Dynamic Simulation of Taxes and Welfare Benefits by Database Imputation

Justin van de Ven

Patryk Bronka

Matteo Richiardi

March 2022 – revised January 2025

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†
January 2025

Abstract

This paper proposes a new method for describing a realistic tax-benefit system in a dynamic microsimulation model. The suggested approach draws upon existing third-party tax-benefit calculators and provides an alternative to integration of tax functions within a dynamic structure. It has the advantages of adapting to the degree of household sector detail described by a model and permits taxes and benefits to be imputed directly from commonly available micro-data sources. A practical application to a dynamic programming model designed to reflect the contemporary UK policy context indicates that the proposed method for projecting taxes and benefits implies a comparable computational burden and generates qualitatively comparable results to a detailed functional description of fiscal policy, while simplifying model development.

1 Introduction

The scope and sophistication of models adopting a micro-approach for projecting household-sector dynamics has grown substantially during the past six decades. A noteworthy feature of such models is that they tend to use highly stylised functional forms to approximate the complexity of modern tax and benefit systems. Although this type of abstraction helps to simplify model specification, it also limits scope for contributing to public policy debate and exposes model projections to distortions that may be difficult to anticipate. This study demonstrates a new method that simplifies the task of projecting realistic tax and benefit payments in a dynamic heterogenous household model.

The modelling approach described in this study delegates the task of reflecting tax and benefit policy to widely available third-party data sources. This task can represent a very significant developmental hurdle in context of complex tax and benefit regulations, particularly in dynamic contexts where such regulations are also subject to frequent change. The suggested approach allows the modeller to leverage the expertise of policy specialists to obtain a detailed approximation of a tax and benefit system, which may be readily modified and shared between diverse model structures.

While our imputation method can be used in conjunction with any degree of household heterogeneity that is simulated, the precision of imputed taxes and benefits increases with simulated household detail. It is our hope that simplifying the problem of modelling the incidence of a tax and benefit system will facilitate increasingly detailed descriptions for households, improving the realism of simulated projections. Such a transition may prompt parallel advances in economic modelling, from macro agent-based models to life-cycle models, dynamic microsimulations, and dynamic programming models.

The suggested approach is demonstrated via practical application to a structural dynamic microsimulation model designed to reflect the UK policy context. Results are reported that show the method is capable of imposing a comparable computational burden, and generating similar simulated projections, to a detailed functional description of contemporary UK tax and benefits policy.

1.1 Scope for application

Economic studies based on dynamic projections of heterogeneity in the household sector often abstract from demographic detail. For instance, macroeconomic agent-based models (Axtell and Farmer, 2022) typically explore household-sector interactions with firms and banks via labour, capital, and product markets. In this context, agents in the household sector are typically distinguished by their skills/earnings potential, capital/savings, and the firms with which they interact; demographic characteristics are ignored.[1] With household demographics

---

[1] 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). Demographic

ignored, these studies often focus on after-tax incomes, or adopt highly stylised representations of existing tax and benefits policy.[2]

Household demographic detail is usually omitted from dynamic models with heterogenous agents 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 education, skills and wealth) include age, relationship status, dependent children and health. [3] Omitting such demographic characteristics can consequently detract from the practical relevance of results, particularly in context of seminal trends in cohabitation, fertility, and healthy life-expectancy.

Given the potential for demographics to influence key margins of economic decision-making, demographic characteristics are likely to feature more prominently 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). The order of the research programme outlined by Huggett (1996) is important, with "household structure" added before "institutional features", thereby suppling the demographic detail upon which tax and benefit payments commonly depend.[4]

---

characteristics (other than age) are also often abstracted in the literature that accounts for precautionary savings incentives in life-cycle frameworks; İmrohoroğlu *et al.* (1995), Hubbard *et al.* (1995), Huggett (1996), Gourinchas and Parker (2002), Low (2005), Aydilek (2013), and Conesa *et al.* (2020).

[2] Where a government sector is included for analysis, most macro-economic agent-based models adopt simple linear functions for net fiscal transfers with households; see 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. These models can be contrasted with those designed expressly to explore tax and benefits policy: see Orcutt et al. (1976), Caldwell (1997), Morrison (1998), Flood (2007), Nelissen (1991,1993), Spielauer (2013), and Dekkers and Van den Bosch (2016) for a review.

[3] 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 Webber (1995), Fernandez-Villaverde and Krueger (2007), and Gustman and Steinmeier (2005) for the importance of household demographics in explaining consumption patterns.

[4] An example is the contemporary literature that accounts for life-cycle precautionary savings incentives. 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.

The method for projecting tax and benefits payments outlined in this study is designed to facilitate point (3) of the research programme set out by Huggett (1996). The objective is to simplify the task of reflecting realistic tax and benefit systems in a way that can be adapted to increasing micro-detail reflected by a simulation model (points 1 and 2 of Huggett). Taken together, these features can then be used to improve our understanding of "market features" referred to in point (4) by Huggett, extending ultimately to the workings of the macro-economy. How do caring responsibilities feed into economic insecurity? What are the implications of alternative funding models for tertiary education on inequality of lifetime incomes? What bearing has the evolution of tax and transfer policy on employment incentives? Realistic descriptions of tax and benefit policy lie at the heart of many questions such as the above, that are of both academic interest and of vital public importance.

1.2 Integration with third-party tax-benefit calculators

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, non-behavioural microsimulation models), but it applies more generally to any micro-data source (including household surveys) that contain information on both gross and net incomes Well-maintained tax-benefit calculators are now publicly available for many countries[5], 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 (Richiardi, 2017), we are aware of just a handful of examples where a static tax-benefit calculator has been integrated with a dynamic microsimulation context (O'Donoghue, 2001, Blundell et al., 2016, 2021; Spielauer et al. 2020; Liégeois and Dekkers, 2014; Liégeois, 2021). 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. For instance, a static microsimulation model for taxes and benefits may not be equipped to reflect intertemporal policy variation or explore potential subjects of interest in a dynamic model. Likewise, concerns may arise regarding the on-going maintenance of the candidate static simulation model that exaggerate risks associated with long-term research investments in dynamic model structures.

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 (as in the dynamic programming literature – see Rust, 2008).

---

[5] For example, static tax-benefit calculators for the UK policy context include UKMOD (Richiardi et al., 2021), 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).

In a series of papers, Blundell and co-authors integrate dynamic programming models of household decision making with a stand-alone tax benefit calculator (Blundell *et al*., 2016, 2021). The tax benefit calculator (Fortax – see Shaw, 2011) provides a functional description of the UK tax and benefit system. The authors directly "embed" this calculator within their structural models to explore the incentive effects of changes to tax and benefits policy.

Embedding one model structure within another, as described by Blundell *et al*., is arguably the most obvious way of integrating functionality from alternative model frameworks. However, the feasibility of such an approach depends upon model architecture(s) and programming language(s).

Liégeois (2021) introduces an interesting alternative perspective to this approach by exploring the feasibility of augmenting a tax-benefit calculator (EUROMOD, see Sutherland and Figari, 2013) with dynamic model functionality based on the LIAM2 framework (de Menten et al. ,2014). 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) describe an approach for combining functionality from disparate models that avoids the complications discussed above. Specifically, Spielauer *et al*. document a continuous time dynamic microsimulation 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 the tax benefit calculator EUROMOD via static aging methods.

Our approach shares with Spielauer *et al*. (2020) the objective of integrating third-party inputs for projecting tax and benefit payments within a dynamic framework. In common with Spielauer *et al*. (2020), the approach that we propose requires no direct (i.e. run-time) communication between alternative model structures. Furthermore, in common with Spielauer *et al*., the approach that we propose uses one model structure to generate a database that is read as input by another model structure.

However, while Spielauer *et al*. (2020) use a dynamic model to "age" the microdata generated by a static tax-benefit calculator (by adjusting sample weights), we envisage using the output of a tax-benefit calculator to impute transfer payments for each agent at each point in time projected by a dynamic model. More specifically, we propose using a third-party source to obtain a reference database 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 used in a similar fashion

to a look-up table for evaluating transfer payments whenever needed, given assumptions about future policy developments.[6]

The approach that we suggest has a number of practical benefits. Relative to the approach of Spielauer *et al*., our approach facilitates the production of panel data rather than temporally dynamic cross-sectional snap-shots. As discussed above, the approach allows the model developer to delegate the task of capturing fiscal policy to dedicated specialists. Furthermore, the micro-data that we suggest be loaded as input into a dynamic model takes a standard form, facilitating reference to alternative static microsimulation models, in addition to commonly available survey microdata with the relevant income information.[7] The approach can also be adapted to reflect temporal transitions, including growth of tax and benefit rates and thresholds, which may be important in dynamic contexts: most simply, changes in the tax-benefit rules over time can be accommodated by loading multiple year-specific look-up tables.

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 need to impute tax and benefit payments in a simulation model is comparable to a missing variable problem. "Hot deck" imputation is a common approach used to replace "missing values of one or more variables for a non-respondent (called the recipient) with observed values from a respondent (the donor) that is similar to the non-respondent with respect to characteristics observed by both cases" (Andridge and Little, 2010, pp. 40-41). Part of the appeal of hot deck imputation is that it delivers imputed values that are guaranteed to be feasible (with respect to the survey population) and mitigates some of the risks of misspecification associated with regression-based methods. [8] It also facilitates *ex post* validation via comparison of the characteristics of donors and their recipients.

---

[6] Schofield et al. (2022) also follow a similar approach by matching simulated individuals in their dynamic microsimulation model of mitochondrial disease in adults with simulated individuals in STINMOD+, a tax-benefit model for Australia, based on a limited set of observable characteristics (age, sex, highest education attained and state of residence). Their objective, however, is to use the microsimulation model as a baseline to evaluate the economic impact of the disease, rather than informing the evolution of the dynamic structure itself.

[7] Nevertheless, use of a tax-benefit calculator does facilitate consideration of policy reform scenarios.

[8] Hot deck imputation does, however, depend upon important assumptions concerning the matching methods used to identify "donors" for "recipients".

2.1 Theoretical framework

Consider the functioning of a real-world tax-benefit system as a collection of mappings from the characteristics of the fiscal unit, X_i, to some (positive or negative) transfer, p_i:

$$p_i = f(X_i) \tag{1}$$

where i is the fiscal unit (time subscripts are omitted to simplify notation). Modelling tax-benefit payments typically involves proxying eq. (1) by:

$$\tilde{p}_i = \tilde{f}(Z_i) \tag{2}$$

where Z represents the model proxy for characteristics X and $\tilde{f}(.)$ is a simulation proxy for the real-world application of fiscal rules $f(.)$.

Projecting tax-benefit payments in a dynamic simulation model, D, involves coding eq. (2), where the arguments of the function are limited to those included in the model: $\tilde{p}_{D,i} = \tilde{f}_D(Z_{D,i})$. Similarly, a (static) tax-benefit calculator involves coding: $\tilde{p}_{S,i} = \tilde{f}_S(Z_{S,i})$.

Static tax-benefit calculators typically enjoy two key advantages, relative to dynamic simulation models, in reflecting real-world transfer policy. First, the set of characteristics used to impute transfer payments can generally be more flexibly chosen to reflect features of transfer policy, so that $Z_D \in Z_S \in Z$.[9] This point must be qualified, however, as dynamic models may generate characteristics, like contributions histories, that have an important bearing on transfer payments but are poorly described by cross-sectional micro-data.[10] Secondly, greater effort can be expended in a static tax-benefit setting in specifying assumed functions to capture real-world policy detail, such that $\left|\tilde{f}_S(.) - f(.)\right| \leq \left|\tilde{f}_D(.) - f(.)\right|$.

Our approach envisages using a tax-benefit calculator to compute $\tilde{p}_{S,i}$ for a reference dataset, and then assign tax-benefit payments in a dynamic model by using matching methods to draw on the reference dataset based on the common variables $Z_C \in (Z_D \cap Z_S)$. It should be clear from the above discussion that the ability of the suggested approach to capture realistic transfer payments will depend upon:

i) The comprehensiveness of the set of input variables considered by the static tax-benefit calculator, Z_S;

ii) The set of common variables used for imputation, $Z_C = Z_S \cap Z_D$;

iii) The realism of the static functional description, $\tilde{f}_S(.)$;

iv) The matching methods considered for imputation.

Concerning points (i) and (ii), where $Z_D \in Z_S$ (as discussed above), then $Z_C \equiv Z_D$ so that the approach we envisage suffers no loss of detail in reflecting the incidence of transfer payments. In this case, the envisaged approach will benefit to the extent that the tax-benefit calculator implements a more realistic function than would otherwise be implemented in the dynamic model; $\left|\tilde{f}_S(.) - f(.)\right| \leq \left|\tilde{f}_D(.) - f(.)\right|$ addressing point (iii). The final point (iv) concerns the

---

[9] Note that Z_D refers exclusively to variables endogenous to the dynamic model that are of use in projecting tax and benefit payments.

[10] One way to mitigate this type of issue is to adopt a hybrid approach for imputing transfer payments, where a database is used to impute "non-contributory" benefits, and functions are used to project the contributory component of a transfer system.

realism associated with methods for imputing missing data. Each of the steps of the suggested method are discussed in turn below.

2.2 Specification of the reference database

A common starting point for static taxes-benefit models is a publicly available microdata set. This dataset describes the set of input variables denoted Z_S in the preceding subsection. The input data include all of the information used by a static tax-benefit model to distinguish transfer payments between individual agents. "Agents" in this context typically refers to people, where the dataset also includes links defining a range of demographic relationships between them. These relationships commonly permit identification of nuclear families (benefit units) – between dependent children and their parents and between cohabiting spouses – and may also indicate where multiple nuclear families share the same accommodation (households).

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 for variables derived from the Living Costs and Food Survey, whereas the Policy Simulation Model (PSM, Department for Work and Pensions) and UKMOD (based on the EUROMOD framework) load in data from the Family Resources Survey. Static tax-benefit models typically generate a range of tax and benefit statistics for each agent represented in the starting microdata set, I, and report results in the form of an augmented data set, $(Z_{S,i}, \tilde{p}_i) \forall i \in I$.

The method we propose involves importing the augmented datasets generated by static microsimulation models – or equivalent survey data – for imputing tax and benefit payments in a dynamic model.[11] The augmented datasets generated by static tax benefit microsimulation models tend to share common structural features, which facilitates switching between them. Furthermore, most static models of taxes and benefits include tools that permit analysis of user-defined case-studies (Hufkens *et al*., 2019), which may be used to tailor the starting database to key subjects of interest.

2.3 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 (donor) individuals in the reference database that possess comparable characteristics to (recipient) simulated agents, in

---

[11] Using a static-tax benefit calculator to impute transfer payments rather than drawing directly from a suitable survey micro-dataset has the advantage of facilitating consideration of policy counterfactuals that are often a focus of interest for dynamic microsimulation models.

[12] These methods range from functional regression specifications, through non-parametric descriptions, to pair-wise matching methods. For studies that apply regression methods to impute net tax and benefit payments, see Bargain *et al*. (2013), Biewen & Juhasz (2012), and Frenette *et al.* (2007).

common with hot deck imputation and the KNN (K-Nearest Neighbours) machine learning algorithm.

Best practice requires that the assumed matching method 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.

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 status, and retirement), suggesting the use of exact matching methods over selected dimensions. Nevertheless, the method should also guarantee that a match is obtained for all conceivable simulated circumstances, which motivates interest in coarsened exact matching (Iacus *et al.*, 2009).

Furthermore, the method should seek to limit differences in relation to a set of selected continuous variables. This last point 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 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 consequently limit exposure to "high influence" observations (outliers).

In response to the above considerations, Spielauer *et al.* (2020) apply coarsened exact matching of discrete valued characteristics to adjust the weights of a database derived from a static tax-benefit model. This involves matching each observation in a dynamic simulation to an observation in the reference database that has the same set of discrete characteristics.

An advantage of this approach is that it can be readily tailored to the level of detail described by a dynamic model, and adapted as the simulated detail is expanded.

Furthermore, the approach permits considerable pre-processing to be conducted on the considered database, which helps to limit the computational overhead associated with the matching process. Specifically, it is possible to "pre-categorise" each observation in the reference database with respect to the feasible combinations of the discrete valued characteristics considered for matching. This implies that matching a dynamically generated individual involves identifying their respective subgroup, and then selecting an observation from the pre-categorised database observations of the respective group.

Here we propose extending on Spielauer *et al.* (2020) in two ways. We require a matching process that is guaranteed to identify a match for all individuals projected in a dynamic simulation while at the same time not wanting to sacrifice the degree of detail used for the matching process. Our first innovation to Spielauer *et al.* is designed to achieve this balance by using stratified coarsened exact matching over discrete valued characteristics. That is, we suggest use of multiple "levels" of exact matching, starting at the most fine-grained (discrete) set of matching criteria, and then only proceeding to more coarse grained matching criteria if a match is not otherwise identified.

Our second innovation to Spielauer *et al.* (2020) recognises that continuous valued variables, like private income, are important for determining tax benefit payments in addition to discrete valued variables. Spielauer *et al.* do not account for continuous variables as their model did not project these. That is not the case in dynamic microsimulation models more generally, and we recommend nearest-neighbour matching where continuous variables that are relevant for determining tax and benefit payments are projected by a dynamic model.

Where the set of continuous variables that could be included for matching extends beyond a single characteristic (e.g. formal childcare costs in addition to private income, and/or income splitting), then we envisage use of a method to reduce the dimensionality of the problem to a single distance. Here we suggest use of Mahalanobis distance matching (MDM – see Rabin, 1980), although it may be noted that Euclidean distance is a more popular method considered elsewhere in the literature (e.g. the KNN algorithm).

The computational overhead associated with matching continuous valued variables can also be economised in the pre-processing stage that is referred to above. Specifically, the tax database observations within pre-categorised subgroups can be stored in ascending order of the continuous variables included in the matching process. 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).

2.4   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 (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. The section finishes with discussion of how the method can be adapted to account for policy variation beyond the scope of a reference database.

2.4.1   Imputations at a point in time
Having matched the characteristics of a benefit unit 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 continuous variables, including private (pre-tax and benefit) incomes. Suppose, for example, that the set of continuous variables is limited to private income, and that a benefit unit in the dynamic model, denoted *A*, is projected to have private income worth £531 per week. Suppose also that the closest observation to this income described by the reference database (after coarse-exact matching), denoted *B*, was £535 per week. How should disposable income be imputed from the reference database to account for the £4 per week difference in private incomes?

One approach would be to match each benefit unit in the dynamic model to multiple units in the reference database and use interpolation methods for imputation. In the above example, the next closest observation described by the reference database, denoted *C*, might have private income of £525 per week (£6 per week less than our target individual). In this case, using linear interpolation would equate disposable income for unit *A* to 0.6 times disposable income of observation *B* plus 0.4 times disposable income of observation *C*.[13]

We reject the interpolation approach described above in favour of an approach that permits each simulated unit to be matched to a single unit in the reference database. The motivation for this preference is that it facilitates post-simulation diagnosis and validation of imputed transfer payments.[14]

Given the above, the problem is how to impute taxes and benefits for a simulated unit from the taxes and benefits reported by a reference database for a (single) matched unit 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 $\tilde{p}_i$ denote, respectively, the private income and transfer payment of individual i, then $\tilde{p}_D = \frac{\tilde{p}_S}{x_S} x_D$, with D denoting the unit simulated in the dynamic model and S the respective matched unit from the reference database.

As x tends toward zero, the fact that $\tilde{p}/x$ tends toward infinity 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, appropriately uprated to consider time differences between the static and dynamic models; that is $\tilde{p}_D = \tilde{p}_S$. Otherwise, the ratio method described in the preceding paragraph is used.

2.4.2 Accommodating policy variation through time

If the policy environment is assumed constant, the use of a single tax database as described above is sufficient to model the evolution of taxes and benefits through time.[15] However, our method also accommodates (real) policy variation over time. It is (conceptually) possible to

---

[13] Alternative interpolation methods could be used. Cubic interpolation, for example, has the advantage of implying a continuous differential at the expense of requiring four database observations for computation (e.g. Keys, 1981).

[14] 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. Validation might also take into account the anticipated incidence of referencing alternative database observations with respect to their associated weights (as pointed out by an anonymous reviewer).

[15] This includes when the system is fixed in either real or nominal terms.

consider a separate reference database for an arbitrarily large number of discrete time periods. This approach provides the analyst with tight control over the assumed temporal evolution of tax and benefit policy.[16]

Between the two extremes of using the same reference database for all simulated years and a different reference database for each year, intermediate solutions are also possible that trade off flexibility for speed of execution. In simulation contexts where execution speed remains a concern[17], changes between policy systems can be implemented as discrete jumps (i.e. the most recent policy system is used, with a step change when a new policy system is introduced), or smoothed using interpolation methods. Suppose, for example, that $y_{i,t}$ denotes disposable (post-tax and benefit) income of unit 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 have:

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

This approach can 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.[18] An added advantage of this limitation is that it would facilitate post-simulation diagnosis of imputed transfer payments, as each unit in the dynamic model would be matched to the same observation in both reference databases.

2.4.3 Accommodating policy variation beyond the scope of a reference database

Many dynamic simulations are currently concerned with features of the evolving economy that may not be accommodated in a reference database. The incidence and costs of long-term care in an aging population, costs of aging public infrastructure, and environmental degradation are all currently subjects of keen public interest, and effects of interest are largely beyond the scope of most existing static tax-benefit models.

Where analysis focusses on fiscal policy that is beyond the scope of a reference database, then it remains possible to augment transfer payments imputed from the database with internally programmed functions that represent the subjects of interest. Suppose, for example, that we are interested in the fiscal reforms necessary to support the costs of long-term care. We have access to a static tax-benefit model (e.g. EUROMOD) that provides a detailed

---

[16] For instance, EUROMOD-based tax-benefit models facilitate creation of policy scenarios for future years, including automated uprating of input data. The resulting output datasets can then be imported in the dynamic model, and the model directed to use the most recent dataset for any given simulated year.

[17] Computation time remains a significant barrier to some classes of dynamic microsimulation model, including those that use dynamic programming methods to project behaviour through time. In such contexts, tax and benefit payments may need to be evaluated for many more sets of circumstances than are actually represented in simulation output, due to the methods used to identify utility maximising decisions.

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

description of the direct tax and transfer system, but omits long-term care costs. Suppose also that we are able to obtain functions that provide plausible projections for the costs of long-term care.

In this context, we can use a database derived from the static tax-benefit model to project much of the tax and transfer system, and supplement this with the functional descriptions for the costs of long-term care. Given the projected costs of long-term care, we can then explore alternative fiscal reform packages capable of supporting the long-term care costs. These reform packages could be explored by obtaining alternative counterfactual databases from the static tax-benefit model, or from additional functions devised specifically to reflect the fiscal reforms of interest. The point is that use of a database to impute taxes and benefits does not preclude the use of dedicated functions to reflect particular fiscal policies of interest.

3   Practical Implementation

LINDA is a dynamic microsimulation model designed to reflect the contemporary UK policy context. The model includes a detailed functional description of the UK tax-benefit system in April 2016 including assumptions for how benefit rates and tax thresholds evolve through time.[19] The current section describes how LINDA was adapted to undertake a validation analysis of the proposed method for projecting tax and benefit payments, which is reported in Section 4.

Following a standard validation strategy, we assume that the functional description of tax and benefit policy included in LINDA is correctly specified, and that the distributions of taxes and benefits simulated using the model are "pseudo-true". We then used data generated by the model based on the functional description for taxes and benefits to create a tax database. This database reflects the type of data that could be obtained from a standard tax-benefit microsimulation model, such as UKMOD. Referring to equations (1) and (2) in Section 2.1, $f(X_i) = \tilde{f}_D(Z_{D,i}) = \tilde{f}_S(Z_{S,i}) = p$, and $X_i = Z_{D,i} = Z_{S,i}$. Finally, LINDA was amended to project tax and benefit payments based upon the tax database using the methodology outlined in Section 2.

The remainder of this section provides some further technical detail of LINDA and how the methodology described in Section 2 was practically implemented. Section 4 returns to describe how the implementation described here was used to validate the proposed method for projecting transfer payments.

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

---

[19] Associated programming code can be accessed as described in Appendix A.1.

concern.[20] 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

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

---

[20] The model is publicly accessible, and free to download from www.simdynamics.org. See also van de Ven (2017b,c).

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

As discussed in Section 2.3, multiple "levels" of matching are used to balance the countervailing objectives of obtaining matches between simulated and database observations that are as similar as possible, while also guaranteeing that a match is obtained for all observations generated by a dynamic microsimulation. 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. This reduction in detail considered for the matching process helps to match a wider set of observations projected by the dynamic model, at the expense of accepting greater diversity between matched observations. 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 contribution and wage histories, health status, care needs, and disability.[21] 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

---

[21] In more recent work, the authors take into consideration a wider range of characteristics for the matching process, including carer status, receipts and costs of social care, and income splitting; see van de Ven *et al*. (2024) for details.

the model. The effects of these features will consequently be indistinguishable from (unexplained) imperfect benefits take-up, tax avoidance, and tax evasion.

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 (due, for example, to imperfect take-up rates, tax avoidance, or tax evasion).

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, 14). Then, only individuals in the matched subgroup with incomes in the set of four income values (7, 9, 11, 12) would be included in the candidate pool.[22] 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.[23]

Step 3: Given the candidate pool and associated selection weights as described in step (2), LINDA is adapted to use one of two methods to impute transfer payments. It can include all candidates for the imputation, with the contribution of each candidate defined by their

---

[22] Those with incomes of 6 and 14 omitted because their income values exhibit greater difference to the target individual than four other income values observed in the group identified from step 1.

[23] 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 candidates' intermediate weights, equal to 10 = 1 + 3 + 3 + 1.5 + 1.5.

respective weight. Alternatively, it can select a single donor by random draw for each individual at each point in time, with probabilities of selection reflecting each candidate's selection weight.

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 Validation

As previously mentioned, we assume that the specification of LINDA based on a functional description of tax and benefit payments is correct, and that associated simulated data derived from the model are "pseudo-true". Given these assumptions, we use projections from LINDA to create a tax database as described in Section 2.1. This tax database mimics information that could be derived from survey data or a standard tax-benefit calculator, such as UKMOD.

4.1 Analytical approach

The starting point for the current analysis is a functional description of tax and benefit payments that was implemented in LINDA. This function was designed to reflect UK policy applicable in April 2016, focussing on variation of payments with respect to the number and age of household members. Tax thresholds and retirement benefits were also assumed to evolve through time in line with real wage growth (1.29% per annum, see Appendix A.1 for details). Given this specification for tax and benefit payments, LINDA was used to simulate panel data for adults in the evolving UK population cross-section between 2017 and 2081. This simulation is referred to below as the "function base scenario".

After evaluating the function base scenario, a "function reform scenario" was simulated. The function reform scenario was specified to be identical to the base scenario, except that all rates of income tax were increased by 10 percentage points. Effects of this reform were then evaluated by subtracting summary statistics evaluated for the function base scenario from the same statistics evaluated for the function reform scenario.

The above analysis was then repeated under identical assumptions, apart from replacing the internally programmed tax and benefit function with database approximations as described in Sections 2 and 3. This involved identifying reference databases for two arbitrarily selected years (2017 and 2057) for each tax and benefit function (base and reform). Two reference databases were required for each tax (and benefit) function to capture the real growth assumed for tax rates and retirement benefits, following the approach described in Section 2.3.2.

All four reference databases (base and reform, 2017 and 2057) were populated from the panel data generated assuming the "function scenarios" described above. For example, the database for 2017 used to approximate the function base scenario was populated with data generated for 2017 under the function base scenario.

This analysis was carefully designed to facilitate evaluation of the database method for imputing tax and benefit payments. First, imputing tax and benefit payments using databases populated with data derived from specific tax and benefit functions permits clear conclusions to be drawn. That is, any differences between projections generated using the function and database descriptions for tax and benefit payments can be interpreted unambiguously as distortions associated with the database method. Such differences arise because of imperfect matches between the simulated agents and records described by the database, and to the variables used for matching donors and recipients, which do not include all of the arguments of the considered tax functions (Section 2.1, point *iv*).

Secondly, the fact that the assumed tax function allows for trend variation in tax thresholds and retirement benefits allows the current analysis to assess the efficacy of the method outlined in Section 2.3.2 for accommodating temporal variation using the database approach.

Finally, use of a structural model of decision-making to consider the effects of a policy reform permits an assessment of the ability of the database method to capture behavioural incentives. Such incentives are interesting to consider here because they are likely to reflect margins of policy detail of interest.

The current validation exercise consequently focusses on the ability of matching methods to reflect a known functional description for transfer payments. This focus concerns the key relative weakness of the proposed methodology, relative to functional descriptions for transfer policy. It is consequently important to bear in mind that the current analysis does not concern the relative advantages that motivate the proposed methodology: greater flexibility in defining input data for a dedicated static tax-benefit model (points *i* and *ii* in Section 2.1), and delegation of functional representation of evolving tax and benefit policy to policy specialists (*iii* in Section 2.1).

Simulation run-times

Relative to the functional description for transfer payments, the database method for imputing net transfer payments required almost identical run-times, for both the base and reform scenarios. Related statistics are reported in Appendix C.1.

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 2037. The year 2037 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 2037 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 reform scenarios. Correlation coefficients between the measures of net transfer payments are between to 0.93 and 0.98 for the full population cross-section and all population subgroups, except for those under state pension age and working couples with no children.[24]

Correlation coefficients are smaller for the population over state pension age (67) primarily due to the smaller net transfer payments identified for this population subgroup. In the case of working couples without children, the matching methods considered here are complicated by income splitting that is important in the UK, where taxes are levied on individual incomes. In related work, we have consequently extended the matching routine considered here to account for individual incomes via the incorporation of Mahalanobis distance metrics, as discussed in Section 2.3 (see Bronka *et al.*, 2023).

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 reform 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).

---

[24] Note that contributory benefits, which may be poorly reflected by the assumed matching methods, do play a role in the UK tax benefit system, but this role has declined appreciably over recent decades.

Table 4.1: Correspondence between net tax burden imputed using functional and database descriptions for transfer payments of simulated population cross-section in 2037

	mean	correlation coefficient	mean difference	standard deviation of difference	mean absolute difference
base policy scenario					
full population	84.26	0.930	-12.07	98.68	36.93
single working aged no children	44.63	0.960	1.01	58.10	23.27
single working aged with children	84.73	0.971	3.50	104.34	55.34
couple working aged no children	212.75	0.816	-72.01	191.77	85.24
couple working aged with children	307.99	0.957	-0.32	128.66	57.83
single pensioner	2.39	0.807	-13.52	45.88	22.00
couple pensioner	33.36	0.786	-22.34	70.08	26.08
reform scenario (all rates of income tax increased by 10 percentage points)					
full population	105.64	0.936	-17.34	112.70	43.10
single working aged no children	57.57	0.947	-1.67	66.99	26.96
single working aged with children	121.36	0.974	-2.28	119.98	62.81
couple working aged no children	245.93	0.868	-87.95	214.95	101.98
couple working aged with children	384.35	0.961	-10.82	153.67	68.49
single pensioner	5.71	0.860	-14.67	47.00	23.86
couple pensioner	41.68	0.847	-27.79	82.25	32.27

Source: Authors' calculations using simulated data

Notes: working age adults between 18 and 66 years; pensioners aged 67 and over; children include all dependents aged 0 to 17 years; all income measured in £ per week; "mean" reports sample arithmetic averages for net taxes evaluated using functional description for transfer payments; "correlation coefficients" report correlations between net tax payments evaluated using functional and database descriptions for transfer payments. "mean 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. "standard deviation of difference" statistics report sample statistics for "mean differences". "mean absolute difference" statistics report sample averages for absolute differences between net taxes evaluated using functional and database descriptions for transfer payments.

4.3 Base policy projections

Simulating a base policy context, representing a "status-quo" scenario, is a common starting point considered for analysing the effects of a policy reform. 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.

Figure 4.1 displays age specific arithmetic averages for selected characteristics of simulated adults born between 1980 and 1989. Consideration of multiple birth cohorts increases the sample of individuals contributing to the reported statistics. This helps to filter-out noise in the projected profiles: projections obtained using data of single-year birth cohorts are qualitatively similar to those reported here. To further clarify, averages displayed in Figure 4.1 for age 40 are based on simulated data for year 2020 for the oldest contributing birth cohort (born in 1980), and 2029 for the youngest birth cohort.

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 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 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.
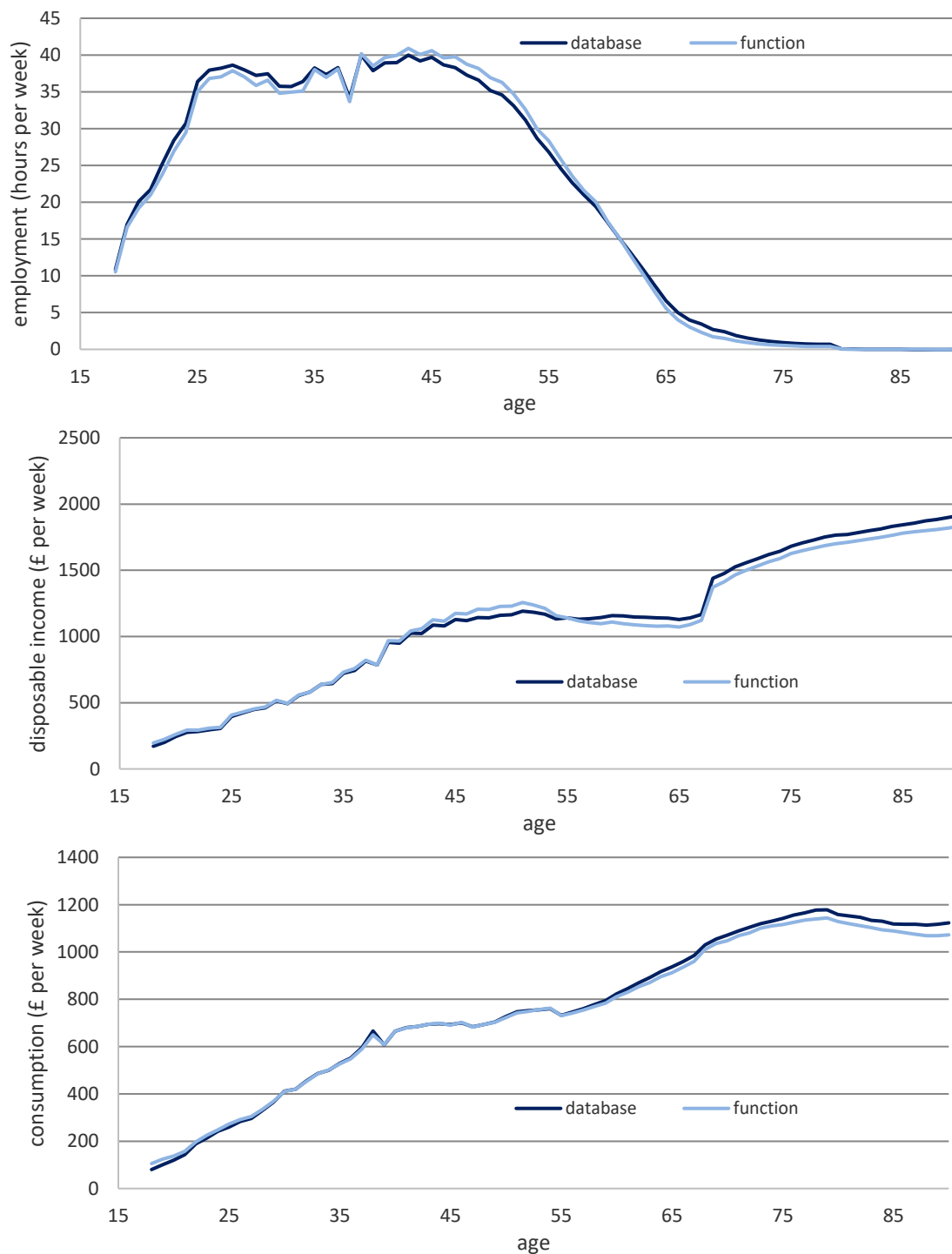
Both panels of Figure 4.1b indicate slightly higher wealth accrual to age 60 under the functional description for transfer payments, relative to the database imputations, followed by a faster decline in pension wealth at higher ages. These statistics help to complete an underlying narrative: distortions 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.

As previously noted, these distortions arise from the fact that some variables considered by the tax functions are not included or imperfectly approximated by the matching procedure. Greater accuracy can be achieved by increasing the individual-specific detail considered for the matching procedure. In the limit, when all variables considered by the tax functions are perfectly reflected by the matching procedure used to identify donors in the tax database, no distortions will arise. Note, however, that this last conclusion depends upon a sufficiently large and diverse tax database, so that associated small sample issues are avoided.

It is also noteworthy that the tax-benefit calculators and survey datasets that we envisage as sources for compiling a tax database typically describe at least as many individual-specific characteristics as a functional representation of taxes and benefits programmed within a dynamic heterogenous household model. This is important because it permits flexibility in specification of the matching methods used to impute taxes and benefits from the reference

database. Given sufficient population diversity in the tax database, the implication is that the imputation method outlined in this study should not suffer from a loss of policy detail, relative to functional methods for representing taxes and benefits.

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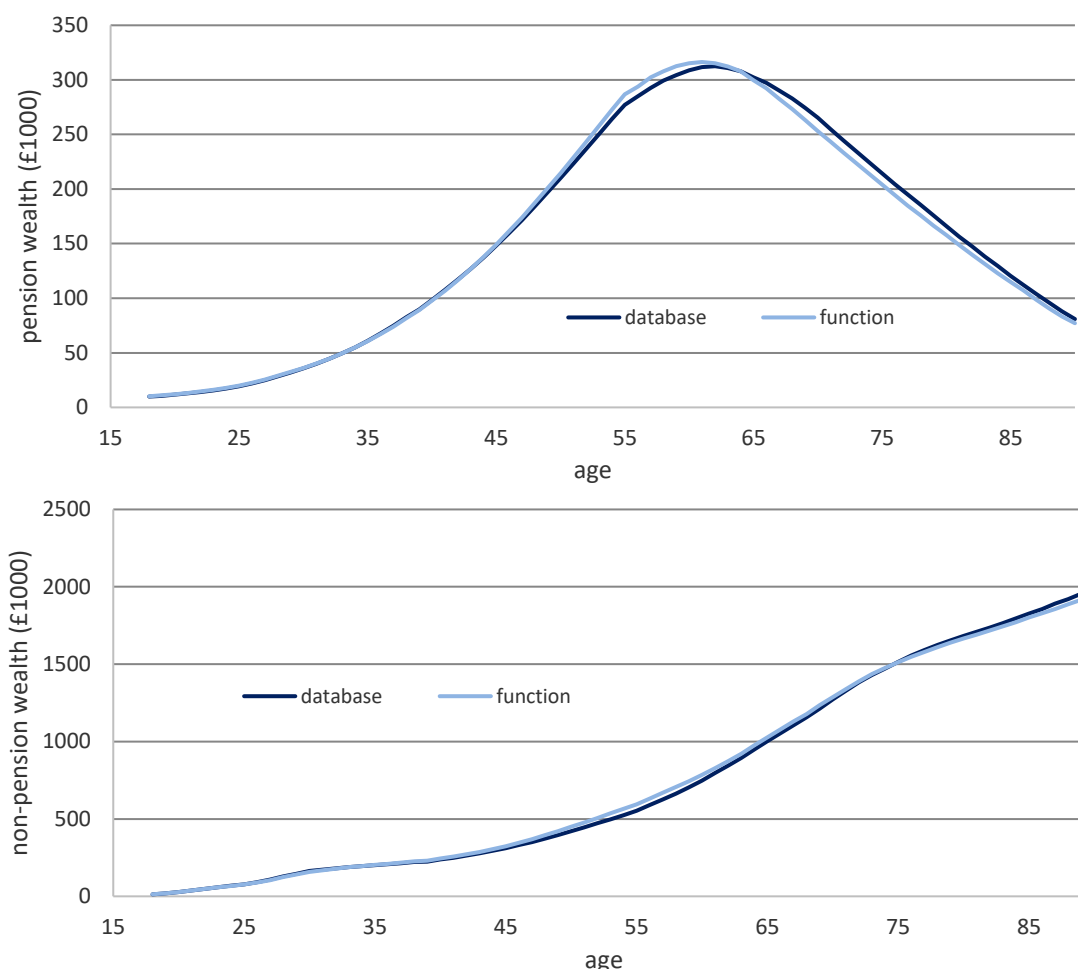

Source: Authors' calculations using simulated data

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 simulated population averages 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: See notes to Figure 4.1a.

4.4   Projected effects of a policy reform

This section reports the effects of a simulated policy reform, in which all rates of income tax are increased by 10 percentage points from 2017. This policy experiment is designed to clarify two related points. First, that the projected profiles reported in Section 4.4 for the "base policy" environment are sensitive to assumed tax and benefit policy parameters; and second that the extent of this sensitivity is similar when tax and benefit payments are simulated using function and database methods. These related points carry over to more complex reforms implemented over alternative time horizons. [25] The current discussion focusses on comparisons between the respective methods for projecting tax and benefit payments, rather than the effects of the simulated reform scenario *per sè*.

---

[25] More complex reforms associated with additional intermediate policy configurations have been considered using the SimPaths model; see Bronka *et al*. (2023).

Comparisons of statistics for the reform 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 reform, as described by the difference between the reform 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 reform 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 reform 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 reform, 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.
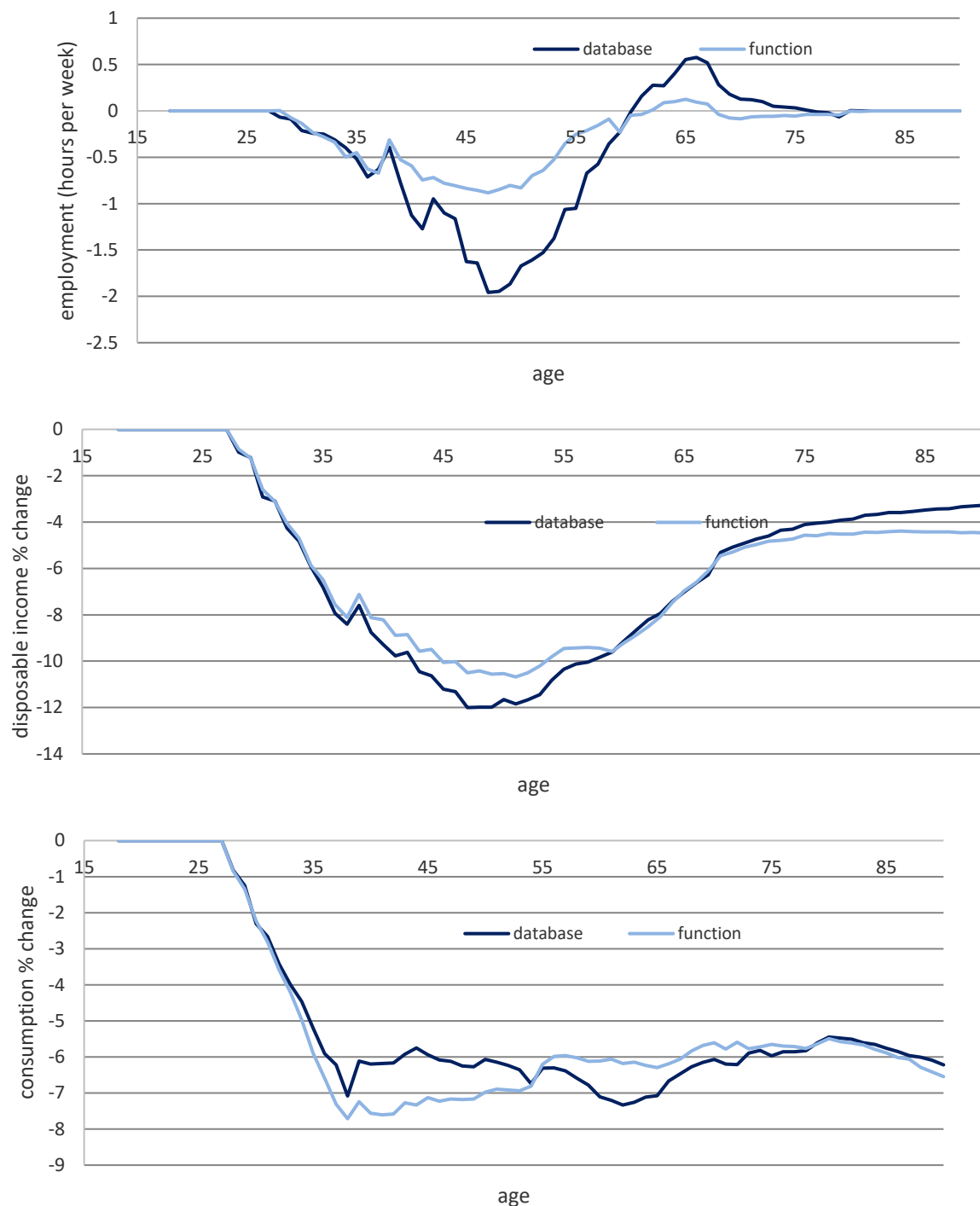
The middle panel of Figure 4.2a reveals very similar age profiles for the effects of the policy reform on disposable income generated using the functional and database descriptions for policy. Disposable income is projected to fall appreciably from age 25 to 50, reflecting the increase in taxes implied by the rise in income tax rates, the projected declines in employment, and – later in life – falls in savings (discussed below). Although maintaining very similar profiles, disposable income is seen to fall slightly more under when simulated using database descriptions for tax and benefit payments between ages 40 and 60, reflecting the more pronounced falls for employment described above. From state pension age, both methods for projecting fiscal transfers suggest similar declines in disposable income, equivalent to approximately 4% relative to the simulation base.

The bottom panel of Figure 4.2a also indicates very similar age profiles for the effects of the policy reform projected using the two methods for reflecting transfer policy; this time for consumption. It is striking that both methods for projecting transfer payments suggest cohort-average declines of consumption between 5 and 8 percent from age 40. This is slightly lower than the 10% increase in all income tax rates considered for analysis, despite the coincident fall in employment, as not all income is subject to income tax.

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. Both figures indicate more prominent declines of wealth to age 55 when transfer payments are projected using database methods, with the differences between the simulated profiles closing later in life. These differences reflect the more pronounced falls in employment projected to age 55 followed by delayed retirement when transfer payments are projected using database methods. Nevertheless, it is important to recognised that in all cases the signs of projected behavioural responses are independent of the method used to project transfer payments, and the scale of differences between the two methods is limited to a few percentage points.

The overall impression made by Figure 4.2 is consequently that projected responses to the policy reform 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 simulating transfer payments, where the latter are subject to some discernible differences in precise scale and timing.
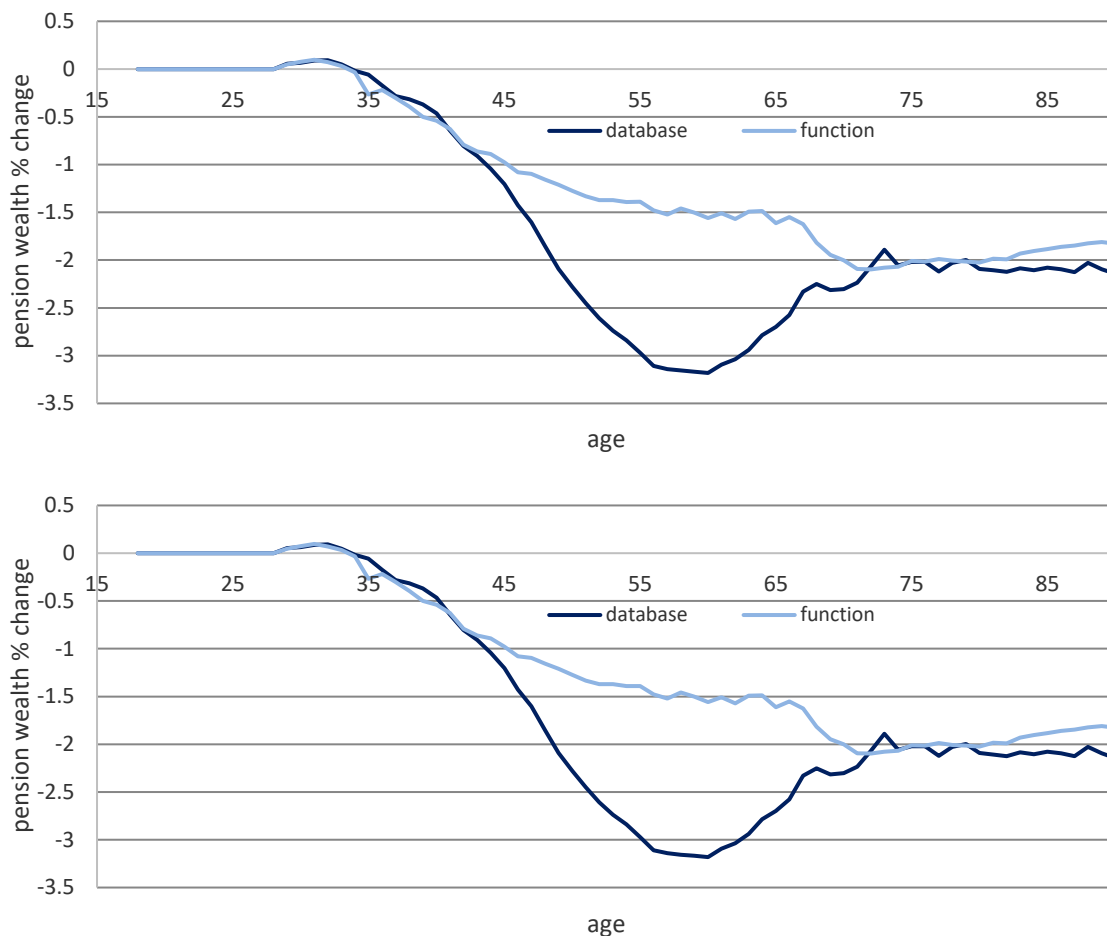
Figure 4.2a: Selected age-specific population average effects of a 10-percentage point increase in income tax rates on cohorts born between 1980 and 1989, by method of transfer payment imputation

Source: Authors' calculations using simulated data

Notes: Age specific population averages for population cohorts born between 1980 and 1989 projected under the policy reform, less the same statistics projected under the base simulation scenario. All financial figures reported as percentage changes relative to the base simulation scenario. 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 population average 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: See notes to Figure 4.2a.

5   Conclusions

This paper describes a new method for imputing taxes and benefits in dynamic heterogenous agent 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 can generate 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.

Notably, a comparison against a functional representation of the tax and benefit system can be a more demanding test than applying the method to real-world data. The functional form reflects an "ideal" version of the tax and benefit system, which excludes the noise and errors stemming from issues such as tax-evasion or incomplete benefit take-up potentially found in real-world data. Even small matching discrepancies become apparent when

compared to the precise functional-form outputs, but might be otherwise indistinguishable from natural variation present in real-world data. Validating against the functional form thus requires the proposed method to precisely replicate an idealised system, making it a more demanding test than one based on real-world data.

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 micro-data, or specialists in the construction of tax-benefit calculators. It is hoped that this will facilitate the development of more realistic dynamic heterogenous agent models (including 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 are 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 (conditionally) 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.

Andridge, R.R. and Little, R.J.A. (2010), "A review of hot deck imputation for survey non-response", *International Statistical Review*, 78, pp. 40-64.

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), "Aggregating 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.

Bargain, O., Dolls, M., Fuest, C., Neumann, D., Peichl, A., Pestel, N., and Siegloch, S. (2013), "Fiscal union in Europe? Redistributive and stabilizing effects of a European tax-benefit system and fiscal equalization mechanism", *Economic Policy*, 28, pp. 375-422.

Biewen, M and Juhasz, A. (2012), "Understanding rising income inequality in Germany, 1999/2000-2005/2006", *Review of Income and Wealth*, 58, pp. 622-647.

Blundell, R., Costa Dias, M., Meghir, C., and Shaw, J. (2016): "Female Labor Supply, Human Capital, and Welfare Reform," *Econometrica*, 84, pp. 1705–1753.

Blundell, R., Costa Dias, M., Goll, D., Meghir, C. (2021), "Wages, Experience, and Training of Women over the Life Cycle", *Journal of Labor Economics* 39(S1): S275-S315.

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.

Bronka, P., van de Ven, J., Kopasker, D., Katikireddi, S.V., and Richiardi, M. (2023), "SimPaths: an open-source microsimulation model for life course analysis", *CeMPA Working Paper Series*, CEMPA6/23.

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 Menten, G., Dekkers, G., Bryon, G., Liégeois, Ph. and O'Donoghue, C. (2014), LIAM2: a New Open Source Development Tool for Discrete-Time Dynamic Microsimulation Models, *Journal of Artificial Societies and Social Simulation*, 17 (3) 9

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

Axtell R.L. and Farmer J.D. (2022), "Agent-Based Modeling in Economics and Finance: Past, Present, and Future", *Journal of Economic Literature*, forthcoming.

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.

Frenette, M., Green, D. and Milligan, K. (2007), "The tail of the tails: Canadian income inequality in the 1980s and 1990s", *Canadian Journal of Economics* 40, pp. 734-764.

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.

Hufkens, T., Goedemé, T., Gasior, K., Leventi, C., Manios, K., Rastrigina, O., Recchia, P., Sutherland, H., Van Mechelen, N. and Verbist, G. (2019), "The hypothetical household tool

(HHoT) in EUROMOD: a new instrument for comparative research on tax-benefit policies in Europe", *JRC Working Papers on Taxation and Structural Reforms* No 05/2019.

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.

Keane, M.P. and Wolpin, K.I. (2009). "Empirical applications of discrete choice dynamic programming models", *Review of Economic Dynamics*, 12 (1), pp. 1–22.

Keys, R.G. (1981), "Cubic convolution interpolation for digital image processing", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29, pp. 1153-1160.

Liégeois, P. and Dekkers, G. (2014), "Combining EUROMOD and LIAM Tools for the Development of Dynamic Cross-sectional Microsimulation Models : a Sneak Preview", in G. Dekkers, M. Keegan and C. O'Donoghue (eds), *New Pathway in Microsimulation*, Burlington : Ashgate.

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.

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. (2001), "Redistribution over the lifetime in the Irish tax-benefit system: an application of a prototype dynamic microsimulation model for Ireland", *Economic and Social Studies*, 32, pp. 191-216.

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

Richiardi, M. (2017), "The future of agent-based modelling", *Eastern Economic Journal*, 43, pp. 271–287.

Richiardi M, Collado D, Popova D (2021). UKMOD – A new tax-benefit model for the four nations of the UK. *International Journal of Microsimulation*, 14(1): 92-101.

Rubin, D. B. (1980). Bias reduction using Mahalanobis-metric matching. *Biometrics*, 293-298.

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

Schofield, D., Lim, K., Tan, O., Shrestha, R., Haque, S., Crawley, K., West, S., Percival, A., Paramar, J., Kraindler, J., Li, J., Tanton, R., and Sue, C. (2022), "The development of a microsimulation model (MitoMOD) to estimate the economic impact of mitochondrial disease in adults", *International Journal of Microsimulation*, 15(2), pp. 102-111.

Shaw, J. (2011), "Fortax: UK tax and benefit system documentation", *IFS Working Paper* W11/08.

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., Bronka, P. and Richiardi, M. (2024), "The life course effects of care", *CeMPA Working Paper* CEMPA7/24.

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(1), pp. 135-166.

van de Ven, J. (2017c), "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. 134-164.", *International Journal of Microsimulation*, 10(3), 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) of the quick-start guide.

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 cohort statistics reported in Section 4.4)

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 – the Stata file copies slightly more data than described below.

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 ninvinc.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. Extend the "reference database" worksheet to reference all data in the "analysis" worksheet

    i. Copy cells A6 to AA6 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

j. Save tax_test3.xls

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

l. Open new workbook

m. Paste values to Cell A1

n. Create subdirectory: MODEL\TAX_DATABASE\base_fn

o. Save the new workbook 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

   p. Close all files

   q. Copy and paste the file tax_test3.xls from MODEL\SIMULATIONS\temp\ to MODEL\ANALYSIS_FILES\

   r. Open MODEL\ANALYSIS_FILES\tax_test3.xls

   s. Worksheet "analysis"

      i. Set cell C2 to 2057

   t. Save file and close

   u. Run SIDD.exe

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

   w. 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

   x. Open new workbook

   y. Paste values to Cell A1

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

      i. CSV format (not UTF8)

   aa. Close all files

6. Set simulation from (4) as base

   a. Open MODEL\Job File.xls

   b. Alt+F8

   c. Run SIDD macro

   d. Form 0 – specify new base using simulation "base2017_fn"

      i. Enter "base2017_fn" in the two empty text boxes at the bottom of the form

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

7. Run reform of 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 (to generate statistics reported in Section 4.5)

  c. Save

  d. Run SIDD.exe

8. Create database for reform of 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

    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 appendix

    b.   Run SIDD.exe

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

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 reform 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 E38 to "10pp_2017.csv" (without quotation marks)

     iii.   Set cell E48 to "10pp_2057.csv" (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

Appendix C.1: 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 reform 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 reform scenarios.[26]

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
Reform	55.7	52.0	107.7
net transfer payments calculated using function			
Base	54.3	87.0	141.3
Reform	54.4	53.3	107.7

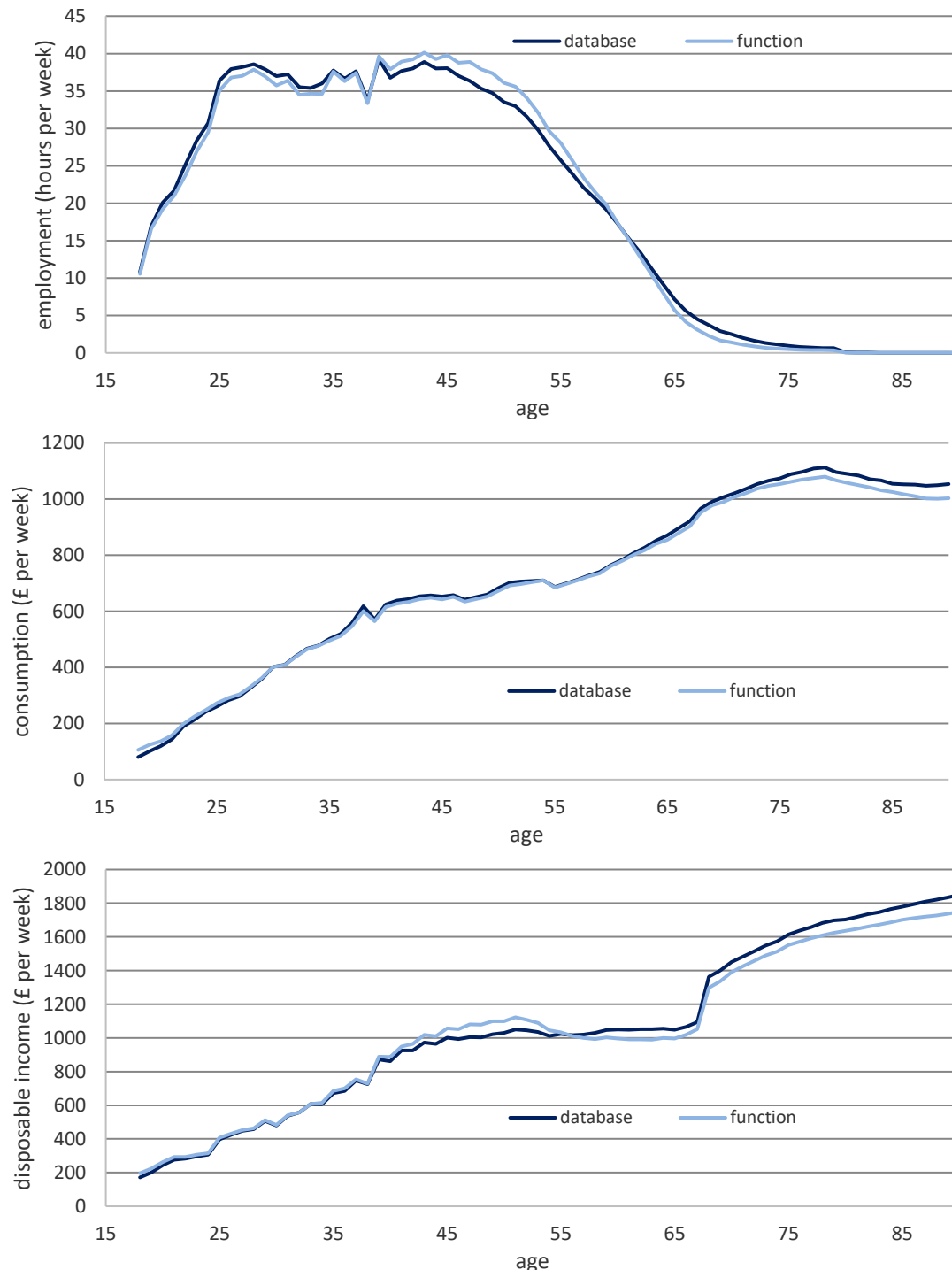Source: Authors' calculations using simulated data
Notes: "base" defines reference policy context; "reform" defines policy context that is identical to "base" in all respects, expect that all income tax rates are increased by 10 percentage points. All times reported in minutes. Simulations run on workstation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

---

[26] The base simulation involves projecting data 65 years forward and 65 years backward through time – 130 years in total – starting from the reference cross-section reported for the UK in 2017. In contrast, the reform simulation starts from the data generated under the base simulation and re-evaluates data only for the 65 years forward from 2017. This explains the longer computation time for population projections under the "base" policy scenario than for the "reform" scenario.

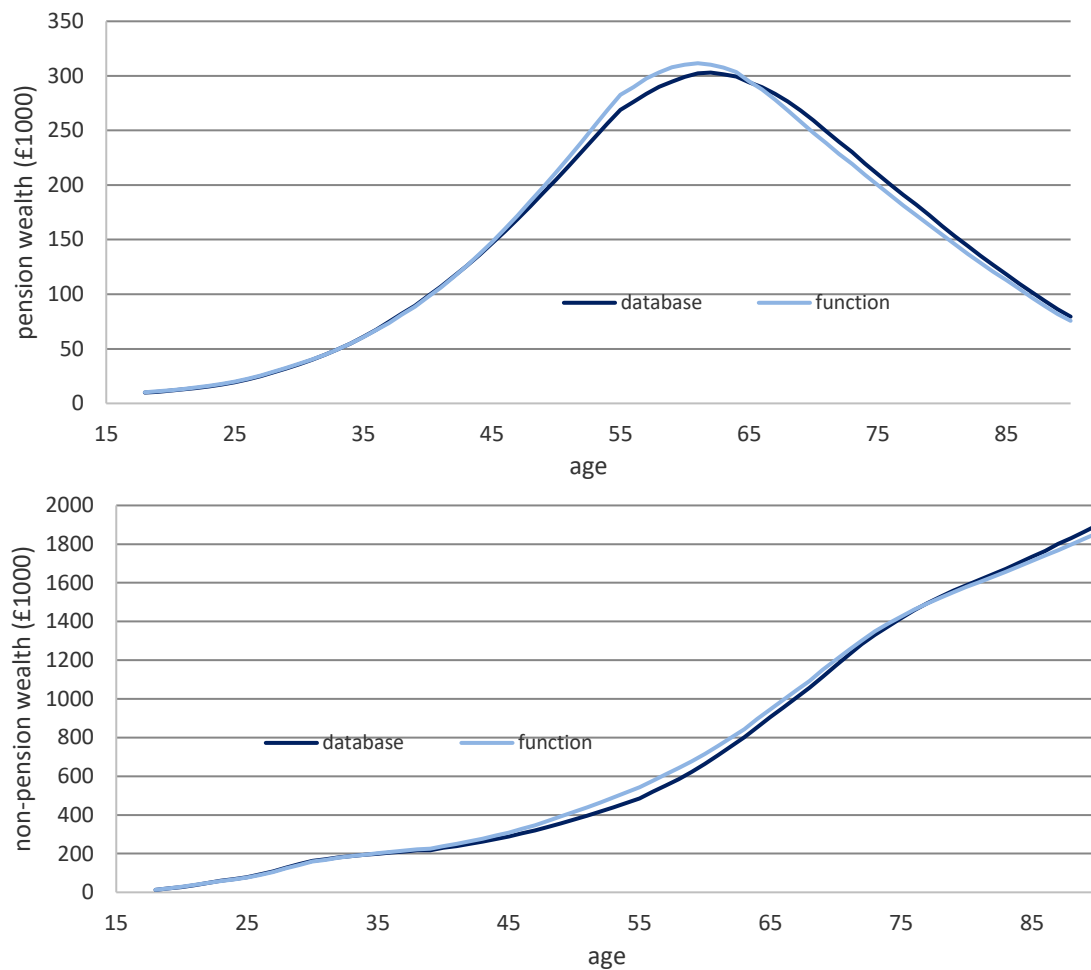# Appendix C.2: Supplementary figures for policy counterfactual

Figure C.1a: Selected age-specific population averages projected under the reform policy scenario 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 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.

Figure C.1b: Selected age-specific population averages projected under the reform policy scenario 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 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.