CeMPA Working Paper Series

CeMPA WP 03/22

Dynamic Simulation of Taxes and Welfare Benefits by Database Imputation

Justin van de Ven

Patryk Bronka

Matteo Richiardi

March 2022

CeMPA
Centre for Microsimulation and Policy Analysis

Dynamic Simulation of Taxes and Welfare Benefits by Database Imputation*

Justin van de Ven[†,‡] , Patryk Bronka† and Matteo Richiardi[†]

December, 2021

* This study benefited from financial support from JPI More Years Better Lives and the UKRI Economic and Social Research Council, grant number ES/W001543/1. Van de Ven also acknowledges financial support from HM Treasury, UK. The views presented in this paper are those of the authors alone. Declarations of interest: none. Corresponding author: j.vandeven@essex.ac.uk; Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ UK.

‡ National Institute of Economic and Social Research, UK.

† University of Essex, UK.

Abstract

This paper proposes a new method for imputing taxes and benefits in dynamic microsimulation and agent-based models, which draws upon existing third-party tax-benefit calculators (e.g. EUROMOD). The suggested approach has the advantages of adapting to the degree of related household sector detail described by a model, mitigating computational burden associated with inter-model communications, and permitting taxes and benefits to be imputed directly from commonly available micro-data sources. A practical application of the proposed method is reported for a dynamic programming model designed to reflect the contemporary UK policy context. Reported results indicate that the proposed method for projecting taxes and benefits can imply a comparable computational burden, and generate qualitatively similar results, to a detailed functional description of fiscal policy.

Key words: static, dynamic, agent-based, microsimulation, taxes, welfare benefits

JEL Classifications: C51, C53, C54

1 Introduction

Economic models designed to project micro-units through time have been growing in scale and sophistication for at least six decades.[1] Within the literature that simulates household-sector dynamics, accurate reflection of financial circumstances is made challenging by the complexity and fluidity of modern tax and benefit systems. The scale of these challenges is such that models in this literature commonly adopt highly stylised approximations of tax and benefit policy, which limits scope for contributing to public policy debate, and exposes associated projections to potentially substantial distortions. In this paper we propose a new method for projecting taxes and benefits within a dynamic heterogeneous household sector model, which delegates the task of reflecting transfer policy to widely available third-party data sources. A practical application of the proposed method is reported, which is found to imply a comparable computational burden, and to generate qualitatively similar results, to a detailed functional description of contemporary UK tax and benefits policy.

Economic studies based on dynamic projections of heterogeneity in the household sector predominantly focus on consumption/saving and/or labour/leisure margins, while abstracting from demographic detail. Macroeconomic agent-based models (MABMs), for example, typically explore household-sector interactions with firms and banks via labour, capital, and product markets. In this context, agents in the household sector are usually considered to differ only with regard to their skills/earnings potential, capital/savings, and the firms with which they interact; demographic characteristics are universally ignored.[2] Demographic characteristics (other than age) are also often abstracted in the literature that accounts for precautionary savings incentives in life-cycle frameworks.[3]

Household demographic detail is usually omitted from dynamic heterogeneous agent projections to mitigate computational complexity and focus on the principal subjects of interest. Yet omitting this detail comes at a cost. Household characteristics that have been identified by the literature as helping to explain empirical regularities of labour/leisure and consumption/savings decisions (in addition to wage potential and wealth) include age, relationship status, dependent children and health.[4] Omitting demographic characteristics

---

[1] The first publication in this literature is commonly attributed to Orcutt (1957). The literature now spans a broad range of methodologies, including dynamic microsimulation and agent-based models; see Richiardi (2014) for discussion. O'Donoghue and Dekkers (2018) review dynamic microsimulation models, and a special issue of the *Journal of Evolutionary Economics* edited by Dosi and Roventini (2019), focusses on agent-based macroeconomics. See also Neugart and Richiardi (2018) for a review of agent-based models of the labour market.

[2] See, for example, Assenza *et al*. (2018), Ashraf *et al*. (2016), Neveu (2013), Dawid *et al*. (2019), Teglio *et al*. (2019), Dosi *et al*. (2015), Botta *et al*. (2021), and Caiani *et al*. (2019).

[3] See, for example, İmrohoroğlu *et al.* (1995), Hubbard *et al.* (1995), Huggett (1996), Gourinchas and Parker (2002), Low (2005), Aydilek (2013), and Conesa *et al.* (2020).

[4] See Attanasio and Webber (2010) and Browning and Lusardi (1996) for discussion of stylised facts underlying the life-cycle framework, including the role of household demographics. See also Attanasio and Browning (1995), Fernandez-Villaverde and Krueger (2007), and Gustman and Steinmeier (2005) for the importance of household demographics in explaining consumption patterns.

like these can consequently detract from the practical relevance of results, particularly in context of seminal trends in cohabitation dynamics, fertility, and healthy life-expectancy.

Given the potential for demographics to influence key margins of economic decision-making, it seems reasonable to suppose that associated detail will become increasingly prominent in prospective dynamic micro-analytic studies of the household. This trend was anticipated by Huggett (1996), who outlines a research programme (attributed to Atkinson, 1971), in which a 'basic' life-cycle framework is augmented to account for "(1) earnings, health and longevity uncertainty, (2) household structure, (3) institutional features such as social security, income taxation, and social insurance, and (4) market features such as borrowing constraints and the absence of some insurance markets" (p. 470). Indeed, a trend toward increasing household detail can be found in the contemporary literature that accounts for life-cycle precautionary savings incentives.[5]

Furthermore, household demographics are typically referenced to determine transfer payments by modern tax and benefit systems. Omitting demographic characteristics from a model consequently limits the capacity to reflect details of transfer schemes that are applied in practice. Hence, existing dynamic micro-analytic studies of the household often focus on after-tax incomes, and/or adopt stylised representations of existing tax and benefits policy.[6] This is also the reason why dynamic models that were specifically designed to analyse detailed descriptions of existing tax and benefits policy include a relatively high level of household demographic detail.[7]

Increasing the demographic detail that is reflected by simulation studies allows for the possibility of obtaining a closer reflection of prevailing tax and benefits policy. Allowing for greater detail in the description of fiscal policy has the dual advantages of capturing a more accurate description of the financial incentives that people face, and generating more practically relevant evidence for public policy debate. The complexity and fluidity of modern tax and benefits policy, however, pose significant challenges to accurate reflection of transfer systems within a model context, even where sufficient household level detail exists to support such a description. The current study proposes a method to facilitate meeting

[5] French (2005), Low and Pistaferri (2015), Keane and Wasi (2016), De Nardi *et al*. (2018) and Albertini *et al.* (2021) accommodate health status, Attanasio *et al*. (2018) reflect education, age, and family composition, and van de Ven (2017a) accounts for uncertainty over a relatively wide range of demographics, including relationship status and number and age of dependent children.

[6] Where a government sector is included for analysis, most MABMs adopt simple linear functions for net fiscal transfers with households. See, for example, Assenza et al. (2018), Ashraf et al. (2016), Teglio et al. (2019), Dossi et al. (2015), Caiani et al. (2020); in contrast, Neveu (2013) and Botta et al. (2021) allow for income tax progressivity. Similarly, for microsimulation models that account for precautionary incentives and adopt linear transfer functions, see Albertini et al. (2021), Low and Pistaferri (2015), Huggett (1996), İmrohoroğlu et al. (1995), and French (2005); in contrast, Kean and Wasi (2016), Conesa et al. (2020), De Nardi et al. (2017) allow for progressive income taxes.

[7] See Orcutt et al. (1976), Caldwell (1997), Morrison (1998), Flood (2007), Nelissen (1991,1993), Spielauer (2013); see also Dekkers and Van den Bosch (2016) for a review of microsimulation models maintained by public and semi-public research agencies in EU member states. Van de Ven (2017a) is an example of a related microsimulation model that accounts for precautionary incentives.

these challenges in a way that can be adapt to the household specific detail accommodated within a given model context.

The method for projecting tax and benefit payments proposed in this paper was devised to permit a dynamic heterogenous household-sector model to draw upon existing third-party tax-benefit calculators (sometimes referred to as static microsimulation models). Well-maintained tax-benefit calculators are now publicly available for many countries[8], and their development involves an appreciably different skill-set to the development of dynamic micro-analytic models more generally. Although the concept of combining complementary model functionality is far from new[9], we are aware of just two examples where a static tax-benefit calculator has been integrated into a dynamic microsimulation context (Liégeois, 2021, and Spielauer *et al*. 2020). At least two considerations may help to explain this lack of integration.

First, relying on third-party inputs for key model components implies a loss of control. It may be, for example, that a static microsimulation model for taxes and benefits is not equipped to reflect intertemporal policy variation or explore potential subjects of interest in a dynamic (microsimulation or agent-based) model. In a similar vein, there may be concerns about on-going maintenance of the candidate static simulation model that exaggerate risks associated with a long-run research investment in a dynamic model structure.

Secondly, integrating a dynamic model with an existing static model of taxes and benefits introduces practical difficulties concerning communications between possibly diverse model frameworks. These problems may appear technically insurmountable, especially where simulation times are a pressing concern (e.g. the dynamic programming literature; Rust, 2008).

Liégeois (2021) explores the feasibility of augmenting a tax-benefit calculator (EUROMOD) with dynamic functionality using the LIAM2 framework. Embedding one model structure (e.g. a static model) within another (e.g. a dynamic model) is one of the most obvious ways of integrating functionality from alternative model frameworks. In the case of Liégeois (2021), interactions between the two considered model structures are facilitated by architectural similarities shared by EUROMOD and LIAM2. Unfortunately, complementarities of this sort can be difficult to identify *ex ante*, and difficult to maintain *ex post*. Furthermore, even where two models share similar programming architecture, they may differ along other dimensions – including periodicity of projected income flows and projected characteristics – which complicate associated integrations between them (see Liégeois for further discussion).

Spielauer *et al*. (2020) also describe a dynamic microsimulation model for households that draws upon EUROMOD for fiscal policy detail. Starting from EUROMOD survey data,

[8] For example, static tax-benefit calculators for the UK policy context include UKMOD (EUROMOD, Sutherland and Figari, 2013), the Intra-Governmental Tax and Benefit Model administered by HM Treasury (IGOTM), Policy Simulation Model (PSM, Department for Work and Pensions), and TAXBEN (Giles and McCrae, 1995).

[9] See, e.g., Richiardi (2014).

the model described by Spielauer *et al*. (2020) is a continuous time model that projects a range of individual-specific (non-financial) characteristics, including age, education, relationship status, fertility and health. Spielauer *et al*. note that the dynamic population projections generated by their model can be used to define weights for imputing associated financial statistics from EUROMOD via static aging methods.

The method set out here is designed to integrate third-party inputs for projecting tax and benefit payments within a dynamic model, in a way that avoids the complicating considerations discussed above. The concept is to obtain a reference database from a third-party source that describes "family specific" characteristics, including private (pre-tax and benefit) and disposable (post-tax and benefit) income. This database is loaded into the dynamic heterogeneous household sector model, and is used in a similar fashion to a look-up table for evaluating transfer payments.

The approach that we describe here builds upon Spielauer *et al*. (2020), by suggesting an algorithm to impute individual specific taxes and benefits for observations within a dynamic model projection (see Section 2.2 for further discussion). The approach that we describe requires no direct communication between alternative model structures during simulation. Furthermore, the micro-data loaded into the dynamic model for projecting taxes and benefits take a standardized form. This allows for the possibility of using alternative static microsimulation models for deriving the source database, in addition to commonly available survey microdata.[10] The structure is also designed to retain features of intertemporal flexibility in the specification of taxes and benefits that can be important in dynamic contexts.

The remainder of the paper is organised as follows. Section 2 describes and justifies principal features of the proposed modelling approach. A practical implementation of the proposed approach is described in Section 3, and an empirical evaluation of the practical implementation is reported in Section 4. Section 5 concludes. Source code that implements the proposed approach is provided in the Appendix, as are details of how to replicate the reported analysis.

2 Proposed Methodology

The proposed methodology proceeds via a number of discrete stages. The starting point for analysis is specification of the database for taxes and benefits. Matching methods are then used to select agents represented in the database to act as donors for agents projected by the dynamic (microsimulation or agent-based) model. Having selected a donor agent from the reference database, values of taxes and benefits are imputed for agents in the dynamic model, in addition to any ancillary statistics of interest. Each of these stages is discussed in turn below, and the section concludes with discussion of simulating policy counterfactuals.

[10] Nevertheless, use of a tax-benefit calculator does facilitate consideration of counterfactual tax and benefits policy; see Section 2.4 for discussion.

2.1 Specification of the reference database

A common starting point for static microsimulation models of taxes and benefits is a publicly available microdata set. In the UK, for example, both the Intra-Governmental Tax and Benefit Model (IGOTM, HM Treasury) and TAXBEN (Institute for Fiscal Studies) load in microdata from the Living Costs and Food Survey, whereas the Policy Simulation Model (PSM, Department for Work and Pensions) and UKMOD (extension of EUROMOD) load in data from the Family Resources Survey. Static microsimulation models typically generate a range of tax and benefit statistics for each individual represented in the starting microdata set, and report results in the form of an augmented data set.

A dynamic model can be designed to read in the augmented dataset generated by a static microsimulation model. It is conceptually possible to perform some pre-import processing of the data generated by a static microsimulation model prior to importation into a dynamic model.[11] That possibility is not considered further here, to limit the disruption associated with switching between alternative tax-benefit calculators and/or survey data sources. The objectives for this stage of the methodology are to collate data for the population of donor agents in a form that is both adaptable to alternative data sources and minimises subsequent computational burden.

The objective to obtain an adaptable method for data importation is facilitated by common features shared by publicly available microdata. Microdata sets that describe economic circumstances can distinguish between alternative levels of aggregation, from the individual, to the tax unit (nuclear family), to the household unit (e.g. this is the case for IGOTM). It is consequently useful to design a dynamic model so that data can be imported from alternative source files in a way that recognises links between files. All file and variable names should be provided as parameters to the dynamic model (rather than hard-coded).

Having defined a dynamic model so that it is capable of loading in data from diverse sources, it is necessary to organise the data in a way that will minimise subsequent computational burden. This objective depends upon the methods used to extract data from the reference database, which are addressed below.

2.2 Matching methods

The econometric literature presents a range of alternative approaches that could be used to impute individual specific proxies for tax and benefit payments from a reference database.[12] The approach proposed here uses matching methods to identify individuals in the reference

---

[11] For example, combining data possibly reported in a number of separate output files reported by a static model (e.g. household level data combined with individual level data), and screening the data to focus exclusively on variables of interest.

[12] These methods range from functional regression specifications, through non-parametric descriptions, to pair-wise matching methods.

database that possess comparable characteristics to individuals in a dynamic model. This approach was chosen because it facilitates post-simulation validation of imputed statistics via inspection of the characteristics of matched individuals.

Best practice requires the assumed matching method to be tailored to the subject of interest (Stuart, 2010). Allocation of the treatment (evaluation of transfer payments by the static microsimulation model) is universal in the current context, which precludes use of propensity score matching. Furthermore, the wide range of variables that may be considered important predictors of tax and benefits payments complicates use of Mahalanobis distance matching.

The matching method should ensure that there is a strict delineation with respect to forms of population heterogeneity that have a substantial bearing on tax and benefit payments (e.g. adult marriage/cohabitation status, health, and retirement), suggesting the use of exact matching methods. Nevertheless, the method should also guarantee that a match is obtained for all conceivable circumstances in the dynamic simulation, which motivates interest in coarsened exact matching (Iacus *et al*., 2009). Furthermore, the method should seek to limit differences in relation to private (pre-tax and benefit) income. This recognises the need to approximate incomes, while acknowledging the impossibility of exact matches to all conceivable measures of income. Finally, factors that affect tax and benefit payments but are not included in a static and/or dynamic model – such as imperfect benefits take-up, tax avoidance and tax evasion (whether sampled or simulated) – can result in noisy observations reported by a reference database. The matching method should limit exposure to "high influence" observations.

The above considerations motivate the use of stratified coarsened exact matching, followed by weighted nearest-neighbour matching over private incomes. Importantly, the features included in the coarsened exact matching routines can be tailored to the household level detail described by a dynamic model, and adapted if this detail is expanded. Spielauer *et al*. (2020) describe a variant of this approach, but omit nearest-neighbour matching as their model does not project private income. Further details concerning the practical implementation of the proposed matching procedure are provided for the example that is discussed in Section 3 (see Section 3.2.2).

2.3 Imputing taxes and benefits

The problem of imputing taxes and benefits for a dynamic model is complicated by the potential for temporal variation of the policy environment. In context of trend (real) wage growth, for example, holding the description of taxes and benefits fixed through time can result in widespread bracket creep (e.g. Bohanon, 1983), and declining relevance of welfare safety-nets. This section begins by discussing issues relating to imputation of taxes and benefits at a point in time, before proceeding to discuss how the approach can be adapted to reflect alternative forms of temporal policy variation.

2.3.1 Imputations at a point in time

Having matched individual-specific characteristics of an agent in a dynamic model to those in a reference database, as discussed in Section 2.2, the objective is to impute taxes and benefits in a way that takes into consideration the potential for inexact matching of private (pre-tax and benefit) incomes. One approach would be to match each simulated individual to multiple individuals in the reference database, and use interpolation methods for imputation. We reject this possibility in favour of an approach that permits each simulated individual to be matched to a single individual in the reference database, as this has the advantage of facilitating post-simulation diagnosis of imputed transfer payments.[13]

Given the above, the problem is how to impute taxes and benefits for a simulated individual from the taxes and benefits reported by a reference database for a (single) matched individual in a way that adjusts for potential differences in private incomes. Where the private incomes of both matched observations (simulated and database) are non-zero, transfer payments can be imputed by assuming that the ratio of transfer payments to private incomes is the same for matched individuals. That is, if x_i and t_i denote, respectively, the private income and transfer payment of individual i, then $t_{sim} = \frac{t_{db}}{x_{db}}.x_{sim}$, with sim denoting the individual simulated in the dynamic model and db the respective matched individual from the reference database.

As x tends toward zero, increasing curvature of t/x complicates use of the ratio suggested above for imputation. We consequently propose a two-step process for matching individuals by private income. In the first step, coarsened exact matching is used to group individuals by broad income bands; e.g. low, middle, and high private incomes. In the second step, weighted nearest-neighbour matching is employed. For individuals in the lowest private income grouping, transfer payments of simulated individuals are set equal to those of the matched individuals identified from the reference database; that is $t_{sim} = t_{db}$. Otherwise, the ratio method described in the preceding paragraph is used.

2.3.2 Accommodating policy variation through time

It is (conceptually) possible to consider a separate reference database for an arbitrarily large number of discrete time periods, which would provide the analyst with tight control over the assumed temporal evolution of tax and benefit policy. A more abstract approach is, however, possible where trend variation of broad aspects of the policy environment is concerned. Here we focus on two discrete cases designed to reflect trend growth of, respectively, private income thresholds (across which fiscal policy treatment varies discontinuously) and benefit payments.

Consider the case where a database has been evaluated that describes taxes and benefits for an assumed "reference year". A fixed growth rate of all private income thresholds can be

[13] The idea is that the output for each individual at each point in time projected by a dynamic model should include identifiers for matched individuals used to impute tax and benefit payments from the reference database.

reflected by discounting simulated private income to the respective reference year for identifying a match from the reference database. In this case, undiscounted private incomes should then be used to impute tax and benefit payments after matching via the approach discussed in Section 2.3.1.

Trend growth in (means-tested) benefit payments is usually accompanied by growth of the cut-off measures of private income, beyond which benefits are no longer payable. The joint determinacy of benefit values and private income cut-offs can be captured by defining two databases, which describe tax and benefit payments for alternative reference years. Each database can then be used to evaluate taxes and benefits that would apply in each of the respective reference years, and an assumed interpolation method used to impute payments for any given year.

Suppose, for example, that $y_{i,t}$ denotes disposable (post-tax and benefit) income of individual i at time t. Two databases are used to impute disposable income, one that would apply at time $t = 0$, and another at time $t = T$. Then assuming geometric interpolation between the two databases, we would have:

$$y_{i,t} = y_{i,0} \cdot \left(\frac{y_{i,T}}{y_{i,0}}\right)^{t/T}$$

This approach could be implemented without increasing the computational demands associated with evaluation of database matching, by limiting variation between the two reference databases to the reported taxes and benefits.[14] An added advantage of this limitation is that it would facilitate post-simulation diagnosis of imputed transfer payments, as each individual in the dynamic model would be matched to the same observation in both reference databases.

2.4 Analysis of policy reforms

Dynamic models are typically used to explore the intertemporal effects of alternative tax and benefit policies via the method of comparative statics. This involves comparing model projections generated under a status-quo policy specification against projections obtained following a counterfactual policy reform. The status-quo and counterfactual simulations are designed to be identical in all respects, with the exception of the assumed policy reform. Where a policy reform of interest can be reflected by a tax-benefit calculator, then the tax-benefit calculator can be used to generated reference databases for both the 'status-quo' and 'counterfactual' policy environments. Analysis would then proceed by comparing projections from the dynamic model generated by assuming each of the respective databases. This approach is adopted for the empirical analysis reported in Section 4.

Nevertheless, tax-benefit calculators usually provide an incomplete description of the policy environment. Key aspects of policy that are seldom reflected by (static) tax-benefit calculators, but are well-suited for analysis with dynamic (agent-based/microsimulation)

[14] That is, each reference database should be derived from a tax-benefit calculator using the same input data.

models, include the treatment of inheritances, wealth, private pension contributions, and housing transactions. If a reform of interest cannot be reflected natively by a tax-benefit calculator, then there are (at least) two options available to the analyst: augment the database derived from the calculator; or include a functional add-on to the database sourced from the calculator for imputing taxes and benefits.

Consider the case of augmenting a tax-benefit calculator to accommodate a specific policy; for example, an inheritance tax. In this case, receipt of inheritances would need to be included in both the tax-benefit calculator and the input data supplied to the calculator. If such data were not described by the original input dataset, then "hypothetical examples" could be appended to the dataset. Having revised by the tax-benefit calculator and input data, a new database could be generated for importing into a dynamic model. The bearing that the augmented reference database had on projections from the dynamic model would then depend upon the matching methods employed.

Continuing the above example, suppose that the matching criteria used to impute tax and benefit payments in a dynamic model failed to discriminate (either implicitly or explicitly) for receipt of inheritances in the revised reference database. Then the incidence of inheritance taxes described by the revised database would be reflected as increased random variation in the transfer payments imputed by dynamic model projections. In this case, imputing deterministic incidence of inheritance taxes would require the set of matching criteria to be extended to distinguish receipt of inheritances.

The complications discussed above could be avoided by programming the influence of inheritance taxes as an add-on to the remainder of the tax and transfer system. Under this approach, a reference database could be used to project most taxes and benefits, with residual schemes accommodated using a functional add-on. This approach also alludes to the possibility of omitting consideration of a tax-benefit calculator entirely. In that case, the default policy environment could be reflected by observations reported by cross-sectional survey microdata. Analysis of a policy counterfactual of interest would then involve adding a function to the dynamic model that would adjust net transfer payments to reflect the (marginal) influence of the counterfactual, given the (status-quo) policy context.

3 Practical Implementation

The methodology for imputing tax and benefit payments outlined in Section 2 was implemented within LINDA, a dynamic microsimulation model designed to reflect the contemporary UK policy context. This section begins with a brief description of LINDA, before providing details concerning the practical implementation of the proposed methodology.

3.1 Model overview

LINDA presents a useful case-study for considering the proposed method for imputing tax and benefit payments because the model is expressly designed to explore behavioural responses to policy alternatives, and associated computational burden is a pressing issue of concern. The model is publicly accessible, and free to download from www.simdynamics.org. As LINDA is not the focus of the current paper, a brief overview of relevant aspects of the model is provided here, and the interested reader is referred to van de Ven (2017a) for technical detail.

LINDA projects panel data at annual intervals for an evolving sample of simulated adults. The decision-making unit assumed by the model is the "benefit unit", defined as a single adult or partner couple and their dependent children. The specification of the model considered for the current study projects consumption, private pension scheme participation, and labour supply decisions as though these are made to maximise expected lifetime utility, where utility takes a nested Constant Elasticity of Substitution form.

The decision of whether to contribute to a private pension in each year is limited to individuals who choose to work, and pension contributions are defined as a fixed proportion of labour income. Labour supply is selected from three discrete alternatives for each simulated adult in each year, representing, full-time, part-time, and non-employment. The decision to supply labour involves a trade-off between leisure time and cash available for (non-durable) consumption. Consumption is chosen with respect to a standard budget constraint, with the upper bound to (unsecured) loans defined by each individual's minimum net present value of all potential future streams of disposable (post-tax and benefit) income.

The model projects relationship status, fertility, and mortality for each adult in each year, all of which are considered to evolve with uncertainty from one year to the next. Wage potential is assumed to evolve as a random walk with drift, and pension and non-pension wealth both evolve with certainty based on standard accounting identities.

Starting with data reported by the Wealth and Assets Survey for the UK population cross-section in 2017, the model is designed to generate panel data forward and backward through time. Entry and exit from the simulated population are designed to reflect forward projections for the evolving population cross-section. The model was parameterised to survey data reported up to 2017, following the method described in van de Ven (2017b).

3.2 Practical implementation of proposed approach

The starting specification for LINDA includes a detailed functional description for imputing tax and benefit payments, designed to reflect UK transfer policy as at April 2016; associated programming code can be accessed as described in Appendix A.1. LINDA was amended for the current study to allow taxes and benefits to be projected using the database approach described in Section 2. The revised code permits taxes and benefits to be projected by

LINDA using a wide range of data sources, including databases supplied by IGOTM and UKMOD. Details of the implementation are provided below, and associated programming code can be accessed as described in Appendix A.2.

3.2.1 Importing the tax-benefit database

LINDA was adapted with parameters and routines that permit data to be imported for up to two reference databases. Data for each database can be supplied in up to three files, which may distinguish between alternative units of analysis (household, tax unit, and individual level data). The files are assumed to be saved in comma-separated-variable format, as this format is supported by a wide range of statistical software packages. The first row of each file is assumed to provide variable names, which are used by LINDA to manipulate the data.

3.2.2 Matching methods

LINDA uses benefit unit specific characteristics to identify three nested "coarsened exact matching levels". The first matching level, which is most detailed, distinguishes between:

- three age categories for the benefit unit reference person: under 45 (child-rearing), under state pension age, and post state pension age;
- relationship status of the benefit unit: single, couple;
- number of children under 5 (schooling age) in the benefit unit (maximum 2);
- number of children aged 5+ in the benefit unit (maximum 3);
- renter status (distinguishing renters from non-renters);
- three labour categories for each adult in the benefit unit (not employed, part-time employed, full-time employed);
- and, three private income categories for the benefit unit: under £225 per week (low), under £710 per week (middle), £710+ per week (high).

The second (intermediate) matching level considers the same population division as the first, but aggregates the two lower age categories, considers a maximum of one child under schooling age, does not distinguish between renters and non-renters, and distinguishes only between employed and not employed adults. Finally, the third (most coarse) matching level considers the same population categorisation as the second, but does not distinguish between child age (subject to a maximum of 3 children), and ignores differences by employment status.

It may be noted that the matching method described above makes no allowance for a range of factors that commonly influence transfer payments, including health status, care needs, and disability. The set of characteristics included for the coarse exact matching reflects those that are explicitly distinguished by the specification of LINDA considered here (discussed in Section 3.1). As discussed in Section 2, any influence that omissions from the matching routine have on tax and benefit payments described by a reference database will appear as random innovations in the transfer payments simulated by the model.

Step 1: Starting with the first matching level described above, LINDA checks whether the respective subgroup implied by an observation's characteristics, as reported by the reference database, is non-empty. If so, then the model proceeds to the next step in the procedure; if not, then the second, and finally the third matching level is consulted to find a matching sub-population. The third matching level is designed to be sufficiently crude that all implied subgroups are non-empty.

Step 2: LINDA searches through the matching subgroup identified in step (1) to select a "candidate pool" of observations that share the closest taxable incomes to the targeted individual. This step identifies a pool of candidates, rather than the single closest individual, to mitigate difficulties that might otherwise arise due to disparate measures of taxes and benefits reported for otherwise similar individuals by the reference database.

Candidates are drawn from the subgroup identified in step (1) based on the proximity of their taxable incomes with respect to the targeted individual. Up to four measures of private income are accommodated for individuals within the "candidate pool". Each candidate included in the pool is associated with a selection weight that combines their sample weight in the reference database with a factor that is inversely proportional to the difference between their taxable income and that of the targeted individual (subject to an income disregard), and to the total number of candidates.

Consider, for example, a targeted individual with taxable income of 10, where the coarsened exact matched subgroup from the reference database (identified in step 1) is comprised of individuals with taxable incomes in the set (6, 7, 9, 11, 12, 12, 13). Then, only individuals in the matched subgroup with incomes in the set (7, 9, 11, 12) would be included in the candidate pool. In this case, with an income disregard of 1 and equal sample weights, the selection weight would be 0.1 for the candidate with income 7, 0.3 for each of the candidates with incomes 9 and 11, and 0.15 for the two candidates with income of 12.[15]

Step 3: Given the candidate pool and associated selection weights as described in step (2), LINDA uses one of two methods to impute transfer payments. When solving the lifetime utility maximisation problem, all candidates are included for the imputation, with the contribution of each candidate defined by their respective weight. In contrast, when projecting individuals through time, a single donor is selected by random draw for each individual at each point in time, with probabilities of selection reflecting each candidate's selection weight. The implication of this approach is that individuals are assumed to ignore

[15] Weights obtained by evaluating for each candidate a test statistic equal to the absolute difference between the candidate's income and the reference income, subject to a minimum defined by the income disregard. The maximum of all candidate test statistics is then identified (3 = abs(10 - 7) in the example where abs(.) denotes the absolute operator), and intermediate weights calculated by dividing the maximum by each candidate's test statistic (for the candidates with income 12, this would equal 1.5 = 3 / abs(10 – 12)). Weights summing to one are then obtained by rescaling all candidate weights; in the example, this involves dividing each intermediate weight by the sum of all candidate intermediate weights, equal to 10 = 1 + 3 + 3 + 1.5 + 1.5.

uncertainty associated with their tax treatment, as described by heterogeneity within their respective candidate pool, when formulating their expectations concerning lifetime utility.

3.2.3 Optimising the matching method

LINDA starts by distinguishing, for each "matching level" (of three, see Section 3.2.2), the coarsened exact matching subgroup of each observation described by data loaded in for a reference database. Within each subgroup, observations are also ranked by their respective private (pre-tax and benefit) incomes. These results are stored in a three-dimensional matrix, $M(k, g, r)$, distinguishing the matching category, $k \in \{1,2,3\}$, the category subgroup, g, and the income rank, r.

Consider any observation projected by LINDA, i, for which tax and benefit payments need to be imputed. Starting from the most detailed matching level, $k = 1$, and the vector of observation i's characteristics (age, relationship status, dependent children, etc.), v_i, a function is used to identify the respective subgroup, $g_{k,i} = f(k, v_i)$. If the respective matching subgroup is empty, $M(k, g_{k,i}, .) = 0$, then the next level, $k + 1$, is considered. Otherwise, starting from $j = 1$, LINDA compares the private income of observation i, x_i, with the private income of the database observation $M(k, g_{k,i}, j)$, x_j. If $x_i > x_j$, then LINDA proceeds to observation $j + 1$. Nearest neighbours to include in the candidate pool are selected about the first rank where $x_i \leq x_j$ or where the matching subgroup $M(k, g_{k,i}, .)$ is exhausted (whichever comes first).

4 Empirical Analysis

This section explores the computational and statistical efficacy of the method for imputing tax and benefit payments described in Sections 2 and 3. All results reported here can be replicated using materials that are freely available for downloaded from the internet: see Appendix B for a step-by-step walk-through of the analysis.

4.1 Analytical approach

LINDA was used to project data for a "base scenario" using an internally programmed functional description specified to reflect UK tax and benefit policy applicable in April 2016 (see Appendix A.1 for details). A "counterfactual scenario" was then considered, specified to be identical to the base scenario in all respects, except with all rates of income taxes increased by 10 percentage points. Effects of increasing income tax rates simulated using the functional description for policy were then evaluated by subtracting summary statistics evaluated for the base scenario from the same statistics evaluated for the counterfactual scenario.

The analysis described above was then replicated, but with tax and benefit payments imputed using the database method described in this paper. Importantly, the databases used to impute taxes and benefits for the base and counterfactual scenarios were each constructed using the same functional descriptions for policy as referred to above.

Specifically, cross-sectional microdata generated by LINDA for simulated adults in an arbitrarily selected year were extracted from the base scenario projected using the functional description for transfer policy. Given these data, LINDA was used to evaluate net transfer payments for four alternative functional specifications: i) the base scenario in 2017; ii) the base scenario in 2057; iii) the counterfactual scenario in 2017; and iv) the counterfactual scenario in 2057. Each of these four projections for tax and benefit payments was saved, along with the associated population cross-sectional data, to a separate data file. LINDA was then directed to import databases comprised of two sets of data reflecting payments for 2017 and 2057, with the differences between them accounting for potential temporal trends as discussed in Section 2.3.2.

The analysis described above is designed so that disparities between projections evaluated using the database and functional descriptions for transfer policy reported below can be interpreted unambiguously as distortions associated with the database method. The focus of the discussion that follows is consequently not the simulated profiles *per sè*, but how the proposed method for projecting transfer payments performs, both in terms of simulation run-times and the influence of projected taxes and benefits on simulated profiles.

4.2 Transfer payments for a population cross-section

A series of test statistics for alternative population cross-sections were evaluated to explore the correspondence between net tax and benefit payments imputed from the functional and database descriptions for policy. Representative statistics from this analysis are reported in Table 4.1, focussing here on simulated data projected by LINDA for 2040. The year 2040 is interesting, because it is intermediate to the two years for which database descriptions for policy were explicitly supplied (2017 and 2057). The correspondence between the functional and database descriptions for tax and benefit payments evaluated for 2040 consequently reflect both the description of policy in the considered reference years (2017 and 2057), and the (geometric) interpolation methods applied between them (described in Section 2.3.2).

The statistics reported in Table 4.1 indicate a reasonably close correspondence between the measures of net transfer payments imputed using functional and database descriptions for the policy environment, for both the base and counterfactual policy scenarios. Correlation coefficients between the measures of net transfer payments are between to 0.97 and 0.98 for the full population cross-section and all population subgroups under state pension age. Correlation coefficients are smaller for the population over state pension age (67), but this is primarily attributable to the smaller net transfer payments identified for the older population subgroup, as is indicated by the remaining statistics reported in the table.

The mean and (sample) standard deviations for differences between the database and functional measures for transfer payments indicate that these differences are not significantly different from zero for any of the population subgroups at any appreciable confidence interval. Although mean absolute differences are higher for the counterfactual policy scenario, relative to the base, they are smaller when the respective statistics are expressed as percentages of the sample means for net taxes evaluated under the functional description for policy (left most column of the table).

Table 4.1: Correspondence between net tax burden imputed using functional and database descriptions for transfer payments of simulated population cross-section in 2040

	mean	correlation coefficient	mean difference	standard deviation of difference	mean absolute difference
base policy scenario					
full population	85.41	0.973	-4.88	67.61	32.45
single working aged no children	57.24	0.977	3.09	55.43	29.14
single working aged with children	62.78	0.973	11.15	98.05	48.80
couple working aged no children	245.66	0.968	-36.52	93.76	51.00
couple working aged with children	326.26	0.980	14.05	97.03	47.64
single pensioner	-4.66	0.876	-10.26	30.78	18.91
couple pensioner	16.23	0.750	-13.86	40.90	18.50
counterfactual scenario (all rates of income tax increased by 10 percentage points)					
full population	115.96	0.977	-7.79	78.48	36.80
single working aged no children	80.89	0.978	0.87	64.61	32.76
single working aged with children	103.37	0.979	9.49	106.01	51.52
couple working aged no children	311.02	0.968	-46.79	117.05	63.49
couple working aged with children	427.56	0.984	13.58	107.68	53.37
single pensioner	-1.95	0.883	-11.42	33.84	20.33
couple pensioner	20.12	0.731	-16.43	51.86	21.32

Source: Authors' calculations using simulated data
Notes: net tax burden defined as aggregate taxes less aggregate benefits excluding state pensions; working age adults between 18 and 66 years; pensioners aged 67 and over; children include all dependents aged 0 to 17 years; transfer payments measured in £ per week; "mean" reports sample arithmetic averages for net taxes evaluated using functional description for transfer payments; "difference" statistics report sample averages for net taxes evaluated using database description for transfer payments less net taxes evaluated using functional description for transfer payments.

4.3 Simulation run times

Each policy context projected using LINDA involves two discrete stages: evaluation of utility maximising behaviour for any feasible combination of individual specific characteristics, and population projections based on the evaluated behavioural solutions. Tax and benefit imputations are important for each of these stages. Table 4.2 reports disaggregated computation times for the base (status-quo) and counterfactual policy scenarios, for each method of imputing net transfer payments.

Table 4.2 indicates that, relative to the functional description for transfer payments, the database method for imputing net transfer payments required almost identical run-times, both to evaluate behavioural solutions, and to project panel data for the population, for both the base and counterfactual scenarios.[16]

[16] The base simulation involves projecting data 65 years forward and backward through time – 130 years in total – starting from the reference cross-section reported for the UK in 2017. In contrast,

Table 4.2: Simulation run-times by policy scenario and method of transfer payment imputation

	behavioural solution	population projection	total
net transfer payments imputed using database			
base	54.5	90.9	145.4
counterfactual	55.7	52.0	107.7
net transfer payments imputed using function			
base	54.3	87.0	141.3
counterfactual	54.4	53.3	107.7

Source: Authors' calculations using simulated data
Notes: "base" defines reference policy context; "counterfactual" defines policy context that is identical to "base" in all respects, expect that all income tax rates are increased by 10 percentage points. All times reported in minutes. Simulations run on workstation with dual Xeon E5-2670 processors and 96GB of RAM.

4.4 Base policy projections

As suggested in Section 2.4, simulating a base policy context, representing a "status-quo" scenario, is a common starting point considered for analysing the effects of a policy counterfactual. Although the statistics reported in Table 4.1 indicate that, all else equal, the two alternative methods for imputing tax and benefit payments considered here are similar, they are not identical. Here we explore the practical bearing that the alternative methods for projecting taxes and benefits had on simulated profiles under the base policy scenario, focussing on projections for birth cohorts born between 1980 and 1989 displayed in Figure 4.1.

The panels of Figure 4.1, taken together, display close similarities between projections based on functional and database methods for projecting tax and benefit payments. There are, nevertheless, some noticeable differences between the two sets of projections reported in Figure 4.1, which it is useful to discuss.

The top panel of Figure 4.1a displays a fairly close correspondence between the labour supply projected using the database and functional descriptions for tax and benefit payments for the ten birth cohorts born between 1980 and 1989. The most obvious disparities between the two series are reported from age 40: between ages 40 and 59 projected labour supply is appreciably higher using functional descriptions for transfer policy, and vice versa at higher ages. The differences between the two series peak at age 50, when labour supply projected under the functional description for transfer policy averages 36.9 hours per week, 2.5 hours higher than under the database description.

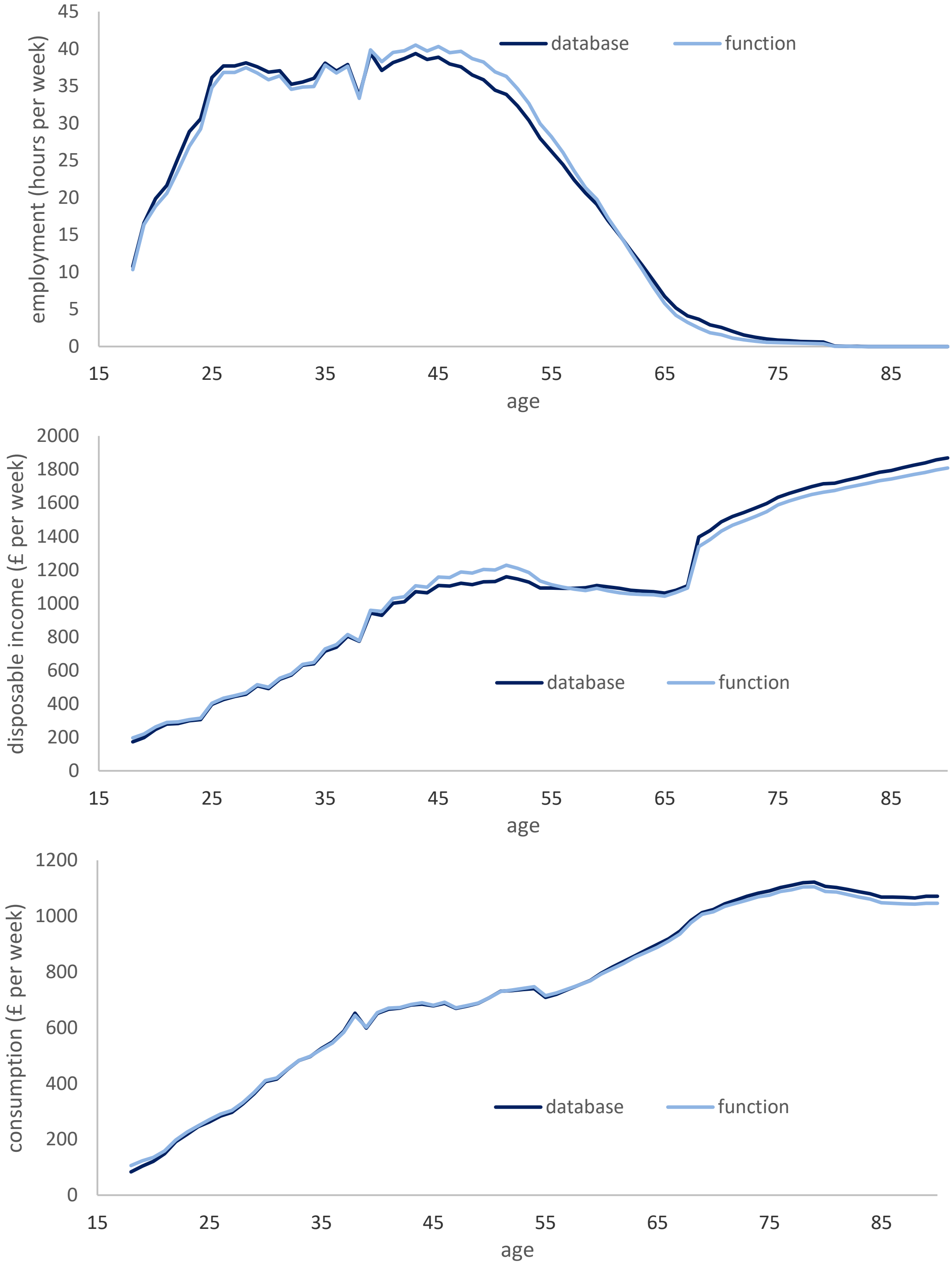
The differences in labour supply projections discussed above are also reflected by the statistics reported for disposable incomes in the middle panel of Figure 4.1a. Like labour, projected disposable income is almost identical under the two methods for projecting

the counterfactual simulation starts from the data generated under the base simulation, and re-evaluates data for the 65 years from 2017. This explains the difference in computation times reported for population projections under the "base" and "counterfactual" policy scenarios.

transfer payments until age 40, is appreciably higher between ages 40 and 59 under the functional method, and vice-versa from age 60. In contrast, measures of consumption projected using the two methods for projecting transfer payments (bottom panel of Figure 4.1a) are almost identical until age 70. From age 70, projected consumption reflects disposable income, with measures based on the functional description for transfer policy systematically lower than those based on database imputations.

Both panels of Figure 4.1b indicate slightly higher wealth accrual to age 60 under the functional description for transfer payments, relative to the database imputations, followed by a faster decline in pension wealth at higher ages. These statistics help to complete an underlying narrative: distortions attributable to imperfections in the database description of functional tax and benefit payments imply slightly stronger savings and work incentives during peak working years (40-60). The higher savings permit slightly earlier retirement, resulting in a faster drawdown of wealth, which is particularly evident for wealth held in private pensions. The faster drawdown of wealth, in turn results in lower disposable incomes, which support lower consumption later in life. Nevertheless, it is important to recognise that the scale of distortions identified here is generally slight.

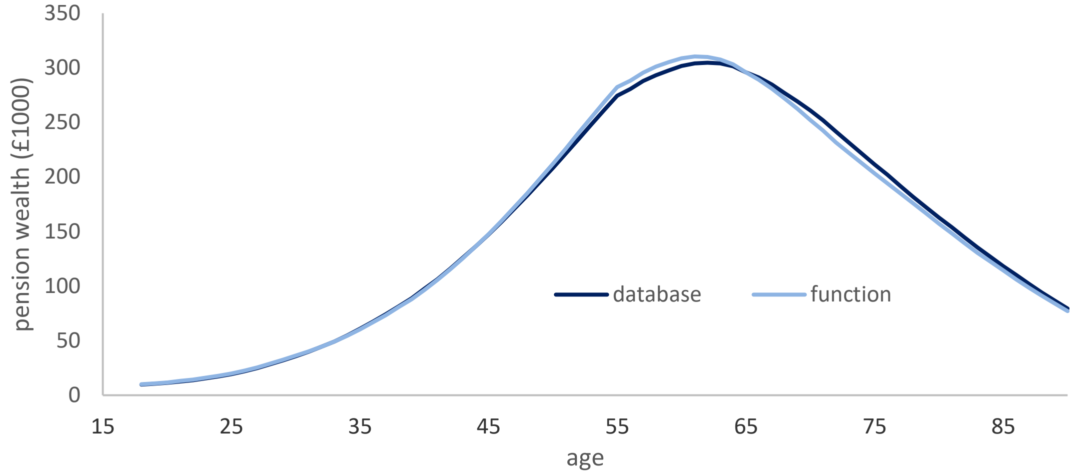
Figure 4.1a: Selected age-specific statistics projected under the base policy scenario for cohorts born between 1980 and 1989, by method of transfer payment imputation

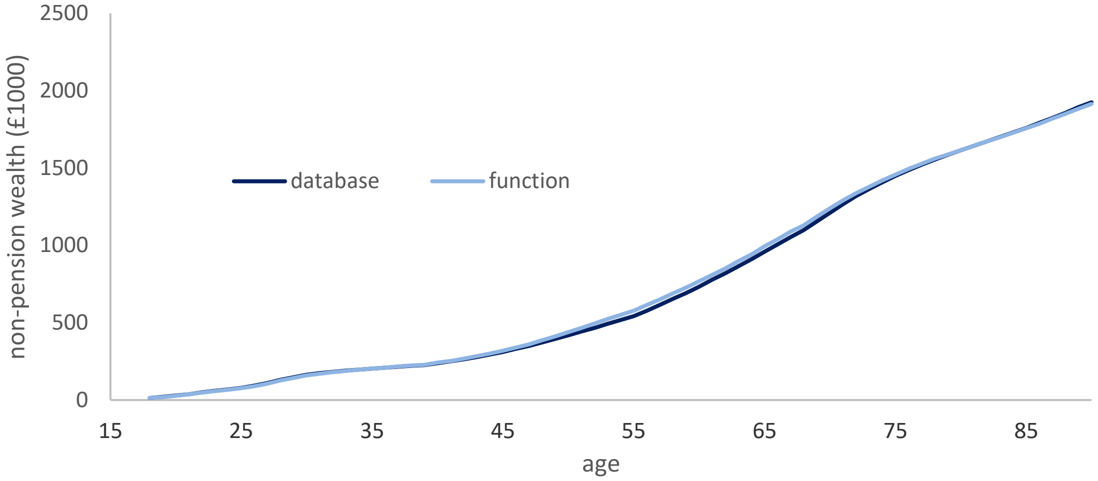


Source: Authors' calculations using simulated data

Notes: Age specific population averages evaluated for population cohorts born between 1980 and 1989. All financial figures reported in 2017 prices. The two series reported in each panel are distinguished by the method used to impute tax and benefit payments. The "base policy scenario" refers to an assumed status-quo policy environment.

Figure 4.1b: Selected age-specific statistics projected under the base policy scenario for cohorts born between 1980 and 1989, by method of transfer payment imputation





Source: Authors' calculations using simulated data

Notes: Age specific population averages evaluated for population cohorts born between 1980 and 1989. All financial figures reported in 2017 prices. The two series reported in each panel are distinguished by the method used to impute tax and benefit payments.

4.5 Projected effects of a policy counterfactual

This section reports the effects of a simulated policy counterfactual, in which all rates of income tax are increased by 10 per centage points from 2017. Comparisons of statistics for the counterfactual projections "in levels", as reported for the base policy projections in Section 4.4, are reported in Appendix C. These do not provide qualitatively useful information that adds to the discussion in Section 4.4. The current analysis consequently focusses on the effects of the policy counterfactual, as described by the difference between the counterfactual and base projections for policy, simulated using the two alternative methods for imputing tax and benefit payments.

Figure 4.2 displays projected effects of a 10-percentage point increase in all income tax rates for the same 10 birth cohorts considered in Section 4.4. These 10 birth cohorts, who were born between 1980 and 1989, were aged between 28 and 37 when the policy reform is assumed to take affect (2017). Echoing discussion in Section 4.4, Figure 4.2 displays age specific averages projected under the counterfactual policy scenario, less the same averages projected under the base policy scenario.

As discussed in Section 3.1, LINDA projects savings and employment decisions as though they are made to maximise expected lifetime utility. The rise in income tax rates consequently imply income and substitution effects underlying the projected behavioural responses. The income effects of the considered policy counterfactual tend to reduce consumption and increase labour supply. The substitution effects, in contrast, tend to reduce labour supply and have ambiguous implications for consumption, as leisure is made less expensive and saving more expensive relative to immediate (non-durable) consumption.

The top panel of Figure 4.2a indicates that the substitution effects tend to dominate labour/leisure responses, as employment is projected to fall under the higher income tax rates. Importantly, this is true for simulations based on both functional and database methods for projecting tax and benefit payments.

Nevertheless, it is notable that the projected falls in employment between ages 40 and 55 are appreciably larger when transfer payments are imputed using the database method, suggesting exaggerated substitution effects for labour supply. The larger falls in peak-working age employment projected using the database description for policy are, however, off-set by a more pronounced delay in the timing of retirement under the counterfactual, relative to the functional description for policy. The overall impact is a widening of the disparities of labour supply projected using the functional and database descriptions for transfer policy, which are identified for the base simulation in Section 4.4.

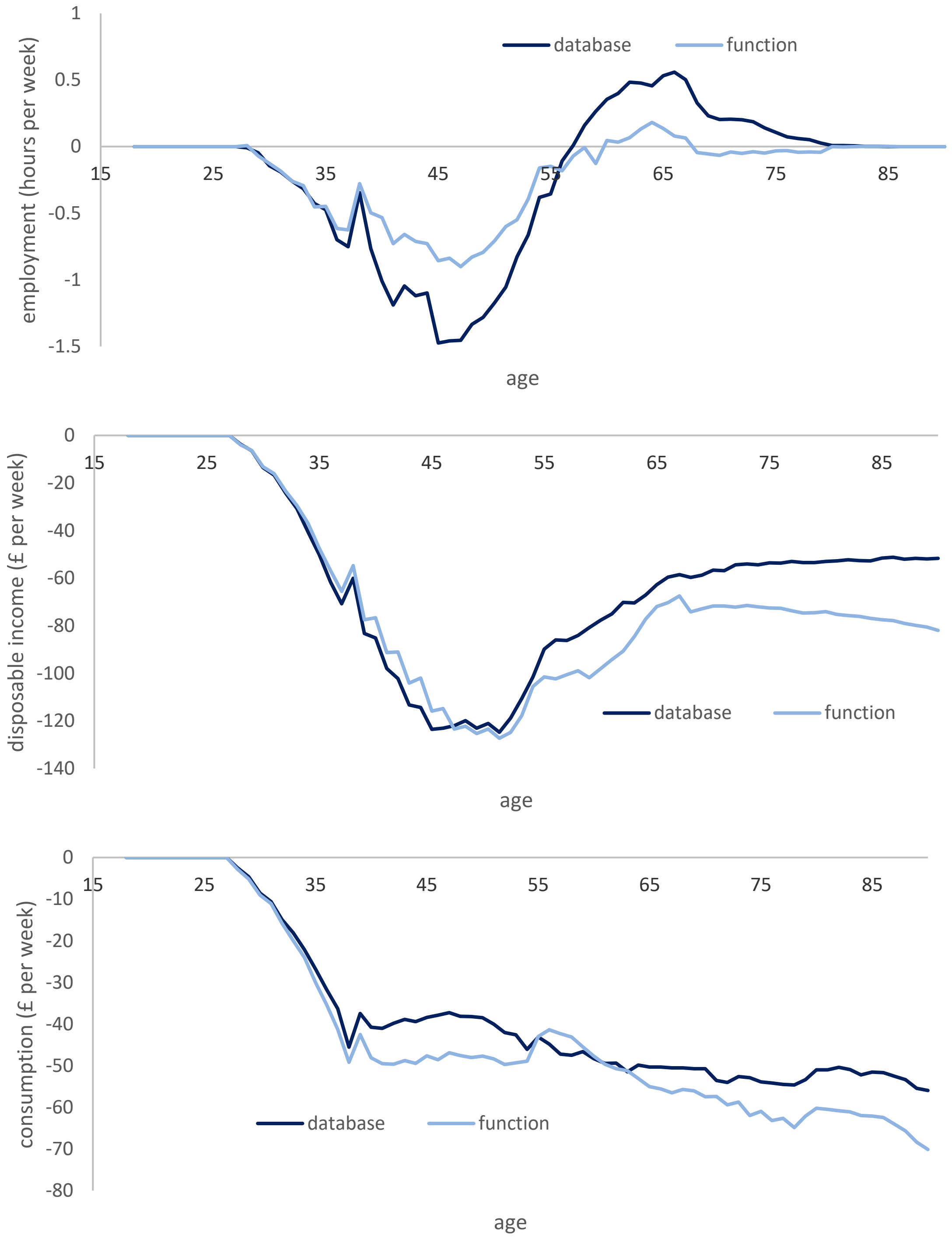
The middle panel of Figure 4.2a reveals very similar age profiles for the effects of the policy counterfactual on disposable income generated using the functional and database descriptions for policy. Disposable income is projected to fall appreciably from age 25 to 50, reflecting the increase in taxes implied by the rise in income tax rates, the projected declines in employment, and – later in life – falls in savings (discussed below). Although maintaining similar profiles, the largest differences between the database and functional projections for the effects of the policy counterfactual are observed from age 55. In this case, although both methods for projecting tax and benefit payments imply a levelling off of the fall in disposable incomes, the functional projections indicate leveling off at an appreciably lower level (£75 c.f. £50 per week).

The bottom panel of Figure 4.2a also indicates very similar age profiles for the effects of the policy counterfactual projected using the two methods for reflecting transfer policy; this time for consumption. Consumption is projected to fall steeply under the policy counterfactual to age 40, following an almost identical profile based on the functional and database descriptions for tax and benefit payments. From age 40, both methods for projecting transfer payments also suggest that the rise in income tax rates will depress consumption by a widening margin with age, but at a lower rate than to age 40. The smaller decline in disposable income projected for the counterfactual under the database description for policy is also seen to support a generally smaller decline in consumption.

Figure 4.2b also indicates broadly similar effects on wealth of the rise in income tax rates projected using the functional and database descriptions for policy. With regard to pension wealth, the top panel of Figure 4.2b indicates trough-shaped declines peaking late in the working lifetime, of between £5,000 and £10,000. With regard to non-pension wealth, projected declines under the counterfactual take a broadly sigmoidal form, are steepest late in the working life, leveling off at approximately £90,000 at age 70. Subject to these broad similarities, projections generated under the database description for transfer policy imply more pronounced declines in both measures of wealth to age 60, and less pronounced declines thereafter. These differences reflect associated statistics reported for labour supply and disposable income discussed above.

The overall impression made by Figure 4.2 is consequently that projected responses to the policy counterfactual derived from the functional and database descriptions for transfer policy are qualitatively similar. These qualitative similarities are complemented with broadly similar quantitative effects projected using the two methods for projecting transfer payments, where the latter are subject to some discernible differences in precise scale and timing.

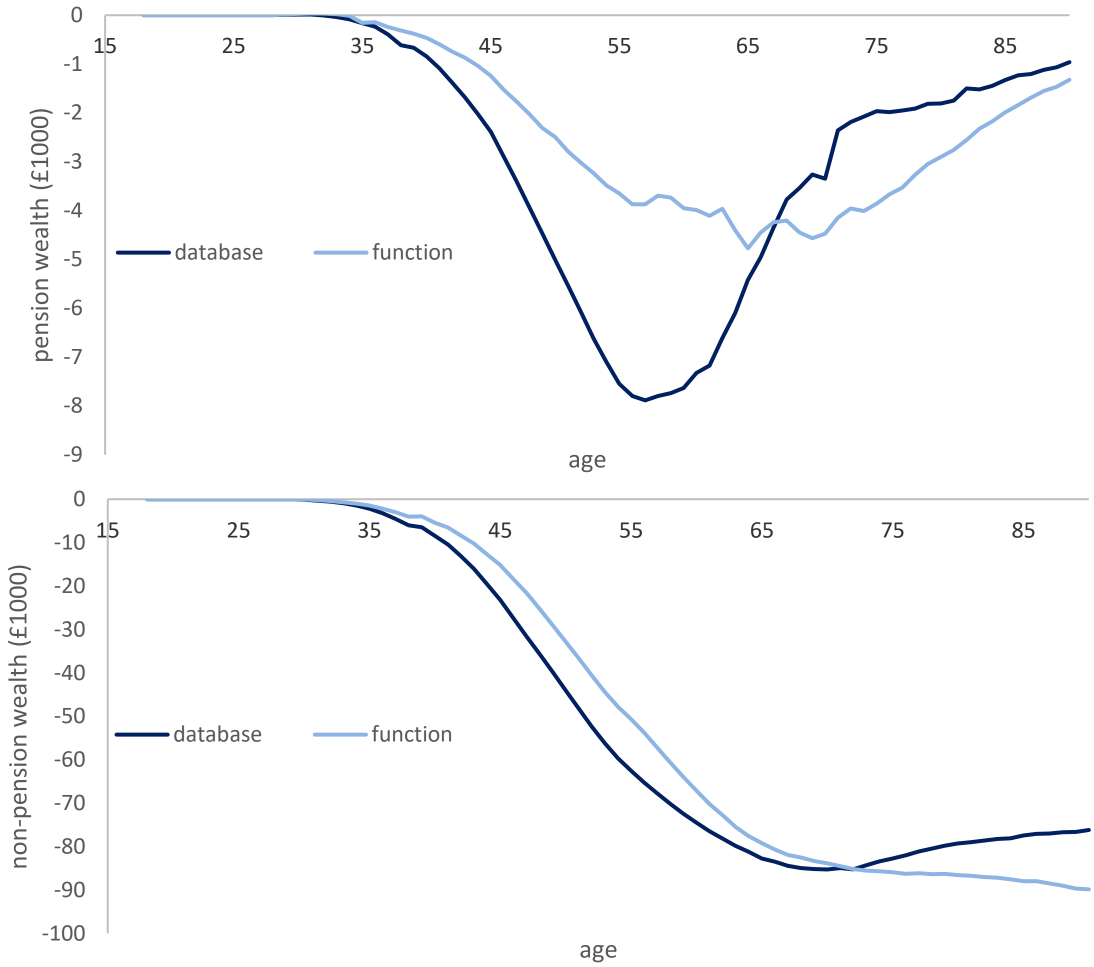
Figure 4.2a: Selected age-specific effects of a 10-percentage point increase in income tax rates on cohorts born between 1980 and 1989, by method of transfer payment imputation



Source: Authors' calculations using simulated data

Notes: Age specific population averages for population cohorts born between 1980 and 1989 projected under the policy counterfactual, less the same statistics projected under the base simulation scenario. All financial figures reported in 2017 prices. The two series reported in each panel are distinguished by the method used to impute tax and benefit payments.

Figure 4.2b: Selected age-specific effects of a 10-percentage point increase in income tax rates on cohorts born between 1980 and 1989, by method of transfer payment imputation



Source: Authors' calculations using simulated data

Notes: Age specific population averages for population cohorts born between 1980 and 1989 projected under the policy counterfactual, less the same statistics projected under the base simulation scenario. All financial figures reported in 2017 prices. The two series reported in each panel are distinguished by the method used to impute tax and benefit payments.

5 Conclusions

This paper describes a new method for imputing taxes and benefits in dynamic agent-based and microsimulation models, that draws upon existing third-party data sources, including (static microsimulation) tax-benefit calculators. A practical implementation of the approach is described for a model parameterised to the UK policy context, and simulated statistics indicate that the proposed method is capable of generating similar results – both qualitatively and quantitatively – to functional methods for imputing tax and benefit payments. These similar results were also obtained in similar simulation run-times.

The proposed method offers the possibility of outsourcing the technical challenges associated with reflecting the complexity and fluidity of modern tax and transfer systems, either to statistical agencies responsible for the collection of widely available survey microdata, or specialists in the construction of tax-benefit calculators. It is hoped that this will facilitate the development of more realistic dynamic (microsimulation and agent-based) models, thereby mitigating a potentially important source of simulation bias.

An advantage of the proposed method is that it can be easily adapted to reflect the influence of an expanded set of simulated characteristics on tax and benefit payments. Consider, for example, an existing dynamic model that used the proposed method to project transfer payments. Suppose that this dynamic model did not distinguish health status, but the reference database used to impute taxes and benefits payments did. Then, the influence on transfer payments of adding health status to the dynamic model would involve updating the matching routine used to impute transfer payments to distinguish by health status: it would not be necessary for the details of health-related tax and benefit programmes to be defined explicitly.

A caveat associated with the proposed method is that it cannot make-up for missing detail in a dynamic model structure. Specifically, a tax or benefit policy will only be suitable for analysis using the proposed approach if (1) it is reflected by the tax-benefit calculator, and its incidence is distinguished by both the (2) assumed matching method adopted for imputation and (3) the wider dynamic model. It is important that all three of these conditions is met for a policy to be eligible for analysis. In this regard, particular care should be exercised in relation to the matching methods employed, as these may be especially opaque to model users in some contexts.

Consider again, the case where a dynamic model did not distinguish health status, but the reference database used to impute taxes and benefits payments did. In this case, health-related benefits described by a tax-benefit calculator will appear as random variations in the tax and benefit payments projected by the dynamic model. These random variations will be indistinguishable from all other features that influence transfer payments in the tax-benefit calculator, but are unaccounted for by the assumed matching methods.

The emphasis on outsourcing in this paper contributes to a modelling paradigm that focusses on modular functionality, and toward the development of hybrid models more generally. Although the current study exclusively considers the problem of simulating tax and benefit payments, the suggested approach could conceivably be adapted to address a wide range of simulation problems, from relationship transitions, to fertility, and the evolution of health status.

References

Albertini, J., Fairise, X. And Terriau, A. (2021), "Health, wealth, and informality over the life cycle", *Journal of Economic Dyanmics and Control*, 129, 104170.

Ashraf, Q., Gershman, B., and Howitt, P. (2016), "How inflation affects macroeconomic performance: an agent-based computational investigation", *Macroeconomic Dynamics*, 20, pp. 558-581.

Assenza, T., Colzani, P., Delli Gatti, D. and Grazzini, J., (2018), "Does fiscal policy matter? Tax, transfer, and spend in a macro ABM with capital and credit", *Industrial and Corporate Change*, 27, pp. 1069-1090.

Atkinson, A.B. (1971), "The distribution of wealth and the individual life cycle", *Oxford Economic Papers*, 23, pp. 239-254.

Attanasio, O., Levell, P., Low, H. and Sánchez-Marcos, V. (2018), "Aggre3gating elasticities: intensive and extensive margins of women's labor supply", *Econometrica*, 86, pp. 2049-2082.

Attanasio, O.P. and Weber, G. (2010), "Consumption and saving: models of intertemporal allocation and their implications for public policy", *Journal of Economic Literature*, 48, pp. 693-751.

Attanasio, O.P. and Weber, G. (1995), "Is consumption growth consistent with intertemporal optimization?", *Journal of Political Economy*, 103, pp. 1121–1157.

Aydilek, A. (2013), "Habit formation and housing over the life cycle", *Economic Modelling*, 33, pp. 858-866.

Bohanon, C.E. (1983), "The tax-price implications of bracket-creep", *National Tax Journal*, 36, pp. 535–538.

Botta, A., Caverzasi, E., Russo, A., Gallegati, M. and Stiglitz, J.E. (2021), "Inequality and finance in a rent economy", *Journal of Economic Behavior & Organization*, 183, pp. 998-1029.

Browning, M. and Lusardi, A. (1996), "Household saving: micro theories and macro facts", *Journal of Economic Literature*, 34, pp. 1797–1855.

Caldwell, S. (1997). *Corsim 3.0 User and Technical Documentation*. Ithaca, New York.

Caiani, A., Russo, A., Gallegati, M. (2019), "Does inequality hamper innovation and growth? An AB-SFC analysis", *Journal of Evolutionary Economics*, 29, pp. 177-228.

Conesa, J.C., Li, B. and Li, Q. (2020), "Welfare implications of switching to consumption taxation", *Journal of Economic Dynamics & Control*, 120, 103991.

Dawid, H., Harting, P., van der Hoog, S., and Neugart, M. (2019), "Macroeconomics with heterogeneous agent models: fostering transparency, reproducibility and replication", *Journal of Evolutionary Economics*, 29, pp. 467-538.

Dekkers, G. and Van den Bosch, K. (2016). Prospective microsimulation of pensions in European Member States. In G. Dekkers and J. Mészáros (eds.), *Applications of Microsimulation Modelling*. Central Administration of National Pension Insurance: Budapest.

De Nardi, M., Pashchenko, S. and Porapakkarm, P. (2018), "The lifetime costs of bad health", *NBER Working Paper* 23963.

Dosi, G. and Roventini, A. (2019), "More is different ... and complex! The case for agent-based macroeconomics", *Journal of Evolutionary Economics*, 29, pp. 1-37.

Dosi, G., Fagiolo, G., Napoletano, M., Roventini, A. and Treibich, T., (2015), "Fiscal and monetary policies in complex evolving economies", *Journal of Economic Dynamics and Control*, 52(C), pp. 166-189.

Fernández-Villaverde, J. and Krueger, D. (2007), "Consumption over the life cycle: facts from consumer expenditure survey data", *Review of Economics and Statistics*, 89, pp. 552-565.

Flood, L. (2007), "Can we afford the future? An evaluation of the new Swedish pension system". In: Harding, A., Gupta, A. (Eds.), *Modelling our future: population ageing, social security and taxation*. Elsevier: Amsterdam, p. 33.

French, E. (2005), "The effects of health, wealth, and wages on labour supply and retirement behaviour", *Review of Economic Studies*, 72, pp. 395-427.

Giles, C. and McCrae, J. (1995), "TAXBEN: the IFS microsimulation tax and benefit model", *IFS Working Paper* W95/19.

Gourinchas, P.O. and Parker, J.A. (2002), "Consumption over the life cycle", *Econometrica*, 70, pp. 47-89.

Gustman, A.L. and Steinmeier, T.L. (2005), "The social security early retirement age in a structure model of retirement and wealth", *Journal of Public Economics*, 89, pp. 441-463.

Hubbard, R.G., Skinner, J. and Zeldes, S.P. (1995), "Precautionary saving and social insurance", *Journal of Political Economy*, 103, pp. 360-399.

Huggett, M. (1996), "Wealth distribution in life-cycle economies", *Journal of Monetary Economics*, 38, pp. 469-494.

Iacus, S.M., King, G. and Porro, G. (2009). "CEM: Software for coarsened exact matching", *Journal of Statistical Software*, 30, pp. 1-27.

İmrohoroğlu, A., İmrohoroğlu, S. and Joines, D.H. (1995), "A life cycle analysis of social security", *Economic Theory*, 6, pp. 83-114.

Keane, M.P. and Wasi, N. (2016), "Labour supply: the roles of human capital and the extensive margin", *Economic Journal*, 126, pp. 578-617.

Liégeois, P. (2021), "Combination of EUROMOD and LIAM2 tools for the development of dynamic microsimulation models: feasibility, example and conditions for sustainability of the linkage", *InGRID Final Report* D8.9.

Low, H.W. (2005), "Self-insurance in a life-cycle model of labour supply and savings", *Review of Economic Dynamics*, 8, pp. 945-975.

Low, H. and Pistaferri, L. (2015), "Disability insurance and the dynamics of the incentive insurance trade-off", *American Economic Review*, 105, pp. 2986-3029.

Morrison, R. (1998). *Overview of DYNACAN*. Retrieved on 16/02/2022 from https://www.actuaries.org/CTTEES_SOCSEC/Documents/dynacan.pdf

Nelissen, J.H.M. (1993), "Labour market, income formation and social security in the microsimulation model NEDYMAS", *Economic Modelling*, 10, pp. 225-272.

Nelissen, J.H.M. (1991), “Household and education projections by means of a microsimulation model”, *Economic Modelling*, 8, pp. 480-521.

Neugart, M., Richiardi M. (2018) “Agent-based models of the labor market”. In: Chen S.H., Kaboudan, M., Du, Y.R. (eds.). *The Oxford Handbook of Computational Economics and Finance*. Oxford University Press: Oxford.

Neveu, A.R. (2013), “Fiscal policy and business cycle characteristics in a heterogeneous agent macro model”, *Journal of Economic Behaviour & Organization*, 92, pp. 224-240.

O’Donoghue, C. and Dekkers, G. (2018), “Increasing the impact of dynamic microsimulation modelling”, *International Journal of Microsimulation*, 11, pp. 61-96.

Orcutt, G. (1957), “A new type of socio-economic system”, *Review of Economics and Statistics*, 39, pp. 116–123.

Orcutt, G.H., Caldwell, S. and Wertheimer, R.F. (1976). *Policy exploration through microanalytic simulation*. Washington: Urban Institute.

Richiardi, M. (2014). The missing link: AB models and dynamic microsimulation. In S. Leitner & F. Wall (Eds.), *Artificial Economics and Self Organization: Agent-Based Approaches to Economics and Social Systems*. Cham: Springer International Publishing.

Rust, J. (2008), Dynamic programming. In Durlauf, S.N. and Blume, L.E. (eds.), *New Palgrave Dictionary of Economics*. Palgrave Macmillan: New York.

Spielauer, M. (2013). *The LifePaths Microsimulation Model: An Overview*. Statistics Canada: Ottawa.

Spielauer, M., Horvath, T. and Fink, M. (2020), “microWELT: A dynamic microsimulation model for the study of welfare transfer flows in ageing societies from a comparative welfare state perspective”, *WIFO Working Papers* 609/2020.

Stuart, E. A. (2010), “Matching methods of causal inference: a review and a look forward”, *Statistical Science*, 25, pp. 1-21.

Sutherland, H. and Figari, F. (2013), “EUROMOD: the European Union tax-benefit microsimulation model”, *International Journal of Microsimulation*, 6, pp. 4-26.

Teglio, A., Mazzocchetti, A., Ponta, L., Raberto, M. and Cincotti, S. (2019), “Budgetary rigour with stimulus in lean times: policy advices from an agent-based model”, *Journal of Economic Behavior & Organization*, 157, pp. 59-83.

van de Ven, J. (2017a), “SIDD: An adaptable framework for analysing the distributional implications of policy alternatives where savings and employment decisions matter”, *Economic Modelling*, 63, pp. 161-174.

van de Ven, J. (2017b), “Parameterising a detailed dynamic programming model of savings and labour supply using cross-sectional data”, *International Journal of Microsimulation*, 10, pp. 135-166.

Appendix A: Code

Appendix A.1: Functional description of UK tax and benefit payments applicable in April 2016

Fortran code is provided in text file UK2016.f90.

Appendix A.2: Database imputations for taxes and benefits

Fortran code is provided in text file taxdb_comms.f90.

Appendix B: Analysis Walk-through

1. Download the LINDA quick-start guide from: https://www.simdynamics.org/index_htm_files/quick%20start%20guide.docx
    a. Note that LINDA is based on a model framework called SIDD
    b. LINDA can be downloaded from:
        i. https://www.simdynamics.org/download.html
2. Work through Sections 1.2 (Loading the model onto a new computer) and 1.3 (Extracting base data from the Wealth and Assets Survey)
3. Implement parameters to project transfer payments based on internally programmed functions for the UK in 2016
    a. Make a copy of MODEL\Job File.xls with name "Job File original.xls"
    b. Open MODEL\Job File.xls
    c. Worksheet "tax params"
        i. Copy columns AJ to AO
    d. Worksheet "input",
        i. Paste to columns AH to AM
        ii. Set cell AO2 to 7
        iii. Set cell AO3 to 8
        iv. Set cell AO5 to 0
4. Run base simulation assuming internally programmed functions for the UK in 2016:
    a. Job File.xls, worksheet "input",
        i. Revise preference parameters to adjust for altered tax system

1. Set cell Y3 to 0.982
2. Set cell Y4 to 0.972
3. Set cell Y10 to 2.3

ii. Set cell A2 to "base2017_fn" (without quotation marks)

iii. Set cell BQ35 to 1 (generate quintile statistics reported in Section 3.3.2)

b. Save

c. Run SIDD.exe

i. Note that this simulation takes appreciably longer to run than most others described below.

5. Create database for functional description of policy in base specification

a. Open MODEL\ANALYSIS_FILES\tax_test3.xls

b. Worksheet "analysis"

i. Set cell C2 to 2017

ii. Clear all data from row 7 in columns A to AA

c. Add input data in columns A to W with values generated by the model for 2017 under the "base2017_fn" simulation, implemented in step (4)

i. TIP: The steps described under 5.c.ii below require Excel to open large data files. Depending on your system, this may not be possible, and a short Stata do file is consequently provided with the appendix materials to facilitate extraction of the required data.

ii. Collate required data

1. Open a new (temporary) Excel file
2. Open file MODEL\SIMULATIONS\base2017_fn\age.csv
3. Copy rows 1 to 40000 from column 66
    a. In Excel, you can change to R1C1 format to see column numbers via the File > Options > Formulas > Working with formulas menu
4. Paste the data to cell A1 of Sheet1 of the temporary Excel file
5. Close file MODEL\SIMULATIONS\base2017_fn\age.csv
6. Repeat for data in columns B to Y of Sheet1 of the temporary Excel file, where:
    a. Data for column B are from file na.csv

b. Data for column C are from file nk.csv
c. Data for column D are from file nk_all1.csv
d. Data for column E are from file nk_all2.csv
e. Data for column F are from file nk_all3.csv
f. Set Cells G1 to G40000 to 0
g. Set Cells H1 to H40000 to 0
h. Set Cells I1 to I40000 to 1
i. Set Cells J1 to J40000 to 1
j. Data for column K are from file emp1.csv
k. Data for column L are from file emp2.csv
l. Data for column M are from file labinc.csv
m. Data for column N are from file peninc.csv
n. Data for column O are from file cpinc.csv
o. Data for column P are from file ppc.csv
p. Data for column Q are from file inv_inc1.csv
q. Data for column R are from file w.csv
r. Data for column S are from file hsgw.csv
s. Data for column T are from file hsgmd.csv
t. Data for column U are from file hsgret.csv
u. Data for column V are from file hsgmr.csv
v. Data for column W are from file comexhs.csv
w. Data for column X are from file psnno.csv (column 1)
x. Data for column Y are from file ben_unit.csv
y. Keep only rows where column X is equal to column Y
    i. TIP: You should be left with a sample of approximately 25000 observations.

7. Delete data in columns X and Y of the temporary Excel file

iii. Copy all remaining data from the temporary Excel file (or the Stata do file) to columns A to W of tax_test3.xls, analysis worksheet, starting at row 7.

iv. Set cell C3 of tax_test3.xls, analysis worksheet to the number of rows of data copied from the temporary Excel file

1. Note that the last row of data in tax_test3.xls, analysis worksheet should be equal to the value in cell C3 + 6

v. Close the temporary Excel file without saving (if necessary)

vi. Save tax_test3.xls and close file

d. Open MODEL\Job File.xls

e. Worksheet "input"

i. Set cell A2 to "temp" (without quotation marks)

ii. Set cell BQ55 to 1

f. Save

g. Run SIDD.exe

h. Open MODEL\SIMULATIONS\temp\tax_test3.xls

i. Worksheet "reference database"

i. Copy cells A6 to Z6 down to reference all rows of data included in the "analysis" worksheet

1. Note that the last row of data in the "reference database" worksheet should be two rows above the last row in the "analysis" worksheet

ii. Copy all data in the "reference database" worksheet, from row 4 to the last row in the worksheet, and from column A to Z

j. Open new workbook

k. Paste values to Cell A1

l. Create subdirectory: MODEL\TAX_DATABASE\base_fn

m. Save file as MODEL\TAX_DATABASE\base_fn\base_2017.csv

i. CSV format (not UTF8)

1. Note: if you receive a warning, proceed by accepting the format

n. Close all files

o. Open MODEL\ANALYSIS_FILES\tax_test3.xls

p. Worksheet "analysis"

i. Set cell C2 to 2057

q. Save file and close

r. Run SIDD.exe

s. Open MODEL\SIMULATIONS\temp\tax_test3.xls

t. Worksheet "reference database"

   i. Copy all data in the "reference database" worksheet, from row 4 to the last row in the worksheet, and from column A to Z

u. Open new workbook

v. Paste values to Cell A1

w. Save file as MODEL\TAX_DATABASE\base_fn\base_2057.csv

   i. CSV format (not UTF8)

x. Close all files

6. Set simulation from (4) as base

   a. Alt+F8

   b. Run SIDD macro

   c. Form 0 – specify new base using simulation "base2017_fn" (without quotation marks)

   d. Press the "CONVERT RUN TO NEW BASE" button

7. Run counterfactual 10% increase in all income tax rates using functional description for policy

   a. Open MODEL\Job File.xls

   b. Worksheet "input",

      i. Set cell A2 to "10pp_fn" (without quotation marks)

      ii. Set cell AI12 to 0.3

      iii. Set cell AI13 to 0.5

      iv. Set cell AI14 to 0.55

      v. Set cell AI106 to 0.55

      vi. Set cell AI107 to 0.55

      vii. Set cell AI108 to 0.55

      viii. Set cell BQ35 to 1 (generate quintile statistics reported in Section 3.3.2)

c. Save
d. Run SIDD.exe

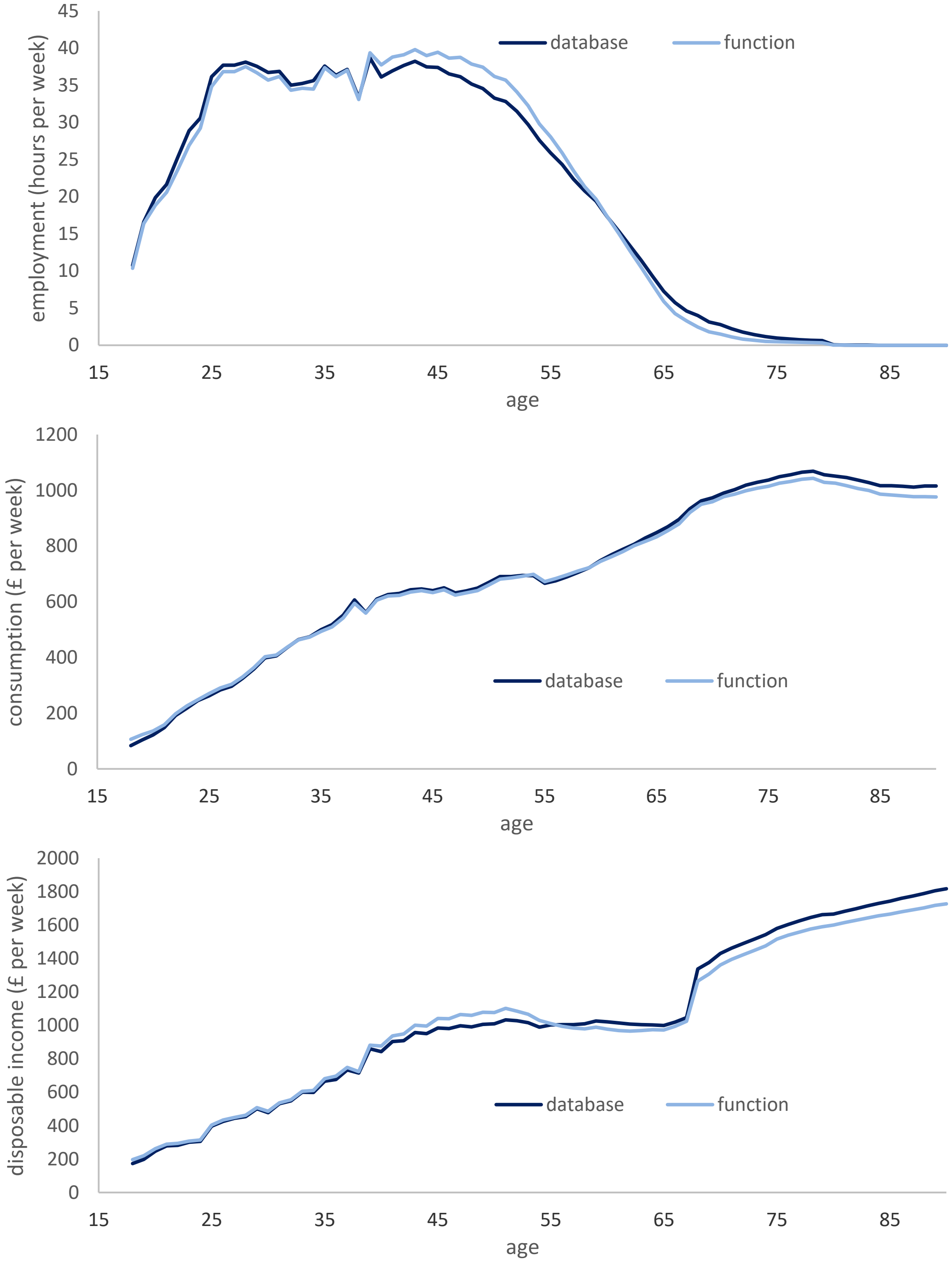
8. Create database for counterfactual 10% increase in all income tax rates using functional description for policy
    a. Open MODEL\ANALYSIS_FILES\tax_test3.xls
    b. Worksheet “analysis”
        i. Set cell C2 to 2017
    c. Save file and close
    d. Open MODEL\Job File.xls
    e. Worksheet “input”
        i. Set cell A2 to “temp” (without quotation marks)
        ii. Set cell BQ55 to 1
    f. Save
    g. Run SIDD.exe
    h. Open MODEL\SIMULATIONS\temp\tax_test3.xls
    i. Worksheet “reference database”
        i. Copy columns A to Z from row 4 to the end
            1. TIP – Ensure that all instances from “analysis” worksheet are included in copied data
    j. Open new workbook
    k. Paste values to Cell A1
    l. Save file as MODEL\TAX_DATABASE\base_fn\10pp_2017.csv
        i. CSV format (not UTF8)
    m. Close all files
    n. Open MODEL\ANALYSIS_FILES\tax_test3.xls
    o. Worksheet “analysis”
        i. Set cell C2 to 2057
    p. Save file and close
    q. Run SIDD.exe
    r. Open MODEL\SIMULATIONS\temp\tax_test3.xls

s. Worksheet "reference database"
    i. Copy columns A to Z from row 4 to the end
        1. TIP – Ensure that all instances from "analysis" worksheet are included in copied data
t. Open new workbook
u. Paste values to Cell A1
v. Save file as MODEL\TAX_DATABASE\base_fn\10pp_2057.csv
    i. CSV format (not UTF8)
w. Close all files

9. Implement parameters to project transfer payments based on database description for internally programmed functions for the UK in 2016
    a. Replace MODEL\Job File.xls with "Job File database.xls" included with this document
    b. Run SIDD.exe
        i. Note that this simulation takes appreciably longer to run than most others described below.
10. Set simulation from (9) as base
    a. Alt+F8
    b. Run SIDD macro,
    c. Form 0 – specify new base using simulation "base2017_db"
    d. Press the "CONVERT RUN TO NEW BASE" button
11. Run counterfactual policy using database description for policy
    a. Open MODEL\Job File.xls
    b. Worksheet "input"
        i. Set cell A2 to "10pp_db" (without quotation marks)
        ii. Set cell E36 to "10pp_2017" (without quotation marks)
        iii. Set cell E36 to "10pp_2057" (without quotation marks)
        iv. Set cell BQ35 to 1 (generate quintile statistics reported in Section 3.3.2)
    c. Save
    d. Run SIDD.exe

Appendix C: Analysis Supplementary Statistics

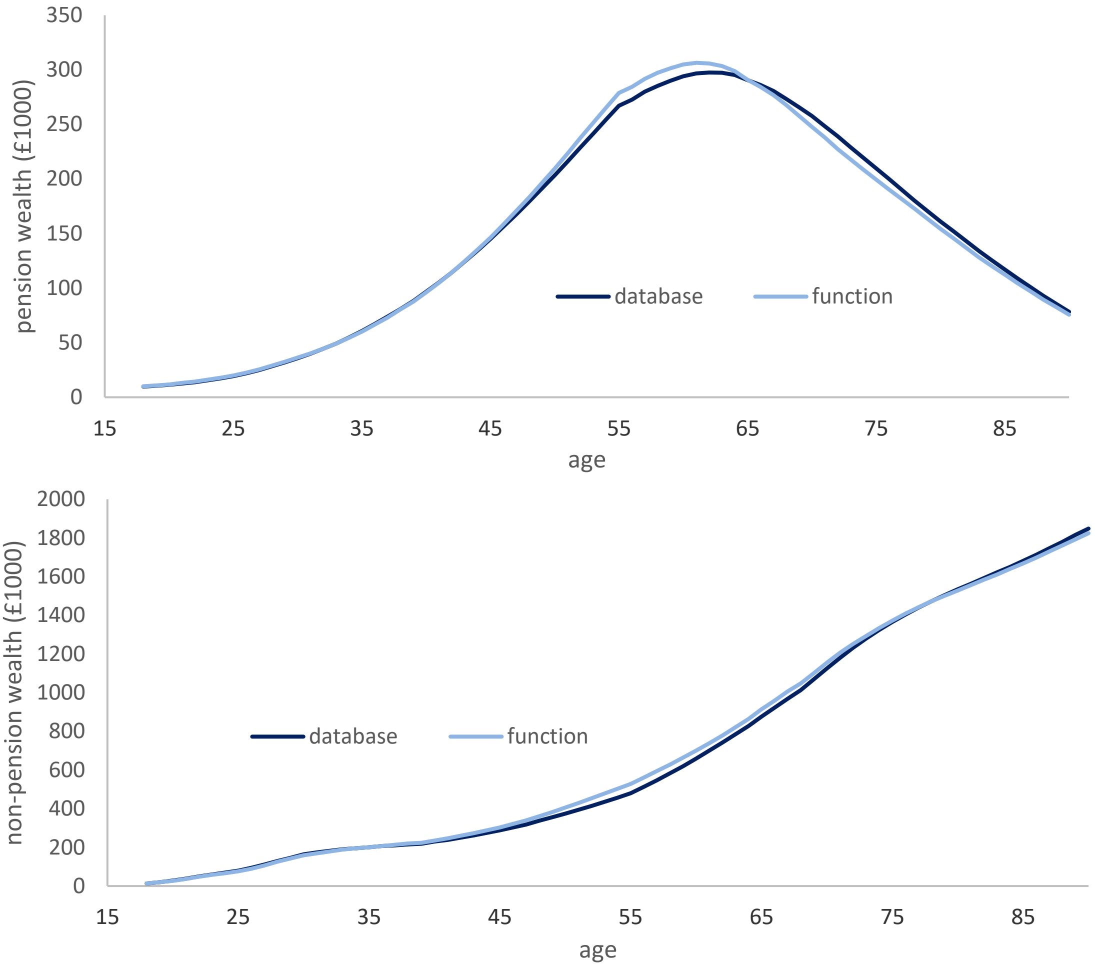
Figure C.1a: Selected age-specific statistics projected under the base policy scenario for cohorts born between 1980 and 1989, by method of transfer payment imputation



Source: Authors' calculations using simulated data

Notes: Age specific population averages evaluated for population cohorts born between 1980 and 1989. All financial figures reported in 2017 prices. The two series reported in each panel are distinguished by the method used to impute tax and benefit payments. The "base policy scenario" refers to an assumed status-quo policy environment.

Figure C.1b: Selected age-specific statistics projected under the base policy scenario for cohorts born between 1980 and 1989, by method of transfer payment imputation



Source: Authors' calculations using simulated data

Notes: Age specific population averages evaluated for population cohorts born between 1980 and 1989. All financial figures reported in 2017 prices. The two series reported in each panel are distinguished by the method used to impute tax and benefit payments.