



Final Report of the Task Team on Microsimulation

March 2022

I. Introduction

The mandate for the Task Team of Microsimulation included a call to consider simulation techniques for aligning HFCS microdata and/or key estimates to a common reference period of a given wave or to another period beyond the survey reference period, potentially under alternative policy or macro scenarios.¹ The mandate also emphasized the importance of assessing the quality of such estimates and for exploring the possibilities for including administrative data in the process. This paper provides a summary of the progress of the task team and a proposal to be considered for additional work.

For purposes of exposition here, we consider the practical approach to microsimulation in terms of what are taken to be polar cases. First, microsimulation could be aimed at projecting all or part of the HFCS micro database (here, “Type 1”) to a period different from the time of data collection, to enable a wide set of possibilities for modelling or estimation that need not be determined a priori. In this approach, attention must be paid to sufficiently unbiased and efficient estimation, but also to the statistical joint error structure of the data. In its purest form, this approach can be seen as creating a type of synthetic panel microdata which, in turn, could be used for estimation.

Alternatively, the microsimulation process may entail making only specific estimates (here, “Type 2”), where the estimates might be either internal or external to the given set of HFCS data. In this case, the objective is focused more narrowly on achieving specific unbiased and efficient estimates, though the process may also involve aspects of Type 1 microsimulation. But even in such cases, there is generally less of a need to consider broad covariance and error structures in the data.

1. *Type 1 microsimulation*

For Type 1 microsimulation, there is an explicit need to characterize and represent the relevant dynamics of the variables to be generated as outputs, given their baseline values. For example, household composition, employment status, and the values of income, assets and debts may alter from the time of the baseline data in inter-related ways, in response to macro-level and micro-level events. To the extent the changes attributed in microsimulation align with the actual changes for a unit, the resulting correlations and the latent error structures remain appropriate for unbiased multivariate modelling or other analysis as if the data were actually panel data.

But because it is unlikely to be the case that all such variables can be constructed exactly from external data or exact formulas (else there would be no need for microsimulation), generally some approximation is needed. Assigning the value of the output variables of microsimulation in this context may be thought of as a type of imputation where the important aspects to be determined are the variables available to condition the imputation and the identification

¹ The Task Team members included Miguel Ampudia (observer, ECB), Marina Asensio (ES), Ludmila Fadejeva (LV), Romina Gambacorta (IT), Sandra García Uribe (ES), Arthur Kennickell (US), Davide Di Laurea (ECB), Peter Lindner (AT), Malgorzata Osiewicz (ECB), Luís Martins (PT), Eleonora Porreca (IT), Veli-Matti Törmälehto (ECB), Matas Vitonis (ECB), Ernesto Villanueva (ES) and Junyi Zhu (DE).

of the imputation mechanism or model.² Intuitively, the more data specific to the individual households in the period simulated that are available, the more reasonable it would be to expect the simulated data to have potential to preserve relevant properties idiosyncratic to both the household and the period of the microsimulation. The mechanism or model should incorporate all such relevant data and other data already simulated, and its estimated parameters should be either specific to the simulated period or reasonably time invariant.

This level of approximation, however, may well be out of reach in many cases. A variety of alternative techniques are available to generate synthetic data that may still be fit for purpose. For example, projection models may be estimated with prior panel data (perhaps together with external data, potentially including administrative data), where available. Models of relationships across variables within the period projected could be estimated using the baseline period or another source if time effects and other potential distortions can be controlled, minimised or ignored.

In any case, it is reasonable to expect that some types of dynamics may go beyond what can reasonably be hoped to be modelled relatively formally. For example, there might be too little information to serve as a basis for simulating the dynamics of household composition; in that case, the relevant composition data could be frozen, and if sufficient external data are available, a calibration routine could be used to adjust the survey weighting to align the distribution of observed households with the distribution of the period simulated.³ A calibration approach might be applicable to other classifications as well, though careful consideration should be given to the possibility the composition of groups may have changed in terms of broader characteristics. Where there are economically important transitions that need to be reflected in the simulation and no other modelling approach is available, transitions could be randomly assigned using whatever more aggregated information is available for conditioning. For example, tables of unemployment rates by occupation class and region could be used to assign unemployment that status on a conditionally random basis.

An important caution in using statistics external to the HFCS data in the microsimulation process is that the two sources may represent populations that are different. In general, the HFCS is taken to be representative of the target household population. In some dimensions, such as age, the survey and external data may align closely, if only in this case because age is typically a factor in weighting adjustments for the HFCS. But especially for variables that are highly skewed, such as income and wealth variables, there is a greater possibility for significant deviations. For example, in the survey the implied holdings of many assets typically fall substantially short of aggregate estimates. Although there might be value in incorporating a means of aligning the two totals as a part of the microsimulation process, it appears that such an effort would add substantially to the complexity of that effort. Moreover, the microsimulation and alignment operations generally appear sufficiently independent that they could be accomplished in sequence as in the DWA (distributional wealth accounts) work, rather than together. Despite the lack of overall alignment of some survey and administrative distributions however, it may still be reasonable to use measures of proportional change in those distributions to adjust survey variables to a different period, when better alternatives are not available. In such cases, it might be better to implement such adjustments by groups, where possible; for example, because evidence indicates that proportional gains in particular assets are not uniformly distributed, it might be better, where possible, to condition such adjustments on wealth group.

² Where the microsimulation aims to model the required range of variables directly, the method is typically referred to as “dynamic”.

³ This approach to microsimulation (adjustments by reweighting and/or inflating/deflating monetary values) is typically referred to as “static”.

2. *Type 2 microsimulation*

For Type 2 microsimulation, the focus is directly on the best way of achieving a specific estimate. As noted above, sometimes achieving such estimates would involve application of Type 1 microsimulation. For example, in one stage of the ECB expert group's construction of the estimates for the DWA, HFCS wealth variables are grossed up for each observation to a period beyond the survey period, according to the percent change in the aggregate value of the variable. In general, Type 2 microsimulation faces less stringent requirements than Type 1 microsimulation, as a result of the focus on a narrower set of objectives, as well as the lesser need in many cases to worry about preserving error structures and other relationships in the record-level data. As a very simple example, if there were sufficient waves of microdata, one might estimate the mean home mortgage payment as of a time after collection of the microdata, using a model of the time series of mean mortgage payments, conditioned on the level of a house price index and the typical mortgage interest rate. For other many estimates, such as a payment-to-income ratio or other indicator of financial fragility, a broader range of actions and assumptions would likely be required, thus moving more in the direction of Type 1 microsimulation.

It should be noted that this approach to microsimulation has already been widely and routinely applied in many policy settings. Policymakers almost always desire to bring the most current information available to bear on decisions. Ad hoc assumptions are often applied to adjust survey data and sometime combine that information with data from other sources to generate a specific set of estimates. The goal of the work proposed by the task team is to develop a more systematic and replicable approach for the HFCS data.

3. *Validation*

To be useful, microsimulation must yield results that reduce bias and variability to a level consistent with the uses to which the data will be put. Sophisticated microsimulation modelling approaches often come at the expense of a large number of assumptions. Thus, some form of validation of the underlying simulation approaches is essential for establishing the credibility of the results. There are various approaches one can take in this regard, including sensitivity analysis, external and internal validation, confidence intervals, and other methods.

Depending on the information available, the structures/models and parameters of a microsimulation model may be determined via a range of approaches and possible parameter values. In some instances, the parameters can be derived from additional information. However, this is often not the case. Exploring a broad range of potential parameters and analysing the impact on results is one of the most widely taken approaches in validation efforts. For example, gauging the variability in the modelling of unemployment via a regression approach could be addressed by altering the type of model and/or explanatory variables.

External as well as internal validation exercises may be used to assess the quality of a microsimulation approach. External (to the data used for the microsimulation model) information might be available to assess the validity of a simulation approach. This could be either for the final results desired by the simulation or an intermediate one, which is derived from a part of the approach. A new or past wave of a survey could also be regarded as an external information simulating the desired information forward or backward and checking the result. If such data are not available, one could also approach the validation by splitting the data used for the simulation, leaving part of the data for validation. This avenue allows for assessing the results of the microsimulation regarding real world data, however, comes at the expense of less data for the simulation.

Investigating standard errors and confidence intervals as a type of validation analysis can be done in a variety of ways. For example, a Monte Carlo type simulation approach can be used to simulate a distribution of a result, instead using only a single estimate. One such approach would be to examine the effects of an overall decline in income or employment through a regression approach and repeat this prediction many times, using either a random draw of a disturbance term or random subsamples for estimation. This type of repetition would allow the derivation

of a distribution of the parameter of interest. The variation of this distribution could provide an indication of how precise a microsimulation might be.

This sketch of possible approaches is intended to provide intuition. More concrete methods should be developed in parallel with the construction of the relevant microsimulation models. Much remains to be learned about how best to quantify the validation of simulated data in general and in the context of the specific analytical objectives.

4. Remainder of this paper

To illustrate all these aspects with a practical example the following section of this paper provides a sketch of an approach to microsimulation of data necessary to assess the financial fragility of households beyond the point at which the most recent HFCS data are available. Although this example involves a type II microsimulation approach, by considering the main variables collected in the survey it may be considered not too far from the type I microsimulation. The final section offers some thoughts on the possibilities for moving forward with microsimulation more broadly for the HFCS.

II. Simulating household financial resilience: an empirical example

Households' financial resilience is the capacity to recover after an unexpected fall in income or increase in undelayable expenditure. It represents the capacity to access financial resources that are sufficient to prevent a substantial worsening in living standards after an adverse economic shock. An adverse economic shock can be represented here by job loss or unwanted reduction in working hours, illness or death of family members, relationship breakdowns, and damages to a household's estates or durable goods.

Households can have access to financial resources through existing or potential resources at the time of the shock. Existing resources refer to all endowments that a household already holds at the time of the shock. Potential resources refer to all possible resources that a household might have access to in case of a shock: public welfare benefits (such as unemployment insurance), insurance pay-outs, loans from the credit market or from family and friends. The financial resources just mentioned can be accessed in the short-term. However, households could rely also on other coping mechanisms that require longer time to be implemented, such as: decreasing consumption, especially of committed expenditure as rent or mortgage payments, increasing home production of goods, sale of properties/goods, increasing labour supply both at the extensive and intensive margins. Therefore, considering short or long persistence of the shock and distinguishing between existing or potential resources leads to different coping strategies and adjustment behaviours.

The relevance of households' financial resilience emerged with the 2008 financial crisis and became even more important with the outbreak of the Covid-19. It affects, not only many dimensions of individual well-being (emotional and material well-being, family stability), but also financial and social stability at the macro-level. In fact, financial resilience influences households' consumption choices, therefore economic growth, and the financial system through the ability to repay debt or to access to a new loan. It potentially affects all types of households, not just those indebted or those already in a status of poverty. Characterising financial resilience entails addressing all potential resources and all possible coping strategies preventing the risk of falling into hardship in case of an adverse income shock. In addition, the set of households' potential tactics and resources may vary according to the duration of the shock. Therefore, the relevant set of resources included should vary according to the short or long run coping capacity. For the sake of this discussion, we assume that aspects of financial resilience can be measured by the three indicators below and we use HFCS wave 3 data to compute those indicators in order to analyse the situation in 2016.

Baseline indicator: a household is resilient for at least n months if its income and financial assets are sufficient to keep the household above the national at-risk-of-poverty line for that period:

$$Y_t \gamma + FA_t \geq \gamma Z_t$$

Where $\gamma = \frac{n}{12}$ is the period of time resilience refers to, Z_t is the at-risk-of-income poverty line (the 60 per cent of the national median of the equivalised total households gross income), Y_t is the equivalent household income (work income from employees and self-employment, rental income from estate properties, income from private and public pensions, social and private transfers or other sources of income) and FA_t is the equivalent household value of liquid financial asset (amounts held in deposits, mutual funds, bonds, and shares).

Short-term indicator: a household is resilient for at least n months (here, $n=3$) if:

$$Y_t \gamma + FA_t - HC_t + NLF_t + NLC_t \geq \gamma Z_t$$

Where HC_t is the equivalised housing cost calculated as the sum of rent and HMR mortgage the household has to pay in three months' time, NLF_t is the equivalised new loan from family and friends (*In an emergency, could (you/your household) get financial assistance of say EUR 5,000 from friends or relatives who do not live with you?*) and NLC_t is the estimated equivalised new non-collateralised loan from the credit market the household can access in the next three months

Long-term indicator: a household is resilient for at least n months (here, $n=12$) if

$$Y_t \gamma + FA_t + RA_t - HC_t^* + NLF_t + NLC_t^* \geq \gamma Z_t$$

Where RA_t is equivalised value of real assets (not used as collateral and not located) other than the main residence, HC_t^* is equivalised housing cost calculated as the sum of rent and HMR mortgage interest payments the household has to pay in the year (estimated assuming a French amortization scheme) and NLC_t^* is the estimated equivalised new non-collateralised loan from the credit market the household can access in the next year.

There are several approaches to extrapolating household resilience in a period of changing incomes. Perhaps the simplest possibility is the static approach: what would be the fraction of households with a given ratio of disposable wealth to income in a situation when the distribution of income differs from that in the benchmark survey. This approach could be achieved by reweighting a prior wave of the survey to impose the distribution of employment (say) had been that observed in another period. By definition, this approach abstracts from other changes (say, asset values, interest rates, debt dynamics). The critical implicit assumption is that if the currently employed lose their job they will behave like the currently unemployed, which may well not hold, depending on the origin and transmission of the shock. On the other hand, that strategy preserves the joint distribution of wealth and income *within* particular groups in the benchmark period—it just changes the population weight of particular income groups.

A dynamic approach would explicitly model changes in income and the other variables relevant for the measures above. The challenge is how to simulate changes that would be sufficiently coherent. To keep the discussion brief, we consider here only the simulation of income and wealth components.

Ideally, external data at the micro level could provide some elements directly, or factors highly correlated with some of the desired elements. Putting this possibility aside for now, we consider a few possibilities.

In the specific case here, the goal is to examine households' situations in light of a shock to income, taken here to be a result of changes in employment. Thus, specifying the incidence of the shock is of primary importance. Where external information is available on patterns of change in employment within sectors is available, that information could be used as a target for attributing changes for survey households within those groups. For example, historical

data might be used to estimate the propensity for unemployment for households within a given sector, and then attributing a change to survey households by taking progressively the households estimated to be most likely to make this transition, until the externally estimated level is reached. A variety of other approaches are possible, depending on the data available.

Addressing changes in wealth variables may require somewhat more bold assumptions or sophisticated modelling. For example, it would be straightforward to allocate the change observed in aggregate wealth measures in a uniform way to the household-level baseline wealth values. However, research indicates that rates of return vary across the population in systematic ways. In addition, households may be on different paths of accumulation or decumulation, which would not be captured at all such an approach. Breaking out the components of change into units that could be modelled would improve the coherence of the results, but the specific approach would depend on what data are available and the degree of precision needed.

III: A possible way forward

As noted in the introduction, microsimulation could focus on simulating microdata for different time periods for broad purposes, on creating specific estimates, or on a combination of the two. Microsimulation for the HFCS offers the possibility of extending the usefulness of the survey in important ways to support research and policy. At the most basic level, projecting all of the country-level surveys or key estimates to a common time period for each wave would eliminate a perennial problem of mismatched reference periods of the surveys. Extending such a model to allow projection to an arbitrary time different from the internal periods of the surveys would obviously be even more useful. Creating a formal structure for microsimulation would also contribute to policy work, which often must apply ad hoc methods to address questions that point to periods outside the survey reference period.

Often the HFCS is thought of only as a basis for computing estimates of distributions or relationships among variables in the survey. But the survey structure also has the capacity to serve as a sort of reaction function to events or rules outside the survey—for example, to compute taxes, or to evaluate stresses in homeownership in light of changes in interest rate policies. Microsimulation, by relying on external data to drive the projections, heightens both the importance and the degree of integration of such data. Developing this angle more fully would align with some of the most advanced working going on in economics to make deeper use of high-frequency data from administrative data and non-traditional sources.

A key fact that has emerged in the discussions of the task team is that the types of data available to be integrated with the HFCS for purposes of simulating survey values for different time periods vary greatly across countries. Thus, at least for the present, it is not feasible to develop a single computer package that could be used for microsimulation in all cases except perhaps the simplest. It is, however, more likely that a variety of formal approaches can be developed to be applied broadly, with data and applications altered locally as needed.

The most important constraint at this point in achieving this goal is the availability of resources. While not every country needs to be actively engaged initially in developing approaches, at least a small number of intensive efforts are needed if progress is to be made without extraordinary delay. The work needed appears to call for at least as much effort as in the case of the expert group on distributional financial accounts or the initial development of multiple imputation for the HFCS. The work of the task team represents a meaningful start on this work. We recommend that the HFCN considers the costs and benefits of proceeding further. Considering an approach focused initially on a core cluster of estimates might be the clearest way to achieve concrete benefits, while proceeding incrementally to develop an infrastructure to support the more advanced microsimulation specified in the mandate to the Task Team.