MARCH 2019

Homelessness

Causes of Homelessness and Rough Sleeping

Review of models of homelessness

alma economics

Homelessness

Causes of Homelessness and Rough Sleeping

Review of models of homelessness

Table of Contents

Non-technical summary

In order to inform a feasibility study for a suite of models about homelessness, we review and assess a set of models used in the UK and internationally for the following purposes:

- predict future levels of various types of homelessness (including statutory and single homelessness),
- appraise policy scenarios with respect to their impacts on homelessness,
- identify households and individuals at risk of homelessness in order to provide prevention services, and
- provide accurate estimates of homeless populations that are possibly hidden and elusive.

Specifically, we review and assess the characteristics of the following models

Economics-based simulation models are the main approach used for policy appraisal purposes. They depend on a solid theoretical framework to quantify behavioural responses to policy and economic changes. They use these quantified relationships to predict how planned policies and changes in economic variables will determine future levels of welfare outcomes.

The model developed by Bramley and colleagues is a complex version of a simulation model. It is the main tool currently used to produce projections of aggregate levels of homelessness across UK regions. The basic idea behind this approach is that housing needs are the outcomes of households, individuals and firms interacting in interconnected housing and labour markets across the UK. The model is based on a set of core functions used to quantify responses to changes in economic and policy variables with respect to outcomes such as migration and household formation that potentially determine housing needs.

Time series models are geared towards producing accurate forecasts of outcomes of interest in the short to medium-term based on past trends. They arrive at forecasts either by simply using past (lagged) values of the outcome of interest or by accounting for relationships between the outcome of interest and other explanatory variables over time. Their main objective is to minimise the errors in forecasting future values of the outcome of interest. They are applied to forecasting welfare outcomes such as homelessness, income inequality and poverty.

Machine-learning methods are also increasingly used to forecast trends in welfare outcomes such as poverty and homelessness. The basic concept behind machine-learning techniques is that model building can be automated without determining relationships between variables in advance. Therefore, machine learning models cannot be used to test hypotheses or estimate of causal relationships.

Homelessness risk models are used to assess the likelihood of facing homelessness at the individual or household level. The models calculate risks of homelessness conditional on a set of predicting factors. They are primarily used by prevention services to identify households and single people who are at priority need for homelessness prevention services.

Non-standard sampling models are used to measure populations that are not straightforward to capture such as homeless groups that are often underestimated, for example, because of varying definitions of homelessness and transient nature of homeless populations. These methods arrive at estimates of population counts by extrapolating parts of the populations that can be observed and measured. They can be used to either guide new data collection or estimate population size using existing survey data.

The **assessment of the models** included in this review is guided by the following set of criteria that reflect characteristics that are important for developing a suite of models of homelessness:

- The **resources-**related group comprises criteria relating to levels of expertise and effort demanded to use the models as well as the required data inputs.
- The **communication**-related group assess how straightforward it is to interpret and replicate model findings.
- The **application**-related group assess the models' relevance to policy, flexibility as well as output accuracy and level of granularity.

Our aim is to discuss the appropriateness of the models in different contexts. We focus on the assessment of the policy model initially developed by Bramley et al. (2010), as this is the main model used in the UK context for homelessness projections and policy appraisal. We group the remaining types of models in two categories according to their objectives: i) models for homelessness measurement at present and future times and ii) models for assessing homelessness risks at the individual and household level.

Below, the most important **findings from the model assessment** are summarised:

The **model developed by Bramley and colleagues** was not developed to be easily accessible by non-experts. It relies on a complex set of assumptions to arrive at projections of housing needs and homelessness as the outcome of the interplay between various policy and economic factors. It follows that updating, revising and assessing the validity of the model's assumptions, for example by using alternative datasets, requires high levels of statistical expertise.
The model is centred around providing a comprehensive theoretical

framework to explore how homelessness trends respond to policy and economic changes rather than arriving to accurate estimation of future values. Its structure allows to capture broader areas of policy that potentially influence homelessness.

The group labelled **models for homelessness measurement** includes the following models:

- time series models,
- machine-learning models, and
- non-standard sampling models.

The models under this category tend to be relatively simple – their development demands standard statistical concepts and can be easy to understand and replicate. They are mainly used to arrive at accurate measurement of homelessness levels in present and future periods rather than evaluate the impact of policy and economic changes. For example, machine learning models arrive at projections based on the iteration of random processes which can be viewed as a “black box”, not allowing users to observe transparent links from explanatory variables to outcomes of interest.

Time series models are the most straightforward method applied to generate accurate short-term forecasts using a limited sets of time series data (either including the dependent variable only or more explanatory variables). Similarly, non-standard sampling methods are simple techniques that can be applied to accurately measure elusive homeless populations concurrently.

The group of **models for assessing homelessness risks** are fairly easy to implement and interpret. While the complexity of these models depends on the particular technique employed to calculate homelessness risks, they can be further simplified to serve as decision tools for workers at prevention services. Moreover, they are shown to contribute to correct targeting of individuals and households that are in priority needs and result in cost savings of prevention services.

In conclusion, the present review highlights the fact that different types of models are used for different purposes. An optimal methodological framework of empirical analysis used to inform homelessness policy should include different types of methods aiming to accommodate various goals. Models included in this framework should be selected according to their comparative advantages in achieving their objectives.

1. Scope

This review is part of a wider study aiming to explore existing methods for the development of an empirical framework used for homelessness policy purposes. Recommendations about suitable models are guided by i. a rapid evidence review that maps existing evidence on a wide set of factors that predict homelessness in the UK and abroad and ii. assessment of characteristics of existing models about homelessness when applied for different purposes. These two distinct tasks (rapid evidence review of literature on homelessness predictors and homelessness models review) provide an evidence base for the next step – the feasibility study.

The main objective of this report is to review and assess the characteristics of empirical models used to inform policy centred around tacking homelessness and providing support to homeless people. This assessment aims to provide insights into relative strengths and weaknesses of different types of models regarding their appropriateness for different purposes.

Empirical methodologies are central to a number of different policy objectives. For example, empirical approaches are adopted to explore the influence of planned policy changes on homelessness. Moreover, quantitative techniques can be implemented to improve the effectiveness of prevention services in targeting households and single people on the verge of homelessness.

This review discusses the relevance of empirical models applied in the UK and overseas with respect to policy purposes about homelessness. We focus on the main model developed by Bramley et al. (2010) and its extensions (for example, Bramley, 2017; Bramley et al., 2016 and Bramley and Watkins, 2016) as this is the key policy model used in the UK context to project homelessness and appraise composite policy scenarios. Additionally, we go beyond this basic model to discuss a broader range of methods that can be adapted to measure and predict homelessness in the UK.

The selection of the models included in the review is limited by the availability and accessibility of relevant documentation. For example, our summary and review of the main UK model is mainly based on published reports rather than actual copies of the model. Additionally, it is reasonable to assume that other countries will have similar ad hoc proprietary models that reside within government departments and are, thus, not accessible by the public.

The models that we were able to identify in the literature are geared towards the following objectives:

- produce reliable predictions of homelessness trends in the future taking under consideration existing trends and forecasted changes in determinants of homelessness,
- evaluate multifaceted policy

scenarios, such as integrating housing support services, such as social rented dwellings provision and housing allowance, as well as other types of welfare assistance, such as unemployment benefits,

- identify households and individuals that are likely to face homelessness in order to offer timely prevention support, and
- accurately measure the size of homeless populations, including hidden homeless or rough sleeping populations.

Our primary objective is to assess whether each model could potentially be included in a methodological framework designed to accommodate a broad range of policy targets in England. To avoid limiting the discussion to narrow applications of each model, we group them in broad classes that share common characteristics. It should be noted that there are usually no strict diving lines between the classes that are discussed in this review. Instead of searching for the perfect typology, we aim to use working definitions that will facilitate the discussion around relative strengths and weaknesses of homelessness models.

We use a set of criteria that reflect model properties related to the resources required to apply the model, the ease of communication of model outputs and its application.

The assessment is focused on the model developed by Bramley and colleagues, as it is the main policy tool currently used by UK public bodies.

For the remaining classes of models, we aim to evaluate relative strengths and weaknesses with respect to how effective they are in reaching their objectives. Models are categorised according to the purpose for which they were designed. Specifically, two groups are identified: i) models for measuring the size of homeless populations both currently and in the future and ii) models for identifying households and individuals at risk of homelessness.

The report is structured as follows: section 2 reviews classes of models used to measure homelessness levels and calculate homelessness risks concurrently and in future periods. Section 3 discusses our approach to assessing these models and sets out the criteria that will guide our assessment. Section 4 discusses the strengths and weaknesses of the classes of models grouped according to their objective. Finally, section 5 concludes.

2. Classes of models

2.1 Economic-based simulation models

Economic-based simulation models are used to project welfare outcomes and explore potential changes in these outcomes under alternative economic and policy scenarios. They rely on a solid theoretical framework to establish causal effects between a set of key determinants, such as demographic characteristics, policy indicators and economic variables, and outcomes of interest.

Simulation models are quite broad – they are able to integrate the full range of policy variables that are relevant to a particular outcome. Their level of complexity depends on the range of paths to the outcomes of interest that policy makers need to explore. They are predominantly used as policy appraisal tools and they are usually ad hoc in the sense that they are designed to accommodate specific policy needs.

According to the description of the simulation model developed by Bramley et al. (2010) to project housing needs in the UK context, a simulation is

> *"an imitation of a real situation or prospect, based on a model, which can be convenient for training or demonstration purposes. When dealing with a very large reality, such as the UK economy or society, a simulation may be a way of carrying out a variety of experiments, to answer 'what if?' questions, without the enormous cost, risk and time involved in doing these things for real. Simulations are often used with economic models, to show the trajectory of a set of relationships under different conditions. In social policy, proposed changes to benefits or taxes are routinely 'simulated' based on large scale survey datasets, …, to demonstrate their cost and impact on different groups" (Bramley et al., 2016)*

Models in this class mainly consist of two pillars:

i. knowledge about parameters that describe behavioural responses to changes in key determinants of the outcomes of interest. These parameters can be either observed in the literature or estimated using econometric models. Past survey and administrative data are used to estimate links from changes in a set of predicting factors to outcomes of interest, and

ii. simulation of future trends in the outcomes of interest, conditional on projected changes in the explanatory variables. Estimated parameters (elasticities) from the first stage are applied to projections of explanatory variables to arrive at predictions.

Simulation models can be used to produce medium to long-term projections and appraise planned policies. While there are models that can do both, such as the model developed by Bramley and colleagues, these two objectives are distinct – in other words, it's possible to have simulation models that can only produce projections assuming a continuation of the policy status quo. It is also possible to have simulation models that estimate the effect of different policies on the 'no change' baseline without themselves producing an estimate of what the 'no change' baseline looks like.

Estimating Housing Needs (EHN) model

The Estimating Housing Needs (EHN) model was developed by Bramley et al. (2010) for the DCLG National Housing Planning Unit to predict housing needs. It is a regional model that makes conditional forecasts of outcomes related to unmet housing needs. While statutory homelessness is one of the needs outcomes that the model estimates, it is not the central focus of the model.

The EHN model is a sophisticated, medium-sized model that goes beyond existing approaches by integrating key demographic, economic and policy processes that influence housing needs. While previous models produced a single estimate of housing needs (measured by numbers of dwellings – mainly social rented – that should be provided over a certain period to meet estimated needs), the EHN model adopts a more comprehensive approach, looking into a broader set of outcomes related to housing needs.

The model is also used to evaluate the effectiveness of the provision of housing support services by calculating the number of welfare recipients. The results are contrasted with the estimated numbers of households and individuals that fall in the housing needs categories assessed by the model.

The major contribution of the model to existing literature (for example, Holmans, 2001; ODPM, 2006) is that it considers the influence of changes in demographic, economic and personal factors on housing needs instead of relying on mechanic extrapolations of existing trends in household numbers and types. Housing needs are modelled as products of dynamic interactions between households, individuals and firms in local housing markets. Moreover, the model considers the influence of the interplay between personal circumstances and external economic conditions on unmet housing needs occurring. The model also has a relatively granular level of geography – it accounts for potential interdependences between labour and housing markets across nine regions in the UK.

The EHN model also features important policy components, considering the influence of existing policies on housing needs and allowing for the appraisal of potential policy scenarios. For example, it simulates levels of housing needs under different housing supply scenarios (for example, low-cost home ownership provision and social rented housing supply). Moreover, it explores how homelessness levels respond to changes in social housing allocation priorities and credit rationing.

First stage of the model

The foundation of the simulation model is a series of econometric 'gross flows' models that estimate housing needs as well as other related outcomes around two key processes; namely, household formation and tenure choice. Findings from these models are integrated into the macro-simulation model that projects the evolution of UK housing market and housing needs at the national and regional level.

The first stage of the EHN model comprises four main modules around:

- *household formation*[1] modelled as a function of housing and labour market conditions (for example, employment/unemployment levels, house prices, supply of social lettings) as well as individual characteristics (for example, marital status, age, children, gender, relationship breakdown, migration, employment etc.),
- *housing market* using inputs from the Reading-CLG Affordability model (NHPAU, 2009) on variables such as house prices, affordability ratios and migration flows to model effects of different housing supply scenarios,
- *tenure flows* conditional on economic (such as credit constraints) and demographic factors. Outcomes of interest are the likelihood of mobility, the choice to buy and the choice/opportunity to move to social housing and private renting, and
- *specific housing needs* including concealed households, sharing households, existing affordability problems, overcrowding, unsuitable accommodation, unsatisfactory house conditions and statutory homelessness.

Figure 1 depicts the model pillars and their connections. As shown in the figure, estimated outcomes related to household changes (household formation, dissolution, migration), tenure flows (determined by affordability and credit constraints) and housing markets are then used as inputs to the housing needs model.

Using past surveys and registry data, the behavioural models estimate elasticities that reflect how sensitive the outcomes of interest are to changes in explanatory factors. Key baseline data for the behavioural models are mainly drawn from the Survey of English Housing (SEH, over

[1] Household formation is measured by the probability that an individual adult will be a household representative.

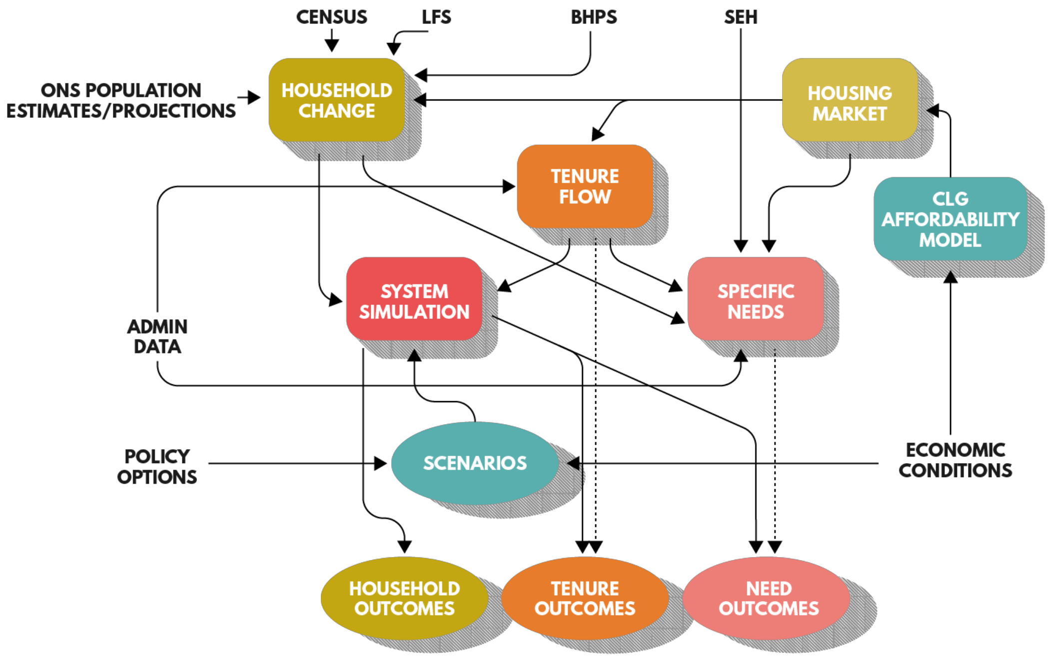


Figure 1. Schematic diagram of the Bramley et al. (2010) simulation model. Source: Bramley et al. (2010).

11 years to 2007/8), Labour Force Survey (LFS, 1992-2008), the British Household Panel Survey and the Continuous Recording System of lettings and sales in social housing (CORE).

Second stage of the model

To project housing needs and changes in household formation and tenure choices, the estimated elasticities from the first stage are applied to forecasted trends of explanatory factors and planned changes in policies. The model produces projections of medium to long-term trends in housing needs across age and tenure groups under various economic and policy scenarios.

At the second stage, macro-simulations are conducted using forecasts of key determinants from two main sources: i. the Reading-CLG Affordability model and ii. simple time series models based on LFS data.

Outputs from the DCLG Affordability model (NHPAU, 2009), which is mainly developed to predict the affordability ratios under different economic and housing supply scenarios, are used for a number of variables – for example, house prices, earnings, migration and employment. This approach is mainly limited in that the Affordability model fails to capture labour market changes accurately.

Trends in demographic factors such as populations in minority groups, distribution of different socio-economic statuses and student populations are forecasted applying simple trend-based techniques using quarterly LFS data. Simple time-series methods are also applied to forecast a set of explanatory variables including private rents, social lettings, vacancies and number of new households. Baseline figures are used for variables for which forecasts are not available or cannot be generated based on available data (for example, deprivation and crime).

The final outputs of the model are:

- size and composition of households at main tenure and age groups (for example, under 40 and over 40 years of age), and
- the incidence of types of housing needs among types of households and age groups at future dates.

Box 1. Homelessness projections in the EHN model

At the first stage, data from local authorities (LAs) covering the 1993-2008 period are used to quantify the contribution of a set of explanatory variables to the incidence of statutory homelessness. A behavioural model that comprises two steps is estimated. In the first step, priority acceptances of homelessness applications are modelled as a function of a set of individual and structural predictors. In the second step, the number of households in temporary accommodation is estimated taking into consideration the estimated acceptance rates from the first step.

The model includes a range of demographic and socio-economic predicting factors along with indicators for policies around homelessness. Particularly, it quantifies the effects of preventive interventions that are shown to reduce homelessness such as home visits, floating support referral, formal external mediation, sanctuary schemes, homeless prevention fund, etc.

Relationships between key predicting factors and homelessness measures (acceptances and households in temporary accommodation) are estimated in the form of elasticities.

It is found that house prices and income are important predictors of homelessness priority acceptances with 0.18 and -0.48 elasticities, respectively. In other words, a 1% increase in house prices (household income) predicts a 0.18% increase (0.48% decrease) in the share of households that are homeless. Moreover, supply of social lettings is positively correlated with homelessness levels potentially reflecting the market effect of increased supply leading to increased demand. On the other hand, increased private renting appears to be associated with reduced homelessness levels. This finding could indicate that increased supply of housing opportunities in the private sector can prevent some people from becoming homeless, thus, resulting in homelessness reduction.

The model also considers the influence of demographic factors that are shown to predict homelessness in existing literature. Particularly, indicators for groups of young adults (elasticity = 0.89), in-migration flows (elasticity = 0.35) and ethnic minorities (elasticity = 0.17) are positively correlated with homelessness. A positive relationship is also found between crime rates and homelessness (elasticity = 0.24). This finding suggests that homelessness is likely to be higher in areas with high deprivation levels.

Estimated elasticities from the two-step behavioural model are incorporated in the macro-economic simulation to arrive at annual projections of homelessness. The parameters are multiplied with forecasts of key predictors such as demographic factors, tenure flows, household formation outcomes and policy variables to predict proportional changes in statutory homelessness levels in the future.

In agreement with previous literature on the predictors of homelessness in the UK, the model reveals that homelessness projections are quite elastic to changes in affordability of housing, poverty and personal characteristics (for example, gender and marital status). Moreover, the findings of the model highlight the contribution of prevention measures in tackling homelessness. Particularly, the introduction of a comprehensive set of homelessness prevention schemes is predicted to reduce acceptance rates by 48% of the mean value.

Sub-Regional Housing Market Model (SRHMM)

The Sub-Regional Housing Market Model is an extension to the EHN model aiming to inform decisions about housing provision in England (Bramley and Watkins, 2016). The model was primarily designed to guide the decisions of LAs regarding new construction. It can also be used to predict future levels of unmet housing needs and appraise composite policy scenarios. Alternative versions of the model were used to predict trends in housing markets in Scotland (Leishman et al., 2008) and New Zealand (Bramley, 2013).

The SRHMM considers housing as a predominantly market-based system wherein individuals, firms and households adjust their behaviour. Housing needs are modelled as the outcomes of a complex process of interactions between different agents in local housing markets.

SRHMM goes beyond the EHN model by explicitly considering spatial interdependence between the regional housing and labour markets in England. Moreover, two new modules are included in the first stage of the model: planned housing supply model and migration model. Estimated in-migration flows are, then, incorporated in other core functions of the model to predict outcomes related to tenure choices and household formation.

In order to arrive at projections of housing needs, the model draws estimated parameters and elasticities from EHN underlying models.

It also produces new estimates of predicting factors using the additional modules for new housing supply and in-migration flows.

The following core functions comprise the first stage of the SHRMM model:

- *migration* modelled as a function of house prices, interest rates, household income, unemployment, the share of social renting, low-income poverty, ethnicity, academic status (e.g. student) and environmental factors (density, sparsity, greenspace, air quality, climate, scenic areas),
- *household formation* modelled as a function of previous tenure, high SEG, being sick or disabled, academic status (e.g. student), ethnicity, previous household type, unemployment, income, house prices and housing supply indicators (including policy variables such as social sector letting rate),
- *new construction* (housing supply) modelled as a function of new planning permissions, the stock of existing permissions, the share of small sites, previously developed land share, the share of green space, house prices and mortgage interest rates,
- *house prices* as a function of the ratio of population to dwellings, private vacancies, lagged flow supply of new mix-adjusted completions, household income, unemployment, crime, climate, prices in surrounding areas,

- o *market rents* as a function of mortgage cost to income ratio (i.e. taking account of house prices, mortgage interest rates, and the inverse of income) and household income level.

These models estimate elasticities that reflect interactions between four key streams of outcomes: labour market, demography, housing market and housing supply. As in the EHN model, these elasticities are then applied to forecasts of key predicting factors to simulate future trends in housing needs conditional on various factors.

Extended versions of SRHMM

The SRHMM model was further expanded to project alternative welfare outcomes such as poverty and inequality (Bramley et al., 2016) and types of homelessness other than statutory homelessness – for example, rough sleeping and sofa surfing (Bramley, 2017).

Newer versions of the model cover 116 sub-regional housing market areas (HMAs) across the whole of the UK. In the extended versions, the behavioural models are re-estimated using survey data that cover more UK countries. For example, the British Household Panel (BHPS, 1992-2008) and the UK Household Longitudinal Study - Understanding Society (UKHLS, 2009-11) are used along with the English Housing Survey (EHS, 1997-2007).

Moreover, a set of modules in the first and second stages of the model were revised and extended. For example, econometric models for demographic components and private rents were developed. The macro-financial framework of the model was also expanded.

Underlying models were revised to consider changes in the housing policy framework. Particularly, policies regarding Right-to-Buy, Low Cost Home Ownership (LCHO), private renting regulation and property and council taxes were modelled. For example, in order to reflect policy shifts from social lettings towards LCHO, the models were extended to include LCHO affordability measures and composite discount parameters integrating various LCHO schemes (such as Starters Home and HomeBuy shared ownership).

In addition to statistic simulations, extended versions integrate dynamic macro-simulations that model processes of behavioural reactions and interactions in response to initial changes. Dynamic simulations are mainly used to model processes related to working status and tenure. For each outcome of interest (for example, first-time buyers), households that are likely to make a transition to a new state depending on observed demographic and economic drivers are identified. Then, propensities for the incidence of such transitions are estimated conditional on this set of drivers and applied to a random selection of households that are likely to undergo changes in status. For example, changes in working status (e.g. employed vs unemployed) are estimated based on incentives varying with respect to changes in taxes and benefits and actual changes.

The extended versions of the model can accommodate the appraisal of composite scenarios incorporating the following set of variables:

- housing supply policies (increase in overall or social housing supply, increase in 'affordable housing' supply – i.e. LCHO)
- rent and policy options for social housing and for subsidised private renting (e.g. "Living Rents" and "Affordable Rents")[2]
- property taxation options (property tax vs council tax)
- migration flows and demographic growth,
- regional economic growth trends,
- housing supply measures,
- pay policies (such as National Living Wage, Full Living Wage, closing the gender or part-time pay gap)
- state benefits scenarios, and
- scenarios on family breakup and childcare.

Micro-simulation feature in versions for projections of poverty outcomes

The version of the model (Bramley et al., 2016) aiming to project poverty outcomes includes one extra feature: a static micro-simulation stage that produces snapshots of welfare outcomes in future periods at a more granular level – for example, projections disaggregated at the individual and household level.

At the micro-simulation stage, snapshots of projected outcomes of interest in future dates are transferred to a set of microdata drawn from the 2011 Understanding Society Survey (UKHLS). The variables drawn from the SRHMM are inserted in the micro-simulation stage as multipliers that quantify projected changes compared to the baseline period. The multipliers are then used to re-weight the UKHLS model population at baseline (2011) in order to reflect these projected changes. For example, if the proportion of working-age single person is projected to have risen from 20% in 2011 to 30% in 2041, the multiplier would be 1.50.

The micro-simulation model does not conduct longitudinal analysis to arrive at long-term projections of welfare outcomes under different scenarios. Instead, it measures sizes of population subgroups and assess their key characteristics in the future by looking at representative observations falling within particular grouping.

[2] "Living Rents" is system proposed in Lupton and Collins, (2015) that linking social housing rents to earnings. "Affordable Rents" are based on 80% of market rents (65% in London) and were used as basis for funding new social housing between 2011 and 2015.

Adaptation of extended SRHMM versions to homelessness projections

The SRHMM was extended to produce projections of homelessness measures under different policy and economic scenarios on behalf of Crisis (Bramley, 2017). The model's outputs have also been used by MHCLG for policy decision-making purposes.

This extended model includes 15 more variables allowing for the estimation of different types of homelessness. The additional variables include indicators for levels of types of homelessness other than statutory homelessness such as people who sleep rough, hostel residents and sofa surfers.

The set of underlying functions of the SRHMM model is augmented by including new behavioural models to estimate the effects of a wide set of predicting factors on other types of homelessness (for example, single homeless people and people who sleep rough). All the model elements that are related to homelessness are based on newer estimations of the core functions using a mix of sub-regional panel datasets and data drawn from surveys such as PSE, UKHLS and BHPS.

For example, to arrive at counts of rough sleeping populations in future periods in England, two behavioural models were used. First, a linear regression model was estimated using data from the Poverty and Social Exclusion (PSE) 2012 survey. The survey includes a retrospective question about homelessness experiences where people who have gone through rough sleeping can be identified. The model featured a set of predictors including crime rates, single person households, low-income poverty, unemployment and unsuitable temporary accommodation.

A logistic regression model was also estimated to predict rough sleeping conditional on covariates such as age, single person households, poverty, past poverty, insecure tenancy problems, criminal records and net-in migration. The final projections of rough sleeping population counts were produced by averaging the outputs of these two models.

Similar approaches were also adopted to model other outcomes related to homelessness, such as counts of sofa surfers and households in temporary accommodation. For example, two logistic models were estimated using different data inputs to arrive at predicted counts of sofa surfers. The first model was estimated using UKHLS data and included the following predictors: age, migration flows, household types, poverty, income, tenure, financial difficulties, unemployment, crowding, social housing supply, job growth or decline, recent changes in levels of temporary accommodation. The second model was estimated using PSE data and included an additional set of indicators for material deprivation, past poverty, criminal records and the availability of mental health institutions in the area.

Projections of alternative homelessness types are produced at the macro-simulation stage based on estimated elasticities from these new models.

Different data sources can be used to generate a range of projections for the same outcome of interest related to homelessness. For example, using administrative homelessness data is likely to result in conservative estimates of homelessness types in future periods. These results might underestimate the actual sizes of homeless populations by depending on data that do not capture households and individuals who do not register to LAs. On the other hand, using data drawn from retrospective surveys is likely to result in more realistic outputs.

Finally, this version of the model can be used to explore variations in homelessness levels under distinct policy scenarios. For example, it has been applied to assess the potential effects of:

- removing the additional planned welfare cuts for the 2016-2021 period,
- increasing the housing supply (including social/affordable housing), and
- implementing best-practice across prevention services in all regions and a policy mix aiming to promote economic growth at the regional level (integrating fiscal, industrial, infrastructural and educational measures).

Box 2. Summary of simpler simulation models: the World Bank approach to projections of welfare outcomes

World Bank models developed to predict outcomes related to poverty and inequality are a typical example of simpler simulation models used to predict welfare outcomes. The basic idea behind this type of models is to measure outcomes for which data is not available by using other key determinants that are observable and easier to forecast.

The World Bank models produce projections of outcomes mainly centred around absolute poverty such as headcount ratios (*share of population that live under the poverty line*) and poverty gaps (*average difference between the poverty line and the income/consumption of those who are below the line expressed as a share of the poverty line – a measure of poverty intensity*).

Poverty outcomes are predicted either at more frequent time intervals or at lower levels of aggregation – for example, at the sub-regional level. As poverty outcomes are not directly observable using existing data, the models rely on the assumption that wealth, income and consumption distributions for which data is available are important determinants of poverty and inequality outcomes.

The EEL model

This approach is based on a model initially developed by Elbers et al. (2002) that uses census and survey data to produce reliable projections of poverty and inequality at disaggregated (sub-regional) levels. The model takes advantage of the high levels of coverage of large-scale data – such as census data – and high levels of granularity of survey data to estimate the distribution of an observed variable that is assumed to determine the outcomes of interest.

The model consists of two stages:

1. the distribution of a key determining variable is estimated using detailed survey data conditional on various demographic and economic determinants, and
2. welfare outcomes are simulated using Monte Carlo techniques – estimated parameters from the first stage feed into this stage as weights.

In the first stage of the model, the distribution of household consumption is estimated conditional on a series of covariates using detailed survey data. Covariates that can be linked to larger scale data – such as census data or national surveys with high coverage – are selected. This estimated distribution is, then, extrapolated to a larger scale – e.g. at the national level – using large-scale data. It can also be used to estimate consumption levels for any subgroup of the population conditional on the selected covariates.

In the second stage, the distribution of poverty measures is simulated based on the estimated distribution of consumption. The final outputs of the model can be disaggregated at lower levels.

The Douidich et al., (2015) model

This approach has been a source of great influence for the development of models estimating welfare outcomes at future time periods and at different levels of aggregation. For example, Douidich et al. (2015) built on the ELL model to predict headcount poverty measures at higher frequencies. This version of the model uses LFS quarterly data instead of population census data and is widely applied to produce World Bank projections of poverty in developing countries. For example, it has been used to provide reliable projections of poverty in Morocco (Douidich et al., (2015) and Jordan (Dang et al., 2017).

The Douidich et al., (2015) model employs a standard imputation technique to arrive at frequent predictions of outcomes when data is not available. For variables that are not available in key datasets, other datasets are used to fill the gaps. A basic prerequisite is that the two datasets share a set of regressors that are correlated with the missing variable. In the first stage, a behavioural log-linear model is used to estimate the distribution of the observed variable of interest – consumption for instance – as a function of household characteristics – such as demographics, education, employment, housing conditions and wealth.

Following the ELL method, estimated relationships between key covariates and variables of interest are inserted in the macro-simulation model as parameters in the second stage. The simulation model is then fitted to data drawn from large-scale, household surveys conducted at frequent intervals – e.g. the LFS – to arrive at frequent predictions of welfare outcomes.

The suggested methodology can be applied to any welfare indicator that is a function of household income or expenditure – such as the poverty gap or Gini coefficient. Given that poverty emerges as a major indicator of homelessness (Anderson and Christian, 2003; Bramley and Fitzpatrick, 2018), the ELL methodology and its extensions can potentially be incorporated in a suite of models around homelessness aiming to produce frequent estimates of homelessness in England as the result of changes in absolute and relative poverty.

While this model was designed to generate predictions of welfare outcomes rather than appraise planned policies, its structure allows for exploring the sensitivity of outputs to changes in policy variables. It can potentially be extended to other applications, including simulation of policy reforms and economic shocks.

Box 3. Policy model: The IFS model for living conditions, inequality and poverty

The Institute for Fiscal Studies (IFS) uses a simulation model to produce projections of *absolute* and *relative* poverty, inequality and livings conditions given existing tax and benefits policies (Browne and Hood, 2016; Hood and Waters, 2017). The model can be also applied for appraisal of potential policy reforms – for example, extending a freeze in benefits, transitions to Universal Credit, etc.

The IFS model estimates poverty and inequality in future periods based on projected distributions of income. It is built on the assumption that the income distribution will change in line with changes in key determinants such as earnings, pensions, benefits and taxes as well as demographics.

It consists of two stages:

1. future income distributions are predicted based on both past trends and changes in key predicting factors – behavioural models are used to quantify relationships between key determinants (e.g. tax and benefits) and income, and

2. a simulation model that projects poverty and inequality outcomes using projected income distribution as input.

In the first stage, projections of the whole income distribution are produced based on past data on household income adjusted for future changes. The samples are drawn from surveys such as the Family Resources Survey (FRS) and the Family Expenditure Survey (FES) and are assigned weights so they match actual populations. Projections are based on changes in these weights that reflect demographic changes with respect to various factors, such as age, region of residence, gender, employment and household type.

Future income distributions also depend on the evolution of income components over time. Therefore, all financial variables are adjusted for projected changes as predicted by Office for Budget Responsibility (OBR).

Additionally, the IFS model quantifies changes in household incomes over time based on policy reforms regarding taxes and benefits. The IFS model draws on outputs from the TAXBEN model regarding projections of tax liabilities and benefits take-up at the household level. TAXBEN is a model commonly used to inform public policy in the UK. It is primarily developed to calculate changes in the fiscal system on a sample of households, that are then extrapolated to the entire population (Giles and McCrae, 1995). The model estimates benefits receipt and entitlement as well as tax liabilities conditional on various demographic and economic factors and their interdependencies.

The outputs from the static simulation stage reflect what will happen to poverty, inequality and living conditions if the latest microeconomic forecasts prove to hold true. It is assumed that while levels of the variables included in the models might vary across UK regions, growth rates are similar.

The IFS simulation model is static – it does not capture dynamic behavioural responses of individuals and households to changes in the economic and policy environment. However, it considers the impact of policy on individual and household characteristics in every step of the process.

For example, potential effects of announced social rents policies are considered when making future levels of poverty through the following process: The impact of changes in social rents policies on housing costs is quantified in the first stage of the model. Then, estimated housing costs conditional on social rents policies are used to arrive at the estimated distribution of household income that is then used to project future levels of poverty.

2.2 Time series models

Time series models are used to forecast outcomes of interest based on existing trends. They mainly use past (lagged) values of the dependent variable to generate forecasts. They can also be extended to account for temporal relationships between the dependent variable and other determinants. They use time series data, i.e. sequence of data points, stored in time order, that are usually recorded at frequent intervals (for example, quarterly data).

The main objective of time series models is to minimise errors in estimates of future values of the dependent variable. Complex relationships and estimation methods are often avoided to improve the accuracy of outputs. Moreover, it is common practice that more than one model designs are used jointly to predict single outcomes to improve accuracy.[3]

Time series models can arrive at forecasts using different techniques – for example, depending on the desired time horizon of the forecasts. In order to generate short-time forecasts, they depend heavily on the latest observations in the sample. On the other hand, when applied for medium-term predictions, they can be adjusted to place more emphasis on longer-term trends.

Some examples of methods commonly applied to produce forecasts of socio-economic outcomes are the Autoregressive Integrated Moving Average (ARIMA) model, the error-correction model, the Autoregressive Conditional Heteroscedastic (ARCH) model and the Box-Jenkins multi-variate time series analysis (see, for example, Box-Steffensmeier et al., 2014; McCleary et al., 1980).

The ARIMA approach is the simplest method applied to forecast welfare outcomes. It has also been used to measure links between key predicting factors and future outcomes. For example, Branas et al., (2015) use ARIMA techniques to forecast suicide rates conditional on adverse economic conditions while Chamlin (1988) applies the ARIMA methodology to explore temporal relationships between crimes and arrest rates. The Box-Jenkins and ARCH approaches, which can be thought of as extensions to the basic ARIMA method, mainly rely on the same principles.

Moreover, the error correction model arrives at forecasts of welfare outcomes considering their relationships with a set of covariates over time. The model assumes that there is a specific, long-run relationship between explanatory variables and the dependent variable – the co-integration relationship. Specifically, it is assumed that the dependent variable deviates from the long-run equilibrium as a result of

[3] More detailed discussion regarding best practices in forecasting techniques can be found in (Armstrong, 1985, 2001)

changes in a set of predicting factors. The model generates forecasts based on observed deviations of welfare outcomes during past periods.

Error correction models are used for both short and longer-term projections. However, identifying variables that are connected with co-integrative relationships is not straightforward – even minor misspecifications might lead to great losses in accuracy.

Scottish Council for Single Homeless model

The Scottish Council for Single Homeless designed an ad hoc model to forecast future levels of needs for temporary accommodation under a policy reform regarding homelessness duties (Scottish Council for Single Homeless, 2003).

Two new homelessness duties have been introduced under the Housing Scotland Act in 2001. Based on this Act, the right to temporary accommodation was given to all households that the City of Edinburgh County believed to be homeless before the assessment process. Moreover, households that were not assessed as being in priority need were granted access to temporary accommodation for a given period following the assessment.

The model calculated the expected number of non-priority households in temporary accommodation at the end of the current quarter based on the number of existing applications during the current and past quarters in Edinburgh. Model outputs were produced at high frequency (every quarter), using data drawn from the Scottish Homelessness Statistics database integrating homelessness records from LAs.[4]

The model generated simple trends-based forecasts, assuming that future temporary accommodation requirements depend on the past number of applications received by the Council. The model applies simple regression analysis to quantify links between levels of temporary accommodation offered by the Council, number of all applications and number of non-priority application prior to the reform. Administrative data prior to the reform are used to estimate parameters that reflect these links. Then, simple time series analysis is used to arrive at forecasts of non-priority households receiving temporary accommodation in every quarter following the policy reform.

In summary, the model presents an example of a simple time series model implemented to estimate future changes in outcomes of interest – in this case, temporary accommodation provision to non-priority households – as a result of policy changes.

[4] Data are drawn from the HL1 dataset – more information on the dataset is available here: https://www.gov.scot/Topics/Statistics/15257/1518

Box 4. Forecasting and policy models[1]

While both policy (simulation models) and forecasting models (time series models) are used to predict welfare outcomes – such as poverty, income inequality and homelessness – in future periods, these classes of models have different relative strengths and weakness depending on the particular purpose they seek to address.

Forecasting (and policy models place different emphasis on the intertemporal dynamics of the data (the way that past values influence current and future values). Forecasting models place the importance of observed intertemporal properties and dynamics above all else. Policy models, place importance on economic and statistical theory to drive the model specification.

Appraisal of planned policies requires models that are able to capture the underlying processes that generate the projections. This largely rules out simple time series models that do not focus on potential influences of economic and policy variables to generate projections of welfare outcomes.

On the other hand, simulation models are mainly centred around quantifying relationships between economic and policy variables and outcomes of interest that play out in the medium to longer-term. Therefore, they are not well-suited for short-term projections.

Such models aim to clearly capture causal relationships between a set of variables to arrive at projections that can inform policy development. However, their complex structure might result in relative losses in predictive power. By emphasising theory, policy models sometimes miss out on potentially useful forecasting information in the dynamics of the data.

Overall, forecasting models such as time series models are generally less suitable for assessing the effect of proposed policy changes but can provide more accurate short to medium-term forecasts. On the other hand, policy models – such as the simulation models discussed in this report – are more well-suited to the generation of medium to long-term projections of welfare outcomes under composite policy and economic scenarios.

The current review exercise aims to contribute to the feasibility study by outlining the benefits from complementing the required suite of models with simple forecasting techniques used to produce accurate projections of homelessness and rough sleeping in future periods.

Notes

[1] Economics-based simulation models are typical examples of policy models while time-series models fall in the category of forecasting models. Simulation and time series models share the characteristics of the broad classes of policy and forecasting models, respectively. Therefore, the discussion in this box highlights the relative strengths and weaknesses of simulation and time series models when applied to accommodate different purposes.

2.3 Machine-learning models

Machine learning methods have recently been used to produce projections of welfare outcomes, such as poverty. These methods identify patterns of connections between explanatory factors and outcomes of interest through iterative processes. High-order interactions between predictive variables and outcomes of interest that do not need to be specified in advance are explored.

Machine learning methods feature black boxes where it is not possible to map links from predicting factors to outcomes of interest. It is argued that they can be effectively applied when the goal is an accurate prediction of welfare outcomes rather than the estimation of separate effects of single causal factors.[5]

A debate has recently emerged as to whether machine learning methods are superior to statistical modelling in forecasting welfare outcomes. For example, Reed (2016) discusses the potential benefits from applying machine learning techniques to predict homelessness levels in the US context.[6] Machine learning is

[5] See here for a comparison between statistical models and machine-learning techniques by Frank E. Harrell: http://www.fharrell.com/post/stat-ml/

[6] For a discussion on the benefits from implementing machine learning models to homelessness predictions see here: https://slideplayer.com/slide/10942310

presented as a flexible and robust option for forecasting homelessness at the population level. It is argued that these methods can project homelessness levels by handling large inputs of administrative and survey data that can cover populations registered as homeless as well as homeless groups not necessarily captured in formal registries.

Random Forest (RF) model

The Random Forest model is a machine learning technique implemented in programming environments – such as Python – that has been recently applied to project welfare outcomes. It models the links between explanatory factors and outcomes of interest using a set of decision trees. Each tree comprises a set of decision nodes for the entire range of values of a subset of predicting factors included in the model. Random sets of data entries feed into each tree.

The data that feed into each tree are split in each node based on how well the driving factors predict the outcomes of interest. The model predicts welfare outcomes by repeating the steps involved in each tree several times and taking average values of the outcomes from all decision trees.

For example, an RF model applied to project homelessness could feature a very simple decision tree about poverty. In the first node of the tree, households in the random sample would be divided into two groups (those above and below the poverty line) based on their income. At the next decision nodes, these two groups of households would further split into smaller groups with respect to other homelessness predictors with the smallest prediction error.[7] Following this process, the subsample is distributed to smaller groups each containing a minimum of observations (for example, households in statutory homelessness and households that are not homeless).

The RF model has been used to calculate poverty measures in Mauritius (Thoplan, 2014) and identify poor households being eligible for welfare assistance in Indonesia (Otok and Seftiana, 2014). Further, Sohnesen and Stender (2016) have shown that while the RF method appeared to generate projections of poverty that were robust within one year, output accuracy started to deteriorate in the longer-term.

2.4 Homelessness risk models

Homelessness risk models assess the likelihood of homelessness occurring at the household and individual level.

[7] Various methods can be implemented to select the variables with the lowest prediction errors for the nodes – for example, Gini impurity loss functions. The Gini impurity factor can be thought of as a criterion to minimise the probability of misclassification. It measures the probability that a selected variable is labelled incorrectly (Venables and Ripley, 2002).

They calculate homelessness risks by quantifying the contribution of demographic, economic and personal predicting factors in homelessness occurring.

Various methods, such as proportional hazard regressions, logistic (or probabilistic) models and linear probability models, can be adopted to estimate the probability that a specific event – in this case, homeless – will take place conditional on a set of predictors. Simpler techniques that rely on correlation analysis to calculate homelessness risks can also be applied. The data for these models is mainly drawn from LAs registries – such as registries regarding social care, housing benefits, applications for temporary accommodation, etc.

Methods applied to calculate household and individual risks of homelessness, either comprising probability models or correlation analysis, are mostly used by prevention services that aim to identify households and individuals in priority need of homelessness prevention services. The outputs of these models help caseworkers target households and single people that are on the verge of homelessness.

Moreover, the models can feed into more complex policy models to measure the contribution of a set of driving factors to the probability of becoming homelessness for households and single people. For example, the logistic and linear probability regression models that were used to predict individual probabilities of rough sleeping at the first stage of the extended SRHMM model (Bramley, 2017) are examples of homelessness risks models.

Trailblazer – Bristol City Council homelessness prevention model

The idea behind the model is to 'predict' the households most at risk of homelessness to:

- target resources effectively to those who seem to have the highest risk of homelessness, and
- proactively offer advice and services early to try to prevent homelessness occurring.

To create the database that delivers *predictive analysis*, a number of data sharing agreements have been established between the Welfare Rights and Money Advice Service and the Council's:

- Housing Benefit Service
- Estate Management Service
- Rent Management Service
- Substance Misuse Team
- Children and Families Department
- Choice Based Lettings System
- Homelessness Prevention Team casework system

The combined household data provides rich information about individual characteristic, financial information, health status and information about family circumstances. A key step in the model is to use local benefits data to identify households in financial distress. There are four key markers, if:

- benefits will be reduced as a result of the benefit cap
- benefits will be reduced as a result of the under occupation charge
- there is a shortfall between a household's rent liability and the housing benefit or universal credit housing cost element entitlement
- a household's benefit entitlement is about to fall or be removed.

A linked step uses other more personal and sensitive data, bearing in mind the client groups identified by the Homelessness Reduction Act, as well as local factors, to prioritise households. For example, the following factors all have weightings which add to households' risk scores. Households:

- with at least one individual experiencing mental health problems
- where there are victims of domestic violence
- where at least one individual is a care leaver
- with a history of substance misuse.

This modelling process results in a list of households in order of the risk of homelessness; who key staff then to offer advice and homelessness prevention services.

Box 5. Newcastle City Council Trailblazer homelessness prevention model

In April 2018, the Active Inclusion Service[1] in Newcastle City Council began using predictive analytics developed with Policy in Practice to identify households who may be at risk of homelessness in the future. This work was funded through the MHCLG Homelessness Prevention Trailblazer programme, of which Newcastle was one of three early adopters. The households identified as being at risk of homelessness are approached by a new multidisciplinary team, also established using Trailblazer funding in October 2017 with the aim of testing new ways of preventing homelessness at an earlier stage.

The team incorporates a housing specialist on secondment from the city's arm's length management organisation, a debt and budgeting specialist and welfare rights specialist out-posted from Newcastle City Council, and an employment specialist on loan from Jobcentre Plus. The team had previously tested their approach by targeting advice and support to residents affected by the "bedroom tax" and the benefit cap.

The predictive analytics are based on a range of local data including Housing Benefit, council rent, Council Tax arrears and Reduction Scheme, and Discretionary Housing Payments. This data is analysed against estimated changes in inflation on household goods and services and expected welfare reforms to predict which residents may be at risk of homelessness in the future. Over 33,000 households are included in the dataset, which the team access through an interactive dashboard that allows them to easily segment the data in various ways.

In July 2018, the team established their first criteria of residents to target using these predictive analytics. As end of an assured shorthold tenancy is the most prominent cause of homelessness nationally they decided to approach residents in private tenancies. Residents were further segmented according to those who are categorised as being 'at risk of financial crisis' by Policy in Practice's analytics, have Council Tax arrears – which is used as a proxy for other forms of debt as rent arrears are not available for private tenants – and are not in receipt of a Discretionary Housing Payment.

After identifying residents through the dashboard, the team screen a variety of available databases to determine the most appropriate approach for each household before offering advice and support across their individual specialisms. As the team only began approaching these households in July, data is not yet available to evaluate the effectiveness of this approach. Such an evaluation will be available in the multidisciplinary team's next quarterly report, due to be published in November 2018.

The team will continue to use the dashboard to find new ways of segmenting residents who may be at risk of homelessness until March 2019, when the Homelessness Prevention Trailblazer pilot ends.

Notes

[1] *For more information the Active about Inclusion Service, see here* https://www.newcastle.gov.uk/sites/default/files/wwwfileroot/housing/housing-advice-and-homelessness/active_inclusion_newcastle_-_briefing_note_2018-19_1.pdf

NY screening tools

An empirical risk model has been recently developed to predict probabilities of ending up in New York City shelters among families (Shinn et al., 2013) and single people (Greer et al., 2016). Particularly, a Cox model was applied to identify households and single people that might face the risk of homelessness conditional on a set of causal factors.

The Cox regression is a proportional hazard model that explores the contribution of a set of triggering factors in the probability of the occurrence of a specific event. The model's outputs are hazard ratios that measure the impact of each predictive factor on the possibility that the event will take place within a specific period. In other words, hazard ratios are the amount by which the rate of the event occurrence is multiplied for people having a particular characteristic.

Screening tools designed to be used by workers in HomeBase homelessness prevention services in New York City (NYC)[8] (see, for example, Greer et al., 2016) were developed based on this model. Using backwards elimination of weak (nonsignificant) predictors, screening tools include fewer variables and thus, require less data. For example, the household screening instrument identifies households in need of prevention services based on 15 predictive factors while the instrument for single people only relies on seven predictors. Evaluation of the ability of the screening tools to assess homelessness risks among shelter applicants reveals that they are almost as efficient as the full models in identifying target families and individuals in need of prevention services (see, for example, Shinn et al., 2013).

The screening tools appeared to be more efficient in identifying households and single people in priority need compared to the judgements of staff working in prevention services (Greer et al., 2016; Shinn et al., 2013). Particularly, using the model resulted in a 26% increase in correct targeting of families entering homelessness shelters and a reduction in failure to provide support to households and single people at risk of homelessness by two thirds (Shinn et al., 2013).

The screening tools draw on data from city records and self-reports obtained through interviews from people working at prevention services (Shinn et al., 2013) to calculate risks of homelessness among people who apply for shelter in NYC prevention services. One basic limitation of this approach is that it can only assess homelessness risks among people that reach out to prevention services for assistance.

[8] A description of the HomeBase programmes can be found here: https://www1.nyc.gov/site/dhs/prevention/homebase.page

2.5 Non-standard sampling models

Non-standard sampling methodologies were primarily developed to estimate the size of populations that are elusive and hard to capture in the areas of biological sciences (for example, to measure animal populations).

They have also been used to measure human populations that are not straightforward to count – e.g. populations of street-working prostitutes in Scotland (Leyland et al., 1993). They rely on probability methods that count populations based on observable fragments (Sudman et al., 1988).

Models in this class have been applied to address the following methodological issues that arise in the measurement of homeless populations:

i. there is not a single definition of different types of homelessness hindering a straightforward classification of homeless households and single people – e.g. statutory homeless, people who sleep rough, couch surfers, etc. Moreover, transitions in and out of homelessness are difficult to track since homeless populations are mobile and continually changing. Therefore, accurate homelessness definitions should integrate restrictions on time and locations that cannot avoid being arbitrary to some extent.

ii. homeless populations are likely to be hidden in the sense they are often not *visible* in official registries – it is not always straightforward to locate and get in touch with homeless households and single people. It is not uncommon that the sizes of particular homeless subgroups, such as drug addicts, people with mental illness or multi-family (concealed) households, are underestimated as they are likely not to turn to LAs for assistance or participate in any survey.

The lack of a definite sampling frame to measure homeless population sizes requires sampling techniques designed to capture elusive populations. For example, Williams (2010) suggests that non-standard sampling methods should replace traditional headcount methods of measuring homeless groups to correct for measurement inconsistencies. For example, in the case of people who sleep rough in London simple enumeration methods are likely to result simultaneously in both *overcount* and *undercount* of rough sleeping populations in a designated area – for example, counting rough sleeping groups more than once or missing them because they move around or because enumerators make identification errors.

Capture-recapture method

The basic concept of the capture-recapture method is simple: obtaining independent observations of the same target populations more than once. Observations can be recorded in two points in time using the same data source. Alternatively, two different sources of data on the same population can be used in such a way that all units of the population have equal chances of being selected. The method produces estimates of the total population by combining the number of these independent observations in two points in time – or drawn by two different sources – and the overlap between them (Sudman et al., 1988).

The validity of the method depends on three core assumptions: i. each member of the population has an equal probability of being selected while this probability changes over time (and data sources), ii. the population does not change in composition and size over time (or data sources) and, iii. the samples are independent, meaning that the observation of one unit in a given time (or data source) should not depend on its selection at earlier times (or other data sources) in case there are only two samples.

Given the changing and mobile nature of homeless populations, these assumptions are easy to be violated. For example, log-linear models are used to control for interdependences between capture and recapture homeless samples aiming to construct the simplest model (Williams, 2010). Based on this model, the size of the population that is not observed in any of the samples (hidden population) can be calculated, resulting in estimations of the total size of such an elusive population as the population of homeless. Given the changing and mobile nature of homeless populations, these assumptions are easy to be violated. For example, log-linear models are used to control for interdependences between capture and recapture homeless samples aiming to construct the simplest model (Williams, 2010). Based on this model, the size of the population that is not observed in any of the samples (hidden population) can be calculated, resulting in estimations of the total size of such an elusive population as the population of homeless.

The capture-recapture method has been applied to measure the size of homeless populations in the UK (Williams, 2010), the US (Cowan et al., 1986) and Hungary (Dávid and Snijders, 2002). Data for the models are either drawn from existing censuses, surveys and registries (as in Dávid and Snijders, 2002) or collected from enumeration teams at different points in time (as in Williams, 2010). In summary, it is shown that applications of the capture-recapture method in these three countries resulted in reliable and valid estimates of homeless populations despite the shortcomings of the method.

3. Assessment strategy

3.1 Approach

Our plan for the assessment is to consider how well each model performs against a set of criteria. The criteria have been selected to be able to guide our discussion about factors that are important in selecting an appropriate methodology for homelessness modelling.

A comprehensive assessment is undertaken for the main methodology used in the UK context to project homelessness in future periods and appraise potential impacts of composite policy scenarios – namely, the model developed by Bramley et al. (2010) and its extended versions. The remaining classes of models are categorised based on their objective and outputs. Models' advantages and disadvantages under each criterion are discussed. We will not attempt to assign weights to any particular criterion leaving that decision to the judgement of the analysts building the model(s) in the future.

Consequently, the assessment does not result in a ranking of models/classes of models. In practice, the models have been built with different purposes and there will be circumstances in which everyone is useful. Instead of numerical scores and rankings, the discussion attempts to shed light to the relative strengths and weaknesses of each model.

3.2 Assessment criteria

Criteria		Description
Resources	Expertise and effort	This criterion relates to the complexity and the specialist knowledge required to use the model. We will also consider the resources and costs of maintaining, updating and adapting the model.
	Data requirements	This criterion refers to how data hungry the model is. It will also consider whether the model can be applied with accessible data sources or requires getting access to/collecting additional data. Where possible this will also be forward-looking by considering potential data additions e.g. case-level data for homeless populations.
Communication	Interpretability	Interpretability refers to the extent a model's results allows us to tell a clear and insightful 'story'. This criterion relates to how intuitive and easily presentable model outputs are – e.g. how simple it is to identify the effects of driving factors, such as changes in housing prices, policy variables etc. – and different policy scenarios in homelessness projections.
	Transparency	Transparency is about understanding how homelessness estimates/projections are calculated. When a model builds on clear and transparent assumptions and the required calculations can be easily comprehended and/or replicated, its results are perceived as more reliable.
Application	Policy relevance	This criterion relates to the extent to which the model is structured to allow for policy variables (e.g. housing allowance, unemployment benefits) that could have an impact on homelessness. We will also consider the breadth of policy relevance (e.g. does the model capture policies relevant to housing market factors? Can broader areas of policy be captured, such as criminal justice system policy and/or immigration policy changes?
	Flexibility	How straightforward would it be to adapt the model to deal with additional policy scenarios, different geographic levels, etc.?
	Accuracy	The accuracy of any forecast outputs will be discussed, e.g. show how accurate model outputs have turned out to be previously.
	Level of analysis	This criterion will describe the different levels of analysis of each model. For example, whether the model generates homelessness projections for the whole population or can be used to predict individuals/households at risk of homelessness. This assessment will include a discussion about the level of analysis for different types of homelessness as well as the geographic area of analysis – e.g. can the model be applied to a UK/regional or lower geographic level?

4. Model assessment

4.1 Complex economic-based simulation models applied in the UK context

The current assessment is centred around the main policy model that is applied to produce medium to long-term projections of poverty in the UK context – the economics-based simulation model developed by Bramley and colleagues. Our main focus is to identify relative strengths and weaknesses of the model developed to project homelessness under composite policy and economic scenarios. The assessment is based on publicly available documentation of various model versions (Bramley, 2017; Bramley et al., 2016, 2010; Bramley and Watkins, 2016), a discussion with Professor Bramley and some unpublished documentation. We complement the assessment by discussing specific characteristics of the broader class of models where useful.

Resources

Level of effort and expertise

The underlying models that are used to estimate elasticities between key determinants and housing needs as well as the macro-simulation exercise require advanced levels of econometric and statistical expertise.

The extensions of the initial EHN model developed by Bramley et al. (2010) have different levels of complexity. For example, the SRHMM is a more complex version of the EHN model, expanded to quantify the influence of a wider range of determinants – for instance, internal migration – and account for the spatial interdependence of housing and labour markets across the UK. Mapping multidimensional paths to homelessness requires technically complex models demanding analysts with a good understanding of econometric modelling.

Additionally, the version of SRHMM used to produce poverty projections includes a micro-simulation model that generates snapshots of poverty trends at the household level. While this feature has not been used thus far to project homelessness, it would be a valuable addition allowing higher granularity of outputs (for example, homelessness levels by age group, gender, and vulnerable groups such as people with drug and alcohol dependence).

The SHRMM micro-simulation model is conducted in a statistical software package that requires analysts with some level of statistical expertise. The model is built using long pieces of code to map complex chains of impacts. For example, approximately 4,000 lines of codes are used to estimate the impact of various policy scenarios on poverty measures at the household level. The process of understanding, running and revising the code is likely to be time-consuming.

For example, micro-economic simulations integrated in the SRHMM to appraise the impact of policy scenarios to future poverty levels are conducted in statistical software packages – such as SPSS – that require some level of statistical expertise. Models are built using long pieces of code to map complex chains of impacts – for example, app. 4,000 lines of code estimate the impact of various policy scenarios on poverty measures. The process of understanding, running, and adapting the code is likely to be time-consuming.

In summary, updating the distinct components of the model developed by Bramley and colleagues (using more recent data, adding new pathways to needs incidence and testing alternative hypotheses between key determinants and outcomes of interest) is a task that would require a team of experts.

Data requirements

Homelessness levels are projected using composite sets of existing data drawn from various sources including administrative registries (such as LAs homelessness acceptances), longitudinal household surveys (e.g. Understanding Society) and national statistics (such as CORE).

Depending on the specific requirements of estimating each outcome of interest, e.g. different types of non-statutory homelessness, alternative datasets can be used to produce ranges of forecasts. Certain datasets, especially administrative datasets, often do not fully capture the extent of an issue which is true for rough sleeping since not all the people who rough sleep will register themselves with a LA. Therefore, using LA administrative data will underestimate the true rough sleeping populations. The version of SRHMM expanded to UK countries outside England was estimated using registry and survey data to explore rough sleeping trends across regions (Bramley, 2017). As expected, using data from LA reporting systems – such as CHAIN – produced conservative forecasts. On the other hand, using data drawn from longitudinal surveys (e.g. British Cohort Study (BCS) and the Poverty and Social Exclusion survey (PSE)) resulted in higher predicted levels of rough sleeping in the long-term.

While the simulation models are able to produce homelessness forecasts based on accessible data, they can also be fitted to new data or existing alternative sources of data, in order to produce reliable projections of different types of homelessness.

It appears that the set of data assembled to project homelessness levels conditional on various demographic, policy and economic variables is comprehensive. However, the reliability of the model depends on testing whether its outputs remain consistent when more detailed data are used – for example, detailed data on the individual and household level, new collections of data on homeless populations.

As discussed before, the task of changing the level of data in the model would demand fundamental changes in the underlying econometric models that would potentially require experts to redesign of the core functions of the simulation framework.

Communication

Interpretability

The outputs of the main model used in the UK context are presented in spreadsheet environments, where users can change key policy and economic parameters and observe their direct effects on homelessness projections. The model front-end enables analysts to test the sensitivity of homelessness outcomes to different assumptions regarding economic growth, labour market trends and policy changes.

However, it appears that the model was not designed to be used by non-experts. Therefore, mapping the links between changes in the policy and economic environment to homelessness trends is unlikely to be a straightforward task. The macro-simulation is a complex process that would require analysts with a detailed understanding of the model design to interpret the quantified paths to homelessness.[9]

Transparency

The paths leading to unmet housing needs and homelessness that users can track are predetermined by the core functions estimated in the first stage of the model. The underlying models quantify complex relationships between various economic, demographic and policy factors that are specific to the UK context. Relying on numerous assumptions and theoretical considerations, their implementation and replication does not appear to be a straightforward task. Particularly, it seems unlikely that the model could be replicated at least based on publicly available documentation.

The model also allows for analysts to test the sensitivity of outputs to different assumptions and model properties. Sensitivity analysis would be performed by running the core functions of the model under alternative designs and assumptions. Therefore, such analysis requires high levels of statistical expertise and familiarity with the model.

In summary, the all-encompassing simulation framework imposes a trade-off between mapping complex mechanisms and transparency.

---

[9] Not having access to any version of the discussed model naturally limits our assessment of the interpretability of the model's outputs. We mainly depend on the authors' discussion of the qualities of the characteristics of the model.

Analysts can observe the direct impacts of demographic, policy and economic changes on homelessness levels that are quantified as elasticities estimated using behavioural models.[10]

Application

Policy relevance

The initial EHN model and its extensions are policy models that rely on various assumptions and theoretical concepts to quantify links from demographic characteristics, economic factors and the policy environment to outcomes of interest. In the past, the model has been adapted and extended to consider different policies and increased disaggregation, e.g. including different types of homelessness. The model identifies direct and indirect impacts of a broad range of policies on homelessness across UK regions.

For example, the model can be used to predict homelessness levels in the medium and longer-term under different levels of social and private housing supply (see, for example, Bramley et al., 2010 and Bramley and Watkins, 2016). Increased supply is expected to result in reduced prices in the housing market, in turn leading to improved affordability (lower housing expenses-to-income ratios). In-migration flows and household formation are also expected to increase as a result of increased affordability. Moreover, higher social supply will lead to more social lettings, directly impacting on household growth. This higher supply of social housing will lead to increased household formation e.g. shared households will decrease, resulting in reduced levels of unmet needs such as homelessness. In summary, the impact of changes in housing supply on unmet housing demands works through the interplay between changes in housing market conditions, affordability and household formation.

Moreover, the models consider the impacts of benefit take up and eligibility as well as tax credits on projections of homelessness. For example, welfare benefits reforms such as the introduction of the Universal Credit (UC) framework can be appraised with respect to their impact on homelessness. However, it appears that there is no depth to modelling behavioural responses to changes in policies regarding taxes and benefits. This limitation can be addressed by revising the model core functions to quantify pathways linking welfare and tax policies to homelessness.

Flexibility

The structure of the model allows for capturing the impacts of additional policy variables using various functional forms to estimate key elasticities. However, the outputs of

[10] Coefficients from estimating the core functions as well as the impacts of changing policies are reported in publicly available documentation. See, for example, Bramley and Watkins (2016).

the macro-simulation depend on the assumptions and methodologies implemented at the first stage of the models. Therefore, revising and extending the model is not straightforward as these would require a revision of the core functions of the model. As discussed before, the model's level of technical complexity requires deep understanding of statistical concepts and methods.

Accuracy

We cannot come up with a definitive assessment of output accuracy as the models have been used on behalf of Crisis to project homelessness levels for years 2021, 2026, 2031, 2036 and 2041 (Bramley, 2017).

The EHN model, which is an earlier and rougher version of the extended SRHMM currently used for homelessness projections, appears to produce more accurate estimates in the longer-term. The model predicted that temporary accommodation cases – used to measure statutory homelessness – would amount to approximately 63,000 in 2009 and rise to the level of 80,000 in 2018 in England (Bramley et al., 2010). The estimates for 2018 are not different to the actual numbers of temporary accommodation cases as reported by the Ministry of Housing, Communities and Local Government,[11] suggesting that the model can potentially result in reliable estimates for long time intervals. On the contrary, the model predicted relatively stable annual increases in the homelessness levels each year of the period between 2009 and 2019, failing to capture actual fluctuations in homelessness levels.

However, it should be noted that the accuracy of the predictions of the policy models can only be assessed on a theoretical level. The idea is not so much to attempt to arrive at precise estimates of homelessness levels in the long-term – there are usually too many economic and policy variables whose changes cannot be easily predicted. The purpose of policy models is mainly to provide a solid theoretical framework that allows us to predict trends in the outcomes of interest under planned changes in policy and economic conditions.

Level of analysis

Policy models allow for high levels of granularity of outputs. The models developed by Bramley and colleagues can produce projections about homelessness across different age and tenure groups. For example, the model produces homelessness projections for groups of people under and over 40 years of age. The model is also used to predict future levels of different types of non-statutory homelessness, e.g. people who sleep rough, sofa surfers, etc.

While outputs for particular population subgroups with similar characteristics

[11] Detailed data on temporary accommodation levels can be found in the temporary accommodation live tables published by MHCLG that are available here: https://www.gov.uk/government/statistical-data-sets/live-tables-on-homelessness

other than age are not reported, the structure of the model potentially allows for further breaking down of projections by specific characteristics. Particularly, using the micro-simulation feature to project various types of homelessness levels would enable further breakdown of the results across various subgroups – such as women, immigrants, etc.

Finally, outputs are broken down across localities (countries, regions and cities) in the UK. The model is shown to be quite robust in producing medium to long-run estimates of aggregated homelessness levels in England. However, its application to other parts of the UK falls behind. This shortcoming is mainly driven by the lack of necessary data for other UK countries outside England – with the exception of Scotland where high-quality, detailed data on homelessness is available. Missing variables are being imputed using available information from similar English areas.

Box 6. Assessment of simpler policy models: the World Bank approach to projections of welfare outcomes

The models developed by World Bank analysts (see, for example, Elbers et al., 2002; Douidich et al., 2015) are mainly centred around projections of welfare outcomes such as poverty and inequality. These models present typical examples of policy models (lacking the level of complexity of the approach predominantly adopted in the UK context).

While they are not specifically designed to measure homelessness, we discuss their characteristics against our group of assessment criteria. The purpose of this assessment is to explore potential benefits from implementing simpler methods to project homelessness levels under a set of assumptions. For example, homelessness projection and policy appraisal tools used by LAs could be designed based on such simple policy models.

Resources

In their current form, World Bank models are easy to implement by analysts with standard levels of statistical understanding. Moreover, they do not require excessive amounts of data.

Mainly two types of datasets (detailed survey data and frequent large-scale data) are used in the two stages of the model to quantify relationships between explanatory variables and outcomes of interest and arrive at projections conditional on a set of covariates.

If the models were to be adapted to homelessness projections they would become more complicated. For example, the distribution of more population parameters should be estimated in the first stage to arrive at reliable aggregate estimates of homelessness.

Communication

The World Bank models are based on clear and transparent assumptions, allowing for easy replication of outputs. Based on these clear assumptions, the models are able to map outcomes as the products of changes in different factors, making it easy for non-expert audiences to observe links between explanatory variables and outcomes of interest.

However, it might not be straightforward to isolate the effect of every single predictor included in the model. The main idea behind the model is to estimate the distribution of key determinants of the outcomes of interest across the entire population conditional on a number of covariates. Then, the model arrives to projections of welfare outcomes using these estimated distributions. Therefore, it is likely that the model does not allow for direct observation of the mechanisms driving the impacts of a single factor on future trends in welfare outcomes.

Application

The models focus on producing reliable estimates of future values of welfare outcomes at various levels of frequency and aggregation rather than appraise changes in policy.

In theory, they can be expanded to include a set of key indicators reflecting prevailing or expected conditions in the economic and policy environment. The models are able to link changes in economic and policy variables and outcomes of interest.

Their adaptation to projections of homelessness remains an open question – it is likely that the models would have to become more complex in order to consider the entire range of homelessness determinants that are specific to the UK contexts and their interdependence.

However, they could potentially be adapted to project homelessness across UK regions where data is not available using simpler assumptions. For example, they could be used to measure current and future levels of homelessness based on a number of observed confounding factors in UK countries outside England where detailed homelessness data are not available. Moreover, they can be simpler alternatives to assessing direct impacts of policy changes at the local level under a simple set of assumptions.

4.2 Models for homelessness measurement

We group models that are developed to estimate the size of homeless populations at present and future periods under this label. The following models are included in this category:

- Time series models generating forecasts based on past temporal trends,
- Machine-learning models used to forecast welfare outcomes based on a broad range of determinants, and
- Non-standard sampling models used to improve measurement of the size of elusive populations.

The assessment is based on our own analysis of the publicly available documentation as well as the discussion found in Box-Steffensmeier et al. (2014) and Branas et al. (2015) for time series models, Reed (2016) and Sohnesen and Stender (2016) for machine-learning methods and Sudman et al. (1988) and Williams (2010) for non-standard sampling models (particularly, the capture-recapture model).

Resources: expertise & effort and data requirements

Time series models are the least demanding models in terms of requirements in data sources and user expertise. They are simple models that can be implemented by analysts with a basic understanding of statistical concepts. They use time series data to forecast future values of a dependent variable based on its past values. Even more complicated versions of time series models (for example, including more than one explanatory variable) require standard levels of expertise and not extensive amounts of data inputs.

The machine learning models are an alternative approach to generating short-term forecasts of welfare outcomes. The design and implementation of machine learning models requires analysts that have an understanding of concepts related to programming.

Moreover, machine learning models are able to produce reliable projections of the outcomes of interest using a quite wide set of data. They can handle hundreds of data entries as inputs, allowing the analysts to use all available information on homelessness drivers to arrive at projections in a relatively fast and reliable way. However, the reliability of the outputs of these models depends on the quality and level of detail of data that will feed in the models. For example, in the English context, homelessness indicators are mainly measured using household= surveys, where homeless people are usually under-represented. A dataset that summarises information from local homelessness registries should feed in the model to maximise forecasting accuracy.

Non-standard sampling methods, such as the capture-recapture technique, can be applied either to existing data or as tools for designing new data collection to measure homeless populations. They are based on simple statistical concepts and are thus, quite straightforward to implement.

Communication: interpretability and transparency

Time series models are quite straightforward to interpret – they produce predictions of key outcomes in future periods based on past trends. Given that the accuracy of the models depends heavily on their simplicity, they are based on simple assumptions that are easy to understand and replicate.

Machine-learning models do not explicitly model the mechanisms that generate outcomes of interest. Instead of relying on predetermined parameters that reflect hypothesised links between outcomes of interest and key covariates, the models identify patterns in outcomes of interest through iterative processes. Such processes can be compared to a black-box - not allowing the user to follow clear paths from changes in key covariates to projected outcomes of interest.

Non-standard sampling techniques are also easy to replicate and understand. The capture-recapture model, which is the main non-standard sampling method applied to measure current sizes of homeless populations, is based on simple concepts and models. It is quite straightforward to understand the process that generates the estimation of population sizes, consisting of multiple sampling (or data drawn from different sources/collected at different points in time) and simple log-linear functions to model the dependence between multiple samples.

Application: policy relevance, flexibility, accuracy and level of analysis

The three types of models in this category are primarily developed to measure welfare outcomes in the short to medium-term and measure the size of homeless populations concurrently. While they are based on different assumptions and arrive at outputs following different processes, they share a common property: output accuracy (see, for example, Box-Steffensmeier et al., 2014; Reed, 2016; Sohnesen and Stender, 2016 and Williams, 2010)

Time series models and machine-learning techniques are the methods that result in the most accurate predictions of outcomes in the short to medium-term. Moreover, non-standard sampling approaches, such as the capture-recapture method, are shown to produce reliable estimates of current sizes of elusive populations that are difficult to measure.

The models are not designed to quantify causal links between the outcomes of interest and related factors. Their objective is mainly to produce out-of-sample forecasts or estimate current population sizes using simple methodologies. For example, multivariate time series models can control for temporal connections between the outcomes of interest and key determinants. However, they are not developed to allow for mapping composite effects of potential changes in the economic and policy environment on outcomes of interest.

Finally, the models can be adapted to generate forecasts and population counts for different types of homelessness and at lower levels of geographical aggregation. Depending on the availability and suitability of local data sources, these models can be fitted to regional and sub-regional data to produce estimates across both UK countries and localities within the countries (for example, big cities). However, given that they do not examine homelessness as the outcome of the interplay between various determinants, they are less suitable for breaking down estimates of homelessness future value for different subgroups – for instance, with respect to age, gender and tenure status.

4.3 Homelessness risk models

In this section, we discuss the appropriateness of models that calculate homelessness risks at the individual and household level. These models are mainly used by prevention services to identify and provide support to people that are likely to face homelessness. We assess the following models:

- Proportional hazard models adopted by prevention services in the US (particularly, HomeBase prevention services in New York City), and
- Predictive analytics (Trailblazer) method used by Bristol City Council for homelessness prevention.

Our assessment of homelessness risk models is based on i. the discussion about the screening tool used by HomeBase prevention services in New York City by Shinn et al.(2013) and ii. documentation about Trailblazer predictive analytics methods made available to us by Bristol City Council.

Resources

Expertise and effort

Both the Bristol City Council Trailblazer model and the model used by HomeBase homelessness prevention services in New York City are developed to allow preventions services to identify households and single people that are in priority need of advice and other homelessness prevention services. The models are relatively simple.

Both models contribute to the reduction of costs in homelessness services by making well-educated decisions regarding prioritisation of interventions with households that are at greater risk of homelessness. Preliminary results indicate that prevention work based on predictive analysis is cost-effective.

Data requirements

Both models use data drawn from LAs and surveys, that are accessible by LAs, to calculate risks of homelessness conditional on a set of important predictors. However, it is not immediately straightforward to use different sources of LA data since they need to be linked for a new purpose. Specifically, to create the database that delivers predictive analysis, it was necessary to establish that legal gateways existed to permit the use of data in this new way and a number of data sharing agreements had to be established between different teams in Bristol, e.g. Welfare Rights and Money Advice Service and the Bristol Council's housing benefit service and social care teams. As a result, the Trailblazer datasets comprises a rich data source.

The New York City model is limited in the sense that it calculates homelessness risks only for people that have contacted the LAs to receive homelessness prevention advice and services – this restriction potentially underestimates the real number of households and single people that are in priority need. This restriction imposed by the data potentially limits the scope of the interventions implemented by prevention services. On the other hand, the Trailblazer model identifies families and single people that face circumstances that might result in homelessness – it is not a prerequisite that deprived households and individuals should contact prevention services first.

Communication

Interpretability

Homelessness risk methods serve as screening tools and are easy to use by workers not required to have any level of expertise in statistical analysis. For example, the New York City model was the product of a more complex proportional hazard model reduced down to the most substantial triggering factors. The Bristol Trailblazer model features a simple and easily operated dashboard that allows caseworkers to observe households and single people with high estimated risks of homelessness as well as their scores in financial risks related to homelessness, such as rent arrears, benefit caps, etc.

Evidently, the calculation of homelessness risks can be conducted using a variety of ways; for example, as discussed previously, Cox regression models are used to calculate hazard ratios for the screening tool used by Homebase prevention services, while the Bristol Trailblazer model calculates total homelessness risk as the sum of individual factors with respect to specific triggers. Developing an intuitive understanding of how the model works naturally depends on the complexity of the adopted method. However, the predictive models that are currently in use in Bristol and New York city allow the users to interpret how different elements affect homelessness probabilities.

Transparency

Models in this class are based on clear assumptions regarding the links from specific triggers to homelessness. Based on available empirical evidence of causal factors of transitions in homelessness, screening tools use a set of predetermined triggers to measure the likelihood of homelessness occurring and prioritise the delivery of prevention services. The assumptions regarding the contribution of each factor to the occurrence of homelessness are transparent, allowing users to identify pathways to homelessness among deprived households and single people.

Application

Policy relevance

The models in this category take into consideration the impacts of policy measures on homelessness occurrence on the household or the individual level; for example, loss in benefits or take-up of benefits such as housing allowance and unemployment benefits.

These models are not designed to capture the dynamic effects of changes in policy and economic variables. The reliability of the model outputs depends on the circumstances prevailing in the period when the data were collected. Changing economic environment and policy shifts might result in accuracy losses in the measurement of homelessness probabilities among deprived households and single people. Therefore, the models should be frequently revised and tested to make sure that they are in line with the existing policy and economic conditions.

Some methodologies in this class can quantify behavioural responses to policy and economic changes. For example, the tool used by HomeBase prevention services in New York is based on a Cox regression model that can potentially capture changes in homelessness risks as a result of changes in economic and policy related factors.

Flexibility

Homelessness risks models used by prevention services can be expanded to either include broader sets of predictive factors or account for changes in policies that can influence homelessness. Moreover, the models can be revised to calculate probabilities of alternative definitions of non-statutory homelessness (for example, sofa surfers and rough sleeping groups).

Risk models rely on simple statistical concepts – therefore, it is quite straightforward to revise or adapt them to calculate risks of different homelessness types, consider changes in policies or include additional predictive factors.

Accuracy

Available evidence for the screening tool used by HomeBase services in New York City suggests that the model has increased correct targeting of prevention services to households and individuals on the verge of homelessness. Specifically, it is shown that the implementation of the screening tools by caseworkers would have increased correct targeting of families entering shelter by 26% and reduced misses by almost two thirds (Shinn et al., 2013). While similar evidence regarding the accuracy of estimated homelessness risk factors is not available for the Trailblazer model, the implementation of predictive analytics is expected to contribute to improving the efficiency of prevention services, suggesting that it can identify households and single people at risk of homelessness reliably.[12]

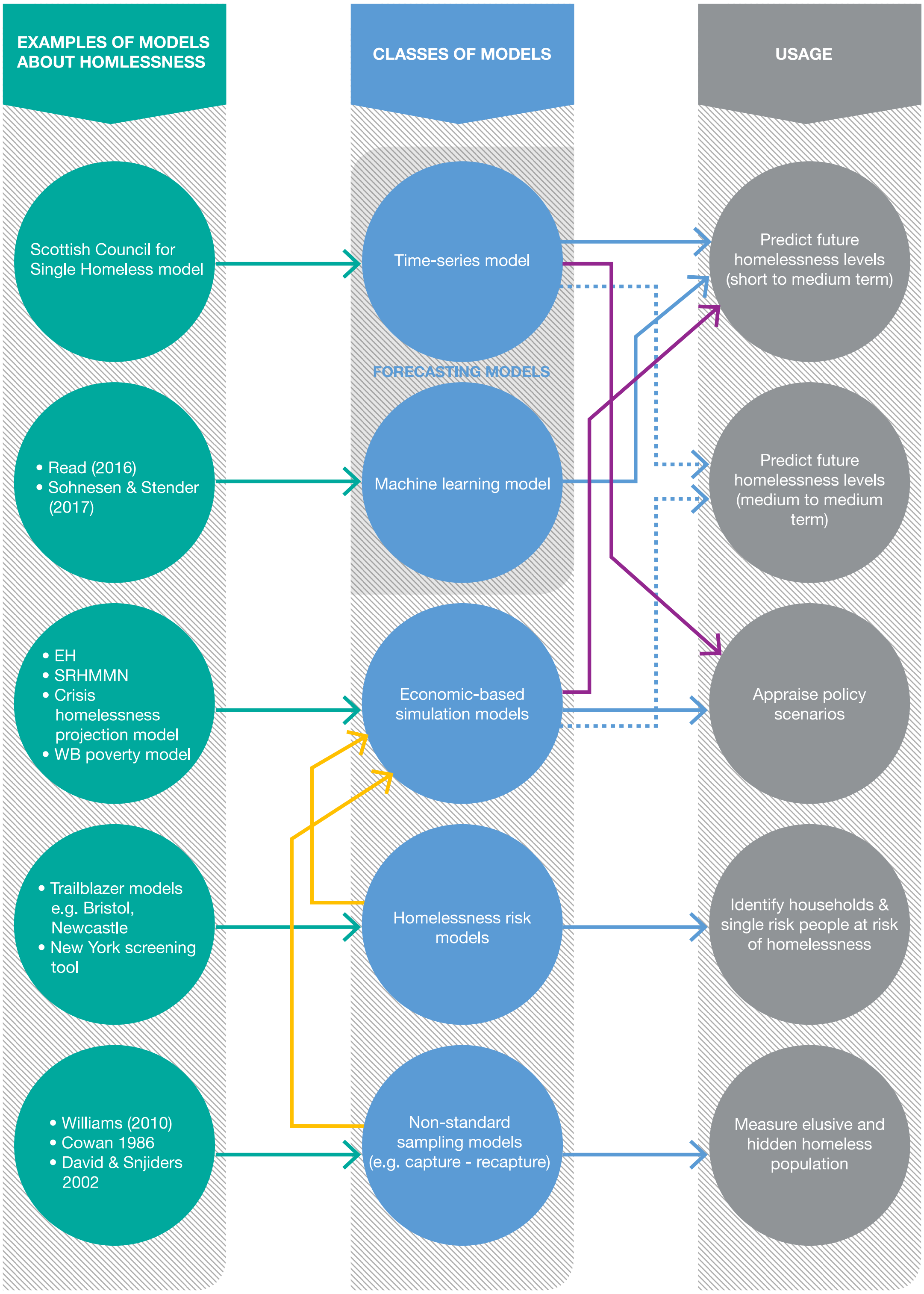
[12] For more information regarding the entire Trailblazer programme in Bristol, see here: https://democracy.bristol.gov.uk/documents/s13065/Funding to Support Homelessness Prevention and Reduction.pdf

Level of analysis

Methods falling in this class are already used by a number of LAs to identify households and individuals that are in priority need of prevention services. The methods are expected to improve the efficiency of prevention services in correct targeting of homeless families and single people. Therefore, their use can potentially be introduced to more areas in the UK.

Moreover, these methods can be applied to distinguish between risks of ending up in various definitions of homelessness – e.g. statutory homeless, rough sleeping groups, etc. The models can be adapted to estimate risk factors of various homelessness types conditional on sets of factors that are identified in the literature as triggers.

Review of Models



Orange arrows: can feed in simulation models | **Blue arrows:** Key output | **Purpple arrows:** Can be used, but not advised | **Dashed line:** can be used, with limitations

Review of Models

Key Characteristics

Time-series models

- Main objective to maximise accuracy (minimise forecasting error)
- Loose/no theoretical framework
- Tend to be relatively simple models
- Future trends estimated using past values of variable of interest or other related factors
- Can be used for policy analysis - but shouldn't

Machine learning models

- Causal links not determined in advance
- Loose/no theoretical framework
- Can handle large amounts of inputs (e.g. administrative data)
- Estimation through iterative process (black box)
- Can't be used for policy analysis

Simulation models

- Main objective to produce projections given a specific scenario
  Strong theoretical framework
- Can be simple or (very) complex
- Can be used for forecasting, but
- shouldn't (especially short to medium term)

Risk models

- Calculate risk of homelessness conditional on a set of predictors
- Used for prevention purposes, or as input to simulation model
- Easy to implement and interpret (can be further simplified to be used by prevention services staff)

Sampling models

- Can be used to guide data collection or with new data
- Extrapolate parts of the population that can be observed and measured
- Based on simple statistical concepts
- Easy to implement
  Shown to produce reliable counts of elusive populations

5. Conclusion

The objective of this review is to summarise and assess the characteristics of the classes of models used to inform policy aiming to tackle homelessness and provide support to homeless people. It is likely that many ad hoc models used for policy purposes regarding homelessness are not available in the public domain. However, we were able to identify a set of key models that are used by different actors, for example LAs and government departments, in the UK and abroad.

In summary, we have identified the following classes of models:

- economics-based simulation models,
- time series models,
- machine learning models,
- homelessness risk models, and
- non-standard sampling models.

These models are designed to accommodate the following set of distinct purposes:

- generate accurate forecasts of homelessness levels,
- evaluate potential homelessness impacts of alternative policy scenarios,
- identify households and single people at risk of homelessness, and
- measure hidden and elusive homeless populations concurrently.

The purpose of our assessment is to discuss the appropriateness of each class of models for different objectives rather than produce a "quality ranking".

Having reviewed all the models, we found that there is merit in applying different models for different purposes. For example, both economics-based simulation models and time series models can be used to produce projections of homelessness types. However, simulation models are more well-suited for appraising the impact of planned policies and changes in the economic environment in the medium to longer-term rather than producing accurate short-term projections. While shifts in exogenous variables such as labour market shocks that cannot be predicted can result in inaccurate projections, this does not affect current policy decisions informed by model outputs. Moreover, simulation models can be expanded to accommodate the appraisal of composite scenarios integrating changes in broad policy areas, including housing and unemployment benefits.

On the other hand, time series models are simple techniques, geared towards generating accurate predictions of future trends in the outcomes of interest. While they can model relationships between predictive factors and outcomes of interest, their main objective is to minimise forecasting error. Therefore, they are more suitable for producing accurate forecasts in the short to medium-term based on past trends rather than evaluating policy impacts.

Machine-learning techniques are an alternative to time series models that can be applied to produce reliable homelessness projections. However, they require detailed datasets – e.g. large administrative or census data – that may not be easy to assemble from LAs in the English context. Therefore, time series forecasting models are likely to be more appropriate for predicting homelessness trends in England and other countries in the UK.

Other types of models can be applied to accommodate different policy objectives. For example, risk models can be applied to calculate homelessness risks at the individual and household level and thus, improve targeting of households and single people in priority need of homelessness prevention services. Further, non-standard sampling techniques such as the capture-recapture method can be used to measure the size of evasive homeless groups that are hard to survey.

Both methods depend on simple statistical concepts and can be further simplified to serve as standalone tools. Additionally, their outputs can feed into simulation models. For instance, risk models quantify the contribution of a set of predictors to the probability of facing homelessness. They can be used to estimate elasticities that reflect the sensitivity of homelessness outcomes to changes in predicting factors. Estimated elasticities can be then inserted in a more complex policy model to simulate outcomes under different scenarios.

Based on these findings, we conclude that a suite of models used for different policy purposes is more suitable than a single model designed to address the entire range of requirements for effective homelessness policy. Policy-making involves a wide set of goals that are not likely to be addressed by a single model regardless of its complexity. For example, an efficient policy mix aiming to tackle homelessness and support homeless groups depends on both reliable appraisal of planned reforms and accurate predictions of the homelessness levels in the short to medium-term.

6. References

Anderson, I., Christian, J., 2003. Causes of homelessness in the UK: a dynamic analysis. Journal of Community & Applied Social Psychology 13, 105–118. https://doi.org/10.1002/casp.714armstrong _ long-range forecasting.pdf

Armstrong, J.S., 2001. Standards and practices for forecasting, in: Principles of Forecasting. Springer US, Boston, MA, pp. 679–732. https://doi.org/10.1007/978-0-306-47630-3_31

Armstrong, J.S., 1985. Long-range forecasting. Wiley – Inter-science, New York.

Box-Steffensmeier, J.M., Freeman, J.R., Hitt, M.P., Pevehouse, J.C.W., 2014. Time series analysis for the social sciences. Cambridge University Press.

Bramley, G., 2017. Homelessness projections: core homelessness in Great Britain. Crisis, London.

Bramley, G., 2013. Embedding economic models of migration and household formation in a housing market model for localities and regions in New Zealand, in: British Society for Population Studies, Annual Conference. Swansea.

Bramley, G., Leishman, C.M., Cosgrove, P.J., Watkins, D., 2016. What would make a difference? Modelling policy scenarios for tackling poverty in the UK. Heriot-Watt University.

Bramley, G., Fitzpatrick, S., 2018. Homelessness in the UK: who is most at risk? Housing Studies 33, 96–116. https://doi.org/10.1080/02673037.2017.1344957

Bramley, G., Pawson, H., White, M., Watkins, D., Pleace, N., 2010. Estimating housing need. DCLG, London.

Bramley, G., Watkins, D., 2016. Housebuilding, demographic change and affordability as outcomes of local planning decisions: exploring interactions using a sub-regional model of housing markets in England. Progress in Planning 104, 1–35. https://doi.org/10.1016/j.progress.2014.10.002

Branas, C.C., Kastanaki, A.E., Michalodimitrakis, M., Tzougas, J., Kranioti, E.F., Theodorakis, P.N., Carr, B.G., Wiebe, D.J., 2015. The impact of economic austerity and prosperity events on suicide in Greece: a 30-year interrupted time-series analysis. BMJ Open 5, e005619–e005619. https://doi.org/10.1136/bmjopen-2014-005619

Browne, J., Hood, A., Payne, J., 2016. Living standards, poverty and inequality in the UK 2015-2016 to 2020-2021. Institute for Fiscal Studies, London.

Chamlin, M.B., 1988. Crime and arrests: an autoregressive integrated moving average (ARIMA) approach. Journal of Quantitative Criminology 4, 247–258. https://doi.org/10.1007/BF01072452

Cowan, C., Breakey, W., Fischer, P., 1986. The methodology of counting the homeless. Proceedings of the Survey Research Methods Section, American Statistical Association 170–175.

Dang, H.-A.H., Lanjouw, P.F., Serajuddin, U., 2017. Updating poverty estimates in the absence of regular and comparable consumption data: methods and illustration with reference to a middle-income country. Oxford Economic Papers. https://doi.org/10.1093/oep/gpx020

Dávid, B., Snijders, T.A.B., 2002. Estimating the size of the homeless population in Budapest, Hungary. Quality & Quantity 36, 291–303. https://doi.org/10.1023/A:1016080606287

Douidich, M., Ezzrari, A., Van der Weide, R., Verme, P., 2015. Estimating quarterly poverty rates using Labor Force Surveys: a primer. The World Bank Economic Review 30(3), 475-500.

Elbers, C., Lanjouw, J.O., Lanjouw, P., 2002. Micro-level estimation of welfare, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-2911

Giles, C., McCrae, J., 1995. TAXBEN: the IFS microsimulation tax and benefit model. IFS Working Paper W95/19. The Institute for Fiscal Studies, London, UK.

Greer, A.L., Shinn, M., Kwon, J., Zuiderveen, S., 2016. Targeting services to individuals most likely to enter shelter: evaluating the efficiency of homelessness prevention. Social Service Review 90, 130–155. https://doi.org/10.1086/686466

Holmans, A.E., 2001. Housing demand and need in England 1996-2016. Town and Country Planning Association.

Hood, A., Waters, T., 2017. Living standards, poverty and inequality in the UK: 2017–18 to 2021–22. The Institute for Fiscal Studies (IFS), London, UK.

Leishman, C., Gibb, K., Meen, G., O'Sullivan, T., Young, G., Chen, Y., Orr, A., Wright, R., 2008. Scottish model of housing supply and affordability. Scottish Government, Edinburgh.

Leyland, A., Barnard, M., McKeganey, N., 1993. The use of capture-recapture methodology to estimate and describe covert populations: an application to female street-working prostitution in Glasgow. Bulletin de Méthodologie Sociologique 38, 52–73. https://doi.org/10.1177/075910639303800104

Lupton, M., Collins, H., 2015. Living Rents – a new development framework for Affordable Housing. Joseph Rowntree Foundation/National Housing Federation.

McCleary, R., Hay, R., Meidinger, E., McDowall, D., 1980. Applied time series analysis for the social sciences. Sage Publications, Beverly Hills.

National Housing and Planning Advice Unit (NHPAU), 2009. A guide to the reading-CLG affordability model.

Office of the Deputy Prime Minister [ODPM], 2006. New projections of households for England and the regions to 2026. ODPM Statistical Release 2006/0042.

Otok, B.W., Seftiana, D., 2014. The classification of poor households in Jombang with random forest classification and regression trees (RF-CART) approach as the solution In achieving the 2015 Indonesian MDGs' targets. International Journal of Science and Research 3, 1497–1503.

Reed, D., 2016. Predicting homelessness with integrated data and machine learning. NYU Wagner and Furman Center for Real Estate and Urban Policy Collaboration with the New York City Center for Innovation through Data Intelligence. NNIP Partnership Meeting.

Scottish Council for Single Homeless, 2003. Forecasting the city of Edinburgh Council's temporary accommodation requirements to 31 March 2004. The City of Edinburgh Council, Edinburgh.

Shinn, M., Greer, A.L., Bainbridge, J., Kwon, J., Zuiderveen, S., 2013. Efficient targeting of homelessness prevention services for families. American Journal of Public Health 103, 324-330. https://doi.org/10.2105/AJPH.2013.301468

Sohnesen, T.P., Stender, N., 2016. Is random forest a superior methodology for predicting poverty? An empirical assessment. Policy Research Working Paper 7612, World Bank, Washington, DC. https://doi.org/10.1002/pop4.169

Sudman, S., Sirken, M.G., Cowan, C.D., 1988. Sampling rare and elusive populations. Science 240, 991–996. https://doi.org/10.1126/science.240.4855.991

Thoplan, R., 2014. Random forests for poverty classification. International Journal of Sciences 17, 9.

Venables, W.N., Ripley, B.D., 2002. Tree-based methods, in: Modern Applied Statistics with S, Statistics and Computing. Springer-Verlag, New York.

Williams, M., 2010. Can we measure homelessness? A critical evaluation of 'capture-recapture.' Methodological Innovations Online 5, 49.1-59. https://doi.org/10.4256/mio.2010.0018

Ministry of Housing, Communities & Local Government

Department for Work & Pensions

Homelessness

Causes of Homelessness and Rough Sleeping

Review of models of homelessness

alma economics