

Nowcasting Household Median Income: A Comparison of Microsimulation and Time Series Approaches

Lee Mallett (Office for National Statistics)

Martin Weale (King's College London)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 2B-1: Timely Indicators of the Distribution of Income and Wealth

Time: Tuesday, August 21, 2018 [14:00-17:30]

Nowcasting Household Median Income: A Comparison of Microsimulation and Time Series Approaches

July 2018

Lee Mallett Office for National Statistics, Newport

Martin Weale

King's College, London and ESCOE

This is an initial contribution to a paper being prepared by Lee Mallett and myself. Discussion with Lee has revealed that some of the data need re-calculating.

This work contains statistical data from ONS which are Crown Copyright.

The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data.

This work uses research datasets which may not exactly reproduce National Statistics aggregates.

1. Introduction

Historically, estimates of household median income have been released just under twelve months later than the end of the financial year in question. The publication lag contrasts with that of aggregate data such as GDP which are made available within a few weeks of the end of the quarter to which they relate. That inevitably means that data on median income, while attracting significant coverage, are less useful to decision makers than they would be if they were more timely.

To address the need for timelier estimates of household median income, the ONS has started to publish early estimates based on information already known about changes to taxes, earnings and employment (ONS, 2016). The early estimates for financial year 2017/2018 will become available at the end of July 2018 - less than four months after the end of the income reference period.

The ONS approach to producing nowcast estimates uses historical data, which are 'uprated' to the period of interest. HM Treasury's Intra-Governmental Tax and Benefit Microsimulation Model (IGOTM) is used to simulate the tax and benefit system, before the population weights are recalibrated to reflect changing demographics and labour market participation over time. This is the first attempt from the ONS to produce quarterly rather than annual nowcast estimates.

The general experience of producing nowcasts and short-term forecasts of aggregates such as GDP is, however, that nowcasting components and then aggregating these to produce the variable in question can deliver a performance worse than that obtained by aggregate time series methods. That makes it of considerable interest to compare with ONS approach to producing nowcasts of median household income with the nowcasts that could be generated using time series/regression methods.

In this paper we therefore present the results of the ONS approach to nowcasting it and compare its performance with two time-series methods. The first involves recursive regression while the second explores the performance of methods based on time-varying parameters. We begin by describing our methods, giving an account of the ONS approach to generating nowcasts of household median income and the two time-series methods with which we compare it. This is followed by presentation of the data we use and the results of the three approaches.

2. Methods

Microsimulation

The term 'microsimulation' is used to describe modelling techniques that apply rules to simulate changes in state or behaviour at individual unit level (Figari et al, 2014). Microsimulation models (MSM) themselves are typically classified as either static or dynamic. Static models are often arithmetic models which look at the direct distributional impacts of policy changes on individuals and households, while not accounting for demographic or behavioural responses (e.g. to changes in government policies). Dynamic models build on static models by providing individuals the ability to change their characteristics due to endogenous factors within the model (cited in Jinjing Li, 2011). Though the distinction between model types is made, Figari et al (2014) note that it is not necessarily useful as modern microsimulation can combine elements of each type.

MSM in the social sciences are often acknowledged as originating in the work of Guy Orcutt in the late 1950's. Orcutt (1957) stated that "existing models of our socio-economic system were of rather limited predictive usefulness". He noted that the models did not predict the long- term effects of alternative government policies; had a limited understanding of the size and location of micro units; and focussed on aggregates whilst failing to predict distributional information of individuals,

households or firms. Orcutt advocated micro-based modelling focussed on individual units, stating in 1957:

"This paper represents a first step in meeting the need for a new type of model of a socioeconomic system designed to capitalize on our growing knowledge about decision-making units."

The 1970s saw large scale microsimulation development, though earlier 'dynamic' models were criticised due to heavy programming, computing and data requirements. These shortcomings led to the development of less ambitious 'static' models in the 1980's (Baroni and Richiardi, 2007).

Although MSM had wide ranging applications in public policy - for instance in transport and healthcare policy - models simulating the effects of social and fiscal policies were not developed until the 1980's with the increased availability of essential inputs and computing power (Figari et al, 2014).

EUROMOD, which was developed in 1996, is a static tax-benefit microsimulation model for the European Union which simulates individual and household tax liabilities and benefit entitlement according to policy rules in each member state. It applies user-defined tax and benefit policy rules to individual and household microdata, calculates the effects of the rules on the household income, and outputs results at the micro level¹. Most of the input data are derived from the European Union Statistics on Income and Living Conditions (EU-SILC), though in some countries national versions of SILC are used directly or as a complement to the EU version (Sutherland and Figari, 2013). The use of updating factors (also referred to as 'uprating factors') for each income source brings the income values from the reference period to the level of the policy year. This process is common when users have "dated" microdata which needs to be adjusted to 'age' the microdata to reflect the period of interest.

Dekkers (2004) states that in assessing distributional effects of alternative scenarios, "one can use static ageing techniques, which age the population by reweighing and uprating, or dynamic ageing, which alter the relevant population by applying deterministic probabilities that a certain event may or may not occur." He notes that static ageing is simpler, and explains "in making abstraction from macroeconomic developments, models with static ageing mimic exogenous future demographic and labour market circumstances by reweighting the dataset".

Kump and Navicke (2014) define reweighting as "a calibration of the survey weights to match the micro-data with the external aggregates", and explain that the process changes the weights whilst leaving other variables unchanged. Immervol et al (2005) provide an example of using static ageing techniques to account for population changes in tax-benefit microsimulation models. In the paper, "static" ageing techniques are defined as "methods attempting to align the available micro-data with other known information (such as changes in population aggregates, age distributions or unemployment rates), without modelling the processes that drive these changes (e.g., migration, fertility, or economic downturn)."

Microsimulation models are found in almost every developed country, with some models - mostly static - also in emerging or developing countries (Baroni and Richiardi, 2007). Examples of tax and/or benefit simulation models in use in the United Kingdom are TAXBEN from the Institute of Fiscal Studies (IFS), HM Treasury's Intra Governmental Tax and Benefit Model (IGOTM), and the

¹ Definition taken from 'How EUROMOD works', <u>https://www.euromod.ac.uk/about/what-is-euromod/how-euromod-works</u>

Department for Work and Pension's (DWP's) policy simulation model (PSM). This paper makes use of HM Treasury's IGOTM, a static model which simulates UK tax and benefit policy changes.

Time Series

We use two time-series methods. Both follow the general principle of nowcasting, that of identifying one or more covariates, x_{it} which are available to earlier than the variable we wish to nowcast, y_{t} , and using these to project y_t . The covariates we use are log of average weekly earnings and log of household disposable income *per capita*, both deflated by the consumer price index. We use these to model the log of household median income, again deflated by the consumer price index. The first method we use is simple regression and the second is a regression model with time-varying coefficients.

A Simple Regression Model

A general regression specification is, with \mathbf{x}_t the covariates in vector form, with a constant term included in the vector

$$y_t = \sum_{k=1}^{n_1} \alpha_k y_{t-k} + \sum_{k=0}^{n_2} \beta_k x_{t-k} + \gamma + \varepsilon_t$$

Here lags are included in both the dependent variable and the explanatory variables. This framework of course includes a model in first differences. If we set $n_1=n_2=1$ we can rewrite the model as $\Delta y_t = \beta_0 \Delta x_t + (\alpha_1 - 1)y_{t-1} - (\beta_1 - \beta_0)x_{t-1} + \gamma + \varepsilon_t$. If we have only one co-variate and the restriction $\alpha_1 + \beta_0 + \beta_1 = 1$ holds then in the long run the covariate and the dependent variable move in step. In our particular context that means, for example, that if log of real disposable income per capita is the explanatory variable, a given proportionate change in it would be reflected one for one in real household median income.

In order to use this model, we estimate the parameters α_k , β_k , and γ using quarterly data. We assume that the covariate accrues on a quarterly basis and use the model to project the dependent variable. Except in the special case where $\alpha_k=0$, if we want to look more than one quarter ahead, we have to generate a projection for y_{t+1} and use this to generate a projection for y_{t+2} etc. In order to produce an annual nowcast once four quarters of the covariate are available, we have to generate four quarterly forecasts in this way. The model also needs to be estimated recursively; all four quarters of median income data become available at the same time, and when this happens the parameters of the model can be re-estimated for use in subsequent projection.

Time-varying Parameters

The approach set out so far assumes that the coefficients of the regression model are stable over time, notwithstanding the fact that different regression coefficients are used at each recursion. An alternative approach, based on the state space model of Harvey (Harvey, 1989, Durbin 2012) models the time variation in the regression coefficients explicitly.

The state space model has the following format. The vector β_t is comprised of unobserved states while y_t represents the variable of interest, log of median real household income. A generic state space model then has the form

$$\beta_{\tau+1} = A_t \beta_t + B_t u_t$$

 $y_t = X_t \beta_t + D_t v_t$

Here u_t is assumed to be a normally distributed random serially uncorrelated vector variable with unit variance of the same dimension as β_t while v_t is a similar vector of the same dimension as y_t . The coefficient matrices, A_t , B_t , X_t and D_t can in principle all be time-varying. In the model used here, however, A_t is assumed to be an identity matrix, so that β_t follows a random walk. B and D are treated as constant β_t contains the time-varying parameters to be estimated while X_t represents the data. Thus $X_t \beta_t$ represents the fitted value of the observed variable y_t and Dv_t is the forecast error.

The model can be estimated using the Kalman filter; MATLAB provides a group of routines to facilitate this. To fit the model it is necessary to have some initial value for β_1 , together with a value of its covariance matrix. However, using the routine *dssm* it is possible to set up a diffuse state space model, which assumes that the covariance matrix of β_1 is infinite. This ensures that the choice of initial value has no implications for the subsequent estimates.

The values of β_t are estimated recursively; nevertheless, the values of B and D are, in our specification, assumed to be invariant. Since their values depend on the whole of the sample, it follows that the estimates of these parameters, and thus of the β_t will depend on the sample over which the model is estimated. We explore models with up to two regression coefficients and a constant term in β_t with the elements of X_t being one or both of log Average Weekly Earnings deflated by the Consumer Price Index and log Household Gross Disposable Income per capita, again deflated by the Consumer Price Index. y_t is the log of median household income.

3. Data

Microsimulation

The microsimulation model is underpinned by data from the ONS' Living Cost and Food Survey (LCFS). To increase sample size, the input data combines three financial years (financial years ending 2013, 2014 and 2015) resulting in a sample of 15,500 households.

Observations are 'uprated' to the quarter of interest using published series produced by the ONS and others for periods where actual data are available. For example, earnings data are uprated to the financial year of interest, using the Annual Survey of Hours and Earnings (ASHE) data on earnings growth at different points across the distribution, as well as the latest average earnings estimates from National Accounts.

Other financial variables are uprated in the following way:

- income from self-employment, incomes from odd jobs and private sector rents are uprated in line with average earnings
- incomes from private pensions and annuities are uprated in line with the Retail Price Index (RPI), with a further adjustment made for growth in pension income at the individual level from ETB for periods where actual data are available
- incomes from the main government benefits are uprated in line with the Consumer Price Index (CPI), or other values as appropriate

Median household disposable income is equivalised to allow comparability between households of different sizes. The equivalisation scale used is a rescaled version of the OECD-modified scale, where two-adult households are given an equivalisation factor of 1. Estimates are deflated using the ONS' Consumer Price Index (CPI) (2015=100). While the CPI is not the ONS' headline measure of inflation –

that is the Consumer Price Index including owner-occupiers' housing costs (CPIH) - it benefits from having a longer time series which is not modelled (the CPIH back series is modelled pre-2005). This is helpful for deflating the time series methods we use. While the results presented here are in terms of 2015 prices, it is a straightforward matter to convert them to current prices.

Nowcast estimates of median equivalised disposable income are compared with outturn data from the ONS Effect of Taxes and Benefits (ETB) publications. Whilst the ETB outturn data are published for financial years only, quarterly estimates are produced by filtering the dataset by a sample quarter variable. This allows a comparison of quarterly nowcast estimates against quarterly outturn data (see table 5).

Given the time lag between the input dataset and the quarters of interest, the original survey weights are recalibrated using known information to improve representiveness of the estimates. Recalibration of the existing weights involves using updated control totals and an additional constraint – economic status, which is taken from the Annual Population Survey (APS). Further information is available in the technical annex.

Time series

The two covariates we use for the time series analysis are shown, together with the median household income itself, in figure 1. For the purposes of econometric analysis we work with logarithms of the variables, denoting that of household median income as LnHmed, the log of average real weekly earnings as LnRAWE and the log of real gross disposable income as LnRGDI. We do not find any evidence of seasonality in LlnHmed. In a regression equation with three seasonal dummies and a constant, we can easily accept the hypothesis that the coefficients on the seasonal terms are zero (F(3,56)=0.35, p=0.79). The other variables we use are seasonally adjusted.

A key question is whether these data are stationary. This is explored using the Dickey-Fuller test, with the BIC criterion employed to establish the order of test. For all three variables, the Dickey-Fuller tests, shown in Table 1 suggest that all the variables are I(1). This is a general feature of economic time series and suggests that regression modelling has to focus on models in first differences.

A separate issue arises from the fact that interest focuses on the median. The econometric models are structured around quarterly data. But, unlike with the mean, the sum of the four quarterly medians is not the same as the annual median household income. Table 4 shows that the discrepancies are not large, and we therefore use the sum of the four quarterly medians to provide estimates of the annual medians generated by the econometric models.



Source: ONS and authors' calculations

Figure 1:Measures of Household Income (£ 2015 prices CPI deflated)

Table 1: Tests for Stationarity

	Levels		First differences	
				p-
	DF	p-value	DF	value
LnHmed	-2.30	0.17	-8.61	0
LnRAWE	-2.32	0.16	-5.08	0
InRGDI	-2.09	0.25	-5.0	0

Variables

LnHmed log Equivalised Median Household Disposable Income deflated by CPI LnRAWE log Average Weekly Earnings deflated by CPI, seasonally adjusted LnRGDI log Gross Disposable income *per capita* deflated by CPI, seasonally adjusted Estimation period 2001Q2-2017Q1

Source; ONS and authors calculations

4. Application of the Different Methods

The Microsimulation Approach

The microsimulation approach can be broken down into three main stages – uprating the base dataset, the simulation of tax and benefit policies, and the reweighting of the survey weights.

First, the income components in the base dataset are uprated to the period of interest using data from the Office for Budget Responsibility (OBR) Economic and fiscal outlook (EFO) publication tables. The OBR back series incorporates data from other sources for periods which are available (see 'Table 1' of the Appendix for further information). When data are not available, uprating relies on OBR

forecasts. For this paper - which covers up to Q1 2017 - outturn data were available. Consequently, though the series used come from OBR's November 2017 EFO, the original source of the data varies. These are described within the EFO tables themselves, and summarised in the Appendix.

These adjustments are necessary to capture changes in income components over time, as there is a time lag between the quarters of interest and the initial input dataset. These 'uprating' factors generally change from quarter to quarter, though some factors remain unchanged between quarters within a financial year (e.g., income from government benefits, statutory sick pay, water rates).

Where possible, tax and benefit policy changes are modelled using HMT's Intra-Governmental Tax and Benefit Model (IGOTM), which is a 'static' microsimulation model for the UK tax and benefit system.

Within the tax system, the main taxes simulated are: Income Tax, employee National Insurance Contributions, Council Tax, VAT, Insurance Premium Tax, Fuel Duty, Alcohol Duty, Tobacco Duty, Stamp Duty Land Tax, and Air Passenger Duty.

Within the welfare system, the most significant welfare benefits simulated are: the State Pension, Pension Credit, Winter Fuel Payments, Attendance Allowance, Jobseeker's Allowance, Employment and Support Allowance, Income Support, Working Tax Credit, Child Tax Credit, Child Benefit, Disability Living Allowance, Personal Independence Payment, Tax-Free Childcare and Housing Benefit.

Not all households claim the benefits to which they are entitled. The microsimulation model accounts for this using information on benefit take-up from the underlying survey data. See HM Treasury (2017) for a more detailed description of IGOTM's underlying methodology and assumptions.

Changes to the tax and benefit system typically take place between financial years. Subsequently, whilst the microsimulation model was applied to each of the eight quarters, many of the parameters remained unchanged within quarters of a financial year, but changed between Q1 and Q2 in 2016 and 2017.

The simulation model outputs datasets at an adult, household and tax benefit unit level. Although very similar, the income measures produced through IGOTM are not conceptually identical to those used by ONS for its ETB publication. For example, the value of employer benefits in-kind such as company cars are included within ETB but not the IGOTM outputs. Therefore, where appropriate and possible, we make further adjustments to align the definition of income measures between IGOTM and ETB.

As a static micro-simulation model - aside from not accounting for behavioural responses to policy changes - IGOTM does not adjust for demographic changes over time. Consequently, the original ETB weights are re-calibrated to account for shifts in labour market participation and demographic characteristics of the UK population between the period when the LCF data were collected and the period for which nowcasts are being produced.

For the main ETB dataset and publication, each household in the microdata is initially given a design weight to account for the probability of selection in the sample. These weights are then adjusted to reduce bias from non-response and the sample distribution is calibrated to match the population distribution in terms of region, age group, sex and employment status - the latter of which allows the incorporation of labour market changes in the analysis.

To ensure consistency between the nowcasts and the actual data, it is desirable for the nonresponse adjusted design weights to be calibrated using new population totals matching those used for the original weights. Hence, the re-calibrated weights are calculated using the same calibration variables as the original ETB weights. Income outliers are then treated in the same way as for ETB. Under the version of the nowcasting methodology presented in this paper, individuals were grouped into twelve categories according to their economic status. Population totals are based on estimates directly from the Annual Population Survey (APS). As the economic status estimates and employment and unemployment rates are drawn from a sample survey (albeit one with a very large sample) the level of precision will be lower. Nevertheless, including this additional calibration constraint is important as changes in levels and patterns of labour market participation are likely to be a key driver in changes to household incomes. A full list of calibration groups can be found in the technical annex.

	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
	2015	2015	2015	2016	2016	2016	2016	2017
i) Base data median	6125	6125	6113	6143	6107	6076	6040	6011
 ii) Base data simulated, old weights, not uprated 	-41	-41	-41	-41	14	14	14	14
 iii) Base data simulated, uprated, original survey weights 	262	290	288	305	387	422	440	443
(iv) Base data simulated, uprated, and recalibrated	123	145	156	162	179	181	172	202
v) Final figures	6469	6518	6515	6568	6686	6693	6666	6670

Table 2: Stages in the Derivation of the Microsimulation Nowcast

Note: The figures in the table are deflated using CPI. The base data median (row i), in nominal terms, is constant across the series. Differences arise as a result of presenting the data in 2015 prices, hence reflecting inflation that took place over the period.

Source; ONS and authors calculations

Error! Reference source not found. shows the results using the microsimulation approach at each stage of the process. Row i) shows the median income in the original input dataset (referred to as the 'base data') - the 3 years of LCFS for financial years ending 2013, 2014 and 2015 - before any 'uprating' takes place.

Row ii) shows the impact of simulating the tax and benefit model on the base data. The figures are identical for Q2 2015 to Q1 2016, and separately for Q2 2016 to Q1 2017. This is because tax and benefit policies do not generally take effect within financial years. The step change between Q1 2016 and Q2 2016 represents a change in financial year, and hence a change in the underlying parameters used in the microsimulation model.

Row iii) shows the impact of 'uprating' the income components of the base dataset to the quarter of interest, and simulating the tax benefit system. This has the largest impact – making up an average of 70.7% of the difference between the base data median (row i) and final figures (row v) over the period. The steps shown so far do not account for any changes in labour status and population demographics over time.

Reweighting changes the original survey weights to allow for changes in demographic and labour market status over time. This is important as changes in levels and patterns of labour market participation are likely to be a key driver in changes to household incomes. Row (iv) includes all the steps up to row (iii), but is reweighted to the quarter of interest using population controls and employment status calibrations (further information is available in the Technical Annex). This has the second largest impact, making up an average of 33.5% of the difference between the base dataset and final values over the entire period. Note that the average differences between the base data and final figures sum to 100% - the negative differences shown in the first four columns of row ii) result in an average negative 4.2% contribution over the whole period.

Row (v) shows the final results once all the microsimulation approach steps have been taken.

The figures in rows (iii) and (iv) increase over time - illustrating the importance of adjusting and recalibrating the base data when there is an increasing time lag between the input dataset and the period of observation. The largest impact comes from uprating the dataset, though it should be noted that this could be due to second round effects coming from the simulation of tax and benefit policies. For example, if income components are 'uprated', the higher incomes observed will most likely impact on taxes paid and household eligibility for certain benefits. The second largest impact comes from reweighting - the calibration groups are detailed in the Technical Annex.

The nowcast results are compared with estimates produced using regression techniques in tables 4 and 5.

Simple Regression

The regression approach follows what was described in section 2. Table 3 shows parameter estimates for various specifications of the general model. Given that tests in Table 1 showed the variables to be I(1) we focus our attention on error correction models specified in first differences but with lagged terms in levels. The first column shows a model with both real wages and real disposable income *per capita* present. The term in the growth rate of wages has a negative coefficient, not an appealing feature of a nowcasting model of this type. We are, however, able to accept the hypothesis that these can be excluded, although only just since the p-value is 0.06.

The second equation therefore excludes the terms in real average earnings. The coefficients on lagged median income and lagged real disposable income add to close to zero, and the hypothesis that this is the case is also readily accepted with a p-value of 0.18. We therefore impose this restriction in the model shown in the third column of the table. The Breusch-Godfrey test for serial correlation allows us to accept the hypothesis that there is no residual serial correlation. This equation provides the structure we use for testing out of sample performance, although of course when we model out of sample, it is estimated recursively.

	(1)	(2)	(3)
	D.InHmed	D.InHmed	D.lnHmed
D.lnRGDI	0.578	0.439	0.466
	(1.58)	(1.41)	(1.71)
D.lnRAWE	-0.454 [*] (-2.35)		

Table 3: Parameter Estimates

LlnRAWE	-0.290		
	(-1.87)		
LlnHmed	-0.684***	-0.644***	-0.645***
	(-6.78)	(-6.16)	(-6.24)
LlnRGDI	0.937***	0.615***	0.645***
	(3.94)	(4.49)	(6.24)
Constant	-0.184	0.412	0.165^{***}
	(-0.20)	(0.51)	(6.56)
N	63	63	63
R^2	0.418	0.363	0.362
D.LnRAWE=0,	F(2,57)=2.93,		
LlnRAWE=0	p=0.06		
LlnHmed	_	F(1,59)=0.09,	
+LlnRGDI=0		P=0.76	
Estimation period 2001	Q3-2017Q1		

t statistics in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Source: ONS and authors' calculations

Using the data vintage of 1st June 2018 we estimate our models recursively. We start with parameter estimates for the period 2001Q2 to 2009Q1. These are used to provide fitted values for log real median income for 2009Q2 to 2010Q1 and the sum of these quarterly median figures is used to provide an estimate of the annual median. This can be compared directly with the deflated annual value of median income computed directly from the survey data. The exercise is repeated with parameter estimates for the period up to 2010Q1 used to compute nowcasts for 2010Q2 to 2011Q1.The model, however, requires a lagged dependent variable. We have an observation for this only in the first quarter of our four-quarter projection. For the other three quarters we have to use the value generated by the model itself.

We show the out of sample annual nowcasts generated in Table 4 of section 5 for the period from 2009/10 onwards. We also present in Table 5 quarterly nowcasts which are directly comparable with those generated by the microsimulation approach.

The Time-varying Approach

We show in Figure 2 the parameters estimated using the log of average earnings and log of gross disposable income per capital, both deflated by the consumer price index as indicator variables. Seasonal dummies and a constant are also included in the time-varying model.



Source: ONS and authors' calculations

Figure 2: The Parameters of the Time-varying Model with LnRGDI as Explanatory Variable

The coefficients are unstable and poorly determined. The coefficient on log gross disposable income is marginally significant at the beginning and end of the period, while the coefficient on log average weekly earnings is insignificant throughout. We therefore estimate a model containing only log real gross disposable income, and the seasonal dummies. As Figure 3 shows, this results in a coefficient which is highly significant throughout, and much more stable than in the earlier model. This finding is consistent with our earlier conclusions over the role of log real average weekly earnings in the regression model.



Source: ONS and authors' calculations

Figure 3: The Parameters of the Time-varying Model with LnRAWE as Explanatory Variable

The out of sample projections are generated as with the regression model. We take the value of the coefficient vector estimated for 2009Q1, and use this to produce projections for household disposable income for the four subsequent quarters. We then rerun the model on data up to 2010Q1 and repeat the exercise. This leads to the sequence of projections shown in Table 4 and Table 5.

5. Results

We assess the performance of our nowcasts by looking at the RMSE measured in logarithmic terms relative to the annual median from the datasets and shown in the final column of Table 4. The microsimulation model, however, stands out. The comparison can be made only over the last two years of the dataset, but over these the root mean square error is 0.01 whereas it is over 0.02 for the regression and time-varying models.

	Regression	Time-			ETB/HDII
	Model	varying		ETB/HDII	outturn
		model	Microsimulation	outturn (a)	(b)
2009/10	25465	25154	-	25460	25525
2010/11	24649	24496	-	25145	25106
2011/12	24146	24778	-	24650	24631
2012/13	24450	24877	-	24049	24061
2013/14	24517	24535	-	24813	24761
2014/15	24797	25077	-	25702	25666
2015/16	26092	26388	26070	26304	26313
2016/17	26022	26154	26716	26965	26986
RMSE	0.021	0.021	0.010		

 Table 4: Nowcasts of Median Equivalised Household Income (£ 2015 CPI deflated)

The nowcasts in the first three columns are the annual totals of the modelled quarterly medians. Outturn (a) shows the arithmetic average of the household median income in each of the four quarters in equation. Outturn(b) shows the household median income calculated from the full annual sample. Both series are deflated by the consumer price index (2015=100). The root mean square errors are calculated as the mean square log deviations relative to outturn (b).

Source: ONS and authors' calculations

Table 5 shows the performance of the four quarterly forecasts for each of the financial years 2015/16 and 2016/17. This provides a longer series between which to compare the forecasts, but gives very similar results. While the RMSEs are larger than those of the annual forecasts, the microsimulation model again outperforms the time-series forecasts.

		Time-		
	Regression	varying	Microsimulation	
	Model	Model	Model	Outturn
2015Q2	6463	6547	6469	6430
2015Q3	6518	6633	6518	6785
2015Q4	6557	6608	6515	6478
2016Q1	6553	6600	6568	6611
2016Q2	6554	6573	6686	6679
2016Q3	6539	6577	6693	6941
2016Q4	6496	6526	6666	6684
2017Q1	6434	6477	6670	6661
RMSE	0.031	0.027	0.02	

Table 5: Quarterly Nowcasts of Median Equivalised Household Income (£2015 CPI deflated)

The Diebold-Mariano statistic, as modified by Harvey, Leybourne and Newbold (1997) provides a statistical means of comparing the performance of different forecasts. A difficulty with the application of it in our case is that an adjustment is needed for serial correlation. Normally this is given by the distance of the forecast horizon. In our case, however, we do not have a fixed forecast horizon, since we are forecasting the four quarters of the financial year on the basis of information from the end of the previous financial year. Thus all we can do is examine the sensitivity of the test result to the assumed forecast horizon.

With a forecast horizon of one quarter, we find a p-value of 0.09 for the hypothesis that the regression model is no worse than the microsimulation model. For the time-varying model the p-value is 0.28. With a horizon set equal to 2, both p-values rise, to 0.21 in the first case and 0.37 in the second case. With longer horizons the p-values rise further. Thus it is not safe to conclude from these results that the microsimulation model out-performs either of the other models, even though the results are encouraging.

A further element of caution is needed. There are good reasons to think that the standard error of the forecast is larger than the RMSE shown in Table 4. There are many sources of uncertainty when compiling estimates using microsimulation techniques, including – on a general level – from a combination of simulation error and sampling variability as explored in Lappo (2015). This paper explained simulation error as the discrepancy between the real value compared to the simulated estimate. The sources of these errors can be broken down into methodological choices in building the model, the mathematical structure of the model and around the estimated model parameters (Bilcke et al., 2011). Lappo (2015) continues that in most cases, measuring simulation error is difficult due to the fact of not knowing the real values of the variables of interest. Therefore, estimates of simulation error to date have focussed on using microsimulation to model past situations and comparing outputs with actual values. Various studies have measured simulation error, including Zhou (2012) and Pudney and Sutherland (1994, p.338).

In addition, as microsimulation is performed using three years of LCFS data - a sample survey - the results are subject to sampling variability. In the ONS' nowcast and Effect of Taxes and Benefits on Household Income (ETB) publications, standard errors for mean and median equivalised disposable income can be calculated.

The average standard error of the two nowcasts produced in this paper was 0.9% - smaller than the comparable estimate for 2016/17 ETB (1.4%). This is due to the standard errors incorporating sampling error only, and the inverse relationship between standard error and sample size. As mentioned. the nowcast input dataset combines 3 years of LCFS data, whereas the ETB publication uses one only year of LCFS. Consequently, the sample for nowcasting is approximately three times larger than ETB.

6. Conclusion

in this paper we have compared the performance of a micro-simulation method of producing nowcasts with the results obtained by using standard regression and time-varying parameter methods. The nowcasts generated using microsimulation are available only for 2015/16 and 2016/17 but over this period their performance is considerably better than that of the two time-series methods.

Both microsimulation and time-series methods in fact produce quarterly rather than annual estimates, and the annual estimates we present are the sum of the four quarterly estimates. This is itself a source of inaccuracy, because, unlike with the mean, the annual median is not exactly equal to the sum of the four quarterly figures. However, it does make it possible for us also to compare the quarterly nowcasts against the data. Once more we find that the microsimulation method performs better than either of the time-series methods.

References

Baroni, E. & Richiardi, M. (2007) "<u>Orcutt's Vision, 50 years on</u>", Centre for Employment Studies, Working Paper no. 65,

Creedy, J. and I. Tuckwell, "<u>Reweighting Household Surveys for Tax Microsimulation Modelling: An</u> <u>Application to the New Zealand Household Economic Survey</u>", Australian Journal of Labour Economics, Vol. 7, No. 1, March 2004, pp 71-88

Dekkers, G. "<u>The simulation properties of microsimulation models with static and dynamic ageing –</u> <u>a brief guide into choosing one type of model over the other</u>", International Journal of Microsimulation (2015), 8(1), 97-109, International Microsimulation Association,

Figari, F., Paulus, A. & Sutherland, H (2014) "<u>Microsimulation and Policy Analysis</u>", ISER Working paper series, No. 2014-23, June 2014,

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*. Vol 13. pp. 281-291

HM Treasury, "Impact on households: distributional analysis to accompany Autumn Budget 2017", 2017:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file /661465/distributional_analysis_autumn_budget_2017.pdf

Immervoll, H. Lindström, K., Mustonen, E., Riihelä, M., & Viitamäki, H. (2005) "<u>Static data "ageing"</u> <u>techniques. Accounting for population changes in tax-benefit microsimulation models</u>", EUROMOD working paper series, EM7/05, March 2005,

Kump, N. & Navicke, J. (2014), "<u>Re-weighting EUROMOD for demographic change: An</u> <u>application on Slovenian and Lithuanian data</u>", Euromod working paper series, EM 13/14, June 2014,

Lappo, S (2015) "Uncertainty in microsimulation: Assessing sampling variability in inequality and poverty indicators through variance estimation", University of Helsinki, Master's Thesis, https://helda.helsinki.fi/bitstream/handle/10138/158960/Gradu Sampo Lappo.pdf

Lay-Yee, R. & Cotterell, G. (2015) "<u>The role of microsimulation in the development of Public Policy</u>", COMPASS Research Centre, The University of Auckland,

Li, J. & O'Donoghue, C. "<u>A Survey of dynamic microsimulation modes: uses, model structure and</u> <u>methodology</u>", International journal of microsimulation (2013), 6(2), 3-55,

Mallett, L. & Webber, D. (2017) "Nowcasting household income in the UK: financial year ending 2017":

https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomea ndwealth/bulletins/nowcastinghouseholdincomeintheuk/financialyearending2017

O'Donoghue, C., 2001a. Dynamic microsimulation: A methodological survey. Brazilian Electronic Journal of Economics 4, 77,

ONS (2015), "Nowcasting Household Income in the UK: Initial Methodology, 2015": <u>https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomea</u> <u>ndwealth/articles/nowcastinghouseholdincomeintheukinitialmethodology/2015-07-01</u> Orcutt, G. (1957) 'A new type of socio-economic system', Review of Economics and Statistics, 39(2), 116-123

Orcutt, G. "<u>A new type of socio-economic system</u>", cited in the International Journal of Microsimulation (2007), 1(1), 3-9

Pudney, S. & Sutherland, H. (1994) "<u>How reliable are microsimulation results?</u> : An analysis of the role of sampling error in a U.K. tax-benefit model," Journal of Public Economics, Elsevier, vol. 53(3), pages 327-365, March

Rastridina, O., Leventi, C. & Sutherland, H. (2015), "<u>Nowcasting: estimating developments in the risk</u> of poverty and income distribution in 2013 and 2014", Euromod working paper series, EM 12/15, August 2015

Stoynova, S. (2016) "Nowcasting household income in the UK: methodology, 2016": <u>https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomea</u> <u>ndwealth/methodologies/nowcastinghouseholdincomeintheukmethodology2016</u>

Sutherland, H. & Figari, F. "<u>EUROMOD: the European Union tax-benefit microsimulation model</u>", International journal of microsimulation (2013), 6(1) 4-26,

Appendix

Table 1: Uprating sources

Variable uprated	Series used
Income from employment, self- employment, odd jobs	Average earnings ¹
Income from annuities, private pensions and other income sources	RPI
Income from banks and building society interest	RDEP ² and NNMP ²
Income from dividends	NDIVHH ³
Private sector rent and rental income	Average Earnings
Income from main government benefits	Uprated in line with actual rates
Income from other government benefits (including JSA)	Uprated in line with actual rates
Statutory sick pay	Statutory Sick Pay rates
Mortgage interest	LHP; RMORT; number outstanding mortgages; interest payment per mortgage per year ⁴
Registered social landlords	Uprated in line with relevant rules in each devolved administration
Local authority rents (before rebates)	Uprated in line with relevant rules in each devolved administration
Water Rates	Average water bill ⁵
Council tax bills	Uprated in line with the average Band D rate in each devolved administration
Household expenditure	Household and non-profit institutions serving households final consumption expenditure (ABJQ + HAYE)

Source: Office for National Statistics

Notes:

1. Average earnings = Source: OBR Economic and Fiscal Outlook (EFO), November 2017. For periods that are not forecast, the back series are taken from Office for National Statistics data. Wages and salaries (Office for National Statistics UK Economic Accounts, Office for National Statistics identifier: DTWM-ROYK) divided by employees (Office for National Statistics Labour Market Statistics, Office for National Statistics identifier: MGRZ-MGRQ)

2. Deposit rates ("RDEP") = Source: Bank of England. Weighted average rates for sight deposits (Bankstats code: CFMHSCV) and time deposits (Bankstats code: CFMHSCW). Household and NPISH currency and deposits ("NNMP") = Source: Office for National Statistics.

3. Dividend receipts of households ("NDIVHH") = Source: Office for National Statistics, UK Economic Accounts Office for National Statistics identifier: NRKU

4. Mortgage debt ("LHP") = Source: Office for National Statistics. Secured debt on dwellings (Office for National Statistics UK Economic Accounts, Office for National Statistics identifier: NNRP); Mortgage rates ("RMORT") - Average interest rate on mortgages (Source: Bank of England; Bankstats code: CFMHSDE)

5. Average water bills are projected forward using price limits set by OfWat

Technical Annex

Calibration

As a standard procedure across the majority of ONS surveys, the LCF is calibrated to known population totals for region, age/sex groups and economic status (from financial year ending 2014). These population totals come directly from projections taken from the most recent Census, which are constantly updated with reliable information derived from birth and death counts, migration rates and immigration counts.

The LCF data are weighted at household level where the design weights represent the inverse probability of selection of a household. The weights are then adjusted to reduce bias from non-response, using scaling factors developed from information taken from the Census Non-Response Link Study (CNRLS). These design weights are then fed into Generalized Estimation System (GES), which adjusts the weights of each household, using information on the region of the household and the age and sex of household members (the latter often gathered by proxy). This calibration process uses known information to improve representiveness of the estimates across these groups. Recalibration of the existing weights involves using updated control totals and an additional constraint – economic status, which is taken from the Annual Population Survey (APS).

The new weights are calibrated to the population totals of the following Sex/Age groups and economic status:

- 1. Male/ female 0 to 15
- 2. Male 16 to 19
- 3. Male 20 to 24
- 4. Male 25 to 29
- 5. Male 30 to 44
- 6. Male 45 to 54
- 7. Male 55 to 64
- 8. Male 65 to 74
- 9. Male over 75
- 10. Female 16 to 19
- 11. Female 20 to 24
- 12. Female 25 to 29
- 13. Female 30 to 59
- 14. Female 60 to 69
- 15. Female 70 to 79
- 16. Female over 80

The following 12 regions:

1. North East

- 2. North West
- 3. Merseyside
- 4. Yorkshire & Humberside
- 5. East Midlands
- 6. West Midlands
- 7. Eastern London
- 8. South East
- 9. South West
- 10. Wales
- 11. Scotland
- 12. Northern Ireland

And the following employment groups:

- 1. Self-employed with children
- 2. Self employed without children
- 3. Full-time employed with children
- 4. Full-time employed without children
- 5. Part-time employed with children
- 6. Part-time employed without children
- 7. Unemployed and work-related government training programmes with children
- 8. Unemployed and work-related government training programmes without children
- 9. Retired or unoccupied and of the minimum National Insurance (NI) Pension age and retired or unoccupied and below the minimum NI Pension age with children
- 10. Retired or unoccupied and of the minimum NI Pension age and retired or unoccupied and below the minimum NI Pension age without children
- 11. Women between 60 and 64 in employment
- 12. Under 16