A Spatial Microsimulation Model of Labor Market Integration in Germany

Monika Obersneider Petra Stein *

Abstract

Microsimulation modeling has gained growing attention in the last years as a useful statistical tool for individual-based population projections. However, only a handful of datasets allow for microsimulation modeling concerning the small level spatial structure the individuals are embedded in. Hence, the interplay of regional disparities with individual and contextual outcomes has received insufficient attention in this field.

The aim of this contribution is to give an insight into the development of a nationwide, spatial microsimulation infrastructure in Germany ('MikroSim')[1], which results in a geo-coded synthetic database. This conference proceeding presents its potential application through regionalized modeling of labor market outcomes within migrant populations. The resulting model can be used for population projections. Special features of this projection will be the integration of regional small area estimates as well as the derivation of estimates from empirical results of longitudinal multilevel analyses. Furthermore, the spatial microsimulation model allows for the simulation of different 'what if' scenarios and future dynamics of the labor market integration of ethnic minorities on several societal levels of interest.

Key Words: Dynamic Microsimulation; Regionalized Modeling; Multilevel Analysis; Labor Market; Migration; Population Dynamics

1. Introduction

Statistical simulation methods can be used as analytical instruments to project potential societal developments and to analyze complex social dynamics. To some extent, using simulation methods means making up the data of interest. The difference in simulation methods lies in the rules that are followed for generating new data and in the base dataset used for this updating process. In the case of prominent macrosimulations, the data-modeling takes place on the contextual-level, whereas the used dataset consists of aggregate data. Macrosimulation is mainly used by statistical offices to estimate population forecasts (e.g., demographic change, effects of immigration for the population development, life-expectancy and care needs, or the share of the working-age population in the future).

According to the major theories in social sciences, the unfolding of social mechanisms, and their change in time takes place on the individual level (Coleman, 1986; Kalter & Kroneberg, 2014). Hence, the striven level of explanation in the social sciences is the micro-level (such as individuals or households). Thus, developments in populations and social structures should be updated only on the base of individual data and on the base of individual information. The method of microsimulation complies with these requirements. In this way, we can use rich data for updating the outcome of interest, as we can rely not only on cross-tabulations of aggregate data but also on individual transition probabilities estimated from social science studies (van Imhoff & Post, 1998, p. 107ff.).

*University of Duisburg-Essen, Lotharstr. 65, Duisburg, Germany

[1]Acknowledgements: This proceeding uses data from the Multi-Sectoral Regional Microsimulation Model ('MikroSim'). The 'MikroSim' project is funded by the German Research Foundation (DFG). The model is developed and implemented by the research group FOR 2559, a cooperation between the University of Trier, the University of Duisburg-Essen and the Federal Statistical Office of Germany.

One further advantage of microsimulations is the possibility to model the projections according to well-established individual-level theories. In this research project, we build our microsimulation on the theories of occupational performance. The specific goal is to investigate the labor market integration of ethnic minorities and its potential supply-side development in Germany. The aim is to investigate individual and spatial characteristics, which both are hypothesized to affect the integration outcomes and their future dynamics. Hence, there is a need for spatially geo-coded individual data for the whole population of Germany to capture this interplay of different explanation levels. The development of such a small-scale database will be the core of a regionalized microsimulation model of labor market integration described in the subsequent pages.

Accordingly, the paper is structured as follows. The focus of this research contribution is the development of a statistically measurable theoretical model for the labor market integration of migrants in Germany (section 2). Undoubtedly, the multidimensional research topic of modeling and projecting the potential labor market integration of migrants is relevant both for scientists as for politicians. Why, however, microsimulation is an appropriate method to reveal such complex social dynamics, will be discussed in section 3. Before concluding and discussing further planned research steps (section 5), we carry out a fictional example of a regionalized labor market integration based on a small-scale data projection and give insights into the ‘MikroSim’ project (section 4).

2. Modeling Labor Market Integration

When it comes to investigating the integration of minority groups, finding explanatory mechanisms of vocational outcomes of migrants is one of the leading research trajectories (Esser, 1999). The search for promoting and inhibiting factors of labor market integration can help to predict and steer the integration of migrants. We are interested in modeling and projecting the long-term development of the labor market integration of individuals. Thereby, we assume that individual outputs and their change are explained ideally on the unit-level (Coleman, 1986). Although the level of explanation is the micro-level, the individual labor market performance can also be affected by contextual and structural conditions, i.e., there is a need to capture the potential influencing factors from a multilevel perspective.

First, estimating the influencing factors on the individual level, the endowment with relevant types of capital has shown to be a powerful predictor (Kalter, 2006). According to this resource-oriented perspective, a significant part of ethnic differences can be traced back to the diverging human capital endowment (Becker, 1962) of migrants compared to the majority group. In line with robust empirical findings, we include individual school and vocational qualifications as central resources for the labor market integration and add these explanatory variables into the analysis. Furthermore, previous research ascribes additional systematic outcome differences within migrants to differences in cultural capital and social capital equipment according to their social origin, i.e., systematic differences in the socio-economic status of the parents, social-networks or language skills. These forms of ‘host country-specific capital’ (Kalter, 2006, p. 145) are also modeled to explain the differences and potential future dynamics of labor market integration more precisely.

Second, in addition to individual characteristics, regional disparities should be taken into account when estimating integration development. The current state of research illustrates the relevance of the spatial perspective on integration dynamics and their long-term relevance for the integration outcomes of migrants. The proportion of migrants in Germany varies regionally: In Western Germany, almost a quarter of the inhabitants has a migration background, whereas in Eastern Germany (excluding Berlin) this is the case for approx. 5

%. Furthermore, there are considerable differences in the proportion of migrants between urban (2015: 26.9 %) and rural (2015: 12.1 %) regions (Federal Statistical Office, 2016). Also, the composition of migrants or migrant groups varies regionally by country or country group of origin (e.g., Granato, 2009). From a theoretical perspective, it can be assumed that the regional share of migrants and the composition of migrant-groups affect the occupational positioning of migrants themselves (so-called endogenous effect). Also, it can be expected that the regional concentration of ethnic minorities correlates with the labor market success of the majority group members (so-called exogenous effect; see overview in Granato, 2009, p. 394).

Third, further contextual characteristics can affect the outcomes of migrants in the labor market, e.g., local unemployment rate, dominant industry sector, or community size (Kogan & Kalter, 2006). Thus, the labor market structure, as well as the labor market situation, can play a role in the vocational outcomes of migrants and their regional dynamics. Similarly, changes in the regional population composition (demographic processes, regional mobility, migration flows, or change of migrant generations; Gresch & Kristen, 2011, p. 213) can have an additional predictive power concerning the long-term integration development of ethnic minorities. In sum, to model and to simulate the phenomena of interest with high precision, multidimensional influencing mechanisms and compositional effects have to be considered.

3. Microsimulation

In this section, we will explain the technical functioning of microsimulations. As mentioned above, the unique advantage of microsimulations in comparison to other simulation methods is the possibility to use individual-level empirical parameters or transition probabilities for the updating processes. These parameters can be estimated with the help of nearly all conventional statistical methods: Linear or logistic regression, multilevel analysis, longitudinal, or dynamic modeling. The usage of information from empirical studies brings not only the advantageous possibility of updating the data on the individual level but also the disadvantage of including possible additional errors. Thus, statistical modeling in microsimulations goes along with specification randomness and misspecification error (see more in van Imhoff & Post, 1998, p. 109ff.). Both need to be addressed correctly and kept at a minimum. Nevertheless, the higher precision and richness of the output gained through individual-level information outweigh these points of caution for the specific interest of the present paper.

Furthermore, microsimulation serves as a tool for experimenting and not for exact predictions. Microsimulation can be seen as a ‘scientific computer game’ to test and compare the consequences of so-called ‘what if’ scenarios (Spielauer, 2011, p. 11). The question that can be answered is, e.g., how would the world look like under different conditions. In microsimulation, the unfolding of such fictitious scenarios can be played through on the individual level, on the contextual level, or they can emerge exogenously, e.g., through political interventions.

3.1 Dynamic Microsimulation

According to the updating process of the base data within a microsimulation, there are two possibilities to project the data into the future: The so-called static and dynamic aging approaches (Spielauer, 2011, p. 12). In both cases, we have the initial set of individuals in our baseline population at time point t (left side of Figure 1). Hence, the baseline data consists of individuals or households with specific characteristics, which will then be updated according to the rules of microsimulation modeling. In this example, we illustrate the two different ways of microsimulation through manipulating the baseline data comprising men and women of different ages.

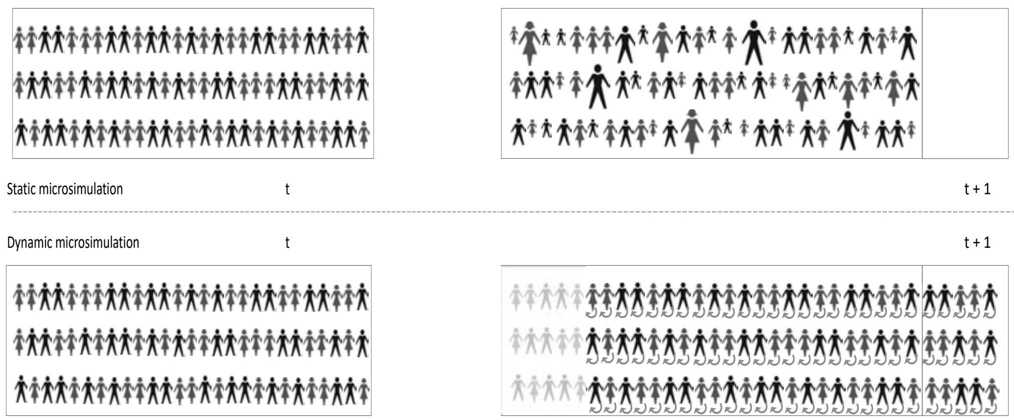

Figure 1: Updating process in a static and in a dynamic microsimulation.

Most of the microsimulation models are static models. Static means that no population aging processes are modeled, but outcomes of the individuals in the baseline data are reweighted according to their characteristics. Hence, static microsimulation is a common approach if the researcher is interested in examining short-term changes (Rephann & Holm, 2004, p. 383). According to this, the baseline dataset remains the same, while the change in individual outcomes becomes visible (see the top-right side of Figure 1). Static microsimulation is mainly used to investigate the so-called 'day-after' outcomes and scenarios, e.g., distributional effects of a tax reform regarding the net household income across different family constellations.

Dynamic microsimulation modeling, on the other hand, means that the individual biographies are updated on the base of individual transition probabilities to capture the long-term developments in populations. The unique feature of dynamic microsimulations is the data-driven projection of individual life-histories: The characteristics of the individuals are updated on bases of individual transition probabilities (see the bottom-right side of Figure 1). Actors in the dataset can have children, get older, become employed or unemployed, get married or divorced, or die. These characteristics reflect a life-cycle behavioral model (Spielauer, 2011, p. 12f.; Rephann & Holm, 2004, p. 383). As we are interested in labor market integration that includes population dynamics, we use the approach of dynamic microsimulation modeling.

3.2 Modular Updating of the Data

Following the definition of a dynamic microsimulation from the previous section, we simulate the individuals and their characteristics from the baseline data into the future. The updating process is clustered into so-called modules, in which updates of the different characteristics of a person are conducted. It is a sequential time-discrete model, meaning that every person will go through the simulation modules annually. Hence, it is a simulation model in which the modular updating of the individuals takes place chronically, whereby the ordering of the modules is held constant (Stein & Bekalarczyk, 2016, p. 243f.; Rephann & Holm 2004, p. 387). Figure 2 shows at what ages actors pass through the modules.

Furthermore, one can see from this illustration, whether the modules include regional predictors in addition to the individual characteristics (marked in blue). According to this information, micro-units in the data set will be updated, which generates a new dataset for each additional year in the microsimulation. The structuring of the modules rests on the assumptions and the modeling of labor market outcomes (see section 2).

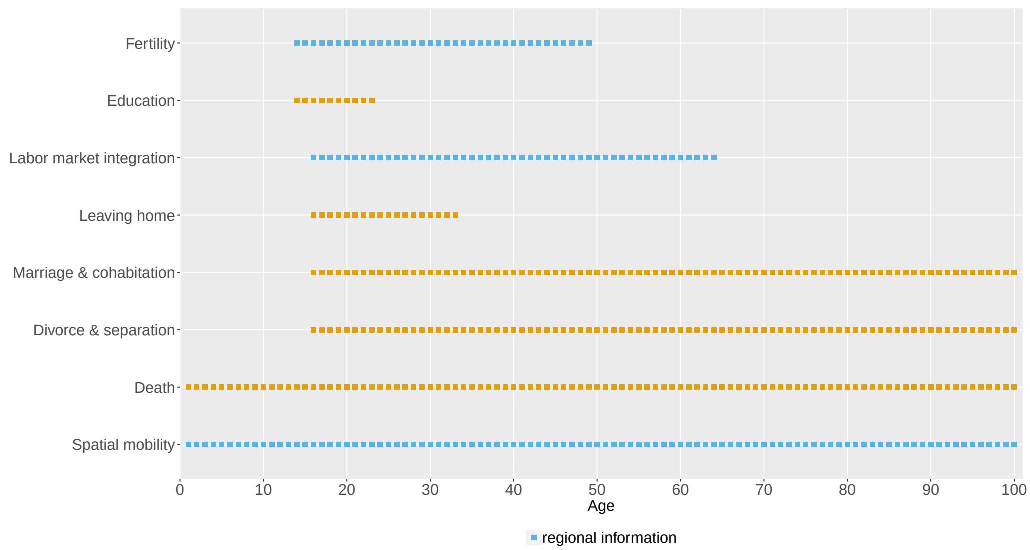

Figure 2: Modular annual updating on the individual-level.

Relevant influencing variables and updating parameters can be modeled on individual and regional levels through multiple regression analysis. We estimate different regression equations for each module, whereas individuals will be updated based on their transition probabilities derived from the regression results. The implementation of the regression results is randomized through Monte Carlo Simulation, whereby the regression-based individual probability is contrasted with a random risk. Specifically, a random number is drawn from a uniform distribution [0, 1] and compared with the estimated probability. If the latter is higher than the random risk, the event takes place and is updated accordingly in the simulation (Rephann & Holm 2004, p. 384).

Labor market integration module, e.g., consists of three different regression equations: Individual monthly income, probability of unemployment, and labor market positioning of migrants. Variables used for estimating these outcomes are, e.g., age, gender, educational level, migration status, number of children, marital status, language skills, length of stay in Germany, local unemployment rate, and regional ethnic concentration. Furthermore, the predictors in different modules can overlap. Thus, a change in one module indicates a change in all other modules, where the same variable is included. This interdependence within the projected individual biographies characterizes the population dynamics, which we can capture through using microsimulations.

4. Developing a Spatial Microsimulation Model

4.1 Construction of the base dataset

Previous sections showed that there is a need for geographical information for projecting labor market outcomes of migrants to analyze possible individual-, regional- or aggregate-level developments in Germany. Apart from this particular research topic, various societal questions show the need for spatial microsimulation projections. One example is the development of elderly care in different regional identities for better political planning, a research topic, also included in the 'MikroSim' project. These examples underline that simulation datasets should include individual, household data, regionalized, and geographical information for the appropriate modeling of the small level spatial structure the individuals are embedded in. So far, spatial base data for conducting microsimulations is missing in Germany. Only a few countries have already conducted spatial microsimulation models. These examples have shown that spatial microsimulation is a useful and promising analysis tool (SVERIGE in Sweden, MOSES in the United Kingdom, SMILE in Ireland, or CORSIM in the United States; see overview in Tanton, 2014, p. 16ff.). Thus, the goal of the 'MikroSim' project is to generate a population database with regionalized geo-coded information for Germany.

This 'MikroSim' base dataset is a synthetic dataset, constructed through various data sources. The dataset consists of the whole population of Germany including their geographical location. The basis for the synthetic data is an anonymized register data with only a few core variables (developed and used for conducting German National Census 2011; Münnich et al., 2012), whereby this register-data is expanded by survey data using (multinomial) logit models (Alfons et al., 2011). The synthetic data is aligned to known totals by applying a combinatorial optimization method of simulated annealing. In this way, integer weights can be gained, and marginal distributions of the small-area identities met (Tanton et al., 2014; Lovelace & Dumont, 2016). All the data is geo-coded in census grid cells (100 meters times 100 meters). Through the application of these data construction methods, we can use the generated synthetic dataset to analyze data from different hierarchical levels of interest: From the German federal states (16) to districts (402) to municipalities (approx. 11.000) to households (approx. 41 Million) to individuals (approx. 80 Million). This constructed synthetic data will be the base for conducting different dynamic discrete-time population projections in Germany.

4.2 A Spatial Microsimulation of Labor Market Integration: A Fictive Scenario

The research focus in this contribution lies in modeling vocational integration of migrants in Germany and its potential supply-side development from a regionalized perspective. As Rephann & Holm (2004) point out, the spatial dimension in microsimulations: '[...] is a useful addition when (1) it improves the accuracy of microsimulation models by introducing spatial and geographical variables as driving forces of change and (2) it allows results to be presented and interpreted for space and regions.' (Rephann & Holm, 2004, p. 389f.). As those aspects are of particular importance in the case of projecting the labor market integration of migrants, we will use the constructed spatial base data and dynamic microsimulation modeling in our further analysis.

The simulating process in a spatial microsimulation has comparable operating principles as the time-discrete dynamic microsimulation - the base dataset will be updated annually resulting in new simulated data for the variable of interest (Leim, 2008, p. 31). The simulation takes place in the modules, where the characteristics of the individuals are updated stochastically based on empirical transition probabilities (see section 3.2). As we use the discrete-time microsimulation approach, the simulation output can be assessed and interpreted for any year of interest (see the central part of Figure 3).

Although the operating principles are comparable with other microsimulations, the updating process has a spatial component. The transition probabilities are estimated for each small area, which requires a high amount of regression equations. This regionalized estimation gives us the possibility to analyze the simulation outputs from a geographical perspective. Not only the comparison between space and spatial dynamics through time can be analyzed, but also the spatial outcome can be disaggregated and compared using cross-tabulations concerning several covariates of interest (Tanton, 2014, p. 5ff.). Figure 3 contrasts the spatial outcome between the years 2011 (base dataset) and 2040. In this fictional example, we have simulated the regionalized unemployment rates of migrants. The individual-level outcomes are aggregated on the district level, yet outputs on geographically smaller areas are feasible. Thus, through using spatial microsimulation modeling, we can compare the outcomes in migrant unemployment between the regions in the same year. Besides, a comparison of one geographical area in time is possible, and we can disaggregate the unemployment rate in a region, e.g., by age, gender, or education of the individuals.

Figure 3: Fictive projection of labor market integration on the district level (2011 vs. 2040).

In addition to regional differences, we can take demographic changes in the composition of the population into account. This makes it possible to examine the interplay of causal individual-level mechanisms with compositional and contextual effects on the development of the labor market integration of ethnic minorities. Further analyses can be carried out based on any simulated year in the database, allowing for the prognosis of small-scale dynamics. The resulting model can then be used for population projections. Special features of this projection will be the integration of regional small area estimates as well as the derivation of estimates from empirical results of longitudinal multilevel analyses. In particular, the spatial microsimulation model enables us to simulate and contrast different ‘what if’ scenarios on several societal levels. In this way, we can compare the possible outcomes by potential changes in individual characteristics in time, regional disparities, or by political interventions. This interplay of the different levels of modeling helps us to capture and to assess future dynamics of the labor market integration of migrants.

5. Conclusion

According to the theories of labor market integration, there exists a complex set of relevant variables influencing the individual outcomes on the labor market. It is possible to explain and model systematic differences of first-generation migrants on the labor market compared to the majority population through variables derived from social science theories and to differentiate between the different countries of origin. Other than that, we can model the possible contextual influence factors, assuming that not only the individual characteristics matter but also the regional differences as diverging unemployment rates or the distribution of ethnic groups in a specific area. Additionally, we expect to capture the different labor market structures in specific areas. To be able to model the interplay of these theoretical levels of interest, a regionalized individual dataset is needed.

Only a handful of datasets in the world allow for microsimulation modeling with regard to the small level spatial structure, where the individuals are embedded in. Spatial microsimulation models are data-intensive and complex as both unit-level data and transition probabilities in small geographical identities are needed to update the data (Tanton, 2014, p. 18). So far, such spatial base data for conducting microsimulations is missing in Germany. Therefore, the goal of the 'MikroSim' project is the generation of a population database with regionalized geo-coded information for Germany. This constructed synthetic small-scale database can then be used to perform microsimulations with a spatial component.

Further steps concerning the research topic of this contribution will be the implementation of the theoretical assumptions on the base of the complete synthetic simulation dataset for Germany and the validation of this model. Whereas, it is not the intention to give an exact prognosis, e.g., about the proportion of migrants in the labor market in 2040, but rather to investigate individual and spatial characteristics, which both are hypothesized to affect the integration outcomes and their potential future dynamics. Furthermore, through the modeling of change processes regionally, the hypothetical 'what if' scenarios can be contrasted at any spatial level. Hence, the presented concept aims at making individual and spatial dynamics in vocational integration more tangible through microsimulation modeling.

REFERENCES

Alfons, A., Kraft, S., Templ, M., & Filzmoser, P. (2011). Simulation of close-to-reality population data for household surveys with application to EU-SILC. *Statistical Methods & Applications, 20*(3), 383-407.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *The Journal of Political Economy, 70*(5), 9-49.

Coleman, J. S. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology, 91*(6), 1309–1335.

Esser, H. (1999). Inklusion, Integration und ethnische Schichtung. *Journal für Konflikt und Gewaltforschung, 1*(1), 5–34.

Federal Statistical Office (2016). *Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2015*. Wiesbaden: Statistisches Bundesamt.

Granato, N. (2009). Effekte der Gruppengröße auf die Arbeitsmarktintegration von Migranten. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 61*(3), 387–409.

Gresch, C., & Kristen, C. (2011). Staatsbürgerschaft oder Migrationshintergrund? Ein Vergleich unterschiedlicher Operationalisierungsweisen am Beispiel der Bildungsbeteiligung. *Zeitschrift für Soziologie, 40*(3), 208–227.

Hannappel, M., & Troitzsch, K.G. (2015). Mikrosimulationsmodelle. In N. Braun & N. J. Saam (Hrsg.), *Handbuch Modellbildung und Simulation in den Sozialwissenschaften* (pp. 455–489). Wiesbaden: Springer Fachmedien Wiesbaden.

Kalter, F. (2006). Auf der Suche nach einer Erklärung für die spezifischen Arbeitsmarktnachteile von Jugendlichen türkischer Herkunft. Zugleich eine Replik auf den Beitrag von Holger Seibert und Heike Solga: "Gleiche Chancen dank einer abgeschlossenen Ausbildung?" (ZfS 5/2005). *Zeitschrift für Soziologie, 35*(2), 144–160.

Kalter, F., & Kroneberg, C. (2014). Between mechanism talk and mechanism cult: new emphases in explanatory sociology and empirical research. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 66*(1), 91–115.

Leim, I. (2008). *Die Modellierung der Fertilitätsentwicklung als Folge komplexer individueller Entscheidungsprozesse mit Hilfe der Mikrosimulation. Social science simulations* (Bd. 5). Marburg: Metropolis-Verlag.

Lovelace, R., & Dumont, M. (2016). *Spatial microsimulation with R*. Chapman and Hall/CRC.

Münnich, R., Gabler, S., Ganninger, M., Burgard, J.P., & Kolb, J.-P. (2012). *Statistik und Wissenschaft. Stichprobenoptimierung und Schätzung im Zensus 2011* (Bd. 21). Wiesbaden: Statistisches Bundesamt.

Rephann, T.J., & Holm, E. (2004). Economic-demographic effects of immigration: results from a dynamic spatial microsimulation model. *International Regional Science Review, 27*(4), 379–410.

Spielauer, M. (2011). What is social science microsimulation? *Social Science Computer Review, 29*(1), 9–20.

Stein, P., & Bekalarczyk, D. (2016). Zur Prognose beruflicher Positionierung von Migranten der dritten Generation. In R. Bachleitner, M. Weichbold & M. Pausch (Hrsg.), *Wissenschaftstheoretische und methodologische Problemlagen empirischer Voraussagen und statistischer Vorhersagen. Grundlagen und Kontroversen erfahrungswissenschaftlicher Prognoseverfahren* (pp. 171–205). Wiesbaden: VS-Verlag für Sozialwissenschaften.

Tanton, R. (2014). A review of spatial microsimulation methods. *International Journal of Microsimulation, 7*(1), 4-25.

Tanton, R., Williamson. P., & Harding, A. (2014). Comparing two methods of reweighting a survey file to small area data. *International Journal of Microsimulation, 7*(1), 76-99.

van Imhoff, E., & Post, W. (1998). Microsimulation methods for population projection. *Population: An English selection, 10*(1), 97–138.