A, B and C include supplementary material for Chapters 3, 4 and 5, respectively.

Chapter 2

Background Literature

This chapter puts the results of Chapters 3 to 5 in the context of the current literature. As these chapters are written in the format of standalone papers suitable for submission to academic journals, all three also incorporate self-contained reviews of the relevant literature.

2.1 Classical linear panel data models

A simple linear panel data model is the standard 'fixed-effects' (FE) model,

$$y_{it} = c_i + x'_{it}\theta + v_{it}, \tag{2.1}$$

where: $i = 1, 2, ..., N$ indexes individuals, or other types of observational units; $t = 1, 2, ..., T$ indexes time periods, or other forms of successive response opportunities; y_{it} is the value of the dependent variable of interest; the c_i are individual-specific intercepts, or 'individual fixed effects'; x_{it} is a length-K column vector of observed covariate values; θ is length-K column vector containing the corresponding partial effects of the covariates; v_{it} is the model error. (See, e.g., Wooldridge, 2012, pp. 459–460, 485.)

By including time dummy variables as additional covariates in (2.1), FE estimation also permits identifying a single set of time-varying intercepts (a 'time profile'), α_t, as in the

two-way fixed-effects (2WFE) model,

$$y_{it} = c_i + \alpha_t + x'_{it}\theta + v_{it}. \tag{2.2}$$

If there is no need to control for N individual-specific intercepts, c_i, then a variant of this model with only time fixed effects is

$$y_{it} = \alpha_t + x'_{it}\theta + v_{it}. \tag{2.3}$$

Bonhomme and Manresa (2015) present the Grouped Fixed-Effects (GFE) estimator, which applies to a class of panel data models for which the most basic specification can be viewed as an extension of (2.3) that allows for multiple sets of time profiles, with distinct time profiles corresponding to different latent groups in the sample. This most basic specification is

$$y_{it} = \alpha_{g_i t} + x'_{it}\theta + v_{it}, \tag{2.4}$$

where $g_i \in \{1, 2, ..., G\}$ is a group assignment variable: if unit i belongs to group $g_i = g$, then its time profile value in period t is α_{gt}. Bonhomme and Manresa (2015) show that their estimator can also be used for a more general version of their model that includes individual-specific intercepts:

$$y_{it} = c_i + \alpha_{g_i t} + x'_{it}\theta + v_{it}. \tag{2.5}$$

Equation (2.5) generalises (2.2): the c_i term controls for heterogeneity in the overall level of the dependent variable, y_{it}, across units, while the $\alpha_{g_i t}$ term models shapes in the heterogeneity over time, which are allowed to vary across individuals. Parsimony is obtained by forcing a grouping structure onto the time profiles, so that the infeasible task of estimating N time profiles, $\alpha_i := (\alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,T})'$, is replaced with the feasible alternative of estimating $G \ll N$ time profiles, $\alpha_g := (\alpha_{g,1}, \alpha_{g,2}, ..., \alpha_{g,T})'$.

A motivation for including the fixed effects, c_i, in panel data models exists if the number of individuals in the panel is large, and there is likely to be substantial heterogeneity across units in the overall levels of the dependent variable, even after accounting for time profiles and the partial effects of observable covariates. This scenario may occur

when modelling observational microeconomic data, where there may be a large source of variation arising from the unobservables. Ignoring the c_i term in panel data models leads to bias in the θ estimates if the c_i parameters are correlated with the observable covariates (Wooldridge, 2012, p. 460), which is the conservative assumption in the absence of experimental controls for the unobservable heterogeneity.

Similarly, including the $\alpha_{g_i t}$ term in the model aids in the general case where the unobservable time-varying heterogeneity is correlated with the observable covariates, removing another potential source of bias in the θ estimates. Including the $\alpha_{g_i t}$ term also permits examining the economic interpretations of the time profiles, which, in the class of applications considered in this thesis, is the primary research aim.

The GFE estimator takes as input a value for the number of latent groups, G. Thus, in the general case where this number is unknown to the researcher, implementing the GFE approach requires estimating the model for multiple values of G, and then performing model selection to estimate the true number of groups, G^0. In their online supplement, Bonhomme and Manresa (2015) suggest an information criterion (IC) to select the number of groups, and show that it recovers the correct number of groups when N and T are both large and of comparable magnitude, but generally overestimates G^0 when T is much smaller than N. Observational microeconomic panels often have a large number of individuals and a small or moderate number of time periods. The application in Bonhomme and Manresa (2015) does not satisfy the requirements for accurate estimation of G using their IC—their dataset has $N = 90$ and $T = 7$—so they instead propose a heuristic approach to estimating G^0 which relies on two theoretical results: a) when $G < G^0$, the GFE estimator is, in general, an inaccurate estimate of θ; b) when $G > G^0$, the estimate of θ is likely to be close to the true value, but the time profiles may be biased (Bonhomme & Manresa, 2015, p. 13 of online supplement). Their heuristic approach plots the covariate effect estimates conditional on different input values of G, and selects a point at which the estimates appear to stabilise. Using this method they choose a value of $G = 4$ for their application, which, as expected, is smaller than the corresponding $G = 10$ estimated by the IC. Zhang (2020) considers alternative ICs, but shows using a simulation study that they do not perform well.

While consistency and asymptotic normality of the GFE estimator are derived using large-N large-T asymptotic results, crucially, T is allowed to grow more slowly than N. Hence, the estimators have good finite-T performance, even with small to moderate T. Good performance for $T = 7$ is shown in the simulation studies given in the online supplement to Bonhomme and Manresa (2015). Their online supplement presents closed-form estimators for the variance of the GFE estimators, including a fixed-T consistent estimator for inference with finite T. The authors also argue that bootstrap inference is valid in this context, where bootstrap resampling is for the raw data, $\{y_{it}, x_{it}\}$, done with replacement at the unit level.

An extension of their basic specification which Bonhomme and Manresa (2015) explore allows for the covariate effects to vary across the groups in the time profiles. However, the heuristic approach to estimating G outlined above is not applicable in such a model; moreover, Chapter 5 shows that constraining groupings in the covariate effects and time profiles in this way is problematic, and presents a more flexible specification addressing this issue.

Related work by Bonhomme, Lamadon, and Manresa (2019) and Bonhomme, Lamadon, and Manresa (in press) highlights a potential limitation of the GFE estimator arising from their implicit assumption that there are a finite number of groups in the population; i.e., grouping units is not only a convenient way to create a parsimonious specification, but accurately represents the underlying generating process. Using a two-step estimator that first classifies the sample into a finite number of groups, then estimates parameters conditional on these groups, Bonhomme et al. (in press) show that approximating a continuous distribution in the unobservable heterogeneity using a finite number of groups may introduce bias into the estimators, and they explore potential bias-reduction methods in this context.

In contrast to the GFE estimator, factor-model approaches to capturing unobservable heterogeneity use more flexible specifications, at the cost of poorer finite-sample results on short- to moderate-T panels. These more flexible models assume that the unobservable heterogeneity has a factor structure, where a finite number of latent factors exist, and can have heterogeneous effects on different units in the panel. Bai (2009) formulates a factor-model framework for analysing panel data, called the

'interactive fixed-effects' (IFE) model, and gives several examples of relationships that can be specified using the IFE model in microeconomics, macroeconomics, and finance. The IFE approach specifies the relationship between the dependent variable, covariates and some latent factors as $y_{it} = x_{it}'\theta + \lambda_i' f_t + v_{it}$, where f_t is a column vector of common, latent, time-varying factor values, and λ_i is a corresponding column vector of individual factor sensitivities. Within this framework, the 2WFE model, (2.2), is a special case: with $f_t = (1, \alpha_t)$ and $\lambda_i = (c_i, 1)$, the model reduces to $y_{it} = x_{it}'\theta + c_i + \alpha_t + v_{it}$. Bonhomme and Manresa (2015) show that the model they consider for the GFE estimator is another special case of an IFE model.

Bai (2009) discusses the assumptions required for consistent estimation of the IFE model parameters. Crucially, the Bai (2009) estimator only has desirable properties when both N and T are large, which is restrictive for our applications. Indeed, Bonhomme and Manresa (2015) claim that the small-T bias in the estimator from Bai (2009) is large compared to the GFE estimator. The models studied by Ando and Bai (2016) and Su and Ju (2018) generalise the model from Bai (2009) and allow for latent groups in the sample with group-specific covariate effects—θ_{g_i} instead of θ—and use regularisation in the estimation procedure to enforce a parsimonious model structure. However, similarly to Bai (2009), favourable estimator performance for both of these more general methods is only proven under a large-N, large-T regime: finite-sample simulation results from Su and Ju (2018) suggest that $T \geq 50$ is necessary for satisfactory performance; the smallest T studied using simulation by Ando and Bai (2016) is $T = 100$. The model studied by Bonhomme and Manresa (2015) can be considered a special case of these factor models; however, for the class of applications studied in this thesis, factor models have a larger level of complexity than necessary to identify the economic effects of interest, and worse finite sample results for small to moderate T.

Finite mixture models are an alternative way to characterise unobserved heterogeneity across groups within a population. These models specify that the data observed for distinct groups may be generated by different distributions within some parametric family, while allowing the group identities of individuals to be unknown. Until recently, panel data mixture models were limited to pooled and 'random-effects' (RE) models

(Deb & Trivedi, 2013), which either assume away unobservable heterogeneity, or constrain it to be uncorrelated with observed regressors, respectively. Expanding this frontier, Deb and Trivedi (2013) introduce finite mixture models with fixed effects (FE-FMM). Their models are mixtures of Normal and mixtures of Poisson random variables, using simulation results to demonstrate estimator consistency for fixed T and large N. A limitation of their method is that it does not incorporate time-varying heterogeneity, and therefore approaches such as Bonhomme and Manresa (2015) are preferred for our applications.

A variant of finite mixture models for panel data are trajectory models, which model trajectories in the dependent variable as parametric functions of time: e.g., polynomials or splines in the time variable. Nagin and Tremblay (2001) describe univariate group-based trajectory modelling (GBTM) as using a mixture model to capture multiple latent groups, estimating distinct trajectories of the dependent variable of interest for each group, as well as the probability of membership in each group. These authors present a bivariate extension to univariate GBTM; however, Nagin, Jones, Passos, and Tremblay (2018) argue that adding more dependent variables quickly makes the estimation intractable.

Nagin et al. (2018) present a tractable extension of univariate GBTM to the multivariate setting, where interest is in the trajectory of multiple dependent variables, rather than one, while Ng and McLachlan (2014) consider a case where there is time-constant heterogeneity with a hierarchical (multi-level) structure: e.g., individuals (level 1) within hospitals (level 2), observed over time. There are no covariates included in the within-component distributions, which are polynomial functions of time; however, across components, they incorporate covariates as risk factors that determine which component the individual is more likely to be drawn from, similar to the gating mechanism in a mixture-of-experts model (e.g., Jacobs, Jordan, Nowlan, & Hinton, 1991; Jordan & Jacobs, 1994).

Rather than modelling individual data points, as is the case in many traditional statistical methods, functional data analysis treats a random function as the object of interest, and collectively views a set of observations as a realisation from an underlying random function (Hall, Müller, & Wang, 2006). Hall et al. (2006) aim to bridge the gap

between functional data analysis and longitudinal data analysis by showing that the two approaches can be reconciled for certain types of underlying data: specifically, when the intervals between observations are random and irregular. However, microeconomic panels commonly measure data at repeated, fixed intervals (e.g., the HILDA data considered in Chapter 5), which violates this irregularity condition.

A review of the classical panel data modelling literature suggests the GFE estimator as the appropriate starting point for this thesis, given the aim of developing methods for inference on time-varying heterogeneity after controlling for covariate effects, to be used in microeconomic applications with observational panel data. Chapter 3 identifies this area as a gap in the literature and demonstrates how the GFE estimator may be adapted to suit such applications.

2.2 Bayesian linear panel data models

Consider a parameter estimation problem in which: a) data, y_i, for $i = 1, 2, ..., N$, is assumed to be generated according to some distribution, $y_i \,|\, \theta \sim F(y_i; \theta)$, which depends on unknown parameters, θ; b) existing knowledge, or lack thereof, about the possible values of θ can be encoded into a 'prior' distribution. Bayesian methods can perform statistical inference on the unknown parameters, θ, taking into account any prior knowledge and the likelihood of observed data, following Bayes' theorem (see, e.g., Gelman et al., 2013, p. 7): $p(\theta \,|\, y) = p(y \,|\, \theta)\, p(\theta) \,/\, p(y)$; the data is $y := \{y_1, y_2, ..., y_N\}$; $p(\theta)$ is the prior density or mass function of θ; $p(y \,|\, \theta)$ is the likelihood of the dataset given values for θ; $p(\theta \,|\, y)$ is the 'posterior' density of the unknown parameters; $p(y) = \int p(y \,|\, \theta)\, p(\theta)\, \mathrm{d}\theta$ is called the marginal likelihood.

The posterior distribution, $p(\theta \,|\, y)$, is often only known up to a constant of proportionality, i.e., $p(\theta \,|\, y) \propto p(y \,|\, \theta)\, p(\theta)$, when the marginal likelihood is intractable. One way to proceed with posterior inference in this case is to use Markov chain Monte Carlo (MCMC; see, e.g., Gelman et al., 2013, p. 275). MCMC generates a sequence of iterates, $\left(\theta^{(1)}, \theta^{(2)}, ...\right)$, that, after a warm-up phase, behave like a sample drawn from the true posterior. Posterior inference then proceeds by estimating posterior moments

using sample statistics computed on the drawn values: e.g., the posterior mean can be estimated using the sample mean of the realised values of the chain.

The Gibbs sampler is an approach to posterior simulation when θ is multidimensional (see, e.g., Gelman et al., 2013, pp. 276–278). In Gibbs sampling, the parameters are separated into non-overlapping 'blocks', then sampled by alternating between the blocks' respective conditional posterior distributions. For example, if θ is blocked as $\theta = \{\theta_A, \theta_B\}$, then the Gibbs sampler alternates between: a) generating θ_A with conditional density $p(\theta_A \mid y, \theta_B)$; b) generating θ_B with conditional density $p(\theta_B \mid y, \theta_A)$. A Gibbs sampler is straightforward to implement when the conditional posteriors for all parameter blocks are given by distributions that are simple to sample from. Given some starting values, the sampler cycles through each of the parameter blocks, generating the next set of values in the chain for each block conditional on the most recently drawn values for that block and all the other blocks.

Another MCMC sampling technique is the Metropolis–Hastings (MH; Hastings, 1970) algorithm. Given a target distribution's density function, $\pi(\theta)$, which is otherwise difficult to sample from, a chain of values for θ that behave like a sample from the target distribution can be generated by first proposing a value, θ_{prop}, given some current value, θ_{curr}, drawn from some known proposal distribution, which is easy to directly sample from, with density function $q(\theta_{curr} \to \theta_{prop})$; second, applying an accept–reject step to determine whether the next value in the chain will take the newly proposed value or repeat the current value. The acceptance probability that ensures this chain has the intended behaviour is given by $\text{Prob}(\theta_{curr} \to \theta_{prop}) := \min\left\{1, \frac{\pi(\theta_{prop})q(\theta_{prop} \to \theta_{curr})}{\pi(\theta_{curr})q(\theta_{curr} \to \theta_{prop})}\right\}$, which can be evaluated even if the target distribution's density function, π, is only known up to a constant of proportionality. A MH sampler can be constructed for the entire parameter vector, θ; however, it can also be incorporated as a step within a Gibbs sampler. For example, if all parameter blocks except one have conditional posterior distributions that can be easily sampled from, then that one parameter block can be sampled using the MH approach applied to its conditional posterior as the target, while all other blocks are simulated directly from their conditional posteriors.

Lancaster (2002) discusses fixed-T estimation of panel data models with individual-specific intercepts. Notably, Bayesian estimation of linear panel data models with individual fixed effects, such as (2.1) or (2.2), after algebraically removing the fixed effects, gives asymptotically identical inference to the classical FE estimator as N grows large with T fixed, given suitable priors. An alternative approach, which places a hierarchical prior on the individual effects, gives a model analogous to the classical random-effects (RE) model (Koop, 2003, p. 152). Placing a hierarchical prior on a variable in a Bayesian model means treating parameters of that variable's prior distribution to be random, rather than fixed; i.e., the prior distribution parameters have their own prior distributions too, instead of taking some fixed values specified by the researcher. If there is reason to believe the individual-specific intercepts may be correlated with the observed covariates, which is the conservative assumption to use with observational panel data, the classical FE approach is generally preferred over the RE approach (Wooldridge, 2012, pp. 492, 495). Rendon (2013) also addresses the relationship between classical and Bayesian approaches to estimating simple linear panel data models, in a restricted scenario with known error term variance. The implications of the results presented there are that any differences in inference based on different linear panel data estimators arise from an explicit or implicit difference in prior assumptions.

Kim and Wang (2019) develop a Bayesian analog to the GFE estimator for (2.4), using a Dirichlet Process Mixture Model (DPMM; see, e.g., Escobar & West, 1995; Ferguson, 1983) to simultaneously allow estimation of unknown group memberships and unknown group-specific parameters. The Dirichlet Process (DP) is introduced by Ferguson (1973) as a random process whose realisations are discrete probability distributions with infinite support.

Practical methods exist to generate a random sample of data using a DP, e.g. Blackwell and MacQueen (1973), avoiding the problem of evaluating an infinite number of probability masses which are needed to explicitly specify a realisation of P; however, in the context of DPMMs, a convenient prior specification for the discrete component of the model uses the 'stick-breaking' construction of the DP (see, e.g., Sethuraman, 1994). Figuratively, the stick-breaking prior divides a unit-length interval into infinitely many

segments, attributing each segment length to a group membership probability for one of an infinite number of potential groups, with group-specific parameters drawn from the base distribution, H_0. The stick-breaking prior is useful because it is an alternative to approaches that integrate the random distribution P out of the joint posterior before estimation; integrating out P prevents identification of the group labels in iterations, and cannot be considered for models that do not admit analytical integration over the probabilities, which limits the flexibility of DPMMs (Papaspiliopoulos & Roberts, 2008).

Some approaches to stick-breaking DPMMs require truncating the total number of stick lengths to some fixed value, as in Ishwaran and Zarepour (2000); other approaches allow for a potentially infinite number of groups using a finite, but variable, set of groups: e.g., the slice samplers from Walker (2007) and Kalli, Griffin, and Walker (2011), or the retrospective sampler from Papaspiliopoulos and Roberts (2008). Comparing the latter two approaches, a slice sampler is analytically simpler to implement, and involves the simulation of fewer auxiliary variables in each iteration (Kalli et al., 2011, p. 95).

Kim and Wang (2019) adopt the slice-sampling approach in their DPMM estimation of (2.4), incorporating the two label-switching steps proposed by Papaspiliopoulos and Roberts (2008) to improve mixing across multiple modes of the group membership posterior probabilities; however, they do not use a third label-switching step, proposed by Hastie, Liverani, and Richardson (2015), which is designed to complement the two from Papaspiliopoulos and Roberts (2008). As advocated by Escobar and West (1995), Kim and Wang (2019) allow the DP prior concentration parameter, β, to be random, rather than fixed.

Critically, the DPMM in Kim and Wang (2019) does not control for individual fixed effects, which, as discussed in Section 2.1, would generally prevent its application to observational microeconomic data. Chapters 4 and 5 identify this area as a gap in the literature and formulate extensions of Kim and Wang (2019) to address this limitation.

Chapter 3

Hidden Group Time Profiles: Heterogeneous Drawdown Behaviours in Retirement

Publication note. This chapter has been adapted into a paper and is submitted to a peer-reviewed academic journal. The authors for the submitted paper are: Igor Balnozan (primary author), Denzil G. Fiebig, Anthony Asher, Robert Kohn and Scott A. Sisson.

Abstract

This chapter investigates retirement decumulation behaviours using the Grouped Fixed-Effects (GFE) estimator applied to Australian panel data on drawdowns from phased withdrawal retirement income products. Behaviours exhibited by the distinct latent groups identified suggest that retirees may adopt simple heuristics determining how they draw down their accumulated wealth. Two extensions to the original GFE methodology are proposed: a latent group label-matching procedure which broadens bootstrap inference to include the time profile estimates, and a modified estimation procedure for models with time-invariant additive fixed effects estimated using unbalanced data.

3.1 Introduction

The importance and prevalence of preference heterogeneity is a pervasive feature of microeconomic modelling. Such heterogeneity is usually unobserved, remaining unexplained after controlling for the observable characteristics of individuals. Where the effects of unobservables have clear economic interpretations, methods of identifying latent groups of individuals that share common behaviours can provide insights into economic phenomena.

There is a need to improve understanding of the behaviours of retirees who fund their consumption by drawing down savings accumulated in personal Defined Contribution accounts. Such understanding can inform appropriate financial advice and product development. Existing theoretical work by Bateman and Thorp (2008) gives reason to expect distinct behavioural groups, where group behaviours correspond to strategies that individuals employ when accessing their retirement savings using phased withdrawal products.

This chapter investigates drawdown behaviours in phased withdrawal retirement income products by studying latent time profiles estimated using the Grouped Fixed-Effects (GFE) estimator of Bonhomme and Manresa (2015). Of primary interest is testing for the presence of multiple behavioural groups in the data, and characterising any groups found. Furthermore, in this application, using the GFE estimator also allows testing whether retirees were heterogeneous in their responses to the Global Financial Crisis and retirement incomes policy changes, demonstrating the value of the GFE estimator to such event studies.

In the presence of group-level time-varying unobservable heterogeneity, using the standard two-way fixed-effects (2WFE) model, defined in Section 3.4, generally gives biased covariate effect estimates and unrepresentative time effect estimates. This is what the present study finds: estimates for the time effects in the standard 2WFE model, which allows for only one set of time effects, obscure the different latent group-level effects which suggest retirees may adopt simple, distinct heuristics when determining how they draw down their accumulated wealth during retirement. The importance of

this comparison with the 2WFE model results motivates our modification to the GFE estimator, described in Section 3.4.

A key novelty in this application of the GFE estimator is the focus on performing statistical inference on the estimated effects of the latent heterogeneity. To robustify the GFE estimator in this scenario, this chapter proposes an extension to the bootstrap method outlined in the supplement to Bonhomme and Manresa (2015). While these authors describe a procedure for using the bootstrap to obtain standard errors for covariate effects, a group label switching problem prevents finding standard errors for the effects of group-level time-varying unobservable heterogeneity. A method to match group labels across independent GFE estimations is required so that the bootstrap can also explore uncertainty in the latent heterogeneity effect estimates.

Related to this, the fixed-T variance estimate formula given in the supplement to Bonhomme and Manresa (2015) provides an alternative approach to performing inference on the effects of the latent heterogeneity, where T is the maximum number of observations on each unit in the sample. The proposed label-matching method also allows studying the properties of the standard errors for the latent group effects derived from this analytical formula, by observing their distribution across a large number of simulated datasets. This is useful in determining the potential applicability of the analytical formula for obtaining standard errors of the latent group effects in finite samples, explored in Appendix A.

The rest of the chapter is organised as follows. Section 3.2 summarises the literature to which this chapter contributes. Section 3.3 describes the available data for the application to retirement decumulation behaviours, with further details given in Appendix 3.A. Section 3.4 outlines the GFE estimator from Bonhomme and Manresa (2015), and explains our proposed extensions. Section 3.5 presents the main results from applying the GFE estimator to the data. Section 3.6 concludes by examining the implications for retirement incomes research and policy. The supplementary material in Appendix A provides robustness checks, a simulation study, descriptive analysis of the data, and further details on the second proposed methodological extension.

3.2 Related literature

3.2.1 Drawdown behaviours in phased withdrawal retirement income products

Recent work studying drawdown behaviours builds on Horneff, Maurer, Mitchell, and Dus (2008). These authors compare three rate-based drawdown strategies against a level guaranteed lifetime annuity, using a model of retiree utility that incorporates stochastic interest returns and retiree lifetimes. The first strategy draws a fixed proportion of the account balance each year; the second determines a terminal time horizon T and draws a proportion $1/T$ of the account balance in the first year, $1/(T-1)$ in the second year, and so on, continuing until the T^{th} year of the plan when all the remaining funds are drawn down; the third draws a proportion of the balance that updates each year based on the surviving retiree's expected remaining lifetime.

Bateman and Thorp (2008) extend Horneff et al. (2008), motivated by newly legislated minimum annual drawdown rates for a phased withdrawal retirement income product in Australia known as the account-based pension. Table 3.1 details these age-based rates, which start at 4% of the account balance for retirees under age 65, and increase as a step function of age to a maximum of 14% for retirees aged 95 or above; these are specified in Schedule 7 of the Superannuation Industry (Supervision) Regulations 1994. The rates, when applied to a retiree's account balance at the start of the relevant financial year, give the dollar amount the retiree must draw down from their account during that financial year.

Bateman and Thorp (2008) use a stochastic lifecycle model to compare five drawdown strategies. Three are based on Horneff et al. (2008); the remaining two are based on legislated minimum drawdown rates in Australia: always drawing at the newly legislated minimum drawdown rates effective from 1 July 2007;[1] and drawing at the

[1] Superannuation Industry (Supervision) Amendment Regulations 2007 (No. 1) Schedule 3.

Table 3.1: Minimum drawdown rates by age for account-based pensions, effective 1 July 2007.

Age at financial year start	<65	65–74	75–79	80–84	85–89	90–94	95+
Minimum drawdown rate	0.04	0.05	0.06	0.07	0.09	0.11	0.14

minima previously in effect. Bateman and Thorp (2008) also use their model to derive the implied optimal drawdown plan, and examine the sensitivity of these results to changes in the model assumptions. The fixed-rate strategy they consider draws a fixed proportion of the remaining account balance each year, where this fixed proportion is the annuity payout rate the retiree would have received in the market at the time of writing by purchasing a guaranteed lifetime annuity using their account balance at retirement. The authors find that while the newly legislated minima outperform the second and third rate-based strategies of Horneff et al. (2008) described earlier, in most cases, simulated utility is increased by initially drawing at a fixed rate higher than the minimum, then switching to the new minimum drawdown schedule once the minimum rate rises higher than the fixed rate.

Based on this theoretical literature alone, researchers can expect, *a priori*, that analysing drawdowns data will reveal groups of retirees exhibiting distinct drawdown behaviours based on simple rules, or heuristics. This chapter uses recent advances in panel data methods to empirically investigate whether such groups exist, and if so, what behaviours they display. Section 3.2.2 summarises recent methodological progress in panel data econometrics.

3.2.2 Capturing time-varying unobserved heterogeneity

Bonhomme and Manresa (2015) present the Grouped Fixed-Effects (GFE) estimator and apply it to a linear panel data model that allows for group-level time-varying unobserved heterogeneity and unknown group membership. Importantly, their method allows for correlation between the latent group effects and the observed regressors. Being able to control for, and estimate, the varying latent group time profiles is the main advantage of using this model over the traditional fixed-effects model, which captures only time-constant unit-level unobservable heterogeneity.

Factor-model approaches to capturing unobserved heterogeneity (e.g., Bai, 2009; Su & Ju, 2018) provide more flexible specifications for the individual-specific heterogeneity. These models assume the presence of common, but unobserved, time-varying factors to which individuals respond heterogeneously. Bonhomme and Manresa (2015) show

that their linear panel model, to which they apply the GFE estimator, is a special type of latent-factor model, where the latent factors are group-specific time effects and the factor loadings are group membership indicator variables. In applications where group time profiles are of economic interest, having this interpretation of the time-varying unobservable heterogeneity is an advantage of using the GFE estimator over a factor-model approach. Furthermore, compared to competing latent-factor model methods, the GFE estimator converges faster to the true parameter values with T, and so has better finite-sample performance (Bonhomme & Manresa, 2015). This makes GFE the preferred estimator when analysing a typical microeconomic panel, which may have a large number of units N, but rarely more than a moderate length T.

A related research area concerns finite mixture models, which can incorporate unobservable heterogeneity when the heterogeneity is constant over time. Finite mixture models are parsimonious solutions to modelling unobserved heterogeneity in populations, wherein data from latent subpopulations can be drawn from different distributions, and the true allocation of specific individuals to subpopulations is unobserved. Deb and Trivedi (2013) develop finite mixture models with time-constant fixed effects. Critically, these models cannot specify the likelihood of being in a particular group as a function of the covariates. By contrast, one of the primary advantages of using the GFE estimator over a standard 2WFE model estimator is the ability to control for correlation between latent group effects and the included covariates. Mixture-of-experts models (e.g., Jacobs et al., 1991; Jordan & Jacobs, 1994) generalise mixture models by parametrically specifying a link between group identity and covariate values; however, these models do not control for unit-level fixed effects in panel data.

The discussion above suggests that the GFE estimator is the most appropriate for the present application, because: the available dataset only has a moderate panel length T; a review of the theoretical drawdowns literature suggests that there may be a finite number of distinct behavioural groups in the population, consistent with the GFE assumption that there are a finite number of latent group time profiles; the time profiles recovered by the GFE estimator are likely to have economically meaningful interpretations.

3.3 Data on superannuation drawdowns

The available superannuation dataset contains information on drawdowns from account-based pensions (ABPs), a phased withdrawal retirement income product in Australia. Using ABPs, retirees generate personal income streams or receive lump-sum payments by drawing down their accumulated savings during retirement. Throughout, the balance remains invested in financial markets based on each retiree's chosen combination of safe and risky exposures.

Following the Global Financial Crisis (GFC), the Australian government introduced temporary minimum drawdown rate reductions as per Schedule 7 of the Superannuation Industry (Supervision) Regulations 1994. These resulted in 'concessional' minimum drawdown rates, which came into effect for the financial year ended 30 June 2009. For financial years 2009–2011 inclusive, the reduction was 50%, so that any retiree's minimum drawdown rate over this period was halved relative to the standard schedule in Table 3.1. For financial years 2012 and 2013, the reduction decreased to 25%, meaning the minimum drawdown rates were three-quarters of those given in the table. For subsequent financial years, the rates returned to their nominal values.

To generate an income stream from an ABP, retirees can elect a payment amount and a frequency, for example $1000 monthly, which the superannuation fund follows in paying the retiree from their account balance. This type of drawdown, which is specified by the beginning of a given financial year, is referred to here as a 'regular' drawdown, and is the primary object of interest in this study. These regular drawdowns are distinct from 'ad-hoc' drawdowns, referring to any additional lump-sum payments requested by the retiree during the financial year. A rigorous analysis of ad-hoc drawdowns is crucial in understanding retirees' needs for flexibility and insurance, but is not the focus of this chapter.

The dependent variable studied here is $y_{it} := \ln \mathrm{DR}_{it}$; DR_{it} is the rate of drawdown for retiree i over financial year t, which is a ratio of the regular drawdown amount over the account balance: $\mathrm{DR}_{it} := \mathrm{DA}_{it}/\mathrm{AB}_{it}$. The dependent variable is the logarithm of this rate because the raw rate variable is strictly positive and right-skewed. Appendix 3.A

gives further details on the relationship between the drawdown rate and the account balance, as well as histograms of the dependent variable and the raw rate variable.

Qualitatively, one interpretation of the drawdown rate is the speed at which retirees deplete their accumulated savings throughout retirement. In ABPs, retirees are exposed to the risk of outliving their savings and their retirement income becoming reliant on non-superannuation assets or taxpayer-funded transfer payments known as the age pension. Policy and product design should support retiree needs for regular income, and so understanding the drawdown behaviours that manifest in this flexible withdrawal product has the potential to inform policymakers, financial advisors and superannuation fund trustees.

The superannuation dataset contains $N = 9516$ individuals, combining source data from multiple large industry and retail superannuation funds. Due to the small number of funds in the sample whose data permitted determining the regular drawdown rate, the economic results in this chapter may not be representative of the Australian superannuation system as a whole.

The data capture window spans the financial years 2004 to 2015 inclusive, for $T = 12$ annual observations. The main results use a balanced subset of each superannuation fund's data, removing individuals with missing observations. However, as the data capture window is not the same for each fund, when combining these balanced subsets, the resulting dataset is unbalanced. Appendix A shows that the results are robust to using a fully balanced subsample of all the available data.

Individuals recorded as dying during the sample observation period do not appear in the data analysed. For some records, obtaining regular and ad-hoc drawdown amounts requires inputting these as estimates based on observed total drawdown amounts using a method which, in periods where the retiree makes an ad-hoc drawdown, may underestimate the amount of the ad-hoc drawdown; this affects fewer than 2.4% of the records in the sample.

The superannuation dataset is derived from administrative data, and so has limited demographic information on members. Table 3.2 provides summary statistics for characteristics that vary across individuals, but not time; Table 3.3 summarises the

time-varying variables in the dataset. The two covariates considered are the minimum drawdown rate for each individual at each time point, and their account balance at the beginning of the respective financial year, both on the log scale. Thus the model, with regular drawdown rate on the log scale as the dependent variable, estimates the elasticity of the regular drawdown rate with respect to the minima and account balances. The remainder of the variation in log regular drawdown rates is attributed to the latent group effects estimated by the GFE procedure, and residual noise.

The lack of demographic information, particularly health and marital status, is also a data-based motivation for using GFE as an estimation procedure. These unobserved characteristics are a potential source of omitted variable bias; they are possibly correlated with the dependent variable as well as the included covariates. Because relevant characteristics such as health and marital status may vary over time, controlling for time-constant unobserved heterogeneity by using a standard fixed-effects model may not remove the biasing effect of all relevant unobservables. For this reason, a model allowing for sources of time-varying unobserved heterogeneity, such as the linear panel model to which the GFE estimator is applied, may be more appropriate; the results in Section 3.5.1 provide evidence for this conjecture.

Table 3.2: Summary statistics – time-invariant variables.

	Age at 31 December 2015	Age at Account Open	Sex: Male	Risk Appetite
Mean	79.4	63.57	0.56	0.41
SD	5.22	4.17	0.5	0.21
Median	79.78	64.23	1	0.46
Q1	76.37	60.9	0	0.25
Q3	83.03	65.39	1	0.54
Min	60.66	48.48	0	0
Max	101.46	85.44	1	1.88
Count	9516	9516	9516	9507

The age at 31 December 2015 represents the individual's cohort, equivalent to measuring a year-of-birth variable. The age at account opening is the age when the retiree initiates a phased withdrawal product and begins drawing down from the account. The sex indicator variable equals 1 if the retiree is male, and 0 otherwise. The risk appetite variable is a proxy for the returns in the account relative to the reference S&P/ASX 200 index; see Appendix A for details. SD refers to the sample standard deviation of the variable. Q1 and Q3 refer to the empirical 1st and 3rd quartiles, respectively.

Table 3.3: Summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.12	6435.91	0.07	0.15	10,216.56	72,686.55
SD	0.12	6121.76	0.25	0.21	24,672.68	78,721.39
Median	0.09	4800	0	0.07	4655.9	52,063
Q1	0.07	2992	0	0.02	1132.33	30,532.5
Q3	0.12	7728	0	0.18	10,000	87,427
Min	0	1	0	0	1	1
Max	2	166,695	1	0.9	600,000	2,427,083
Count	107,935	107,975	108,717	7450	7454	108,635

The regular drawdown rate is the dependent variable. The account balance is defined as the position at the start of the relevant financial year; this is the denominator for the rate computations. The ad-hoc drawdown indicator variable equals 1 if the retiree made an ad-hoc withdrawal from their account during a given financial year. SD refers to the sample standard deviation of the variable. Q1 and Q3 refer to the empirical 1st and 3rd quartiles, respectively.

3.4 Methodology

3.4.1 The Grouped Fixed-Effects (GFE) Estimator

The GFE estimator introduced by Bonhomme and Manresa (2015) considers a linear model of the form

$$y_{it} = x'_{it}\theta + \alpha_{g_i t} + v_{it}; \tag{3.1}$$

in our application, $i = 1, 2, ..., N$ indexes individual retirees; $t = 1, 2, ..., T$ indexes financial years; y_{it} is the dependent variable; x_{it} is a vector of covariates; θ are the partial effects of covariates on the dependent variable after controlling for group-level time profiles; $g_i = 1, 2, ..., G$ identifies the group membership for unit i, where G is the chosen number of latent groups in the sample; $\alpha_{g_i t}$ is a term representing time-varying, group-specific unobserved heterogeneity—these are time-varying model intercepts that can differ across individuals, depending on the value of the individual's group identifier, g_i; v_{it} represents the residual effect of all other unobserved determinants of the dependent variable.

For the consistency and asymptotic Normality results of Bonhomme and Manresa (2015) to apply, the distribution of the errors, v_{it}, must have finite fourth moment and be uncorrelated with the covariates, x_{it}; errors can be correlated across cross-sectional units; the unit-level time-sequences of errors can be autocorrelated if the sequences are strongly mixing processes; the tails of the error distribution must exhibit a faster-than-polynomial decay rate. The time profile values, α_{gt}, may be correlated with x_{it} but not

v_{it}. For the complete list of assumptions required for the GFE estimator, see Bonhomme and Manresa (2015), pp. 1156–1159.

The linear model in (3.1) can be extended to include time-invariant, individual-specific fixed effects, c_i, so that

$$y_{it} = x'_{it}\theta + c_i + \alpha_{g_i t} + v_{it}, \tag{3.2}$$

because applying the within transformation (centering the variables around individual-specific means) reduces (3.2) to (3.1). To see this, for any variable z_{it}, define $\dot{z}_{it} := z_{it} - \bar{z}_i$, where $\bar{z}_i := T^{-1} \sum_{t=1}^{T} z_{it}$; then,

$$\dot{y}_{it} = \dot{x}'_{it}\theta + \dot{\alpha}_{g_i t} + \dot{v}_{it}, \tag{3.3}$$

which has the same form as (3.1). In (3.3), the time profiles, $\dot{\alpha}_{g_i t}$, have a different economic interpretation to the $\alpha_{g_i t}$ from (3.1); Section 3.4.2 explains the treatment and interpretation of the group time profile estimators in the transformed model (3.3).

Throughout this chapter, 'GFE model' refers to (3.2). A useful way to consider the GFE model is as a generalisation of the 2WFE model,

$$y_{it} = x'_{it}\theta + c_i + \alpha_t + v_{it}, \tag{3.4}$$

where the GFE model allows distinct time profiles for G groups, with group membership unobserved.

Bonhomme and Manresa (2015) state the assumptions needed for large-N, large-T consistency and asymptotic normality of the GFE estimator, in models both with, and without, time-invariant fixed effects. Moreover, the asymptotics show that the estimator converges in distribution even when T grows substantially more slowly than N. For any fixed value of T, however, the θ and α estimators are root-N consistent not for the true population parameters, but for 'pseudo-true' parameter values, and although these pseudo-true values may differ from the true population values for any fixed T, any difference vanishes rapidly as T increases. Specifically, the pseudo-true and true values may differ when group classification is imperfect; because of this possibility, the authors derive a fixed-T consistent variance estimate formula which takes into account potential

group misclassification, providing a practical alternative to the large-T variance estimate formula (Bonhomme & Manresa, 2015, p. 1161). To examine the finite-T properties of the GFE estimator, Bonhomme and Manresa (2015) use a Monte Carlo exercise with simulated datasets that match the $N = 90$, $T = 7$ panel used in their empirical application. In the present study, the superannuation dataset has large $N = 9516$ but moderate $T = 12$; Appendix A finds that the GFE estimator and the fixed-T consistent variance estimator perform well on simulated datasets of this size.

When using the standard 2WFE model, if there is group-level time-varying unobservable heterogeneity, $\alpha_{g_i t}$, correlated with observable characteristics, x_{it}, two problems arise: first, estimates of the covariate effects, θ, may suffer from omitted variable bias; second, the group time profiles, which in many applications have interesting economic interpretations, remain hidden. The 2WFE model is a special case of the GFE model with $G = 1$; hence, the GFE model is an appealing alternative to the 2WFE model in applications where there is the possibility of unobservable time-varying heterogeneity in addition to time-invariant unobservable heterogeneity. While determining the precise number of groups, G, lacks a general solution, estimating the GFE model can provide evidence for whether only controlling for one set of time effects, as in the 2WFE model, is adequate; Section 3.5.1 discusses this issue.

The GFE estimator for (3.1) minimises the sum of squared residuals, giving

$$(\widehat{\theta}, \widehat{\alpha}, \widehat{\gamma}) = \underset{(\theta,\alpha,\gamma)\in\Theta\times\mathcal{A}^{GT}\times\Gamma_G}{\arg\min} \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\theta - \alpha_{g_i t})^2, \tag{3.5}$$

where the vector $\gamma = (g_1, g_2, ..., g_N)$ defines the grouping of each of the N units into one of the G groups; Γ_G is the set of all possible groupings of N units into G or fewer groups; Θ is a subset of $\mathbb{R}^k$, the k-dimensional real space, where k is the number of covariates or equivalently the dimension of x_{it}; $\mathcal{A}$ is a subset of $\mathbb{R}$; the α_{gt} parameters belong to $\mathcal{A}^{GT}$.

Bonhomme and Manresa (2015) present an iterative algorithm for estimating (3.5). The algorithm initialises values for the grouping vector γ and the model parameters (θ, α); it then alternates between the following two steps until convergence: 1) a grouping update step, which allocates each unit to the group minimising the sum of squared residuals

given the most recent estimate of the model parameters θ and α; 2) a parameter update step, which estimates the parameters (θ, α) conditioning on the most recent estimate of the grouping vector γ.

Estimating the GFE model does not require a balanced panel; the algorithm can be adjusted to run even when the sample contains unit–period combinations with missing data. However, when considering the relationship between the 2WFE model (3.4) and the GFE model (3.2) with $G = 1$, the unbalanced data case presents subtleties requiring further discussion.

The original GFE estimator for (3.2) with $G = 1$ gives the same estimates as applying the within transformation to the data and running a least-squares regression on the time-demeaned data, including T time dummy variables and no constant term. In balanced samples, this returns the same numerical results as standard implementations of the 2WFE model: for example `Stata`'s `xtreg` command with the `fe` option and time dummy variables as covariates. However, in unbalanced panels, the results differ.

With unbalanced data, obtaining the results from standard implementations of the 2WFE model requires time-demeaning the time dummy variables at the unit level. Hence, this chapter proposes and utilises a modified GFE estimation procedure having the following property: when $G = 1$, it recovers precisely the same estimates as a standard implementation of the 2WFE model even when the data is unbalanced and the model includes time-invariant unit-level unobservable heterogeneity. Section 3.4.5 describes this alternative estimation procedure. When the panel is balanced, or when there are no time-invariant unit-level unobservables, the results from this method are identical to the unmodified GFE estimation procedure results. However, panel data models in microeconomic applications generally control for time-invariant unit-level unobservable heterogeneity. Hence, this is a useful extension when applying the GFE estimator to microeconomic applications more broadly. To examine the sensitivity of the present results to using the modified algorithm, Appendix A provides a robustness check using the unmodified GFE estimation procedure in place of the proposed modified algorithm.

Bonhomme and Manresa (2015) explain that in the absence of covariates, their algorithm outlined above reduces to k-means clustering. Similarly to standard implementations of k-means clustering algorithms, the results depend on the starting values. Running the algorithm multiple times with randomly generated starting values increases the likelihood of finding solutions corresponding to smaller values of the objective function in (3.5).

In this study, the algorithm is initialised by drawing starting values for each covariate effect from independent Gaussian distributions, centred at the coefficient estimates from a 2WFE regression fit to the data, and with standard deviations equal to the magnitude of these coefficient estimates. Results for the application take the most optimal solution across 1000 independent runs of the algorithm using randomly drawn starting values to initialise the parameters. Appendix A also tests the sensitivity of the results to the number of starting values used.

To facilitate inference on the group time profiles, this study implements the fixed-T variance estimate formula in the supplement to Bonhomme and Manresa (2015), as well as an extension to their non-parametric bootstrap method. Notably, approximating the variability in time profile estimates using the non-parametric bootstrap method described by Bonhomme and Manresa (2015) requires solving a label-switching problem. Section 3.4.4 explains this problem and proposes an extension to the bootstrap procedure which matches group labels across replications to allow for bootstrap inference on the time profiles.

3.4.2 Time profiles in the transformed model

Within-transforming the data to remove the effect of any time-constant unobservable heterogeneity results in time-demeaned group time profiles, $\dot{\alpha}_{gt}$. As the $\dot{\alpha}_{gt}$ contain only information about changes in the group effects over time relative to their mean, and no information about the absolute level of the effect, the estimates of the $\dot{\alpha}_{gt}$ are modified in order to obtain the desired economic interpretations. All estimated time profiles are shifted to begin at a value of 0 at $t = 1$, so that the interpretations of the values are as changes relative to the first time period. This is analogous to estimating a set of $T - 1$

coefficients for the time effects in the standard 2WFE model, where each estimate is interpreted as a change relative to the omitted reference period.

This study uses two methods to estimate the uncertainty in the shifted time-demeaned group time profiles, defined as $\tilde{\dot{\alpha}}_{gt} := \dot{\alpha}_{gt} - \dot{\alpha}_{g,1}$, for all g and t. First, Normal-approximation 95% confidence intervals are constructed for the $\tilde{\dot{\alpha}}_{gt}$, using variance estimates which are computed from the variance–covariance matrix components for the time-demeaned time profiles, $Var(\dot{\alpha}_{gt})$, $Var(\dot{\alpha}_{g,1})$ and $Cov(\dot{\alpha}_{gt}, \dot{\alpha}_{g,1})$; these are obtained by applying the fixed-T variance estimate formula after estimating the GFE model on the within-transformed data. Second, these confidence intervals are compared to the corresponding intervals obtained from the bootstrap, using the proposed extension for matching group labels across bootstrap replications.

A comparison of the time profile standard errors derived from the fixed-T variance estimate formula versus simulated standard errors from a Monte Carlo exercise shows that the fixed-T variance estimate formula performs well on simulated data calibrated to the superannuation dataset; Appendix A provides details of this comparison. Moreover, in the present study, confidence intervals constructed from the bootstrap procedure results are broadly similar to those derived from the fixed-T variance estimate formula, implying practically equivalent inference on the estimated parameters. For smaller datasets, where asymptotic results are less likely to provide accurate standard errors, and where the computational cost of running the bootstrap is lower, it may be preferable to use the bootstrap. However, the results suggest that in larger datasets, using the fixed-T variance estimate formula may be sufficiently accurate while keeping the computational burden low compared to using the bootstrap.

3.4.3 Selecting G

Bonhomme and Manresa (2015) propose an information criterion that correctly selects the number of groups when N and T are both large and similar in magnitude; however, for large-N moderate-T panels, like the superannuation dataset, this criterion may overestimate the true value of G (Bonhomme & Manresa, 2015). An alternative method to choose the number of groups, which Bonhomme and Manresa (2015) employ in their

empirical application, uses the result that underestimating G leads to omitted-variable bias in the θ parameters, to the extent that the unobserved effects are correlated with the included covariates. Conversely, overestimating G does not bias the θ estimates, although it does increase the number of unknown parameters and the parameter space of possible groupings, resulting in noisier parameter estimates.

This suggests plotting the coefficient estimates against G can reveal the point at which the coefficients have been de-biased. Visually, the coefficients might vary significantly between $G = 1$ and some higher value, $G^\star$, after which they may appear to stabilise, albeit becoming noisier as G continues to increase. One can then select $G^\star$ as the number of groups suggested by the data.[2] This is how Bonhomme and Manresa (2015) decide on $G = 4$ in their application, rather than $G = 10$ as suggested by the information criterion—which, given the dimensions of their panel, may overestimate the true number of groups.

For applications focusing only on estimating unbiased effects of the covariates, the above suggests choosing the number of groups by selecting the smallest value of G that places the covariate effect estimates close to their stable values. However, the present application is interested in identifying all distinct behavioural groups with economically meaningful interpretations, which may require searching beyond the smallest G that de-biases the regression coefficients.

Conversely, overestimating G can result in the separation of a particular time profile into multiple biased representations of itself, where the difference between the representations is spuriously driven by noise (Bonhomme & Manresa, 2015). This means that the more similar a pair of time profiles appear, the more sceptical the researcher should be in regarding them as distinct effects.

Hence, this study posits a selection rule that picks the greater of two values of G: the smallest G that appears to result in unbiased estimates of the regression coefficients; or the last value of G for which the marginal change from $G - 1$ to G groups finds an economically important behaviour, sufficiently distinct from all others and exhibited by

[2] This method would not apply to a model with group-specific covariate effects.

a non-trivial portion of the sample. Section 3.5 discusses this selection process alongside the main results.

3.4.4 Extension 1: Bootstrapping time profile estimates

Bonhomme and Manresa (2015) address how to obtain bootstrap standard errors for the covariate effects, but not the time profiles. As the covariate effects are unambiguously labelled across multiple runs of the GFE procedure, estimating their standard errors by comparing estimates across a large number of bootstrap replications requires no special consideration. By contrast, group-level results from the procedure are unique only up to a relabelling of the groups, and the group labels are determined at random during estimation. This means that the same economic group may receive different labels across multiple independent estimation runs, e.g., when performing the bootstrap. Hence, estimating the standard errors of group-level time effects by comparing time profile estimates across bootstrap replications requires a method to match labels across runs.

The bootstrap procedure outlined in the supplement to Bonhomme and Manresa (2015) involves running the GFE procedure B times, each time on a different bootstrap replicate dataset. To create the bootstrap replicate datasets, all N units in the original data are sampled with replacement N times, taking all observations corresponding to a given unit into the replicate dataset when that unit is sampled. Each of the B model runs produces a set of coefficient estimates for the model covariates, as well as an estimated grouping and a set of time profiles for each group.

To match group labels across replications, this chapter proposes the following label-matching procedure, similar to the solution Hofmans, Ceulemans, Steinley, and Van Mechelen (2015) use for the analogous problem in k-means clustering.

1. Fix the labelling generated by the GFE model estimation completed on the original sample. These labels act as the reference labels to which all subsequent estimates will align their group labels. Define the initial time profile estimates labelled using the reference labels as $\widehat{\alpha}_g^r := \left(\widehat{\alpha}_{g,1}^r, \widehat{\alpha}_{g,2}^r, ..., \widehat{\alpha}_{g,T}^r\right)'$ for $g = 1, 2, ..., G$.

2. For $b = 1, 2, ..., B$, find the permutation of group labels for the b^{th} bootstrap run which minimises the sum of Euclidean distances, aggregated across all G time profiles, between the time profiles estimated in the b^{th} bootstrap run and the original sample estimates as identified by the reference labels; i.e., for $b = 1, 2, ..., B$, do:

   2.1. Using the unmodified labels generated by the b^{th} bootstrap run, define the estimated time profiles from the b^{th} bootstrap run with these labels as the length-T vectors $\widehat{\alpha}_g^{b,0}$ for $g = 1, 2, ..., G$.

   2.2. Index the $G!$ permutations of the label sequence $(1, 2, ..., G)$ by $p = 1, 2, ..., G!$. Each permutation p for the b^{th} bootstrap run defines a permuted set of time profiles, $\widehat{\alpha}_g^{b,p}$, for $g = 1, 2, ..., G$, and a corresponding mapping function, $m_p : \{1, 2, ..., G\} \to \{1, 2, ..., G\}$, such that $m_p(g)$ is the g^{th} element of the p^{th} permuted label sequence, and $\widehat{\alpha}_g^{b,p} = \widehat{\alpha}_{m_p(g)}^{b,0}$.

   2.3. For $p = 1, 2, ..., G!$, do:[3]

      2.3.1. For $g = 1, 2, ..., G$, compute the Euclidean distance between the length-T vectors $\widehat{\alpha}_g^{b,p}$ and $\widehat{\alpha}_g^r$, given by $\|\widehat{\alpha}_g^{b,p} - \widehat{\alpha}_g^r\|$; here $\|\cdot\|$ is the L_2 norm.

      2.3.2. Sum all G Euclidean distances to compute an aggregate distance metric for that permutation, $\sum_{g=1}^{G} \|\widehat{\alpha}_g^{b,p} - \widehat{\alpha}_g^r\|$.

   2.4. Select the permutation, $p^\star$, with the smallest aggregate distance metric:

$$p^\star = \underset{p \in \{1,2,...,G!\}}{\arg\min} \sum_{g=1}^{G} \|\widehat{\alpha}_g^{b,p} - \widehat{\alpha}_g^r\|. \tag{3.6}$$

   Relabel the G groups in the b^{th} bootstrap run using the labelling given by the selected permutation.

This method matches time profiles across different bootstrap replications in the main results presented in Section 3.5, as well as across different simulated datasets and bootstrap replications in the simulation study reported in Appendix A. In general, it could be seen as an issue that time profile bootstrap results are contingent on some method to match labels, because, in principle, the bootstrap results may be sensitive to

[3] If $G!$ is too large to check each permutation in an acceptable amount of time, an alternative is to use the more computationally efficient approach given in Munkres (1957) and used by Hofmans et al. (2015).

the choice of the label-matching method. The results from this chapter, which use the procedure described here, show that in this application, the bootstrap and analytical formula results are similar, and so little is gained by using the bootstrap over the analytical formula. Thus, we did not pursue this question further, but this general problem may be of interest to future research.

Deriving standard errors for the shifted time-demeaned group time profiles, $\tilde{\dot{\alpha}}_{gt}$, using the fixed-$T$ variance estimate formula requires first estimating variance–covariance matrix components for the time-demeaned time profile estimators, $\dot{\alpha}_{gt}$, then using these variance–covariance matrix components to estimate standard errors for the shifted time profiles to construct Normal-approximation element-wise confidence intervals. By contrast, the bootstrap results directly estimate the variability of the shifted time profile estimators, by matching shifted time profiles across the B bootstrap replications, then using the empirical distribution of the matched time profile estimates to obtain bootstrap intervals.

3.4.5 Extension 2: Alternative estimation method for unbalanced data

Consider the model in (3.4), which, after time-demeaning, has the form

$$\dot{y}_{it} = \dot{x}'_{it}\theta + \dot{\alpha}_t + \dot{v}_{it}. \tag{3.7}$$

The primary estimands of interest are the $T-1$ relative effects $\tilde{\dot{\alpha}}_t = \dot{\alpha}_t - \dot{\alpha}_1$ for $t = 2, 3, ..., T$. Since $\dot{\alpha}_t := \alpha_t - \bar{\alpha}$, where $\bar{\alpha} = T^{-1}\sum_{t=1}^{T}\alpha_t$, the set of T values of $\dot{\alpha}_t$ satisfy $\sum_{t=1}^{T}\dot{\alpha}_t = 0$.

In balanced panels, the following three methods obtain identical estimates for the $\tilde{\dot{\alpha}}_t$:[4]

1. regress $\dot{y}_{it}$ on $\dot{x}_{it}$ and T time dummy variables with no constant term to estimate all T values for $\dot{\alpha}_t$ as the coefficients of the time dummies; then compute $\tilde{\dot{\alpha}}_t = \dot{\alpha}_t - \dot{\alpha}_1$ for $t = 2, 3, ..., T$;

2. regress $\dot{y}_{it}$ on $\dot{x}_{it}$ and $T-1$ time dummy variables excluding the dummy variable for the first time period, including a constant term, after which the $T-1$ desired

[4]This claim is justified in Appendix A.

estimates of $\tilde{\dot{\alpha}}_t$ are the estimated coefficients of the included $T-1$ time dummy variables;

3. first, time-demean the time dummy variables corresponding to periods $2, 3, ..., T$; then, regress $\dot{y}_{it}$ on $\dot{x}_{it}$ and the $T-1$ time-demeaned time dummy variables, not including a constant term, after which the $T-1$ desired estimates of $\tilde{\dot{\alpha}}_t$ are the estimated coefficients of the included $T-1$ time dummy variables.

For an unbalanced panel, methods 1 and 2 described above for obtaining estimates of $\tilde{\dot{\alpha}}_t$ no longer guarantee that the corresponding values of $\dot{\alpha}_t$ sum to zero; a nonzero sum contradicts a property of the model given by (3.7). Conversely, method 3 obtains the desired estimates while ensuring that the T estimates of $\dot{\alpha}_t$, derived from the $T-1$ estimates for $\tilde{\dot{\alpha}}_t$, sum to 0, consistent with the model.

Using the GFE estimator implementation provided by Bonhomme and Manresa (2015) with $G = 1$ on unbalanced data gives estimates for the time effects in line with those produced by methods 1 and 2, whereas standard implementations of the 2WFE model give the same estimates as method 3. This motivates our modification to the GFE methodology, which aligns the $G = 1$ results with the standard 2WFE model estimates, regardless of whether the panel is balanced or not: in the update step of the modified procedure, the algorithm interacts the group identity variables with $T-1$ time-demeaned time dummy variables, analogous to using method 3 for $G = 1$.[5]

3.5 Main results

3.5.1 Covariate effects

Figure 3.1 plots the covariate effect estimates versus G for the GFE model fitted to the superannuation drawdowns dataset. The values of G on the horizontal axis correspond to the number of latent groups specified; the vertical axis denotes partial effect values for the two included covariates: log minimum drawdown rate and log account balance. The lines connect covariate effect estimates, while the shaded regions around these

[5] Appendix A provides further details of this modification, as well as results using the unmodified GFE estimation procedure to compare with the results presented here.

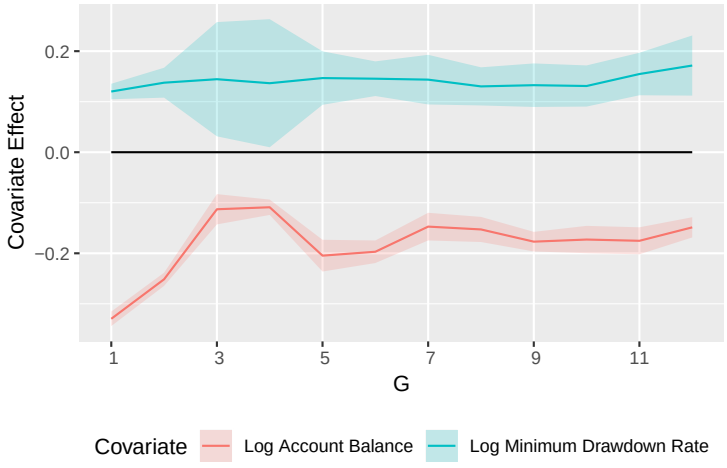
correspond to 95% confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula in the supplement to Bonhomme and Manresa (2015).

As the dependent variable and covariates are on the log scale, the covariate estimates represent the elasticity of regular drawdown rates to changes in account balances and the minimum drawdown rates. Section 3.5.2 provides the economic interpretation for the corresponding estimates after choosing the value of G.

A critical feature of Figure 3.1 is that as G increases, changes in the log account balance covariate effect estimates are large relative to the confidence intervals for the first few values of G. After $G = 7$, the point estimates stabilise and successive 95% confidence intervals show significant overlap. This suggests that from $G = 7$, the GFE procedure removes the bias arising from correlation between group-level latent time effects and the included covariates.

Using the proposed modified GFE estimator, the point estimates corresponding to $G = 1$ are identical to the results from a standard 2WFE estimator on the drawdowns dataset. Hence, Figure 3.1 implies that the estimators using a more traditional analysis including

Figure 3.1: Point estimates and 95% confidence intervals for the partial effects of the log minimum drawdown rate and log account balance covariates on the log regular drawdown rate, controlling for group-level time-varying unobservable heterogeneity for a range of values of G. The shaded regions denote 95% confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula in the supplement to Bonhomme and Manresa (2015).



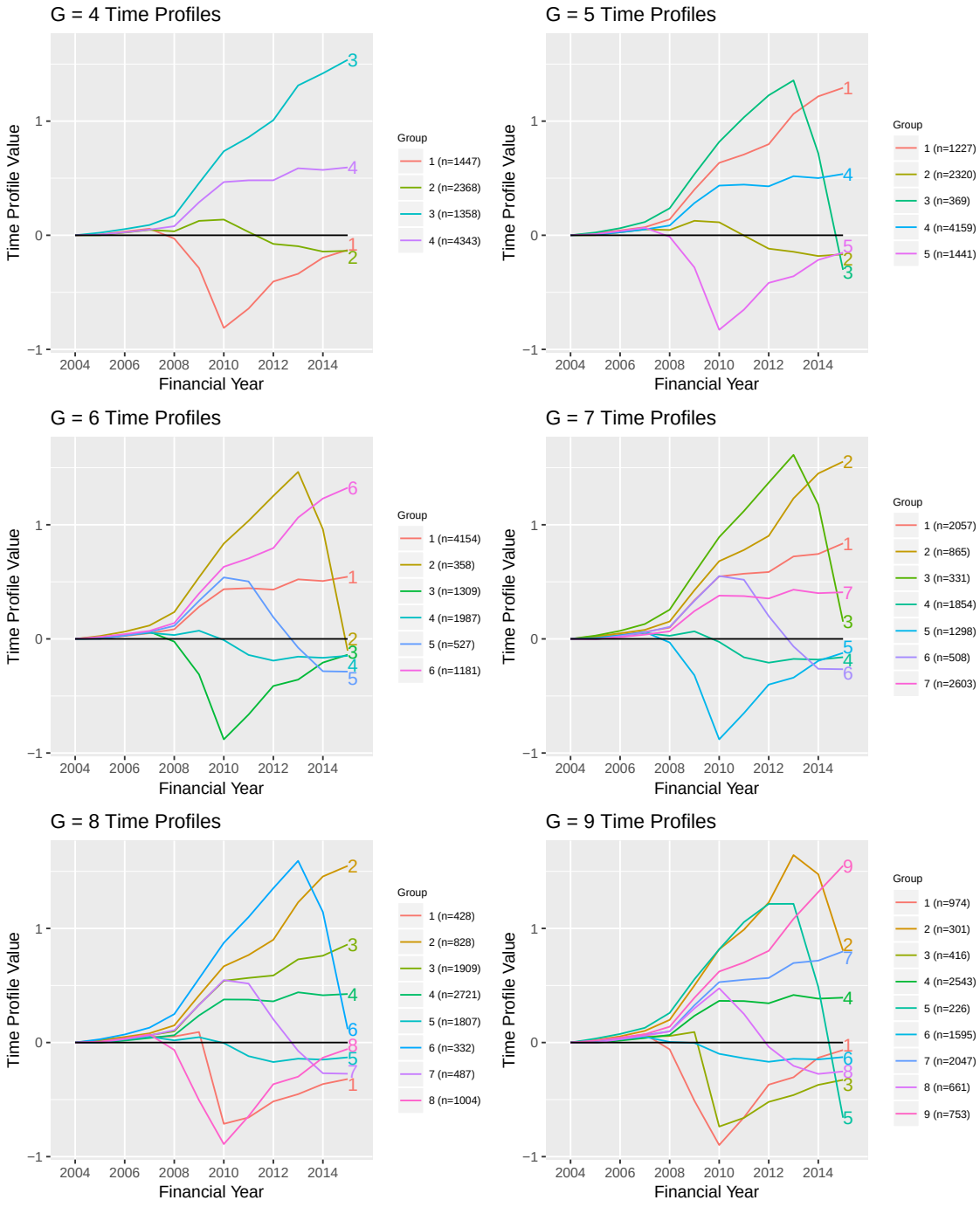
only one time profile are biased compared to the estimators in the GFE model for $G \geq 7$. Appendix A shows that this result holds in the fully balanced case, where both the modified and unmodified GFE procedures give results identical to the 2WFE model when $G = 1$.

3.5.2 Choice of G

Figure 3.2 shows plots of the time profiles estimated for $G = 4, 5, ..., 9$, all shifted to begin at a y-axis value of 0, and omitting confidence interval bounds for clarity. This figure suggests that while increases in G initially produce new and economically distinct time profiles, at $G = 7$, incremental moves to a larger value of G split existing time profiles into highly similar representations of the same trajectory. This is consistent with a theoretical result from Bonhomme and Manresa (2015) that implies overestimating G can create spurious copies of the true time profiles, differing by random noise. Hence, the number of latent groups to use in the model is set to $G = 7$. Section 3.5.3 explains in detail the economic interpretations of the resulting time profile estimates.

Selecting $G = 7$ groups corresponds to estimated covariate effects of 0.144 and -0.147 for the log minimum drawdown rate and log account balance, respectively. These are statistically significant at the usual levels, e.g. 5%, with respective standard errors of 0.0251 and 0.0140. As the dependent variable and covariates are on the log scale, the covariate effects represent partial elasticities of the regular drawdown rate with respect to the minimum drawdown rate and account balances. For example, consider an increase of 0.1 in the log minimum drawdown rate, which translates to a proportional increase in the level of the minimum drawdown rate of approximately 10.5%. This increase in the log minimum drawdown rate is expected to raise regular drawdown rates proportionally by an estimated $e^{0.0144} - 1 = 1.45\%$ on average, holding account balance equal and controlling for group-level time-varying heterogeneity. A similar proportional increase in a retiree's account balance is expected to effect a proportional decrease of 1.46% $(1 - e^{-0.0147})$ in their regular drawdown rate on average, holding the minimum drawdown rate constant and controlling for group effects.

Figure 3.2: Point estimates for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 4, 5, ..., 9$. The estimated time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.



3.5.3 Time profiles

Figure 3.3 shows the time profiles estimated using the selected value of $G = 7$ with element-wise confidence interval bounds constructed using standard errors derived from the fixed-T variance estimate formula. Figure 3.4 presents the equivalent plot with confidence interval bounds computed from the empirical 2.5 and 97.5 percentiles of the time profile estimates from 1000 bootstrap replications.[6] While qualitatively similar overall to the intervals constructed using standard errors derived from the fixed-T variance estimate formula, a comparison highlights some differences. In particular, the intervals are wider around financial years 2011–2013 for group 6; the interval is wider for financial year 2010 for group 4; the intervals are tighter for group 5 throughout; the interval is wider in the terminal financial year for group 3; and the point estimates from the original sample tend towards the lower end of the 95% bootstrap confidence intervals for groups 1, 2 and 7. However, due to the estimated effect magnitudes, these numerical differences in the sets of confidence intervals do not meaningfully change inferential conclusions regarding the estimates or their economic interpretations.

[6]Each bootstrap replication uses distinct sets of randomly generated starting values for the estimation algorithm.

Figure 3.3: Point estimates and 95% confidence intervals from the analytical formula for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 7$. The shaded regions denote 95% element-wise confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula in the supplement to Bonhomme and Manresa (2015). The time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.

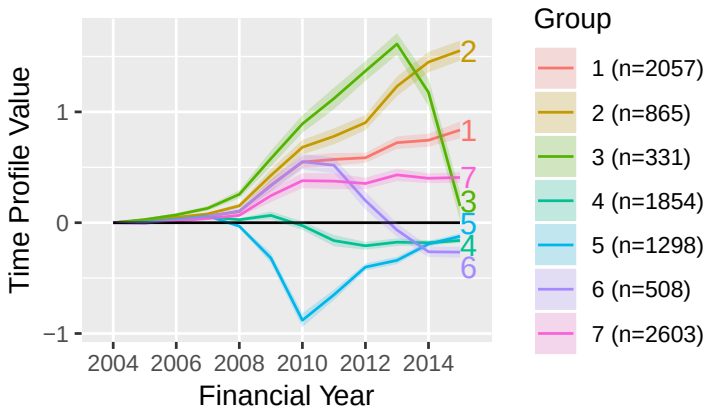
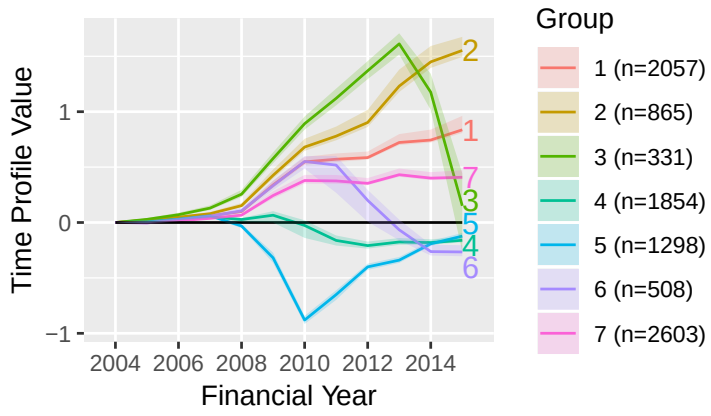


Figure 3.4: Point estimates and 95% bootstrap confidence intervals for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 7$. The shaded regions denote 95% element-wise confidence intervals computed from the empirical percentiles across 1000 bootstrap replications. The time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.



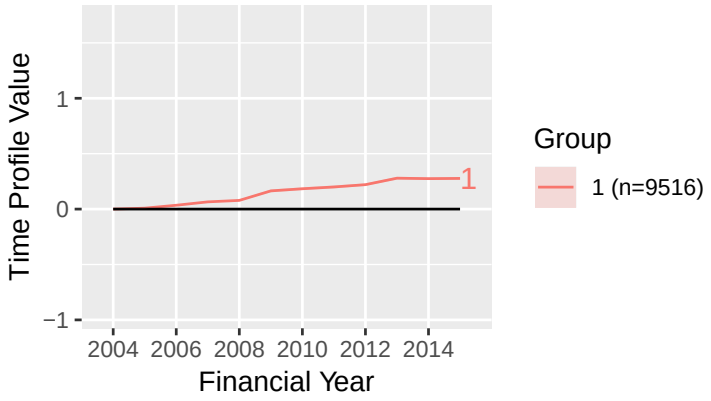
The values of the time profiles represent behavioural effects on the proportional changes in a retiree's regular drawdown rate, relative to their own rate of drawdown in 2004. As the dependent variable is on the log scale, a value of, say, 0.5 on the y-axis represents a proportional change in the regular drawdown rate of $e^{0.5} \approx 1.65$, or roughly a 65% increase in the drawdown rate, relative to 2004 levels.

Figure 3.5 shows the time profile estimates for the $G = 1$ model estimated using the modified GFE procedure, with the y-axis scale set to the same scale used for the $G = 7$ model group time profile plot (Figure 3.3). Comparing these shows that a model controlling for only one set of time effects fails to capture the shape and magnitude of any of the time profiles in the seven-group model. Visually, the resulting single time profile appears to 'average over' the seven markedly distinct time profiles. Hence, not only are the covariate effect estimates biased when $G = 1$, but the single time profile is entirely unrepresentative of the latent group-level time profiles.

3.5.4 Behavioural interpretations

The following refers to the group time profiles in Figure 3.3. All the time profiles exhibit similar, slowly rising trajectories until the financial year ended 30 June 2008, after which they diverge in statistically and economically significant ways.

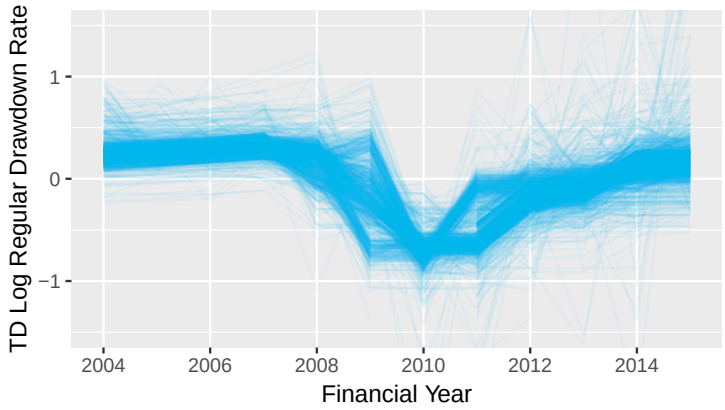
Figure 3.5: Point estimates and 95% confidence intervals for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 1$. The shaded regions (indistinguishable from point estimates in the plot) denote 95% element-wise confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula in the supplement to Bonhomme and Manresa (2015). The time-demeaned time profile is shifted to begin at 0 on the vertical axis. Owing to the modification of the GFE estimation algorithm, point estimates here are identical to those recovered from estimating a standard 2WFE model on the data. Appendix A gives the corresponding results using the unmodified GFE estimator as a comparison.



Consider group 5, whose time profile describes a rapid reduction in drawdown rates between financial years 2008 and 2010, and then a gradual return towards pre-reduction levels. The timing of these movements follows closely the progression of concessional minimum drawdown rates (see Section 3.3). This implies that members of group 5 were the most responsive to the changing minimum drawdown rates over this period, compared to the rest of the sample. Figure 3.6 confirms that many individuals in group 5 show a step-like pattern in their log regular drawdown rate series during the second half of the observation period. This pattern suggests use of the new step-function schedule for minimum drawdown rates which came into effect 1 July 2007.

Groups 1 and 2 display time profiles rising steadily following financial year 2008 for the remainder of the sample. Group 7 follows a similar rising trend with a reduced magnitude for the first few financial years after 2008, before stabilising for the remainder of the observation window. The panel plots in Figure 3.7 suggest a tendency for individuals in these groups to draw constant dollar amounts, while their account balances gradually decline over time. The corresponding plots for group 3 show a similar preference for constant dollar amounts, but after financial year 2013, individuals in this group suffer a rapid decline in both the amounts drawn down and account

Figure 3.6: Panel plot showing all individual time series for group 5 of the time-demeaned (TD) log regular drawdown rate variable vs. financial year.

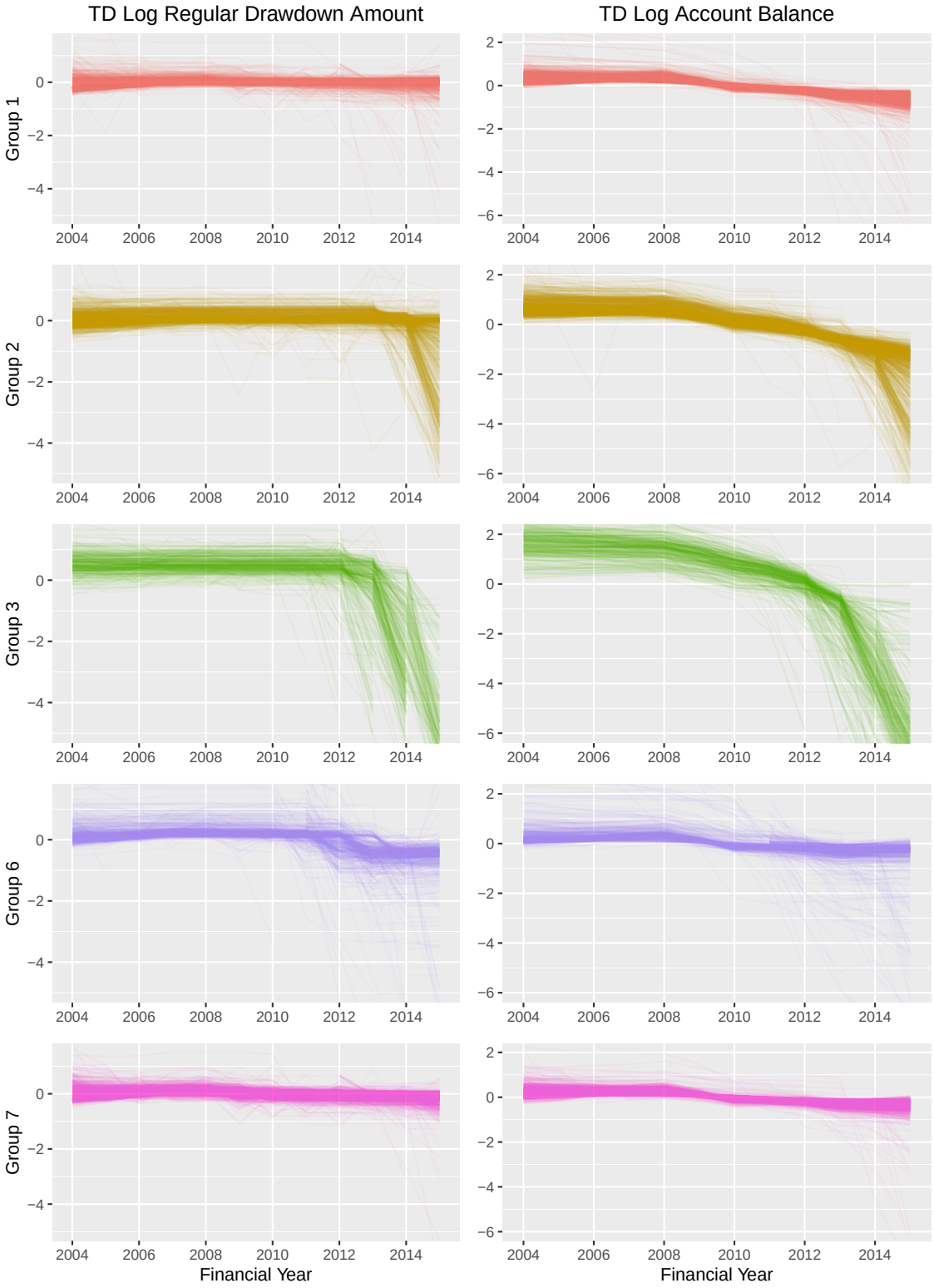


balances. Many members of group 6 undergo a similar evolution, characterised by mostly constant drawdowns initially and a subsequent reduction in the level in later years. The magnitudes of the downwards revisions appear smaller for this group, and occur earlier for many retirees. Moreover, while group 3 members continue to revise down over successive financial years, it appears that many individuals in group 6 revise down once and then continue to draw at the new, reduced level. The group 7 plots show gradual downward revisions in the amounts year on year.

Thus, groups 1, 2, 3, 6 and 7 appear to be similar in that members draw mostly constant dollar amounts over time; however, these groups seem to differ in how many members make a downwards revision in the level of their drawdown amounts, and whether this revision is once-off or the beginning of a downward trend. Crucially, the timing of downwards revisions often appears to align with periods where account balances begin falling at an accelerated rate. Hence, the heuristic of drawing constant amounts over time may adversely impact retiree financial security by contributing to a premature exhaustion of account balances.

Figure 3.3 shows that the time profile for group 4 closely follows the zero line before financial year 2011. A time profile with all values near zero suggests the model $y_{it} = x'_{it}\theta + c_i + v_{it}$ approximates the group's data well. Correspondingly, a possible interpretation for group 4's behaviour is that before financial year 2011, they were constantly tuning their drawdown rates as they progressively faced higher minimum drawdown rates and as

Figure 3.7: Panel plots showing all individual time series for groups 1, 2, 3, 6 and 7 (top to bottom) of the time-demeaned (TD) log regular drawdown amount (left column) and TD log account balance (right column) variables on the vertical axes vs. financial year on the horizontal axes. The plots in each column maintain the same y-axis scale for comparability.



their account balances changed. This potentially represents a group of 'engaged' retirees, who regularly occupy themselves with determining their desired rate of drawdown; however, it is impossible to further investigate this hypothesis using this dataset.

Group 4's time profile drops significantly below 0 after financial year 2010. These movements, while of a smaller magnitude compared to the time profile values of other groups, are still economically significant; they suggest a behavioural response, after netting out the effect of covariates, of reducing drawdown rates relative to earlier levels from 2011 onwards.

Appendix A provides summary statistics and descriptive plots for all seven groups. Based on these, some of the ways in which the groups differ in terms of observable characteristics are: the proportion of males in groups 1 to 3 are 62%, 62% and 69% respectively, while the sample on aggregate is 56% male; the members of groups 1 and 3 tend to be older, with median retirees aged 81 at 31 Dec 2015; members in groups 4, 5 and 6 have the youngest median retirees, aged 78 at 31 Dec 2015; group 5 members have the highest median risk appetite, a variable defined in Appendix A as a summary measure of the relative sizes of derived equity returns within individual accounts compared to the S&P/ASX 200 market index; group 7 members have the lowest median risk appetite; group 7 members make ad-hoc drawdowns least frequently, in roughly 3% of person–years observed, followed by group 1, for 5% of person–years observed—for all other groups this frequency was 9–11%; in years where they make ad-hoc drawdowns, group 3 members tend to draw down the largest proportions of their account balance using ad-hoc drawdowns over the course of a year—on average, these ad-hoc drawdowns over the year amount to 35% of their account balance at the start of the respective financial year.

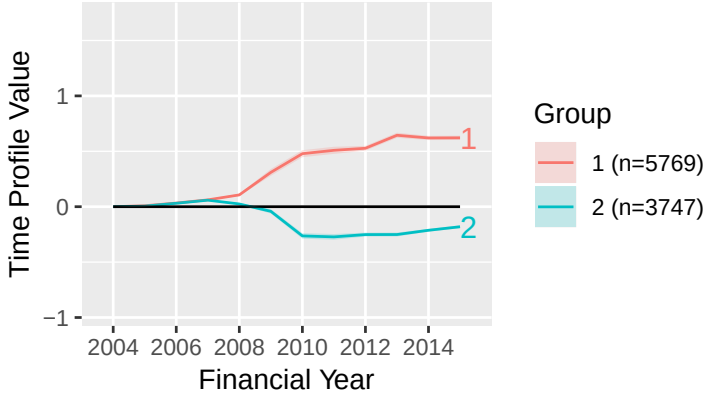
3.5.5 Prior expectations

Section 3.1 presents this study's prior expectations for finding at least two types of strategies in the data: drawing constant dollar amounts; following closely the minimum drawdown rates. Most of the group behaviours found under the seven-group assumption can be interpreted as cases of these two hypothesised behaviours. Moreover, a restricted

model assuming only two groups already begins to show evidence for the existence of both of these groups. Figure 3.8 shows the time profiles for the two-group model. As in Figure 3.5, the y-axis scale allows for direct comparison of the estimated magnitudes with the $G = 7$ model results. The time profile of the first group appears similar to those for the groups in the seven-group model that seem to target constant dollar amounts for the regular drawdowns; the second profile shows a dip comparable to the group that seems to follow the minimum drawdown rate.

Appendix A also provides descriptive panel plots for these two groups. These plots are broadly consistent with the interpretation that the first group often attempts to hold drawdown amounts constant, although with downward revisions in the amount common, as well as rapid declines in account balance towards the end of the period; the second group mainly makes decisions regarding their drawdown rate. This analysis, alongside the observation that moving to a two-group model already removes much of the bias in the covariate effect estimates, suggests that accounting for both rate-based and amount-based strategies seems to be the most important step in controlling for unobservable heterogeneity in the data.

Figure 3.8: Point estimates and 95% confidence intervals from the analytical formula for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 2$. The shaded regions (indistinguishable from point estimates in the plot) denote 95% element-wise confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula in the supplement to Bonhomme and Manresa (2015). The time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.



3.6 Discussion

3.6.1 Retirement incomes

A key implication for retirement incomes research and policy is that there is now a statistical basis for empirical results that indicate different behavioural stances explain much of the variation in the drawdowns from phased withdrawal retirement income products. The significant change in covariate effect estimates and the emergence of distinct time profiles, as the number of groups assumed by the GFE estimator increases, is evidence for the presence of multiple behavioural groups in the data. The magnitudes of the estimated group time profile values suggest a sizeable contribution of this behavioural component to the observed drawdowns.

Retirees need to trade-off current against future consumption in the face of investment, longevity and other, background, risks that include inflation rates and health expenditure (Yang & Huang, 2009). Retirement fund trustees and financial advisors could use the findings from this research when aiding retirees who are navigating the trade-offs in apportioning their accumulated retirement savings between consumption, income-generating assets and precautionary assets. The results do not find clear evidence for a group following the optimal heuristic strategy from Bateman and Thorp (2008), which is to initially draw at a fixed rate until the rising minima require drawing larger proportions. Two-thirds of the sample belong to the groups exhibiting a preference for drawing down a constant dollar amount over time. There is no evidence as to whether the initial level was optimal, but two of the groups targeting level income streams had their account balances begin a rapid decline following the GFC; retirees with this preference who are unable to respond to periods of low or negative investment returns by curtailing their income streams may be at risk of prematurely depleting their account balances. 14% of the sample belong to the group who most closely follow the next-most optimal heuristic strategy, which is to always draw at the legislated minimum rates.

Those providing financial advice may wish to convey these findings to retirees in the form of cautionary tales, or as part of a standard set of recommendations presented

to retirees. Practitioners and policymakers may also wish to consider these findings when designing retirement income products and policy to align with the Comprehensive Income Products for Retirement framework, which places an emphasis on generating income streams without compromising the ability to retain superannuation savings until late into retirement, by including some life annuity products (The Australian Government the Treasury, 2016). Financial advisors may benefit retirees by predicting their individual needs for income and precautionary savings later in life, monitoring their actual experience relative to this forecast, and recommending a reduction in expenditure during years of unfavourable investment returns, to avoid prematurely exhausting their accounts.

The present study clarifies how a robust behavioural analysis of drawdowns data is vital for informing the various stakeholders in the retirement incomes space, including retirees, the government, and the financial services industry. However, Section 3.3 notes that despite having data from multiple superannuation funds, due to the small number of funds in the estimation sample, the results may not describe the population of Australian retirees in general. In particular, using the current dataset may overlook behavioural patterns present in funds not included in the sample.

3.6.2 GFE in behavioural microeconomics and event studies

We believe that the GFE model is valuable for research in behavioural microeconomics and event studies. While this chapter's focus is on the behavioural interpretations of the latent group effects, the superannuation data covers a period in which multiple common macroeconomic and policy shocks affect all members in the sample. Hence, the superannuation application also invites an event-study interpretation, although it is impossible to disentangle the effects of several common shocks occurring in close proximity. These include the Global Financial Crisis, the introduction of a new schedule of minimum drawdown rates for account-based pensions, and temporary tweaks made to the minimum drawdown rates. Taken together, however, the time profile plots presented in Figure 3.3 clearly show how otherwise homogeneous-looking trends diverge radically, plausibly driven by one or more of these shocks. Hence, the results broadly show how researchers can utilise the same methodology in an event-study framework.

In the superannuation application, the GFE assumption of a group structure to the time-varying unobservable heterogeneity is tenable because it was *a priori* expected that an important source of heterogeneity in the drawdowns arises from the latent motivations that drive individuals to choose one of a finite number of drawdown strategies. In general, other behavioural economics studies with panel data may have analogous motivations to assume that there is some true, finite number of groups in the population, which the GFE estimator will attempt to characterise. This prior assumption may be important as it is unknown precisely how the estimator behaves when a group structure only approximates, rather than accurately captures, the nature of the unobservable heterogeneity. For a discussion on approximating more general forms of heterogeneity, see e.g. Bonhomme et al. (in press).

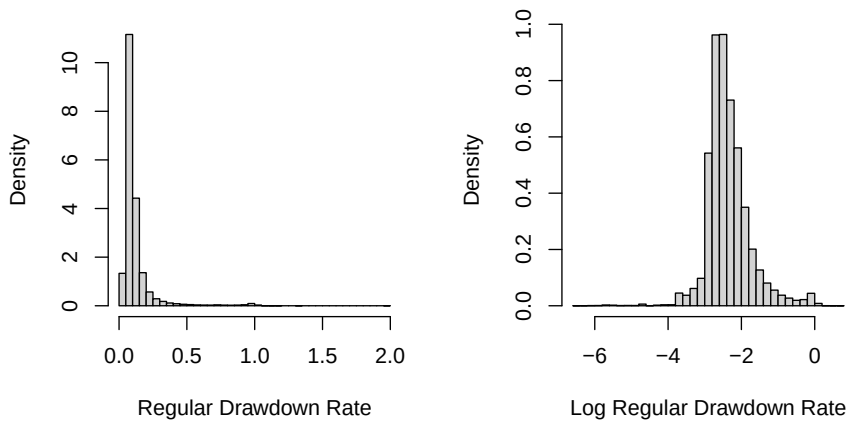
3.A Appendix: Account balance decomposition

The ABP account balance at 1 July of calendar year $t+1$ for an arbitrary retiree (i subscripts omitted) is given by

$$\mathrm{AB}_{t+1} = \mathrm{AB}_t - \mathrm{DA}_t - \mathrm{ADA}_t + \mathrm{NIRA}_t - \mathrm{AF}_t + \mathrm{NT}_t,$$

where:

Figure 3.9: Histograms of the regular drawdown rate variable and its logarithm in the superannuation data.



- t indexes financial years;
- AB_t is account balance at start of financial year t;
- DA_t is the regular drawdown amount. This is the total dollar amount, determined in advance of the financial year, received over the course of the financial year as regular payments; e.g., the retiree nominates, before a financial year starts, to receive \$1000, once a month, each month, giving $\mathrm{DA}_t = 12{,}000$;
- ADA_t is the ad-hoc drawdown amount. This is the total dollar value of any lump-sum commutations made throughout the year, not including the regular payment stream nominated;
- NIRA_t are net investment returns, in dollars, over financial year t;
- AF_t are account fees deducted;
- NT_t are net transfers, e.g., rollover amount from another super fund, or receipt of spouse's death benefit.

The dependent variable in the estimated model is $y_{it} := \ln \mathrm{DR}_{it}$, where $\mathrm{DR}_{it} = \mathrm{DA}_{it}/\mathrm{AB}_{it}$; $\mathrm{DA}_{it} \neq \mathrm{AB}_{it} - \mathrm{AB}_{i,t+1}$. Figure 3.9 shows histograms of DR_{it} and $\ln \mathrm{DR}_{it}$ for all non-missing unit–time combinations in the superannuation data.

Chapter 4

Nonparametric Bayesian Estimation of Latent Group Time Profiles in Linear Panel Data Models with Fixed Effects

Abstract

This chapter presents a nonparametric Bayesian estimation method for linear panel data models with latent group-level time-varying unobservable heterogeneity and individual-specific intercepts. The method is designed for applications to observational panel datasets that track many units over a short to moderate period of time. A simulation study shows that the method outperforms the closest classical alternative, the Grouped Fixed-Effects estimator, at estimating the time profiles and group allocations. An empirical application to retirement decisions using Australian panel data suggests retirees may be more likely to consider decumulation strategies in terms of dollar amounts, rather than drawdown rates.

4.1 Introduction

Latent group time profiles flexibly account for the effect of time-varying unobserved heterogeneity in panel data models. A variety of economic scenarios can provide a practical motivation for such grouped heterogeneity: in behavioural economics, for instance, one might assume the presence of multiple unknown subgroups in a population, each having distinct preferences that influence the values of an outcome variable over time. Alternatively, an event study can investigate whether common shocks, such as macroeconomic disruptions or policy changes, affect subgroups of observational units differentially. A data-driven motivation for grouped heterogeneity can be the existence of known and relevant, but unobserved, variables, whose values define subgroups with different patterns of time-varying heterogeneity.

A linear panel data model incorporating such unobserved heterogeneity is

$$y_{it} = c_i + x'_{it}\theta + \alpha^*_{g_i t} + u_{it}, \tag{4.1}$$

where $i = 1, 2, ..., N$ indexes cross-sectional units; $t = 1, 2, ..., T$ indexes time periods;[1] y_{it} is the value of the dependent variable; c_i is a unit-specific intercept, or 'individual fixed effect'; x_{it} is a column vector of size K containing covariate values; θ is a column vector of corresponding covariate effects; $\alpha^*_{g_i t}$ is the contemporaneous effect of group-specific time-varying heterogeneity, or 'group time profiles'; $g_i \in \{1, 2, ..., G\}$ is the group membership variable for unit i, with G the number of groups in the sample; u_{it} is an idiosyncratic error term with zero mean and unknown variance; errors can potentially be heteroskedastic across both groups and time.

In applications with economic motivations such as those listed above, performing statistical inference on the group-specific heterogeneity term may be of primary interest, in addition to inference on the covariate effects. For example, in an event study, while point estimates for the group time profile values, in particular around a time period when a common shock was experienced by all sampled individuals, may be useful for characterising different subgroups and determining the potential economic significance of

[1] Adapting the methodology in this chapter to accommodate unbalanced panels, with missing data for some unit–time combinations, would be a valuable extension.

the latent time effects, the researcher may also want to consider the uncertainty in these time effects, in order to assess whether there is sufficient statistical evidence to conclude differential responses by group.

To aid in the economic interpretation of the group time profiles, it is instructive to consider (4.1) as an extension of the two-way fixed effects (2WFE) model,

$$y_{it} = c_i + x_{it}'\theta + \alpha_t^* + u_{it}, \tag{4.2}$$

where the difference is that (4.1) allows for G heterogeneous group-specific time profiles.

A classical approach to estimating (4.1) that performs well even when T is moderate is the Grouped Fixed-Effects (GFE) estimator from Bonhomme and Manresa (2015). Some limitations of this approach are: the researcher must run the procedure for each potential value of G considered; for moderate T, estimating the true G requires subjective analysis of how covariate effects change for different input values of G; estimation becomes computationally difficult for larger G, especially for large N, which makes it hard to determine whether a given dataset supports an assumption that G is small or moderate.

In principle, a nonparametric Bayesian approach can overcome these limitations. Kim and Wang (2019) develop a nonparametric Bayesian method for a special case of (4.1) that does not include individual fixed effects;[2] i.e., they consider the model

$$y_{it} = x_{it}'\theta + \alpha_{g_it}^* + u_{it}. \tag{4.3}$$

Kim and Wang (2019) flexibly model the latent group time profiles using a Dirichlet Process Mixture Model (DPMM; see, e.g., Escobar & West, 1995; Ferguson, 1983). The grouped structure for the time profiles arises as a result of using a Dirichlet Process (DP; Ferguson, 1973) as the prior for the time-varying heterogeneity parameters. A realisation from a DP is a discrete probability distribution; N data points drawn from this discrete distribution have $G \leq N$ unique values, and given suitable DP parameters, G can be considerably smaller than N.

[2] If some time-invariant confounders are observed, and their partial effects are of economic interest, using the Kim and Wang (2019) method with these covariates included in x_{it} may be preferred, since the approach used in this chapter estimates θ only for time-varying covariates.

For large-N applications using observational data, such as typical microeconomic applications, omitting the individual fixed effects term, c_i, from panel data models is generally undesirable, due to an endogeneity concern from potential correlation between the unobserved c_i and the observed covariates, x_{it}. Moreover, large-N panel data may show considerable heterogeneity in the absolute levels of unit-level time series, even in applications where a comparatively small number of groups can capture the relative changes in the heterogeneity over time. These features motivate specifying two heterogeneity components: the c_i term modelling the absolute level of the heterogeneity, and the $\alpha^*_{g_it}$ term modelling changes over time. Hence, extending the Kim and Wang (2019) method to control for individual fixed effects is valuable as it allows estimating models with group time profiles using a nonparametric Bayesian approach for a broader class of applications.

The contributions of this chapter are: a) formulating (4.1) as a DPMM, implementing a Bayesian estimation methodology for this model, and proposing a method for performing inference on the group time profiles; b) studying the behaviour of the Bayesian estimators using simulated data from a data generating process (DGP) based on a motivating economic scenario for which the model is suitable, and comparing the results with the classical GFE alternative; c) presenting results that strengthen the claim made by existing empirical research that retirees consider decumulation strategies in terms of dollar amounts, rather than drawdown rates, obtained by applying the proposed methodology to Australian panel data on retirement decumulation.

This chapter has the following structure. Section 4.2 describes DPMMs, specifies the proposed Bayesian method for estimating (4.1), and proposes a method for inference on the time profiles. Section 4.3 presents the results of a simulation study, comparing the proposed method with the classical alternative. Section 4.4 applies the new method to panel data on retirement decumulation. Section 4.5 discusses the implications of the simulation study and empirical application results. Appendix 4.A contains derivations and implementation details for the proposed method. Appendix B contains additional results from the simulation study and empirical application.

4.2 Methodology

4.2.1 'BFG': The Bayesian First-differenced Grouped fixed-effects procedure

Consider the distribution of the dependent variable from the linear model (4.1) as $y_{it} \sim N\left(c_i + x'_{it}\theta + \alpha^*_{g_i t}, \sigma^{2*}_{g_i t}\right)$, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, where $y_{it} \sim N(m_y, v_y)$ denotes the random variable y_{it} having a Normal distribution with mean m_y and variance v_y, and where the idiosyncratic error variance, $\sigma^{2*}_{g_i t}$, allows for heteroskedastic errors across both groups and time. Modelling errors in this way allows the second moment of the dependent variable to experience heterogeneity across groups and time, in addition to the first-moment heterogeneity induced through the time profiles.

Stacking observations over time at the unit level leads to the equivalent representation, $y_i \sim MVN_T\left(c_i \iota_T + x_i\theta + \alpha^*_{g_i}, \Sigma^*_{g_i}\right)$, where $y_i \sim MVN_T\left(m_y, S_y\right)$ denotes the random length-T vector y_i having a multivariate Normal distribution with mean vector m_y and covariance matrix S_y; $\Sigma^*_{g_i} := \text{diag}\left(\sigma^{2*}_{g_i}\right)$; $\text{diag}\left(v\right)$ denotes the matrix with diagonal entries given by the corresponding entries of some vector v and zeros elsewhere; $\sigma^{2*}_{g_i} := \left(\sigma^{2*}_{g_i 1}, \sigma^{2*}_{g_i 2}, ..., \sigma^{2*}_{g_i T}\right)'$; $\alpha^*_{g_i} := \left(\alpha^*_{g_i 1}, \alpha^*_{g_i 2}, ..., \alpha^*_{g_i T}\right)'$; $y_i := (y_{i1}, y_{i2}, ..., y_{iT})'$; $x_i := [x'_{i1}, x'_{i2}, ..., x'_{iT}]'$; ι_T is a length-T column vector of ones.

Defining D as the $(T-1) \times T$ first-differencing matrix, with typical row $[0, ..., 0, -1, 1, 0, ..., 0]$, and premultiplying the unit-level model by D, produces $Dy_i \sim MVN_{T-1}\left(D\alpha^*_{g_i} + Dx_i\theta, D\Sigma^*_{g_i}D'\right)$; i.e., first-differencing removes the c_i term from the model. Condensing notation gives the equivalent form, $\overset{\Delta}{y}_i \sim MVN_{T-1}\left(\overset{\Delta^*}{\alpha}_{g_i} + \overset{\Delta}{x}_i\theta, D\Sigma^*_{g_i}D'\right)$, where $\overset{\Delta}{y}_i := \left(\overset{\Delta}{y}_{i2}, \overset{\Delta}{y}_{i3}, ..., \overset{\Delta}{y}_{iT}\right)'$, and $\overset{\Delta}{y}_{it} := y_{it} - y_{i,t-1}$ for $t = 2, 3, ..., T$. The motivation for first-differencing originates from the observation in Lancaster (2002) that in the linear panel data model, Bayesian inference using the first-differenced data likelihood is equivalent to inference using the classical 'fixed effects' estimator for large N, given a diffuse prior for the fixed effects.

The conditional distribution of the observed data is

$$\overset{\Delta}{y}_i \,|\, \theta, \overset{\Delta^*}{\alpha}_{g_i}, \delta^*_{g_i} \sim MVN_{T-1}\left(\overset{\Delta^*}{\alpha}_{g_i} + \overset{\Delta}{x}_i\theta, D\Sigma^*_{g_i}D'\right), \tag{4.4}$$

independently for $i = 1, 2, ..., N$, with $\sigma_{g_i t}^{2*} := \exp\left(\delta_{g_i t}^*\right)$; $\exp(v) := e^v$;
$\delta_{g_i}^* := \left(\delta_{g_i 1}^*, \delta_{g_i 2}^*, ..., \delta_{g_i T}^*\right)'$.

The priors, using the stick-breaking representation of the DP prior (see, e.g., Sethuraman, 1994), are:

- $\overset{\Delta}{\alpha}_g^* \sim MVN_{T-1}(m_\alpha, S_\alpha)$, for $g = 1, 2, ...$;
- $\delta_g^* \sim MVN_T(m_\delta, S_\delta)$, for $g = 1, 2, ...$;
- $g_i \,|\, \pi \sim \sum_{g=1}^{\infty} \pi_g \delta(g)$, for $i = 1, 2, ..., N$, with $\pi := (\pi_1, \pi_2, ...)$ and $\delta(g)$ being the degenerate discrete measure with unit mass at value g; $p(g \,|\, V) := \Pr(g_i = g \,|\, V) = \pi_g$, for $i = 1, 2, ..., N$ and $g = 1, 2, ...$, with $V := (V_1, V_2, ...)$;
- $\pi_g = V_g \prod_{j=1}^{g-1}(1 - V_j)$, for $g = 2, 3, ...$, and $\pi_1 = V_1$;
- $V_g \,|\, \beta \sim \text{Beta}(1, \beta)$, for $g = 1, 2, ...$;
- $\beta \sim \text{Gamma}(a_\beta, b_\beta)$. Escobar and West (1995) recommend treating the DP prior concentration parameter β as random, rather than fixed.[3] $\text{Gamma}(a, b)$ denotes the distribution with density function $\text{Gamma}(v; a, b) \propto b^a v^{a-1} e^{-bv}$;
- $p(\theta) = \prod_{k=1}^{K} p(\theta_k)$, with $p(\theta_k) \propto 1$ to express a lack of prior knowledge regarding the true distribution of the covariate effects.[4]

Selection of the fixed prior parameters is detailed in Appendix 4.A. Priors for the time profiles are centered around standard panel regression estimates for a one-group model, using large variance values compared to the magnitudes of the estimates. The prior parameters for β are chosen to shrink the expected number of groups under the prior.

The input data comprise the first-differenced series of dependent variables, $\overset{\Delta}{y} := \left\{\overset{\Delta}{y}_1, \overset{\Delta}{y}_2, ..., \overset{\Delta}{y}_N\right\}$, and observed covariates, $\overset{\Delta}{x} := \left\{\overset{\Delta}{x}_1, \overset{\Delta}{x}_2, ..., \overset{\Delta}{x}_N\right\}$.

A Gibbs sampler (see, e.g., Gelman et al., 2013, pp. 276–278) explores the joint posterior distribution, implementing the slice sampler from Kalli et al. (2011), which improves

[3] Kim and Wang (2019) use this prior also.

[4] Changing the prior on the covariate effects to a proper multivariate Normal distribution requires only minor modifications to the sampler derived in Appendix 4.A.

the original DPMM slice sampler in Walker (2007).[5] The slice sampler avoids the need to truncate the number of stick lengths to some fixed value, while affording easier implementation than the 'retrospective' sampling approach of Papaspiliopoulos and Roberts (2008).

The sampler incorporates three label-switching steps to facilitate exploration of multiple modes in the posterior distribution of group probabilities, π_g: the two proposed by Papaspiliopoulos and Roberts (2008) and used in Kim and Wang (2019), as well as a third proposed by Hastie et al. (2015).

Appendix 4.A details the Gibbs sampling algorithm and initialisation. All parameter blocks are sampled directly from their respective conditional posteriors except the δ_g^* parameters, which determine the idiosyncratic error variances. The sampler uses a tailored Metropolis–Hastings (MH; Hastings, 1970) step for this block; see Appendix 4.A.2 for details.

Posterior inference for parameters that do not vary across groups proceeds by examining the empirical distribution or sample moments of the respective set of iterates (parameter values) drawn across iterations of the Gibbs sampler; the next section discusses inference for the group time profiles.

4.2.2 Time profile inference with the BFG method

In each iteration m of the sampler, individual unit time profiles, $\overset{\Delta}{\alpha}_i = \left(\overset{\Delta}{\alpha}_{i2}, \overset{\Delta}{\alpha}_{i3}, ..., \overset{\Delta}{\alpha}_{iT}\right)'$, are grouped into $G^{(m)} \leq N$ groups, so that $\overset{\Delta}{\alpha}_i^{(m)} = \overset{\Delta}{\alpha}_{g_i}^{*(m)}$; $\phi^{(m)}$ denotes the sampled value for an arbitrary parameter ϕ in the m-th iteration. However, the number of groups, and the group identities of each unit, may vary across iterations, as the sampler explores the joint posterior. This feature implies that while the sampler output allows identification of individual-level first-differenced time profiles, $\overset{\Delta}{\alpha}_i$, the group-level parameters, $\overset{\Delta}{\alpha}_g^*$, are not clearly defined outside of any single iteration, making it challenging to perform inference on group-level parameters.

[5] Here, the 'Gibbs sampler' is actually a Metropolis-within-Gibbs sampler, meaning a Gibbs sampler in which at least one step involves Metropolis–Hastings sampling.

While Kim and Wang (2019) propose a heuristic method for computing group time profile point estimates based on the individual-level iterate series, they do not consider quantifying the uncertainty around these point estimates, unlike the method proposed below. Furthermore, they do not test the accuracy of their method for getting point estimates, e.g., by examining estimator bias and root mean squared error, which is possible with a simulation study such as the one reported in Section 4.3.

The proposed method for time profile inference with the BFG model uses the following three steps, with details following.

1. Determine a 'final' grouping of the units; i.e., convert the multiple clusterings generated by the sampler into a single, post-processed allocation of the individuals into distinct groups.

2. For each final group, select an 'exemplar' unit, which is assumed to be representative of how members in that group behave.

3. Infer properties, e.g., posterior means and 95% credible intervals, about group time profiles by summarising the respective exemplar unit time profile iterates.

Step 1. Given the output of M iterations of the sampler, the objective is to determine a single clustering of the N units into $G \ll N$ groups. One possibility is to partition the units by performing k-means clustering (Forgy, 1965) on the N estimated posterior means of the unit-level time profiles. Kim and Wang (2019) use this approach as an intermediate step in their method for obtaining point estimates of the time profiles: they cluster the sample into G_{pmax} groups, with G_{pmax} being an estimate for the number of groups in the sample computed by finding the modal number of unique, non-empty groups across all iterations.

An alternative is to use the 'PEAR' method from Fritsch and Ickstadt (2009), implemented in the `R` package `mcclust`, which estimates the group allocation for which the posterior expectation of the adjusted Rand index (Hubert & Arabie, 1985; Rand, 1971) is maximised. For a recent application of the PEAR method with DPMMs, see, e.g., Chin, Lee, Ryan, Kohn, and Sisson (2019).

While both methods perform well in the simulation study, researchers may wish to substitute alternative post-processing methods; the remaining steps are agnostic to the specific method used here.

Step 2. Given the output of step 1, this step proposes to find, for each final group, the unit that, across the iterations, is most often grouped with other units from the same final group. Call this the exemplar unit, in the sense that they are assumed to be a typical example of how that group behaves.

This method for finding exemplar units is in part motivated by the method Kim and Wang (2019) use to find 'benchmark' units in the context of estimating group allocation probabilities, which has broadly the same rationale, although the implementation differs. Examining the Kim and Wang (2019) replication code reveals that their benchmark unit procedure finds, for each final group, the unit that, across the iterations, is most likely to not be grouped with any other arbitrary unit in the sample.

The following step is agnostic to how the prototypical unit for each group is selected, and so alternative methods could be considered here also.

Step 3. Given the exemplar units, the proposed method infers properties for the time profiles by computing sample statistics on the set of time profile iterates for each exemplar unit. Economically, the first-differenced time profiles, $\overset{\Delta}{\alpha}_i$, may be of less interest than 'relative' time profiles, defined as $\tilde{\alpha}_{it} = \sum_{s=2}^{t} \overset{\Delta}{\alpha}_{is}$ for $t = 2, 3, ..., T$; $\tilde{\alpha}_{it} := \alpha_{it} - \alpha_{i,1}$. The relative time profiles, $\tilde{\alpha}_{it}$, represent the change in the time profile value relative to the first time period, which is a similar interpretation to that of the coefficients on the $T-1$ time dummy variables included in a standard 2WFE model.

In this chapter, interest is primarily in: a) the posterior mean estimate $M^{-1} \sum_{m=1}^{M} \tilde{\alpha}_{it}^{(m)}$ for $t = 1, 2, ..., T$ and $i \in \mathcal{I}_e$, with $\mathcal{I}_e$ being the set of unit indices for the exemplar units, and $m = 1, 2, ..., M$ indexing selected iterations of the sampler; b) the 95% credible interval estimate, obtained by finding empirical 2.5 and 97.5 percentiles of the iterates $\left\{\tilde{\alpha}_{it}^{(m)}\right\}$.

Two alternative estimators for the group effects are also considered: a simple averaging approach and a weighted averaging approach. The simple averaging approach computes, for each iteration, the group-level averages of the individual-level time profile iterates, with equal weight given to each unit when computing the averages. The group-level averages are then considered to represent iterates of the latent group effects, and sample statistics are used to estimate posterior means and credible intervals. The weighted average approach is similar, but differs in that each unit is assigned a weight proportional to the number of iterations in which that unit is clustered with its final group's exemplar unit, so that units less strongly associated with the group are down-weighted. Both of these alternatives consistently performed worse than the exemplar unit approach in the simulation study, and so are not considered further, although they both outperform the GFE estimator in all metrics considered. Appendix B provides simulation results for these two alternative approaches that can be compared with the results presented below.

4.3 Simulation Study

The first goal of this simulation study is to see whether the proposed BFG estimation procedure can identify and estimate covariate effects and time profiles reasonably well on an example of the kind of hypothesised economic model that would motivate its use: models where multiple latent subgroups, which vary in their behaviour through the values of time profiles, are plausible.

The second goal is to compare the BFG method with the classical GFE alternative from Bonhomme and Manresa (2015) on the same sets of simulated data. Although the GFE estimator does not use the exemplar unit method for inference, both the BFG and GFE approaches provide point estimates and measures of estimator uncertainty, so the two methods can be directly compared using these quantities. Comparisons consider estimator bias, root mean squared error (RMSE) and 95% credible/confidence interval coverage probabilities.

Estimator bias for an arbitrary parameter ϕ is estimated using $\widehat{\text{Bias}}\left(\widehat{\phi}\right) := S^{-1}\sum_{s=1}^{S}\left(\widehat{\phi}_s - \phi^0\right) = \left[S^{-1}\sum_{s=1}^{S}\widehat{\phi}_s\right] - \phi^0$. For the BFG method, $\hat{\phi}_s$ is

the posterior mean estimate; i.e., $\hat{\phi}_s$ is the sample average of iterates for parameter ϕ in replicate s. For the GFE estimator, $\hat{\phi}_s$ is the point estimate returned by the estimation procedure. ϕ^0 is the true parameter value used in the data generating process (DGP).

RMSE is estimated using $\widehat{\text{RMSE}}\left(\widehat{\phi}\right) = \left[S^{-1}\sum_{s=1}^{S}\left(\widehat{\phi}_s - \phi^0\right)^2\right]^{1/2}$.

Interval coverage probabilities are estimated as the proportion of replicates in which the 95% credible or confidence intervals contain the true parameter value.

The results from the simulation study suggest that the BFG and GFE alternatives perform similarly at estimating covariate effects, but that the BFG method outperforms the GFE approach at estimating the unknown group allocation and the time profiles.

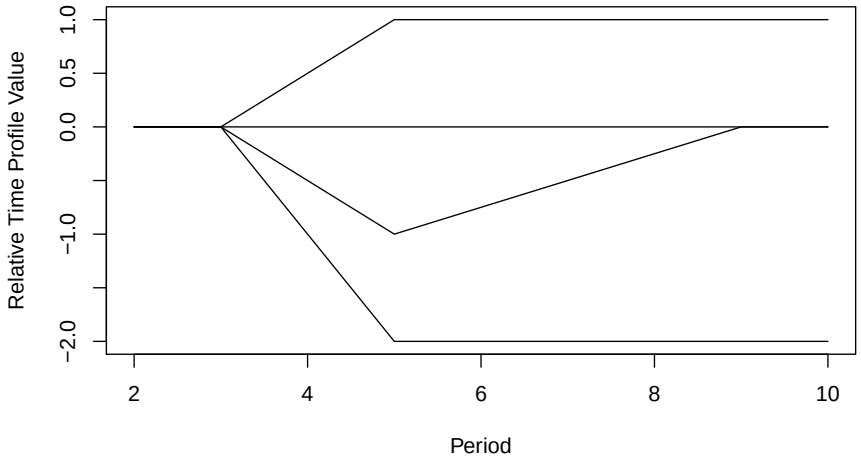
4.3.1 Design

$S = 500$ independently simulated datasets are generated using the DGP governed by (4.1). The true number of equally sized groups is $G = 4$, and there are $T = 10$ time periods observed for $N = 1000$ units. Individual-specific intercepts follow $c_i \sim N(0,1) + 1$ independently for $i = 1, 2, ..., 1000$, and each of the 500 replicates generates an independent set of intercepts.

The group time profiles, α_g^*, are generated so that relative time profiles have the shapes given in Figure 4.1. These shapes are motivated by an event study-style scenario in which the latent group time profiles diverge after a common shock at $t = 3$. At $t = 5$, the potential impact of the shock on the mean of the dependent variable attains its maximum magnitude, with the effect being permanent for two groups, and temporary for one group. There is a group with no latent effect (the 'no-change' group); two groups respond in the downward direction, which in this case is the intended policy effect, with the effect being temporary for one group ('down-up'), and persistent for the other ('down-flat'); one group shows an unintended consequence of the policy, moving in the opposite direction intended ('up-flat').

There are three covariates, $x_{it} = (x_{1,it}, x_{2,it}, z_{it})'$, for $i = 1, 2, ..., 1000$ and $t = 1, 2, ..., 10$. The $x_{1,it}$ are independently distributed as $N(0,1) + 0.3c_i + 0.3\alpha^*_{g_i t}$, while the $x_{2,it}$ follow $N(0,1) - 0.3c_i - 0.3\alpha^*_{g_i t}$. This specification induces correlation between the observable and

Figure 4.1: Time profiles for the simulation study. The values are relative to period 1.



unobservable determinants of the dependent variable, y_{it}, so that failing to account for either individual-specific intercepts or time-varying group-specific heterogeneity results in an endogeneity problem. The third covariate, z_{it}, is uncorrelated with c_i and $\alpha^*_{g_i t}$; $z_{it} \sim N(0,1)$ independently. The covariate effects are $\theta = (1,1,1)'$.

The model errors are heteroskedastic and independent, $u_{it} \sim N\big(0, \sigma^{2*}_{g_i t}\big)$. Error standard deviations change both over time and across groups, either starting at $(\sigma^{2*}_{g,1})^{1/2} = 0.3$ and rising linearly to $(\sigma^{2*}_{g,T})^{1/2} = 0.9$ over time, or starting at 0.9 and falling linearly to 0.3. The up-flat and no-change groups have rising error variances; the down-up and down-flat groups falling. The BFG method is designed to be used even in the presence of heteroskedastic errors, as this feature is generally present in microeconomic observational panel data.

To estimate the model parameters using the BFG method, for each simulated dataset, the Gibbs sampler is run for a total of 20,000 iterations, with the first 10,000 discarded as burn-in. To speed up the post-processing, the remaining iterations are thinned, keeping every 10th iteration. The realised parameter values in the resulting set of $M = 1000$ iterations are treated as a sample from the joint posterior, and are used to infer properties of the marginal posterior distributions for parameters of interest—in particular, posterior means and 95% credible intervals—using sample statistics, such as sample means and empirical quantiles.

The GFE results use the implementation described in Chapter 3; this implementation of the GFE estimator computes the fixed-T variance–covariance matrix estimate to construct element-wise 95% confidence intervals for the covariate effects and the relative time profiles. Results take the most optimal solution across 10,000 independent runs of the procedure, each initialised using independently generated random starting values.

Appendix B reports the results from applying a third estimator to each simulated dataset, referred to as the Oracle estimator. The Oracle is an (infeasible) frequentist 2WFE estimator with multiple time profiles, estimated using known group membership, by interacting the true group identity dummies with $T-1$ time dummies, omitting the first period.

4.3.2 Results

BFG group identification accuracy is similar for the two proposed post-processing techniques, and both outperform the GFE grouping. The PEAR clustering accuracy is marginally superior to the k-means approach. Table 4.1 gives the mean accuracy, and empirical 2.5 and 97.5 percentiles, across replications. Computing group allocation accuracy requires matching the arbitrarily labelled replication-level groupings with the true DGP labels. The GFE and BFG k-means clustering results follow the label-matching procedure in Chapter 3, which finds, for each set of estimated time profiles, the permutation of group labels that minimises the sum of Euclidean distances between the estimated time profiles and a common set of reference time profiles with fixed labels; here, the reference time profiles are the DGP profiles.[6] The BFG PEAR clustering results use the permutation of group labels that maximise the permutation-specific group classification accuracy with respect to the DGP groupings.

The BFG time profile estimators outperform the GFE estimators, having smaller bias and RMSE estimates, and interval coverage closer to nominal values. The BFG time profile estimators follow the exemplar unit approach described in Section 4.2.2. Figures 4.2a and 4.2b summarise the bias estimates using the same heatmap scale, Figures 4.2c and 4.2d are the RMSE estimates, and Figures 4.2e and 4.2f are the interval coverage

[6] This label-matching procedure adapts a similar method used by Hofmans et al. (2015) for matching k-means clustering results across independent bootstrap replications.

Table 4.1: Simulation study group identification accuracy comparison.

	Mean Accuracy	2.5 Percentile	97.5 Percentile
BFG: PEAR	0.919	0.901	0.935
BFG: k-means	0.912	0.892	0.929
GFE	0.781	0.705	0.824

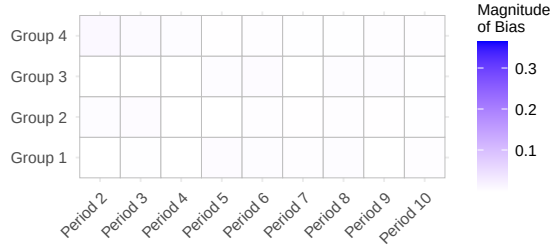
Note: Mean accuracy is the mean proportion of units correctly grouped across replications. 2.5 and 97.5 percentiles are the corresponding empirical percentiles of the group allocation accuracy across replications.

probabilities. These comparisons also require matching labels between the DGP time profiles and the estimated BFG and GFE time profiles from each replicate.

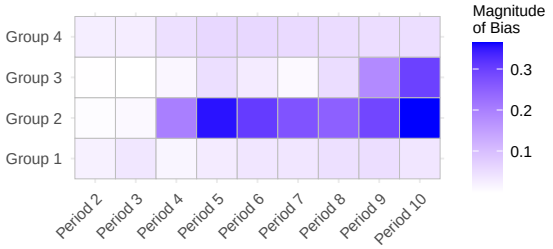
Appendix B contains additional figures which corroborate the summary result that BFG time profile estimators outperform those from the GFE approach, by examining time profile estimates for randomly selected replicates.

Figure 4.2: BFG exemplar unit (left column) and GFE (right column) time profile estimated bias (top row), RMSE (middle row) and 95% credible/confidence interval coverage probabilities (last row). The time profile estimators are identified relative to period 1. For the last row, the heatmap is white at the nominal coverage probability of 0.95.

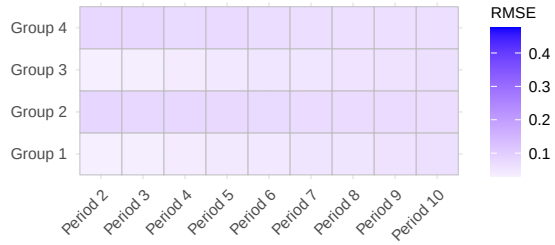
(a) BFG bias



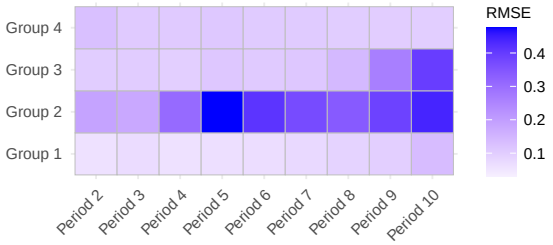
(b) GFE bias



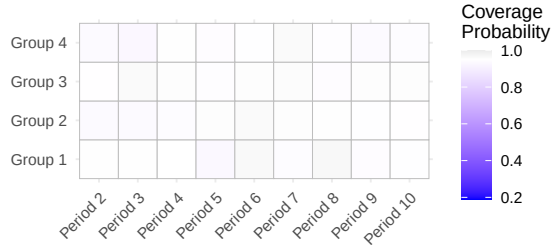
(c) BFG RMSE



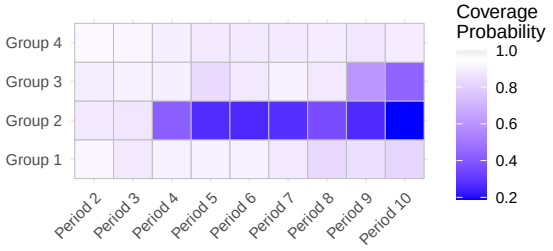
(d) GFE RMSE



(e) BFG interval coverage probability



(f) GFE interval coverage probability



The above comparisons assume that the GFE estimator takes as input the correct number of latent groups. Selecting the number of groups, G, for the GFE estimator when T is moderate is challenging; the method in Bonhomme and Manresa (2015) for selecting G in their application involves a subjective judgement based on how covariate effects vary as G increases from an initial value of 1 to some chosen maximum value. Thus, in general, it is possible to pick the wrong number of groups when using the GFE estimator; similarly with the BFG method, the posterior mode number of groups, G_{pmax}, may not equal the true number of groups. In this simulation study, $G_{pmax} = 4$ for 464 (92.8%) of the 500 replicate datasets, with only these replicates factoring into the reported results; $G_{pmax} = 5$ in 35 (7%) of the replicates; $G_{pmax} = 6$ in 1 (0.2%) of the replicates.[7] On the other hand, the PEAR groupings consistently generate four large clusters—together comprising, on average, 99.5% of the sample—and tend to place the remaining units into groups containing only one unit (singletons). The 2.5 and 97.5 percentiles of the proportion of units in the four largest PEAR groups across the 500 replicates are 98.9% and 99.9%, respectively. For the GFE estimator, it is unclear how to quantify the probability of choosing the wrong value of G. Examining the GFE results for various values of G from the first three simulated datasets reveals that the plots of covariate effects versus G arguably support $G = 3$ for all three datasets, so underestimation of the true number of groups is a risk.

Covariate effect estimators perform similarly between the BFG and GFE methods in their estimated bias, RMSE and interval coverage probabilities; Appendix B provides details.

4.4 Application: Retirement decumulation behaviours in Australian superannuation funds

This section uses the BFG method to analyse observational panel data on drawdowns (withdrawals) from account-based pensions, a phased withdrawal retirement income product in Australia in which drawdowns are subject to age-dependent minimum

[7] The BFG method run on 100 datasets simulated from a similar DGP but with $G = 1$ finds $G_{pmax} = 1$ in 95 of the replicates.

drawdown rates. Existing theoretical literature (e.g., Bateman & Thorp, 2008) motivates a prior assumption for grouped behavioural effects in the drawdown rates, which could manifest through the time-varying unobservable heterogeneity component of a linear panel data model. Chapter 3 finds some evidence for such latent groups effects with the GFE estimator of Bonhomme and Manresa (2015).

Examining the same dataset using the BFG method is motivated by: a) the simulation evidence for improved clustering and time profile estimation compared to the GFE alternative; b) the ability to examine the validity of the assumption that there are a small number of distinct groups, G, in the sample. The BFG method can more feasibly do the latter compared to the GFE estimator, as the BFG approach considers a posterior distribution over the number of groups in a single estimation, rather than comparing the output of multiple estimations with different input values of G, as is required by the GFE estimator. Also, estimating a model using the GFE estimator quickly becomes computationally difficult as G increases, which, in practice, limits the largest value of G that can feasibly be considered.

The dataset used here is a subset of the data considered in Chapter 3, which observes $T = 12$ years of drawdown experience for retirees from multiple large industry and retail superannuation funds. A balanced subsample, first removing individuals with missing observations then randomly selecting $N = 1000$ retirees, is taken to satisfy the computational demands of our `R` implementation of the BFG method;[8] in Chapter 3, robustness checks confirm that using a fully balanced subsample of their main estimation sample does not materially change the economic interpretation of the results. Table 4.2 reports summary statistics on the sample used for BFG estimation.

As in Chapter 3, the present study considers 'regular' drawdowns, which are determined before the relevant financial year; 'ad-hoc' drawdowns, which are supplementary lump-sum commutations taken during a financial year, are not considered here.

The estimated model is

$$\ln \mathrm{DR}_{it} = c_i + \theta_1 \ln \mathrm{MDR}_{it} + \theta_2 \ln \mathrm{AB}_{it} + \alpha^*_{g_i t} + u_{it}, \tag{4.5}$$

[8] A more efficient code implementation may allow estimation of models on $N > 1000$ observational units, but larger samples prevent obtaining results in a reasonable timeframe using the current implementation.

Table 4.2: Summary statistics on the superannuation application sample.

	Drawdown Rate	Drawdown Amount	Account Balance	Age at 31 Dec 2015	Age at Account Open	Sex: Male
Mean	0.11	6249.23	70,118.70	80.22	63.55	0.59
SD	0.11	5915.88	74,195.34	4.37	4.06	0.49
Median	0.08	4656	51,046	80.38	64.34	1
Q1	0.07	2976	31,366	77.55	60.99	0
Q3	0.12	7428	82,908	83.32	65.43	1
Min	0.01	1	1	67.67	50.85	0
Max	1.33	84,422	1,274,233	92.5	78.32	1
Count	12,000	12,000	12,000	1000	1000	1000

Note: The log of the drawdown rate is the dependent variable. The account balance is defined as the position at the start of the relevant financial year; this is the denominator for the rate computations. The age at 31 December 2015 represents the individual's cohort, equivalent to measuring a year-of-birth variable. The age at account opening is the age when the retiree initiates a phased withdrawal product and begins drawing down from the account. The sex indicator variable equals 1 if the retiree is male, and 0 otherwise. SD refers to the sample standard deviation of the variable. Q1 and Q3 refer to the empirical 1st and 3rd quartiles, respectively.

where DR_{it} is the regular drawdown rate for individual i for financial year t; MDR_{it} is the corresponding minimum drawdown rate, computed based on the age of the retiree, following Schedule 7 of the Superannuation Industry (Supervision) Regulations 1994; AB_{it} is the individual's account balance at the start of financial year t; the c_i are individual-specific intercepts; $\alpha^*_{g_i t}$ represents time-varying group heterogeneity, the shapes of which form the time profiles of interest; θ_1 and θ_2 are partial effects of the covariates; u_{it} is the model error, assumed to follow a zero-mean Normal distribution in which the error variance varies across both groups and time.

A feature of model (4.5) is that there exist two particular covariate effect configurations, $(\theta_1, \theta_2) = (0, -1)$ and $(1, 0)$, reducing the model to different equations where the time profiles have unique economic interpretations. Substituting $(\theta_1, \theta_2) = (0, -1)$, model (4.5) becomes

$$\ln \text{DA}_{it} = c_i + \alpha^*_{g_i t} + u_{it}, \tag{4.6}$$

using $\text{DR}_{it} = \text{DA}_{it}/\text{AB}_{it}$, where DA_{it} is regular drawdown amount for individual i for financial year t. Hence, while the time profiles in the base specification (4.5) are interpreted in terms of drawdown rates, time profiles in the reduced model (4.6) are interpreted as the effect of unobserved heterogeneity on drawdown amounts, not drawdown rates.

Conversely, substituting $(\theta_1, \theta_2) = (1, 0)$, model (4.5) becomes

$$\ln\left(\frac{\text{DR}}{\text{MDR}}\right)_{it} = c_i + \alpha^*_{g_i t} + u_{it}. \tag{4.7}$$

This equation represents a model for proportional changes in the drawdown rate relative to the minima, which gives the time profiles an interpretation as proportional changes in the multiplicative factor by which drawdowns exceed the minima; more loosely, the time profiles represent the extent to which individuals draw above the minimum drawdown rates.

To observe which regions of the posterior distribution the sampler explores when no particular covariate effect configuration is targeted, the BFG method is run five times, using random starting values independently generated as described in Appendix 4.A. Each run generates 120,000 iterations, with the first 90,000 discarded as burn-in.[9] The remaining iterations are thinned, keeping every 30th iteration. Realised parameter values from the remaining $M = 1000$ iterations are used to infer properties of the marginal posterior distributions for parameters of interest.

All five runs locate modes at either one of the special covariate effect configurations and then stay close to the identified mode, making only a local exploration of the posterior near the mode. For four of the five independent runs, the covariate effect posterior means are estimated as $\left(\widehat{\theta}_1, \widehat{\theta}_2\right) = (0, -1)$ to 6 decimal places, with the widest sets of 95% credible intervals (CIs) given by $0 + \left[-3.6 \times 10^{-7}, 7.3 \times 10^{-7}\right]$ and $-1 + \left[-8.2 \times 10^{-7}, 8.3 \times 10^{-7}\right]$, respectively. Only one of the five runs estimates the covariate effect configuration close to $(1, 0)$, with $\left(\widehat{\theta}_1, \widehat{\theta}_2\right) = (0.977, -0.006)$, and 95% CIs given by $[0.967, 0.989]$ and $[-0.007, -0.004]$.

Since the economic interpretation of the time profiles is different depending on the mode, the finding that no single run mixes across both the modes is convenient for interpreting the results, as the results from a sampler that visited both modes would have a less clear interpretation of the time profiles. However, the lack of mixing does require selecting a preferred mode. The Widely Applicable Information Criterion

[9] Convergence to a stable mean for the number of groups in each iteration, $G^{(m)}$, was assessed visually to select the end point of the burn-in period.

(WAIC; Watanabe, 2010) favours results from the runs that locate the mode at $(0, -1)$, which have lower WAIC values than the run locating the mode near $(1, 0)$.[10]

For the $(0, -1)$ configuration results with the best fit according to WAIC, estimating the number of groups by taking the posterior mode number of groups across iterations, G_{pmax}, or alternatively, taking the number of groups in the optimal PEAR clustering, G_{pear}, gives $G_{pmax} = G_{pear} = 27$. Group sizes are highly skewed, with the 6 largest groups covering 83% of the sample; the remaining groups each have less than 3% of the sample; 11 groups contain 5 or fewer individuals and there are 6 singletons. Contrastingly, for the $(1, 0)$ configuration run, the estimated model is considerably less parsimonious, with $G_{pmax} = 34$ and $G_{pear} = 40$. It takes the 17 largest groups to cover 80% of the sample, primarily because there are no large groups, with the largest group comprising less than 10% of the sample.

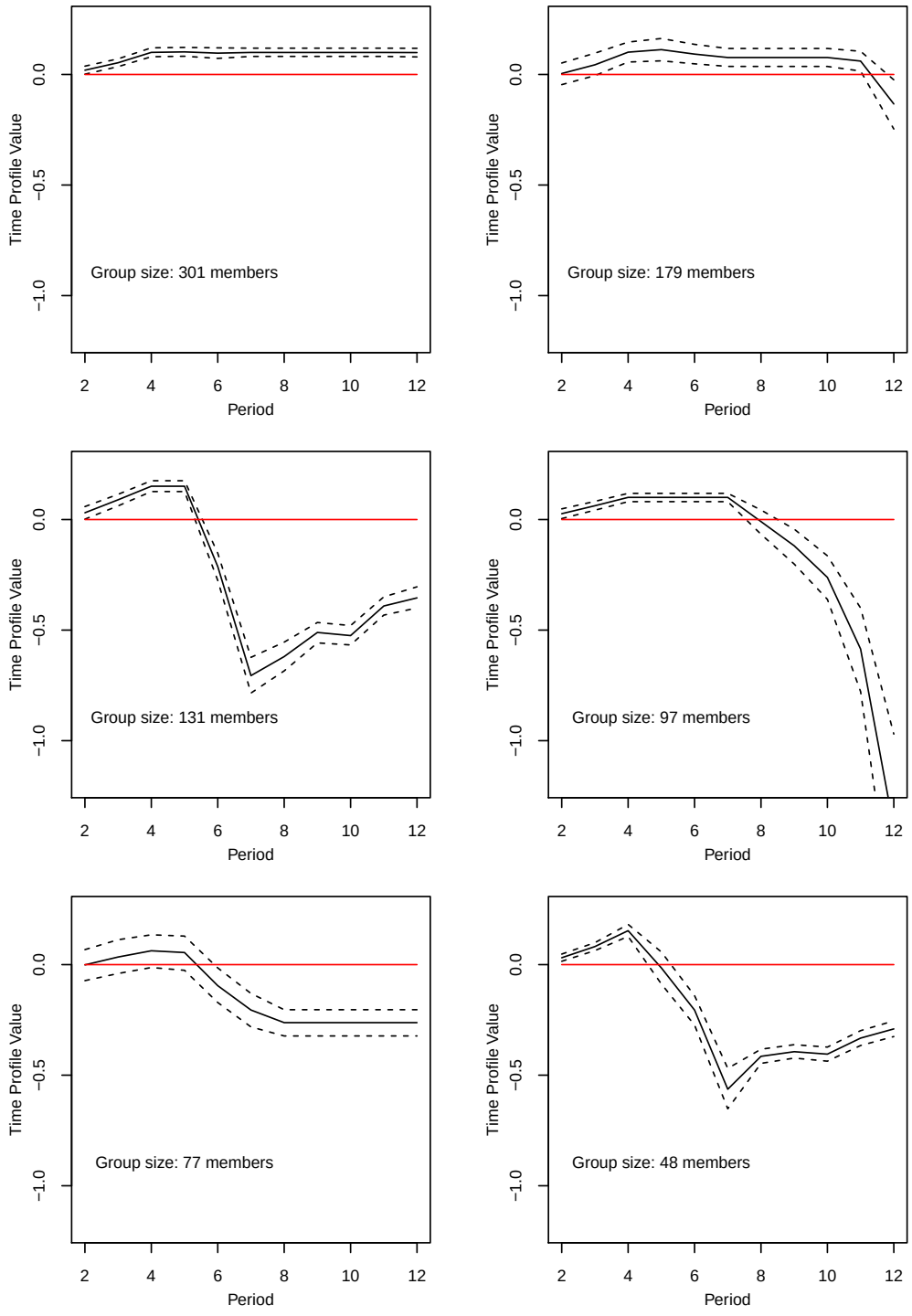
Figure 4.3 plots the estimated time profile posterior means and element-wise 95% credible intervals for the 6 largest groups from the best $(0, -1)$ configuration run, following the exemplar unit inference method described in Section 4.2.2. The time profiles roughly divide into three types based on the economic interpretation of their shapes: 1) targeting constant drawdown amounts; 2) initially targeting constant drawdowns, with subsequent downward revisions; 3) adjusting frequently during financial years 2009–2013, a time interval when the legislated minimum drawdown rates were temporarily reduced, then restored.[11]

Appendix B shows the time profiles for the $(1, 0)$ configuration run, which suggest the model is fitting noise in the dependent variable of the reduced model (4.7), rather than revealing distinct behavioural groups like in the $(0, -1)$ configuration runs. This visual comparison of the results is consistent with the WAIC values which more favourably rate the $(0, -1)$ configuration runs. Appendix B also provides MCMC diagnostics for the estimation runs.

[10] The WAIC for the $(1, 0)$ configuration run is -14306.81, while the $(0, -1)$ configuration run with the best fit has WAIC $= -26057.01$.

[11] The estimated time profiles for the other three $(0, -1)$ configuration runs also divide into the same three categories of behaviour.

Figure 4.3: Estimated time profile posterior means and 95% credible intervals for the 6 largest groups from the best $(0, -1)$ configuration run. Each plot corresponds to a different group. The time profile estimators are identified relative to period 1. Zero lines are plotted in red. The dashed lines depict element-wise 95% credible interval bounds.



4.5 Discussion

4.5.1 Superannuation drawdowns

The $(0, -1)$ configuration runs in the superannuation application, which are favoured by WAIC, algebraically reduce model (4.5) to model (4.6), suggesting that the variation in the drawdowns is best explained not through the drawdown rates chosen by retirees, expressed as a proportion of remaining account balance, but rather by the dollar amounts chosen. Thus, the results from the $(0, -1)$ configuration runs indicate that amount-based heuristics, such as drawing constant dollar amounts over time, seem to dominate rate-based heuristics, which are the focus of the theoretical literature (e.g., Bateman & Thorp, 2008). The results from the $(1, 0)$ configuration run, where the time profiles denote relative deviations in drawdown rates from the minima, support this story too, because the $(1, 0)$ configuration run does not clearly identify behavioural groups, and lacks parsimony in the grouping structure, suggesting that characterising the drawdowns as rates alone does not satisfactorily explain the variation in the drawdowns observed.

The results from the present study broadly align with related results from Chapter 3, which uses the GFE estimator. Although the GFE results find grouped time profiles with rate-based interpretations, for 5 of the 7 groups estimated, which collectively cover 66% of the sample, inspecting the drawdown amount series shows evidence for the use of amount-based heuristics: namely, tending to draw constant dollar amounts, possibly followed by subsequent downward revisions in the amounts.

The GFE estimator also finds a group that closely follows the minimum drawdown rates as they change during the sample observation period; the BFG results from the $(0, -1)$ configuration runs also find this behaviour, despite the time profiles not having a rate-based interpretation. Therefore, both the GFE and BFG approaches uncover the same behavioural groups, even though the estimated time profiles from the two methods have, respectively, rate-based and amount-based interpretations, and the identified behaviours themselves are a mix between rate-based and amount-based behaviours. Both approaches also support the conclusion that amount-based heuristics

are more prevalent than rate-based heuristics.[12] However, it is unclear whether the results generalise beyond the current dataset, given that the available data comes from a small number of superannuation funds.

The large number of groups, $G = 27$, estimated by the BFG method may seem problematic, since: a) one of the starting hypotheses was that the time-varying heterogeneity can be captured using a relatively small number of group time profiles; b) the DP prior specification used favours a small number of groups under the prior. Yet the largest 6 groups in the best $(0, -1)$ configuration run comprise over 80% of the sample, and a considerable portion of the estimated groups have very few members, or are singletons, so many of the estimated 'groups' can be viewed as noise, rather than representing real groups with behaviours of direct interest. Partitioning outlier units into small groups and singletons may have the added benefit of improving the quality of time profile estimates for the substantive groups, which are freed from the influence of outlier units. By contrast, the GFE estimator with some small, fixed number of groups is more likely to include outlier units in the substantive groups, which may affect estimation of the time profiles undesirably. However, this conjecture is based solely on a heuristic argument, and future research could formally investigate this hypothesis.

Even if small groups are treated as having economically meaningful time profiles like the larger groups, the initial hypothesis of a small number of groups in drawdown rate behaviours need not translate into an equivalent assumption on groups in the drawdown amounts. In any case, concluding that a small value for G is not defended by the data can be an economically useful result for future research. The plausibility of a small-G assumption can be questioned by examining the BFG results, whereas the computational demands of running the GFE approach with, say, $G \geq 27$, are prohibitive, so the GFE estimator might not feasibly be able to consider such a large model.

Aside from the economic implications, such a large G deserves the consideration of whether the BFG method is inherently biased towards overestimating the true number of groups. The simulation study described in Section 4.3 finds that overestimation can occur, but that it is attributed to the generation of singletons, rather than estimating

[12] The correspondence between the distinct groupings obtained by the BFG and GFE approaches is further considered in Appendix B.

too many groups of meaningful sizes; on average across the replicates, the PEAR method partitions 99.5% of the sample into four groups, which is the true number of groups in the DGP. This result supports the approach of restricting the analysis to the largest groups, and assuming that small groups represent overfitted noise.

Finally, the existence of at least two modes in the covariate effect configurations, and that the interpretation of the time profiles as effects on drawdown amounts, or as effects on excess drawdown rates, changes between these modes, suggests that a more complete characterisation of the data may be obtained by using a more flexible specification: one that allows for heterogeneous covariate effects across groups. Such a model allows, for each individual, observing time profiles with whichever economic interpretation fits their behaviour, rather than constraining the time profiles for all individuals to have the same interpretation in each estimation run. Developing such heterogeneous-coefficient models represents a useful extension to the current work.

4.5.2 Methodological considerations

Important distinctions between the BFG approach and the classical GFE estimator are: in the simulation study reported, BFG results show better group identification accuracy and time profile estimators than those from the GFE approach; the BFG method, in principle, only requires running the estimation procedure once, while the GFE estimator requires estimating the model once for each value of G considered; estimating models for larger G can be prohibitive with the GFE approach, and so the BFG method is better suited to test the assumption of a small number of latent groups in the data. However, the proposed method for inference on the group time profiles using the BFG method relies on a heuristic approach that assumes each group identified has a prototypical unit that, across iterations of the sampler, is attributed a time profile relating to the group often enough that iterates of this unit's time profile can be used to accurately describe the true group time profile. While this method performs well in the simulation study, it lacks a formal characterisation.

Future work can consider: extensions to nonlinear models, e.g., for binary response data; modelling the group allocation mechanism as a function of the observed covariates, e.g.,

similarly to mixture-of-experts models (Jacobs et al., 1991; Jordan & Jacobs, 1994); or a framework for predicting values of the dependent variable for out-of-sample time periods, similarly to the models with no individual-specific intercepts considered by Zhang (2020).

4.A Appendix: Gibbs sampling details

4.A.1 Joint posterior

The joint posterior is

$$p\left(\overset{\Delta}{\alpha}^{*}, \delta^{*}, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right) \propto p\left(\overset{\Delta}{y} \,|\, \overset{\Delta}{\alpha}^{*}, \delta^{*}, \theta, \gamma\right) \times p\left(\overset{\Delta}{\alpha}^{*}, \delta^{*}, V, \gamma, \theta, \beta\right)$$

$$\propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_{i} \,|\, \overset{\Delta}{\alpha}^{*}_{g_i}, \theta, \delta^{*}_{g_i}\right) p\left(g_i \,|\, V\right)\right] \left[\prod_{j=1}^{\infty} p\left(V_j \,|\, \beta\right) p\left(\overset{\Delta}{\alpha}^{*}_{j}\right) p\left(\delta^{*}_{j}\right)\right] p\left(\theta\right) p\left(\beta\right),$$

where $\gamma := \{g_1, g_2, ..., g_N\}$ is the set of group indicators for all units in the sample; $\overset{\Delta}{\alpha}^{*} := \left\{\overset{\Delta}{\alpha}^{*}_{1}, \overset{\Delta}{\alpha}^{*}_{2}, ...\right\}$.

The slice sampling approach (Kalli et al., 2011; Walker, 2007) augments the parameter space to include the auxiliary variables, $\eta = \{\eta_1, \eta_2, ..., \eta_N\}$, having independent uniform priors, $\eta_i \sim U\left(0, 1\right)$, for $i = 1, 2, ..., N$, so that $p\left(\eta\right) = 1$. The extended joint posterior is

$$p\left(\eta, \overset{\Delta}{\alpha}^{*}, \delta^{*}, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right) \propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_{i} \,|\, \overset{\Delta}{\alpha}^{*}_{g_i}, \theta, \delta^{*}_{g_i}\right) \mathbb{I}\left\{\eta_i < \pi_{g_i}\right\}\right]$$

$$\times \left[\prod_{j=1}^{\infty} p\left(V_j \,|\, \beta\right) p\left(\overset{\Delta}{\alpha}^{*}_{j}\right) p\left(\delta^{*}_{j}\right)\right] p\left(\theta\right) p\left(\beta\right).$$

Equivalently, the rewritten extended joint posterior is

$$p\left(\eta, \overset{\Delta}{\alpha}^{*}, \delta^{*}, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right) \propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_{i} \,|\, \overset{\Delta}{\alpha}^{*}_{g_i}, \theta, \delta^{*}_{g_i}\right) p\left(\eta_i \,|\, \pi_{g_i}\right) \pi_{g_i}\right]$$

$$\times \left[\prod_{j=1}^{\infty} p\left(V_j \,|\, \beta\right) p\left(\overset{\Delta}{\alpha}^{*}_{j}\right) p\left(\delta^{*}_{j}\right)\right] p\left(\theta\right) p\left(\beta\right),$$

with $p\left(\eta_i \,|\, \pi_{g_i}\right) = \pi_{g_i}^{-1} \mathbb{I}\left\{0 < \eta_i < \pi_{g_i}\right\}$.

Integrating η out of the extended joint posterior, $p\left(\eta, \overset{\Delta}{\alpha}^{*}, \delta^{*}, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right)$, recovers the original target joint posterior, $p\left(\overset{\Delta}{\alpha}^{*}, \delta^{*}, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right)$.

4.A.2 Gibbs sampling steps

Each iteration of the sampler executes the following sequence of steps. During an iteration, until a particular parameter has been drawn in that iteration, references to that parameter indicate its previous value. The first iteration uses starting values generated as in Appendix 4.A.4.

Step 1: Draw V and compute π

$$p\left(V \mid \overset{\Delta}{y}, \eta, \delta^*, \overset{\Delta^*}{\alpha}, \gamma, \theta, \beta\right) \propto p\left(\eta, \overset{\Delta^*}{\alpha}, \delta^*, V, \gamma, \theta, \beta \mid \overset{\Delta}{y}\right)$$

$$\propto \left[\prod_{i=1}^{N} p\left(\eta_i \mid \pi_{g_i}\right) \pi_{g_i}\right] \left[\prod_{j=1}^{\infty} p\left(V_j \mid \beta\right)\right].$$

This is a partially collapsed Gibbs sampler (see, e.g., Van Dyk & Park, 2008), with η integrated out of this step.

$$p\left(V \mid \overset{\Delta}{y}, \delta^*, \overset{\Delta^*}{\alpha}, \gamma, \theta, \beta\right) = \int p\left(V \mid \eta, \overset{\Delta}{y}, \sigma^2, \overset{\Delta}{\alpha}, \gamma, \theta, \beta\right) p\left(\eta\right) \mathrm{d}\eta$$

$$\propto \left[\prod_{j=1}^{\infty} \left(1 - V_j\right)^{\beta - 1}\right] \left[\prod_{i=1}^{N} \left(V_{g_i} \prod_{l=1}^{g_i - 1} \left(1 - V_l\right)\right)\right].$$

For $g = 1, 2, ..., C^*$,

$$p\left(V_g \mid \overset{\Delta}{y}, \delta^*, \overset{\Delta^*}{\alpha}, \gamma, \theta, \beta\right) \propto \left[\prod_{i=1}^{N} \left(V_{g_i} \prod_{l=1}^{g_i - 1} \left(1 - V_l\right)\right)\right] \left[\left(1 - V_g\right)^{\beta - 1}\right]$$

$$\propto \mathrm{Beta}\left(V_g; 1 + N_g, \beta + N_{>g}\right),$$

with $N_g := \sum_{i=1}^{N} \mathbb{I}\left\{g_i = g\right\}$, the number of units in the group labelled g; $N_{>g} := \sum_{i=1}^{N} \mathbb{I}\left\{g_i > g\right\}$, the number of units in groups with label value $> g$.

After drawing V_g for $g = 1, 2, ..., C^*$, update $\pi_g = V_g \prod_{l=1}^{g-1} \left(1 - V_l\right)$ for $g = 2, ..., C^*$, and $\pi_1 = V_1$.

Step 2: Draw $\overset{\Delta^*}{\alpha}$

$$p\left(\overset{\Delta^*}{\alpha} \mid \overset{\Delta}{y}, \eta, \delta^*, V, \gamma, \theta, \beta\right) \propto p\left(\eta, \overset{\Delta^*}{\alpha}, \delta^*, V, \gamma, \theta, \beta \mid \overset{\Delta}{y}\right)$$

$$\propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_i \mid \overset{\Delta^*}{\alpha}_{g_i}, \theta, \delta^*_{g_i}\right)\right] \left[\prod_{j=1}^{\infty} p\left(\overset{\Delta^*}{\alpha}_j\right)\right].$$

For $g = 1, 2, ..., C^*$,

$$p\left(\overset{\Delta}{\alpha}{}^*_g \,|\, \overset{\Delta}{y}, \eta, \delta^*, V, \gamma, \theta, \beta\right) \propto \left[\prod_{i \in \mathcal{I}_g} p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta}{\alpha}{}^*_g, \theta, \delta^*_g\right)\right] p\left(\overset{\Delta}{\alpha}{}^*_g\right),$$

where $\mathcal{I}_g$ is the subset of $\mathcal{I} = \{1, 2, ..., N\}$ which contains all i such that $g_i = g$.

Then,

$$p\left(\overset{\Delta}{\alpha}{}^*_g \,|\, \overset{\Delta}{y}, \eta, \delta^*, V, \gamma, \theta, \beta\right) \propto MVN_{T-1}\left(\overset{\Delta}{\alpha}{}^*_g; A_{\alpha,g}, B_{\alpha,g}\right),$$

with $B_{\alpha,g} := \left[S_\alpha^{-1} + N_g \left(D\Sigma^*_g D'\right)^{-1}\right]^{-1}$;
$A_{\alpha,g} := B_{\alpha,g}\left[S_\alpha^{-1} m_\alpha + \left(D\Sigma^*_g D'\right)^{-1}\left(\sum_{i \in \mathcal{I}_g} \overset{\Delta}{y}_i - \overset{\Delta}{x}_i \theta\right)\right]$.

Step 3: Draw δ^*

$$p\left(\delta^* \,|\, \overset{\Delta}{y}, \eta, \overset{\Delta}{\alpha}{}^*, V, \gamma, \theta, \beta\right) \propto p\left(\eta, \overset{\Delta}{\alpha}{}^*, \delta^*, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right)$$

$$\propto \left[\prod_{i=1}^N p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta}{\alpha}{}^*_{g_i}, \theta, \delta^*_{g_i}\right)\right]\left[\prod_{j=1}^\infty p\left(\delta^*_j\right)\right].$$

For $g = 1, 2, ..., C^*$,

$$p\left(\delta^*_g \,|\, \overset{\Delta}{y}, \eta, \overset{\Delta}{\alpha}{}^*, V, \gamma, \theta, \beta\right) \propto \left[\prod_{i \in \mathcal{I}_g} p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta}{\alpha}{}^*_g, \theta, \delta^*_g\right)\right] p\left(\delta^*_g\right)$$

$$\propto \left[\det\left(\Lambda^*_g\right)\right]^{-N_g/2} \exp\left\{-\tfrac{1}{2}\left[\sum_{i \in \mathcal{I}_g} \tilde{y}_i' \left(\Lambda^*_g\right)^{-1} \tilde{y}_i\right]\right\}$$

$$\times \left[\det\left(S_\delta\right)\right]^{-1/2} \exp\left\{-\tfrac{1}{2}\left(\delta^*_g - m_\delta\right)' S_\delta^{-1}\left(\delta^*_g - m_\delta\right)\right\},$$

with $\Lambda^*_g := D\Sigma^*_g D'$; $\tilde{y}_i := \overset{\Delta}{y}_i - \overset{\Delta}{\alpha}{}^*_g - \overset{\Delta}{x}_i \theta$.

Here, an option is to use a Metropolis–Hastings (MH) step to sample δ^*_g, where the MH proposal density function, $q\left(\delta^{*(m)}_g \to \delta^{*(m+1)}_g\right)$, at an arbitrary iteration $m+1$, is the Laplace approximation to $p\left(\delta^*_g \,|\, \overset{\Delta}{y}, \eta, \overset{\Delta}{\alpha}{}^*, V, \gamma, \theta, \beta\right)$; i.e.,

$$q\left(\delta^{*(m)}_g \to \delta^{*(m+1)}_g\right) := MVN_T\left(\delta^{*(m+1)}_g; \widehat{\delta}^{MCP}_g, \Sigma\left(\widehat{\delta}^{MCP}_g\right)\right) =: q\left(\delta^{*(m+1)}_g\right),$$

and similarly,

$$q\left(\delta^{*(m+1)}_g \to \delta^{*(m)}_g\right) := MVN_T\left(\delta^{*(m)}_g; \widehat{\delta}^{MCP}_g, \Sigma\left(\widehat{\delta}^{MCP}_g\right)\right) =: q\left(\delta^{*(m)}_g\right),$$

where $\widehat{\delta}_g^{MCP}$ is the estimated value of δ_g^* that maximises $\ln p\left(\delta_g^* \mid \overset{\Delta}{y}, \eta, \overset{\Delta^*}{\alpha}, V, \gamma, \theta, \beta\right)$; $\Sigma\left(\widehat{\delta}_g^{MCP}\right)$ is the inverse of the negative of the Hessian matrix of $\ln p\left(\delta_g^* \mid \overset{\Delta}{y}, \eta, \overset{\Delta^*}{\alpha}, V, \gamma, \theta, \beta\right)$ evaluated at $\delta_g^* = \widehat{\delta}_g^{MCP}$; $\delta_g^{*(m)}$ is the iteration-m iterate of parameter δ_g^*.

The probability of accepting a proposed value $\delta_g^{*(m+1)}$, drawn from $q\left(\delta_g^{*(m)} \to \delta_g^{*(m+1)}\right)$, is

$$\text{Prob}\left(\delta_g^{*(m)} \to \delta_g^{*(m+1)}\right) := \min\left\{1, \frac{p\left(\delta_g^{*(m+1)} \mid \overset{\Delta}{y}, \eta, \overset{\Delta^*}{\alpha}, V, \gamma, \theta, \beta\right) q\left(\delta_g^{*(m+1)} \to \delta_g^{*(m)}\right)}{p\left(\delta_g^{*(m)} \mid \overset{\Delta}{y}, \eta, \overset{\Delta^*}{\alpha}, V, \gamma, \theta, \beta\right) q\left(\delta_g^{*(m)} \to \delta_g^{*(m+1)}\right)}\right\}.$$

Step 4: Label-switching move

Three label-switching move proposal generators are considered: two introduced by Papaspiliopoulos and Roberts (2008), and the third introduced by Hastie et al. (2015). In each iteration of the sampler, one of the three generators is selected at random with equal probability.[13] A corresponding label-switching move is proposed from the selected generator, and implemented with the move-specific acceptance probability specified below.

Move 1 (from Papaspiliopoulos & Roberts, 2008). Randomly select two non-empty groups, call them groups j and l. The groups initially have N_j and N_l members, respectively. The proposed action is to switch the labels of groups j and l, by labelling all N_j units that were in group j as now belonging to group l, and labelling the N_l units that were in group l as now belonging to group j; also swap the labels of $\overset{\Delta^*}{\alpha}_j$ and $\overset{\Delta^*}{\alpha}_l$, as well as δ_j^* and δ_l^*. The acceptance probability for this move is

$$\min\left(1, \left(\frac{\pi_l}{\pi_j}\right)^{N_j - N_l}\right).$$

Move 2 (from Papaspiliopoulos & Roberts, 2008). Randomly select two sequentially labelled groups, call them groups j and $j+1$. Switch the labels of the two groups and group-specific parameters, as in move 1, and also switch the values of V_j and

[13] Kim and Wang (2019) also do this, but only considering the two label-switching moves from Papaspiliopoulos and Roberts (2008).

V_{j+1}. Accept this action with probability

$$\min\left(1, \frac{(1-V_{j+1})^{N_j}}{(1-V_j)^{N_{j+1}}}\right).$$

Move 3 (from Hastie et al., 2015). Randomly select two sequentially labelled groups, call them groups j and $j+1$. Swap group labels and group-specific parameters, as in moves 1 and 2, then set π_j to π'_j, and π_{j+1} to π'_{j+1}, where:

$\pi'_j := \pi_{j+1}\frac{\pi^+}{\Pi'}R_1$;

$\pi^+ := \pi_j + \pi_{j+1}$;

$\Pi' := \pi_{j+1}R_1 + \pi_j R_2$;

$R_1 := \frac{1+\beta+N_{j+1}+\sum_{l>j+1}N_l}{\beta+N_{j+1}+\sum_{l>j+1}N_l}$;

$R_2 := \frac{\beta+N_j+\sum_{l>j+1}N_l}{1+\beta+N_j+\sum_{l>j+1}N_l}$;

$\pi'_{j+1} := \pi_j\frac{\pi^+}{\Pi'}R_2$.

Accept the proposed move with probability

$$\min\left(1, \left(\frac{\pi^+}{\Pi'}\right)^{N_j+N_{j+1}} R_1^{N_{j+1}} R_2^{N_j}\right).$$

Step 5: Draw η

$p\left(\eta \,|\, \overset{\Delta}{y}, \gamma, \delta^*, \overset{\Delta^*}{\alpha}, V, \theta, \beta\right) \propto p\left(\eta, \overset{\Delta^*}{\alpha}, \delta^*, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right)$

$\propto \left[\prod_{i=1}^N \mathbb{I}\{\eta_i < \pi_{g_i}\}\right]$.

For $i = 1, 2, ..., N$,

$p\left(\eta_i \,|\, \overset{\Delta}{y}, \gamma, \delta^*, \overset{\Delta^*}{\alpha}, V, \theta, \beta\right) \propto \mathbb{I}\{\eta_i < \pi_{g_i}\}$

$\propto$ Uniform $(\eta_i; 0, \pi_{g_i})$.

Step 6: Draw β

Following Walker (2007) and Kim and Wang (2019),

$$p\left(\beta \,|\, \overset{\Delta}{y}, \gamma, \tilde{\sigma}^2, \overset{\Delta}{\alpha}, V, \theta, \eta\right) = p\left(\beta \,|\, G; N\right) \propto p\left(\beta\right) p\left(G \,|\, \beta; N\right),$$

where G is the number of non-empty groups in the current iteration. Following Escobar and West (1995), draw $p\left(\beta \,|\, G; N\right)$ by first introducing a latent variable, κ, and sampling $\kappa \,|\, \beta, G$, then $\beta \,|\, \kappa, G$, using the following two results.

1. $\kappa \,|\, \beta, G \sim \text{Beta}\left(\beta + 1, N\right)$.
2. $p\left(\beta \,|\, \kappa, G\right) = \xi \text{Gamma}\left(\beta; a_\beta + G, b_\beta - \ln \kappa\right)$
   $+ \left(1 - \xi\right) \text{Gamma}\left(\beta; a_\beta + G - 1, b_\beta - \ln \kappa\right)$, where $\xi := \frac{a_\beta + G - 1}{a_\beta + G - 1 + N \times \left(b_\beta - \ln \kappa\right)}$.

To draw from $p\left(\beta \,|\, \kappa, G\right)$, which is the density of a two-component mixture of Gamma random variables, in each iteration, with probability ξ, draw from the first component, and with probability $\left(1 - \xi\right)$, draw from the second component.

Step 7: Recompute η^* and C^*

Compute η^* and C^* for the current iteration's set of η and V, as per $\eta^* := \min\left\{\eta_1, \eta_2, ..., \eta_N\right\}$, and $C^* := \min\left\{c : \sum_{j=1}^{c} \pi_j > 1 - \eta^*\right\}$.

If the C^* for a given iteration is larger than for the previous iteration, draw more $\overset{\Delta}{\alpha}{}^*_g, \delta^*_g$ and V_g from their respective priors until there are C^* of each in total. To draw fresh values of V_g, use the current iterate of β as input to the prior.

Step 8: Draw θ

$$p\left(\theta \,|\, \overset{\Delta}{y}, \gamma, \delta^*, \overset{\Delta}{\alpha}{}^*, V, \eta, \beta\right) \propto p\left(\eta, \overset{\Delta}{\alpha}{}^*, \delta^*, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right)$$

$$\propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta}{\alpha}{}^*_{g_i}, \theta, \delta^*_{g_i}\right)\right] \times 1$$

$$\propto MVN_K\left(\theta; A_\theta, B_\theta\right),$$

$$\text{with } B_\theta := \left[\sum_{i=1}^{N} \overset{\Delta}{x}{}'_i \left(D \Sigma^*_{g_i} D'\right)^{-1} \overset{\Delta}{x}_i\right]^{-1}; \; A_\theta := B_\theta \left[\sum_{i=1}^{N} \overset{\Delta}{x}{}'_i \left(D \Sigma^*_{g_i} D'\right)^{-1} \left(\overset{\Delta}{y}_i - \overset{\Delta}{\alpha}_{g_i}\right)\right].$$

Step 9: Draw γ

$$p\left(\gamma \,|\, \overset{\Delta}{y}, \eta, \delta^*, \overset{\Delta}{\alpha}^*, V, \theta, \beta\right) \propto p\left(\eta, \overset{\Delta}{\alpha}^*, \delta^*, V, \gamma, \theta, \beta \,|\, \overset{\Delta}{y}\right)$$

$$\propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta}{\alpha}_{g_i}, \theta, \delta^*_{g_i}\right) \mathbb{I}\left\{\eta_i < \pi_{g_i}\right\}\right].$$

For $i = 1, 2, ..., N$ and $g = 1, 2, ...,$

$$\Pr\left(g_i = g \,|\, \overset{\Delta}{y}, \eta, \delta^*, \overset{\Delta}{\alpha}^*, V, \theta, \beta\right) \propto p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta}{\alpha}^*_g, \theta, \delta^*_g\right) \mathbb{I}\left\{\eta_i < \pi_g\right\}.$$

For $g > C^*$, $\mathbb{I}\left\{\eta_i < \pi_g\right\} = 0$ for all i; therefore, draw g_i according to

$$\Pr\left(g_i = g \,|\, \overset{\Delta}{y}, \eta, \delta^*, \overset{\Delta}{\alpha}^*, V, \theta, \beta\right) = \begin{cases} \frac{p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta}{\alpha}^*_g, \theta, \delta^*_g\right)\mathbb{I}\{\eta_i < \pi_g\}}{\sum_{j=1}^{C^*} p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta}{\alpha}^*_j, \theta, \delta^*_j\right)\mathbb{I}\{\eta_i < \pi_j\}}, & \text{for } g = 1, 2, ..., C^*; \\ 0, & \text{otherwise.} \end{cases}$$

4.A.3 Fixed prior parameters

$\overset{\Delta}{\alpha}^*_g$ **prior.** m_α is set to the first-differenced time profile obtained by manipulating the 2WFE relative time profile point estimates, $\tilde{\alpha}_t$, for $t = 2, 3, ..., T$. The $\tilde{\alpha}_t$ are the estimated coefficients on the time dummy variables from the 2WFE results, obtained by a least-squares regression using the time-demeaned data, $\dot{y}_{it}$ and $\dot{x}_{it}$, including a time-demeaned time dummy variable for each time period except the first; $\dot{y}_{it} := y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$, and similarly for $\dot{x}_{it}$. The first-differenced time profile values, $\overset{\Delta}{\alpha}_t$, for $t = 2, 3, ..., T$, can be recovered from the relative time profiles, using $\overset{\Delta}{\alpha}_2 := \tilde{\alpha}_2$, and $\overset{\Delta}{\alpha}_t := \tilde{\alpha}_t - \tilde{\alpha}_{t-1}$, for $t = 3, 4, ..., T$. S_α is a $(T-1) \times (T-1)$ diagonal matrix with diagonal elements equal to 30 times the magnitude of the corresponding elements of m_α; all off-diagonal elements are zero.

δ^*_g **prior.** m_δ is set to a length-T vector with all elements equal to $\ln\left(\frac{A^2}{\sqrt{A+A^2}}\right)$, with A set to the 2WFE error term variance point estimate,
$A := (NT - N - K)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{u}^2_{it}$; $\widehat{u}_{it} := \dot{y}_{it} - \dot{x}'_{it}\widehat{\theta}^{(2WFE)} - \tilde{\alpha}_t$ for all (i, t), with $\tilde{\alpha}_1 = 0$; $\widehat{\theta}^{(2WFE)}$ is the 2WFE estimate of θ. S_δ is a $T \times T$ diagonal matrix with all diagonal elements equal to $\ln\left(\frac{A}{A^2} + 1\right)$. This is equivalent to putting a log-Normal prior on $\sigma^{2*}_{gt} = \exp\left(\delta^*_{gt}\right)$ for all (g, t), with prior mean and variance equal to A.

β prior. (a_β, b_β) is set to $(0.4, 10)$ as in Kim and Wang (2019). A Gamma distribution with these parameters concentrates its mass towards 0, acting as a shrinkage prior on the DP concentration parameter β. These parameter values encode a prior expectation that the number of groups in the sample is small.

4.A.4 Starting value generation

The following starting value generation procedure is based on that in Kim and Wang (2019), with some adaptations to suit the BFG specification. The first and third steps generate candidate group time profiles and error variance parameters. Steps 4 and 5 then filter these group parameters down to a smaller set of time profiles and error variances that are more likely to fit the data. This procedure worked well in our simulation study and application, so we did not pursue alternative starting value generation procedures; however, a valuable contribution of future research could be to propose and compare other procedures.

1. Generate G_{start} sets of $\overset{\Delta}{\alpha}{}^*_g$ vectors from the prior, $\overset{\Delta}{\alpha}{}^*_g \sim MVN_{T-1}(m_\alpha, S_\alpha)$, where G_{start} is selected by the researcher. Refer to the collection of these vectors as $\left\{\overset{\Delta}{\alpha}{}^{*(j)}\right\}_{j=1}^{G_{start}}$.
2. Draw θ_k, for $k = 1, 2, ..., K$, from $N\left(\widehat{\theta}_k^{(2WFE)}, 30 \times \left|\widehat{\theta}_k^{(2WFE)}\right|\right)$, where the scalar $\widehat{\theta}_k^{(2WFE)}$ is the 2WFE estimate of θ_k.
3. Draw δ_g^*, for $g = 1, 2, ..., G_{start}$, independently from the prior, $\delta_g^* \sim MVN_T(m_\delta, S_\delta)$.
4. Assign each unit to one of the G_{start} vectors $\overset{\Delta}{\alpha}{}^{*(j)}$ by drawing g_i, for each i, from the conditional posterior for g_i but ignoring the terms $\mathbb{I}\{\eta_i < \pi_g\}$ and $\mathbb{I}\{\eta_i < \pi_j\}$.
5. Eliminate any of the $\overset{\Delta}{\alpha}{}^{*(j)}$ which had no units assigned to it; i.e., eliminate the empty groups.
6. Set C^* as the number of non-empty groups.
7. Relabel the set of surviving $\overset{\Delta}{\alpha}{}^{*(j)}$ as $\left\{\overset{\Delta}{\alpha}{}^*_g\right\}_{g=1}^{C^*}$, maintaining the order from the initial set of vectors.

8. Set the grouping vector, $\gamma = \{g_1, g_2, ..., g_N\}$, according to the new labels.

9. Set β to 1.

Chapter 5

Bayesian Heterogeneous Panel Models with Latent Group Time Profiles

Abstract

This chapter proposes nonparametric Bayesian estimation methods for linear panel data models with group-specific covariate effects and group-specific time-varying heterogeneity, while controlling for individual-specific time-invariant effects. Group membership is unobserved, and groupings in the covariate effects can differ from groupings in the time-varying heterogeneity. Estimator performance is examined using a simulation study. Two empirical applications demonstrate the value of such models in a broad class of microeconomic applications using observational panel data: one application extends an existing analysis of superannuation drawdown behaviours which assumes homogeneous covariate effects; the second application finds heterogeneous smoking behaviours in Australia surrounding multiple government policy interventions.

5.1 Introduction

A significant portion of the variability in outcomes measured in observational microeconomic data is driven by heterogeneity in individual preferences. These preferences may manifest as behavioural changes over time, driven by unobservable factors, as well as heterogeneous responses to variables observable by the researcher. In a panel (longitudinal) regression model, these features are captured by allowing for differential time effects, and heterogeneous coefficients on the included covariates, respectively.

A panel data model with individual-specific intercepts, group-specific time profiles, and group-specific regression coefficients is

$$y_{it} = c_i + \alpha^*_{g_i t} + x'_{it}\theta^*_{g_i} + u_{it}, \tag{5.1}$$

where $i = 1, 2, ..., N$ indexes observational units; $t = 1, 2, ..., T$ indexes successive time periods; y_{it} is the value of an observed outcome for unit i in period t; c_i is a unit-specific effect, called a 'fixed effect' in econometrics; $\alpha^*_{g_i t}$ is the contemporaneous contribution of time-varying unobservable heterogeneity, where the heterogeneity across units adheres to a grouping structure in which the N units are grouped into G groups, and the time-varying heterogeneity is common within groups, but can differ across groups; $g_i \in \{1, 2, ..., G\}$ identifies the group that unit i belongs to; x_{it} is a $K \times 1$ column vector of observed covariates; $\theta^*_{g_i}$ are the corresponding partial effects of the covariates, which, in this restricted model, have the same grouping structure as the time profiles, $\alpha^*_{g_i} = \left(\alpha^*_{g_i 1}, \alpha^*_{g_i 2}, ..., \alpha^*_{g_i T}\right)'$; u_{it} is the model error, assumed throughout this chapter to follow a Normal distribution with a variance that can change across groups and time.

Model (5.1) generalises the linear panel data model considered in Chapter 4,

$$y_{it} = c_i + \alpha^*_{g_i t} + x'_{it}\theta + u_{it}, \tag{5.2}$$

where the entire sample is assumed to adhere to the same set of covariate effects; i.e., $\theta^*_{g_i} = \theta$ for all i. However, results from the application to the superannuation data presented in Chapter 4 suggest a model with heterogeneous coefficients, such as (5.1),

may be more suitable. More broadly, considering the existence of latent groups with different behaviours related to unobservables motivates a similar consideration of whether there are also subgroups defined by differential responses to observed covariates.

A class of applications for which there is a clear economic motivation to consider a heterogeneous-coefficients model is in the relatively general scenario where some individuals are expected to modify behaviours in response to a set of observed covariates, while other individuals may entirely ignore these covariates. Alternatively, or in addition, observed factors may have a much weaker influence on an observed outcome for some subgroups when compared to others, or some covariates may be associated with an increase in the measured response for some groups, and a decrease in the response for other groups. Furthermore, in some instances, such as the application to retirement decumulation considered in this chapter, a model can be specified such that certain configurations of the covariate effects reduce the model into distinct, special cases, with each case giving rise to different economic interpretations of the remaining parameters.

While (5.1) allows for grouped heterogeneity in the covariate effects, assuming the latent groups in the time profiles, $\alpha^*_{g_i}$, are the same groups which correspond to the heterogeneous covariate effects, $\theta^*_{g_i}$, is somewhat restrictive; moreover, simulation results presented here and in Appendix C suggest that this model performs badly when this assumption is violated. For these reasons, this chapter also develops an alternative formulation of the heterogeneous-coefficients model,

$$y_{it} = c_i + \alpha^*_{g_i t} + x'_{it}\theta^*_{h_i} + u_{it}, \tag{5.3}$$

where $h_i \in \{1, 2, ..., H\}$, for some number of groups H, is a group assignment variable analogous to g_i. In (5.3), individual units are classified into groups along two 'dimensions': the first dimension is the grouping with respect to the time profiles; the second is with respect to partial effects of the observed covariates. A multi-dimensional grouping structure is not only an abstract tool to specify model (5.3), but a realistic characterisation of human preferences: e.g., two people may respond similarly to a change in personal income, but their demand for some product may evolve differently

over time in response to economy-wide shocks, in which case they could be seen as belonging to the same group in the covariate effect dimension, but different groups in the time profile dimension.

There is an active literature that compares heterogeneous and homogeneous panels (see, e.g., Baltagi, Bresson, & Pirotte, 2008). The former involves allowing all parameters to vary over i, while the latter refers to a standard two-way fixed effects (2WFE) specification. The 2WFE model can be viewed as a special case of model (5.1), assuming that the sample is one 'group'; i.e., there are no latent subgroups in the sample with distinct covariate effects and time profiles.

A common theme in the heterogeneous panel modelling literature is that even when T is large enough to allow this level of flexibility, homogeneous specifications tend to perform well in terms of forecasting and parameter estimation. This chapter proposes a partial pooling approach that arguably provides more flexibility than a homogeneous specification, without the degradation in performance typical of individual-specific models. Moreover, the type of specification considered here remains feasible when T is moderate.

The main contribution of this chapter is to propose a nonparametric Bayesian model for (5.1) and (5.3); in particular, a Dirichlet Process (DP; Ferguson, 1973) Mixture Model (DPMM; see, e.g., Escobar & West, 1995; Ferguson, 1983) and derive suitable estimation procedures. Such a model extends the model used in Chapter 4, which in turn extends the model presented in Kim and Wang (2019) by controlling for individual fixed effects. DPMMs allow for flexible data-driven clustering, jointly estimating a latent grouping structure along with the model parameters, without requiring the researcher to specify the number of latent groups.

A second contribution compares the performance of two proposed methods using a simulation study, which highlights certain conditions under which one approach may be preferred. Two other contributions relate to the application results: applying the methods presented in this chapter to extend an existing application to retirement decumulation confirms previous results in addition to identifying new behavioural

groups; a second, novel application to smoking behaviours shows how the methods may be used more broadly for policy analysis.

Section 5.2 presents DPMMs for (5.1) and (5.3); Section 5.3 compares the performance of the proposed estimation procedures using simulated data; Section 5.4 presents the results from the application to retirement decumulation; Section 5.5 shows the novel application to smoking policy and behaviours in Australia; Section 5.6 concludes the chapter with a discussion of the simulation and application results. Appendix 5.A gives implementation details for the proposed methods, and Appendix C contains additional results and details from the applications and simulation study.

5.2 Methodology

5.2.1 'BFG-HC': A heterogeneous-coefficient extension of the Bayesian First-differenced Grouped fixed-effects (BFG) procedure

Assume observations are generated according to $y_{it} \sim N\left(c_i + \alpha^*_{g_i t} + x'_{it}\theta^*_{g_i}, \sigma^{2*}_{g_i t}\right)$, for $i = 1, 2, ..., N$, and $t = 1, 2, ..., T$; the notation $y_{it} \sim N\left(m_y, v_y\right)$ means the random scalar y_{it} has a Normal distribution with mean m_y and variance v_y; errors u_{it} may be heteroskedastic across groups and time. For convenience, especially in terms of simplifying notation, we assume a balanced panel; i.e., all units are observed in all time periods. Stacking over time gives the individual-level vectors $y_i \sim MVN_T\left(c_i\iota_T + \alpha^*_{g_i} + x_i\theta^*_{g_i}, \Sigma^*_{g_i}\right)$; $y_i \sim MVN_T\left(m_y, S_y\right)$ means the random length-T vector y_i has a multivariate Normal distribution with mean vector m_y and covariance matrix S_y; $\alpha^*_{g_i}$ are the time profile vectors as defined in Section 5.1; $\Sigma^*_{g_i} := \text{diag}\left(\sigma^{2*}_{g_i}\right)$; $\text{diag}\left(v\right)$ is a diagonal matrix with nonzero entries given by the values in some vector v; $\sigma^{2*}_{g_i} := \left(\sigma^{2*}_{g_i 1}, \sigma^{2*}_{g_i 2}, ..., \sigma^{2*}_{g_i T}\right)'$; $y_i := (y_{i1}, y_{i2}, ..., y_{iT})'$; $x_i := [x'_{i1}, x'_{i2}, ..., x'_{iT}]'$; ι_T is a length-T column vector of ones. First-differencing the model leads to $Dy_i \sim MVN_{T-1}\left(D\alpha^*_{g_i} + Dx_i\theta^*_{g_i}, D\Sigma^*_{g_i}D'\right)$, where D is the $(T-1)$ by T first-differencing matrix with typical row $[0, ..., 0, -1, 1, 0, ..., 0]$; compressing notation gives the equivalent $\overset{\Delta}{y}_i \sim MVN_{T-1}\left(\overset{\Delta^*}{\alpha}_{g_i} + \overset{\Delta}{x}_i\theta^*_{g_i}, \Lambda^*_{g_i}\right)$, using $\overset{\Delta}{y}_i := Dy_i$ and similarly for $\overset{\Delta^*}{\alpha}_{g_i}$ and $\overset{\Delta}{x}_i$; $\Lambda^*_{g_i} := D\Sigma^*_{g_i}D'$.

Since the grouping structure for the θ^* and $\overset{\Delta}{\alpha}^*$ parameters are the same, the model can be simplified by writing $\overset{\Delta}{y}_i \sim MVN_{T-1}\left(\overset{\Delta}{w}_i \phi^*_{g_i}, \Lambda^*_{g_i}\right)$; $\overset{\Delta}{w}_i := \left[I_{T-1} \; \overset{\Delta}{x}_i\right]$; I_{T-1} is the $(T-1)$ by $(T-1)$ identity matrix; $\phi^*_{g_i} := \left[\overset{\Delta}{\alpha}^*_{g_i}{}' \; \theta^*_{g_i}{}'\right]'$. Thus, the conditional density for the data is

$$\overset{\Delta}{y}_i \,|\, \phi^*_{g_i}, \delta^*_{g_i} \sim MVN_{T-1}\left(\overset{\Delta}{w}_i \phi^*_{g_i}, \Lambda^*_{g_i}\right), \tag{5.4}$$

independently for $i = 1, 2, ..., N$. The error variance terms are reparameterised using $\sigma^{2*}_{g_i t} := \exp\left(\delta^*_{g_i t}\right)$; $\delta^*_{g_i} := \left(\delta^*_{g_i 1}, \delta^*_{g_i 2}, ..., \delta^*_{g_i T}\right)'$.

Most of the priors are identical to the corresponding priors in the BFG procedure, following the stick-breaking construction of the DP prior (see, e.g., Sethuraman, 1994). The prior specification is presented in full here for completeness; however, the only difference between the specification below and the BFG specification is that the prior for the ϕ^*_g terms in the list below replaces the priors for the θ and $\overset{\Delta}{\alpha}^*_g$ terms from the BFG procedure.

- $\phi^*_g \sim MVN_{T-1+K}\left(m_\phi, S_\phi\right)$, for $g = 1, 2, ...$;
- $\delta^*_g \sim MVN_T\left(m_\delta, S_\delta\right)$, for $g = 1, 2, ...$;
- $g_i \,|\, \pi \sim \sum_{g=1}^{\infty} \pi_g \delta\left(g\right)$, for $i = 1, 2, ..., N$, with $\pi := \left(\pi_1, \pi_2, ...\right)$ and $\delta\left(g\right)$ being the degenerate discrete measure with unit mass at value g; $p\left(g \,|\, V\right) := \Pr\left(g_i = g \,|\, V\right) = \pi_g$, for $i = 1, 2, ..., N$ and $g = 1, 2, ...$, with $V := \left(V_1, V_2, ...\right)$;
- $\pi_g = V_g \prod_{j=1}^{g-1}\left(1 - V_j\right)$, for $g = 2, 3, ...$, and $\pi_1 = V_1$;
- $V_g \,|\, \beta \sim \text{Beta}\left(1, \beta\right)$, for $g = 1, 2, ...$;
- $\beta \sim \text{Gamma}\left(a_\beta, b_\beta\right)$, following Escobar and West (1995), who recommend treating the DP prior concentration parameter β as random, rather than fixed. $\text{Gamma}\left(a, b\right)$ denotes the distribution with density function $\text{Gamma}\left(v; a, b\right) \propto b^a v^{a-1} e^{-bv}$.

5.2.2 'BFG-HC2': A two-dimensional extension of the BFG-HC procedure

The setup for the BFG-HC2 procedure differs from the BFG-HC approach in two ways: a) the covariate effect parameters and time profile parameters do not collapse into one parameter block; b) new, independent sets of stick-breaking parameters are defined to track group membership in the covariate effects. The conditional density of the data is

$$\overset{\Delta}{y}_i \,|\, \overset{\Delta^*}{\alpha}_{g_i}, \theta^*_{h_i}, \delta^*_{g_i} \sim MVN_{T-1}\left(\overset{\Delta^*}{\alpha}_{g_i} + \overset{\Delta}{x}_i\theta^*_{h_i}, \Lambda^*_{g_i}\right), \tag{5.5}$$

independently for $i = 1, 2, ..., N$. The full stick-breaking prior specification is:

- $\overset{\Delta^*}{\alpha}_g \sim MVN_{T-1}(m_\alpha, S_\alpha)$, for $g = 1, 2, ...$;
- $\delta^*_g \sim MVN_T(m_\delta, S_\delta)$, for $g = 1, 2, ...$;
- $g_i \,|\, \pi_\alpha \sim \sum_{g=1}^{\infty} \pi_{\alpha,g}\delta(g)$, for $i = 1, 2, ..., N$, with π_α defined analogously to π in the BFG and BFG-HC procedures; $p(g \,|\, V_\alpha) := \Pr(g_i = g \,|\, V_\alpha) = \pi_{\alpha,g}$, for $i = 1, 2, ..., N$ and $g = 1, 2, ...$, with V_α acting like V in the BFG and BFG-HC procedures;
- $\pi_{\alpha,g} = V_{\alpha,g}\prod_{j=1}^{g-1}(1 - V_{\alpha,j})$, for $g = 2, 3, ...$, and $\pi_{\alpha,1} = V_{\alpha,1}$;
- $V_{\alpha,g} \,|\, \beta_\alpha \sim \text{Beta}(1, \beta_\alpha)$, for $g = 1, 2, ...$;
- $\beta_\alpha \sim \text{Gamma}(a_\beta, b_\beta)$;
- $\theta^*_h \sim MVN_K(m_\theta, S_\theta)$, for $h = 1, 2, ...$;
- $h_i \,|\, \pi_\theta \sim \sum_{h=1}^{\infty} \pi_{\theta,h}\delta(h)$, for $i = 1, 2, ..., N$, with π_θ analogous to π_α; $p(h \,|\, V_\theta) := \Pr(h_i = h \,|\, V_\theta) = \pi_{\theta,h}$, for $i = 1, 2, ..., N$ and $h = 1, 2, ...$, with V_θ analogous to V_α;
- $\pi_{\theta,h} = V_{\theta,h}\prod_{j=1}^{h-1}(1 - V_{\theta,j})$, for $h = 2, 3, ...$, and $\pi_{\theta,1} = V_{\theta,1}$;
- $V_{\theta,h} \,|\, \beta_\theta \sim \text{Beta}(1, \beta_\theta)$, for $h = 1, 2, ...$;
- $\beta_\theta \sim \text{Gamma}(a_\beta, b_\beta)$.

5.2.3 Estimation and inference with the BFG-HC and BFG-HC2 procedures

Appendix 5.A provides implementation details for the BFG-HC and BFG-HC2 procedures. Similarly to the BFG procedure presented in Section 4.2, both methods are implemented as Gibbs samplers (see, e.g., Gelman et al., 2013, pp. 276–278), using slice sampling to avoid truncating the number of groups to some predetermined limit (Kalli et al., 2011; Walker, 2007).[1] Three label-switching moves are used to enhance mixing between modes in the group probability distributions, two of which are proposed by Papaspiliopoulos and Roberts (2008), and the third proposed by Hastie et al. (2015). The samplers generate dependent Markov chains for each parameter block, with sufficiently long chains assumed to adequately explore the joint posterior distribution, so that summary statistics on drawn realisations give accurate inference on the true parameter values. All conditional posteriors for each parameter block are sampled from directly, with the exception of the error variance parameters, which use tailored Metropolis–Hastings (MH; Hastings, 1970) accept–reject steps.

Inference on the parameters of interest, namely the group-specific covariate effects and time profiles, uses the 'exemplar unit' method, proposed for time profile inference with the BFG procedure in Section 4.2. The exemplar unit inference method overcomes a difficulty inherent in analysing the DPMM estimation output, which is that group-specific parameters are not easily identified after estimation. First, the multiple random group allocations from the Gibbs sampler output are converted into a set of 'PEAR-groups'. A set of PEAR-groups is a single, post-processed group allocation using the 'PEAR' method from Fritsch and Ickstadt (2009), as implemented in the `R` package `mcclust`. For the BFG-HC2 method, the only difference compared to the BFG-HC approach is that there are two post-estimation PEAR groupings computed: one corresponding to the groups in the time profiles, and the other for the covariate effects. In the simulation study done for the BFG procedure, the clustering accuracy of PEAR performs marginally better than an approach based on k-means clustering individual-level time profile posterior means—the latter approach is how Kim and Wang

[1] As in Chapter 4, 'Gibbs sampler' is used here even though the implementations are actually Metropolis-within-Gibbs samplers.

(2019) post-process the sampler output to determine point estimates for a fixed number of group time profiles.

Given the PEAR groupings, for each PEAR-group, the exemplar is the unit which, across iterations of the sampler, is most often clustered with other members of the same PEAR-group. Inference on group-specific parameters is then performed by summarising unit-level parameter draws (iterates) produced by the sampler for the exemplar units: point estimates are the posterior mean estimates, computed as the mean of the iterates for the corresponding parameters; 95% central credible interval bounds are estimated as the empirical 2.5 and 97.5 percentiles of the iterates. The simulation study presented in Section 5.3 suggests this method works well for performing inference on group-level parameters for both the BFG-HC and BFG-HC2 methods.

For the BFG-HC procedure, covariate effects at the unit level are extracted directly from the set of iterates $\left\{\phi_i^{(m)}\right\}_{m=1}^{M}$; $\phi_i^{(m)} = \phi_{g_i}^{*(m)}$ is the group-specific parameter vector drawn for unit i in iteration m from a set of M iterations of the sampler, collected after a suitable burn-in (warm-up) period. For analysing the time profiles, while the first-differenced time profile parameters, $\overset{\Delta}{\alpha}_i = \left[\overset{\Delta}{\alpha}_{i,2}, \overset{\Delta}{\alpha}_{i,3}, ..., \overset{\Delta}{\alpha}_{i,T}\right]' = \overset{\Delta}{\alpha}^*_{g_i}$, can be directly extracted from ϕ_i, the $\overset{\Delta}{\alpha}_i$ are manipulated to construct relative time profile values, $\tilde{\alpha}_{it} = \alpha_{it} - \alpha_{i,1}$, using $\tilde{\alpha}_{it} := \sum_{s=2}^{t} \overset{\Delta}{\alpha}_{is}$ for $t = 2, 3, ..., T$. The relative time profiles have an appealing economic interpretation as effects relative to the first period, analogous to the interpretation of the regression coefficients on $T - 1$ time period dummy variables in a classical 2WFE panel data regression model (see, e.g., Wooldridge, 2012, pp. 484–487).

For the BFG-HC2 procedure, PEAR groupings are determined separately for the covariate effects and the time profiles. Conditional on these groupings, independent inference for the covariate effects and time profiles proceeds analogously to the BFG-HC procedure.

5.3 Simulation Study

This simulation study compares two data-generating processes (DGPs), with identical parameters except for the grouping structures. The results suggest distinct situations

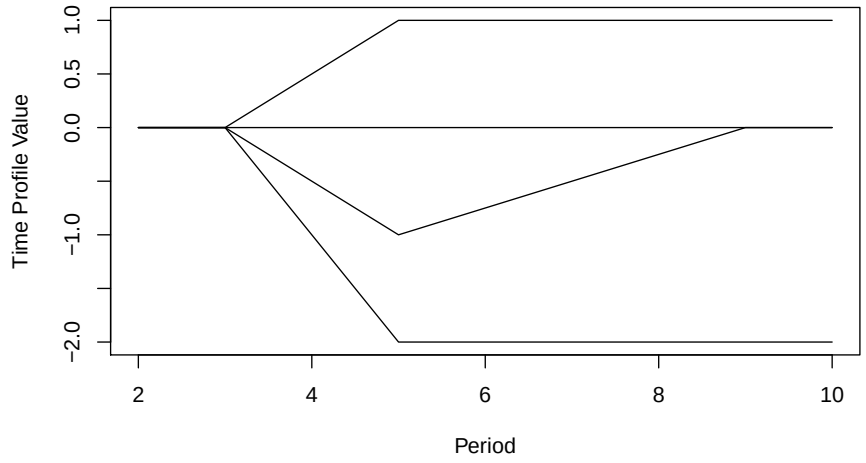
in which the BFG-HC procedure may be preferred to the BFG-HC2 method, and vice-versa. The study compares the two procedures in terms of the magnitude of estimated biases for parameters of interest, root mean-squared error (RMSE), and 95% credible interval coverage probabilities. The bias estimate for a parameter is the mean difference between the point estimate—here, the posterior mean estimate—and the corresponding true DGP value across replicates; the RMSE is the square root of the average squared differences between point estimates and DGP values; the interval coverage probability estimate is the proportion of replicates for which the DGP value lies in the estimated central 95% credible intervals. Point estimates and credible interval estimates are computed using the exemplar unit method outlined in Section 5.2.3.

5.3.1 Design

The first DGP (DGP1) generates data as per (5.1), while the second DGP (DGP2) uses (5.3). For both DGPs, we simulate $N = 1000$ individuals over $T = 10$ time periods, with $K = 2$ relevant covariates and $G = 4$ equally sized groups with distinct group effects. The design generally follows that used for testing the BFG procedure (Chapter 4), but with some modifications to allow studying the novel features of the BFG-HC and BFG-HC2 procedures. Figure 5.1 shows the four relative time profiles, $\alpha^*_{gt} - \alpha^*_{g,1}$, for $t = 2, 3, ..., T$ and $g = 1, 2, 3, 4$; these groups are referred to as the 'up-flat', 'no-change', 'down-up', and 'down-flat' groups, ordered from the largest relative time profile value observed in period 5 to the smallest. As in Chapter 4, these shapes are chosen because they have sensible economic interpretations as varied responses to a common shock, representing either temporary or sustained changes to behaviour. Heteroskedastic errors are generated as $u_{it} \sim N\left(0, \sigma^{2*}_{g_i t}\right)$ independently; group-specific, time-varying error standard deviations either begin at $(\sigma^{2*}_{g,1})^{1/2} = 0.3$ and increase at a constant rate to $(\sigma^{2*}_{g,T})^{1/2} = 0.9$, or vice-versa. The individual fixed effects are given by $c_i \sim N(0,1) + 1$ independently for all i, with an independent set of c_i randomly generated for each replicate.

Under DGP1, the covariate effects, $\theta^*_g = \left(\theta^*_{g,1}, \theta^*_{g,2}\right)'$, are grouped so that each of the time profiles is associated with a unique set of covariate effects. The up-flat group has covariate effects $(0.5, -0.5)$, the no-change group $(0, -1)$, the down-up group $(1, 0)$,

Figure 5.1: Time profiles, $\alpha^*_{gt} - \alpha^*_{g,1}$, for the simulation study. The values are relative to period 1 effects. DGP1 and DGP2 use the same time profiles. These are the same profiles used in the DGP for the BFG simulation study.



and the down-flat group $(-0.5, 0.5)$. These covariate effect configurations are chosen to include as true covariate effects the two configurations found when applying the BFG procedure to the superannuation data.

The covariates, $x_{it} = (x_{1,it}, x_{2,it})'$, are generated independently according to the following distributions, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$: $x_{1,it} \sim N(0,1) + 0.3c_i + 0.3\alpha^*_{g_i t}$; $x_{2,it} \sim N(0,1) - 0.3c_i - 0.3\alpha^*_{g_i t}$. This construction induces correlations between the observable and unobservable determinants of y_{it}, as might be expected to occur in general for observational microeconomic data, which lack experimental controls.

For DGP2, $H = 4$, so that the number of groups in the covariate effects is the same as in DGP1. The possible covariate effect configurations are unchanged, and there are an equal number of units adopting each of the possible configurations. However, among the members of any one of the groups in the time profiles ('g-groups'), there are approximately an equal number of units divided among the groups in the covariate effects ('h-groups'): either 62 or 63. Due to this, there are a total of $GH = 16$ unique 'types' of unit, with each type corresponding to a distinct combination of the time profiles and covariate effects. Since the BFG-HC procedure is unable to model groups in the time profiles and covariate effects independently, a reasonable prior expectation is that it will attempt to estimate 16 groups, one for each of the unique combinations of g- and h-groups.

This simulation study generates $S = 500$ replicate datasets each for DGP1 and DGP2, and runs the BFG-HC and BFG-HC2 estimation procedures on each dataset for both DGPs. The sampling algorithms and initialisation procedures are detailed in Appendix 5.A. The samplers are run for 20,000 iterations, with the first 10,000 discarded as a burn-in sample, and the remaining iterates thinned by a factor of 10, leaving $M = 1000$ iterates of each parameter for post-processing and inference.

The exemplar unit method outlined in Section 5.2.3 is used for inference on the group-specific parameters. In simulation results, the PEAR method for determining post-estimation groupings of the units tends to overestimate the total number of groups in the BFG-HC ϕ_g^* vectors, and in the BFG-HC2 time profiles or covariate effects; however, beyond the four largest groups, additional groups tend to be singletons (groups with only one member), so for generating the comparisons with the DGPs, only the four largest groups for each replicate are considered, with the singletons ignored. To justify this approach, the results below consider what proportion of the sample is covered by the four largest groups across the replicates. More generally, in practical applications, this approach is justified if the researcher is interested primarily in explaining the majority of the systematic variation in the data, and is willing to treat singletons as 'noise', or 'outliers'.

5.3.2 Results

Under DGP1, for the BFG-HC procedure, in 416 (83.2%) of the 500 replicates, 100% of the units are allocated to one of the four largest PEAR-groups; for 78 (15.6%) of the replicates, the corresponding coverage is 99.9%, meaning that only one unit is not allocated to one of the four largest groups; for the remaining 6 (1.2%), the coverage is 99.8%, or two units not in the four largest groups. For the BFG-HC2 procedure, the sample coverage of the four largest g-groups is lower: on average across the replicates, 99.4% of the sample belongs to one of the four largest groups; the 2.5 and 97.5 empirical percentiles of the four-group coverage rates are given by 98.8% and 99.8% of the sample, respectively. For the h-groups, the corresponding mean coverage across all replicates is 99.7%, and the respective percentiles are 98.8% and 100%. Under DGP2, the corresponding BFG-HC2 coverage rate distributions are similar: for the g-groups,

the mean coverage, and the 2.5 and 97.5 percentiles, are the same as for DGP1 to one decimal place; for the h-groups, the respective figures are 99.8%, 99.1% and 100%.

Collectively, these summary results support the approach used to allow for direct comparison between estimated group effects and the DGP values, which is to only consider the four largest groups for the comparisons.

Table 5.1 compares the group classification accuracies for DGP1 and DGP2 across all 500 replicates. The BFG-HC groupings are highly precise and outperform the BFG-HC2 groupings under DGP1; however, under DGP2, the BFG-HC method is consistently unable to identify the 16 types of individuals, making it unclear how to compute classification accuracy. Figure 5.2 depicts the problem by plotting group time profiles and covariate effects combined into one vector for the BFG-HC results for a randomly selected replicate under DGP2, showing the procedure's inability to estimate 16 unique combinations of time profiles and covariate effects with only one grouping dimension. Appendix C gives similar plots for two other randomly selected replicates, supporting the same conclusion.

Table 5.1 also suggests that under both DGPs, the BFG-HC2 h-groupings are more accurate than the corresponding g-groupings. This seems reasonable given that for any dataset size, the influence of a covariate can be observed at T times as many data points as a single time period effect.

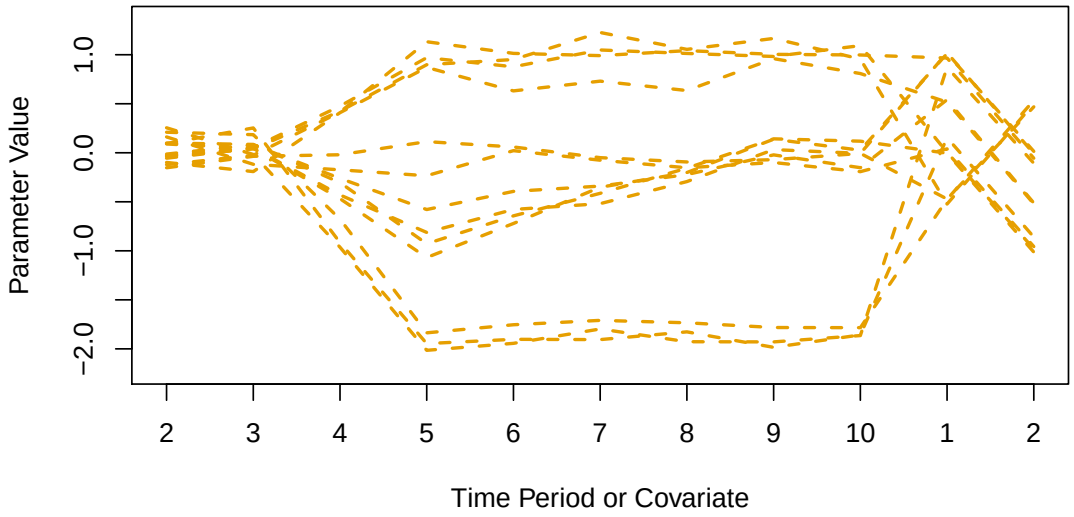
Table 5.2 summarises three comparison metrics across replicates: bias, RMSE, and 95% credible interval coverage probability. Under DGP1, the time profile and covariate effect estimators for the BFG-HC method outperform those from the BFG-HC2 method in

Table 5.1: BFG-HC and BFG-HC2 group classification accuracy across replicates under DGP1 and DGP2.

DGP	Grouping	Mean Accuracy	2.5 Percentile	97.5 Percentile
DGP1	BFG-HC	0.988	0.980	0.994
	BFG-HC2 h-groups	0.942	0.928	0.956
	BFG-HC2 g-groups	0.905	0.886	0.923
DGP2	BFG-HC2 h-groups	0.944	0.928	0.959
	BFG-HC2 g-groups	0.906	0.890	0.924

Note: Group classification accuracy determined by permuting through group labels in the estimation results and selecting the permutation that maximises grouping accuracy relative to DGP group labels.

Figure 5.2: BFG-HC time profile and covariate effect estimates for a randomly selected replicate from DGP2. The first 9 x-axis tick marks, labelled 2–10, represent time periods; the following 2 tick marks, labelled 1 and 2, correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively. The time profile values are relative to period 1.



terms of bias magnitude and RMSE, while both methods perform similarly for interval coverage. Although the results show the overall performance of the BFG-HC method is numerically superior, descriptive plots for randomly selected replicates suggest that differences tend to be small relative to the magnitude of the corresponding effects. Figures 5.3 and 5.4 illustrate this by comparing the BFG-HC and BFG-HC2 estimates of the time profiles and covariate effects, respectively, against the DGP values for one randomly selected replicate. Appendix C presents corresponding plots for two more randomly selected replicates, which also support this claim.

Figures 5.5 and 5.6 give similar plots for one randomly selected replicate under DGP2, while Appendix C provides these plots for two other randomly selected replicates.

Overall, these plots, and similar plots in Appendix C, showing BFG-HC and BFG-HC2 results for randomly selected replicates under both DGPs, are descriptive evidence suggesting that the BFG-HC procedure performs poorly under DGP2, whereas the BFG-HC2 procedure performs similarly under both DGP1 and DGP2. Moreover, the plots suggest that under DGP2, the BFG-HC procedure is consistently unable to identify all latent types, and can give misleading estimates of the group-specific parameters, whereas the BFG-HC2 procedure performs well on the same data. Across both DGPs, BFG-HC2 time profile and covariate effect estimators perform similarly overall.

Table 5.2: Simulation study summary statistics.

Time Profiles:

Measure	Summary Statistic	DGP1 BFG-HC	DGP1 BFG-HC2	DGP2 BFG-HC2
Magnitude of Bias	Mean	0.0016	0.0061	0.0038
	SD	0.0013	0.0049	0.0021
RMSE	Mean	0.0573	0.0634	0.0629
	SD	0.0144	0.0153	0.0156
95% CI Coverage Probability	Mean	0.9476	0.9478	0.9479
	SD	0.0108	0.0106	0.0099

Covariate Effects:

Measure	Summary Statistic	DGP1 BFG-HC	DGP1 BFG-HC2	DGP2 BFG-HC2
Magnitude of Bias	Mean	0.0004	0.0042	0.0042
	SD	0.0002	0.0023	0.0006
RMSE	Mean	0.0112	0.0136	0.0131
	SD	0.0004	0.0008	0.0010
95% CI Coverage Probability	Mean	0.9450	0.9432	0.9518
	SD	0.0086	0.0220	0.0090

Note: Values for time profiles are means and sample standard deviations (SDs) computed across $G(T-1) = 36$ comparison metrics each; values for covariate effects computed over $GK = HK = 8$ comparison metrics each. CI stands for 'credible interval'. For comparison purposes, group labels for estimates of group-specific parameter vectors—ϕ_g^* for the BFG-HC method, and time profiles and covariate effects for the BFG-HC2 method—are matched to DGP group labels using the label-matching procedure proposed in Chapter 3, which finds the permutation of group labels that minimises the sum of Euclidean distances between sets of group-specific parameters with matching labels. This method is similar to the label-matching approach used by Hofmans et al. (2015).

Figure 5.3: BFG-HC and BFG-HC2 time profile estimate comparison with the DGP values for a randomly selected replicate from DGP1. The values are relative to period 1.

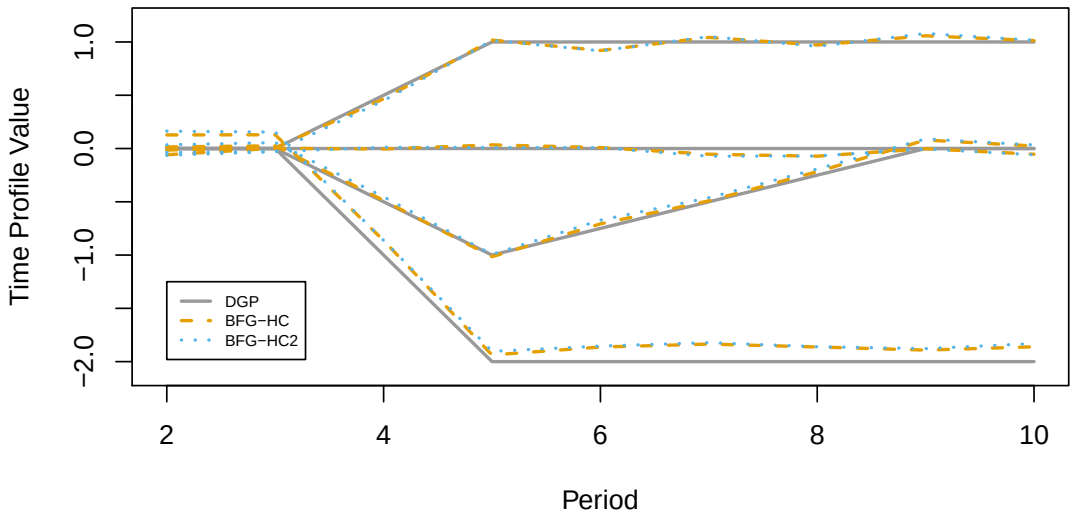


Figure 5.4: BFG-HC and BFG-HC2 covariate effect estimate comparison with the DGP values for a randomly selected replicate from DGP1. x-axis labels 1 and 2 correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively.

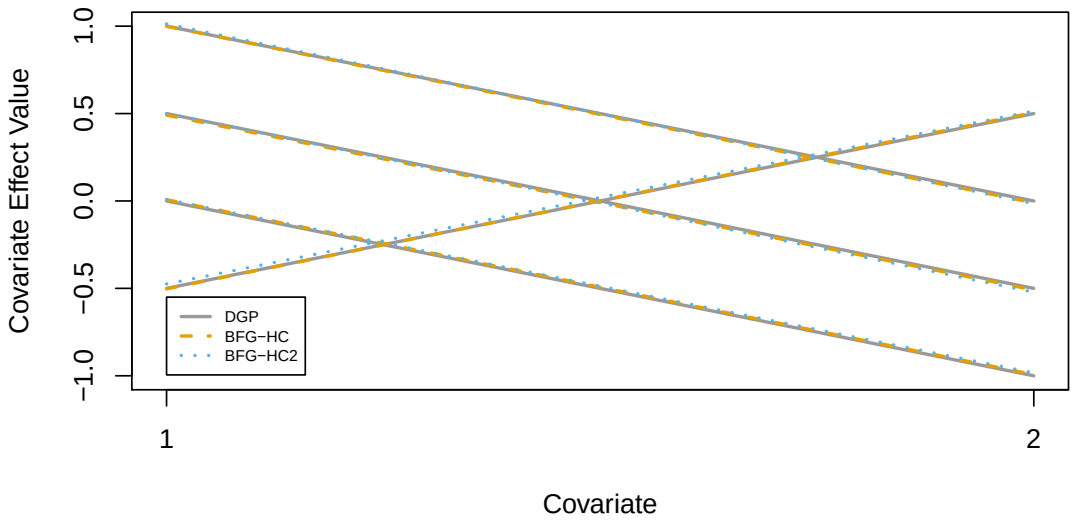


Figure 5.5: BFG-HC and BFG-HC2 time profile estimate comparison with the DGP values for a randomly selected replicate from DGP2. The values are relative to period 1.

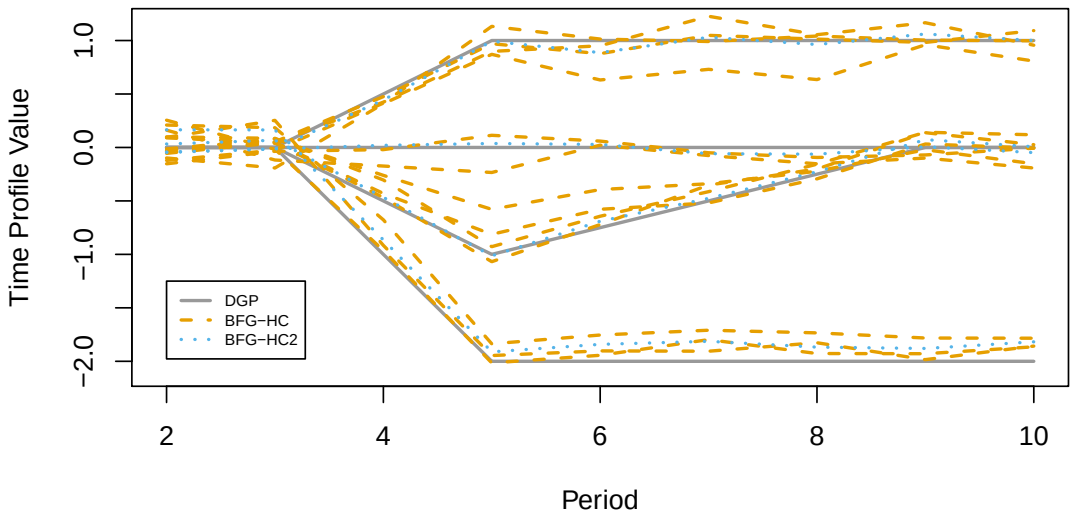
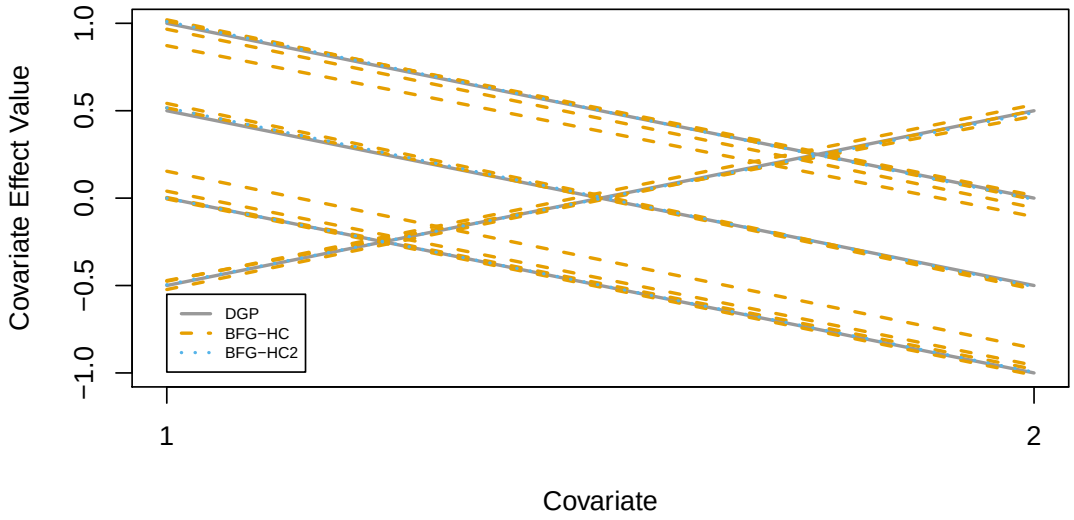


Figure 5.6: BFG-HC and BFG-HC2 covariate effect estimate comparison with the DGP values for a randomly selected replicate from DGP2. x-axis labels 1 and 2 correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively.



5.4 Application 1: Retirement decumulation in Australia

This section extends the application in Chapter 4, which studies drawdowns (withdrawals) from account-based pensions (ABPs) in Australia using the BFG method. ABPs are a relatively flexible option for accessing wealth that is stored in Defined-Contribution accumulation accounts.

The balance of an ABP can remain invested in a combination of safe and risky financial assets, based on the retiree's preferences, and the account balances and investment earnings are subject to preferential tax treatment, so long as the retiree draws down their balance at a rate at least as high as the government-legislated minimum drawdown rates. These drawdown rates are age-dependent, ranging from 5–14% of the corresponding account balance at the start of the financial year to which they apply.

In principle, the preferential tax treatment of superannuation assets means that retirees have an incentive to decumulate other liquid sources of wealth before drawing down their ABPs at a rate higher than the legislated minima.

The available superannuation dataset reports drawdown experience in ABPs for $N = 9516$ retirees for up to $T = 12$ years and is compiled from multiple large industry and retail superannuation funds in Australia. This dataset is investigated in Chapter 3 using the classical Grouped Fixed-Effects (GFE) estimator from Bonhomme and Manresa

(2015), and the previous BFG analysis which uses a random balanced subsample of size $N = 1000$ of the data, due to limitations of the existing implementation. In this context, 'balanced' means dropping from the panel any individuals with missing data in the dependent variable or covariates.

The dataset used in this application is the same as for the BFG application, because similar limitations apply to the present implementations of the BFG-HC and BFG-HC2 procedures. The use of an identical dataset as for the BFG application also allows for a direct comparison of the estimation results on the same set of observational units. Table 5.3 provides summary statistics for this estimation.

The estimated model is

$$\ln \mathrm{DR}_{it} = c_i + \theta^*_{1,h_i} \ln \mathrm{MDR}_{it} + \theta^*_{2,h_i} \ln \mathrm{AB}_{it} + \alpha^*_{g_i t} + u_{it}, \tag{5.6}$$

for the BFG-HC2 procedure; the BFG-HC approach enforces the special case where $h_i = g_i$ for all i. With the exception of allowing heterogeneous regression coefficients, θ^*_{1,h_i} and θ^*_{2,h_i}, (5.6) is the same equation estimated in the previous BFG application, as well as the GFE application in Chapter 3: $t = 2004, 2005, ..., 2015$ indexes financial years, starting on July 1 each year; DR_{it} is the chosen rate of the drawdowns, expressed

Table 5.3: Summary statistics for the superannuation application. This table is reproduced from Chapter 4 (Table 4.2).

	Drawdown Rate	Drawdown Amount	Account Balance	Age at 31 Dec 2015	Age at Account Open	Sex: Male
Mean	0.11	6249.23	70,118.70	80.22	63.55	0.59
SD	0.11	5915.88	74,195.34	4.37	4.06	0.49
Median	0.08	4656	51,046	80.38	64.34	1
Q1	0.07	2976	31,366	77.55	60.99	0
Q3	0.12	7428	82,908	83.32	65.43	1
Min	0.01	1	1	67.67	50.85	0
Max	1.33	84,422	1,274,233	92.5	78.32	1
Count	12,000	12,000	12,000	1000	1000	1000

Note: The log of the drawdown rate is the dependent variable. The account balance is defined as the position at the start of the relevant financial year; this is the denominator for the rate computations. The age at 31 December 2015 represents the individual's cohort, equivalent to measuring a year-of-birth variable. The age at account opening is the age when the retiree initiates a phased withdrawal product and begins drawing down from the account. The sex indicator variable equals 1 if the retiree is male, and 0 otherwise. SD refers to the sample standard deviation of the variable. Q1 and Q3 refer to the empirical 1st and 3rd quartiles, respectively.

as a proportion of the account balance at the start of the respective financial year, that the retiree has nominated, in advance of the financial year, to receive at a regular frequency over the corresponding twelve-month period; MDR_{it} is the age-dependent legislated minimum drawdown rate that the retiree must match or exceed in order to retain preferential tax treatment of their superannuation assets; AB_{it} is the retiree's account balance at the start of the relevant financial year; the model error is assumed to be independently distributed as $u_{it} \sim N\left(0, \sigma^{2*}_{g_it}\right)$. In addition to the prespecified 'regular' drawdowns DR_{it}, retirees can make 'ad-hoc' withdrawals from their accounts at any time during the financial year. These ad-hoc drawdowns are not studied here.

The feature from the BFG results that motivated the BFG-HC and BFG-HC2 methods is that inference performed with the BFG procedure on this sample had covariate effect parameters that became trapped in local modes close to either $(\theta_1, \theta_2) = (0, -1)$ or $(1, 0)$. The BFG application shows that substituting the $(0, -1)$ covariate effects reduces the estimated model to

$$\ln \text{DA}_{it} = c_i + \alpha^*_{g_it} + u_{it}, \tag{5.7}$$

where DA_{it} is the dollar amount of the drawdowns requested by unit i for receipt in financial year t, while the $(1, 0)$ covariate effects reduce the model to

$$\ln \left(\frac{\text{DR}}{\text{MDR}} \right)_{it} = c_i + \alpha^*_{g_it} + u_{it}. \tag{5.8}$$

Both (5.7) and (5.8) imply different economic interpretations of the group time profiles, α^*_{gt}, estimated by the model: in (5.7), the estimated time profiles are interpreted as changes in drawdown amounts, not drawdown rates; in (5.8), they represent changes in how drawdown rates evolve relative to the age-dependent minimum drawdown rates, in proportional terms. Indeed, neither of these are the interpretation implied by the original model, (5.6), where they are the partial effects of grouped heterogeneity on the absolute drawdown rates—not relative to the minima, and not in dollar terms.

The implementation details for this chapter are as follows. The procedure for generating starting values is presented in Appendix 5.A. The BFG-HC and BFG-HC2 procedures are each run five times on the estimation sample; each run supplies different randomly generated starting values to the sampler, which is run for 120,000 iterations; the

first 90,000 iterations are discarded as a burn-in sample; to conduct post-estimation analysis, the remaining iterations are thinned by a factor of 30, leaving a final sample of $M = 1000$ 'production' iterates, with sampled parameter values in these production iterates treated as samples from their posterior distributions for the purpose of statistical inference. As in the BFG application, visual assessment of the output suggests that the distribution of the number of groups has stabilised within each run by iteration 90,000.

The BFG-HC results show duplication of common covariate effect estimates across different PEAR-groups. Among the 8 largest groups in any given run—which capture between 79–86% of the sample, across the five independent runs—some groups have estimated posterior means for the covariate effects, $\left(\theta^*_{1,g}, \theta^*_{2,g}\right)$, close to $(0, -1)$, while the remaining groups have values close to $(1, 0)$. Of these 8 largest groups, the estimated covariate effects which deviate the most from one of these two pairs of values across the runs are $(0.954, -0.005)$. The duplication of common covariate effects in the BFG-HC results suggests a preference towards the BFG-HC2 method, which, by construction, allows for sharing common covariate effects across distinct time profile groups.

Across the five independent BFG-HC2 runs, the two largest groups in the covariate effects (h-groups) cover between 80–88% of sample. The estimated covariate effects for these two h-groups using the exemplar unit method are always precisely estimated as either $\left(\widehat{\theta}^*_{1,h}, \widehat{\theta}^*_{2,h}\right) = (0, -1)$ or $(1, 0)$, which is a reasonable prior expectation based on the BFG-HC results, as well as the BFG results. Similarly to the BFG results, these covariate effect configurations define specific, reduced form versions of (5.6): (5.7) or (5.8).

The 11 largest groups in the time profiles (g-groups) cover 78–85% of the sample across the independent runs, followed by small groups and singletons: across the five runs, the 12th largest group in any single run contains at most 3% of the sample. Appendix C examines group membership probability estimates at the unit-level, showing that the largest groups appear well separated, in the sense that there are generally few units with non-negligible probability mass on g-groups other than their own.

Since a) the same time profile can be adopted by units with different covariate effects when using the BFG-HC2 method, and b) in this application, the time profiles have different interpretations if the covariate effect configurations are close to $\left(\theta^*_{1,h}, \theta^*_{2,h}\right) = (0, -1)$ or $(1, 0)$, the economic interpretation of the group behaviours differs depending on whether the time profiles represent the time-varying intercepts in (5.6), (5.7) or (5.8). The different possible interpretations for the same time profiles motivates a three-partition, or 'tripartite', analysis of the BFG-HC2 time profiles, involving two main steps. First, the sample is partitioned into three, with each partition having a distinct interpretation of time profiles in its corresponding model. The sample can be split in this way at the unit level by considering the estimated posterior means of each unit's covariate effects, $\left(\widehat{\theta}_{1,i}, \widehat{\theta}_{2,i}\right) = \left(\widehat{\theta}^*_{1,h_i}, \widehat{\theta}^*_{2,h_i}\right)$. Second, the substantive time profiles found among the members of each partition are interpreted conditional on the model.

The three partitions are defined as:
$\mathcal{P}_1 = \left\{i \in \{1, 2, ..., N\} : \left(\widehat{\theta}_{1,i}, \widehat{\theta}_{2,i}\right) = (0 \pm 0.1, -1 \pm 0.1)\right\}$, which are units the results suggest may consider drawdowns in terms of dollar amounts; $\mathcal{P}_2 = \left\{i \in \{1, 2, ..., N\} : \left(\widehat{\theta}_{1,i}, \widehat{\theta}_{2,i}\right) = (1 \pm 0.1, 0 \pm 0.1)\right\}$, containing units that are likely to be following a drawdown rate pattern and making reference to the minimum drawdown rates; $\mathcal{P}_3 = \{1, 2, ..., N\} \setminus \{\mathcal{P}_1 \bigcup \mathcal{P}_2\}$, with the units who do not align to specific covariate effect configurations, and have the intended interpretation of the time profiles as effects on the drawdown rates, holding the covariates constant. $\mathcal{P}_1 \bigcup \mathcal{P}_2 \bigcup \mathcal{P}_3 = \{1, 2, ..., N\}$, and $\mathcal{P}_a \bigcap \mathcal{P}_b = \varnothing$ for $a \neq b$; i.e., the three partitions of the sample are collectively exhaustive and mutually exclusive.

The rest of this section presents the tripartite results from the first run; however, the results are economically similar across all five independent runs. Broadly, the results find all the behaviours identified with the GFE and BFG procedures, in addition to new behaviours not found using those methods.

Figure 5.7 plots the estimated posterior means, $\left(\widehat{\theta}_{1,i}, \widehat{\theta}_{2,i}\right)$, by partition for all $N = 1000$ units. The sizes of the three partitions, as a proportion of the sample covered, are 68.2%, 16.4% and 15.4%, respectively, for $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$. By construction, data points for $\mathcal{P}_1$ and $\mathcal{P}_2$ are concentrated near the coordinates $(0, -1)$ and $(1, 0)$, respectively. For $\mathcal{P}_3$, the distribution of points indicates that a substantial number of partition members

have small or zero estimated sensitivity to changes in the minimum drawdown rates (x-axis), but among these members, there is considerable heterogeneity in responsiveness to changing account balances (y-axis).

Figure 5.8 shows the subset of the largest 11 exemplar unit time profiles corresponding to the g-groups that are associated with units in Partition 1, where the plotted relative time profile values represent changes in the (log) drawdown amounts. As well as tending towards drawing a constant dollar amount over time, potentially with a downward revision, which are behaviours previously found with the GFE and BFG procedures, Figure 5.8 also contains groups whose time profiles suggest some individuals adjusted their drawdown amounts coincident with the implementation, and subsequent unwinding, of temporary concessional minimum drawdown rates, applying in the 2009–2013 financial years.

Figure 5.9 gives the corresponding time profiles for Partition 2, where movements in the time profiles represent proportional changes in the retiree's chosen drawdown rate in excess of the legislated minima. The timing and relative magnitudes of the step-like movements in the first panel of Figure 5.9, as well as the collapsing of the credible intervals from 2011 onward, suggest a group of retirees initially following the minimum rates, but continuing to draw at the non-concessional minimum rates during the period when the concessional rates were active. This behaviour may be intentional, but it could also be that some retirees drawing at the minima were unaware of the temporary reduction in the legislated rates, and so missed an opportunity to slow the depletion

Figure 5.7: All individual-level estimated BFG-HC2 covariate effect posterior means by sample partition for the superannuation application. The respective partition sizes are 68.2%, 16.4% and 15.4% of the sample. Individual dots are partly transparent so that darker regions correspond to a higher concentration of data points.

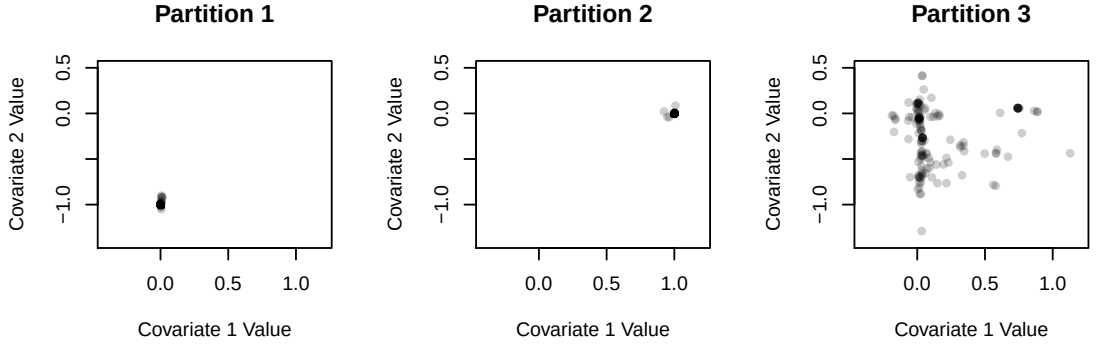
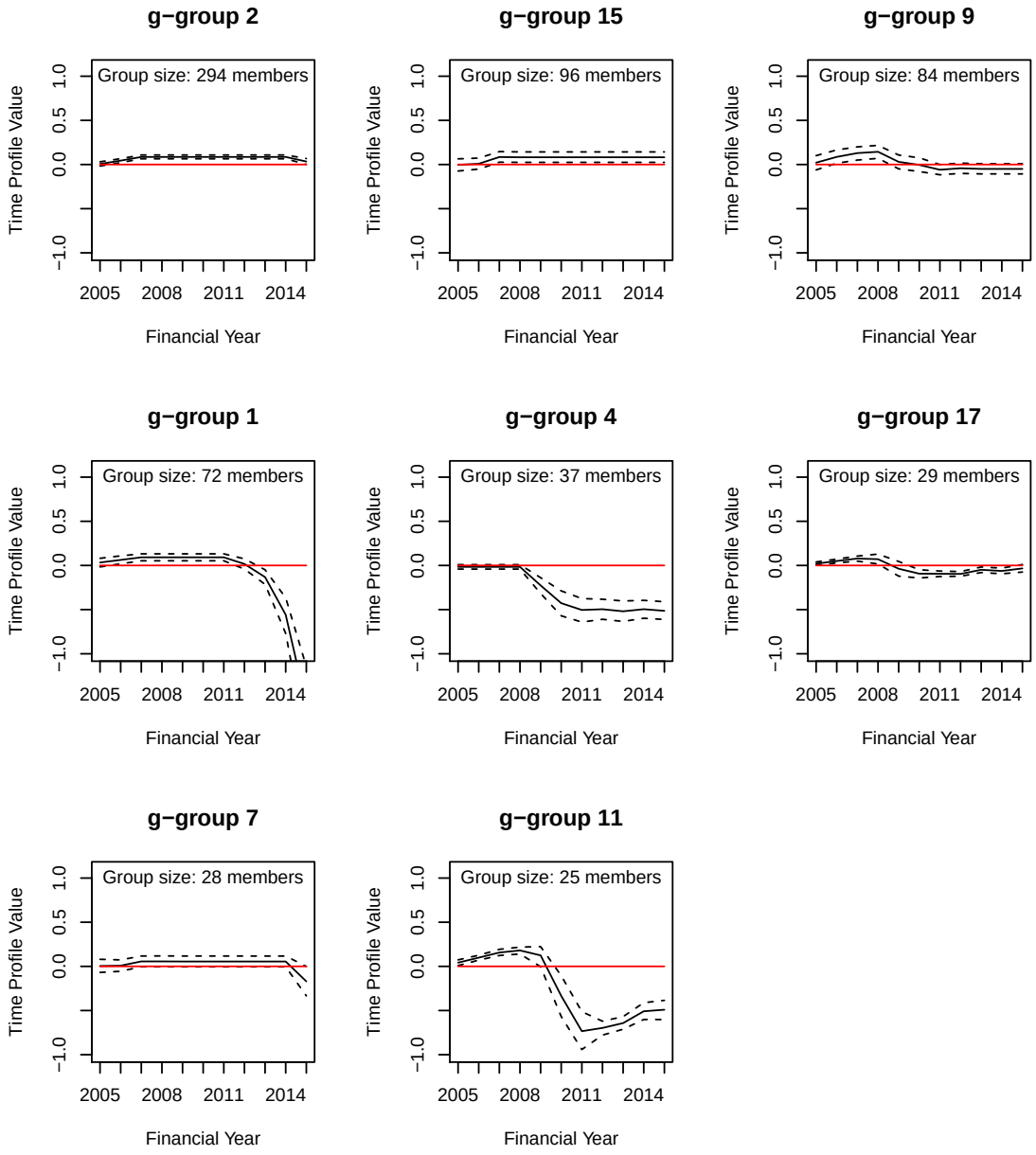


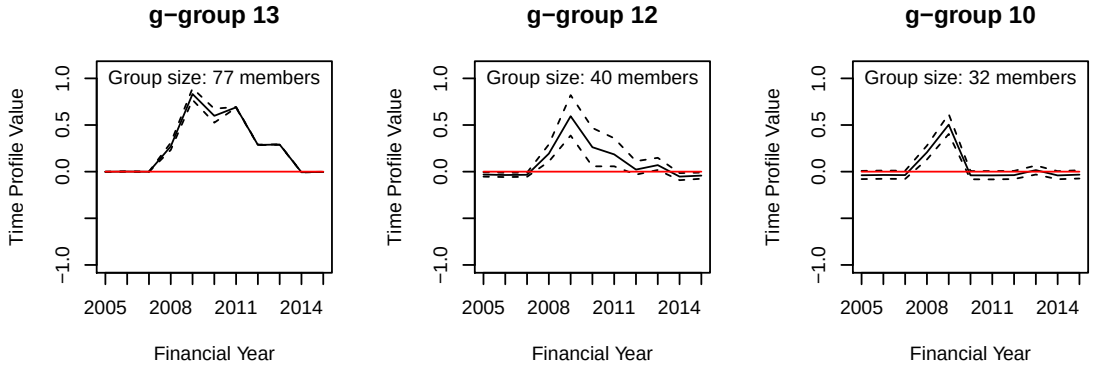
Figure 5.8: Substantive BFG-HC2 time profiles for Partition 1 in the superannuation application. The values are relative to 2004 effects. The g-groups are listed in order of size from left to right, top to bottom. The solid black lines depict the g-group exemplar unit posterior mean estimates; the dashed lines connect element-wise 95% credible intervals. Zero lines are plotted in red.



of their balances in the years immediately following the Global Financial Crisis (GFC). In addition to this behaviour, which is not identified in the previous GFE and BFG analyses, the latter two panels in Figure 5.9 are more consistent with individuals that follow the minima and adopt the concessional minima, although with an apparent lag in adoption time.

Unlike Partitions 1 and 2, the time profiles associated with units from Partition 3 have the standard rate-based interpretation intended by the initial model formulation, (5.6). To varying degrees, units in this partition are sensitive to changes in the minimum drawdown rates and their account balances when determining their rate of drawdown, but there may also be systematic changes in the drawdowns driven by unobserved factors, including latent behavioural preferences and responses to macroeconomic conditions and policy shocks. Because no single BFG run escapes either of the local modes in the covariate effect configurations that define Partitions 1 and 2, the BFG procedure is unable to estimate a model appropriate for the units in Partition 3. Figure 5.10 gives the substantive time profiles associated with units from Partition 3, showing groups that have a tendency to make adjustments to their drawdown rates coincident with the introduction of the temporary concessional minima. One interpretation of the kind of retirees in this group is that they are the most engaged and responsive to relevant factors determining their drawdowns: their covariate effects imply adjustments based on the increasing minimum drawdown rates over time as well as changing account

Figure 5.9: Substantive BFG-HC2 time profiles for Partition 2 in the superannuation application. The values are relative to 2004 effects. The g-groups are listed in order of size from left to right, top to bottom. The solid black lines depict g-group exemplar unit posterior mean estimates; the dashed lines connect element-wise 95% credible intervals. Zero lines are plotted in red.



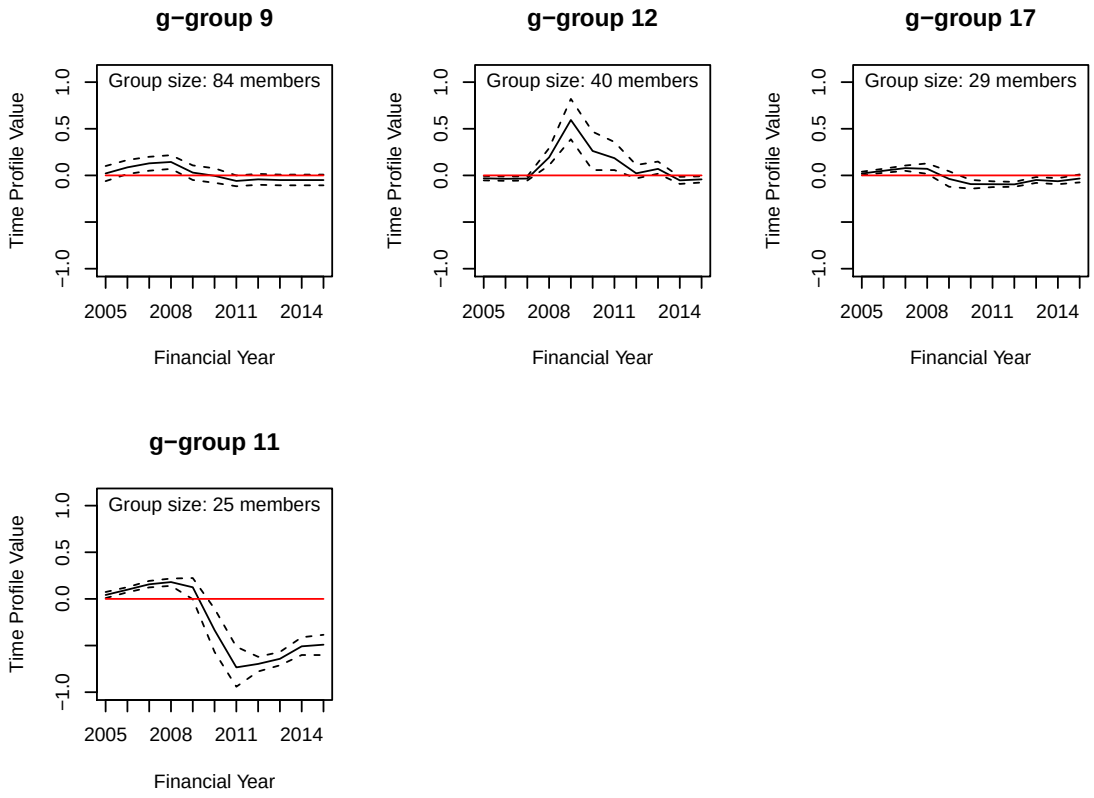
balances, which are both reasonable factors to include in decision-making; their time profiles suggest a responsiveness to unobserved factors, which, based on the timing of shifts in the time profiles, are likely to relate to macroeconomic events such as the GFC, or government policy-related shocks.

5.5 Application 2: Smoking policy in Australia

In recent decades, the Australian government has enacted several policies targeting a reduction in smoking.[2] One of these policies is the plain-packaging mandate (PPM), introduced in December 2012, which requires that cigarette packaging use uniform font and colour, and display graphic health warnings on the majority of the packaging

[2]The Australian Government Department of Health's tobacco control timeline is available on the Department of Health website: https://www1.health.gov.au/internet/publications/publishing.nsf/Content/tobacco-control-toc~timeline. Page viewed 25 October 2021.

Figure 5.10: Substantive BFG-HC2 time profiles for Partition 3 in the superannuation application. The values are relative to 2004 effects. The g-groups are listed in order of size from left to right, top to bottom. The solid black lines depict g-group exemplar unit posterior mean estimates; the dashed lines connect element-wise 95% credible intervals. Zero lines are plotted in red.



(Bonfrer, Chintagunta, Roberts, & Corkindale, 2020). Using aggregated scanner data on cigarette sales, Bonfrer et al. (2020) estimate the causal effect of the PPM as a roughly 7.5% decrease in monthly sales volume of cigarettes (sticks), which can be considered the average treatment effect of the PPM intervention. Their method relies on a Difference-in-Differences strategy (see, e.g., Angrist & Pischke, 2009), using New Zealand as the control group.

Bonfrer et al. (2020, p. 235) state that due to data limitations, "heterogeneous effects across consumers or segments cannot be explored." The search for heterogeneous impacts of a policy on latent subgroups, in the style of an event study, is one of the motivating scenarios for the models proposed in this chapter. While these models are not necessarily able to isolate causal effects, they are able to characterise differential behaviours in response to both observed covariates and common shocks occurring at a known time: i.e., heterogenous effects across consumers. Since the period surrounding the PPM was subject to multiple policy changes, these methods are unlikely to isolate the effect of the PPM alone; however, it can be argued that any group behaviours that are found may be useful in understanding distinct changes in smoking behaviours over time—as net effects of multiple related policies—while flexibly controlling for the estimated effects of observable characteristics.

The Household, Income and Labour Dynamics in Australia (HILDA) survey (Watson & Wooden, 2012) tracks households over successive annual waves, collecting a range of demographic and behavioural data. Using data from 2009–2018, we construct a 10-year panel that allows searching for heterogeneous patterns in smoking behaviours over time, controlling for a small set of potentially relevant demographic covariates. The aim of this application is not to produce a thorough empirical analysis, but rather to outline how practitioners or applied researchers might use the methods presented in this chapter in a more general class of applications.

Here, the HILDA data is filtered to keep only individuals that stated, at least once in the range 2009–2018, that they smoked a nonzero number of cigarettes per week, on average, over the year. The three demographic covariates selected are: a) net income, in thousands of AUD; b) self-assessed health, rated on a scale from 1–5, with 5 being the healthiest rating; c) a binary indicator for whether there was a resident child living

at the respondent's premises.[3] Units with missing data for the dependent variable, or any of the covariates, are dropped from the sample, resulting in a balanced panel with $N = 1125$ and $T = 10$. Appendix C lists all filters applied to the raw HILDA data to create the estimation sample.

Table 5.4 reports summary statistics on this sample. Because the distribution of the dependent variable is skewed and left-censored, inferential conclusions from this application are likely to be inaccurate, so we reiterate that this application is illustrative, and not intended to be taken as complete empirical analysis of the problem. Instead, the results motivate further research: empirically, into the relationship between behavioural and policy effects on smoking patterns, and methodologically, into more sophisticated versions of the models for such data.

As a point of comparison for the main results, the frequentist 2WFE estimator is applied to the smoking data, giving a single set of estimates of the time profile and covariate effects. The 2WFE estimator is analogous to estimating (5.1) under the more restrictive assumption that there are no latent subgroups in the sample with heterogeneous responses to covariates or smoking policy: $y_{it} = c_i + \alpha_t + x_{it}'\theta + u_{it}$.

[3] A parent or guardian of a resident child may consider the potential impacts of their smoking habit on the child.

Table 5.4: Summary statistics on the HILDA smoking application sample.

	Weekly Number of Cigarettes Smoked	Net Income (AUD 000s)	Self-Assessed Health Rating	Resident Child Indicator
Mean	55.3	52.1	2.8	0.43
SD	65.4	52.2	0.93	0.49
Median	30	43	3	0
Q1	0	22	2	0
Q3	100	67	3	1
Min	0	-444	1	0
Max	750	1505.9	5	1
Count	11250	11250	11250	11250

Note: The weekly number of cigarettes smoked is the dependent variable of interest. Annual income is reported in thousands of AUD. Business and investment losses can result in negative annual income. The health rating takes integer values from 1–5. The resident child indicator variable takes value 1 if the respondent has a resident child living at their premises, and 0 otherwise. SD refers to the sample standard deviation of the variable. Q1 and Q3 refer to the empirical 1st and 3rd quartiles, respectively.

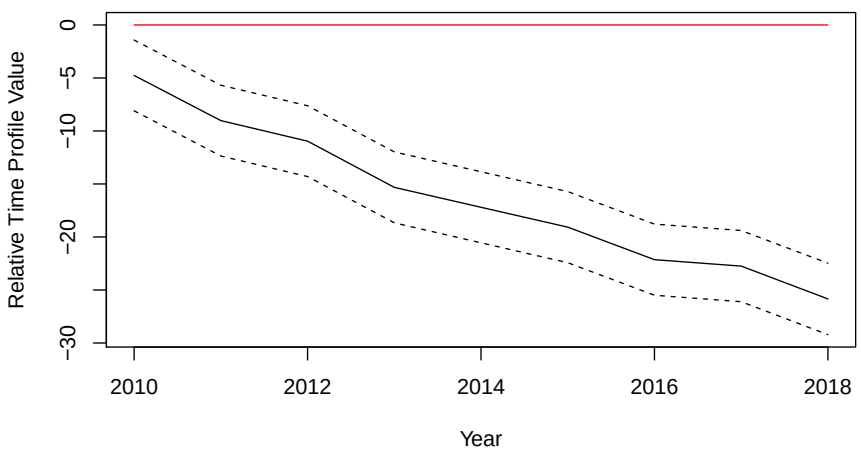
This approach provides an example of the type of results one might find by using a traditional panel data approach to investigate possible policy effects in this application.

In the 2WFE results, the resident child indicator covariate is statistically insignificant at common levels of significance, with a p-value of 0.78. The income and health rating covariates are statistically significant at the 5% level, but the estimated economic effects are quite weak: a $100,000 increase in net income is associated with an estimated increase in smoking by an average of 2.7 sticks per week; a one-point increase in the five-point health rating is associated with an estimated average increase of 1.6 sticks smoked per week. These effects are estimated on a sample of current or potential smokers, and may contradict intuitions about smoking that apply to the general population.

Figure 5.11 plots the corresponding 2WFE relative time profile estimates. The estimated average 10-year change in smoking behaviours, after controlling for the partial effects of the covariates, is a reduction of approximately 26 sticks smoked per week. This effect is statistically significant at the 5% level, and economically large: the mean number of cigarettes smoked, across the entire sample, is roughly 55 sticks per week.

The BFG-HC2 procedure is run on the same data, with starting values initialised using the method in Appendix 5.A. A more complete analysis of the data would compare the results from both procedures proposed in this chapter; however, for brevity, only the BFG-HC2 results are considered here.

Figure 5.11: 2WFE time profile in the smoking application. The values are relative to the 2009 effect. The dashed lines indicate 95% confidence interval bounds.



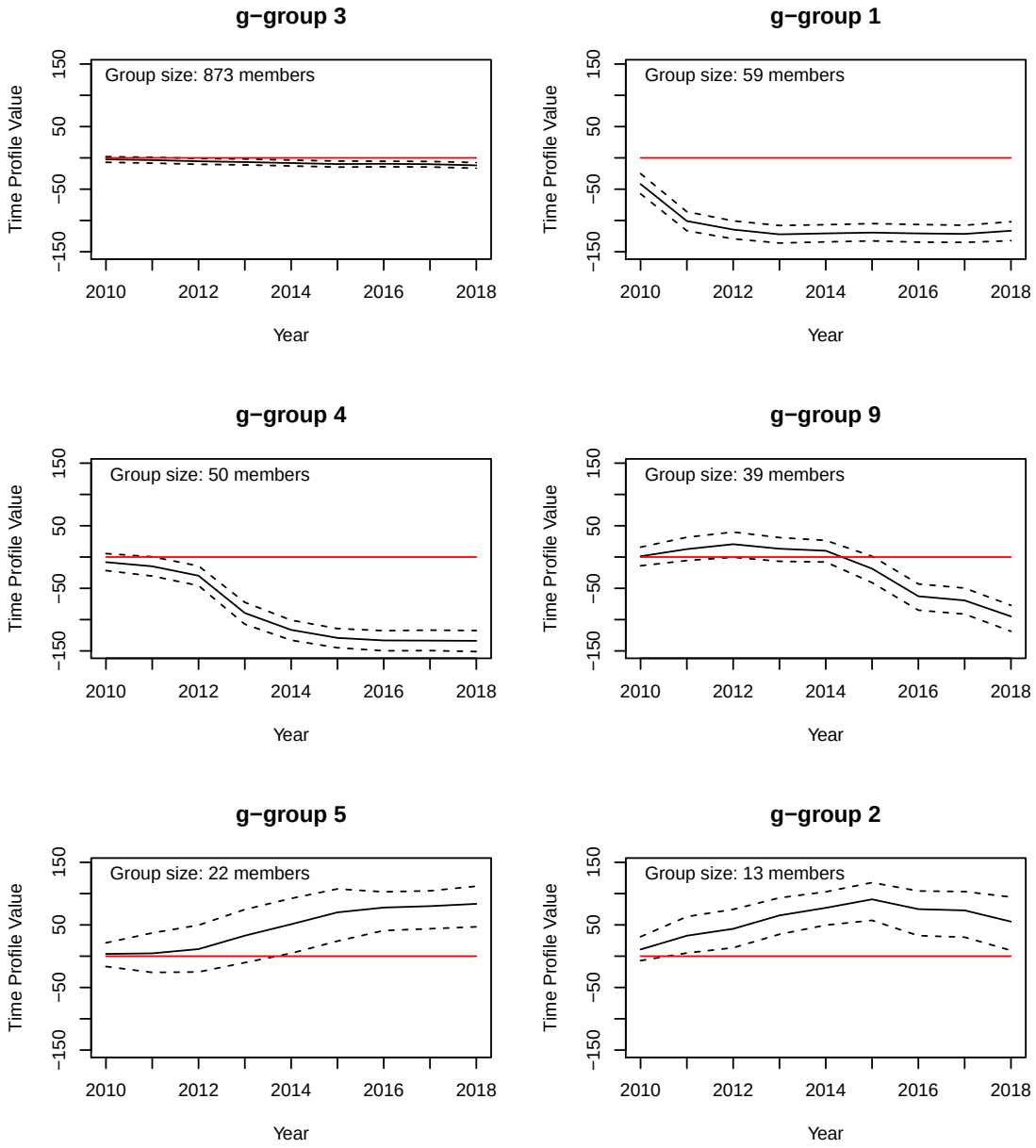
To check the stability of the results, the BFG-HC2 procedure is run five times for 120,000 iterations, with the first 90,000 iterations discarded, and the remaining iterations thinned by a factor of 30, leaving a set of $M = 1000$ production iterations. Statistical inference uses sample statistics computed on the drawn parameter values in the production iterations.

Across the five independent runs, the economic results are similar. The largest h-group (covariate effect group) covers 96.4–97.5% of the sample. In all five runs, estimated covariate effect magnitudes for the largest group are all closer to 0 than the 2WFE results, which are already economically small. Furthermore, the estimated 95% credible intervals all contain the value 0. The remaining h-groups have relatively negligible size in all runs: the next-largest h-group has at most four members, and singletons begin from the third-largest h-group onward. Thus, the covariate effect results suggest that for the majority of the sample, all significant variation in smoking behaviours over time at the individual level is driven either by individual-specific (time-invariant) characteristics and traits, or by latent group time effects.

The latent time effects are concentrated into a handful of groups: across the five runs, the largest g-group (time profile group) covers 77.1–77.6% of the sample; the largest 6 g-groups cover 92.6–94.0% of the sample; the 7th largest group has at most 10 individuals; the 8th largest has at most 3. Figure 5.12 plots the 6 largest time profiles for one of the independent runs, with this run being representative of the results in all five runs. In contrast to the superannuation application, different covariate effect configurations do not impact the interpretation of the time profiles here. Thus, all time profiles have the intended interpretation as changes in the expected number of cigarettes smoked compared to 2009, holding the effects of the included covariates constant.

The largest group exhibits the weakest average effects, which are comparatively dwarfed by the effect magnitudes for the other five groups. However, the first profile is most similar to the single time profile estimated by 2WFE. Figure 5.13 compares these two sets of point estimates: the estimated average 10-year effect on the largest g-group is approximately 12 fewer sticks per week (approximate 95% credible interval: $[8, 17]$), which is roughly half of the corresponding 2WFE estimate of 26 fewer sticks (approximate 95% confidence interval: $[22, 29]$). Furthermore, the kinks produced in

Figure 5.12: Substantive BFG-HC2 time profiles for the smoking application. The values are relative to 2009 effects. The g-groups are listed in order of size from left to right, top to bottom. The solid black lines depict g-group exemplar unit posterior mean estimates; the dashed lines connect element-wise 95% credible intervals. Zero lines are plotted in red.



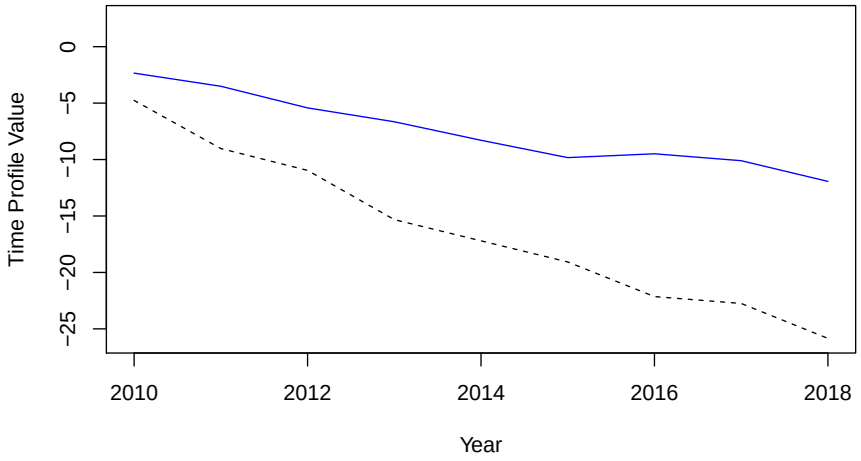
the 2WFE time profile due to faster estimated declines in years 2010, 2011, 2013, 2016 and 2018, compared to other years, are absent from the largest g-group profile; however, the timing of those kinks align with much larger movements made by other, smaller g-groups. Whereas the kinks in the 2WFE profile might reasonably be attributed to a sequence of policy changes creating small, temporary increases in the speed at which cigarette usage reduced, the BFG-HC2 results suggest a different story: most smokers, belonging to the largest g-group, continually decreased their consumption at a relatively steady rate until 2015, while the next three largest groups of smokers at some point began to decrease consumption at a much faster rate. This interpretation is corroborated by examining the proportion of individuals in each of the first four groups that quit smoking over the 10-year window: these proportions are approximately 26%, 70%, 72% and 64%, respectively, for the four largest g-groups, ordered by size. These figures are the proportion of units in these four groups that: a) reported a nonzero weekly number of sticks in 2009; b) reported 0 sticks per week in 2018.

Due to the timing at which the largest reduction in sticks smoked occurred for g-group 4 in Figure 5.12, it is possible that this group was immediately impacted by the PPM introduced in December 2012; however, because similar jumps are seen for two other groups, this alignment in timing could be entirely coincidental, meaning that this group may have chosen that point in time to quit for unrelated reasons. This alternative explanation is explored further in Appendix C.

The 2WFE results also hide the fact that some nonsmokers may have taken up smoking over the same period, as suggested by the last row of panels in Figure 5.12. In the fifth- and sixth-largest g-groups, approximately 46% and 39% of units, respectively, are nonsmokers in 2009 and smokers in 2018—computed as the proportion of units in these two groups that: a) reported 0 sticks per week in 2009; b) reported a nonzero number of sticks per week in 2018.

These results corroborate the idea that the estimated 2WFE time profile acts like an unrepresentative average over multiple distinct latent subgroups. The 2WFE profile masks meaningful changes exhibited by several smaller groups, and is larger in magnitude than the most common behavioural trend, by a factor of 2.

Figure 5.13: Comparison of the 2WFE time profile and the largest BFG-HC2 g-group time profile in the smoking application. The values are relative to 2009 effects. The solid blue line depicts the largest g-group exemplar unit posterior mean estimates; the dashed black line plots the corresponding 2WFE profile.



Appendix C contains additional results to complement the smoking application presented here, including: a) results from a 2WFE estimation of a linear probability model for binary smoking outcomes, run on a sample that does not select units based on the dependent variable value; b) a 'placebo test', which runs the BFG-HC2 procedure on data collected before the introduction of the PPM, to observe whether similar patterns of quitting and commencing a smoking habit are seen before the policy; c) estimates of the distribution of group membership probabilities across the sample.

5.6 Discussion

5.6.1 BFG-HC vs. BFG-HC2

The simulation study in Section 5.3 suggests that if each group in the time profiles corresponds to only one group in the covariate effects, the BFG-HC procedure can outperform the BFG-HC2 approach; however, an economically significant difference is only seen in clustering accuracy. In the less restrictive case where units with the same time profile can still exhibit heterogeneity in their response to covariates, the simulation study suggests that the BFG-HC results can be unreliable, while the BFG-HC2 procedure performs well. Given that BFG-HC2 inference appears similar overall to BFG-HC inference under the more restrictive set of assumptions, the BFG-HC2

procedure is likely to be the preferred approach in general; however, in cases where the more restrictive assumption is defensible, then the BFG-HC procedure may be preferred for its superior performance in classifying individuals into latent subgroups. The similarity in BFG-HC and BFG-HC2 time profile and covariate effect estimates under the first DGP considered, and their noticeable differences under the second, suggests that running both procedures and comparing output may help to defend the assumptions made by the researcher regarding the structure of the latent groups.

5.6.2 Superannuation in Australia

The novel superannuation results in this chapter raise questions that can motivate future applied research into drawdown behaviours. The first question is what type of individual has a preference for following the legislated minimum drawdown rates but did not adopt the temporary concessional rates from 2009–2013. These individuals were not affected by this government intervention, so it may be of interest to policymakers or practitioners to understand the demographic and financial characteristics of such individuals. An open-response survey could collect such data, as well as inquire about the motivations for not using the concessional rates. While some retirees may have done so consciously, others may have been entirely unaware of the concessional minima, which is an argument to explore methods for improving communication between fund trustees and members.

A second question is whether the retirees allocated to Partition 3 are a collection of 'engaged' retirees, in the sense that they actively monitor their own account balances and broader economic concerns, and use these factors to determine their optimal rate of drawdown. If so, understanding whether there is a way to predict engagement status of an arbitrary new retiree can be useful for fund trustees and policymakers. Again, a survey may be the tool to collect data to inform such analysis.

5.6.3 Smoking policy

The smoking application analyses a select sample, not only because the panel is balanced, but also because only people that reported smoking at some point in the

period 2009–2018 are included. Since the PPM is intended to reduce the number of sticks lit by smokers, or to incentivise smokers to quit entirely, having individuals selected into the sample based on smoker status seems appropriate. While this chapter uses a sample of current or potential smokers, an alternative way to construct a dataset for such analysis is to take only those who were smokers at or before 2012, the year in which the PPM was introduced; however, the current set of filters applied to the raw data also allows for potentially observing new smokers commencing the habit after the policy was implemented, and these individuals might attenuate the magnitude of an estimated aggregate policy effect. In a 2WFE analysis, with only one time profile, there can be compelling arguments for either the inclusion or exclusion of these people, but since the models proposed in this chapter can separate the time profiles for the groups that quit from the groups that commenced smoking, leaving them in the sample allows for a more comprehensive analysis of smoking behaviours over time.

The results presented in Section 5.5 should serve as motivation for future investigations into smoking behaviours. Access to the HILDA data may be requested for research purposes (see Watson & Wooden, 2012), and the results presented in this chapter can be replicated independently to act as a starting point. An analysis of the demographic characteristics of the substantive groups in the time profiles identified may help policymakers determine how to target policies and education towards the groups most likely to quit. It may also be useful to closely examine the groups that take up a smoking habit, as this may help to target preventative campaigns.

The skewed and censored distribution of the dependent variable in the smoking application also provides a motivation for future methodological research to adapt the techniques presented in this chapter, to more suitably model such data.

5.6.4 Other considerations

The current BFG-HC and BFG-HC2 implementations require balanced panels, which in general can cause sample selection concerns in applications. Thus, a valuable extension to the class of methods studied here would be to implement the procedures for unbalanced panels.

Furthermore, this chapter does not consider the impact of varying T on the relative performance of the BFG-HC and BFG-HC2 procedures. With small T, say $T < 5$, the time profiles themselves may be less economically interesting; however, considering the performance of both methods across a range of moderate values of T, say, $T = 5$–15 or larger, can be a valuable contribution of future work.

Finally, while using the PEAR method for allocating units into distinct groups provides a workable solution to the general problem of identifying outliers and reducing their impact on substantive groups—namely, some units are isolated into singletons or small groups, and so do not impact other, substantive groups—a more parsimonious approach, left for future research, is to develop a method that uses error distributions with heavier tails than the Normal distribution to account for outlier observations.

5.A Appendix: Implementation details

5.A.1 BFG-HC

5.A.1.1 Joint posterior

Following the arguments in Chapter 4 and Kim and Wang (2019), the joint posterior in the extended parameter space including the auxiliary slice variables, $\eta = \{\eta_1, \eta_2, ..., \eta_N\}$, is

$$p\left(\eta, \phi^*, \delta^*, V, \gamma, \beta \,|\, \overset{\Delta}{y}\right) \propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_i \,|\, \phi^*_{g_i}, \delta^*_{g_i}\right) \mathbb{I}\left\{\eta_i < \pi_{g_i}\right\}\right]$$
$$\times \left[\prod_{j=1}^{\infty} p\left(V_j \,|\, \beta\right) p\left(\phi^*_j\right) p\left(\delta^*_j\right)\right] p\left(\beta\right).$$

The slice variables have independent uniform priors, $\eta_i \sim U\left(0, 1\right)$, for $i = 1, 2, ..., N$; $\overset{\Delta}{y} := \left\{\overset{\Delta}{y}_1, \overset{\Delta}{y}_2, ..., \overset{\Delta}{y}_N\right\}$ are the first-differenced individual-level series of the dependent variable; $\overset{\Delta}{x} := \left\{\overset{\Delta}{x}_1, \overset{\Delta}{x}_2, ..., \overset{\Delta}{x}_N\right\}$ are the differenced covariates, which are suppressed from the notation where possible; $\gamma := \{g_1, g_2, ..., g_N\}$ is the set of group indicators for all units in the sample; $\phi^* := \{\phi^*_1, \phi^*_2, ...\}$.

This extended joint posterior can be equivalently expressed as

$p\left(\eta, \phi^*, \delta^*, V, \gamma, \beta \,|\, \overset{\Delta}{y}\right) \propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_i \,|\, \phi^*_{g_i}, \delta^*_{g_i}\right) p\left(\eta_i \,|\, \pi_{g_i}\right) \pi_{g_i}\right]$

$\times \left[\prod_{j=1}^{\infty} p\left(V_j \,|\, \beta\right) p\left(\phi^*_j\right) p\left(\delta^*_j\right)\right] p\left(\beta\right),$

with $p\left(\eta_i \,|\, \pi_{g_i}\right) = \frac{1}{\pi_{g_i}} \mathbb{I}\left\{0 < \eta_i < \pi_{g_i}\right\}$.

Integrating this extended joint posterior over the set of slice variables with respect to the uniform prior recovers the target posterior of interest, $p\left(\phi^*, \delta^*, V, \gamma, \beta \,|\, \overset{\Delta}{y}\right)$.

5.A.1.2 Gibbs sampling steps

There is significant overlap between the sampling algorithms for BFG-HC and BFG, and details will be omitted where practical to avoid duplication.

Step 1: Draw V and compute π

Identical to BFG.

Step 2: Draw ϕ^*

$p\left(\phi^* \,|\, \overset{\Delta}{y}, \eta, \delta^*, V, \gamma, \beta\right) \propto \left[\prod_{i=1}^{N} p\left(\overset{\Delta}{y}_i \,|\, \phi^*_{g_i}, \delta^*_{g_i}\right)\right] \left[\prod_{j=1}^{\infty} p\left(\phi^*_j\right)\right].$

For $g = 1, 2, ..., C^*$,

$p\left(\phi^*_g \,|\, \overset{\Delta}{y}, \eta, \delta^*, V, \gamma, \beta\right) \propto MVN_{T-1+K}\left(\phi^*_g; A_{\phi,g}, B_{\phi,g}\right),$

with $B_{\phi,g} := \left[S_\phi^{-1} + \left(\sum_{i \in \mathcal{I}_g} \overset{\Delta}{w}_i' \left(\Lambda^*_g\right)^{-1} \overset{\Delta}{w}_i\right)\right]^{-1}$; $\mathcal{I}_g = \{i \in \mathcal{I} = \{1, 2, ..., N\} : g_i = g\}$;
$A_{\phi,g} := B_{\phi,g}\left[S_\phi^{-1} m_\phi + \left(\sum_{i \in \mathcal{I}_g} \overset{\Delta}{w}_i' \left(\Lambda^*_g\right)^{-1} \overset{\Delta}{y}_i\right)\right].$

Step 3: Draw δ^*

This step is similar to the corresponding step in BFG, noting that $\overset{\Delta}{w}_i \phi^*_g$ replaces $\overset{\Delta}{\alpha}^*_g + \overset{\Delta}{x}_i \theta$ in the computations.

A simpler alternative to the Metropolis–Hastings (MH) proposal distribution used in BFG is a tailored random walk which considers the local curvature of the conditional posterior at the current iterate value; this proposal distribution is easier to implement and performs well in the simulation study presented in this chapter. At some iteration $m+1$, the MH proposal density function, q, is defined by

$q\left(\delta_g^{(m)} \to \delta_g^{(m+1)}\right) := MVN_T\left(\delta_g^{(m+1)}; \delta_g^{(m)}, \Sigma\left(\delta_g^{(m)}\right)\right)$ and $q\left(\delta_g^{(m+1)} \to \delta_g^{(m)}\right) := MVN_T\left(\delta_g^{(m)}; \delta_g^{(m+1)}, \Sigma\left(\delta_g^{(m+1)}\right)\right)$,

with $\Sigma(v)$ defined as the inverse of the negative of the Hessian matrix of the log conditional posterior, $\ln p\left(\delta_g^* \mid \overset{\Delta}{y}, \eta, \phi^*, V, \gamma, \beta\right)$, evaluated at $\delta_g^* = v$; $\delta_g^{*(m)}$ is the iterate of parameter δ_g^* drawn in iteration m.

The acceptance probability has the form

$$\text{Prob}\left(\delta_g^{*(m)} \to \delta_g^{*(m+1)}\right) := \min\left\{1, \frac{p\left(\delta_g^{*(m+1)} \mid \overset{\Delta}{y}, \eta, \phi^*, V, \gamma, \beta\right) q\left(\delta_g^{*(m+1)} \to \delta_g^{*(m)}\right)}{p\left(\delta_g^{*(m)} \mid \overset{\Delta}{y}, \eta, \phi^*, V, \gamma, \beta\right) q\left(\delta_g^{*(m)} \to \delta_g^{*(m+1)}\right)}\right\}.$$

Step 4: Label-switching move

This requires only minor modification from the corresponding step in BFG: references to group-specific parameter vectors $\overset{\Delta}{\alpha}_g^*$ need to be replaced with ϕ_g^*.

Step 5: Draw η

Unchanged.

Step 6: Draw β

Unchanged.

Step 7: Recompute η^* and C^*

Minor modification: references to $\overset{\Delta}{\alpha}_g^*$ need to be replaced with ϕ_g^*.

Step 8: Draw γ

Minor modification: $\overset{\Delta}{w}_i \phi_g^*$ replaces $\overset{\Delta}{\alpha}_g^* + \overset{\Delta}{x}_i \theta$ in the computations.

5.A.1.3 Fixed prior parameters

There is now a prior for the group-specific ϕ_g^*, instead of $\overset{\Delta}{\alpha}_g^*$. The first $T-1$ elements of m_ϕ are identical to m_α from BFG; the remaining K elements are the estimates of the covariate effects from the same two-way fixed effects (2WFE) estimation used to generate the fixed prior parameters for the $\overset{\Delta}{\alpha}_g^*$ in BFG. Similarly, the first $T-1$ diagonal elements of S_ϕ are the diagonal elements of S_α, and the remaining K terms on the

diagonal are 30 times the magnitude of the respective elements of m_ϕ. All off-diagonal elements of S_ϕ are set to zero.

Other fixed prior parameters are unchanged.

5.A.1.4 Starting values

Identical to BFG except for the following modifications: step 1 draws the $\phi_g^* \sim MVN_{T-1+K}(m_\phi, S_\phi)$ terms instead of $\overset{\Delta}{\alpha}{}_g^*$; step 2 is unnecessary.

5.A.2 BFG-HC2

In contrast with BFG-HC, the sampler for BFG-HC2 is a more involved extension to the original BFG algorithm.

5.A.2.1 Joint posterior

The extended joint posterior now contains two sets of slice variables, η_ψ, stick-length variables, V_ψ, grouping allocation variables, γ_ψ, and DP concentration parameters, β_ψ, for $\psi \in \{\alpha, \theta\}$. The roles and definitions of these are analogous to the one-dimensional case, noting that $\gamma_\alpha := \{g_1, g_2, ..., g_N\}$, and $\gamma_\theta := \{h_1, h_2, ..., h_N\}$. The extended posterior becomes

$$p\left(\eta_\alpha, \overset{\Delta}{\alpha}{}^*, \delta^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \eta_\theta, \theta^*, V_\theta, \gamma_\theta, \beta_\theta \mid \overset{\Delta}{y}\right)$$

$$\propto \left[\prod_{i=1}^N p\left(\overset{\Delta}{y}_i \mid \theta_{h_i}^*, \overset{\Delta}{\alpha}{}_{g_i}^*, \delta_{g_i}^*\right) \mathbb{I}\left\{\eta_{\alpha,i} < \pi_{\alpha,g_i}\right\} \mathbb{I}\left\{\eta_{\theta,i} < \pi_{\theta,h_i}\right\}\right]$$

$$\times \left[\prod_{j=1}^\infty p\left(V_{\alpha,j} \mid \beta_\alpha\right) p\left(\overset{\Delta}{\alpha}{}_j^*\right) p\left(\delta_j^*\right)\right] p\left(\beta_\alpha\right) \left[\prod_{j=1}^\infty p\left(V_{\theta,j} \mid \beta_\theta\right) p\left(\theta_j^*\right)\right] p\left(\beta_\theta\right),$$

or alternatively,

$$p\left(\eta_\alpha, \overset{\Delta}{\alpha}{}^*, \delta^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \eta_\theta, \theta^*, V_\theta, \gamma_\theta, \beta_\theta \mid \overset{\Delta}{y}\right)$$

$$\propto \left[\prod_{i=1}^N p\left(\overset{\Delta}{y}_i \mid \theta_{h_i}^*, \overset{\Delta}{\alpha}{}_{g_i}^*, \delta_{g_i}^*\right) p\left(\eta_{\alpha,i} \mid \pi_{\alpha,g_i}\right) \pi_{\alpha,g_i} p\left(\eta_{\theta,i} \mid \pi_{\theta,h_i}\right) \pi_{\theta,h_i}\right]$$

$$\times \left[\prod_{j=1}^\infty p\left(V_{\alpha,j} \mid \beta_\alpha\right) p\left(\overset{\Delta}{\alpha}{}_j^*\right) p\left(\delta_j^*\right)\right] p\left(\beta_\alpha\right) \times \left[\prod_{j=1}^\infty p\left(V_{\theta,j} \mid \beta_\theta\right) p\left(\theta_j^*\right)\right] p\left(\beta_\theta\right),$$

with $p\left(\eta_{\alpha,i} \mid \pi_{\alpha,g_i}\right) = \frac{1}{\pi_{\alpha,g_i}} \mathbb{I}\left\{0 < \eta_{\alpha,i} < \pi_{\alpha,g_i}\right\}$, and similarly for $p\left(\eta_{\theta,i} \mid \pi_{\theta,h_i}\right)$.

Integrating out both sets of slice variables with respect to independent uniform priors recovers the joint posterior of primary interest,

$$p\left(\overset{\Delta}{\alpha}^*, \delta^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \theta^*, V_\theta, \gamma_\theta, \beta_\theta \mid \overset{\Delta}{y}\right)$$

$$\propto \left[\prod_{i=1}^N p\left(\overset{\Delta}{y}_i \mid \theta^*_{h_i}, \overset{\Delta}{\alpha}^*_{g_i}, \delta^*_{g_i}\right) p\left(g_i \mid V_\alpha\right) p\left(h_i \mid V_\theta\right)\right]$$

$$\times \left[\prod_{j=1}^\infty p\left(V_{\alpha,j} \mid \beta_\alpha\right) p\left(\overset{\Delta}{\alpha}^*_j\right) p\left(\delta^*_j\right)\right] p\left(\beta_\alpha\right) \left[\prod_{j=1}^\infty p\left(V_{\theta,j} \mid \beta_\theta\right) p\left(\theta^*_j\right)\right] p\left(\beta_\theta\right).$$

5.A.2.2 Gibbs sampling steps

Step 1: Draw V_α and compute π_α

For $g = 1, 2, ..., C^*_\alpha$,

$$p\left(V_{\alpha,g} \mid \overset{\Delta}{y}, \overset{\Delta}{\alpha}^*, \delta^*, \gamma_\alpha, \beta_\alpha, \eta_\theta, \theta^*, V_\theta, \gamma_\theta, \beta_\theta\right) \propto \text{Beta}\left(V_{\alpha,g}; 1 + N_{\alpha,g}, \beta_\alpha + N_{\alpha,>g}\right);$$

$N_{\alpha,g} := \sum_{i=1}^N \mathbb{I}\left\{g_i = g\right\}$; $N_{\alpha,>g} := \sum_{i=1}^N \mathbb{I}\left\{g_i > g\right\}$.

Update π_α using: $\pi_{\alpha,g} = V_{\alpha,g} \prod_{l=1}^{g-1}\left(1 - V_{\alpha,l}\right)$ for $g = 2, ..., C^*_\alpha$; $\pi_{\alpha,1} = V_{\alpha,1}$.

Step 2: Draw $\overset{\Delta}{\alpha}^*$

For $g = 1, 2, ..., C^*_\alpha$,

$$p\left(\overset{\Delta}{\alpha}^*_g \mid \overset{\Delta}{y}, \eta_\alpha, \delta^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \eta_\theta, \theta^*, V_\theta, \gamma_\theta, \beta_\theta\right) \propto MVN_{T-1}\left(\overset{\Delta}{\alpha}^*_g; A_{\alpha,g}, B_{\alpha,g}\right);$$

$B_{\alpha,g} := \left[S_\alpha^{-1} + N_{\alpha,g}\left(\Lambda^*_g\right)^{-1}\right]^{-1}$; $A_{\alpha,g} := B_{\alpha,g}\left[S_\alpha^{-1} m_\alpha + \left(\Lambda^*_g\right)^{-1}\left(\sum_{i \in \mathcal{I}_{\alpha,g}} \overset{\Delta}{y}_i - \overset{\Delta}{x}_i \theta^*_{h_i}\right)\right]$;
$\mathcal{I}_{\alpha,g} = \left\{i \in \mathcal{I} : g_i = g\right\}$.

Step 3: Draw δ^*

For $g = 1, 2, ..., C^*_\alpha$,

$$p\left(\delta^*_g \mid \overset{\Delta}{y}, \eta_\alpha, \overset{\Delta}{\alpha}^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \eta_\theta, \theta^*, V_\theta, \gamma_\theta, \beta_\theta\right)$$

$$\propto \left[\det\left(\Lambda^*_g\right)\right]^{-N_{\alpha,g}/2} \exp\left\{-\tfrac{1}{2}\sum_{i \in \mathcal{I}_{\alpha,g}} \tilde{y}'_i \left(\Lambda^*_g\right)^{-1} \tilde{y}_i\right\}$$

$$\times \left[\det\left(S_\delta\right)\right]^{-1/2} \exp\left\{-\tfrac{1}{2}\left(\delta^*_g - m_\delta\right)' S_\delta^{-1}\left(\delta^*_g - m_\delta\right)\right\};$$

$\tilde{y}_i := \overset{\Delta}{y}_i - \overset{\Delta^*}{\alpha}_g - \overset{\Delta}{x}_i \theta^*_{h_i}$.

Use a MH step to sample from this conditional posterior; the MH proposal described in Appendix 5.A.1.2 for the corresponding Gibbs sampling step also performs well for BFG-HC2 in the simulation study.

Step 4: g-group label-switching move

'g-groups' refer to the grouping structure stored in $\gamma_\alpha = \{g_1, g_2, ..., g_N\}$. This step is identical to the label-switching step defined for BFG after making the following substitutions in the formulas presented: $N_j \to N_{\alpha,j}$ and similarly for N_l and N_{j+1}, where these group size variables are defined above in step 1; $\pi_j \to \pi_{\alpha,j}$ and similarly for π_l and π_{j+1}; $V_j \to V_{\alpha,j}$ and similarly for V_{j+1}; $\pi'_j \to \pi_{\alpha,j}$ and similarly for π'_{j+1}; $\pi^+ \to \pi^+_\alpha$ and similarly for Π'; $R_1 \to R_{\alpha,1}$ and similarly for R_2; $\beta \to \beta_\alpha$.

Step 5: Draw V_θ and compute π_θ

For $h = 1, 2, ..., C^*_\theta$,

$$p\left(V_{\theta,h} \,|\, \overset{\Delta}{y}, \overset{\Delta^*}{\alpha}, \delta^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \eta_\theta, \theta^*, \gamma_\theta, \beta_\theta\right) \propto \text{Beta}\left(V_{\theta,h}; 1 + N_{\theta,h}, \beta_\theta + N_{\theta,>h}\right);$$

$N_{\theta,h} := \sum_{i=1}^N \mathbb{I}\{h_i = h\}$; $N_{\theta,>h} := \sum_{i=1}^N \mathbb{I}\{h_i > h\}$.

Update π_θ analogously to π_α in step 1.

Step 6: Draw θ^*

For $h = 1, 2, ..., C^*_\theta$,

$$p\left(\theta^*_h \,|\, \overset{\Delta}{y}, \eta_\alpha, \overset{\Delta^*}{\alpha}, \delta^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \eta_\theta, V_\theta, \gamma_\theta, \beta_\theta\right) \propto MVN_K\left(\theta^*_h; A_{\theta,h}, B_{\theta,h}\right);$$

$$B_{\theta,h} := \left[S_\theta^{-1} + \sum_{i \in \mathfrak{I}_{\theta,h}} \overset{\Delta'}{x}_i \left(\Lambda^*_{g_i}\right)^{-1} \overset{\Delta}{x}_i\right]^{-1};$$

$$A_{\theta,h} := B_{\theta,h}\left[S_\theta^{-1} m_\theta + \sum_{i \in \mathfrak{I}_{\theta,h}} \overset{\Delta'}{x}_i \left(\Lambda^*_{g_i}\right)^{-1} \left(\overset{\Delta}{y}_i - \overset{\Delta^*}{\alpha}_{g_i}\right)\right]; \; \mathfrak{I}_{\theta,h} := [i \in \mathfrak{I} : h_i = h].$$

Step 7: h-group label-switching move

Analogous to step 4 but for the 'h-groups' stored in γ_θ: replace the 'α' subscripts in step 4 with 'θ' subscripts.

Step 8: Draw η_α

For $i = 1, 2, ..., N$,

$$p\left(\eta_{\alpha,i} \mid \overset{\Delta}{y}, \overset{\Delta^*}{\alpha}, \delta^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \eta_\theta, \theta^*, V_\theta, \gamma_\theta, \beta_\theta\right) \propto \text{Uniform}\left(\eta_{\alpha,i}; 0, \pi_{\alpha,g_i}\right).$$

Step 9: Draw η_θ

Same as step 8 but replace: α subscripts $\rightarrow \theta$ subscripts; g_i subscript $\rightarrow h_i$ subscript.

Step 10: Draw β_α

Analogous to step 6 of the BFG algorithm but replace: $\beta \rightarrow \beta_\alpha$ and $\kappa \rightarrow \kappa_\alpha$.

Step 11: Draw β_θ

Same as step 10 but replace: α subscripts $\rightarrow \theta$ subscripts; $G \rightarrow H$.

Step 12: Recompute η_α^* and C_α^*

Same as BFG step 7 with replacements: $\eta^* \rightarrow \eta_\alpha^*$; $\eta_i \rightarrow \eta_{\alpha,i}$ for all i; $\pi_j \rightarrow \pi_{\alpha,j}$; $C^* \rightarrow C_\alpha^*$; $V_g \rightarrow V_{\alpha,g}$; $\beta \rightarrow \beta_\alpha$.

Step 13: Recompute η_θ^* and C_θ^*

Same as step 12 but replacing α subscripts $\rightarrow \theta$ subscripts, and the new group-specific parameters to draw from the prior, if required, are the θ_h^* instead of the $\overset{\Delta^*}{\alpha}_g$ and δ_g^*.

Step 14: Draw γ_α

For $i = 1, 2, ..., N$, draw g_i according to

$$\Pr\left(g_i = g \mid \overset{\Delta}{y}, \eta_\alpha, \overset{\Delta^*}{\alpha}, \delta^*, V_\alpha, \beta_\alpha, \eta_\theta, \theta^*, V_\theta, \gamma_\theta, \beta_\theta\right)$$

$$= \begin{cases} \dfrac{p\left(\overset{\Delta}{y}_i \mid \overset{\Delta^*}{\alpha}_g, \delta_g^*, \theta_{h_i}^*\right)^{\mathbb{I}\{\eta_{\alpha,i} < \pi_{\alpha,g}\}}}{\sum_{j=1}^{C_\alpha^*} p\left(\overset{\Delta}{y}_i \mid \overset{\Delta^*}{\alpha}_j, \delta_j^*, \theta_{h_i}^*\right)^{\mathbb{I}\{\eta_{\alpha,i} < \pi_{\alpha,j}\}}}, & \text{for } g = 1, 2, ..., C_\alpha^*; \\ 0, & \text{otherwise.} \end{cases}$$

Step 15: Draw γ_θ

For $i = 1, 2, ..., N$, draw h_i according to

$$\Pr\left(h_i = h \mid \overset{\Delta}{y}, \eta_\alpha, \overset{\Delta^*}{\alpha}, \delta^*, V_\alpha, \gamma_\alpha, \beta_\alpha, \eta_\theta, \theta^*, V_\theta, \beta_\theta\right)$$

$$= \begin{cases} \frac{p\left(\overset{\Delta}{y}_i \,|\, \theta_h^*, \overset{\Delta^*}{\alpha}_{g_i}, \delta_{g_i}^*\right) \mathbb{I}\left\{\eta_{\theta,i} < \pi_{\theta,h}\right\}}{\sum_{j=1}^{C_\theta^*} p\left(\overset{\Delta}{y}_i \,|\, \theta_j^*, \overset{\Delta^*}{\alpha}_{g_i}, \delta_{g_i}^*,\right) \mathbb{I}\left\{\eta_{\theta,i} < \pi_{\theta,j}\right\}}, & \text{for } h = 1, 2, ..., C_\theta^*; \\ 0, & \text{otherwise.} \end{cases}$$

5.A.2.3 Fixed prior parameters

$\overset{\Delta^*}{\alpha}_g$ **and** δ_g^* **priors.** Same as BFG.

θ_h^* **prior.** m_θ are the estimates of the covariate effects from the same 2WFE estimation used to generate the fixed prior parameters for the $\overset{\Delta^*}{\alpha}_g$. S_θ is a diagonal matrix with nonzero entries equal to 30 times the magnitude of the respective elements of m_θ.

β_α **and** β_θ **priors.** The prior parameters for both β_α and β_θ are the same as for the β prior in BFG.

5.A.2.4 Starting values

Some modifications to the procedure used for BFG are required to support the second clustering dimension. The following procedure can be used to initialise starting values for BFG-HC2.

1. Same as step 1 for BFG.
2. Similarly to step 1, generate H_{start} sets of θ_h^* vectors from $\theta_h^* \sim MVN_K\left(m_\theta, S_\theta\right)$; H_{start} is selected by the researcher; refer to these as $\left\{\theta^{*(j)}\right\}_{j=1}^{H_{start}}$.
3. Same as step 3 for BFG.
4. Similarly to step 4 for BFG, assign each unit to one of the $\overset{\Delta^{*(j)}}{\alpha}$ by drawing the g_i from their conditional posteriors, but ignoring the terms $\mathbb{I}\left\{\eta_{\alpha,i} < \pi_{\alpha,g}\right\}$ and $\mathbb{I}\left\{\eta_{\alpha,i} < \pi_{\alpha,j}\right\}$. During the computation of the conditional posteriors for each g_i, it is necessary to evaluate the term $p\left(\overset{\Delta}{y}_i \,|\, \overset{\Delta^*}{\alpha}_g, \delta_g^*, \theta_{h_i}^*\right)$ for $g = 1, 2, ..., G_{start}$; however, this term depends on the as-yet uninitialised values of h_i. The solution

used here is to randomly sample values with replacement from $\{1, 2, ..., H_{start}\}$ for each element of γ_θ before commencing this step; this approach works well in the simulation study.

5. Same as step 5 for BFG.
6. Set C^*_α as the number of surviving $\overset{\Delta^{*(j)}}{\alpha}$ vectors.
7. Same as step 7 for BFG, but replace $C^* \to C^*_\alpha$.
8. Same as step 8 for BFG, but replace $\gamma \to \gamma_\alpha$.
9. Perform the equivalent of step 4 but for the $\theta^{*(j)}$; the g_i required have been initialised in step 8.
10. As step 5, but for the $\theta^{*(j)}$.
11. Set C^*_θ as the number of surviving $\theta^{*(j)}$ vectors.
12. As step 7, but for the $\theta^{*(j)}$; replace $C^*_\alpha \to C^*_\theta$ and $g \to h$.
13. As step 8, but for γ_θ.
14. Set $\beta_\alpha = \beta_\theta = 1$.

Chapter 6

Discussion and Conclusions

6.1 Reconciling superannuation application results

Broadly, the results from all models run on the superannuation data in this thesis show that capturing latent group behaviours in the recorded drawdowns from account-based pensions explains much of the variation observed in the data.

In the application results presented in Chapter 3, 66% of units in the $N = 9516$ sample are deemed to have used amount-based drawdown heuristics to decumulate their balances over time. This percentage figure is similar to the 68% of the $N = 1000$ sample that is allocated to Partition 1 in Chapter 5, which is the partition corresponding to individuals whose estimated covariate effects imply they consider their drawdown patterns in terms of dollar amounts, rather than proportions of their account balance.

The two distinct covariate effect groups found separately among the five independent BFG runs in Chapter 4 are both found using the methods in Chapter 5 too, this time in all independent estimation runs. Partition 3 in Chapter 5 shows that the BFG-HC2 procedure can also find groups where the covariate effects do not collapse to the Partition 1 or Partition 2 configurations; i.e., the method captures a set of individuals that are sensitive to changes in their account balances and the minimum drawdown rates. Estimating time profiles for individuals like those in Partition 3 is the intention implied by the base model specification for the superannuation applications in Chapters

3, 4 and 5, rather than estimating time profiles for reduced-form models that do not simultaneously estimate individual responses to the covariates.

In addition to replicating all the behaviours found in Chapters 3 and 4, the BFG-HC2 method also identifies two behaviours that are not found in the results from those chapters: a) a group of retirees who prefer drawing at the minimum rates, but who did not adopt the temporary concessional minima that came into effect from 2009–2013; b) groups that did adjust drawdowns coincident with this same time period, but who otherwise do not show evidence of following the minimum rates as their usual strategy.

Taken together, the comparison of BFG-HC2 results against the other results in this thesis suggests that the BFG-HC2 method is the most appropriate choice considered for modelling the heterogeneity in the drawdowns data.

6.2 Modelling time-varying heterogeneity in observational microeconomic data

Chapters 3, 4 and 5 show the value of using models that control for time-varying heterogeneity in microeconomic applications, not only when latent group behavioural effects are unwanted sources of endogeneity, but also in cases where the time profiles are the primary estimands of interest. These chapters demonstrate how researchers may use such models for observational panel data in these scenarios.

In applications where the need to capture time-varying latent group effects is unclear, the methods considered in this thesis may be used as diagnostic tools. The GFE estimator is able to both test for the presence of time-varying unobservable heterogeneity, and show the impact of failing to account for it, by observing how covariate effect estimates evolve as the number of assumed groups, G, increases from an initial value of $G = 1$, with the one-group model equation corresponding to the standard 2WFE model. However, the question of which G to use as input for the GFE estimator, when the true number of latent groups is unknown, only has a clear answer when N and T tend to infinity at the same rate, as described in the online supplement to Bonhomme and Manresa (2015). In other cases, including typical microeconomic panels like those

used in this thesis, as well as the application in Bonhomme and Manresa (2015), a heuristic approach is required to estimate the number of groups in the sample. Thus, there is a motivation to explore alternative solutions to the group estimation problem in such applications. The Bayesian methods in Chapters 4 and 5 are useful because, in principle, they only require running the estimation procedure once, with the number of groups treated as random, rather than fixed as with the GFE estimator. A more direct Bayesian analog to the GFE estimator is to run multiple Bayesian finite mixture models (see, e.g., Gelman et al., 2013, Ch. 22) with fixed values for G, then perform some model selection procedure to choose the preferred value of G; e.g., using WAIC as described in Gelman et al. (2013), p. 173. The DPMM approach taken in this thesis can be considered a more convenient alternative to selection among multiple finite mixture models, as it treats G as random, which avoids the problem of selecting a single value of G, while still permitting estimation of G after running the method—if this is of interest to the researcher. If the BFG class of methods finds multiple substantive groups, this can be treated as evidence that a one-group analysis of the data is inappropriate. Further, if the results suggest an unexpectedly large number of nontrivial groups, this may be evidence that a grouped-data modelling assumption is inappropriate. If, in a particular application, a rigid upper limit on the possible number of groups is desired, one can consider adapting the methods in this thesis by following the principles in Miller and Harrison (2018), which shows that DPMM Gibbs sampling can in many cases be suitably adapted for the Bayesian estimation of finite mixture models.

When using the GFE estimator, if bootstrap standard errors are desired for the time profile estimators, a solution to a label-matching problem, such as the one proposed in Chapter 3, is required to align group labels across independent model runs. A label-matching algorithm also enables using simulation to test the accuracy of standard errors derived from the fixed-T variance estimator for a panel dataset with specific values of N and T, as in Appendix A. Also, in applications where the model includes individual fixed effects and the data is unbalanced, the proposed modification to the estimation procedure maintains the desirable equivalence between the 2WFE estimation results and the GFE estimation results when $G = 1$.

Compared to the `Stata` implementation of the GFE estimator provided by Bonhomme and Manresa (2015), the `R` implementation developed for Chapter 3 is able to run the GFE estimator for an order of magnitude larger N, permitting the analysis of the $N = 9516$ and $T = 12$ superannuation dataset. While the theory derived in Bonhomme and Manresa (2015) describes the asymptotic performance of the GFE estimator when T grows substantially more slowly than N, the theory does not provide precise guarantees for how well the procedure works with any given values of N and T. Hence, Appendix A uses a simulation study to see if the dimensions of the available superannuation dataset pose a barrier to the analysis. The simulation study finds that the procedure works well on the simulated datasets, which use $G = 7$, $N = 9516$ and $T = 12$. Using a simulation approach, Chapters 4 and 5 verify that the corresponding BFG class of methods proposed have good performance in panels of size $N = 1000$ and $T = 10$; moreover, Chapter 4 compares the GFE and BFG estimators on the same simulated datasets, finding that the BFG procedure noticeably outperforms the GFE approach in terms of group allocation accuracy and time profile inference.

The set of possible solutions to the optimisation problem that produces the GFE estimates grows quickly with N and G, and the solutions are in general sensitive to the starting values used to initialise the algorithm (Bonhomme & Manresa, 2015). For the superannuation dataset, a robustness check provided in Appendix A shows that the results appear relatively insensitive to changing the number of starting values; hence, that chapter does not further address this general problem. However, Bonhomme and Manresa (2015), and authors they cite, develop exact and heuristic methods for finding solutions to the difficulties in optimising functions that determine groupings of individual units. Developing better solutions to this problem in the context of GFE estimation will increase the estimator's suitability when applied to large-N datasets, which may not always prove as robust to increasing the number of starting values as the superannuation dataset. The BFG class of methods explored in Chapters 4 and 5 suffer from a similar problem: using the `R` code written to implement the Gibbs samplers, estimating models on datasets with N much larger than 1000, or T much larger than 10, proves impractical due to the large amount of time required to run many iterations of the sampler with those values of N and T, unless the number of groups is small. More efficient code than the current implementation may partly ameliorate this issue.

6.3 Future methodological considerations

An avenue for methodological contributions that would broaden the applicability of the methods considered in this thesis is the extension to nonlinear panel data models. A nonlinear model allows the researcher to more accurately capture non-continuous dependent variables, e.g., binary or count data. Nonlinear models are discussed by Bonhomme and Manresa (2015) as a potential extension to their GFE estimator, but there is currently no algorithm for a GFE estimator that can jointly estimate all the parameters of such a model, nor is there a statistical treatment that would allow for inference on parameter estimates. For nonparametric Bayesian estimation, nonlinear models may leverage work by Lancaster (2000, 2002) and Arellano and Bonhomme (2009), who make progress in this area for panel data models without latent groups. The main obstacle to overcome is an incidental parameters problem that arises when individual-specific fixed effects cannot be algebraically removed from the model, as was possible for the linear models considered in this thesis.

A limitation of the implementations of the BFG class of methods presented in Chapters 4 and 5 is the current requirement that the panel be balanced. In practice, observational panel datasets may suffer from gradual attrition in the number of individuals observed, or the addition of a top-up sample, meaning that these methods as currently implemented must necessarily be estimating parameters on a select sample: those individuals that are observed in all time periods. Extending the implementations to handle unbalanced panels is therefore a useful path for future work on these methods.

The development of the BFG-HC2 method in Chapter 5, and its success in the simulation study, paves the way for similar multidimensional clustering models for microeconomic applications. The following ideas are all relatively direct extensions or adaptations of the BFG-HC2 specification, which have similar economic motivations but may be more flexible than existing methods.

- A variant of the BFG-HC2 procedure where the covariate effects and the time profiles are on the same grouping dimension, with error variances on a second dimension.

- ‘3D-BFG-HC’: like the BFG-HC2 procedure but including a third clustering dimension for the error variances.

- ‘BGFE-HC’: the method from Kim and Wang (2019), but where covariate effects vary by group.

- ‘2D-BGFE-HC’: like the BGFE-HC procedure but covariate effects are on a different grouping dimension than the time profiles.

- ‘3D-BGFE-HC’: like the BFG-HC2 procedure but where the model is not first-differenced; instead, the individual-specific intercepts are placed on their own clustering dimension, to more parsimoniously model the heterogeneity in levels.

- ‘4D-BGFE-HC’: like the 3D-BGFE-HC procedure, but error variances are on a separate clustering dimension.

6.4 Conclusions and main contributions

This thesis considers the development of flexible and general approaches to the econometric analysis of observational panel datasets, motivated by novel microeconomic applications and the limitations of existing approaches. The models studied allow for both time-constant and time-varying unobservable heterogeneity. With these methods, the time-varying heterogeneity is modelled through the use of latent variables that define the membership of units into distinct groups, permitting a more parsimonious specification of heterogeneity.

Classical and Bayesian techniques are considered, with several advantages of the Bayesian approaches demonstrated, both in simulation studies and in empirical applications. Relative to the closest classical alternative, the key advantages of the Bayesian approaches shown in this thesis are: a) more accurate latent group membership identification; b) more accurate time profile inference; c) allowing the number of unknown groups to be a random variable in the estimation, rather than a fixed value.

This thesis contributes to the econometrics literature on linear panel data modelling in two ways. First, it demonstrates how an existing classical method, the Grouped Fixed-Effects (GFE) estimator from Bonhomme and Manresa (2015), can be repurposed to answer questions typical of applied microeconomics research. Extensions to the estimator are proposed to facilitate its novel use in performing inference on the time-varying heterogeneity component of models, which in many applications may have useful economic interpretations.

Second, this thesis develops nonparametric Bayesian methods for analysing linear panel data models with time-varying unobservable heterogeneity and individual-specific intercepts. The first new method, the BFG estimator, can be viewed as one possible Bayesian analog to the specification considered in Bonhomme and Manresa (2015) that includes individual-specific intercepts, in which the number of unknown groups remains unspecified *a priori*. The BFG procedure can also be considered a direct extension of the method in Kim and Wang (2019), tailored for use in microeconomic applications by also controlling for individual fixed effects. Another two Bayesian methods are developed for more general panel data models that also allow for heterogeneity in the partial effects of observed covariates: the BFG-HC and BFG-HC2 procedures. The BFG-HC2 approach allows for the most flexible specification, where groups in the covariate effects can differ from groups in the time profiles. Simulation studies validate the proposed Bayesian methods and show that: a) the BFG procedure can outperform the GFE approach on the same data; b) the BFG-HC2 procedure is preferred to the BFG-HC approach if groups in the covariate effects may be different to groups in the time-varying heterogeneity.

There are various economic scenarios that motivate the class of models studied in this thesis, including the search for latent behavioural patterns that are partly obstructed by the effects of observed covariates, and policy analysis using event studies. As illustrations of their broader applicability, the methods provide applied contributions through novel insights in two distinct applications: understanding retirement decumulation behaviours using Australian superannuation data, and studying policy impacts and behavioural groups in smoking patterns using Australian longitudinal survey data. The superannuation application shows only weak evidence

that retirees consider their withdrawal patterns in terms of heuristic strategies based on the rate of drawdown, with such rate-based heuristics being the object of study in theoretical papers by Horneff et al. (2008) and Bateman and Thorp (2008). The main rate-based strategy identified is to closely follow the government-legislated minimum drawdown rates. Instead, retirees seem considerably more likely to frame drawdowns in terms of dollar amounts, and to target drawdowns of constant dollar amounts over time. Diminishing balances then either necessitate a downward revision of the income stream, or result in rapid account depletion. These findings broadly align with empirical results in Balnozan (2018), which uses a less formal approach that combines traditional panel data models and cluster analysis on similar data.

The smoking application complements work by Bonfrer et al. (2020), which estimates the average policy effect of the plain packaging mandate introduced in Australia in 2012. The preliminary empirical analysis presented in this thesis suggests that there is heterogeneity in the smoking behaviours observed, and that individuals were not homogeneously impacted by smoking policies around this time period; however, this application is mainly intended as an illustration of how the methods in this thesis can be applied more broadly. Further research is required to draw robust inferential conclusions in this application.

References

Ando, T., & Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, *31*(1), 163–191.

Angrist, J., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Arellano, M., & Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, *77*(2), 489–536.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, *77*, 1229–1279. doi: 10.3982/ECTA6135

Balnozan, I. (2018). *Slow and steady: Drawdown behaviours in phased withdrawal retirement income products.* Unpublished manuscript, UNSW Business School, University of New South Wales, UNSW Sydney, Australia. Retrieved from https://sites.google.com/view/igorbalnozan/research

Baltagi, B. H., Bresson, G., & Pirotte, A. (2008). To pool or not to pool? In L. Mátyás & P. Sevestre (Eds.), *The econometrics of panel data: Fundamentals and recent developments in theory and practice* (Vol. 46, pp. 517–546). Springer Berlin Heidelberg.

Bateman, H., & Thorp, S. (2008). Choices and constraints over retirement income streams: Comparing rules and regulations. *Economic Record*, *84*, S17–S31. doi: 10.1111/j.1475-4932.2008.00480.x

Blackwell, D., & MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, *1*, 353–355.

Bonfrer, A., Chintagunta, P. K., Roberts, J. H., & Corkindale, D. (2020). Assessing the sales impact of plain packaging regulation for cigarettes: Evidence from Australia. *Marketing Science*, *39*(1), 234–252.

Bonhomme, S., Lamadon, T., & Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, *87*(3), 699–739.

Bonhomme, S., Lamadon, T., & Manresa, E. (in press). Discretizing unobserved heterogeneity. *Econometrica*.

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, *83*, 1147–1184. doi: 10.3982/ECTA11319

Chin, V., Lee, J. Y. L., Ryan, L. M., Kohn, R., & Sisson, S. A. (2019). Multiclass classification of growth curves using random change points and heterogeneous random effects. *arXiv preprint arXiv:1909.07550*.

Deb, P., & Trivedi, P. K. (2013). Finite mixture for panels with fixed effects. *Journal of Econometric Methods*, *2*, 35–51. doi: 10.1515/jem-2012-0018

Escobar, M. D., & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, *90*, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 209–230.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of Normal distributions. In *Recent Advances in Statistics* (pp. 287–302). Elsevier.

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, *21*, 768–769.

Fritsch, A., & Ickstadt, K. (2009). Improved criteria for clustering based on the posterior similarity matrix. *Bayesian Analysis*, *4*, 367–391.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC press.

Hall, P., Müller, H.-G., & Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 1493–1517.

Hastie, D. I., Liverani, S., & Richardson, S. (2015). Sampling from Dirichlet process mixture models with unknown concentration parameter: Mixing issues in large data implementations. *Statistics and Computing*, *25*, 1023–1037.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97.

Hofmans, J., Ceulemans, E., Steinley, D., & Van Mechelen, I. (2015). On the added value of bootstrap analysis for k-means clustering. *Journal of Classification*, *32*, 268–284. doi: 10.1007/s00357-015-9178-y

Horneff, W. J., Maurer, R. H., Mitchell, O. S., & Dus, I. (2008). Following the rules: Integrating asset allocation and annuitization in retirement portfolios. *Insurance: Mathematics and Economics*, *42*, 396–408. doi: 10.1016/j.insmatheco.2007.04.004

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Ishwaran, H., & Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, *87*(2), 371–390.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, *3*, 79–87. doi: 10.1162/neco.1991.3.1.79

Jordan, M. I., & Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, *6*, 181–214. doi: 10.1162/neco.1994.6.2.181

Kalli, M., Griffin, J. E., & Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, *21*, 93–105.

Kim, J., & Wang, L. (2019). Hidden group patterns in democracy developments: Bayesian inference for grouped heterogeneity. *Journal of Applied Econometrics*, *34*, 1016–1028. doi: 10.1002/jae.2734

Koop, G. (2003). *Bayesian econometrics*. Wiley.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, *95*(2), 391–413.

Lancaster, T. (2002). Orthogonal parameters and panel data. *The Review of Economic Studies*, *69*, 647–666.

Miller, J. W., & Harrison, M. T. (2018). Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, *113*(521), 340–356.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, *5*, 32–38. doi: 10.1137/0105003

Nagin, D. S., Jones, B. L., Passos, V. L., & Tremblay, R. E. (2018). Group-based multi-trajectory modeling. *Statistical Methods in Medical Research*, *27*(7), 2015–2023.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods*, *6*(1), 18.

Ng, S.-K., & McLachlan, G. J. (2014). Mixture models for clustering multilevel growth trajectories. *Computational Statistics & Data Analysis*, *71*, 43–51.

Papaspiliopoulos, O., & Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, *95*, 169–186.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850.

Rendon, S. R. (2013). Fixed and random effects in classical and Bayesian regression. *Oxford Bulletin of Economics and Statistics*, *75*, 460–476.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, *4*, 639–650.

Su, L., & Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics*, *206*, 554–573. doi: 10.1016/j.jeconom.2018.06.014

Suits, D. B. (1984). Dummy variables: Mechanics v. interpretation. *The Review of Economics and Statistics*, *66*, 177–180. doi: 10.2307/1924713

The Australian Government the Treasury. (2016). *Development of the framework for comprehensive income products for retirement* (Discussion Paper). Retrieved from https://consult.treasury.gov.au/retirement-income-policy-division/comprehensive-income-products-for-retirement/supporting_documents/CIPRs_Discussion_Paper_1702.pdf

Van Dyk, D. A., & Park, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, *103*, 790–796.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation*, *36*, 45–54.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(12), 3571–3594.

Watson, N., & Wooden, M. P. (2012). The HILDA survey: a case study in the design and development of a successful household panel survey. *Longitudinal and Life Course Studies*, *3*(3), 369–381.

Wooldridge, J. (2012). *Introductory econometrics: A modern approach.* Cengage Learning.

Yang, S. S., & Huang, H.-C. (2009). The impact of longevity risk on the optimal contribution rate and asset allocation for defined contribution pension plans. *The Geneva Papers on Risk and Insurance - Issues and Practice*, *34*, 660–681. doi: 10.1057/gpp.2009.18

Zhang, B. (2020). Forecasting with Bayesian grouped random effects in panel data. *arXiv preprint arXiv:2007.02435*.

Appendix A

Supplement for *Hidden Group Time Profiles: Heterogeneous Drawdown Behaviours in Retirement*

Abstract

This material supplements Chapter 3 with robustness checks to examine the sensitivity of the main results to changes in the dataset and estimation strategy; a simulation study to compare simulated standard errors with those estimated from the analytical formula and the bootstrap; additional details on the superannuation drawdowns dataset; further details on the proposed modification to the estimation algorithm for unbalanced data; additional descriptive results for the seven-group model; and panel plots for the two-group model.

A.1 More details on 'Extension 2: Alternative estimation method for unbalanced data'

We now justify the claims made in Section 3.4.5 about the relationship between the three methods for estimating time profiles with balanced and unbalanced data.

Lemma 1. Methods 1–3 estimate the same time profiles in the balanced data case.

Proof. The regression equation estimated by method 3 (no constant term; $T-1$ time-demeaned time dummy variables, $\dot{\delta}_s$) is

$$\dot{y}_{it} = \dot{x}'_{it}\theta + \sum_{s=2}^{T} \tilde{\dot{\alpha}}_s \dot{\delta}_s + \dot{v}_{it}, \tag{A.1}$$

where $\dot{\delta}_s := \delta_s - T^{-1}\sum_{l=1}^{T}\delta_l$; $\delta_l = \mathbb{I}\{t=l\}$. $\dot{\delta}_s = \delta_s - T^{-1}$, as $\sum_{l=1}^{T}\delta_l = 1$ for all t. Thus, (A.1) can be rewritten using time dummy variables, δ_s, instead of time-demeaned time dummy variables, $\dot{\delta}_s$, as

$$\dot{y}_{it} = \dot{x}'_{it}\theta + \dot{\alpha}_1 + \sum_{s=2}^{T} \tilde{\dot{\alpha}}_s \delta_s + \dot{v}_{it}, \tag{A.2}$$

where $\dot{\alpha}_1 := -T^{-1}\sum_{s=2}^{T}\tilde{\dot{\alpha}}_s$.

(A.2) is precisely the regression equation being estimated by method 2 (including constant term; $T-1$ time dummy variables, δ_s). Defining $\dot{\alpha}_s := \dot{\alpha}_1 + \tilde{\dot{\alpha}}_s$ for $s = 1, 2, ..., T$, (A.2) can be rewritten as

$$\dot{y}_{it} = \dot{x}'_{it}\theta + \sum_{s=1}^{T} \dot{\alpha}_s \delta_s + \dot{v}_{it}, \tag{A.3}$$

because $\dot{\alpha}_1 + \sum_{s=2}^{T}\tilde{\dot{\alpha}}_s\delta_s = \dot{\alpha}_1\delta_1 + \sum_{s=2}^{T}\dot{\alpha}_s\delta_s$.

(A.3) is the regression equation estimated by method 1 (no constant term; T time dummy variables, δ_s).

Combining $\dot{\alpha}_1 = -T^{-1}\sum_{s=2}^{T}\tilde{\dot{\alpha}}_s$ with $\dot{\alpha}_s = \dot{\alpha}_1 + \tilde{\dot{\alpha}}_s$ implies $\sum_{t=1}^{T}\dot{\alpha}_t = \dot{\alpha}_1 + \sum_{t=2}^{T}\dot{\alpha}_t$ $= \dot{\alpha}_1 + \sum_{t=2}^{T}\left(\dot{\alpha}_1 + \tilde{\dot{\alpha}}_t\right) = \dot{\alpha}_1 + (T-1)\,\dot{\alpha}_1 - T\dot{\alpha}_1 = 0.$

The conversion between $\dot{\alpha}_s$ and $\tilde{\dot{\alpha}}_s$ terms is similar to the transformation used in Suits (1984, p. 178), but applied to a model with no constant term.

Lemma 2. With unbalanced data, method 1 does not guarantee that $\sum_{t=1}^{T} \widehat{\dot{\alpha}}_t = 0$.

Proof. Consider the special case with no covariates and $T = 2$ as an illustration. The corresponding regression equation is

$$\dot{y}_{it} = \dot{\alpha}_1 \delta_1 + \dot{\alpha}_2 \delta_2 + \dot{v}_{it},$$

or equivalently, $\dot{y}_{i1} = \dot{\alpha}_1 \delta_1 + \dot{v}_{i1}$ and $\dot{y}_{i2} = \dot{\alpha}_2 \delta_2 + \dot{v}_{i2}$.

The least-squares estimators of $\dot{\alpha}_1$ and $\dot{\alpha}_2$ are

$$\widehat{\dot{\alpha}}_s = N_s^{-1} \sum_{i \in \mathcal{I}_s} (y_{i,s} - \bar{y}_i),$$

for $s = 1, 2$, where $\mathcal{I}_s$ is the set of units observed in period s; N_s is the number of elements in $\mathcal{I}_s$; $\bar{y}_i := T_i^{-1} \sum_{t \in \tau_i} y_{it}$; τ_i is the set of periods in which unit i is observed; T_i is the number of elements in τ_i. Then, $\widehat{\dot{\alpha}}_1 + \widehat{\dot{\alpha}}_2 =: A - B$, where

$$A := N_1^{-1} \sum_{i \in \mathcal{I}_1} y_{i,1} + N_2^{-1} \sum_{i \in \mathcal{I}_2} y_{i,2};$$

$$B := N_1^{-1} \sum_{i \in \mathcal{I}_1} \bar{y}_i + N_2^{-1} \sum_{i \in \mathcal{I}_2} \bar{y}_i.$$

When the panel is balanced, $N_1 = N_2 = N$, $\mathcal{I}_1 = \mathcal{I}_2 = \{1, 2, ..., N\}$, and $T_i = T = 2$ for all i. Then, $A = B = T\bar{\bar{y}}$, where

$$\bar{\bar{y}} := (N_1 + N_2)^{-1} \left(\sum_{i \in \mathcal{I}_1} y_{i,1} + \sum_{i \in \mathcal{I}_2} y_{i,2} \right) = (NT)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \right),$$

so that $\widehat{\dot{\alpha}}_1 + \widehat{\dot{\alpha}}_2 = 0$. In general, however, with unbalanced data this zero-sum is not guaranteed: a counter-example is the dataset

$$\{y_{it}\} = \begin{bmatrix} 1 & 2 \\ 3 & . \\ 4 & 5 \end{bmatrix},$$

which has $A = 6 + \frac{1}{6} \neq B = 6 + 1.5 \neq T\bar{\bar{y}} = 6$.

Lemma 3. Method 3 can be used on unbalanced data while ensuring that $\sum_{t=1}^{T} \widehat{\dot{\alpha}}_t = 0$.

Proof. By construction, method 3 estimates $T-1$ terms, $\widehat{\tilde{\dot{\alpha}}}_t$ for $t = 2, 3, ..., T$, then computes

$$\widehat{\dot{\alpha}}_1 = -T^{-1} \sum_{t=2}^{T} \widehat{\tilde{\dot{\alpha}}}_t,$$

and

$$\widehat{\dot{\alpha}}_t = \widehat{\dot{\alpha}}_1 + \widehat{\tilde{\dot{\alpha}}}_t,$$

for $t = 2, 3, ..., T$, which ensures that

$$\sum_{t=1}^{T} \widehat{\dot{\alpha}}_t = 0.$$

Other comments. The unmodified GFE model estimate obtains the time profiles arising from interacting group identifier dummy variables with a set of T time dummy variables, analogous to using method 1 in the $G = 1$ case. In the modified procedure's parameter update step, the algorithm interacts the group identifier dummy variables with $T-1$ time-demeaned time dummy variables, generating $G(T-1)$ estimates of relative effects $\tilde{\dot{\alpha}}_{gt} := \dot{\alpha}_{gt} - \dot{\alpha}_{g,1}$; this is analogous to method 3 for $G = 1$. As the GFE algorithm requires GT estimates of the $\dot{\alpha}_{gt}$ to compute the sum of squared residuals across all i and t, it is necessary to perform the conversion step from the $T-1$ estimates of the $\tilde{\dot{\alpha}}_{gt}$ to T estimates of the $\dot{\alpha}_{gt}$, for each of the G time profiles.

A.2 Robustness checks

This section explores the sensitivity of the main results to changes in data composition and estimation methodology.

A.2.1 Data composition

The first robustness check determines whether the results depend materially on using a balanced subsample of each fund's data. It uses the same model as the main results, but the data is filtered down to retain a fully balanced sample, leaving $N = 8274$ units in the sample.

Covariate effects

Figure A.1 plots covariate effect estimates for different values of G using the fully balanced subsample. Like the dataset used for the main results in the chapter, the fully balanced subsample supports a seven-group model using the selection method in the chapter.

Time profiles

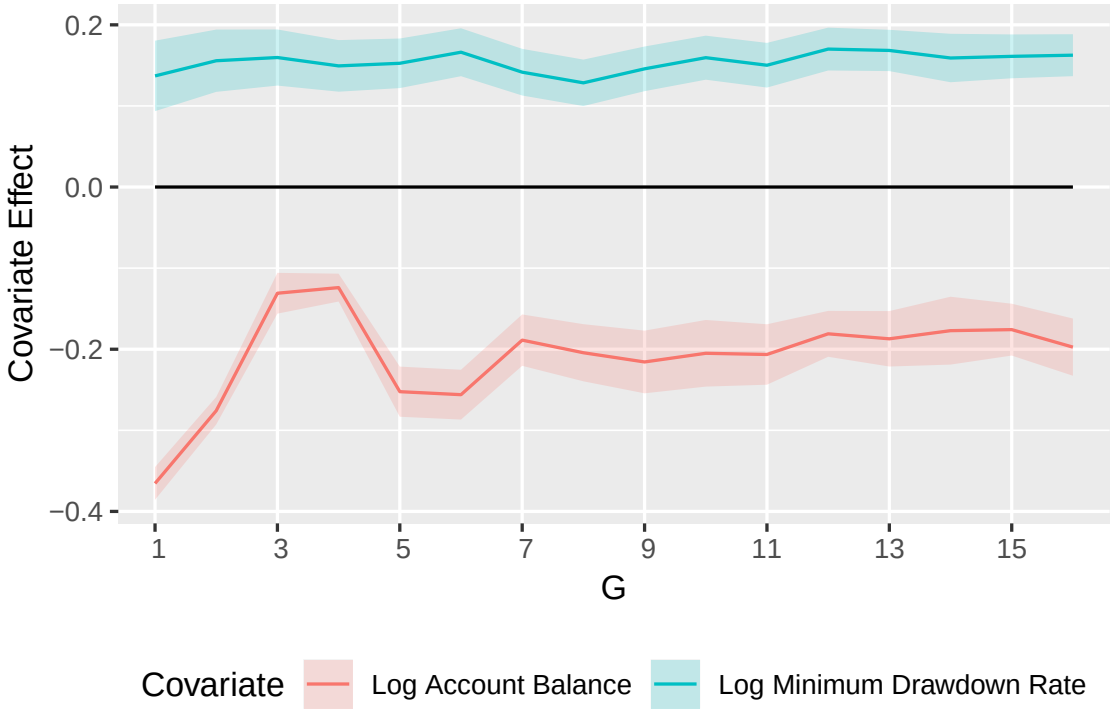
Figure A.2 plots the time profiles for $G = 4, 5, ..., 9$ using the fully balanced subsample. Figure A.3 shows the time profiles for $G = 7$, with element-wise confidence intervals computed using standard error estimates derived from the fixed-T variance estimate formula. While the numerical values differ, the economic interpretation of these results is similar to the interpretation of the main results in the chapter.

A.2.2 Estimation methodology

A.2.2.1 Number of starting values

The number of possible allocations of N units into G groups is G^N, and solutions found by the GFE procedure depend on starting values for the algorithm (Bonhomme &

Figure A.1: Fully balanced subsample – point estimates and 95% confidence intervals for the partial effects of the log minimum drawdown rate and log account balance covariates on the log regular drawdown rate, controlling for group-level time-varying unobservable heterogeneity assuming $G = 1, 2, ..., 16$. The shaded regions denote confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula.



Manresa, 2015); hence, as N and G increase, the GFE procedure is more likely to find a local, rather than global, optimum. This may require an increasing number of randomly selected starting values for the algorithm to adequately explore the solution space, if individual runs become trapped in neighbourhoods of local optima.

To test the sensitivity of the main results to the choice of 1000 starting values, Figure A.4 provides the estimated group time profiles from the equivalent estimation using 1 million starting values. With 1000 starting values, the objective function value is 2673.732, while with 1 million starting values the value is 2673.716. This change in the sum of squared errors across 107,935 data points is small, suggesting there is little gained from running the estimation procedure for more than 1000 starting values; the following paragraphs investigate whether there are any economically meaningful differences in the results with 1 million starting values compared to the main results.

The time profiles are nearly identical to those from the main results in the chapter. The largest absolute difference in any pair of corresponding time profile value estimates is

Figure A.2: Fully balanced subsample – point estimates for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 4, 5, ..., 9$. The estimated time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.

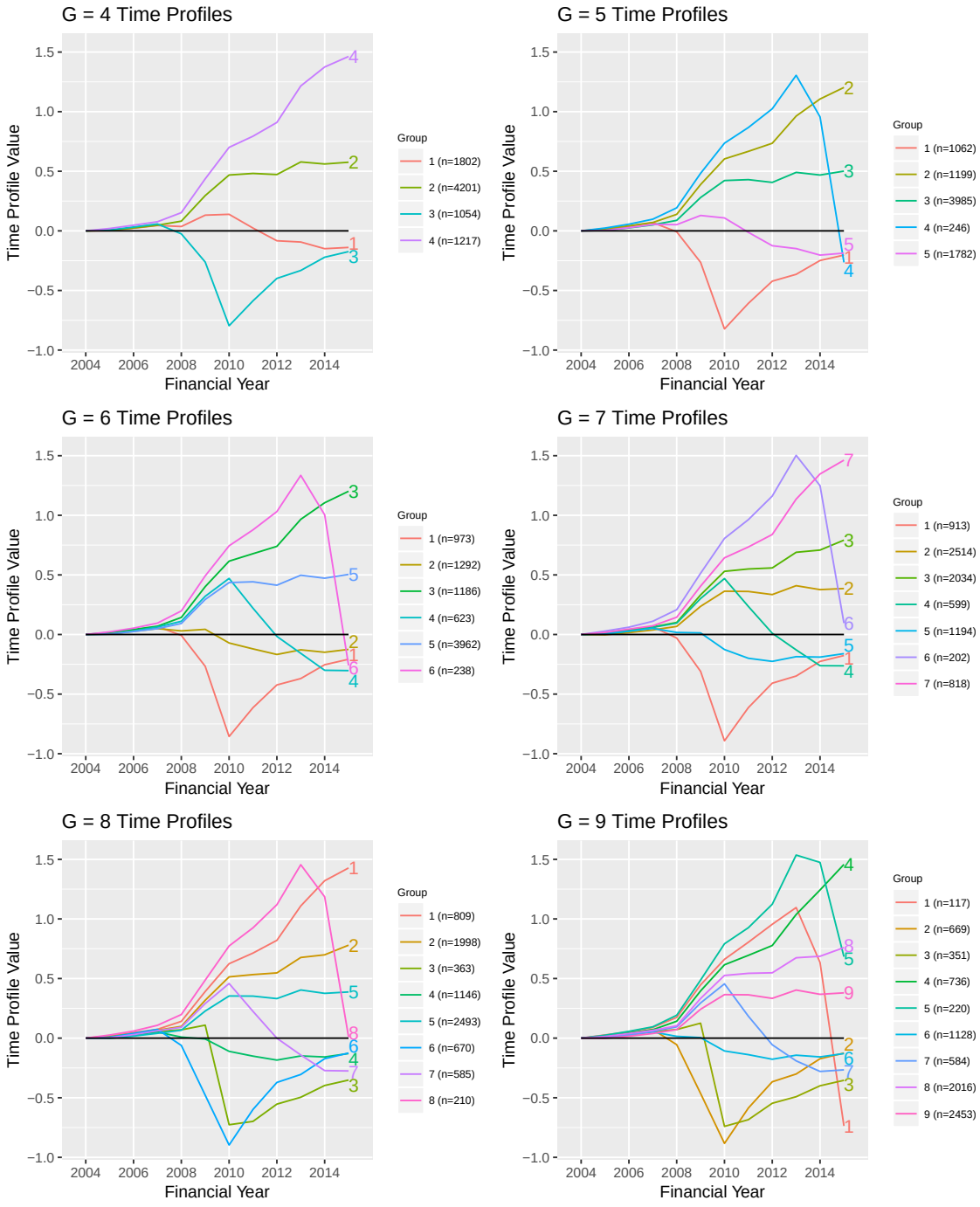


Figure A.3: Fully balanced subsample – point estimates and 95% confidence intervals from the analytical formula for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 7$. The shaded regions denote 95% element-wise confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula. The time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.

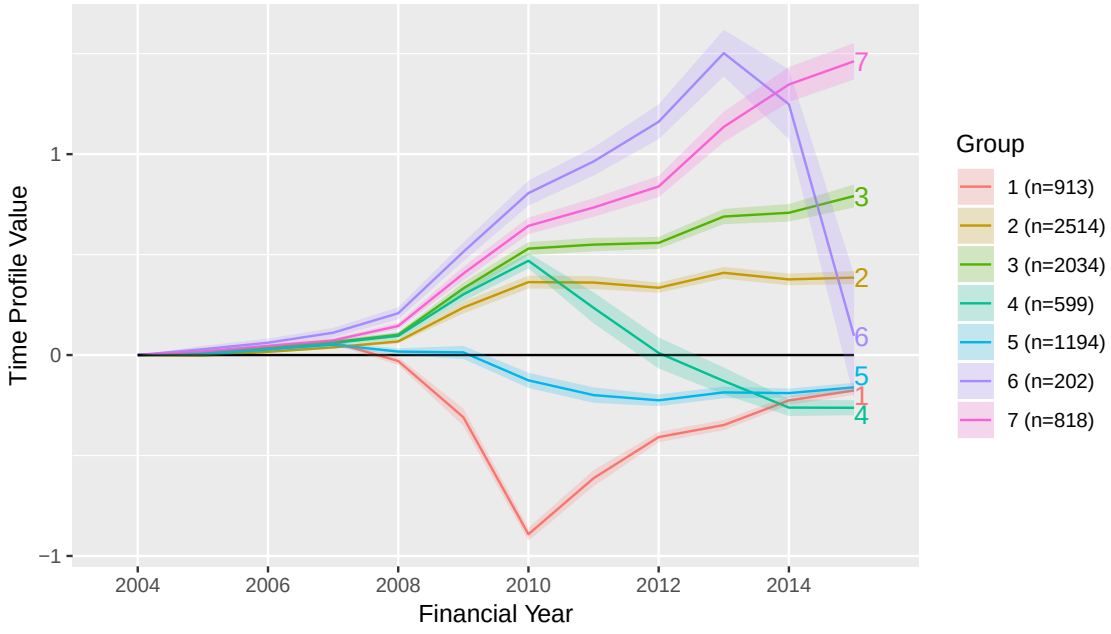
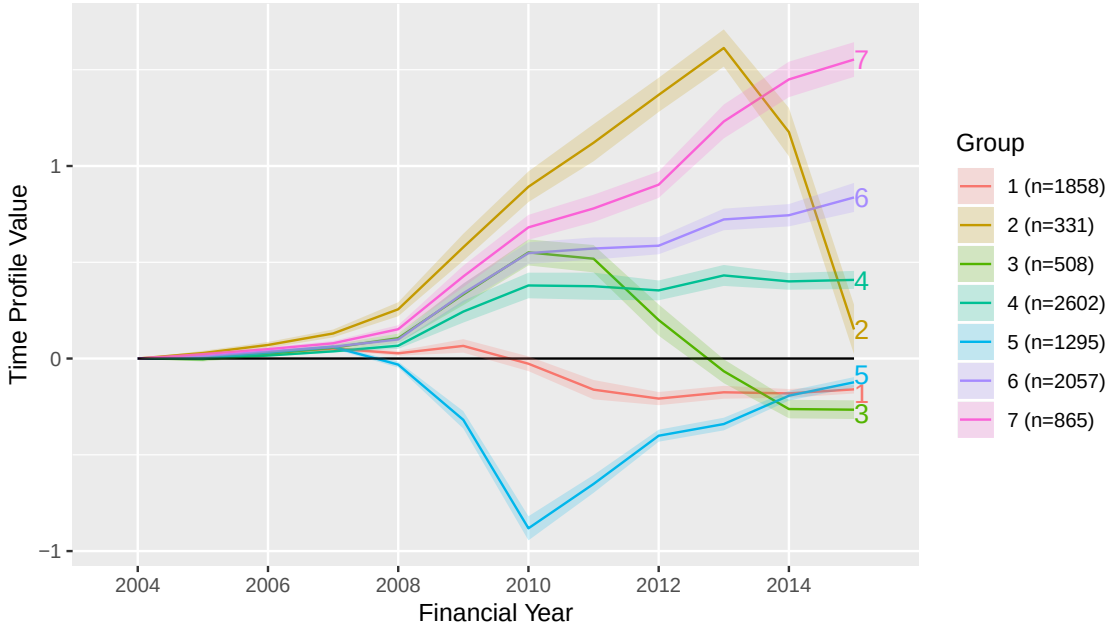


Figure A.4: One million starting values – point estimates and 95% confidence intervals from the analytical formula for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 7$. The shaded regions denote 95% element-wise confidence intervals derived from the fixed-T variance estimate formula. The time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.



approximately 1.49×10^{-3}. With 1 million starting values, using the group labels as in the main results in the chapter, group 4 contains four more people, group 7 has one person less and group 5 has three fewer people, compared to the run with 1000 starting values.

Table A.1 also provides the covariate effect estimates for estimation with 1 million starting values. The estimates and standard errors are identical to those in the main results to three decimal places. The largest difference in estimated effect magnitudes is approximately 2.87×10^{-4}, for the coefficient on the log minimum drawdown rate variable. These results suggest that the economic interpretation of the main results in the chapter is robust to the number of starting values used; however, it is still possible that there exists a more optimal solution with materially different results for the covariate effects or time profiles that was not found using 1 million starting values.

A.2.2.2 The unmodified estimation procedure

The chapter uses a modified estimation method for unbalanced data, having the property that for $G = 1$, the results align precisely with those obtained by running a standard two-way fixed-effects regression model. Comparison output is presented here, using the unmodified algorithm to test the sensitivity of the main results in the chapter to this alternative procedure.

Figure A.5 shows how the covariate effect estimates evolve with G. Overall, the results for the log account balance variable appear almost identical to those in the main results. Although the results for the log minimum drawdown rate differ more significantly in magnitude to those in the main results, their economic implications are similar.

Table A.1: One million starting values – point estimates and 95% confidence intervals for the partial effects of the log minimum drawdown rate and log account balance covariates on the log regular drawdown rate, controlling for group-level time-varying unobservable heterogeneity assuming $G = 6$. The standard errors are derived from the fixed-T variance estimate formula.

Covariate	Estimate (Standard Error)
Log Minimum Drawdown Rate	0.144 (0.0248)
Log Account Balance	−0.147 (0.0139)

Figure A.5: Unmodified estimation procedure – point estimates and 95% confidence intervals for the partial effects of the log minimum drawdown rate and log account balance covariates on the log regular drawdown rate, controlling for group-level time-varying unobservable heterogeneity assuming $G = 1, 2, ..., 16$. The shaded regions denote confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula.

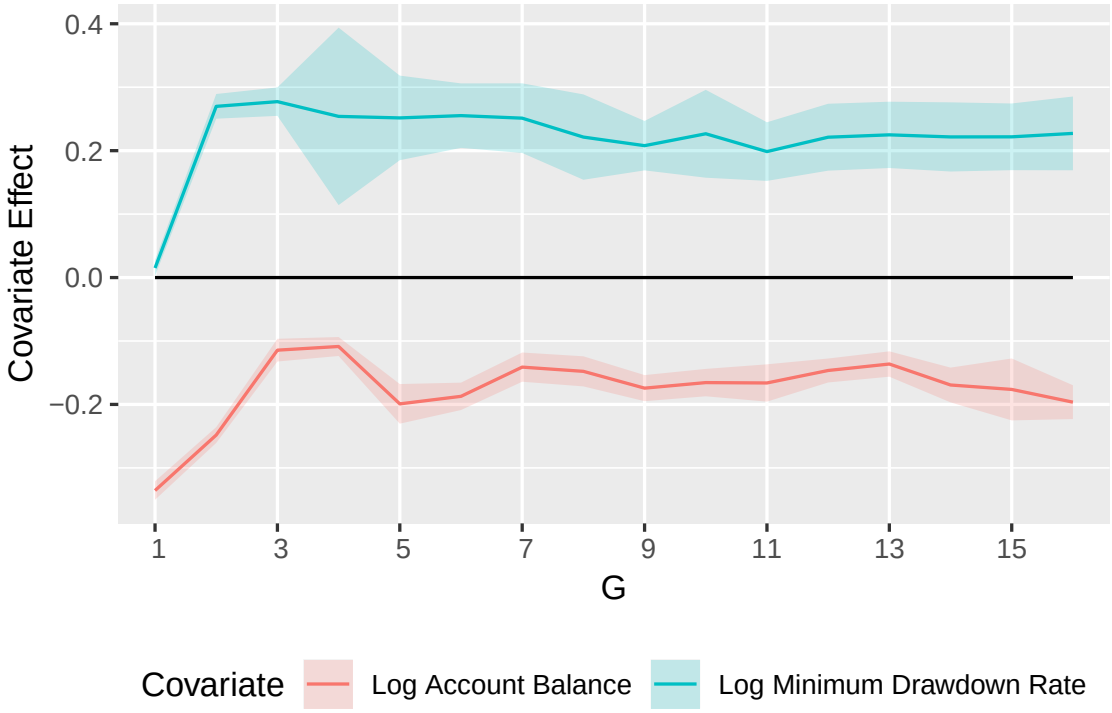


Figure A.6 presents time profile point estimates for $G = 4, 5, ..., 9$. Figure A.7 shows the time profile plot for $G = 7$, including 95% element-wise confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula. The plots show that the time profiles obtained from both estimation strategies are similar.

Figure A.8 shows the time profile plot for $G = 1$, with axes identical to the corresponding time profile plot in the main results section of the chapter. Comparing the plots reveals that the point estimates differ depending on the algorithm used, and the economic interpretations of time effects around financial year 2008 are different. Using the unmodified procedure suggests a small downward effect in 2008 followed by a gradual rise, while this initial drop is absent in the corresponding plot created using the modified algorithm in the chapter.

Figure A.6: Unmodified estimation procedure – point estimates for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 4, 5, ..., 9$. The estimated time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.

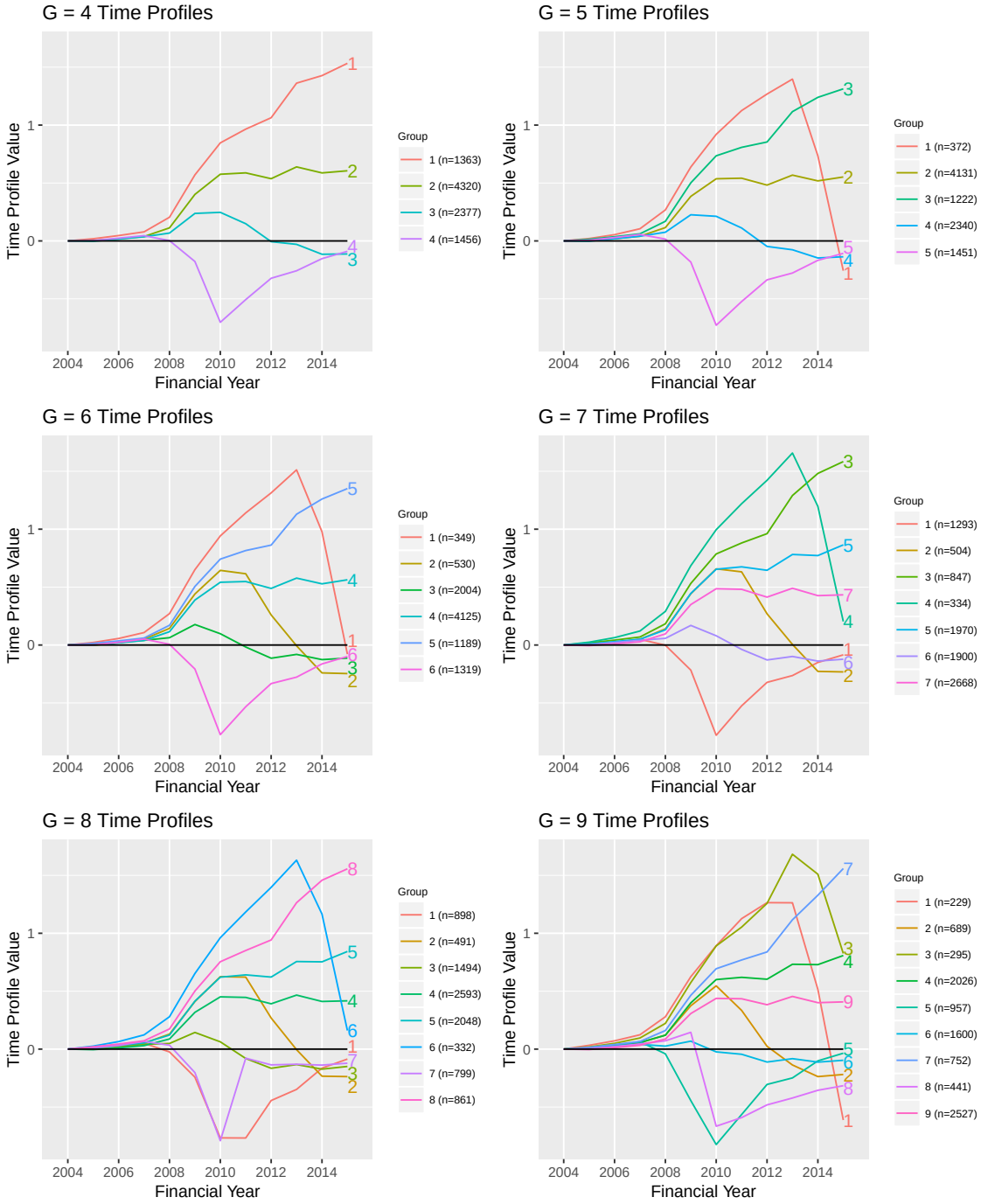


Figure A.7: Unmodified estimation procedure – point estimates and 95% confidence intervals from the analytical formula for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 7$. The shaded regions denote 95% element-wise confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula. The time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.

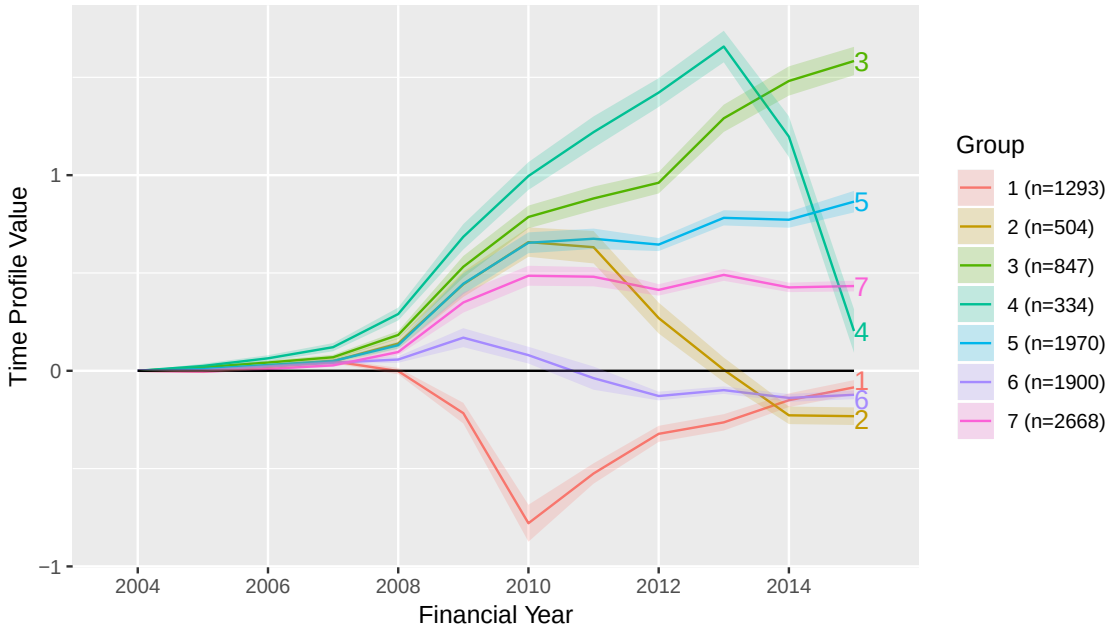
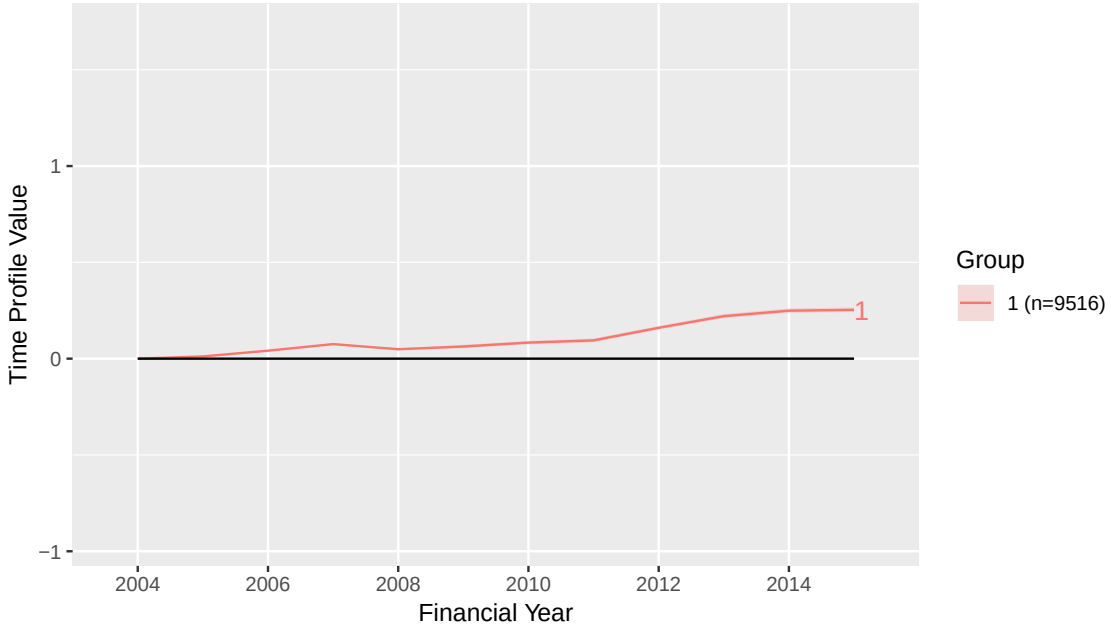


Figure A.8: Unmodified estimation procedure – point estimates and 95% confidence intervals from the analytical formula for the effects of group-level time-varying unobservable heterogeneity on the log regular drawdown rates assuming $G = 1$. The shaded regions (indistinguishable from point estimates in the plot) denote 95% element-wise confidence intervals constructed using standard errors derived from the fixed-T variance estimate formula. The time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.



A.3 Simulation study

This section uses simulation evidence to argue for the validity of the GFE procedure in applications with data resembling the superannuation dataset. It also uses the proposed method for matching labels between estimations to investigate how standard errors derived from the fixed-T variance estimate formula and the bootstrap compare to the simulated standard errors estimated across replicates.

A.3.1 Simulation methodology

Creating the simulated datasets and a framework for analysing the results are now described.

A.3.1.1 Generating simulated data

Consider the data generating process (DGP)

$$\dot{y}_{it}^{\star} := \dot{x}_{it}^{\star\prime}\widehat{\theta} + \widehat{\dot{\alpha}}_{g_i t} + \dot{v}_{it}^{\star}, \tag{A.4}$$

where:

- $\dot{x}_{it}^{\star} = (\dot{x}_{1,it}^{\star}, \dot{x}_{2,it}^{\star})'$ is a column vector of simulated covariate values for unit i at time t;
- $\widehat{\theta}$ and $\widehat{\dot{\alpha}}_{g_i t}$ are the GFE estimates for the covariate effects and time-demeaned group time profiles from the main results in the chapter, respectively;
- the simulated covariates have mean zero and there is no time-constant individual-specific fixed effect—that is, the generated data resembles the true data after time-demeaning;
- $\dot{x}_{k,it}^{\star} \sim N(0, \widehat{\sigma}^2_{\dot{x}_k})$, where $\widehat{\sigma}^2_{\dot{x}_k}$ is the sample variance of all values of the $\dot{x}_{k,it}$, for $k = 1, 2$;
- $\dot{v}_{it}^{\star} \sim N(0, \widehat{\sigma}^2_{\widehat{\dot{v}}})$, where $\widehat{\sigma}^2_{\widehat{\dot{v}}}$ is the sample variance of the empirical residuals $\widehat{\dot{v}}_{it} := \dot{y}_{it} - \dot{x}_{it}'\widehat{\theta} - \widehat{\dot{\alpha}}_{g_i t}$;

- $\dot{x}^{\star}_{k,it}$ are generated using a method that induces correlation between the time profile values $\widehat{\alpha}_{g_i t}$ and the covariates $\dot{x}^{\star}_{k,it}$, for $k = 1, 2$; this approximates the correlation observed in the data. The details are given below.

A simulation study using data simulated from (A.4) with the $\dot{x}^{\star}_{k,it}$ uncorrelated with $\widehat{\alpha}_{g_i t}$ would be unfaithful to the challenges involved in estimating the model on the original data. Recall that the GFE method allows for arbitrary correlation between the covariates and the unobserved grouped fixed effects. Moreover, in the absence of correlation between $\dot{x}^{\star}_{k,it}$ and $\widehat{\alpha}_{g_i t}$, a standard two-way fixed-effects regression of $\dot{y}$ on the $\dot{x}$-es directly obtains unbiased estimates $\widehat{\theta}$. Thus, recovering accurate estimates using the simulated data is unrealistically easy if the covariates are uncorrelated with the time profiles.

A realistic exercise simulates correlation between the covariates and the time profiles to match that observed in the data. Using the data to estimate the correlation statistics $\widehat{\rho}_{k,g}$, for all (k, g), allows a flexible correlation structure. The $\widehat{\rho}_{k,g}$ values are the correlations between values of $\dot{x}^{\star}_{k,it}$ and values of $\widehat{\alpha}_{g_i t}$; i.e., $\widehat{\rho}_{k,g_i}$ is the correlation between the value of covariate k and the value contributed to the dependent variable by individual i's group time profile. The $\widehat{\rho}_{k,g}$ are estimated for all (k, g) by:

1. filtering the observed data to keep only records where $g_i = g$;
2. computing the sample correlation statistic between the observed values of $\dot{x}^{\star}_{k,it}$ and corresponding estimated values of $\widehat{\alpha}_{g_i t}$; call this value $\widehat{\rho}_{k,g}$.

Having estimated the correlation statistics, the aim is to generate Gaussian random variables $\dot{x}^{\star}_{k,it}$ which have correlation $\widehat{\rho}_{k,g_i}$ with the $\widehat{\alpha}_{g_i t}$. The following procedure induces this correlation structure, treating the model estimates of $\widehat{\alpha}_{gt}$ for $(g, t) \in \{1, 2, ..., G\} \times \{1, 2, ..., T\}$ as if they had been drawn from a Gaussian distribution. For all $(k, i, t) \in \{1, 2\} \times \{1, 2, ..., N\} \times \{1, 2, ..., T\}$:

1. set $W_{1,kit} = \widehat{\alpha}_{g_i t} / \widehat{\sigma}_{\alpha}$, where $\widehat{\sigma}_{\alpha}$ is estimated using the sample standard deviation of the set of $G \times T$ estimated values $\widehat{\alpha}_{gt}$;
2. draw $W_{2,kit} \sim N(0, 1)$;

3. set $W_{3,kit} = \widehat{\rho}_{k,g_i} W_{1,kit} + \sqrt{1 - \widehat{\rho}^2_{k,g_i}} W_{2,kit}$;

4. set $\dot{x}^{\star}_{k,it} = \widehat{\sigma}_{\dot{x}_k} W_{3,kit}$, where $\widehat{\sigma}_{\dot{x}_k}$ is the sample standard deviation of all observed values of covariate $\dot{x}_k$.

The error term $\dot{v}^{\star}_{it} \sim N(0, \widehat{\sigma}^2_{\hat{v}})$, and (A.4) gives the simulated values of the dependent variable $\dot{y}^{\star}$. The GFE procedure with $G = 7$ is then run on a large number of simulated datasets. Comparing the GFE estimation results to the DGP values checks the validity of the GFE procedure applied to this setting and the code implementing the method.

A.3.1.2 Framework for interpreting the results

The simulation study investigates the following properties of the GFE estimator:

1. Whether the GFE procedure applied to a known DGP, constructed from the results of applying the GFE procedure to the superannuation dataset, estimates the DGP accurately in simulated datasets whose dimensions are the same as in the superannuation dataset.

2. The closeness of the confidence intervals derived from the fixed-T variance estimate formula to the 'simulated' confidence intervals—the intervals obtained by matching time profile estimates across a large number of simulated datasets and observing the empirical spread of estimates.

3. The closeness of the simulated confidence intervals to the bootstrap confidence intervals—the intervals obtained by matching time profile estimates across a large number of bootstrap samples drawn from the first simulated dataset and observing the empirical spread of the estimates.

The following steps are used to generate the simulation results:

1. $M = 1000$ datasets are generated independently from the DGP, each with $N = 9516$ units and covering $T = 12$ time periods. The GFE procedure is then run on each of these using 1000 random starting values for each estimation, and the standard errors are obtained from the fixed-T variance estimate formula.

2. Using the resulting set of M estimates of $\widehat{\theta}$ and $\widehat{\alpha} := \{\widehat{\alpha}_{gt}\}_{(g,t)\in\{1,2,\ldots,G\}\times\{1,2,\ldots,T\}}$ as samples $\{\widehat{\theta}^{(m)}\}_{m=1}^{M}$ and $\{\widehat{\alpha}^{(m)}\}_{m=1}^{M}$ from the sampling distributions of the estimators, simulated standard errors and 95% element-wise confidence intervals are estimated.

3. All bootstrap results are obtained by creating $B = 1000$ bootstrap replicate datasets using the method outlined in the methodology section of the chapter—except that here the first simulated dataset is treated as the source dataset for the bootstrap sampling. The GFE procedure with $G = 7$ is then run on each of the resulting bootstrap replicate datasets. The bootstrap results that follow are different to the results obtained in the main results section of the chapter, which use the observed data as the source dataset for bootstrap sampling.

4. Estimated time profiles are compared to the DGP time profiles after shifting all time profiles to begin at 0.

A.3.2 Results

For the first property listed in Section A.3.1.2, the input values used for the DGP are compared to those estimated using the GFE procedure on the first simulated dataset. Figure A.9 shows this comparison for the time profiles; Table A.2 compares the covariate effects numerically. The results are close and suggest that the GFE procedure recovers the parameters of the DGP with a high degree of accuracy.

Figure A.10 summarises the distribution of time profile estimates across all 1000 simulated datasets. The 95% element-wise confidence interval bounds represent the empirical 2.5 and 97.5 percentiles of each estimated value. The tightness of these confidence intervals around the DGP values suggests that in most simulated datasets, the time profile estimates are numerically close to the true values, and economically indistinguishable; i.e., the interpretations of the time profiles are the same. The DGP

Table A.2: DGP covariate effects vs. the first simulated dataset estimates.

	Log Minimum Drawdown Rate	**Log Account Balance**
DGP	0.1436	−0.1472
First simulated dataset	0.1445	−0.1490

Figure A.9: DGP time profiles vs. the first simulated dataset estimates. The red series represent DGP values; the blue series are time profile estimates derived from the first simulated dataset. The time-demeaned group time profiles are shifted to begin at 0 on the vertical axis.

time profile values for group 7 tend to be close to the upper bounds of the simulated empirical 95% intervals, and for the terminal time period, are slightly above the upper bound.

Table A.3 makes the corresponding comparison for the covariate effects. The simulated confidence intervals surround the DGP values; however, for both covariates, the DGP estimates are relatively close to the upper bounds of the intervals.

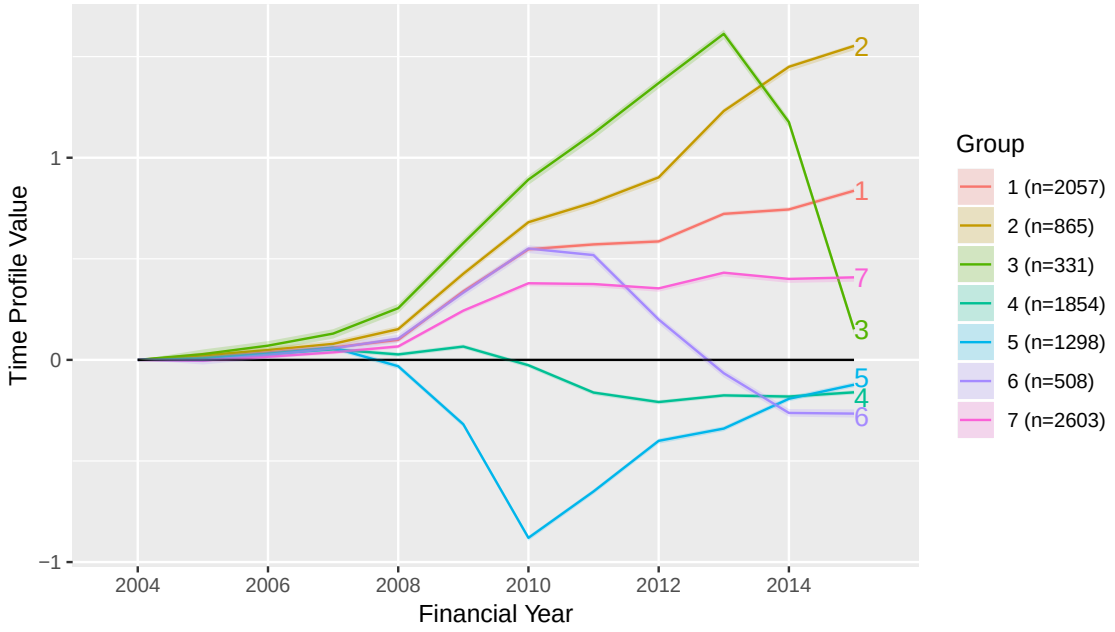
For the second property listed in Section A.3.1.2, Figure A.10 is compared to Figure A.11. Figure A.11 shows time profile estimates and 95% element-wise confidence intervals constructed from standard errors derived from the fixed-T variance estimate formula, where the input data is the first simulated dataset. The plots are nearly indistinguishable up to group relabelling, suggesting that the fixed-T variance estimate formula applied to the first simulated dataset estimates the true standard errors with

Table A.3: DGP covariate effects and simulated 95% CIs.

	Log Minimum Drawdown Rate	**Log Account Balance**
DGP value	0.1436	−0.1472
Simulated 95% CIs	[0.1340, 0.1459]	[−0.1522, −0.1471]

Simulated 95% CI bounds represent empirical 2.5 and 97.5 percentiles of the estimated covariate effects across 1000 simulated datasets.

Figure A.10: DGP time profiles and simulated 95% CIs. The time-demeaned group time profile values are the inputs to the DGP and here are shifted to begin at 0 on the vertical axis. The shaded regions denote 95% element-wise confidence intervals computed from the empirical percentiles of the shifted, time-demeaned group time profile estimates across 1000 simulated datasets.



reasonable precision. Table A.4 gives a similar comparison for the covariate effects, and has the same interpretation.

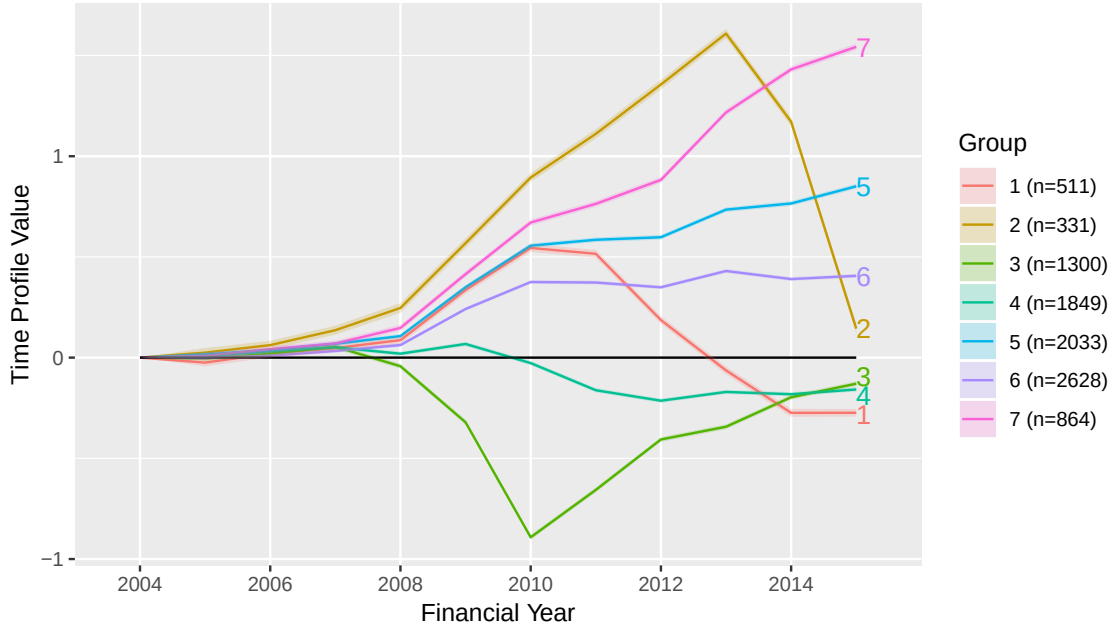
As with the previous comparison, all 1000 simulated datasets are considered next, by comparing the empirical distribution of the estimated standard errors—derived from the results of applying the fixed-T variance formula to 1000 simulated datasets—to the simulated standard errors—computed by calculating the sample standard deviation of parameter estimates across the 1000 simulated datasets. Table A.5 provides the figure references for the standard error distribution plots by group of the time-demeaned group time profile values, shifted to begin at zero in the first time period, corresponding to the financial year ended 30 June 2004. For each group, the plots show the empirical distribution of the standard errors for estimates corresponding to financial years 2005 to 2015, inclusive. The group labels follow Figure A.10, which shows the group time

Table A.4: Covariate effect estimates from the first simulated dataset, CIs from the formula.

	Log Minimum Drawdown Rate	**Log Account Balance**
Covariate Effect Estimate	0.1445	−0.1490
95% CI based on formula	[0.1415, 0.1475]	[−0.1515, −0.1465]

95% CIs derived from the fixed-T variance estimate formula.

Figure A.11: Time profile estimates from the first simulated dataset, CIs from formula. The time-demeaned group time profile values are from the GFE estimator run on the first simulated dataset and are shifted to begin at 0 on the vertical axis. The shaded regions denote 95% element-wise confidence intervals derived from the fixed-T variance estimate formula.



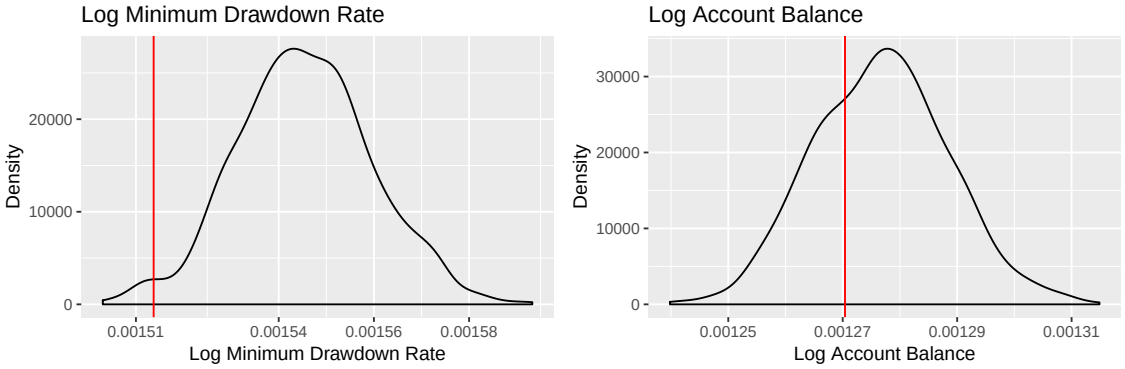
profiles with element-wise confidence intervals derived from the simulated standard errors. These are the same group labels presented in the figures for the main results in the chapter. In general, the simulated standard error value is in an area of nontrivial density in the corresponding empirical standard error distribution.

Figure A.12 provides the corresponding plots for standard errors of the covariate estimates. The fixed-T variance estimate formula tends to overestimate the true standard error for the first covariate; for the second covariate, the true standard error is more centrally located in the distribution of the empirical standard errors.

Table A.5: Lookup table – standard error distributions for the time profile estimates. The groups are labelled as per Figure A.10.

Group Label	Figure
1	A.18
2	A.19
3	A.20
4	A.21
5	A.22
6	A.23
7	A.24

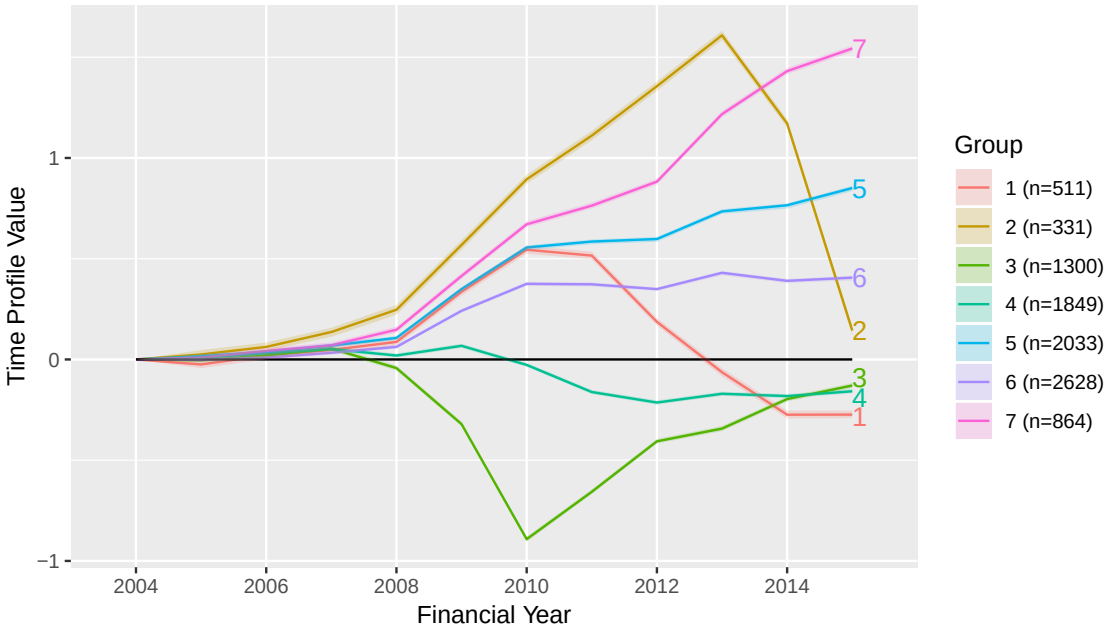
Figure A.12: Covariate effect analytical SE distributions across the simulated datasets. The black lines plot the kernel density estimates for standard errors derived from the fixed-T variance estimate formula after estimating the GFE model on 1000 simulated datasets. The red vertical lines represent the value of the simulated standard error.



Figures A.10 and A.13 are compared for the third property listed in Section A.3.1.2. Figure A.13 gives time profile estimates and 95% element-wise confidence intervals constructed from the empirical distribution of estimates across 1000 bootstrap replications, where the input data is the first simulated dataset. The results are again nearly indistinguishable up to group relabelling. This suggests that inference conducted using the bootstrap procedure gives almost identical results to that using the fixed-T variance estimate formula, which is a good approximation to the true standard errors. Ideally, the bootstrap procedure would be performed on all 1000 simulated datasets, and the resulting distributions of bootstrap standard errors compared to the simulated standard errors, as done for the fixed-T variance estimate formula. However, computational constraints prevent this.

Considering the simulation evidence, the fixed-T variance estimate formula performs well in datasets simulated from the generating process implied by the GFE estimation results for the superannuation dataset. This suggests that if the GFE assumptions are satisfied, the analytical formula may be adequate in datasets of similar size to the superannuation dataset, not requiring the bootstrap procedure for inference on the parameter estimates. This is important as using our code to perform the bootstrap on the superannuation dataset is currently too computationally intensive to run on a standard machine; it requires access to a high-performance computing cluster. As for the bootstrap results on the simulated data, these suggest that the bootstrap

Figure A.13: Time profile estimates from the first simulated dataset, CIs from the bootstrap. The results are from point estimates aggregated over 1000 bootstrap replications using the first simulated dataset to generate the bootstrap replicate datasets. The time-demeaned group time profile values are from the GFE estimator run on the first simulated dataset and are shifted to begin at 0 on the vertical axis. The shaded regions denote 95% element-wise confidence intervals computed from the empirical percentiles across the 1000 bootstrap replications.



may also perform comparably well; however, due to computational constraints, our implementation is unable to test this as rigorously as for the analytical formula.

A.4 Superannuation drawdowns dataset

A.4.1 More on covariate selection

Possible covariates for the analysis included in the dataset are limited to age, account balance and gender. From the available information, two derived covariates are also constructed: the minimum drawdown rate, and a crude estimate of an individual's risk appetite over the observation period. For each person–year observation, the age of the member maps to the relevant minimum drawdown rate the retiree is constrained by, with concessional reductions in the rates applying to the 2009–2013 financial years.

As the risk appetite variable is only used descriptively, and does not enter into the model estimation, its preciseness does not affect the main results. Its construction involves observing movements in account balances and comparing these with the

amounts drawn down and contributed to the funds by retirees. From this, ignoring administrative fees on the accounts, one can roughly estimate the return on assets. As the source data is at a monthly frequency, this return is computed monthly and then annualised; comparing it to the S&P/ASX 200 market index over matching time periods gives an approximate measure of sensitivity to market returns. Taking the magnitude of the average of these sensitivities then serves as a proxy for risk appetite.

Applying the within transformation—centering all variables around their time averages for each individual—prevents estimating the effect of any time-invariant covariates, which do not show within-individual variation; this includes gender as well as the constructed risk appetite variable. Thus, gender and risk appetite do not enter in the GFE estimation, although they are used to qualitatively characterise the groups that the GFE method finds in the data.

Furthermore, age is not included as a covariate because the focus is on estimating group effects for each time period; these time effects cannot easily be separated from the effect of ageing after within-transforming the data.

A.4.2 Exploratory data analysis

The remainder of this section presents a preliminary descriptive analysis of the dataset used to obtain the main results.

A.4.2.1 Summary statistics

Table A.6 summarises characteristics that vary across individuals, but not time. The age at 31st December 2015 represents the individual's cohort, equivalent to measuring a year-of-birth variable. The median retiree in the sample was born in 1936, with more than 50% of the sample born in an interval of four years on either side.

The age at account opening is the age when the retiree initiates a phased withdrawal product and begins drawing down from the account. In the superannuation dataset, the median retiree was aged 64 when opening their account. In general, opening an account before age 65 requires an individual to cease employment.

Table A.6: Summary statistics – time-invariant variables.

	Age at 31 December 2015	Age at Account Open	Sex: Male	Risk Appetite
Mean	79.4	63.57	0.56	0.41
SD	5.22	4.17	0.5	0.21
Median	79.78	64.23	1	0.46
Q1	76.37	60.9	0	0.25
Q3	83.03	65.39	1	0.54
Min	60.66	48.48	0	0
Max	101.46	85.44	1	1.88
Count	9516	9516	9516	9507

The sex indicator variable equals 1 if the the retiree is male, and 0 otherwise. The mean value of 0.56 represents the proportion of retirees in the sample that are male.

The risk appetite variable is a proxy for the returns in the account relative to the reference S&P/ASX 200 index. The median retiree earned approximately 46% of the index returns in the sample period, with 75% of the sample earning less than 55% of the index returns. This variable suggests that most retirees have asset mixes that are conservative or balanced, with few retirees seeking aggressive returns in these accounts.

Table A.7 summarises the time-varying variables in the dataset. The regular drawdown rate is the dependent variable of interest. The median drawdown rate in the sample is 9% of the account balances annually, while the mean is 12%. In absolute terms, the median regular drawdown amount is $4800 while the average is $6436.

The ad-hoc drawdown indicator variable equals 1 if the retiree made an ad-hoc withdrawal from their account balance during a given financial year; its mean value of 0.07 indicates that 7% of the observations recorded contain an ad-hoc drawdown. An interpretation is that the average retiree in the sample makes an ad-hoc drawdown roughly once every 14 years. Conditional on making an ad-hoc drawdown, the median ad-hoc drawdown rate is 7% of the account balance at the start of the year, while the

Table A.7: Summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.12	6435.91	0.07	0.15	10,216.56	72,686.55
SD	0.12	6121.76	0.25	0.21	24,672.68	78,721.39
Median	0.09	4800	0	0.07	4655.9	52,063
Q1	0.07	2992	0	0.02	1132.33	30,532.5
Q3	0.12	7728	0	0.18	10,000	87,427
Min	0	1	0	0	1	1
Max	2	166,695	1	0.9	600,000	2,427,083
Count	107,935	107,975	108,717	7450	7454	108,635

mean is 15%. In dollars, the median ad-hoc drawdown is $4656 and the average is $10,217.

Median account balances, as measured at the beginning of each financial year, are $52,063, with roughly 50% of balances lying in the interval ($30,000, $87,000).

A.4.2.2 Histograms

Figure A.14 plots the histograms of the time-invariant covariates. Most notable is the spike around age 65 for the account open age distribution, corresponding to the age at which individuals can open a phased withdrawal account unconditionally. Also instructive are the multiple peaks in the risk appetite distribution, suggesting a bunching of retirees into distinct asset mix options.

Figure A.15 plots the histograms for the time-varying variables, which show some evidence of ad-hoc drawdown rates bunching around several modes for the larger values.

Figure A.14: Histograms – time-invariant variables. These are graphical representations of the data summarised in Table A.6.

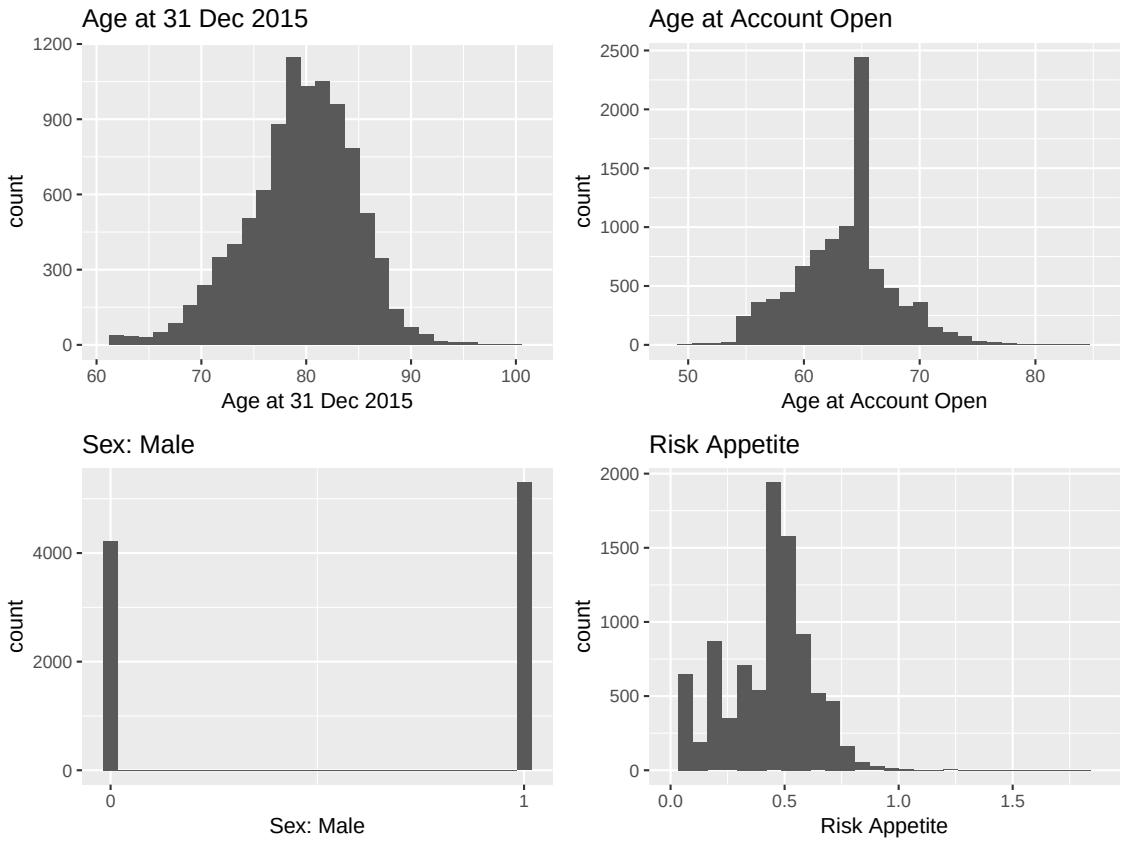
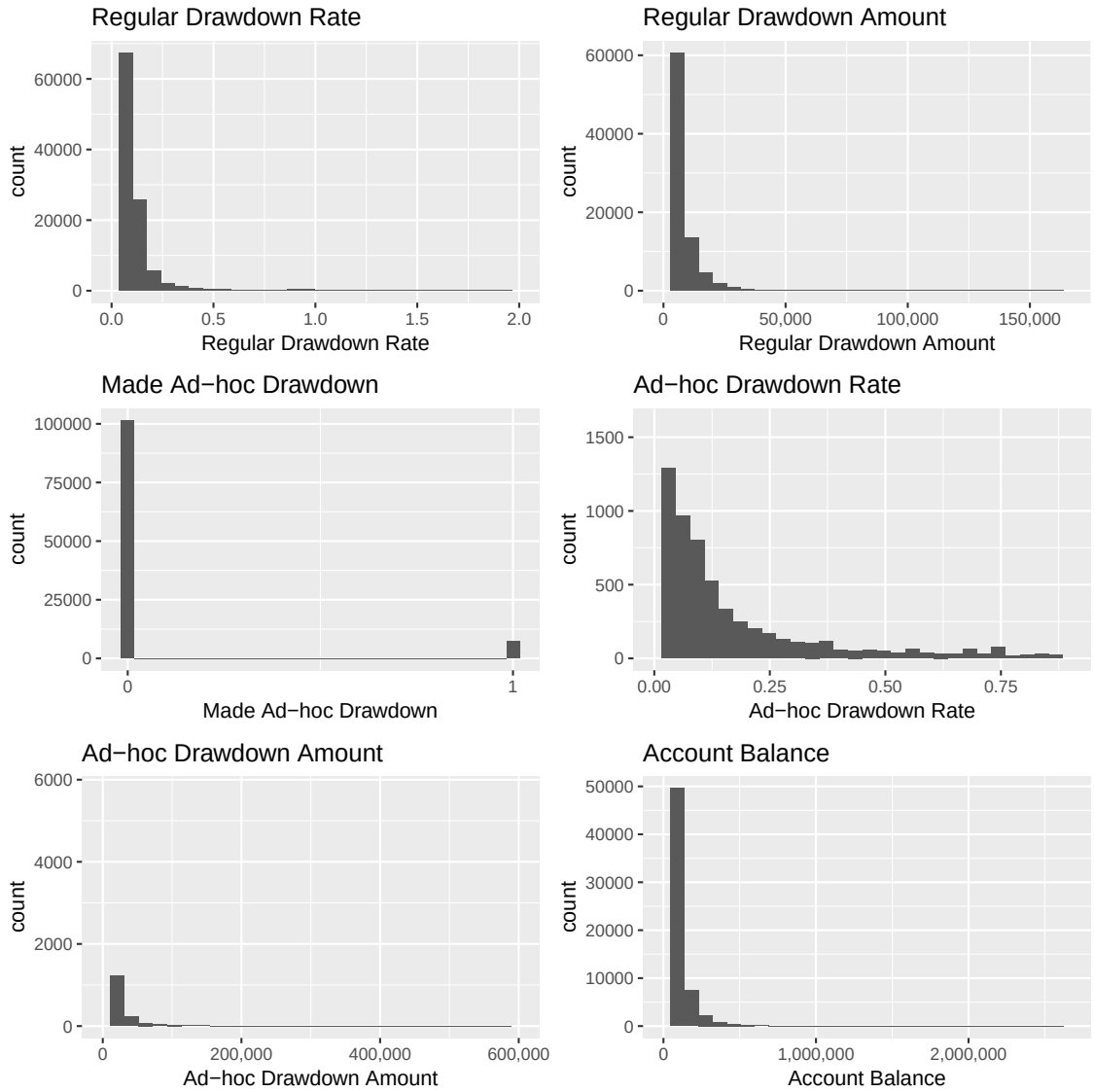


Figure A.15: Histograms – time-varying variables. These are graphical representations of the data summarised in Table A.7.



A.5 Characterising groups in the seven-group model

This section provides summary statistics and histograms created for each of the estimated groups from the main results. This allows comparison of each of the groups' characteristics against other groups, using the following set of tables and plots. Alternatively, it is possible to compare group-level characteristics against the aggregate sample, by comparing with the results in Section A.4.2. We also present panel plots to supplement those given in the chapter. Throughout this section, group labels follow the main results section in the chapter.

A.5.1 Summary statistics by group

A.5.1.1 Time-invariant variables

Table A.8 lists table references for the summary statistics of the time-invariant variables by group.

A.5.1.2 Time-varying variables

Table A.9 lists table references for the summary statistics of the time-varying variables by group.

A.5.2 Histograms by group

A.5.2.1 Time-invariant variables

Table A.10 lists figure references for the histograms of the time-invariant variables by group.

Table A.8: Lookup table – summary statistics for the time-invariant variables by group. The group labels follow the main results section in the chapter.

Group	Table
1	A.13
2	A.14
3	A.15
4	A.16
5	A.17
6	A.18
7	A.19

Table A.9: Lookup table – summary statistics for the time-varying variables by group. The group labels follow the main results section in the chapter.

Group	Table
1	A.20
2	A.21
3	A.22
4	A.23
5	A.24
6	A.25
7	A.26

Table A.10: Lookup table – histograms of the time-invariant variables by group. The group labels follow the main results section in the chapter.

Group	Figure
1	A.25
2	A.26
3	A.27
4	A.28
5	A.29
6	A.30
7	A.31

A.5.2.2 Time-varying variables

Table A.11 lists figure references for the histograms of the time-varying variables by group.

A.5.3 Time-demeaned panel plots by group

Table A.12 lists figure references for time-demeaned panel plots by group. Each figure shows four variables after time-demeaning by unit: the log regular drawdown rate; the log regular drawdown dollar amount; the log account balance at the financial year start; composite residuals from the estimation. We define the composite residuals as $\widehat{\dot{\alpha}}_{g_i t} + \widehat{\dot{v}}_{it} := \dot{y}_{it} - \dot{x}'_{it}\widehat{\theta}$; a composite residual is the estimated group time profile value plus the model residual, obtained by subtracting the estimated effect of covariates from the dependent variable. The black line represents the estimated time-demeaned group time profile values $\widehat{\dot{\alpha}}_{gt}$. The composite residual plots provide information similar to that of a residual plot of $\widehat{\dot{v}}_{it}$, but also convey a sense of how large the estimated

Table A.11: Lookup table – histograms of the time-varying variables by group. The group labels follow the main results section in the chapter.

Group	Figure
1	A.32
2	A.33
3	A.34
4	A.35
5	A.36
6	A.37
7	A.38

idiosyncratic unobserved component $\widehat{v}_{it}$ is compared to the estimated systematic unobserved component $\widehat{\alpha}_{g_i t}$.

A.6 Panel plots for the two-group model

Figures A.16 and A.17 show panel plots of time-demeaned variables for the two groups in the two-group model. Group labels in these plots follow the two-group model results in the chapter.

Table A.12: Lookup table – time-demeaned (TD) panel plots by group. The group labels follow the main results section in the chapter.

Group	Figure
1	A.39
2	A.40
3	A.41
4	A.42
5	A.43
6	A.44
7	A.45

Table A.13: Group 1 summary statistics – time-invariant variables.

	Age at 31 Dec 2015	Age at Account Open	Sex: Male	Risk Appetite
Mean	81.53	64.67	0.62	0.43
SD	4.22	3.84	0.49	0.18
Median	81.76	65	1	0.46
Q1	78.87	62.69	0	0.34
Q3	84.57	66.67	1	0.54
Min	67.67	50.45	0	0
Max	94.7	80.22	1	0.92
Count	2057	2057	2057	2055

Figure A.16: $G = 2$ model – group 1 time-demeaned (TD) panel plots. The account balances are as at the financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.

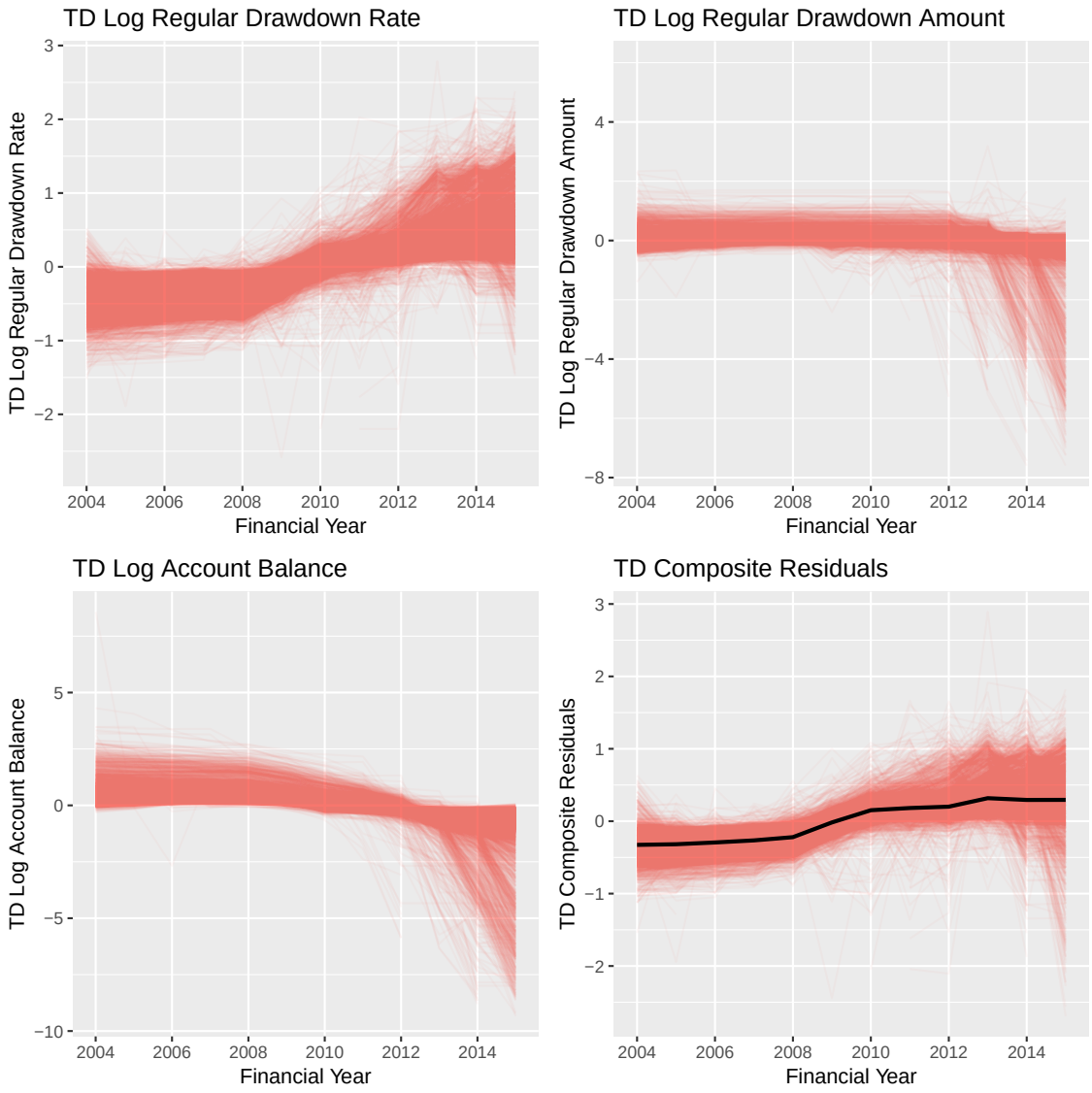


Table A.14: Group 2 summary statistics – time-invariant variables.

	Age at 31 December 2015	Age at Account Open	Sex: Male	Risk Appetite
Mean	80.03	63.36	0.62	0.41
SD	5.23	5	0.48	0.2
Median	79.41	63.67	1	0.46
Q1	76.42	59.82	0	0.24
Q3	83.54	65.57	1	0.54
Min	67.53	50.26	0	0.01
Max	96.23	81.78	1	1.02
Count	865	865	865	863

Figure A.17: $G = 2$ model – group 2 time-demeaned (TD) panel plots. The account balances are as at the financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.

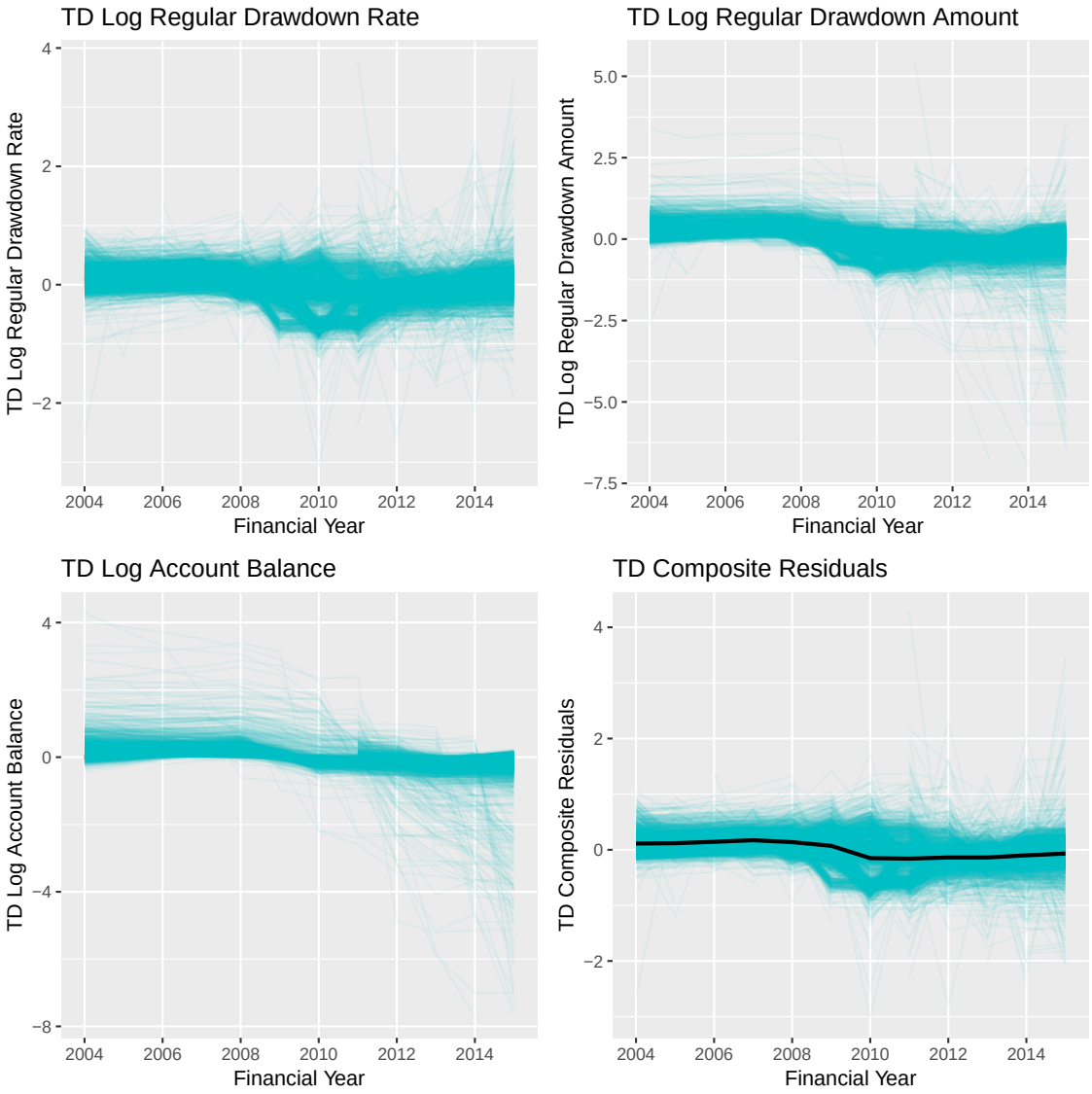


Table A.15: Group 3 summary statistics – time-invariant variables.

	age_at_31DEC15	age_at_account_open	sex_male	risk_appetite
Mean	81.75	64.66	0.69	0.35
SD	4.43	4.29	0.46	0.22
Median	81.39	64.94	1	0.4
Q1	79.24	61.93	0	0.17
Q3	83.88	65.74	1	0.5
Min	65.82	55.92	0	0
Max	99.87	85.44	1	1.52
Count	331	331	331	331

Table A.16: Group 4 summary statistics – time-invariant variables.

	Age at 31 December 2015	Age at Account Open	Sex: Male	Risk Appetite
Mean	78.17	63.54	0.51	0.44
SD	6.09	4.19	0.5	0.21
Median	78.93	64.1	1	0.47
Q1	74.25	60.79	0	0.31
Q3	82.57	65.46	1	0.57
Min	60.66	51.3	0	0
Max	97.06	79.79	1	1.59
Count	1854	1854	1854	1852

Table A.17: Group 5 summary statistics – time-invariant variables.

	Age at 31 December 2015	Age at Account Open	Sex: Male	Risk Appetite
Mean	78.12	63.92	0.49	0.47
SD	6.17	4.25	0.5	0.21
Median	78.8	64.35	0	0.49
Q1	73.35	61.48	0	0.35
Q3	82.77	65.93	1	0.6
Min	60.81	51.58	0	0
Max	101.46	81.91	1	1.88
Count	1298	1298	1298	1298

Table A.18: Group 6 summary statistics – time-invariant variables.

	Age at 31 December 2015	Age at Account Open	Sex: Male	Risk Appetite
Mean	78.19	63.16	0.53	0.45
SD	5.7	4.16	0.5	0.21
Median	78.99	63.78	1	0.48
Q1	74.56	60.5	0	0.34
Q3	82.34	65.29	1	0.58
Min	61.74	52.68	0	0
Max	94.21	76.01	1	1.09
Count	508	508	508	508

Table A.19: Group 7 summary statistics – time-invariant variables.

	Age at 31 December 2015	Age at Account Open	Sex: Male	Risk Appetite
Mean	78.97	62.56	0.54	0.35
SD	3.93	3.77	0.5	0.2
Median	79.08	63.06	1	0.39
Q1	76.49	60.15	0	0.22
Q3	81.83	65.02	1	0.5
Min	65.27	48.48	0	0
Max	92.5	78.32	1	0.88
Count	2603	2603	2603	2600

Table A.20: Group 1 summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.13	6687.54	0.05	0.14	10434.4	60246.23
SD	0.07	5615.68	0.21	0.16	20751.13	56052.49
Median	0.11	5172	0	0.09	6000	44855
Q1	0.09	3360	0	0.05	3432	27314
Q3	0.15	7944	0	0.15	10013	73574.5
Min	0.01	1	0	0	1	1
Max	1.22	70800	1	0.9	500000	812479
Count	24638	24644	24684	1144	1145	24675

Table A.21: Group 2 summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.25	6933.94	0.1	0.18	10224.05	46450.67
SD	0.23	5594.9	0.3	0.17	19495.22	47647.52
Median	0.15	5460	0	0.12	5000	33775
Q1	0.11	3372	0	0.06	2000	15689.5
Q3	0.3	8928	0	0.25	10075	62153.75
Min	0.01	2	0	0	1	1
Max	2	42000	1	0.9	277831	543370
Count	10344	10345	10372	1114	1115	10368

Table A.22: Group 3 summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.34	6076.35	0.1	0.35	9789.94	33754.28
SD	0.3	6305.35	0.3	0.3	19162.35	46781.99
Median	0.2	4400	0	0.23	4013	19960
Q1	0.14	2291	0	0.09	42	5732.25
Q3	0.46	7481.25	0	0.6	10000	42861.5
Min	0	1	0	0	1	1
Max	1.38	48000	1	0.9	233305	632325
Count	3716	3716	3944	461	461	3914

Table A.23: Group 4 summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.07	6412.78	0.09	0.11	8635.62	90976.64
SD	0.03	6524.65	0.29	0.2	25313.44	94553.22
Median	0.07	4588.91	0	0.02	2253.33	65456
Q1	0.06	2712	0	0.01	732.5	39864
Q3	0.08	7560	0	0.09	7500	104543
Min	0	30	0	0	1	15
Max	0.94	99768	1	0.9	430000	1573153
Count	19460	19472	19609	1831	1833	19597

Table A.24: Group 5 summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.06	5898.44	0.11	0.11	9308.75	99415.18
SD	0.04	6219.52	0.31	0.19	26711.86	103532.05
Median	0.06	4201.49	0	0.03	2565.83	71760
Q1	0.05	2310	0	0	622.93	43445.75
Q3	0.07	7094.75	0	0.09	8705.42	115423.25
Min	0	10	0	0	3	53
Max	0.75	109320	1	0.9	600000	1514586.47
Count	13184	13192	13294	1403	1403	13286

Figure A.18: Group 1 time profile analytical SE distributions across the simulated datasets. The black lines plot the kernel density estimates for standard errors derived from the fixed-T variance estimate formula after estimating the GFE model on 1000 simulated datasets. The red vertical lines represent the value of the simulated standard error.

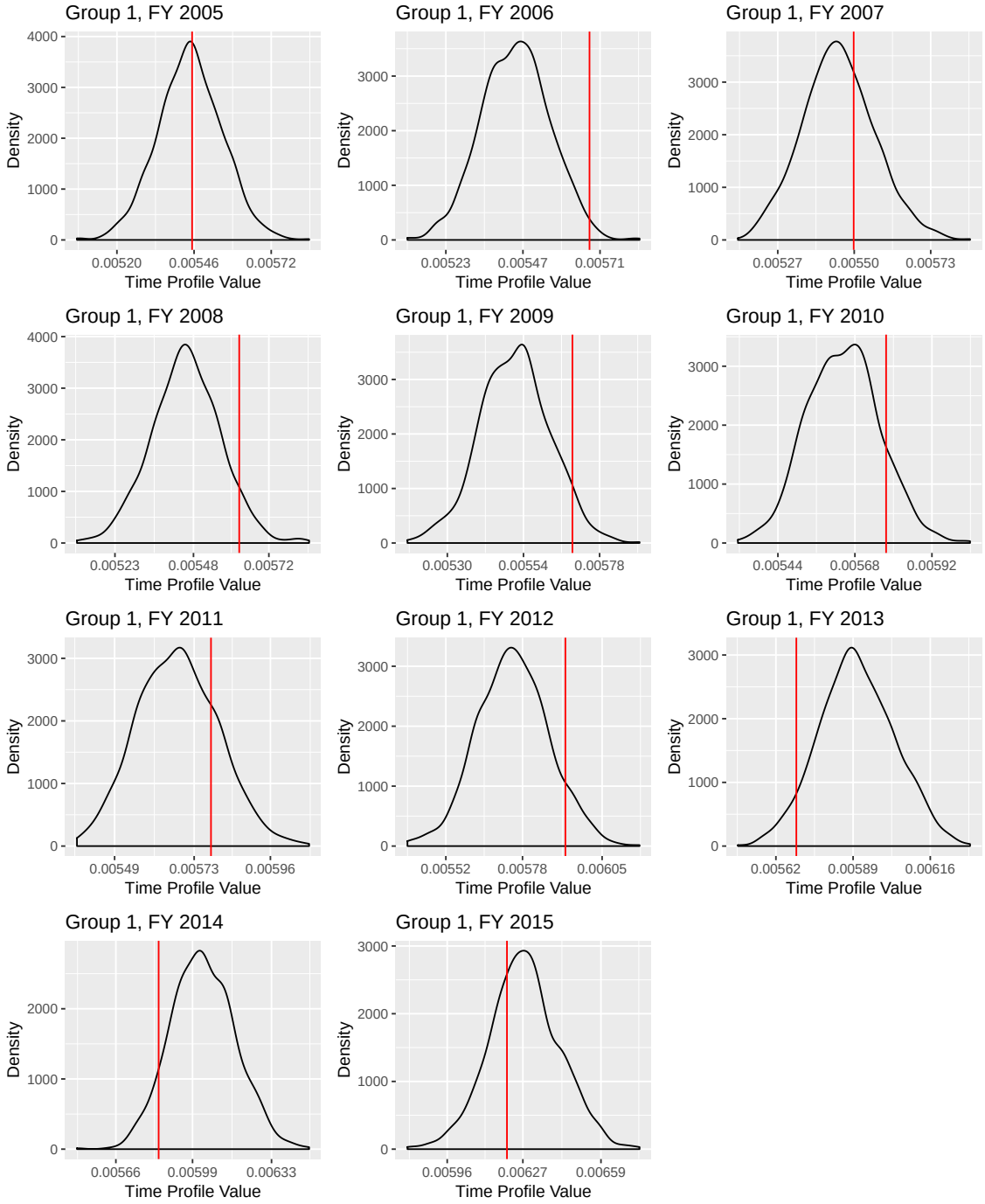


Figure A.19: Group 2 time profile analytical SE distributions across the simulated datasets. The black lines plot the kernel density estimates for standard errors derived from the fixed-T variance estimate formula after estimating the GFE model on 1000 simulated datasets. The red vertical lines represent the value of the simulated standard error.

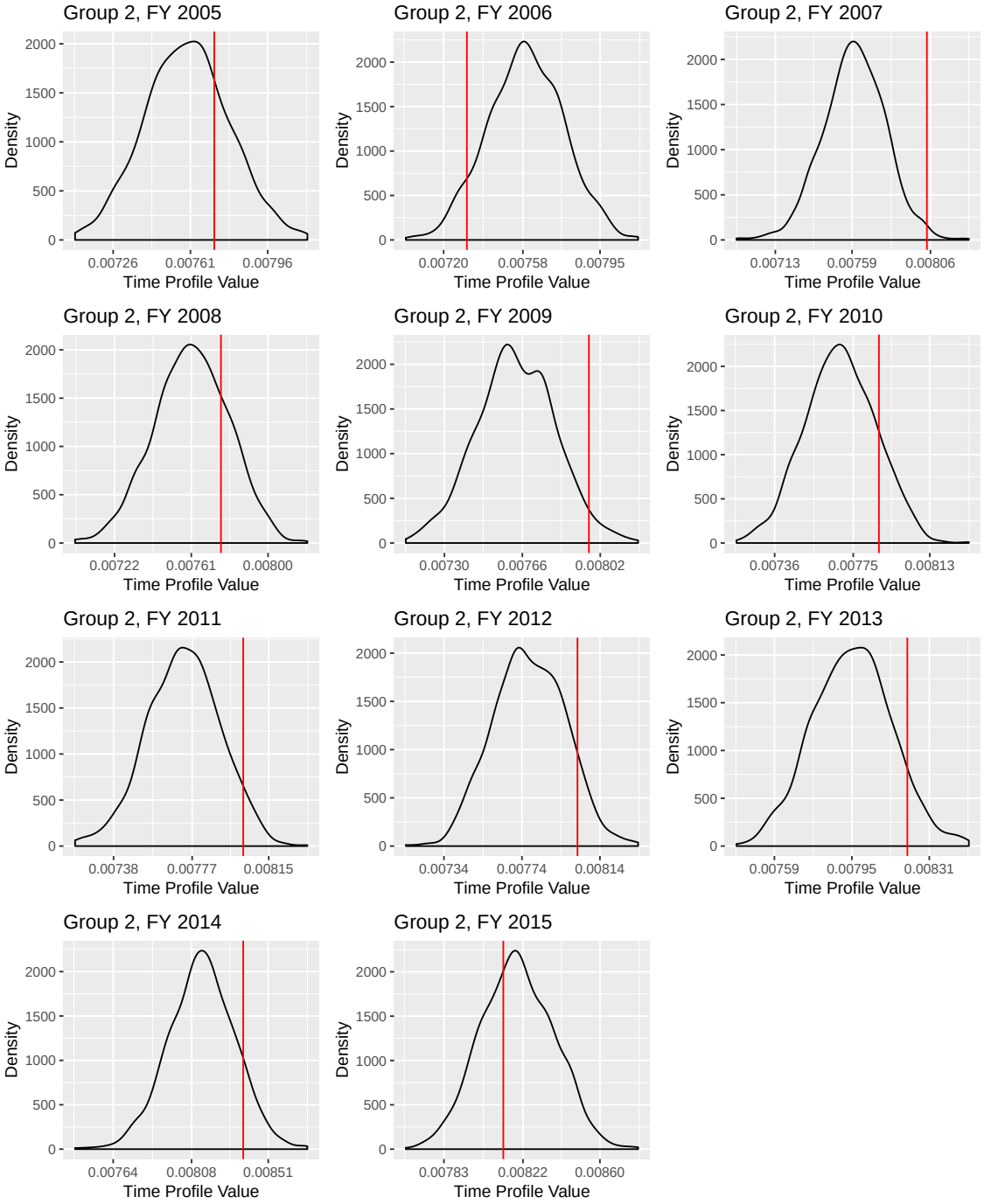


Figure A.20: Group 3 time profile analytical SE distributions across the simulated datasets. The black lines plot the kernel density estimates for standard errors derived from the fixed-T variance estimate formula after estimating the GFE model on 1000 simulated datasets. The red vertical lines represent the value of the simulated standard error.

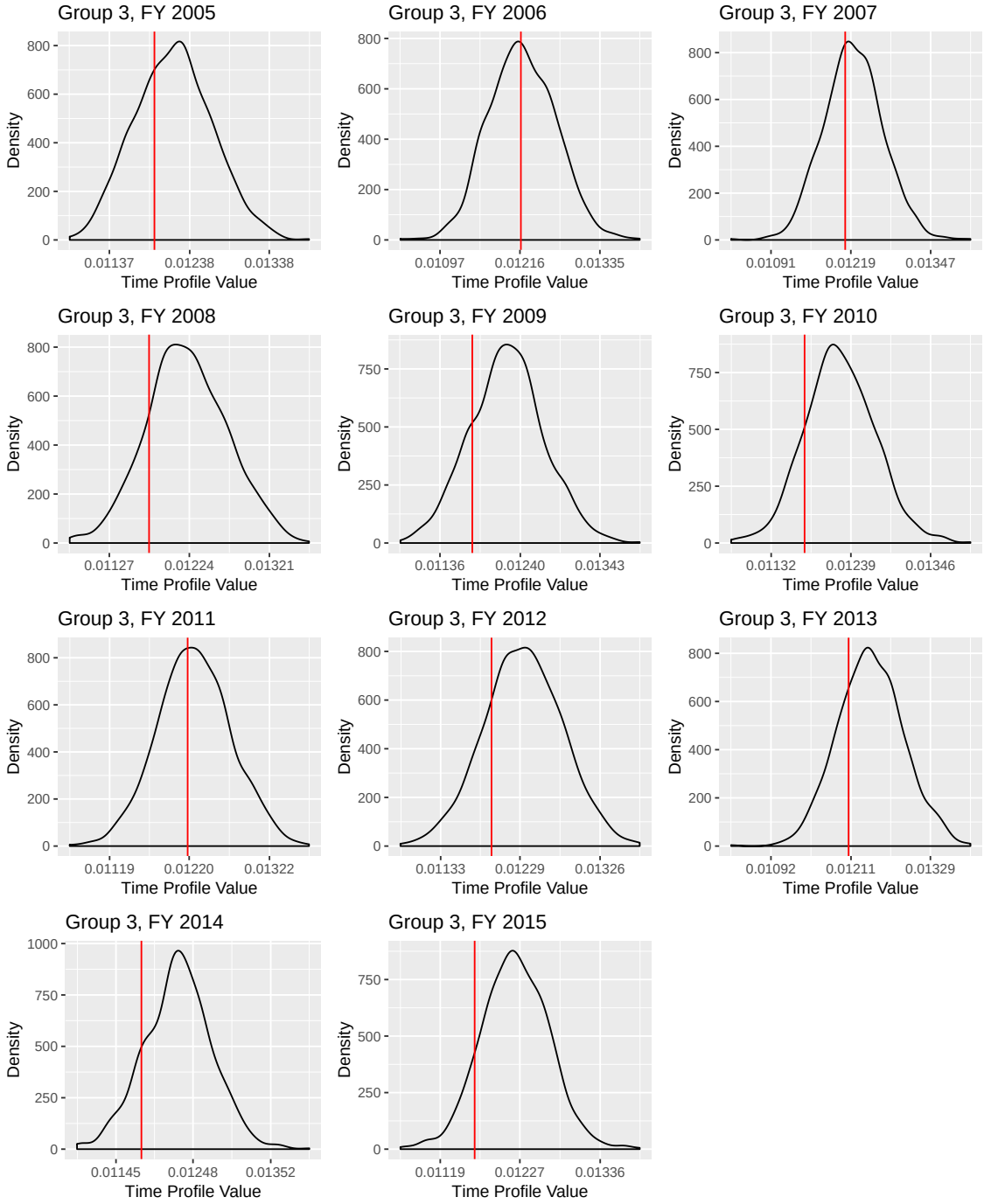


Figure A.21: Group 4 time profile analytical SE distributions across the simulated datasets. The black lines plot the kernel density estimates for standard errors derived from the fixed-T variance estimate formula after estimating the GFE model on 1000 simulated datasets. The red vertical lines represent the value of the simulated standard error.

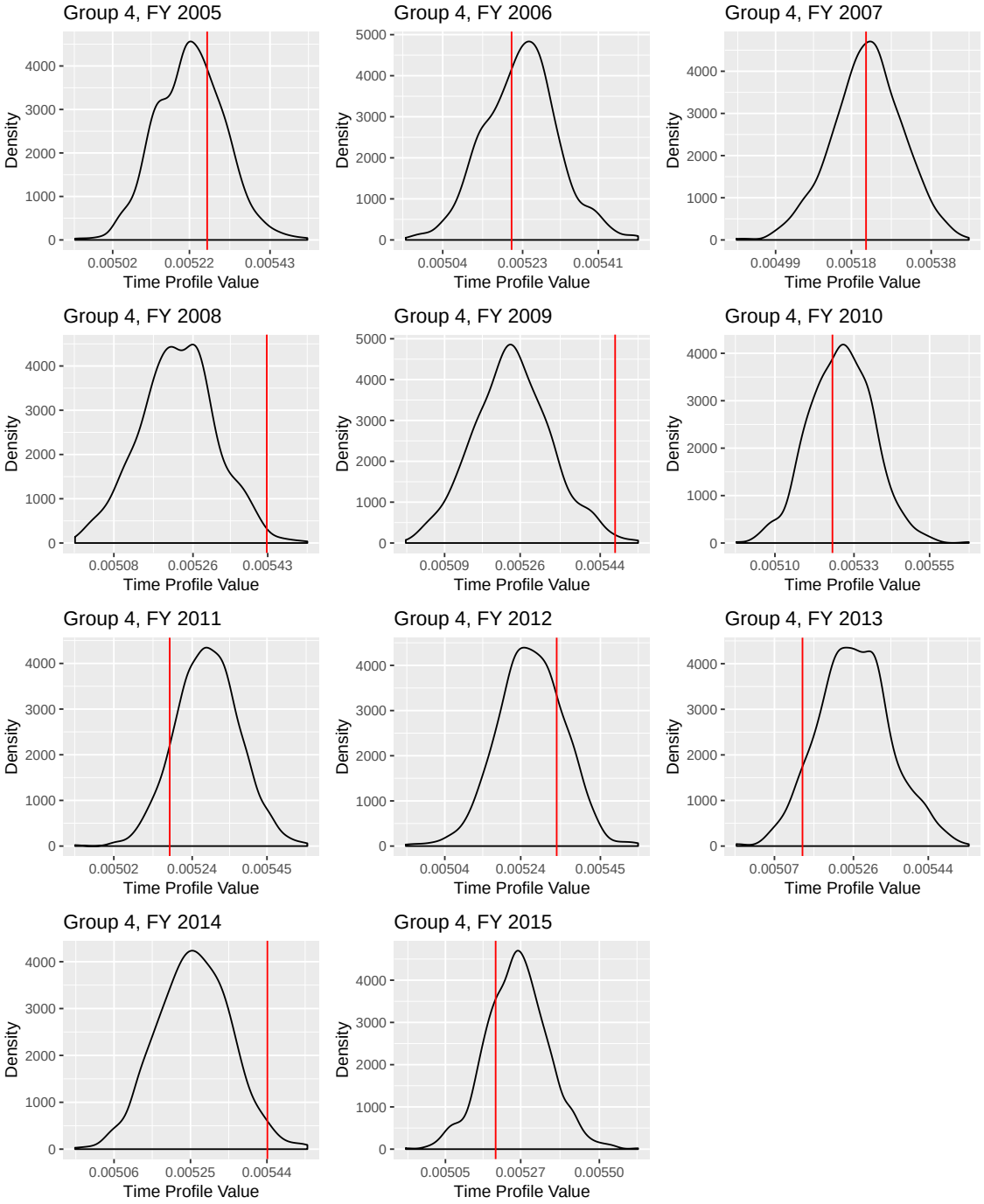


Figure A.22: Group 5 time profile analytical SE distributions across the simulated datasets. The black lines plot the kernel density estimates for standard errors derived from the fixed-T variance estimate formula after estimating the GFE model on 1000 simulated datasets. The red vertical lines represent the value of the simulated standard error.

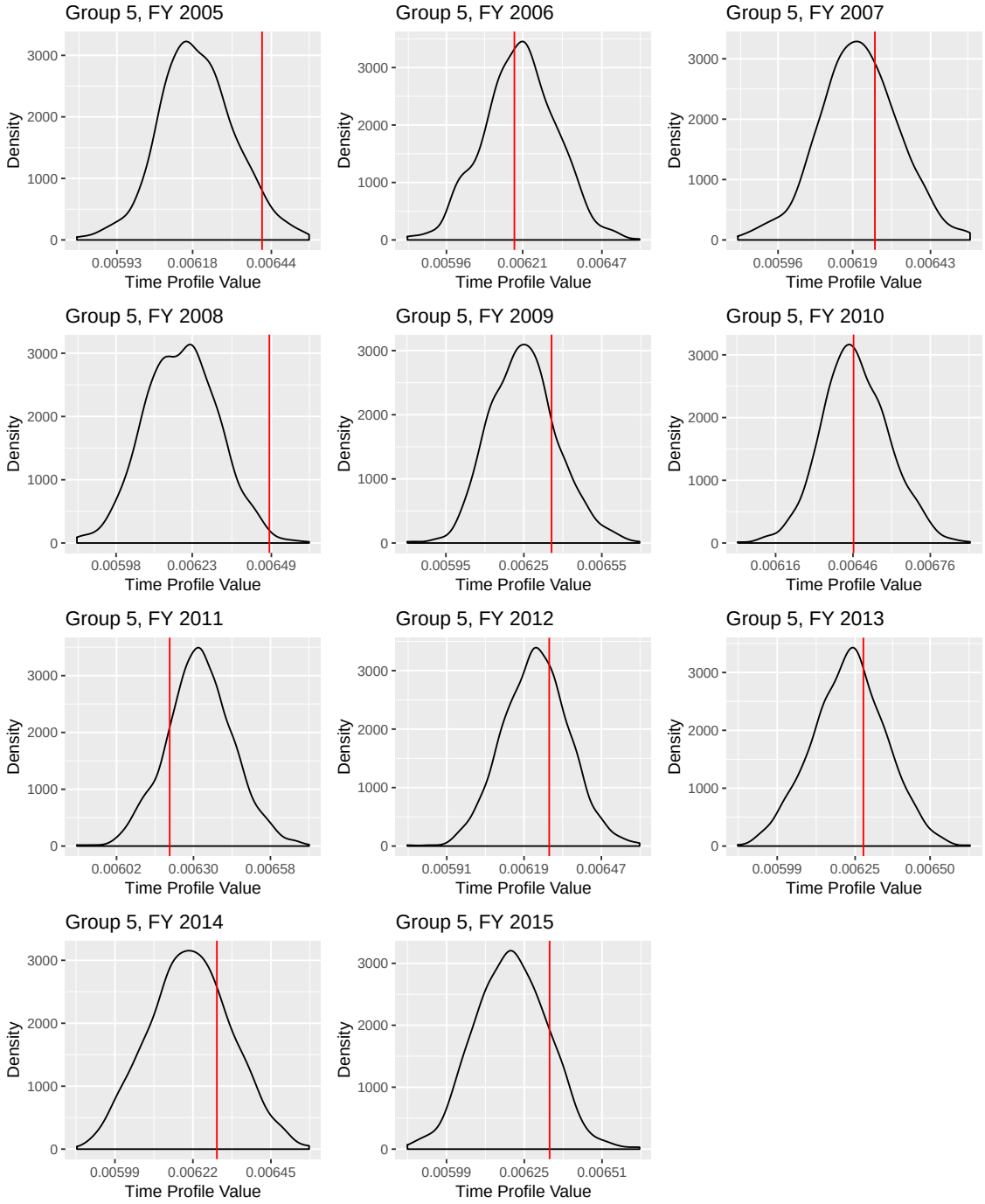


Figure A.23: Group 6 time profile analytical SE distributions across the simulated datasets. The black lines plot the kernel density estimates for standard errors derived from the fixed-T variance estimate formula after estimating the GFE model on 1000 simulated datasets. The red vertical lines represent the value of the simulated standard error.

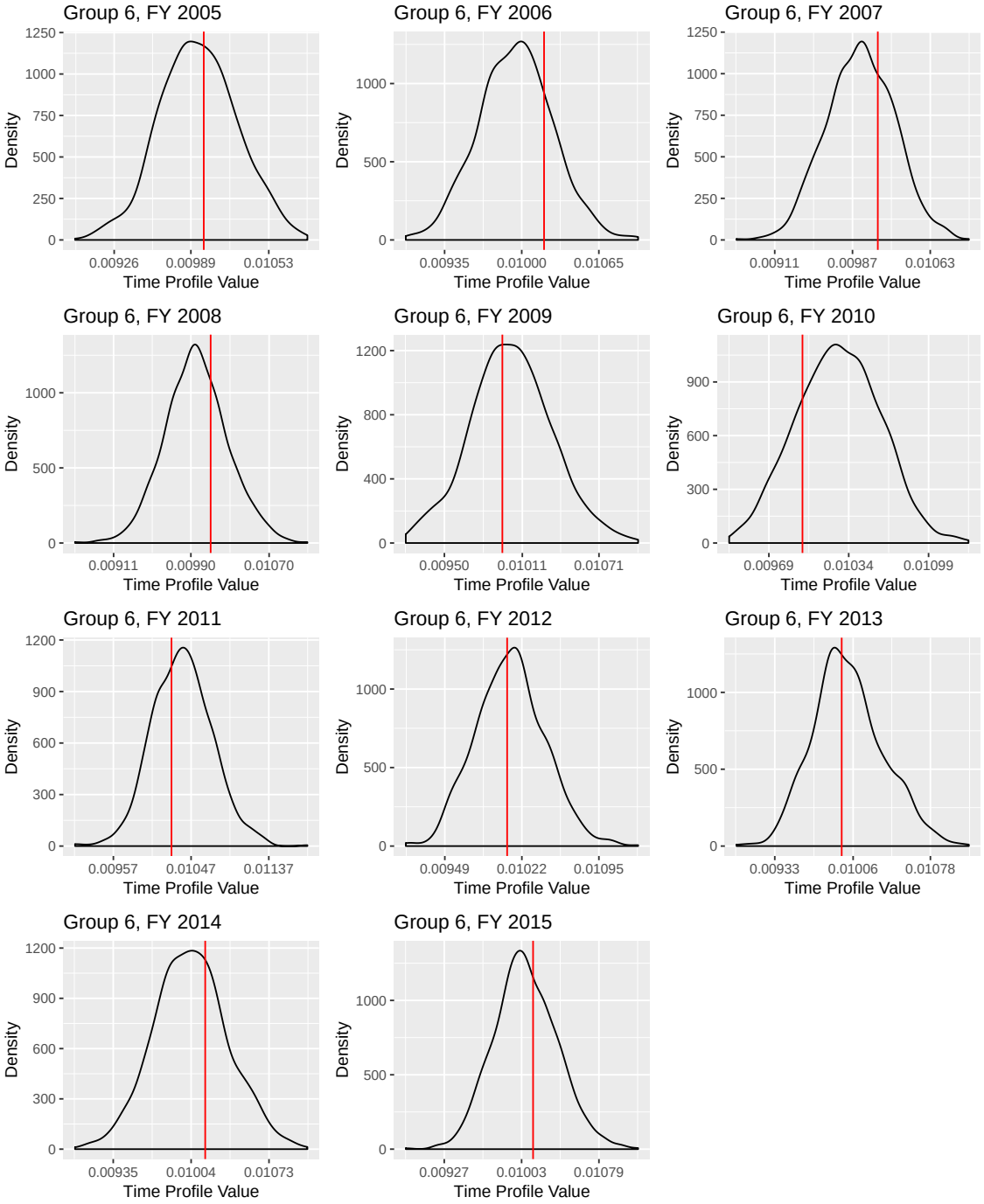


Figure A.24: Group 7 time profile analytical SE distributions across the simulated datasets. The black lines plot the kernel density estimates for standard errors derived from the fixed-T variance estimate formula after estimating the GFE model on 1000 simulated datasets. The red vertical lines represent the value of the simulated standard error.

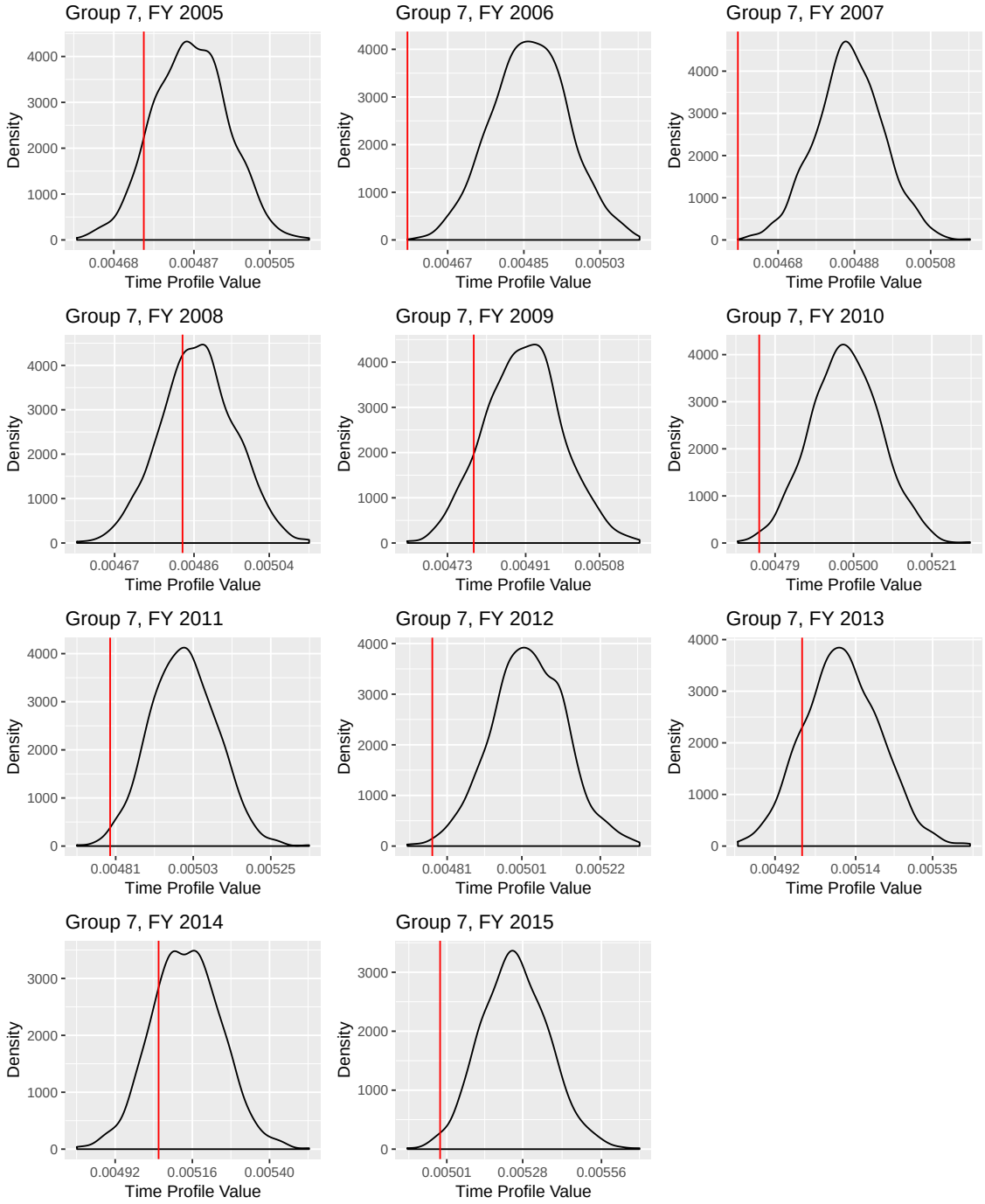


Table A.25: Group 6 summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.1	6377.06	0.11	0.21	13530.6	73469.07
SD	0.08	5853.41	0.31	0.27	32379.36	71795.16
Median	0.08	4800	0	0.09	5000	53776
Q1	0.07	2670	0	0.02	1159.84	30503
Q3	0.11	8060	0	0.31	13318.59	89727
Min	0	1	0	0	1	1
Max	1.57	86688.13	1	0.9	455548	781753.8
Count	5437	5438	5592	606	606	5585

Table A.26: Group 7 summary statistics – time-varying variables.

	Regular Drawdown Rate	Regular Drawdown Amount	Ad-hoc Drawdown Indicator	Ad-hoc Drawdown Rate	Ad-hoc Drawdown Amount	Account Balance
Mean	0.09	6366.74	0.03	0.16	12575.79	73117.21
SD	0.03	6367.71	0.17	0.2	26520.99	76839.21
Median	0.09	4764	0	0.09	6000	53267.5
Q1	0.07	3036	0	0.04	3000	33820
Q3	0.11	7416	0	0.18	12000	87103.25
Min	0.02	12	0	0	31	13
Max	0.92	166695	1	0.9	520802	2427083
Count	31156	31168	31222	891	891	31210

Figure A.25: Group 1 histograms – time-invariant variables.

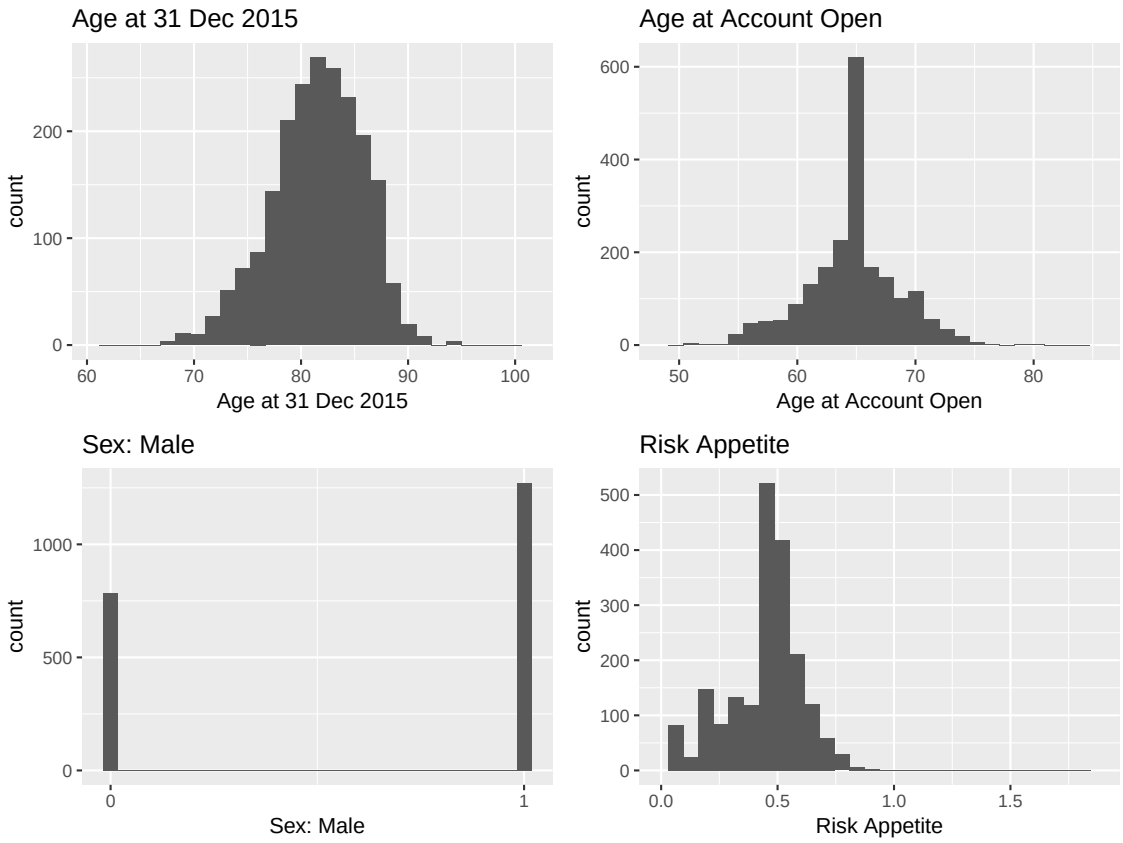


Figure A.26: Group 2 histograms – time-invariant variables.

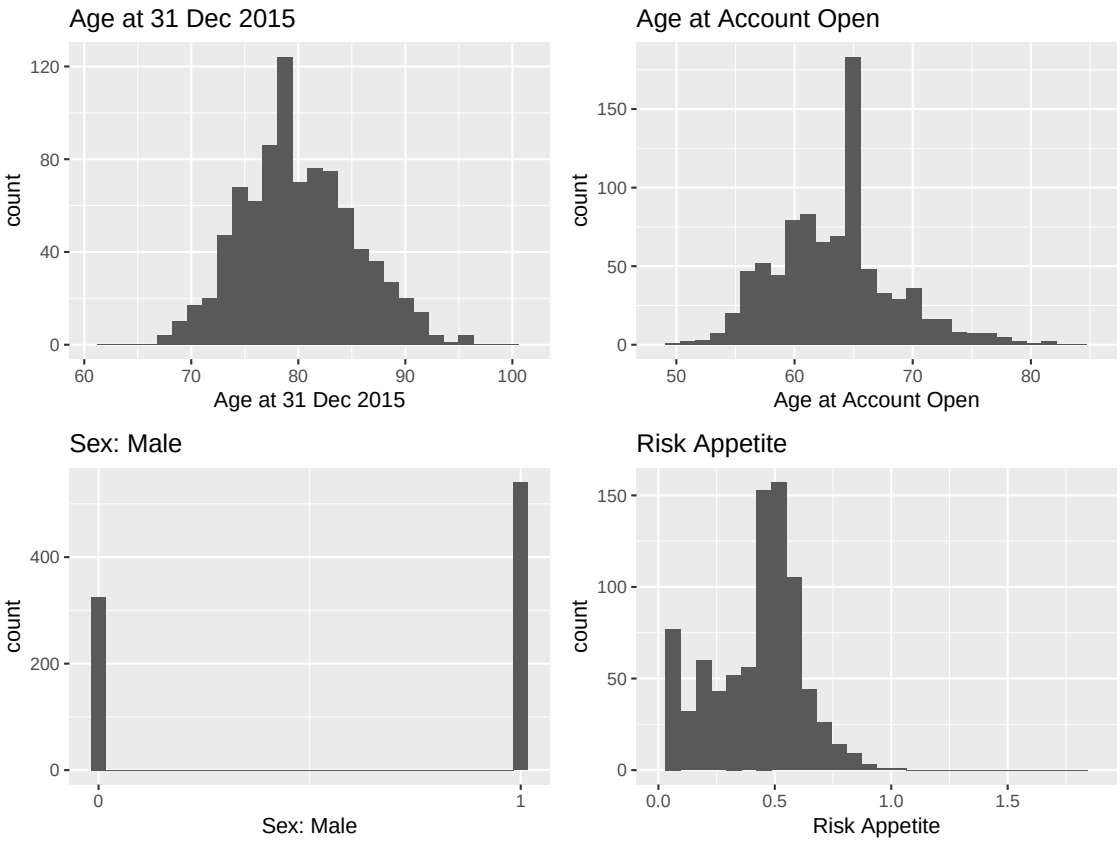


Figure A.27: Group 3 histograms – time-invariant variables.

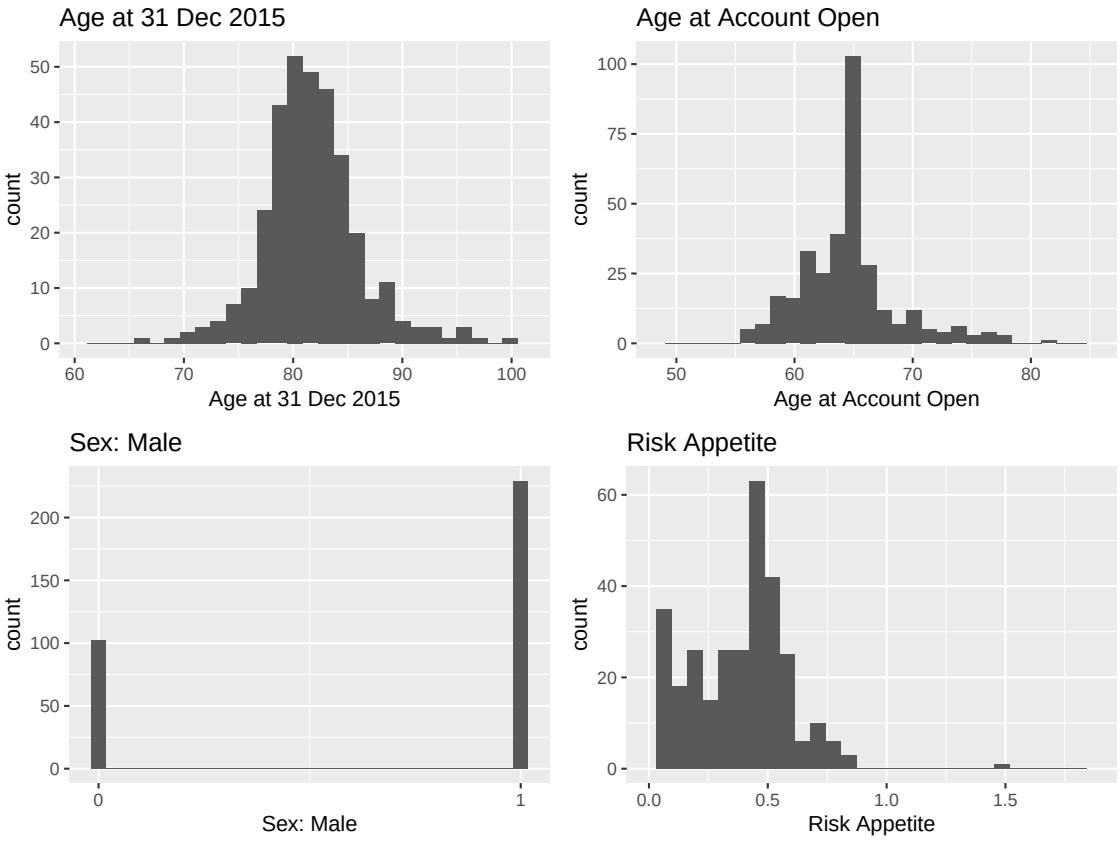


Figure A.28: Group 4 histograms – time-invariant variables.

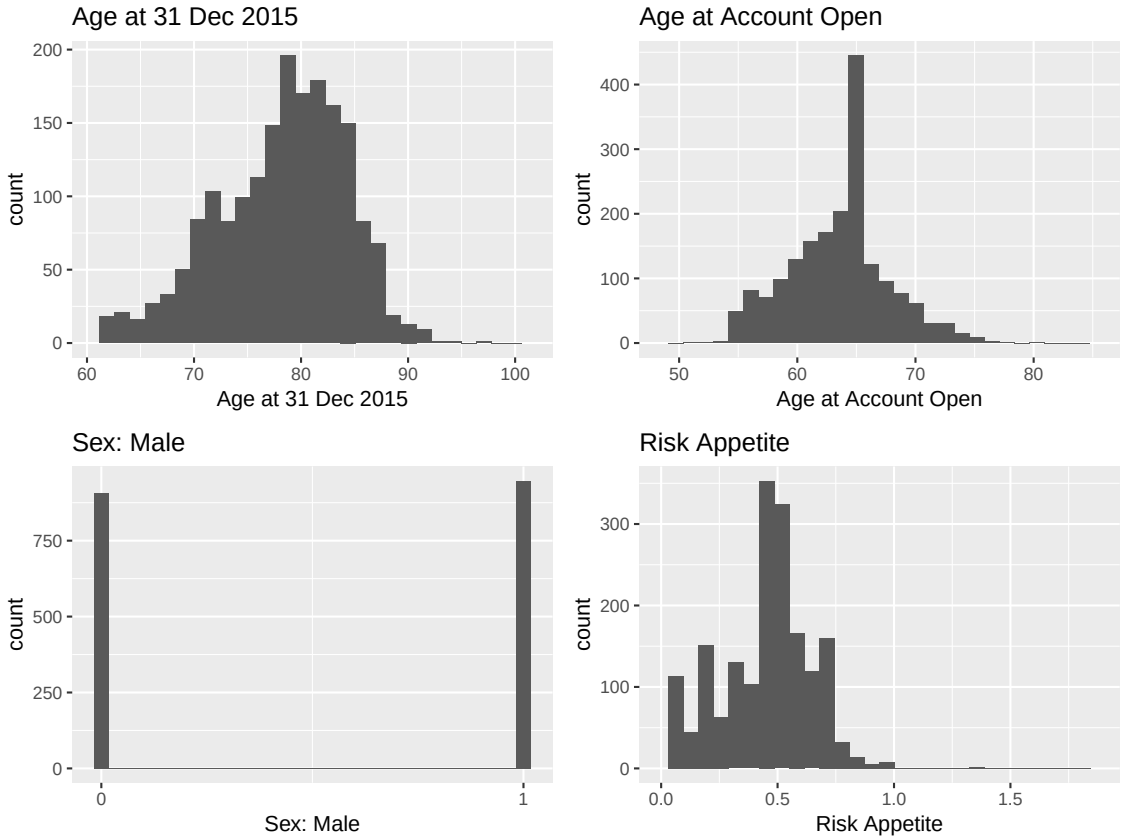


Figure A.29: Group 5 histograms – time-invariant variables.

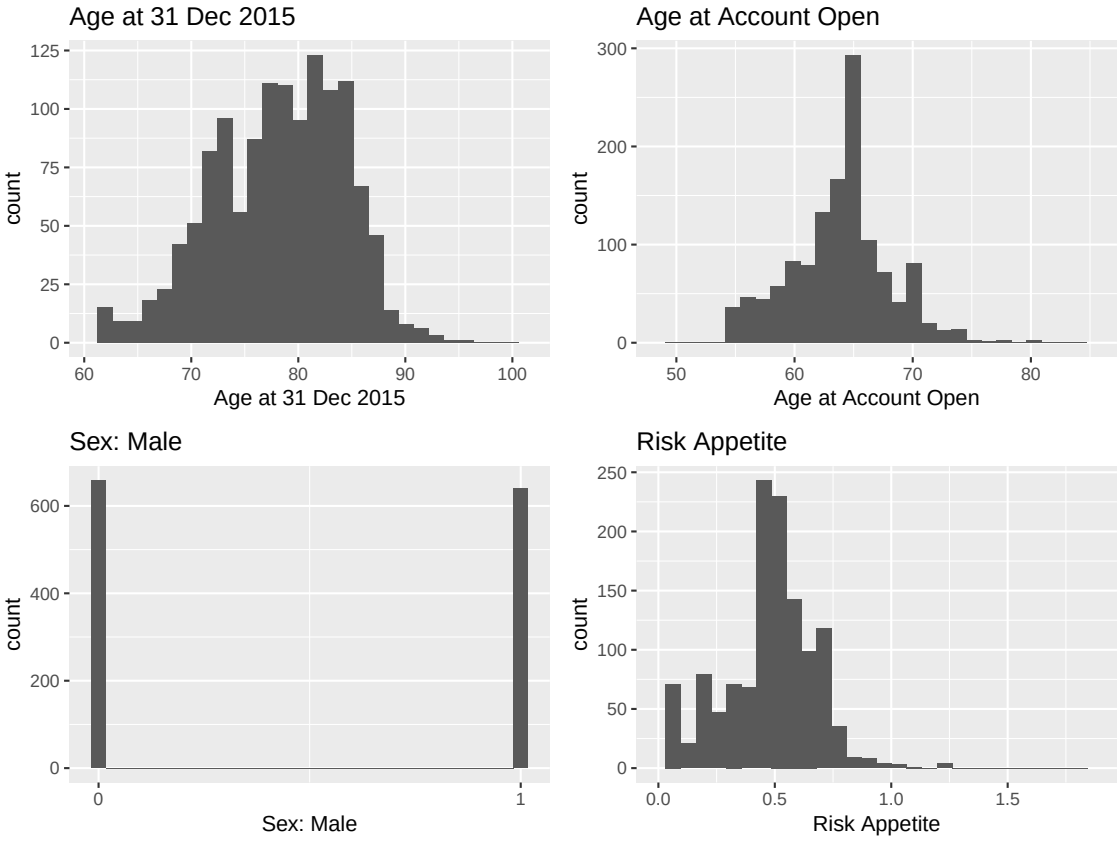


Figure A.30: Group 6 histograms – time-invariant variables.

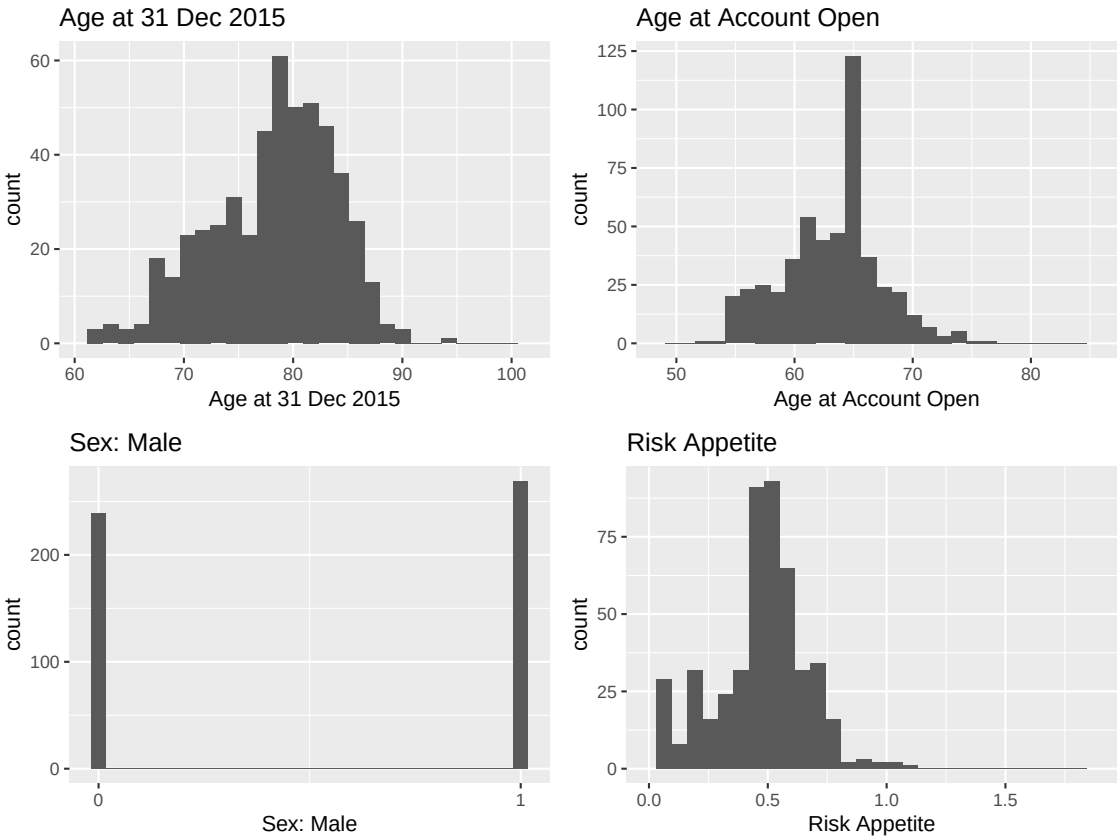


Figure A.31: Group 7 histograms – time-invariant variables.

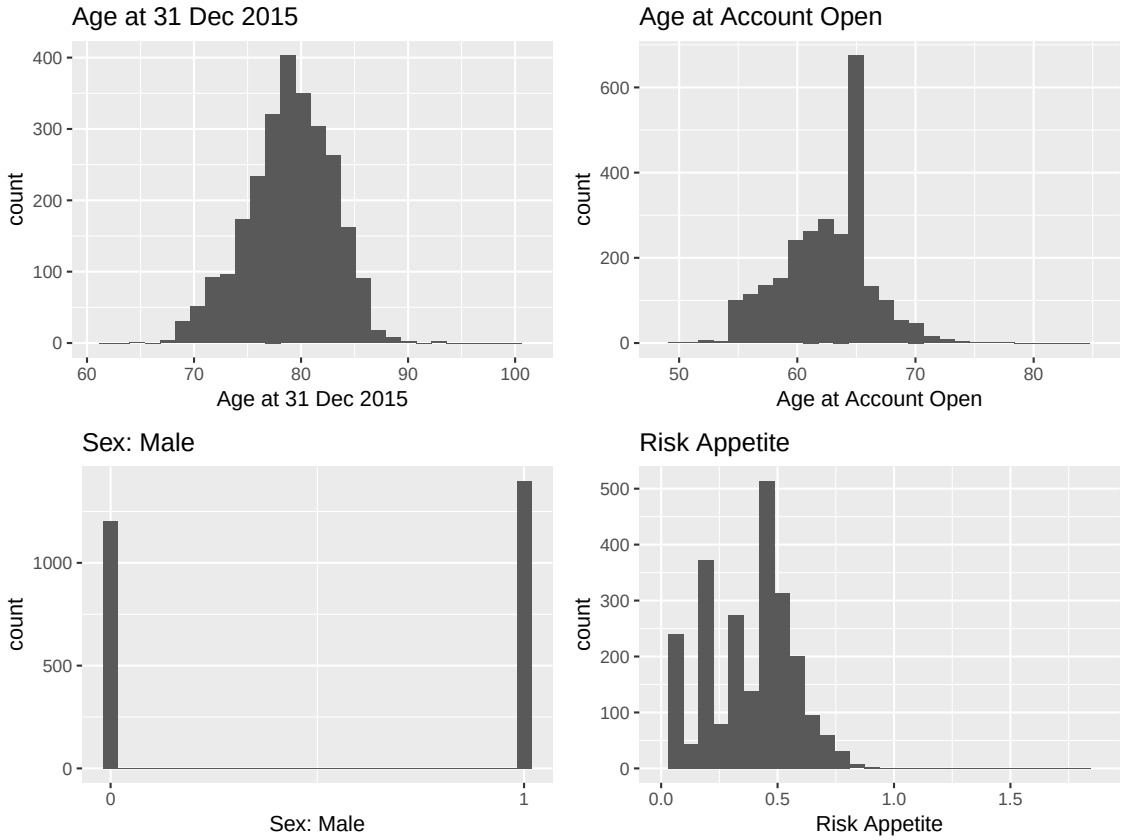


Figure A.32: Group 1 histograms – time-varying variables.

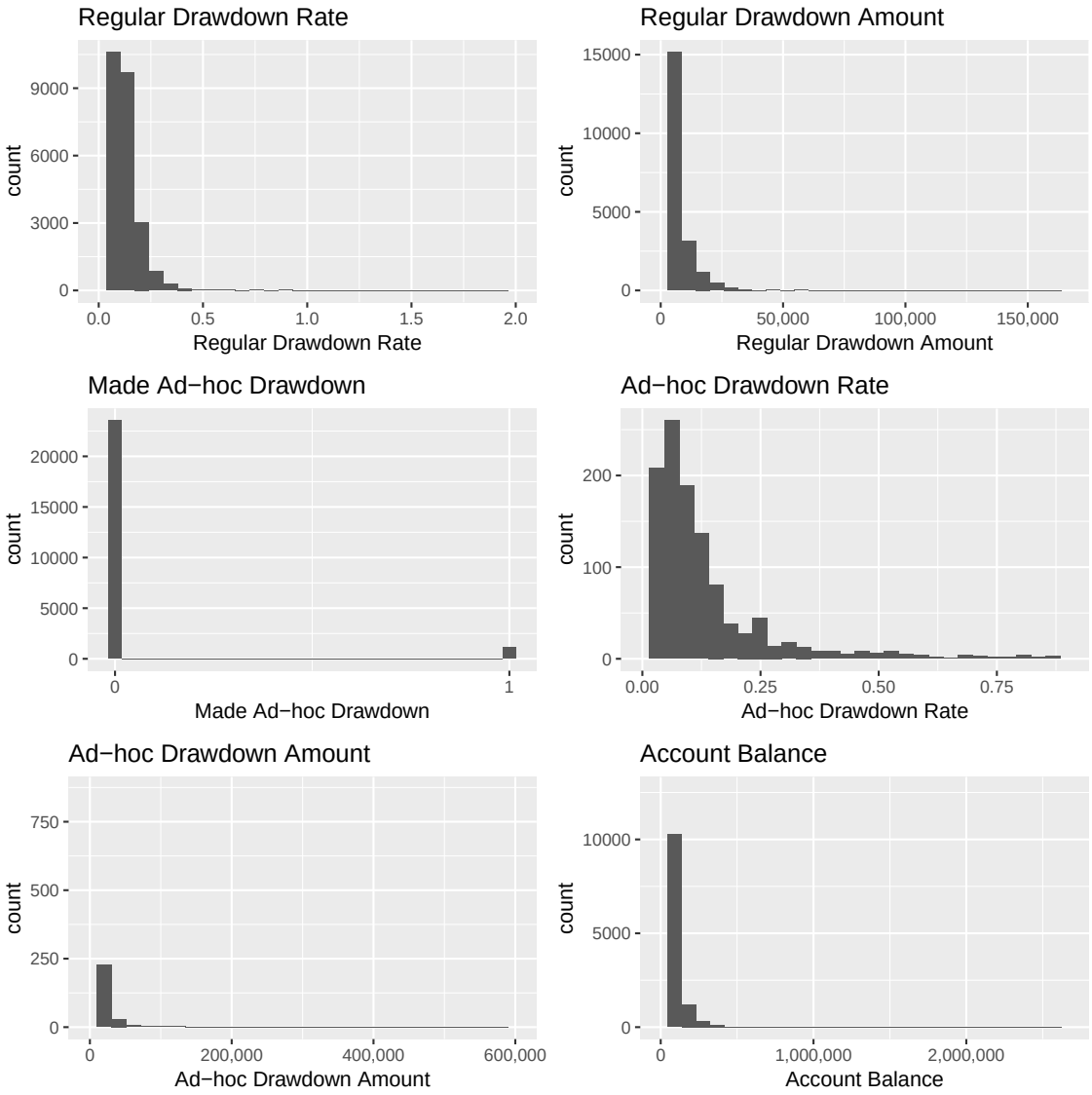


Figure A.33: Group 2 histograms – time-varying variables.

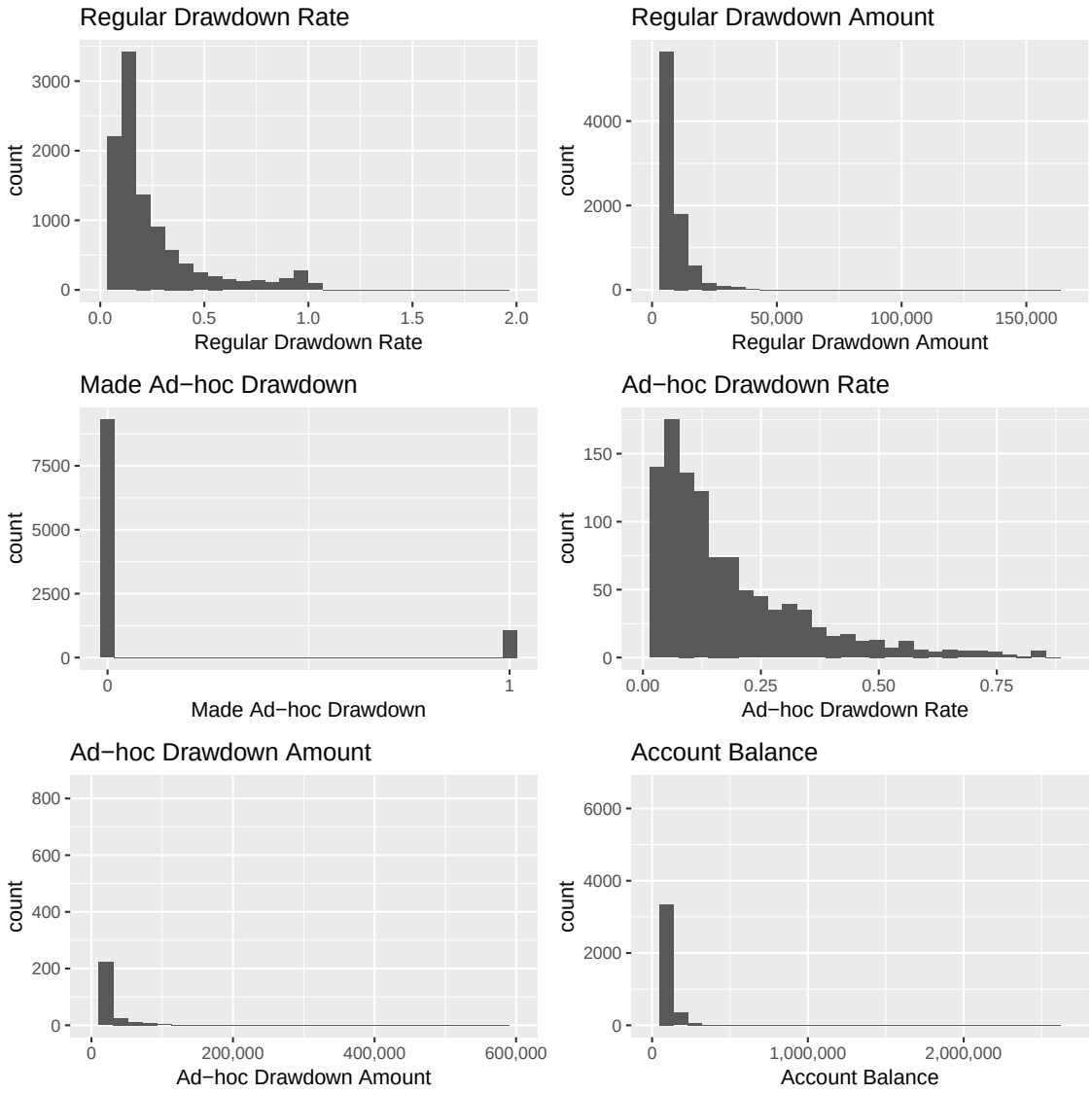


Figure A.34: Group 3 histograms – time-varying variables.

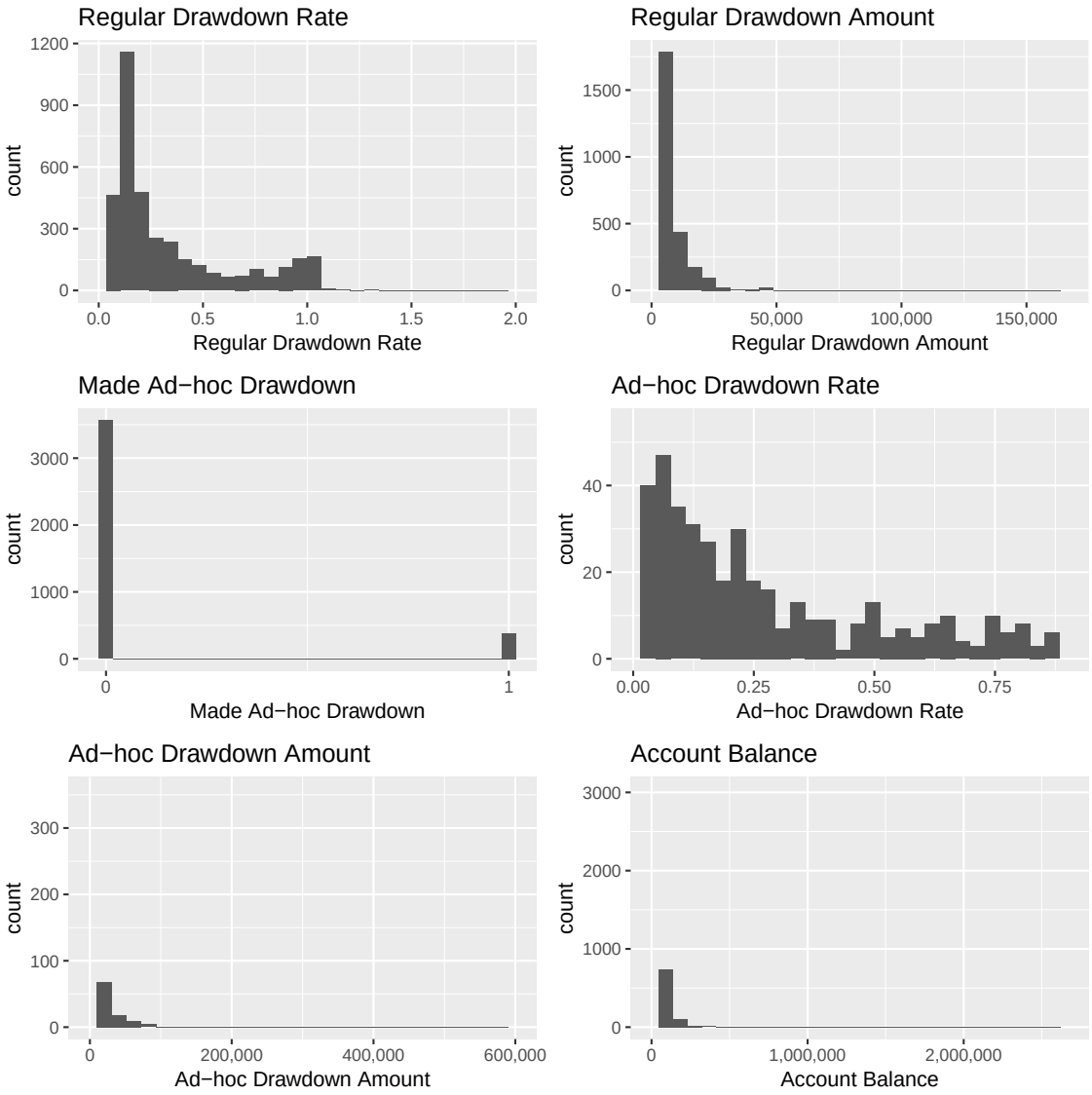


Figure A.35: Group 4 histograms – time-varying variables.

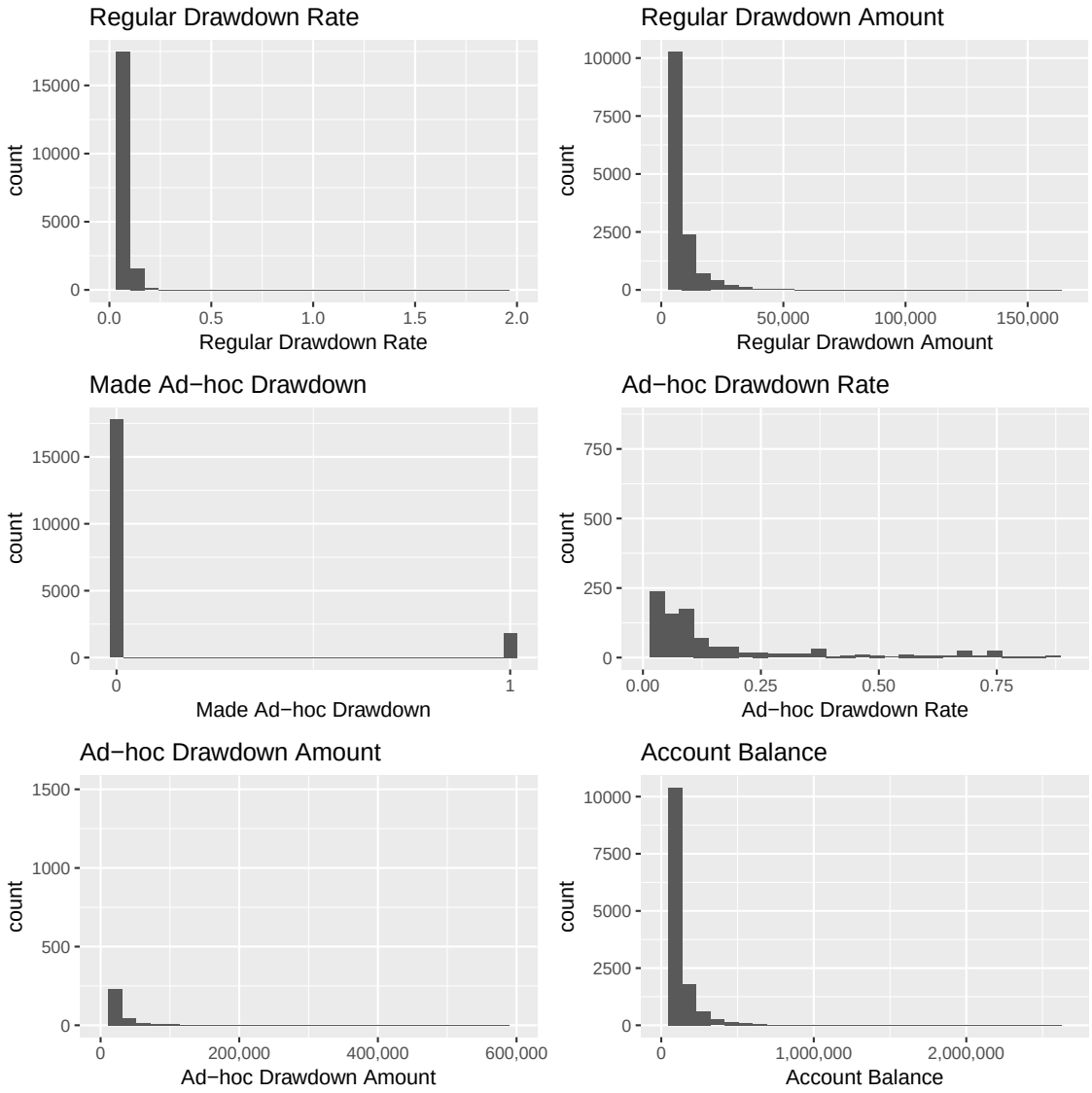


Figure A.36: Group 5 histograms – time-varying variables.

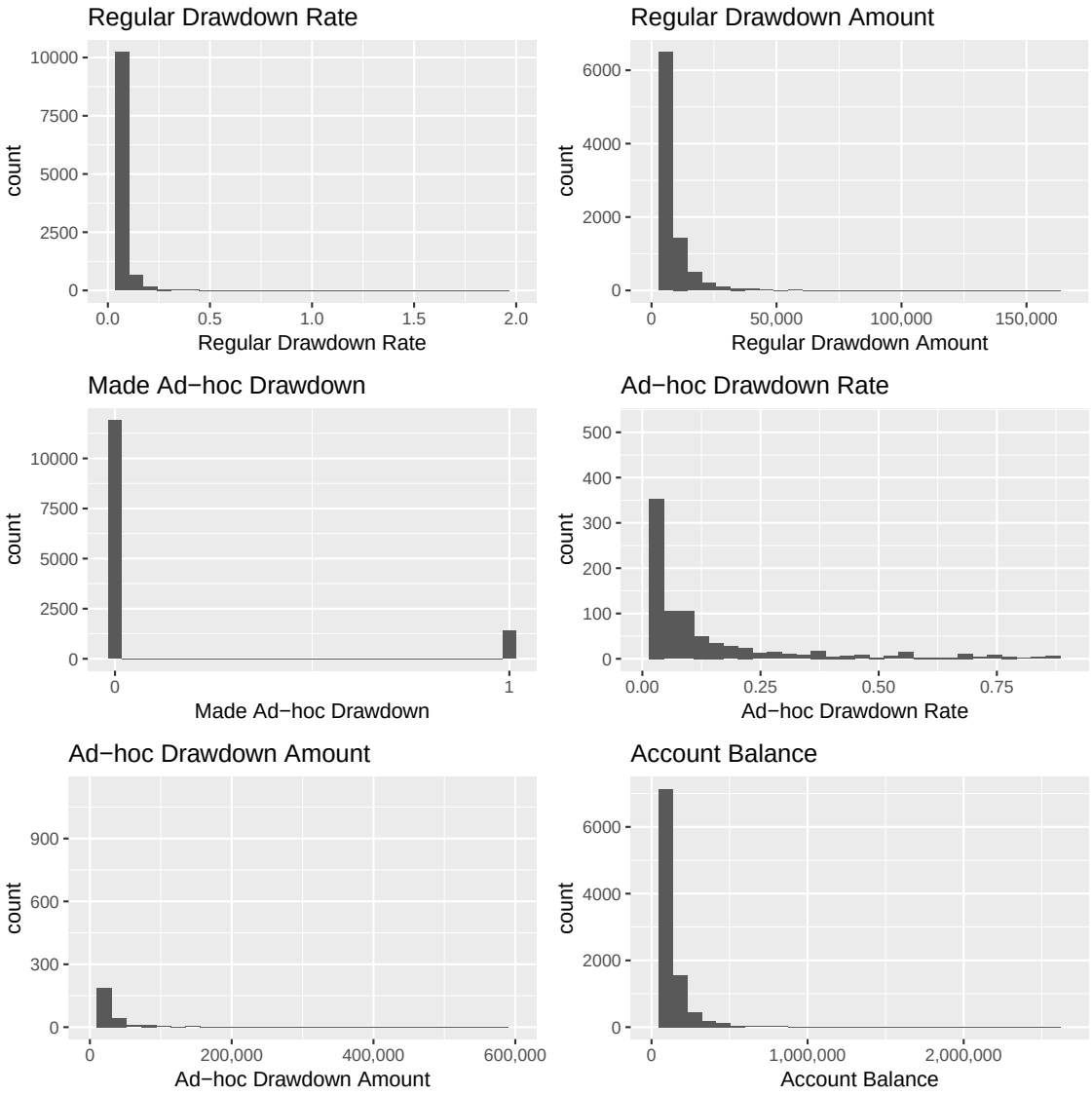


Figure A.37: Group 6 histograms – time-varying variables.

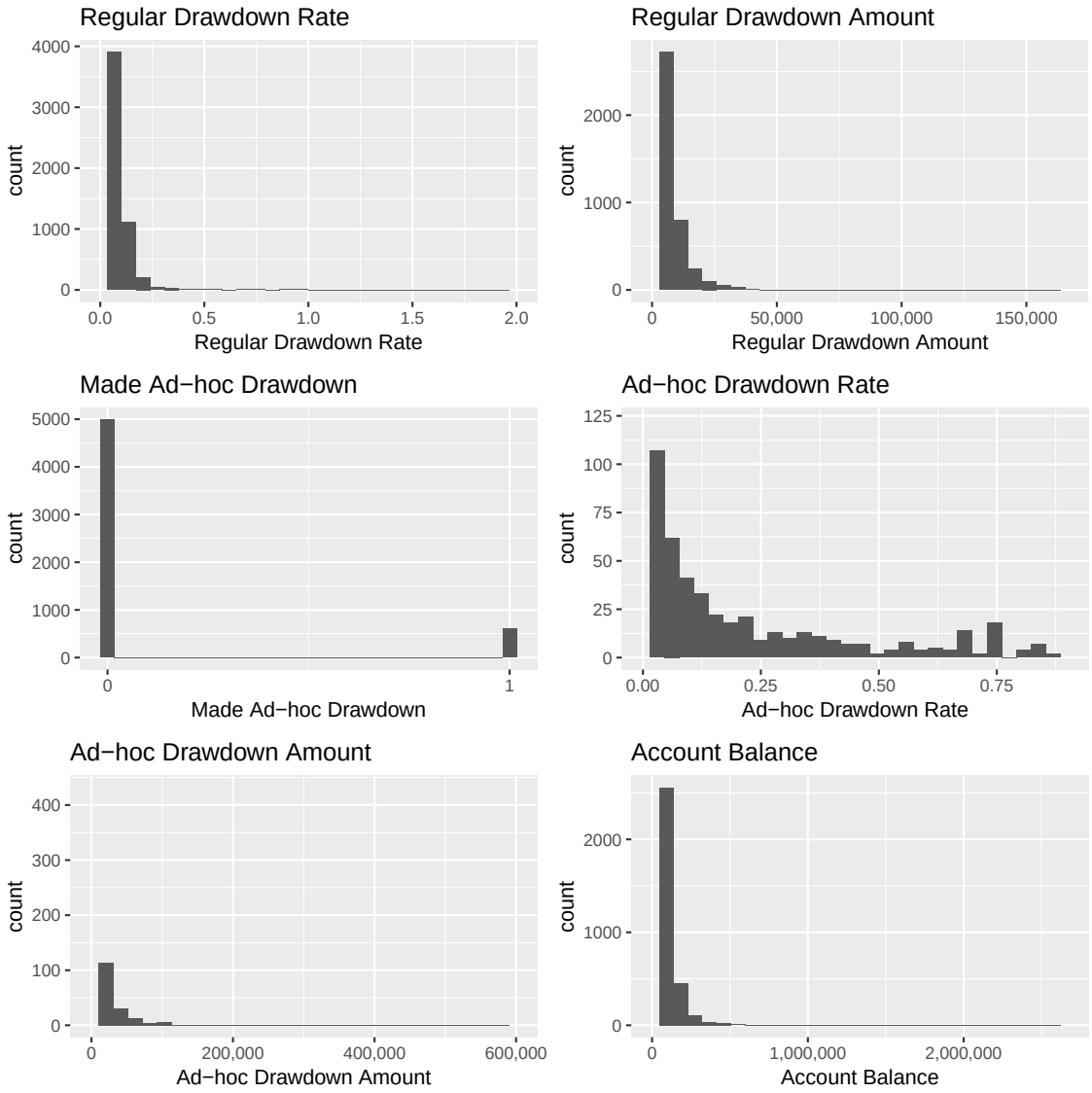


Figure A.38: Group 7 histograms – time-varying variables.

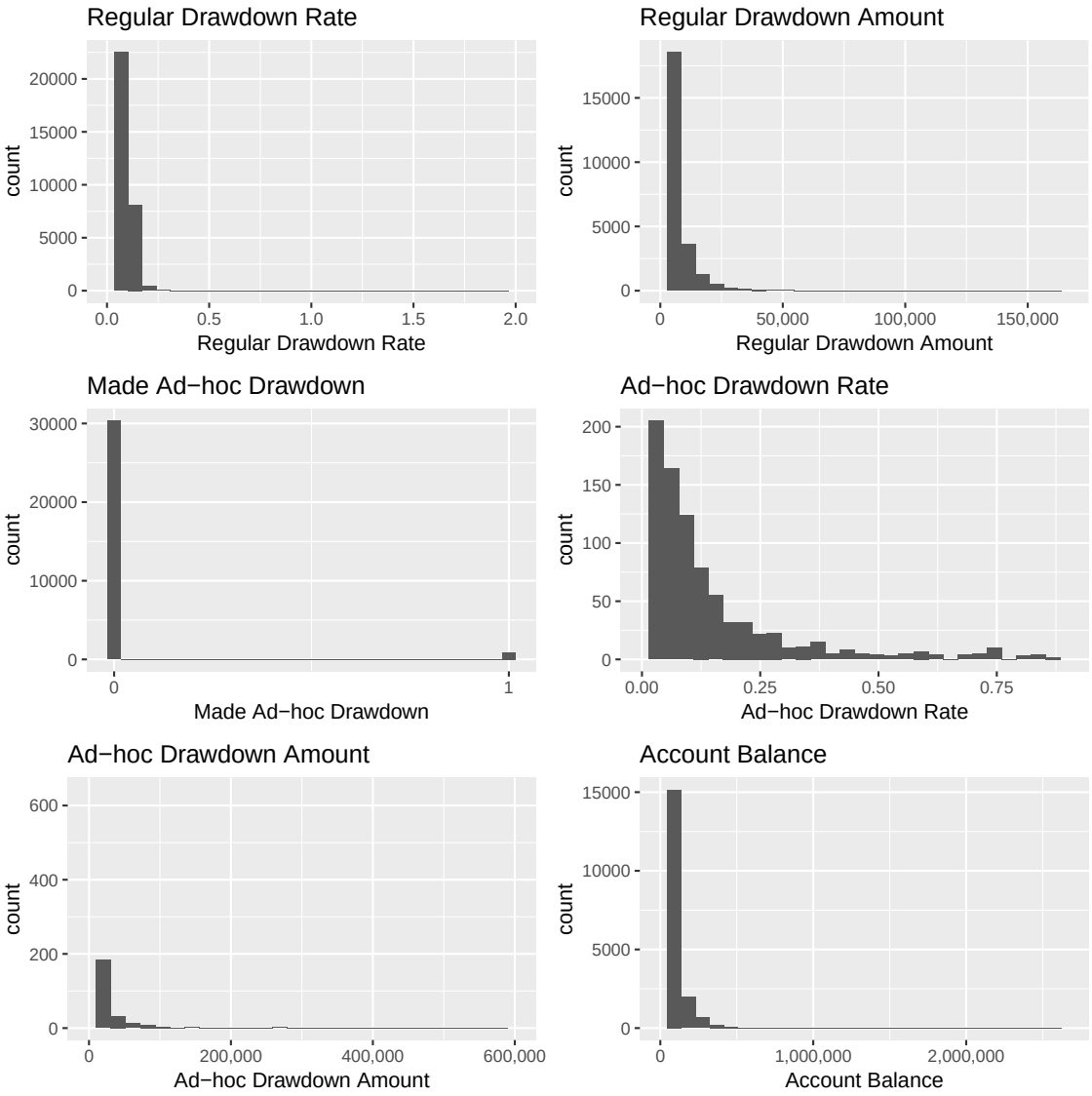


Figure A.39: $G = 7$ model – group 1 time-demeaned (TD) panel plots. The account balances are as at financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.

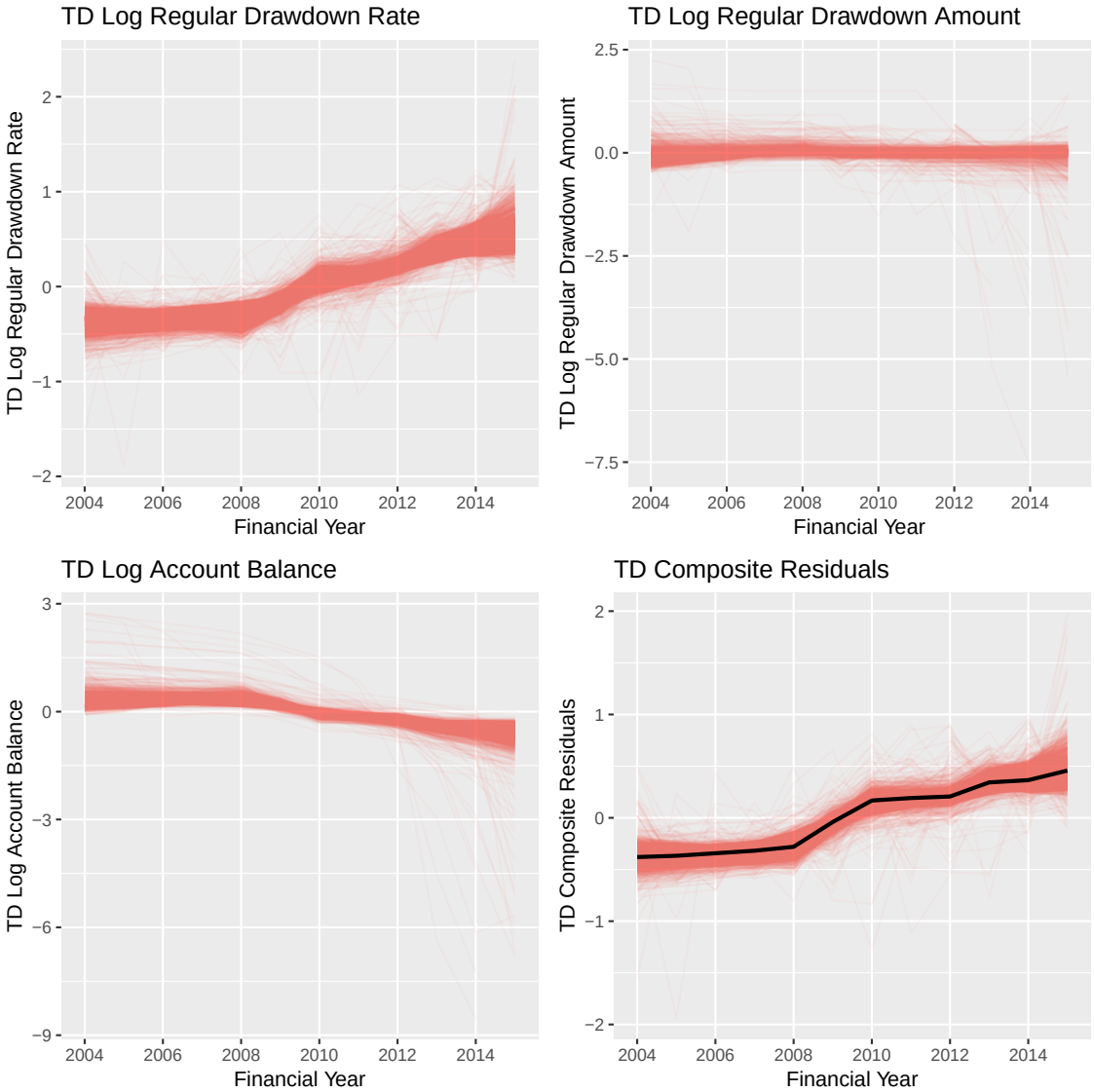


Figure A.40: $G = 7$ model – group 2 time-demeaned (TD) panel plots. The account balances are as at financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.

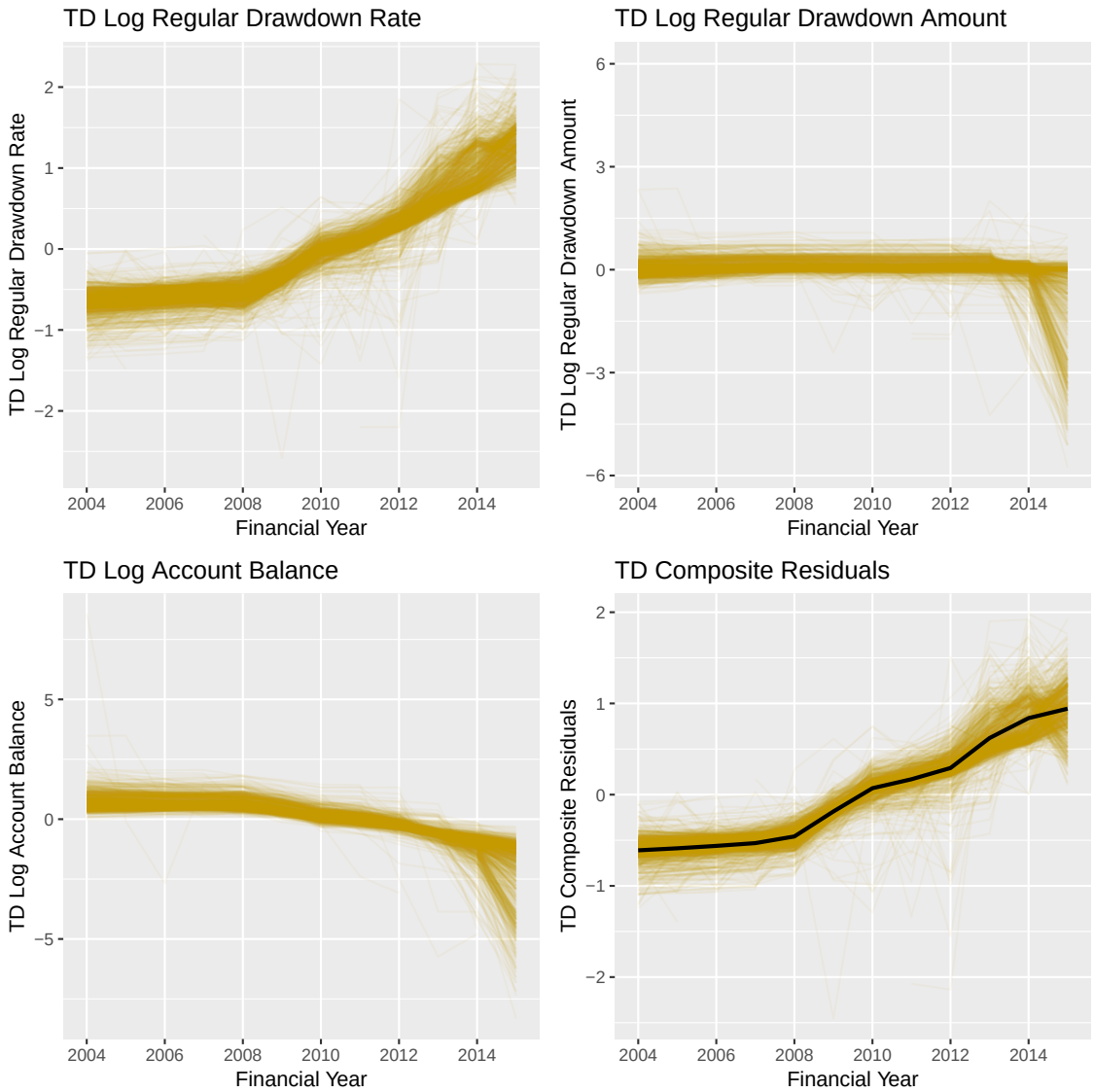


Figure A.41: $G = 7$ model – group 3 time-demeaned (TD) panel plots. The account balances are as at financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.

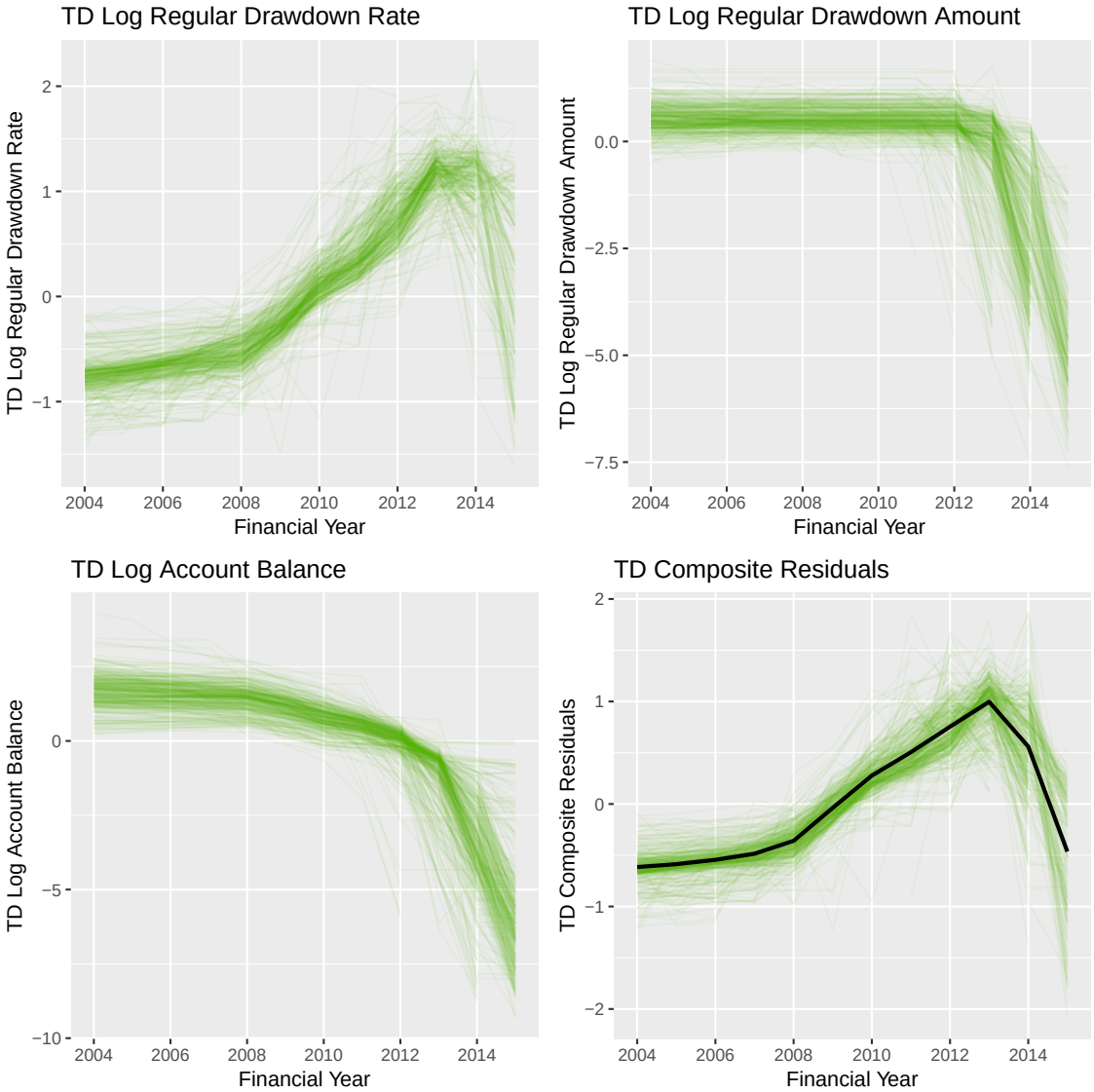


Figure A.42: $G = 7$ model – group 4 time-demeaned (TD) panel plots. The account balances are as at financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.

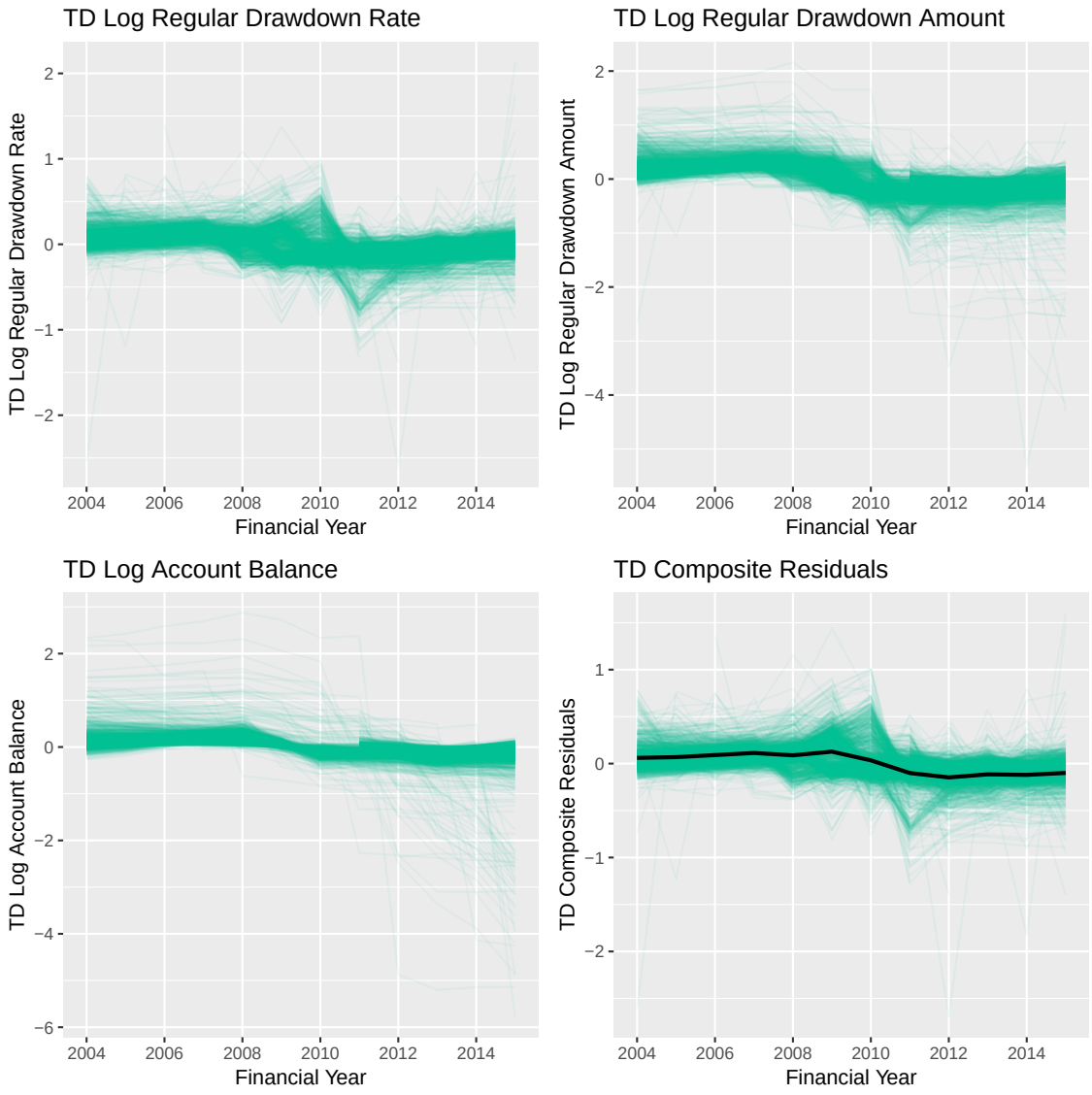


Figure A.43: $G = 7$ model – group 5 time-demeaned (TD) panel plots. The account balances are as at financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.

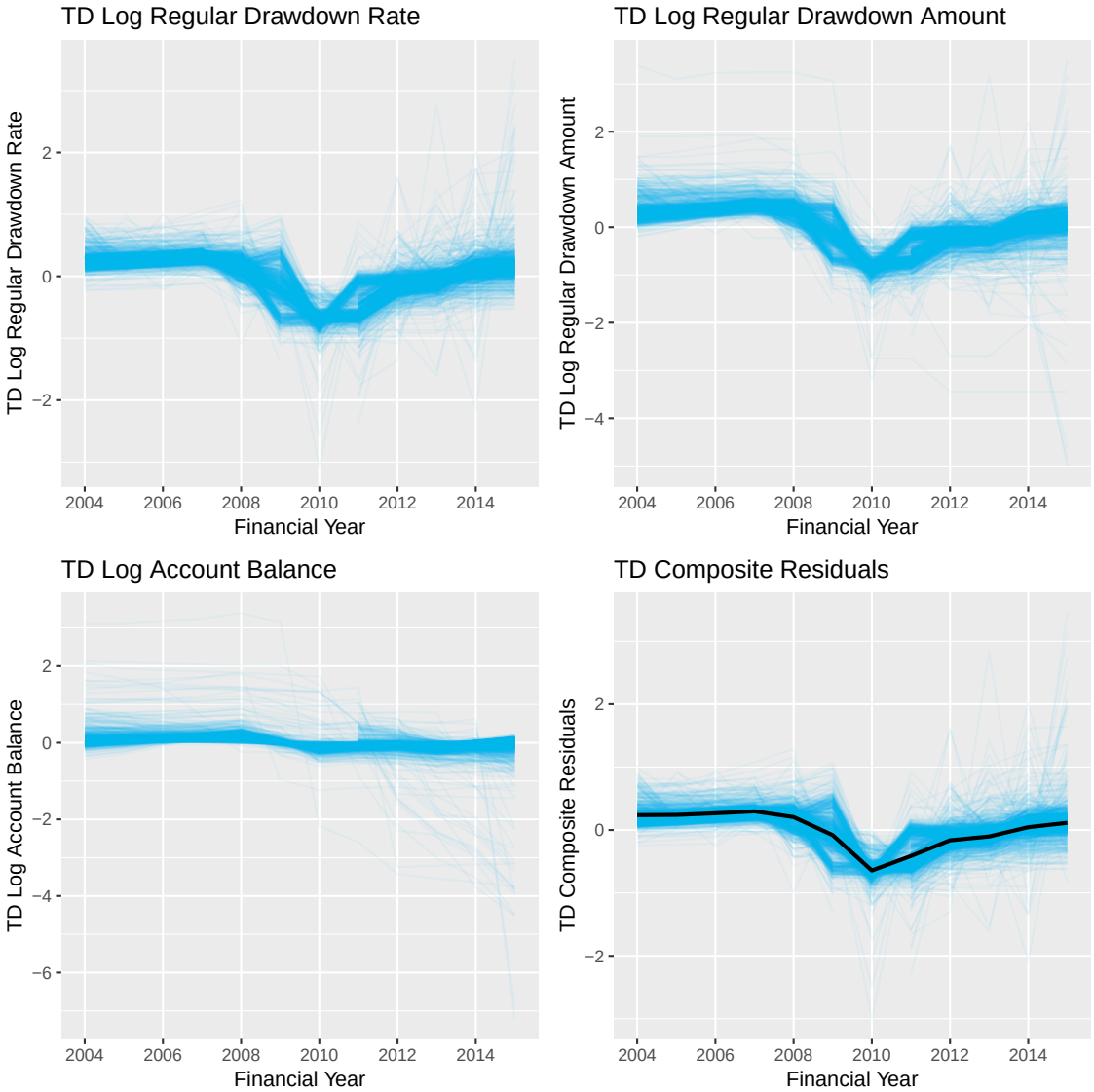


Figure A.44: $G = 7$ model – group 6 time-demeaned (TD) panel plots. The account balances are as at financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.

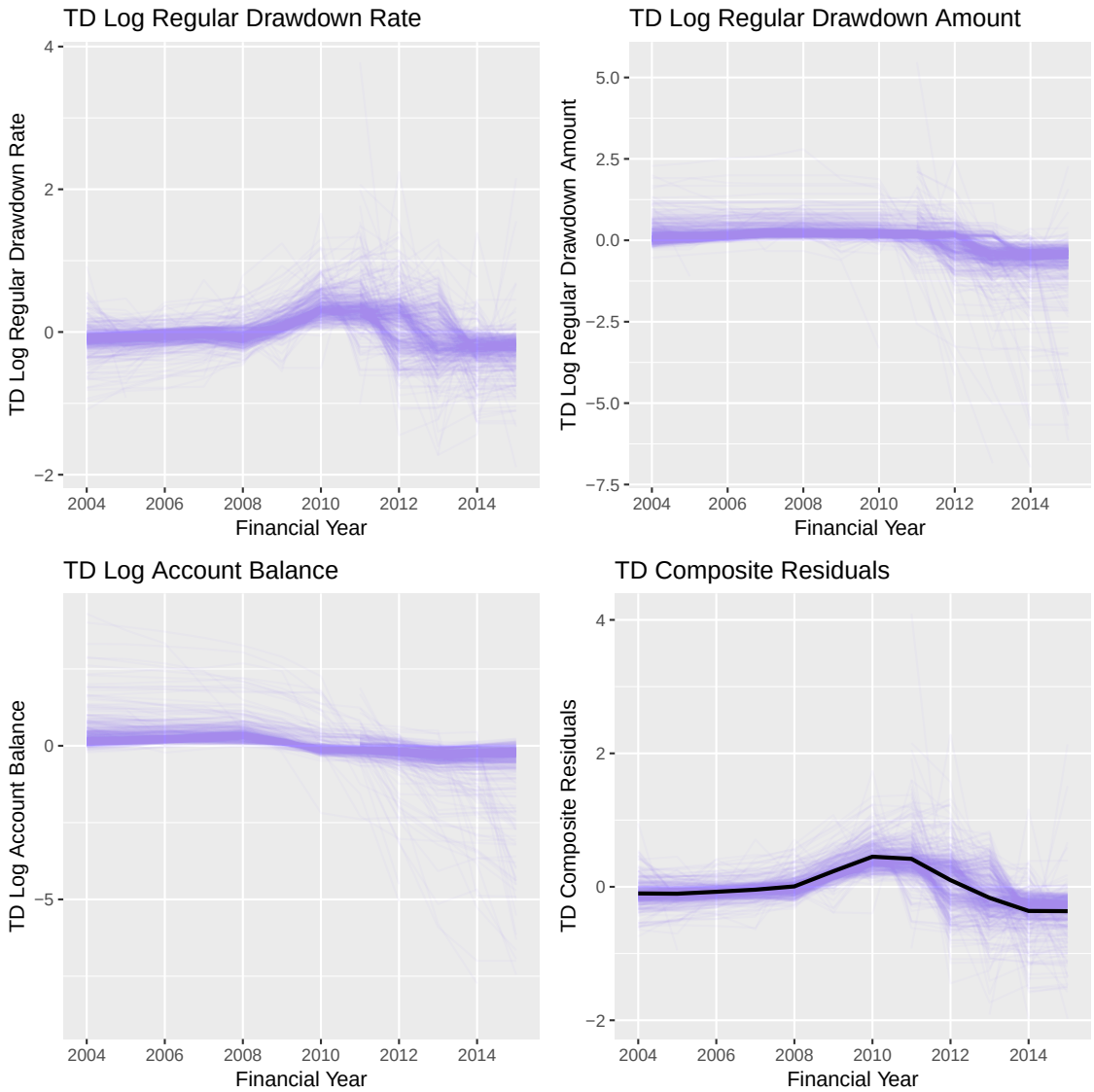
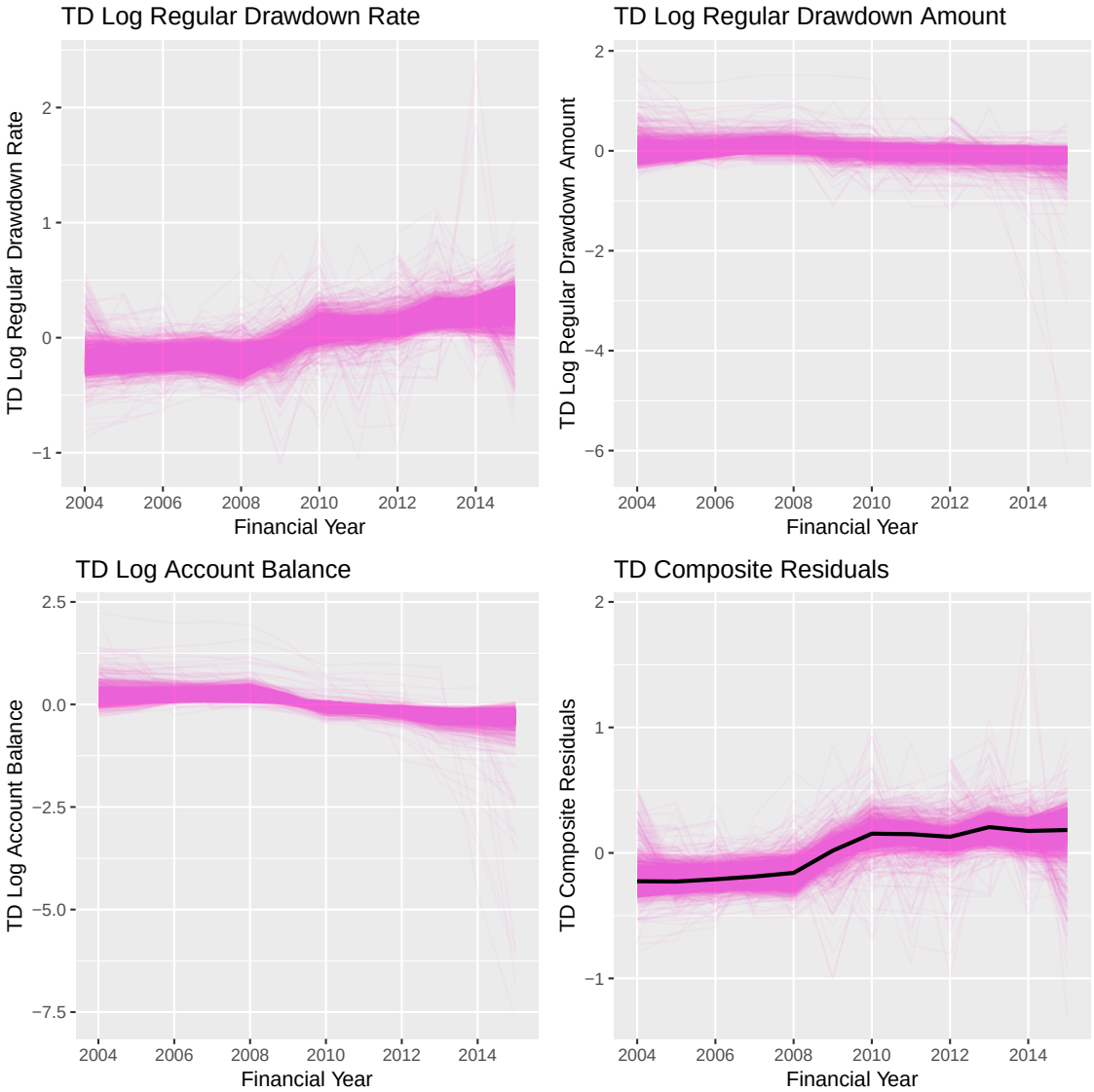


Figure A.45: $G = 7$ model – group 7 time-demeaned (TD) panel plots. The account balances are as at financial year start. The black series in the bottom-right panel represents the estimated time-demeaned group time profile values.



Appendix B

Supplement for *Nonparametric Bayesian Estimation of Latent Group Time Profiles in Linear Panel Data Models with Fixed Effects*

Abstract

This supporting material provides additional results for the simulation study and empirical application reported in Chapter 4.

B.1 Additional simulation results

Figure B.1 shows the simulation results for the two alternative time profile estimators considered, the simple averaging and weighted averaging approaches. For all three metrics, both of these alternatives perform noticeably worse than the exemplar unit approach.

Tables B.1, B.2 and B.3 compare the estimated bias magnitude, RMSE, and 95% confidence/credible intervals, respectively, for the Oracle, BFG and GFE estimators of the covariate effects, $\theta = (\theta_1, \theta_2, \theta_3)'$, across replicates. BFG and GFE perform similarly across these three metrics.

Figures B.2, B.3 and B.4 summarise the Oracle time profile estimator bias, RMSE, and 95% confidence interval coverage probabilities, respectively. The heatmap scales match the corresponding figures for BFG and GFE in the chapter.

Figures B.5, B.6 and B.7 corroborate the finding, based on the summary results in the chapter, that the BFG time profile estimators outperform the GFE estimators, by comparing the BFG, GFE and Oracle time profile estimates for three randomly selected replicates.

B.2 Additional results from the empirical application

Figures B.8 and B.9 give traceplots of the $M = 1000$ thinned, post-burn iterates for the covariate effect parameters, θ_1 and θ_2, respectively, from the best $(0, -1)$ configuration run in the superannuation application presented in the chapter. These plots are representative of the iterate behaviour across all four of the $(0, -1)$ configuration runs. Figure B.10 presents the largest 17 time profiles for the $(1, 0)$ configuration run, which collectively cover 80% of the sample; Figures B.11 and B.12 give the traceplots for θ_1 and θ_2 in this run.

B.3 Correspondence between the BFG and GFE groupings

The estimation run reported in the chapter finds 27 groups using the PEAR method; however, many of these are small groups or singletons: 83.3% of the sample belongs to 6 groups, while there are 13 groups containing fewer than 10 individuals, 6 of which are singletons. In Chapter 3, the GFE results suggest that 7 groups adequately partition the sample based on the time-varying heterogeneity. The following steps allow the comparison of these two distinct clusterings.

1. The GFE group identities for the $N = 1000$ units found in the BFG sample—which is a random balanced subsample of the original $N = 9516$ panel used for the GFE application—are matched to one of three 'supergroups' defined by the economic interpretation of the expressed behaviour: 1) following the minimum drawdown rates; 2) drawing constant dollar amounts over time, potentially with a downwards revision in the level; 3) any behaviour that does not fit the previous two classifications.

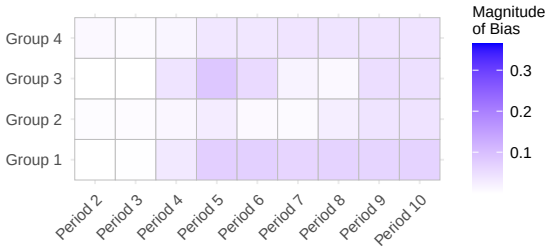
2. The BFG group identities are similarly aggregated into these three supergroups.

3. The correspondence between the two sets of supergroup allocations is computed as the proportion of the $N = 1000$ sample that is classified in the same way by both procedures.

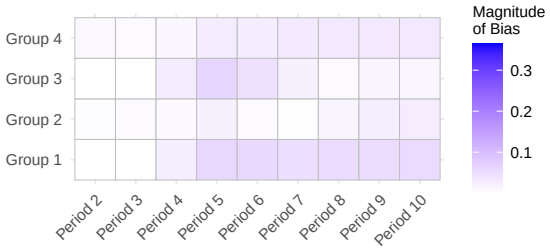
This method gives a correspondence rating of 77.1%, which is considerably higher than the 33.3% that is expected by random chance, but still suggests a meaningfully different allocation of the behaviours. An advantage of the Bayesian procedure in this context is its tendency to estimate a few 'core' groups, and separate outlier units into smaller clusters or singletons; by contrast, the classical method must leave outliers inside the core groups, which partly explains the correspondence result. Furthermore, the simulation results in the chapter show that BFG is able to achieve considerably higher group allocation accuracy than GFE on the same data.

Figure B.1: BFG simple average (left column) and weighted average (right column) time profile estimated bias (top row), RMSE (middle row) and 95% credible interval coverage probabilities (last row). The time profile estimators are identified relative to period 1. For the last row, the heatmap is white at the nominal coverage probability of 0.95.

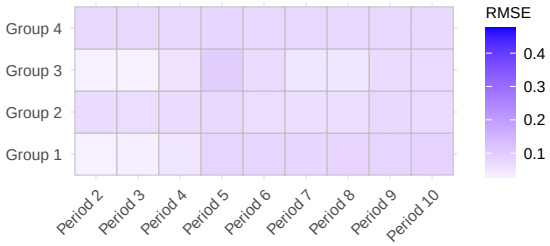
(a) Simple average bias.



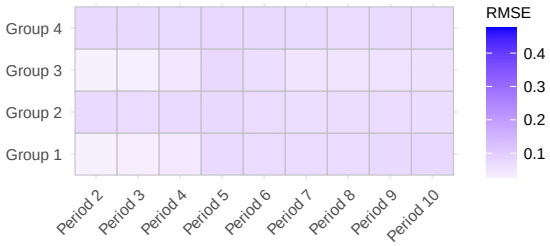
(b) Weighted average bias.



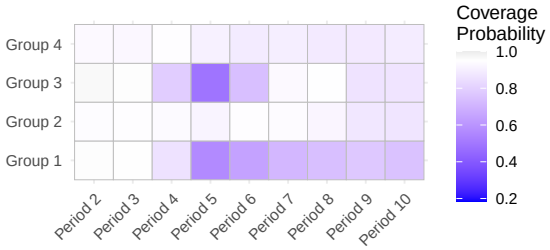
(c) Simple average RMSE.



(d) Weighted average RMSE.



(e) Simple average coverage probability.



(f) Weighted average coverage probability.

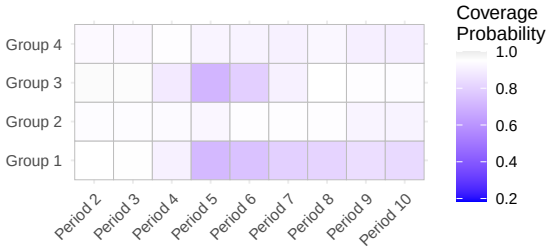


Table B.1: Covariate effect estimated bias magnitude across replicates.

	θ_1	θ_2	θ_3
Oracle	0.18×10^{-3}	0.04×10^{-3}	0.16×10^{-3}
BFG	3.96×10^{-3}	3.92×10^{-3}	0.09×10^{-3}
GFE	5.70×10^{-3}	5.95×10^{-3}	0.09×10^{-3}

Table B.2: Covariate effect estimated RMSE across replicates.

	θ_1	θ_2	θ_3
Oracle	6.3×10^{-3}	6.8×10^{-3}	6.5×10^{-3}
BFG	6.9×10^{-3}	7.1×10^{-3}	5.8×10^{-3}
GFE	8.6×10^{-3}	9.6×10^{-3}	6.9×10^{-3}

Table B.3: Covariate effect estimated 95% credible/confidence interval coverage probabilities across replicates.

	θ_1	θ_2	θ_3
Oracle	0.97	0.94	0.96
BFG	0.89	0.88	0.95
GFE	0.89	0.85	0.94

Figure B.2: Oracle time profile estimated bias across replicates. The time profile estimators are identified relative to period 1.

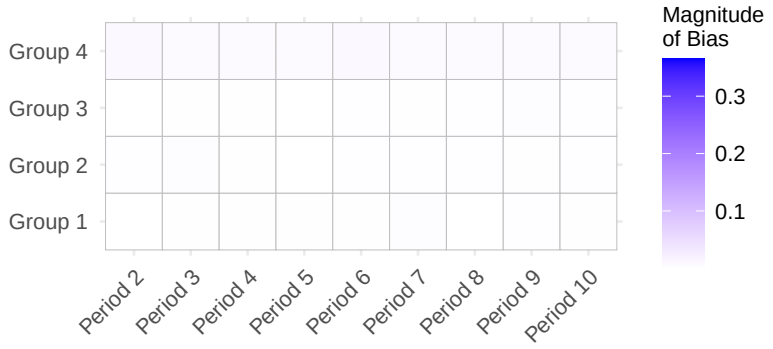


Figure B.3: Oracle time profile estimated RMSE across replicates. The time profile estimators are identified relative to period 1.

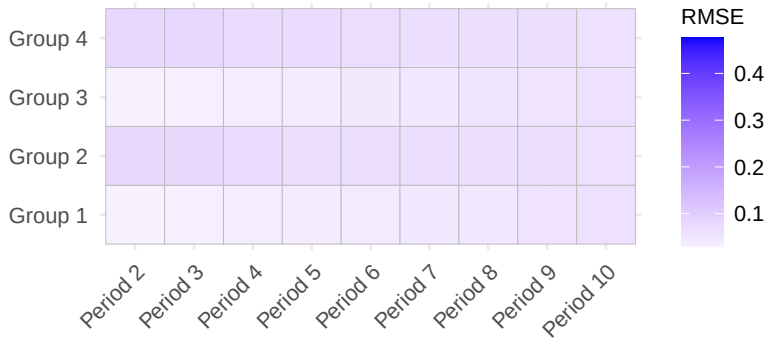


Figure B.4: Oracle time profile estimated 95% confidence interval coverage probabilities across replicates. The heatmap is white at the nominal coverage probability of 0.95. The time profile estimators are identified relative to period 1.

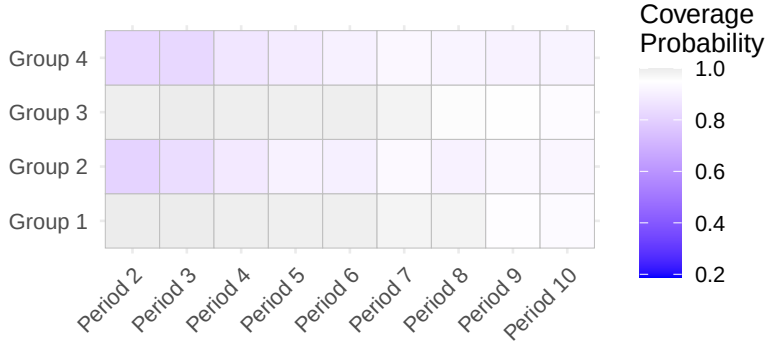


Figure B.5: Time profile estimate comparison for the first randomly selected replicate. The time profile estimators are identified relative to period 1.

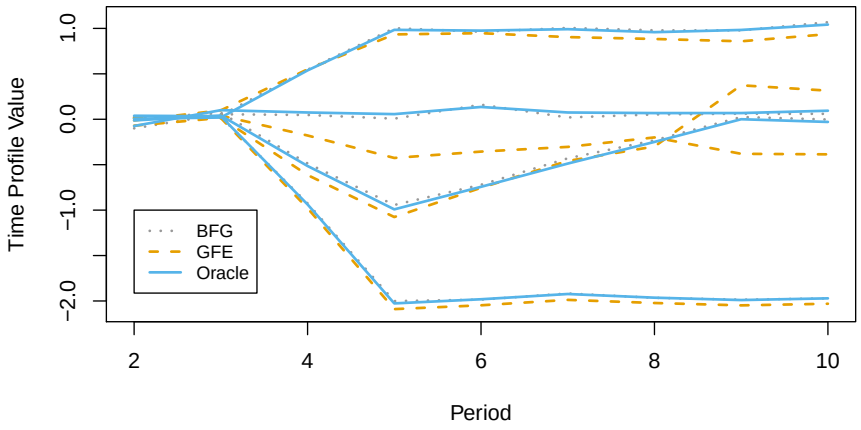


Figure B.6: Time profile estimate comparison for the second randomly selected replicate. The time profile estimators are identified relative to period 1.

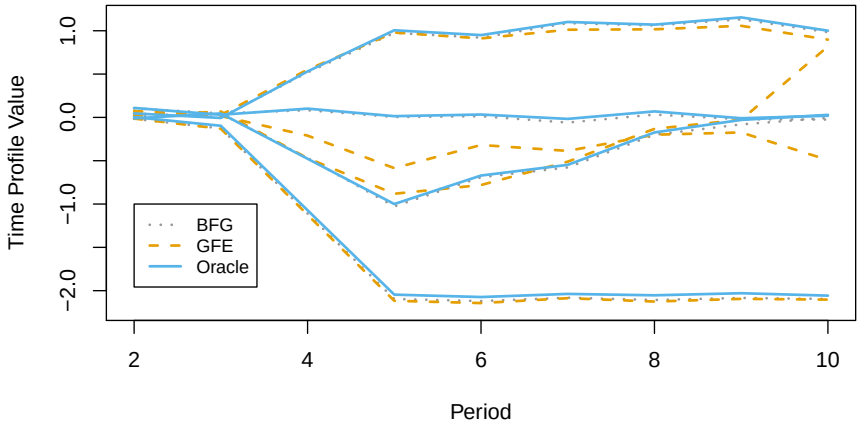


Figure B.7: Time profile estimate comparison for the third randomly selected replicate. The time profile estimators are identified relative to period 1.

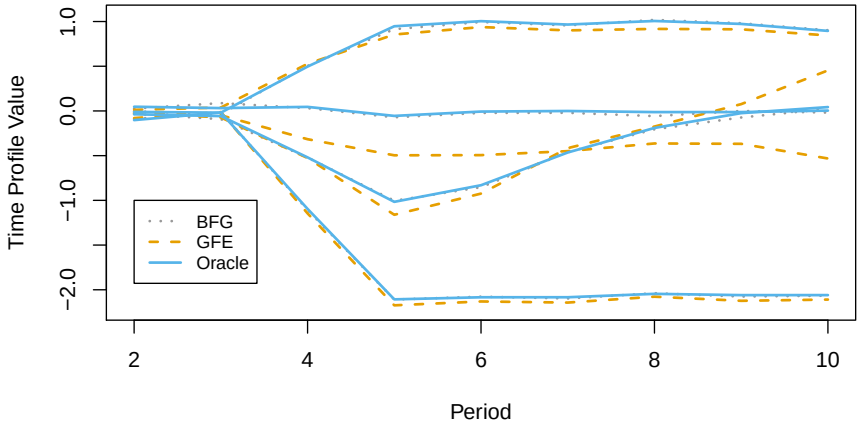


Figure B.8: Traceplot of the θ_1 iterates from the best $(0, -1)$ configuration run in the superannuation application.

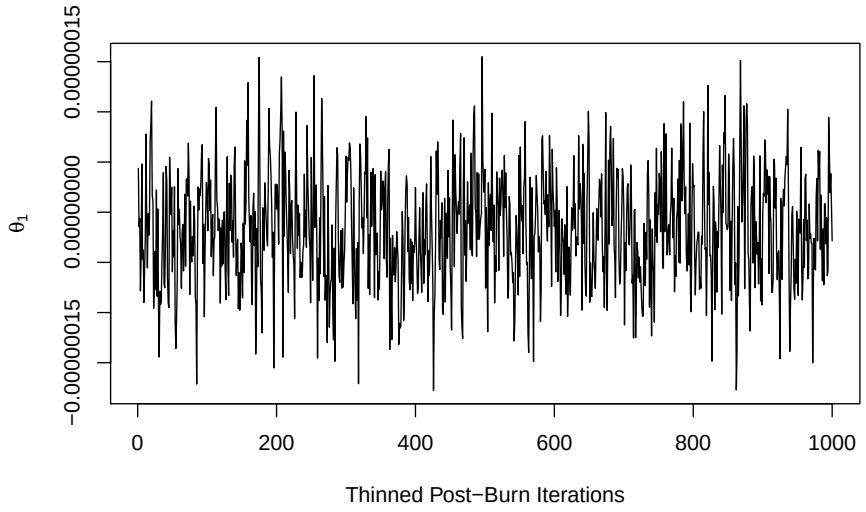


Figure B.9: Traceplot of the θ_2 iterates from the best $(0, -1)$ configuration run in the superannuation application.

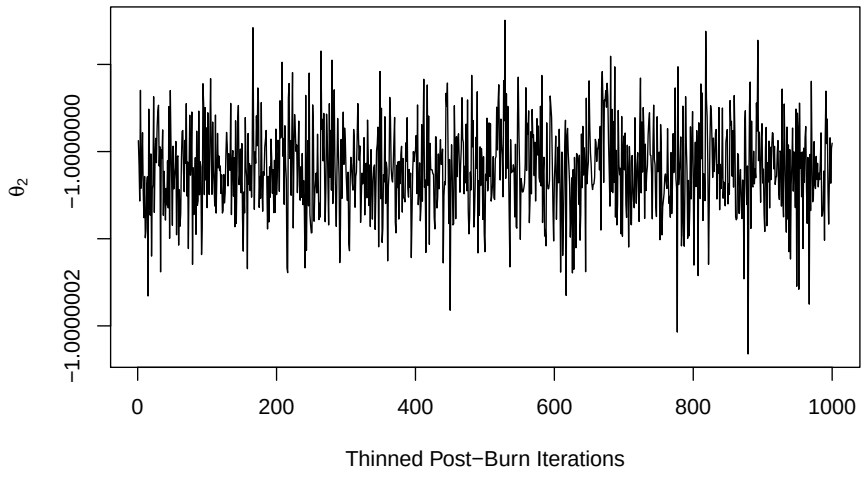


Figure B.10: Time profile posterior means for the 17 largest groups from the $(1, 0)$ configuration run in the superannuation application. The values are relative to period 1 effects.

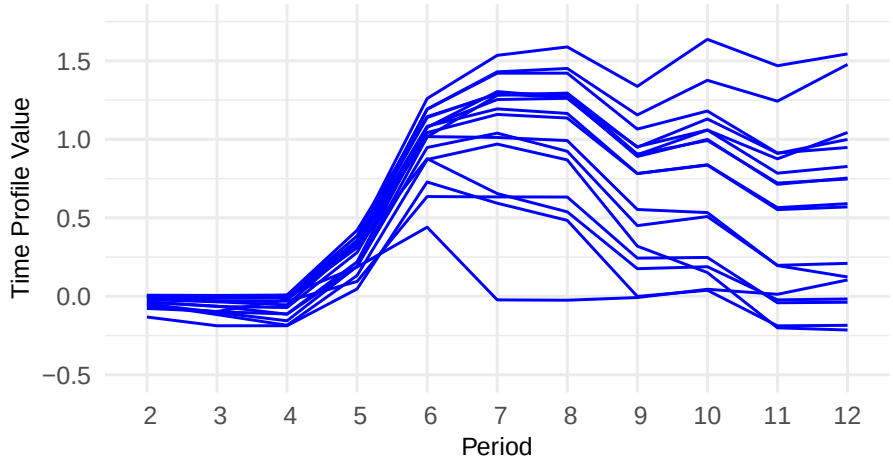


Figure B.11: Traceplot of the θ_1 iterates from the $(1, 0)$ configuration run in the superannuation application.

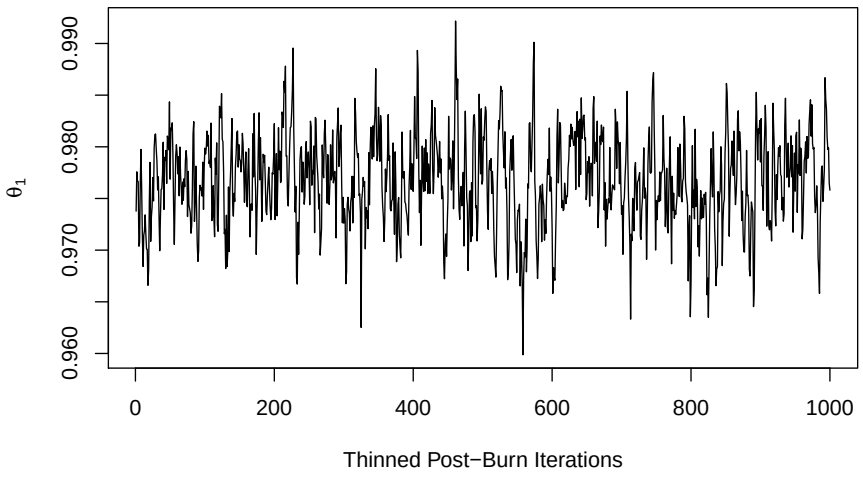
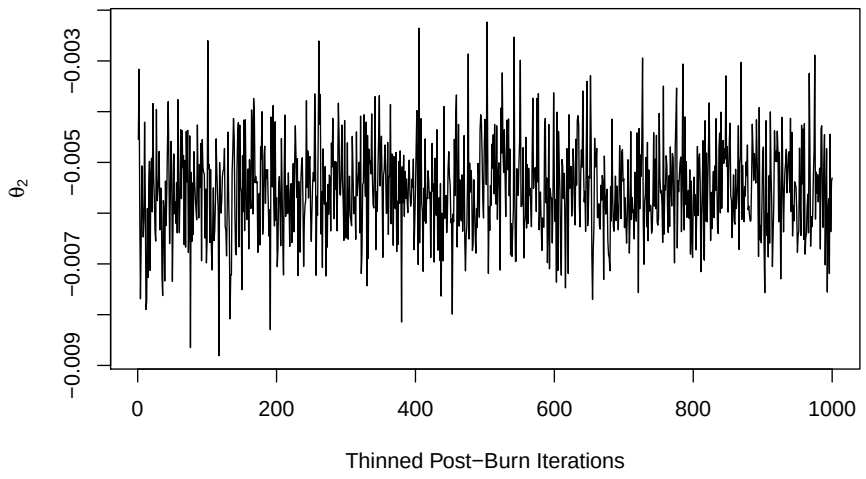


Figure B.12: Traceplot of the θ_2 iterates from the $(1, 0)$ configuration run in the superannuation application.



Appendix C

Supplement for *Bayesian Heterogeneous Panel Models with Latent Group Time Profiles*

Abstract

This supplementary material contains additional results and details from the applications and simulation study presented in Chapter 5.

C.1 The HILDA smoking dataset

The following steps can be used to recreate the $N = 1125$, $T = 10$ balanced panel used in the smoking application in the chapter.

1. Keep variables `lstbcn`, `tifeftp`, `tifeftn`, `gh1`, `rchave` and `lssmkf` from waves 9–18 of the 'responding person' ('Rperson') data files. The weekly number of cigarettes smoked (the dependent variable in the models estimated) is `lstbcn`; the income variable is derived from variables `tifeftp` and `tifeftn`; the self-assessed health rating is `gh1`; the resident child indicator is `rchave`; `lssmkf` is used to find individuals who have missing data in the `lstbcn` variable because they are not smokers.

2. Create a binary indicator for whether the respondent is currently a smoker using variable `lssmkf`. This indicator takes value 0 if the variable `lssmkf` is either '[1] No, I have never smoked' or '[2] No, I no longer smoke'; it takes value 1 if the response is one of '[3] Yes, I smoke daily', '[4] Yes, I smoke at least weekly (but not daily)' or '[5] Yes, I smoke less often than weekly'.

3. If the variable `lstbcn` has missing data but the smoking indicator is 0, set `lstbcn` = 0.

4. Construct the income covariate (in thousands of AUD) as `income_000s` = (`tifeftp` – `tifeftn`) / 1000.

5. Drop any units that have missing data in the dependent variable or any covariates.

6. Drop any unit that never has `lstbcn` > 0 in any of the 10 waves observed.

C.2 Additional application results

C.2.1 Smoking application: 2WFE estimation of LPM

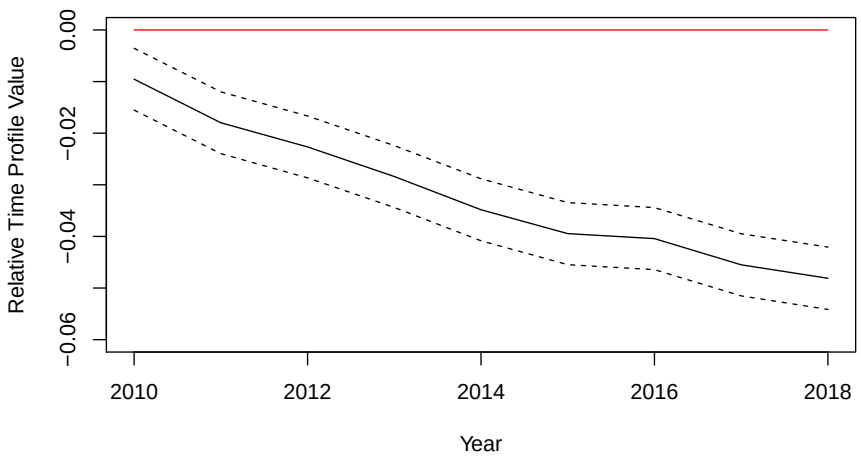
A second dataset is constructed by using the first five steps listed above, excluding the final filter which drops units that never smoke while under observation. Ignoring

this filter leaves $N = 5645$ units in the sample. The dependent variable used for the linear probability model (LPM) is a binary indicator which is equal to 1 if `lstbcn` $>$ 0, and otherwise 0. The LPM is fit using the classical 2WFE estimator with the binary smoking indicator used as the dependent variable, and the same covariates as the model in the chapter.

The interpretations of the regression coefficients for the time dummy variables, plotted in Figure C.1, are as estimated changes in the probability of being a smoker, holding the covariates constant. Comparing this to the corresponding 2WFE estimates in the chapter shows that the overall time trend is similar for both estimates, but that the distribution of kinks in the curve across time are different. This suggests that, on aggregate, the intensive and extensive margin effects of recent smoking policies in Australia follow a similar trajectory.

The estimated income and resident child effects in the LPM are economically small and statistically insignificant, with p-values of 0.11 and 0.31, respectively. The self-assessed health rating is highly statistically significant, with a p-value of 9.0×10^{-4}, but the estimated effect economically negligible: the estimated effect is a decrease of 4.5×10^{-3} sticks smoked per week on average for a one-point increase in health rating (out of 5).

Figure C.1: 2WFE time profile for the LPM run on the smoking dataset including nonsmokers. The values are relative to the 2009 effect.



C.2.2 Smoking application: Placebo test using main specification

To examine whether smoking cessation patterns found in the smoking application in the paper are unique to the period 2009–2018, a 'placebo test' is considered here. The placebo test performs five independent runs of the BFG-HC2 model, run for 120,000 iterations each, on a sample covering a 10-year period immediately before the introduction of the PPM, from 2003–2012. Apart from the different time window, the steps used to construct the dataset are identical to those used for the original estimation sample. There are $N = 1075$ individuals selected into this sample.

Similarly to the main results presented in the chapter, across the 5 runs, the largest h-group covers 93.1–94.4% of the sample, followed by small groups or singletons. In all the runs, the exemplar unit covariate effect 95% credible intervals contain 0 for all three included covariates.

The placebo test time profile results are also broadly similar to the main results in the chapter: the percentage of units classified into the largest g-group ranges from 80.2–81.0% across the five independent runs; the largest 5 cover 95.5–96.4% of the sample; the sixth-largest groups onwards only contain singletons. These results suggest there are 5 substantive groups in the time profiles, which are plotted in Figure C.2 for the first estimation run. All five independent runs have estimated time profiles with identical economic interpretations.

The largest g-group time profile estimates an average decline in smoking of approximately 6 sticks per week over a 10-year period for this group, whereas in the paper the largest group had an estimated 10-year effect of 12 fewer sticks per week. Three of the remaining groups behave similarly to the three 'quitting' groups in the paper: the percentage of units allocated to g-groups 3, 5 and 4 in Figure C.2 who were smokers in 2003 but nonsmokers in 2012 is approximately 69%, 66% and 87%, respectively; the corresponding percentage in the largest group is roughly 29%. In the placebo test results there is also one group that appears economically similar to the two groups in the main results that increased smoking habits during the sample. Approximately 34% of the units in this group were nonsmokers in 2003 and smokers in 2012.

Figure C.2: Substantive BFG-HC2 time profiles for the smoking application placebo test. The values are relative to the 2003 effects. The g-groups are listed in order of size from left to right, top to bottom. The solid black lines depict g-group exemplar unit posterior mean estimates; the dashed lines connect element-wise 95% credible intervals. Zero lines are plotted in red.

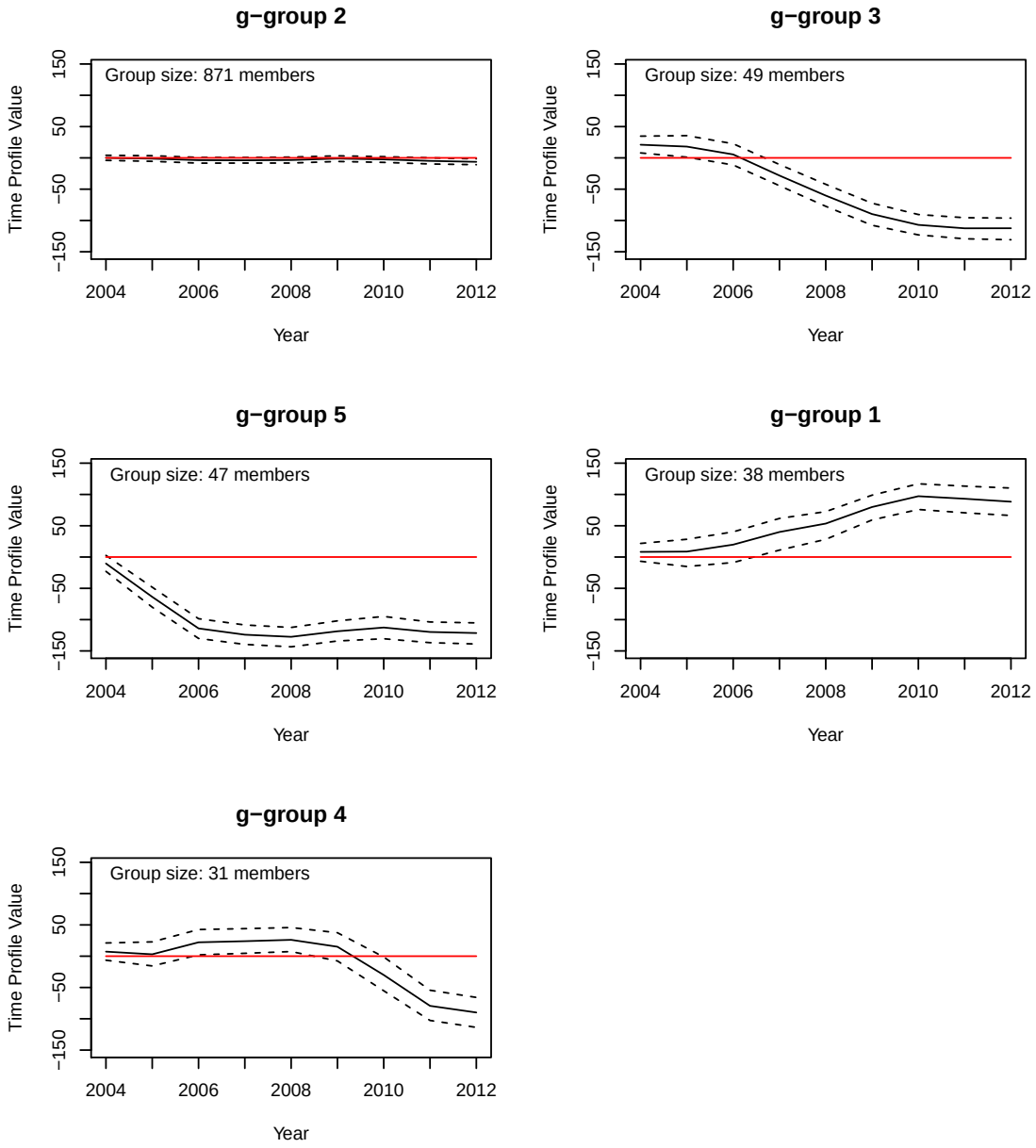


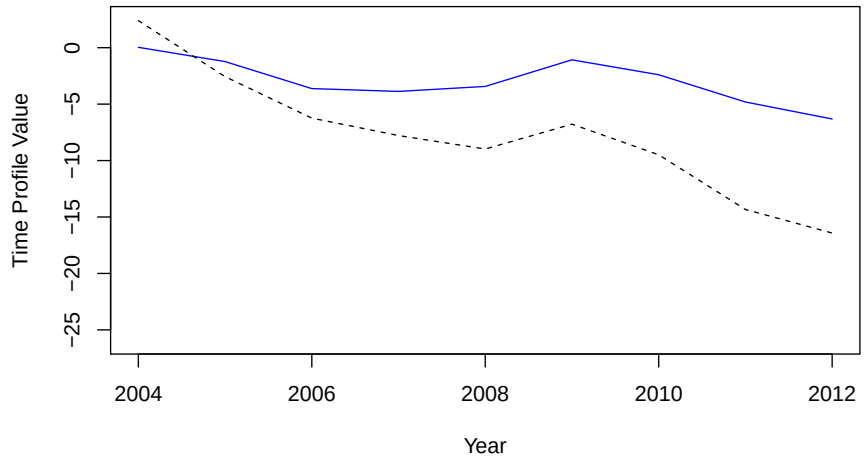
Figure C.3 compares the largest g-group time profile with the single time profile from a 2WFE estimation on the same sample. As in the main results, the 2WFE time profile appears generally unrepresentative of the sample: the 2WFE time profile is most like the time profile from the largest g-group, however the 2WFE time profile magnitudes tend to be larger than the magnitudes for this g-group's profile; the 2WFE time profile also fails to indicate that there are a few small groups with large changes in smoking patterns over time.

Overall, the placebo test results are economically similar to the main results in the chapter, suggesting that smoking cessation results similar to those observed in the chapter may have been found from 2009–2018 even in the absence of the PPM policy. However, the sustained decline in smoking behaviours exhibited by the largest group in the main results is larger than the decline in the corresponding group in the placebo sample; thus, some policy, or a combination of policies, implemented in 2009 or later—which includes the PPM—may explain this difference in 10-year trends between the samples.

C.2.3 Smoking application: Group membership probabilities

To complement the main results in the chapter, Figure C.4 shows, for each unit in the sample, the estimated probabilities of being allocated to any of the six largest g-

Figure C.3: Comparison of the 2WFE time profile and the largest BFG-HC2 g-group time profile in the smoking application placebo test. The values are relative to the 2003 effects. The solid blue line depicts the largest g-group exemplar unit posterior mean estimates; the dashed black line plots the corresponding 2WFE profile.



groups estimated by PEAR. This probability is computed as the proportion of the $M = 1000$ production iterations that the unit is grouped with the exemplar unit corresponding to each of these g-groups. The plotted estimates are from the first run of BFG-HC2 on the data, with the other four independent runs returning similar results. Among the units clustered into one of the four largest g-groups, most have an estimated probability close to 1 of being allocated, in any given iteration, to a group containing their g-group exemplar unit, which implies a relatively high level of consistency in groupings across the iterations. By contrast, the fifth- and sixth-largest groups are less consistently defined, as seen by the nontrivial probability for units in g-group 5 to be allocated with the exemplar unit for g-group 2 in a given iteration, and vice-versa. A possible explanation for the overlap in these two groups across iterations is that the corresponding time profiles (plotted in the chapter) are economically and numerically similar, and can be treated as one behavioural group.

C.2.4 Superannuation application: Group membership probabilities

Figure C.5 gives the corresponding allocation probabilities for the superannuation results in the paper. Most of the largest groups are well separated, in the sense that within each group, there are generally few units with probability mass allocated to other groups. The most obvious exception is g-group 11, where many units from the next-largest group are still often allocated, across iterations, with the exemplar unit for g-group 11. These two groups both cover only 2.5% of the sample each; the fact that the clusters are no longer well separated for such small groups is another argument in favour of focusing the interpretation of results on the largest groups, and ignoring the smaller groups.

C.3 Additional simulation results

The paper reports descriptive plots comparing time profile and covariate effect estimates from the BFG-HC and BFG-HC2 methods with the true DGP values for one randomly selected dataset simulated from DGP1. Figures C.6 and C.7 give the time profile and

Figure C.4: Estimated group allocation probabilities for all units in the smoking application from the chapter. The units are arranged on the vertical axis by g-group, and the g-groups are arranged on the horizontal axis by size.

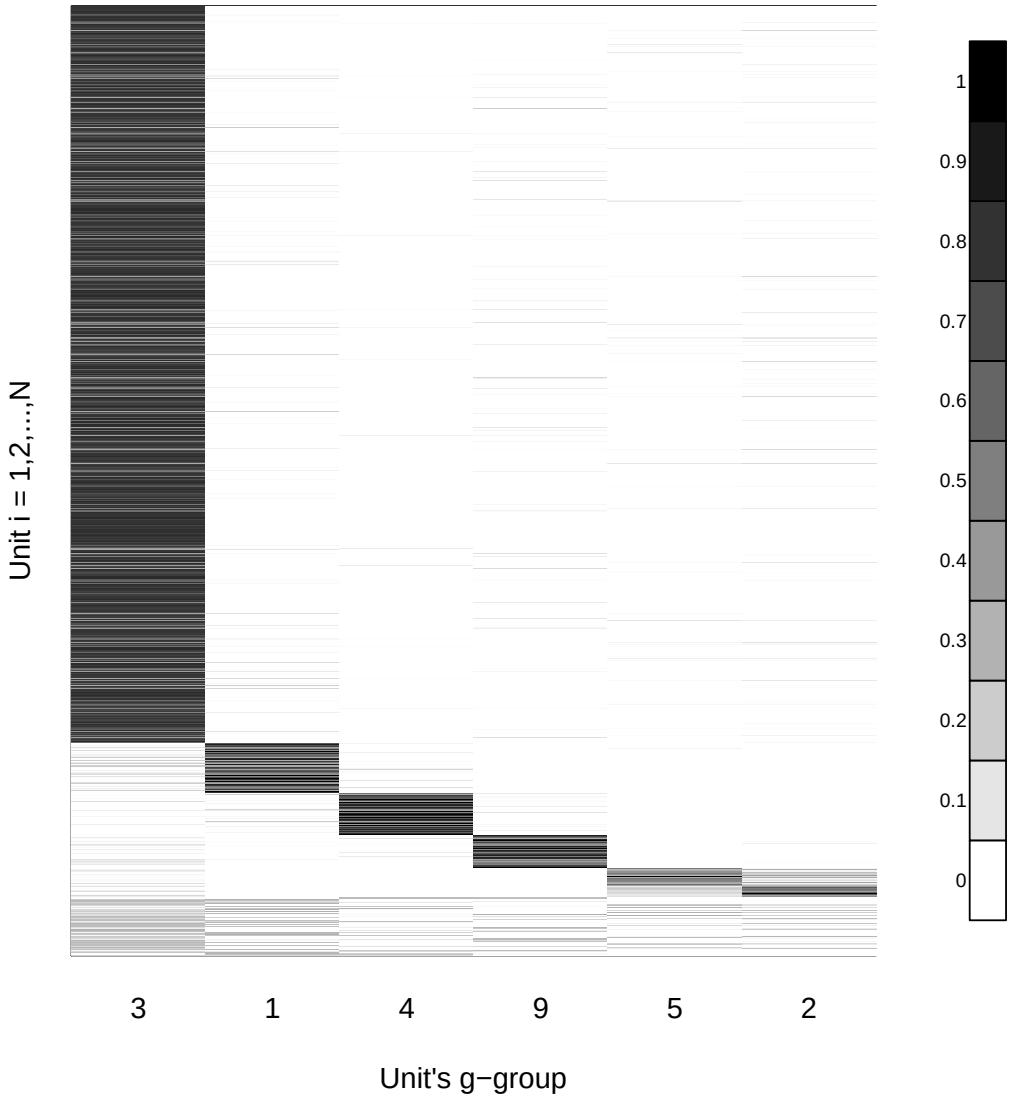
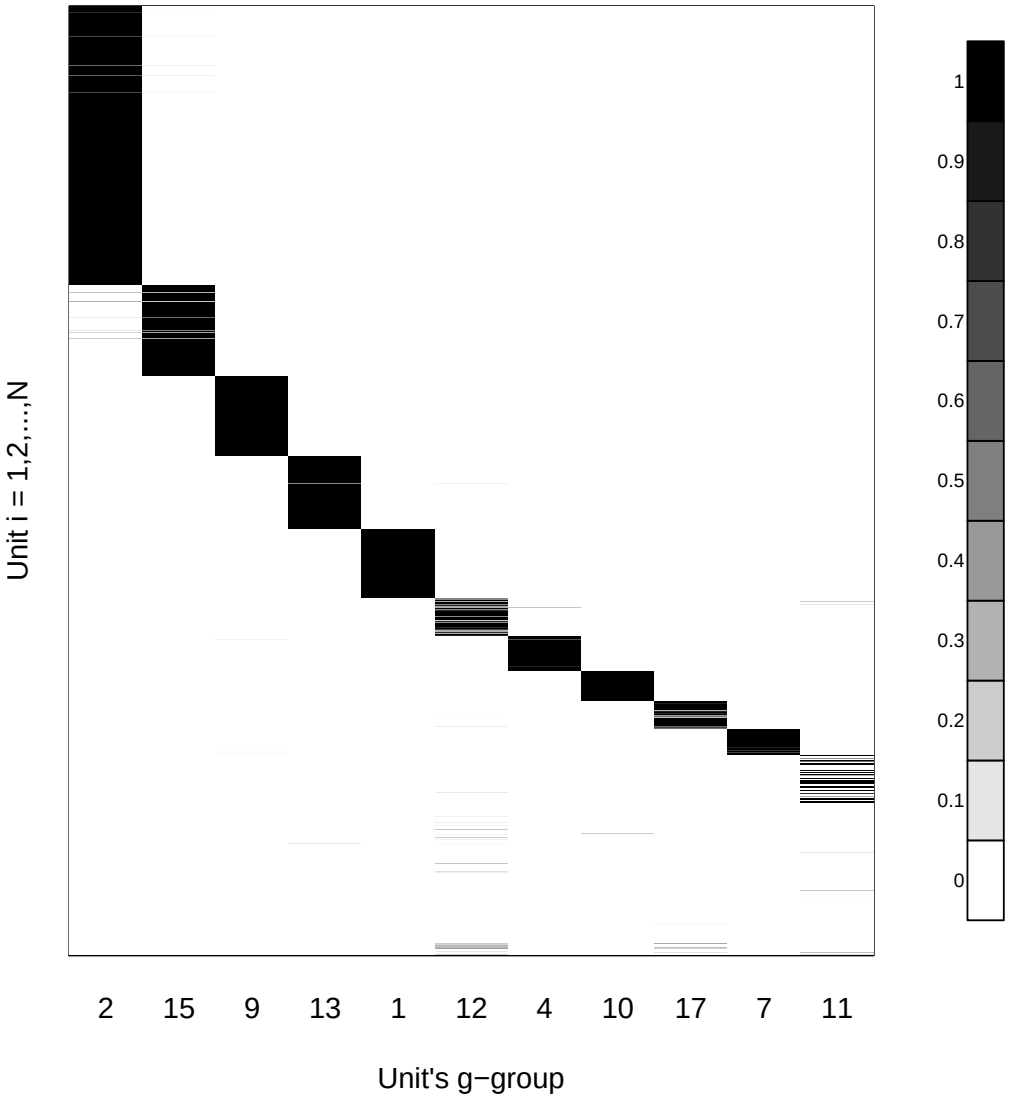


Figure C.5: Estimated group allocation probabilities for all units in the superannuation application from the chapter. The units are arranged on the vertical axis by g-group, and the g-groups are arranged on the horizontal axis by size.



covariate effect comparisons, respectively, for a second randomly selected replicate; Figures C.8 and C.9 for a third.

The paper also presents a similar comparison for one randomly selected dataset generated from DGP2, as well as a plot of BFG-HC time profiles and covariate effects combined into one vector. Figures C.10, C.11 and C.12 plot the corresponding comparisons for time profiles, covariate effects, and their unique combinations, respectively, for a second randomly selected replicate; Figures C.13, C.14 and C.15 for a third.

Figure C.6: BFG-HC and BFG-HC2 time profile estimate comparison with the DGP values for the second randomly selected replicate from DGP1. The values are relative to period 1.

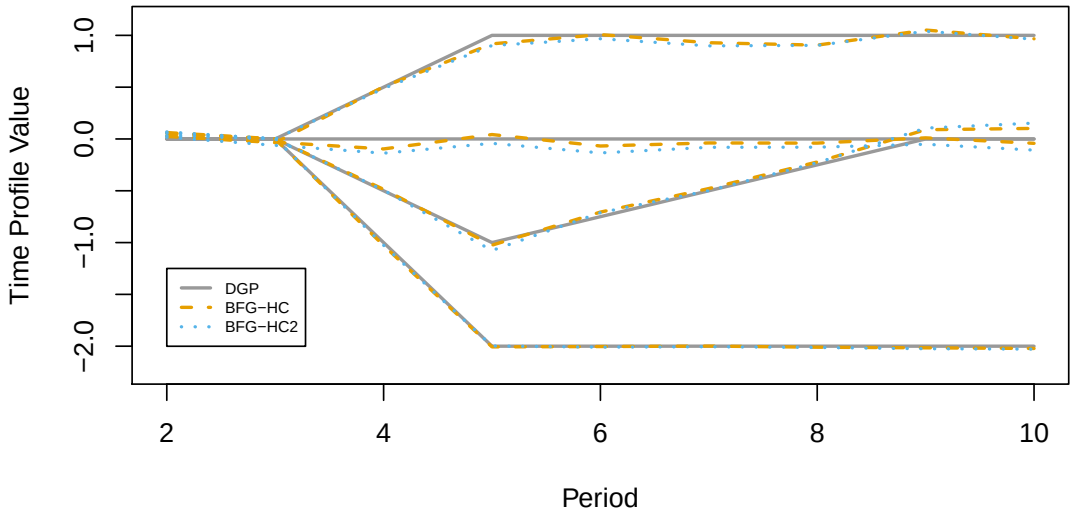


Figure C.7: BFG-HC and BFG-HC2 covariate effect estimate comparison with the DGP values for the second randomly selected replicate from DGP1. x-axis labels 1 and 2 correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively.

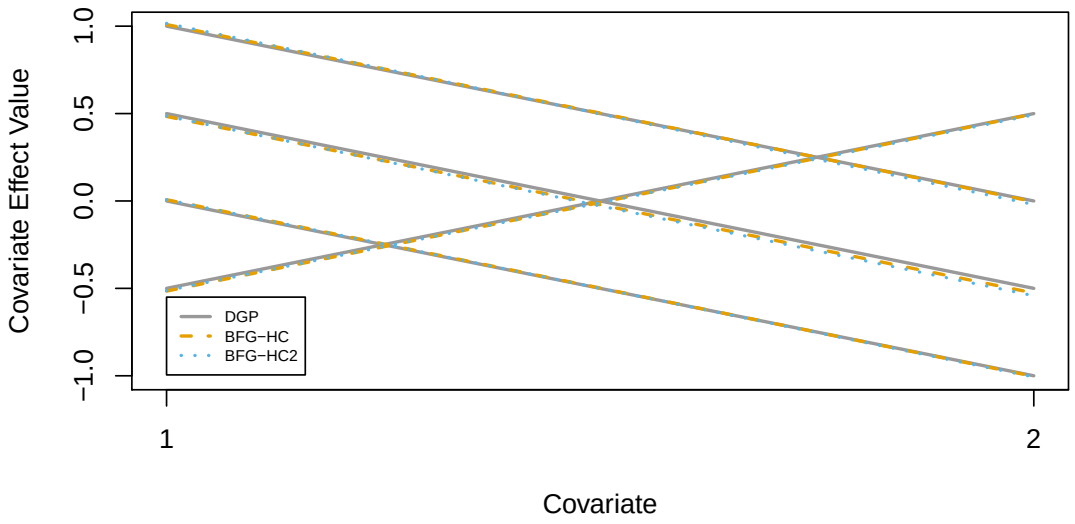


Figure C.8: BFG-HC and BFG-HC2 time profile estimate comparison with the DGP values for the third randomly selected replicate from DGP1. The values are relative to period 1.

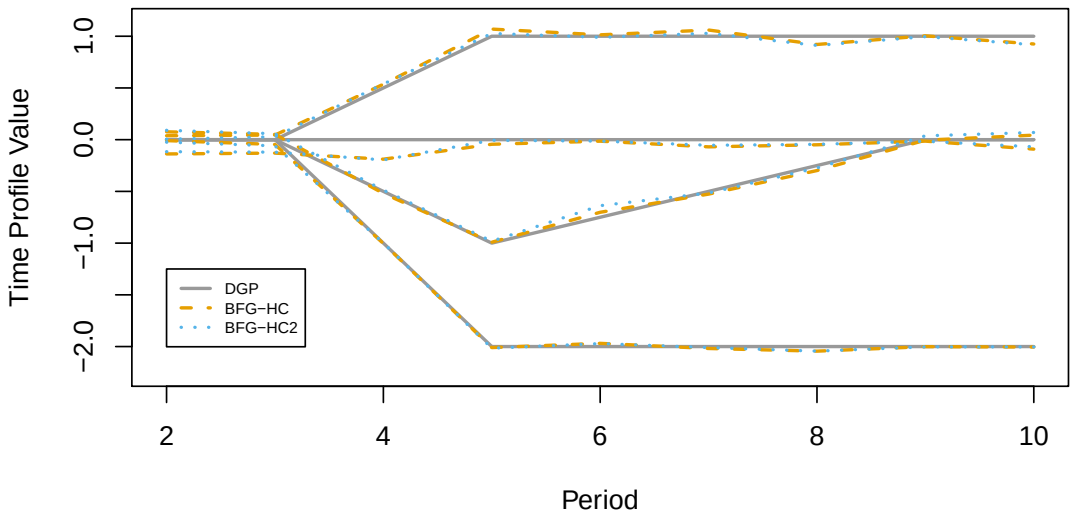


Figure C.9: BFG-HC and BFG-HC2 covariate effect estimate comparison with the DGP values for the third randomly selected replicate from DGP1. x-axis labels 1 and 2 correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively.

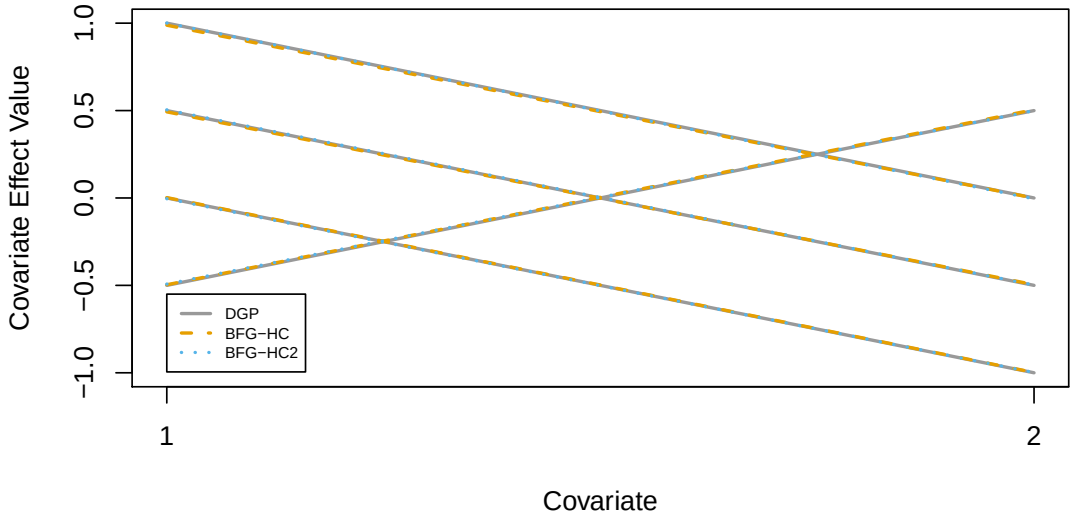


Figure C.10: BFG-HC and BFG-HC2 time profile estimate comparison with the DGP values for the second randomly selected replicate from DGP2. The values are relative to period 1.

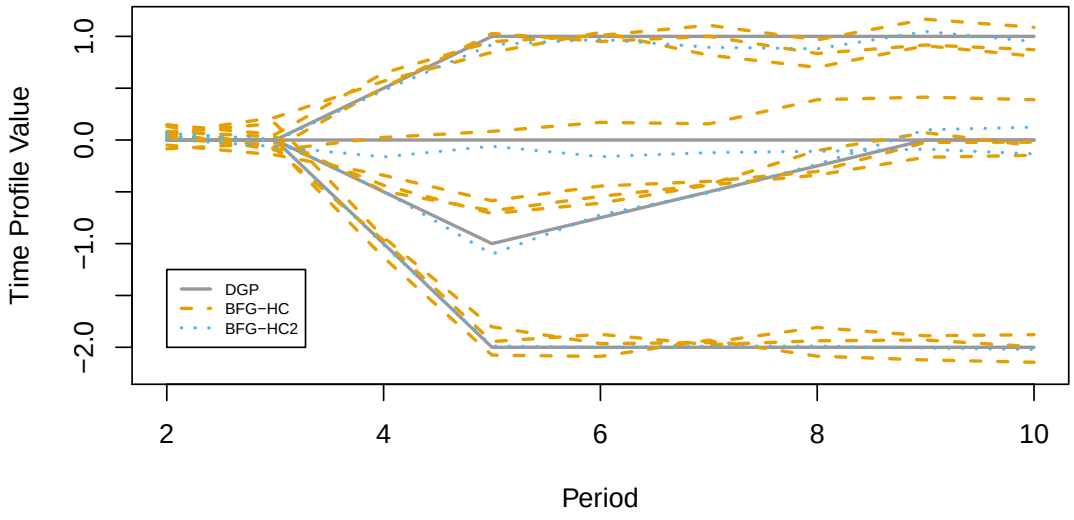


Figure C.11: BFG-HC and BFG-HC2 covariate effect estimate comparison with the DGP values for the second randomly selected replicate from DGP2. x-axis labels 1 and 2 correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively.

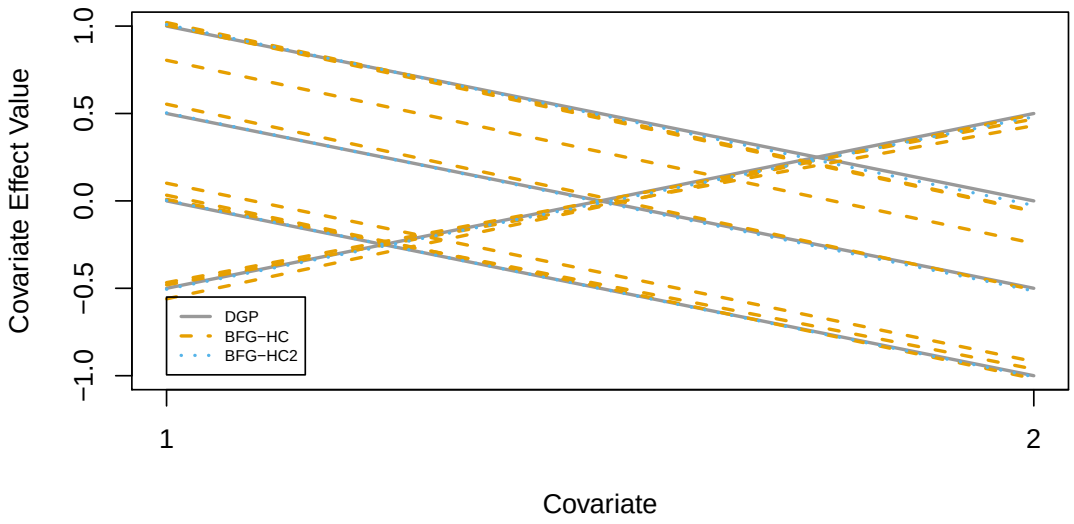


Figure C.12: BFG-HC time profile and covariate effect estimates for the second randomly selected replicate from DGP2. The first 9 x-axis tick marks, labelled 2–10, represent time periods; the subsequent 2 tick marks, labelled 1 and 2, correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively. The time profile values are relative to period 1.

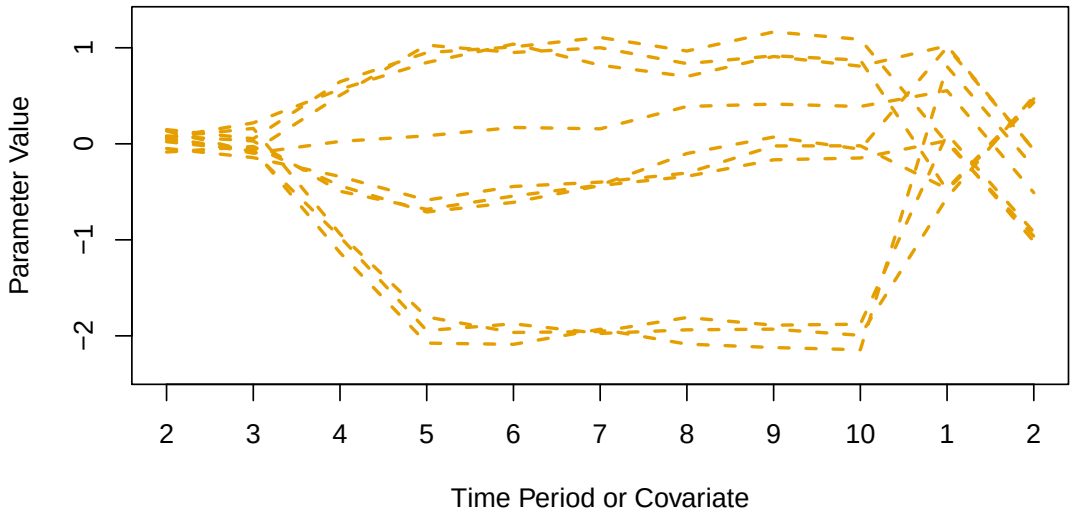


Figure C.13: BFG-HC and BFG-HC2 time profile estimate comparison with the DGP values for the third randomly selected replicate from DGP2. The values are relative to period 1.

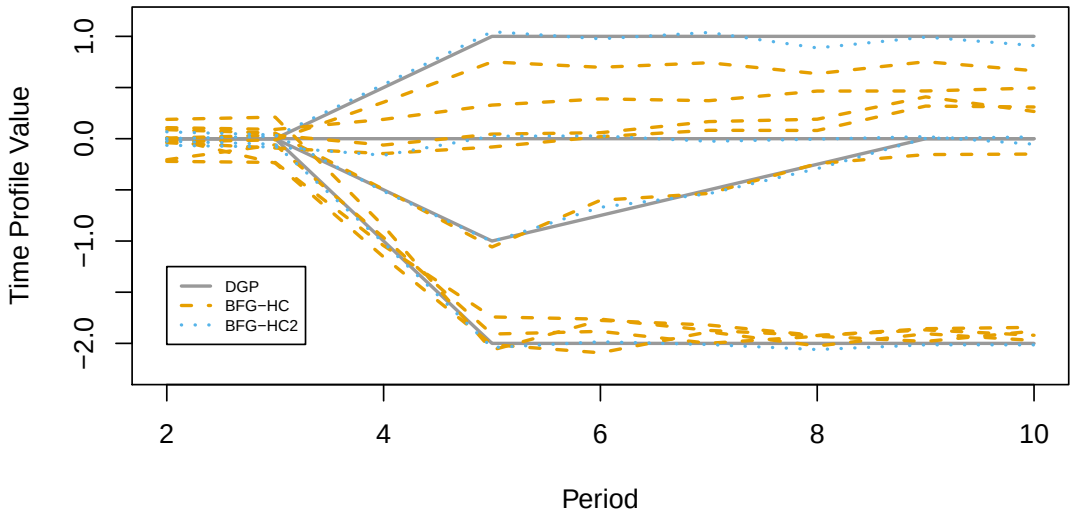


Figure C.14: BFG-HC and BFG-HC2 covariate effect estimate comparison with the DGP values for the third randomly selected replicate from DGP2. x-axis labels 1 and 2 correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively.

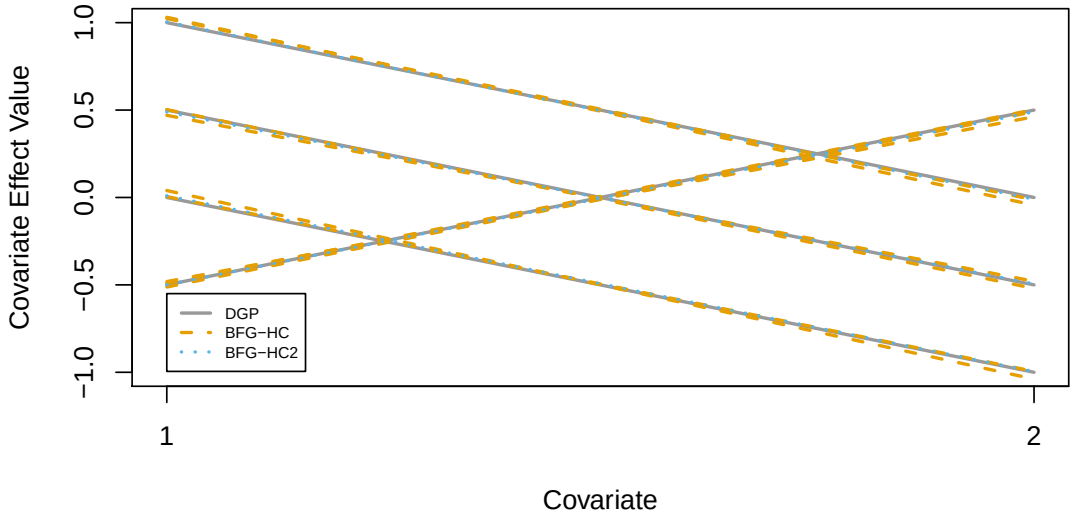


Figure C.15: BFG-HC time profile and covariate effect estimates for the third randomly selected replicate from DGP2. The first 9 x-axis tick marks, labelled 2–10, represent time periods; the subsequent 2 tick marks, labelled 1 and 2, correspond to covariates $x_{1,it}$ and $x_{2,it}$, respectively. The time profile values are relative to period 1.