Chapter 10
Model-Based Demography in Practice: I

Eric Silverman, Jakub Bijak, and Jason Hilton

10.1 Introduction

In the previous chapter, we outlined a new research agenda for demography inspired by the *model-based science* approach in disciplines such as population biology (Godfrey-Smith 2006). This paradigm is constructed as a cumulative extension of the previous four paradigms in demography – period, cohort, event-history and multilevel – and makes use of computational simulation approaches, particularly agent-based modelling. Ultimately, we propose that model-based demography will allow the discipline to develop a better understanding of the complex processes underlying population change at both the micro and macro levels, while retaining demography's characteristic empirical focus.

However, as with any massive change in disciplinary practices, we cannot expect demography to shift wholesale toward a model-based approach without some illustrative proofs-of-concept to draw from. Learning to design and implement simulations is no mean feat, and as outlined throughout this volume, tackling this task requires in many cases a significant shift in the type of research questions one seeks to ask. That being the case, how might demography move forward from here?

In this chapter we will focus on a few examples of demographic modelling which have successfully used agent-based modelling to add to demographic knowledge. We will start by discussing two well-known examples of demographic simulation, and transition from there into a detailed discussion of a more recent simulation project targeted specifically at the population sciences.

10.2 Demographic Modelling Case Studies

The history of demographic simulation is rather short, and in general relatively few demographers have made the jump to computational approaches. However, since the early 2000s there have been a few seminal papers which have capably illustrated the

E. Silverman, *Methodological Investigations in Agent-Based Modelling*,
Methodos Series 13, https://doi.org/10.1007/978-3-319-72408-9_10

possibilities inherent in these approaches, and these have influenced further work in subsequent years. Before diving into a detailed discussion of a recent model, we will outline two of these seminal papers, one of which in particular is closely related to the 'Wedding Doughnut' model to be discussed later in this chapter.

10.2.1 The Decline of the Anasazi

Axtell et al.'s 2002 model of the Kayenta Anasazi (Axtell et al. 2002) is well-known not just in demographic and archaeological circles, but in social simulation circles more generally. The model attempts to paint a historical demographic portrait of the Kayenta Anasazi tribe, who lived in Long House Valley in Arizona between around 1800 BC and 1300AD. At the end of that time, there was a precipitous drop in population, leading to a mass migration of Anasazi people out of the valley.

Axtell et al. used this scenario as a test case for the computational simulation of agricultural societies and their cultural and economic evolution (Axtell et al. 2002). The model reconstructs the Long House Valley landscape and places agents within it, though given the limitations of the historical data available agents represent households rather than individuals. As is typical for an agent-based model, the modellers developed rules of behaviour for the agents to follow when choosing locations to settle and cultivate. In brief, the agents seek locations which offer potential for successful maize cultivation, taking into account the location of other resources such as water and the proximity of neighbours. Agents can change their location in subsequent years based on the success of their efforts in providing nutrition; the level of nutrition received from these crops in turn affects agents' fertility.

This simulated population produced population dynamics which reflected the development, and eventual reduction, of the real Anasazi. Intriguingly, results also show that the northern reaches of Long House Valley could have still supported a reduced population, even during the late stages of the Anasazi's cultivation of the valley and the resultant degradation in soil quality. While the reason for this difference in outcomes is unclear, the results of the simulation show that in order to survive and stay sustainable the virtual Anasazi had to disperse into smaller communities; perhaps the real Anasazi were unwilling to fundamentally shift a settlement pattern and community organisation that had persisted for centuries, and chose to move on rather than make these significant adjustments.

A subsequent replication of the model[1] confirms Axtell et al.'s results, and reconfirms that environmental factors alone cannot account for the abandonment of

[1]Unfortunately model replications are surprisingly rare in interdisciplinary simulation research. The constant push for 'new' research results de-emphasises the role of replication in the scientific process, which in our view is damaging to our collective efforts. A good replication can confirm or challenge previous results, illuminate areas for improvement, and generally help the development of a rigorous and well-founded model-based science.

Long House Valley by the Anasazi (Janssen 2009). The model results are sufficiently close to the historical data to suggest that the simple behavioural and environmental rules in place provide a plausible explanation for the demographic changes seen in the Anasazi population. At the same time, we see that further details are needed to understand why the valley was abandoned.

Most importantly for our purposes, the explanation provided by this model takes us beyond what a statistical study founded in traditional demographic methods could have given us. By implementing an agent-based social system, we are able to illustrate and analyse the interactions between environmental pressures and low-level decisions about building homes and tending crops. Over the course of many simulated generations, we can see how these low-level processes drove a gradual migration northward through the valley as the environment became less suitable for farming, until finally only a small subsistence population could be supported. A statistical model with well-founded assumptions could have shown us the resultant population dynamics, but would not have facilitated this kind of detailed look at the low-level changes in community structure and behaviour over time.

Further, we can see that the model follows some of the core tenets of the model-based approach we outlined previously. The model avoids 'the beast' of expensive and extensive data demands (Silverman et al. 2011) by making use of archaeological data already in use for other projects, so there was no need for significant expenditures of money and time on feeding vast amounts of data into the simulation beyond what was already available. Qualitative data drawn from ethnographic studies of the Anasazi was also brought into the modelling process to help formulate the agents' behavioural rules. The model thus strikes a balance between staying empirically relevant – by using a digital reconstruction of the real landscape with behavioural rules developed from real data – and providing higher-level theoretical insight.

10.2.2 The Wedding Ring

Our next example of demographic simulation takes a rather different approach, focusing on a core concern to modern demography at a broader level rather than a specific historical case study. This well-known "Wedding Ring" model by Billari et al. (2007) attempts to bridge the micro-macro link discussed in the last chapter, in this case investigating the phenomenon of partnership formation. The authors describe the gap between the macro-level statistical explanations provided by demographers and the micro-level, individual studies on the process of searching for a partner, and view this model as an attempt to "account for macro-level marriage patterns while starting from plausible micro-level assumptions" (Billari et al. 2007, p. 60).

The Wedding Ring uses the phenomenon of 'social pressure' as the underpinnings of the partnership search process. In this formulation, social pressure originates from married agents within a social network, who then exert pressure on their non-married peers, who as the pressure increases feel an increasing urgency to find a partner. While most of us would prefer to think our partnership decisions are made purely out of love and not being pressured by family and friends, there is evidence from social research that social pressure does have a strong influence on our partnership choices (Bernardi 2003; Bernardi et al. 2007; Bohler and Fratczak 2007).

The virtual space Wedding Ring agents reside in is deliberately highly simplistic. Effectively they live in a cylindrical space, consisting of a one-dimensional spatial component and the vertical dimension of time (Billari et al. 2007). Agents each have a social network of what Billari et al. refer to as 'relevant others', which we can conceptualise as a separate space overlaid on the physical space defined by social distance between agents. As the number of married agents within the network of relevant others increases, social pressure on unmarried agents in that social neighbourhood also increases. Higher levels of social pressure directly influence an agent's partnership decisions, as agents under greater pressure will search a greater area for their prospective partners.

Billari et al. note that this effectively represents marriage as a diffusion process; however, marriage in this context differs significantly from other diffusion processes in that it requires another agent to participate, meaning that higher levels of social pressure do not guarantee that an agent will find a partner (Billari et al. 2007). In another simplification, though perhaps reflective of social norms, agents will only agree to marriage when both partners are within a certain age range. Once a partnership is formed, agents can decide to reproduce, which adds more agents to the Wedding Ring who will then undergo the same process as they age. The Wedding Ring model does attempt to allow for heterogeneity in agent decision-making by dividing agents into five different classifications according to which age ranges of agents they trust most. In other words, some agents most listen to the opinions of older agents, others listen to younger ones, and some agents listen equally to both.

Despite the overall simplicity of the model, the Wedding Ring produces hazard functions for partnership formation highly reminiscent of the patterns seen in the real world, and sensitivity analyses performed by the authors seem to indicate that the results are relatively robust to parameter changes (Billari et al. 2007). In the context of a nascent model-based demography seeking to avoid 'the beast', this study offers some reasons to be optimistic. The model's initial population was generated to mirror the age distribution of America in the 1950s, but notably this is the only appearance of any empirical data within the model. The simplistic world of the Wedding Ring is one dimensional, and the agents' behaviours are very simple, and do not take into account individual agent characteristics beyond age, location and social pressure classification.

The lack of data in this case does not result in an irrelevant model, however. The Wedding Ring is able to replicate patterns of partnership formation that are reflective of patterns in the real world, and in the process offer social pressure as a possible low-level explanation of those higher-level patterns. A statistical model of the same process would need substantially more investment in terms of data collection to provide similar insights. As we can see from the Wedding Ring, constructing our agents' behaviours on a foundation of social theory supported by evidence allows us to avoid that time-consuming aspect of traditional approaches, while still retaining empirical relevance and avoiding entirely arbitrary assumptions.

10.3 Extending the Wedding Ring

The Wedding Ring model, exciting as it may be, remains a proof-of-concept of sorts. The model provides a useful test-case for the application of relatively abstract agent-based models to the expansion of demographic knowledge, but there is plenty of room for further expansion. The exclusion of some more detailed social factors from the model also raises the question of whether the model results will remain robust when additional elements are added to more closely reflect real-world conditions.

Here we will discuss in detail a modelling project which aimed to replicate and expand upon the initial Wedding Ring model.[2] The expanded model sought to use the Wedding Ring as a basis for developing a multilevel simulation which makes use of the strengths of both statistical demography and social simulation. In essence, the model takes the framework presented by Billari et al. and adds demographic data and predictions based on population data from the United Kingdom. Along with partnership formation, the model also includes a simplistic representation of health status, which showcases the capacity of models like this to be used as policy tools.

We will begin by discussing the motivations and basic components of the model, before describing the specifics of the implementation.

10.3.1 Model Motivations

As we have discussed previously, demography and social simulation have been on a collision-course for some time. This model aims to take this process further by providing an exemplar model that combines several complex elements, namely demographic change, population health, and partnership formation. None of these elements by themselves are new topics for agent-based demography; a number of studies have focused on partnership formation, for example Billari and Prskawetz (2003), Todd et al. (2005), Billari et al. (2006, 2007), Billari (2006) and Hills and Todd (2008). However, we attempt to take things further here, by building a model that successfully ties together the strengths of statistical demography with social simulation. Such an enterprise can help the further development of model-based demography by demonstrating how the differing perspectives of these two

[2]For significantly more detail on this project, please refer to Bijak et al. (2013) and Silverman et al. (2013), a pair of papers intended to illustrate the expanded model for both demographers and simulation researchers. This chapter is based on that work but does not include all the analyses present in those two papers.

disciplines can be reconciled into a model that offers both novelty and policy relevance.

The 2003 volume by Billari and Prskawetz (2003) was influential in demography for presenting an approach to what they call 'agent-based computational demography', or ABCD. They present ABCD as focusing on the development of theories in demography, rather than focusing exclusively on prediction. We might extend this further, following Epstein who provided numerous examples of uses for simulation beyond prediction, many of which are as applicable to population change as to any other complex phenomena (Epstein 2008).

This model aims to further flesh out this approach in the context of model-based demography. While we outlined in the last chapter some principles by which the empirical focus of demography and the explanatory focus of simulation can be reconciled, putting this into practice requires the development of exemplar models that demonstrate the implementation of some of these principles. Some of this movement is already taking place, as we see in the popularity of microsimulation in demography, which incorporates some elements of mechanistic explanation (Gilbert and Troitzsch 2005), or increasingly empirically-focused models in various simulation domains (see Silverman and Bullock 2004 for a broad discussion, Grim et al. 2012 for an example).

Despite these incremental steps, there are still significant challenges to overcome. Microsimulation models are embedded in the event-history paradigm, and as such suffer from the excessive data demands carrying over from statistical approaches (Silverman et al. 2011). They also rarely include elements of spatiality, or detailed representations of social processes (Gilbert and Troitzsch 2005). Similarly, integrating data directly into agent-based models is a difficult process, and as a consequence many agent-based models rely on assumptions about agent behaviour and interactions between processes (sometimes to their detriment, and sometimes not). Here we will present a model that attempts to marry these two streams of work together into a cohesive whole, while also presenting some methods of model analysis that make the results more tractable.

10.3.2 Simulated Individuals

In attempting to resolve the gaps between statistical and agent-based models, it can be useful to reflect on the ways in which these methods conceptualise individual members of a population. In the case of statistical demography, we use observations such as survey or census data in order to understand *statistical individuals* (Courgeau 2012). When we take a further step into demographic microsimulation models, we start considering *synthetic individuals* (Willekens 2005), each consisting of a set of transition probabilities between different possible states during their life course. Upon taking the final step into agent-based models, we then need to consider the concept of the *simulated individual*.

The simulated individual consists of a set of rules governing their behaviour, which must be parameterised according to our available knowledge of those behaviours. This knowledge may be founded entirely in empirical data, or it may be built on a framework of assumptions derived from well-tested social theories. In either case, our relationship with population data changes its character in the agent-based context, as we may not be able to directly translate our available data into parameters that govern our agents' behaviours.

However, taking a page from the discussion of the data-based 'beast', we can make an effort to incorporate demographic data when it is appropriate to the research question posed. Agent-based models don't necessarily benefit from infusions of large amounts of data, which means we should incorporate data only when it adds to the empirical relevance of the simulation in question, not simply because it is desirable to do so. In order to alleviate concerns about our choices of parameter settings, particularly when data is missing or absent, we can use methods of uncertainty quantification to explore the parameter space and better characterise the behaviour of the simulation across a wide range of scenarios.

Figure 10.1 below illustrates our synthesis of statistical demography and social simulation. In this framework, elements of statistical demography can work in combination with an agent-based approach. The agent-based model allows us to investigate possible scenarios of population change by exploring our parameter space, letting us look beyond the one-generation time horizon often cited in the demographic literature (Keyfitz 1981; Bijak 2010; Wright and Goodwin 2009). In turn, statistical demography contributes empirical relevance through the incorporation of population data that can shape the development of our virtual population.

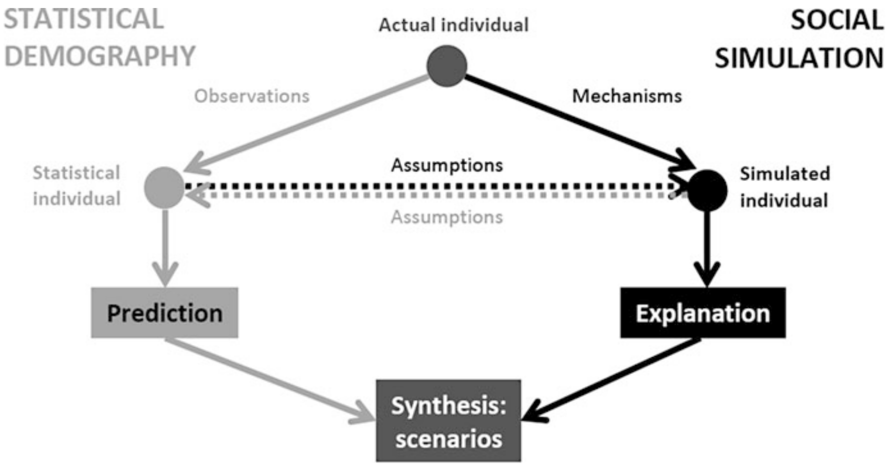

Fig. 10.1 Two approaches to modelling social systems: statistical demography and agent-based social simulation (Source: own elaboration, reprinted from Figure 1 in Silverman et al. 2013, drawing from Willekens 2005 and Courgeau 2012)

This approach thus brings together some of the key strengths of statistical demography and agent-based modelling that we have identified thus far. We are able to maintain demography's empirical relevance and access to rich, detailed population data. We can use simulated agents to investigate the complex relationships between social factors at the micro level and macro-level population trends. By investigating our parameter space, we can study possible futures of population change that extend beyond the notional one-generation predictive horizon, putting these explorations into context by applying uncertainty quantification techniques. All the while we make sure to incorporate data that enhances our model but does not unnecessarily complicate our simulation; we feed 'the beast' only what is absolutely necessary, and we maintain a careful balance of tractability and realism, in keeping with a modified Levinsian framework.

10.4 Extension Details

The model presented here is an extension of the Wedding Ring (Billari et al. 2007) designed to demonstrate this modelling framework. The extended Ring takes the form of toroidal physical space rather than the one-dimensional ring of the original, henceforth dubbed the Wedding Doughnut. The Doughnut was developed in the context of a five-year research project, The Care Life Cycle, devoted to the study of social care in an ageing UK population; the partnership modelling aspects of the Ring form a useful foundation here, because understanding family structures are key to this area given that the majority of social care in the UK is provided by family members (Vlachantoni et al. 2011).

10.4.1 Spatial and Demographic Extensions

Our decision to expand the agents' world into a grid-based, toroidal shape was driven by the hypothesis that situating agents in a ring may restrict their ability to form diverse networks of 'relevant others' during the course of a simulation run. The Doughnut world is square space 72 grid spaces on a side which wraps onto a toroidal topology, meaning agents can form networks across the vertical and horizontal boundaries. The initial population of agents was set at 800 individuals, a significant increase over the Ring's initial population of only 100 agents (Billari et al. 2007).

This major change to the virtual space the agents inhabit necessitated a number of significant changes to the model code. The spatial locations of agents had to be recorded in a different way, and calculations for spatial separation also had to change. A number of tests were run before settling on the initial population of 800; ultimately this figure was chosen as the model seemed to produce varied dynamics without unduly lengthy running times. Finally, virtually all model parameters had to be completely re-tuned, as the defaults presented in the original paper were constructed specifically for the 1D ring.

In order to connect the model more closely with real-world processes of population change, in keeping with the framework above we allowed the agents a much more varied life course and brought in statistical demographic elements.[3] In the original Ring, agents all lived to the age of 100 and died only at that age; in the Doughnut, we allow agents to die at any time, driven by age-specific probabilities from the Human Mortality database (2011). Given our interest in family structures for this model, we allow partnered agents to reproduce, with fertility rates taken from the Office of National Statistics (1998) and Eurostat (2011). Once the simulation advanced beyond our available population data, these rates were projected using bi-linear modelling methods derived from Lee and Carter (1992). Finally, the initial population was structured according to data from the 1951 census in England and Wales.

10.4.2 Health Status Component

In order to connect this model with the urgent debates on social care provision in the UK and other ageing societies, we further expanded the model to include a health status component. This aspect is kept intentionally simplistic, as it is intended more as a proof-of-concept to illustrate how models of this type can be policy-relevant.

In the health status component, agents may become ill at any given time according to a transition probability that increases as they age. We are restricting our focus here to long-term limiting illnesses which would require long-term social care in order to manage. Once agents contract such an illness in the simulation we assume they continue to display those symptoms until they eventually die. The transition probabilities are generated slightly differently for male and female agents, to account for the real-world trends in which males are somewhat more likely to enter a state of long-term limiting illness than females:

$$\begin{aligned} p(x) &= 0.0001 + 0.00041 \cdot \exp(x/16)(\textit{for males}) \\ p(x) &= 0.0001 + 0.00039 \cdot \exp(x/14)(\textit{for females}) \end{aligned} \tag{10.1}$$

10.4.3 Agent Behaviours and Characteristics

Agents in the Wedding Doughnut have the same characteristics as in the original Ring as described above, though additions were made in accordance with our

[3] For more details on the demographic elements, please see Bijak et al. (2013).

extensions to their behaviours and the new shape of the space. Given that agents can move during their life course in this model, we record their spatial positions as coordinates on the grid.

When agents form a partnership they are able to move on the grid to be with their new partner. In order to capture the effects of each partner's social and familial connections, we place the new couple on the grid in proportion to the size of each agent's social network of relevant others. Once the agents reproduce, any child agents are placed on the grid next to them; in this way we begin to see the formation of virtual households. Partnerships are permanent and can only form once both agents have reached the age of 16 or higher. In order to facilitate data analysis we record extensive records of every agent, including their spatial location; location and IDs of any partners or children; years of birth, death and partnership formation; and health status.

10.4.4 Simulation Properties

The simulation is written in the Java programming language using the Repast simulation toolkit. This free Java library and graphical interface is specialised for the implementation of agent-based models, and includes built-in functions for data recording, real-time visualisation, and data analysis. While the simulated population grows substantially during any given run, the simulation is not particularly demanding in terms of computing resources; any given run would take 2–3 min on a mid-range home desktop PC.

The simulation runs in discrete time-steps of one year. A complete simulation run consists of 300 years, starting from an initial population based on 1951 census figures, meaning that the final simulation year corresponds to the year 2250. In contrast, the original Wedding Ring ran for 150 years in each run.

Each time-step follows an identical sequence of events:

1. Agents age one year
2. Agents without partners:
   a. Identify relevant others
   b. Calculate social pressure
   c. Identify potential partners
   d. Form a partnership if possible
3. Agents with partners:
   a. Check fertility status
   b. If applicable, agents can produce children

4. All agents:
    a. Check health status
    b. If results indicate, transition agent to state of ill health
    c. Check mortality status
    d. Agents who die have year of death recorded
5. Remove dead agents from population
6. Add new child agents next to their parents
7. Record agent statistics in output files

During the final statistics-recording step, we keep a summary file for the simulated population, as well as detailed logs of events from specific agents. We also record the results of continuous hazard-rate calculations throughout the run; these are further summarised every decade and at the end of the run.

10.5 Simulation Results

10.5.1 Population Change

The demographic results derived from the simulation match the expected outcomes for an ageing population quite closely. While there are some notable differences due to the lack of international migration in our model, the results closely reflect the underlying processes of population change in this context.

Figure 10.2 below shows a population pyramid with a direct comparison between the averaged outputs of 250 simulation runs and the 2011 census data from the UK. The simulated results follow the real results very closely, and provide some assurance that the simulation is accurately reflecting the core demographic processes of fertility and mortality (and domestic migration, if not international). Note that these results reflect the 'base scenario' of health status; the other scenarios will be detailed later in this section.

10.5.2 Simple Sensitivity Analysis

Following the example of Billari et al. (2007), we performed a simple sensitivity analysis of some core parameters defining the social pressure functions of the simulation. Figure 10.3 includes four different model scenarios: default parameter settings; constant social pressure that does not vary in proportion to an agent's network of relevant others; a restricted spatial distance in which relevant others can be found; and a constant age influence, meaning that agents no longer differ in their responses to relevant others of differing age groups.

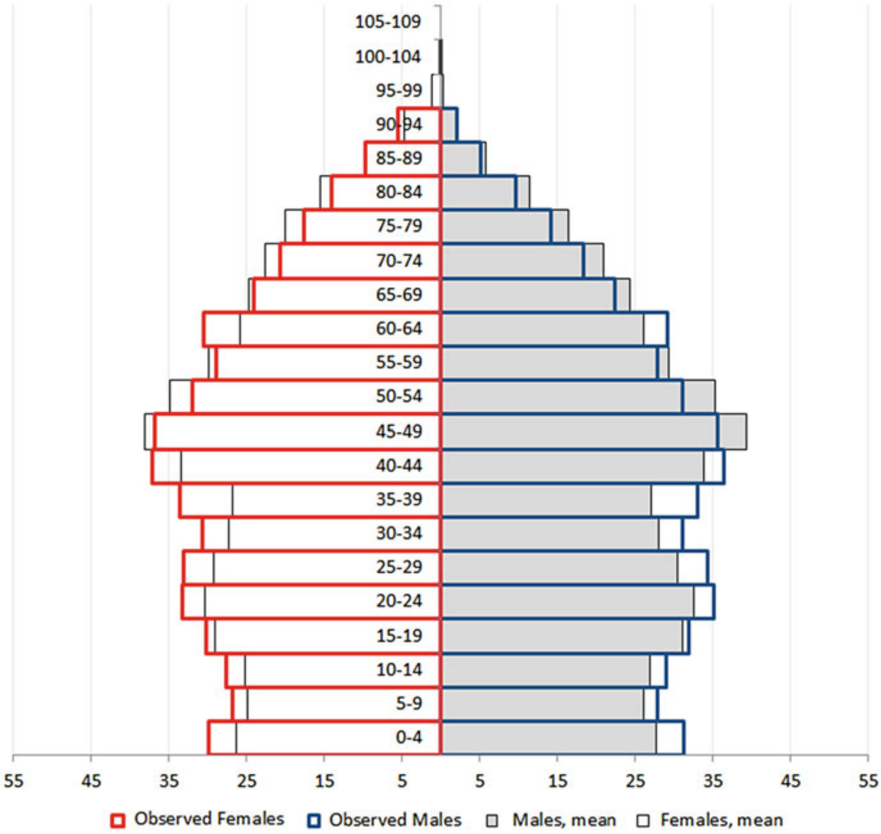

Fig. 10.2 Population pyramid compared with UK census data (Source: Figure 2, Silverman et al. 2013)

The graphs in Fig. 10.3 are each the result of an average of ten simulation runs. The hazard graphs for each of the four implemented scenarios closely replicate the identical tests seen in the original Wedding Ring (Billari et al. 2007), though in the Doughnut we see agents tending to form partnerships significantly earlier than in the 1D case. This suggests that further tweaks of the parameter settings would be necessary to identify portions of the parameter space that more closely match partnership formation patterns in the real world.

Further, we can see that the shape of the agents' virtual space can influence the macro-level population patterns we observe. This is not unexpected, given that the micro-level processes we model here are heavily influenced by the spatial positions of agents, and those positions change over each agent's life course as they form partnerships. As we will see in the next set of results, spatiality also plays a role in our analysis of health status.

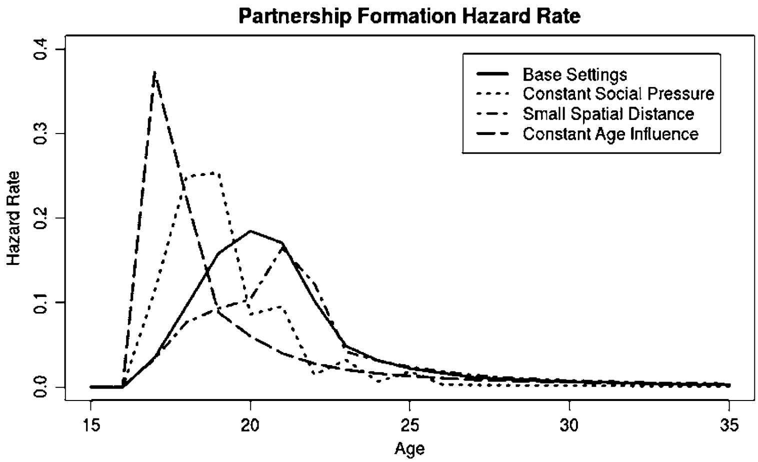

Fig. 10.3 Hazard rates for four different simulation scenarios (Source: Figure 3, Silverman et al. 2013)

10.5.3 Scenario Generation Example

One specific advantage of agent-based models in a context like the study of social care is that we can investigate the impact of the concept of 'linked lives'. This is the idea that agents in social simulations can exhibit long-term connections to other simulated individuals, allowing us to investigate the impact of social interactions and relationships on social and health outcomes at both the micro and macro levels.[4]

Traditional statistical models in demography do not have this capacity, given their reliance on aggregate population data in many cases. Even multilevel microsimulation models generally conceptualise their agents as individuals following personal life-course trajectories without the capacity for much interaction with other agents. This model is based on an agent-based framework, and as we will illustrate, even our very simple health status component allows us some rudimentary exploration of the linked-lives phenomenon in the context of social care provision.

For our simple analysis, we once again collected the results of 250 simulation runs, this time under three health scenarios:

1. Base Health Scenario: default values for ill-health transition probabilities
2. Good Health Scenario: halved values for ill-health transition probabilities
3. Bad Health Scenario: doubled values for ill-health transition probabilities

In implementing this simple comparison, we might imagine presenting this simulation to a group of policy-makers seeking insight on the expected levels of social care demand in the UK, extending quite far into the future. In such a case,

[4]For further discussion of this concept, see Noble et al. (2012) and Chap. 11 in this volume.

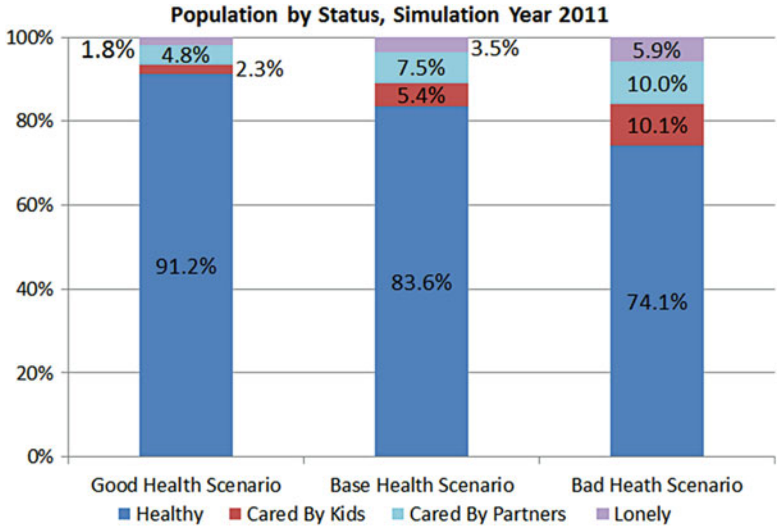

Fig. 10.4 Agent health status and care availability (results for 2011) (Source: Figure 4, Silverman et al. 2013)

given the reliance *informal social care* provided by family members, we may also want to provide details on the expected supply of informal care under different possible scenarios of general population health. Even with this highly simplified example, we can use the simulated data to provide some level of insight into these key questions:

1. How many ill agents will have access to informal care from their health spouses or children?
2. What proportion of agents may have *unmet care needs*?

The latter question in particular is of interest to policy-makers, as unmet care needs will have to be met with other resources, most probably with state assistance.

In Fig. 10.4 we present the outcomes of these three scenarios. Using our extensive records of the life-course of each simulated agent, we are able to determine the proportions of agents who are healthy or ill, and we can further subdivide the ill group into agents who could be cared for by their partners, children, or who have unmet care needs. We note here however that this is an optimistic virtual world in which we assume that any ill agents with available healthy spouses or children will receive care; in reality, this is unlikely to be the case, as partnerships may break down or children may move away or have other obstacles that prevent their availability for care.

As we can see from the results for each scenario, the proportion of ill agents in the population grows massively as the transition probabilities increase, from 9% in the best case to 26% in the worst. Additionally, we observe that as the scenario worsens, the burden on children for providing care grows faster than for spouses; this is due to the increased age-specific transition probabilities for ill health making it more likely that spouses become ill together, rather than one being left in a sufficiently healthy state to provide care.

10.5.4 Spatial Distribution of Informal Care

We can also use our recording of agents' spatial locations to investigate the distribution of healthy carers across our simulated Doughnut. Again this is a highly simplified example, but this gives us an illustration of how a model of this type might offer additional data over its statistical demographic counterparts. Using an agent-based methodology we are able to illustrate how agents are distributed spatially after simulation runs, and then draw conclusions about how those distributions vary under different scenario settings. In the case of social care, the availability of informal care within a reasonable distance could be an essential factor in determining whether an individual in need of care will need to seek state assistance in order to cope.

Figure 10.5 shows the fraction of ill agents with healthy carers available within a given distance on the Doughnut. These values were calculated using agents' spatial locations and health status across 250 simulation runs for each of our three health status scenarios. As we can see from the results, significantly fewer ill agents have access to carers in the same household (distance = 0) in the Bad Health Scenario than the Good Health Scenario. In a real-world context, this would tend to leave a greater burden on adult children to return home to provide care, as the absence of

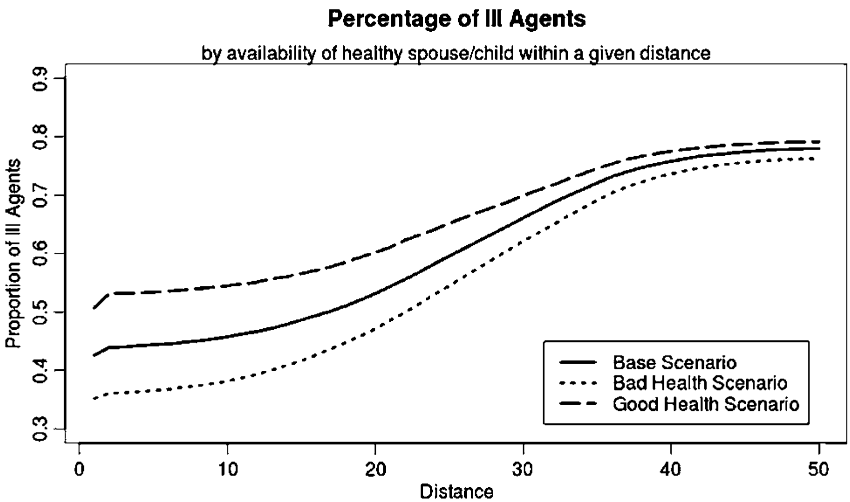

Fig. 10.5 Cumulative availability of care for ill agents by distance, simulation year 2011 (Source: Figure 5, Silverman et al. 2013)

an in-house informal carer implies that both spouses are in need of care (or that the ill parent is alone with their spouse having died). This has significant implications for policy-makers, who would need to consider how to support adult children who wish to make a contribution to care, or to 'nudge' somewhat reluctant children to consider helping out. For those ill individuals unlucky enough to have neither a healthy spouse nor healthy adult children within a reasonable distance, we would then need to consider supporting those individuals with state-funded formal care.

10.5.5 Sensitivity Analysis Using Emulators

One area of difficulty for social simulations, discussed previously in Part II, is the analysis of simulation results. When investigating multilevel simulations incorporating complex interactions between agents and their environment, understanding the specific impact of our parameter settings can be a very difficult undertaking. As a consequence, researchers more accustomed to the well-established methodologies of statistical demography can find this aspect of agent-based modelling concerning; after all, if we build a fascinating model but are unable to determine what actually happened during a run, have we gained enough knowledge of the mechanisms at play to justify the time and effort involved?

However, recent innovations in statistics have provided new ways of understanding the impact of model parameters on results, even in very complex simulations. Here we have followed the example set by the 'Managing Uncertainty in Complex Models' project, undertaken at the University of Sheffield until 2012, and used *Gaussian process emulators* to examine the impact of simulation parameters in the Wedding Doughnut.

We will only summarise the approach briefly here, as a certain level of intensive statistical exposition is required to explain these emulators in detail; for a more detailed examination in the context of this specific model, please see Silverman et al. (2013). In essence, a Gaussian process emulator creates a statistical model of the computational model, known as the *simulator*, and then decomposes the variance of our main output variable of interest into a constant term, a series of *main effects* related to our input parameters, and interaction effects between those parameters (Oakley and O'Hagan 2002, 2004; Kennedy and O'Hagan 2001). Effectively we are then left with an illustration of the amount of output variance that can be accounted for by each of our input parameters.

Kennedy then took this method further (Kennedy 2004), noting that simulations may contain additional variability due to uncertainty within the computer code itself. This can be accounted for through an additional term, the *nugget*, which addresses this additional uncertainty.

In order to implement a Gaussian process emulator for our simulation results, we identified four main input parameters: α and β, two parameters defining the social pressure function as formulated in the original Ring model (Billari et al. 2007); c, the scaling term in our transition probability function (set at 0.00041 and 0.00039 in

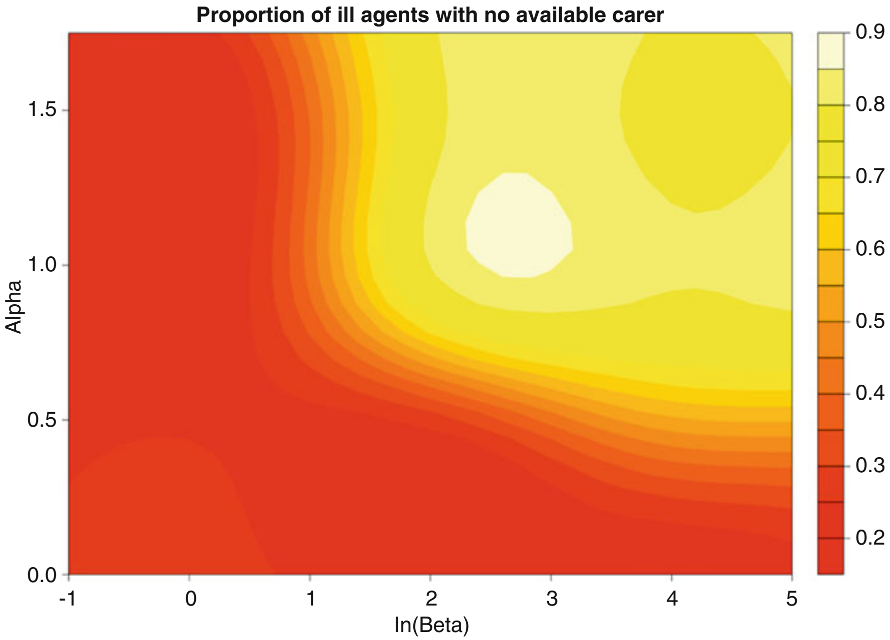

Fig. 10.6 Share of ill agents with no available carers (Source: Figure 6, Silverman et al. 2013)

Eq. 10.1); and d, the spatial distance in which partner search is allowed. Our output of interest is defined as the share of ill agents without access to a healthy spouse or adult children. We then run the simulation 400 times at a range of values for these four key parameters, and recorded the final output values for each run. We input those results into the free software GEM-SA (Gaussian Emulation Machine for Statistical Analysis) by Kennedy and O'Hagan to run this initial emulator, and the results of 41,000 emulator runs are shown in Fig. 10.6 in the form of a heat map.

The emulator produced a mean of 55.4% of ill agents with no access to informal carers; note however that this mean is generated from results across the entire segment of the parameter space, and many areas of this space will have very unrealistic values of the four input parameters. The vast majority of the variance in the final output values originated from α and β, at 29.8% and 48.6% of the variance respectively.

In order to more clearly illustrate the link between partnership formation and informal care provision in the model, we implemented a second emulator, this time using the share of agents who partnered over the course of the simulation as our final output of interest. In the model an agent's health status has no impact on partnership decision-making, so in this iteration we only used α, β, and d as input parameters.

Figure 10.7 illustrates the results, again the form of a heat map. The values of α and β are again accounting for most of the output variance, 30.2% and 45.6% respectively, and 17.7% of the remaining variance is due to the interaction of those two terms.

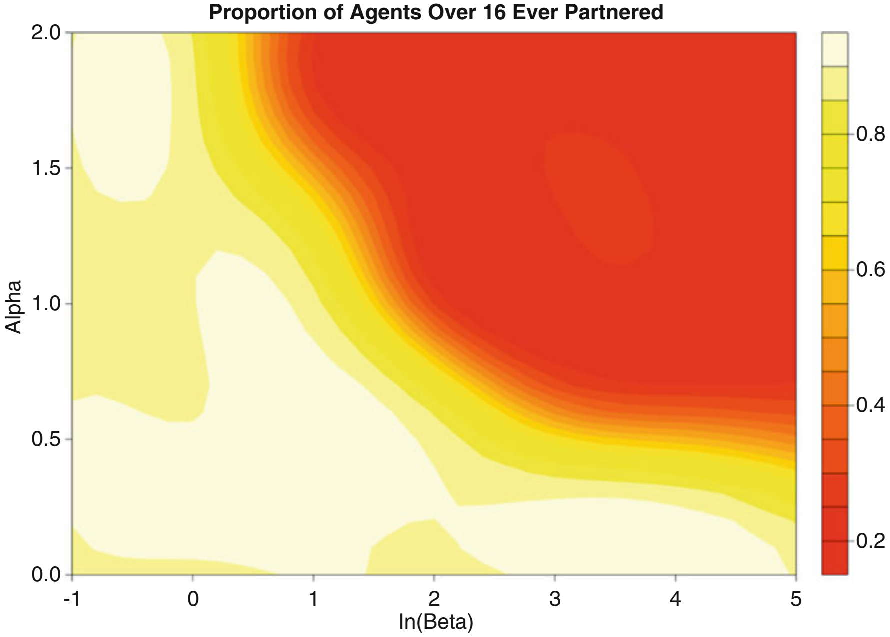

Fig. 10.7 Share of agents who have ever been married (Source: Figure 7, Silverman et al. 2013)

Taking these two visualisations together, we see a clear correspondence in the parameter space between high partnership formation levels correlating to lower unmet care need, and low partnership formation correlating to high unmet care need. The emulator thus serves as a comprehensive analysis of a segment of the simulation's parameter space, allowing us to investigate an enormous range of possible scenarios with a relatively small number of simulation runs. In the context of more complex simulations, we can easily see how the Gaussian emulator method puts our input parameter values in context and allows us much greater insight into their impact on simulation outputs.

10.6 The Wedding Doughnut: Discussion and Future Directions

As mentioned above, these results are intended as a clear and simple illustration of how even relatively abstracted models informed by real-world population data can answer policy questions that traditional demographic methods struggle to address. Even in our simplistic case, we are able to discuss possible scenarios of population change and population health at both the micro and macro levels. Through detailed study of simulation results, we are able to explore the possible consequences of policy changes, and draw some conclusions about where potential policy solutions are most urgently needed.

The model results here also indicate that spatiality plays an important role in the function of these social processes at the micro level. While there may be circumstances in which modern technology such as telemedicine systems can be of some help, in the case of social care we are studying individuals who need direct, hands-on help with simple activities of daily living, from changing clothes to going to the bathroom. As such, in cases like this where direct interaction is paramount, these simulations underline the importance of considering spatiality when making predictions regarding care demand and supply. Fortunately, such data is often available through the census or large-scale survey studies, but we would suggest that agent-based models are better able to directly represent the impact of spatiality then statistical, population-level models.

Of course there are clear areas where this simulation falls short of being realistic. As we outlined in previous chapters, this is not necessarily a negative in cases where we are investigating general social theories, but in the case of a policy-relevant area like social care, additional details implemented sensitively would enable us to investigate additional important social factors that can significantly impact the demand for and supply of social care. In this model we have only a very simplistic representation of ill health, for example, whereas in the real world, different individuals will have highly differentiated levels of care need, which will demand different amounts of investment from their families or the state. Partnerships also continue for life in this model, which doubtless inflates the numbers of agents having access to healthy carers; in reality partnership dissolution is commonplace, which could very well have a significant effect on the levels of unmet care need in the population.

In the next chapter, we will discuss a simulation which leaves the Wedding Ring framework behind, and adds additional levels of detail for each agent's life-course. In this model, partnerships can both form and dissolve, health status is not simply a binary state of ill-or-healthy but involves five increasing states of severity, and agents can migrate out of the family home for a much larger array of reasons than just finding a partner. We will also make use of more sophisticated methods of sensitivity analysis to investigate the model's parameter space.

10.7 General Conclusions

In this chapter we have examined some practical examples of the application of agent-based social simulation methods to the discipline of demography. Our starting examples of the modelling of the Anasazi abandonment of Long House Valley (Axtell et al. 2002) and Billari's examination of partnership formation in a simple one-dimensional space (Billari et al. 2007) illustrated some of the benefits of agent-based methods for the generation of new demographic knowledge. These two models demonstrate how model-based demography has the potential to answer questions that traditional statistical demography cannot, and how avoiding 'the beast' of demographic data demands can allow us to maintain tractability even in models of complex processes of population change.

Our detailed examination of the extended Wedding Doughnut model of partnership formation and health status takes additional steps toward integrating the richness of demographic data with the flexibility and power of simulation. This model has a fundamentally simple foundation, providing the essential processes necessary to generate a realistic illustration of an ageing population based on UK census data. By integrating statistical demographic data and predictions into this model, we are able to ensure the results remain relatively realistic; however, the amount of data incorporated into the model is small by the standards of multilevel approaches in demography, as we leave the partnership decision-making within the model to the simple social pressure function derived from research results in the social sciences.

A further consequence of this approach to modelling is that we are able to investigate the actions of agents embedded in both physical and social spaces. This opens the door to better understanding of the impact of linked lives – the substantive connections between individuals that can have both social and physical consequences. The Wedding Doughnut model presents a simple example of how these social connections and agents' spatial distribution interact in the context of social care, in which these two aspects are critical in driving the demand and supply of social care in an ageing society.

As with any simulation of social systems, we are left with substantial archives of data after running the model many hundreds of times; in this particular case we generated several gigabytes of detailed logs of every agent action across numerous possible scenarios. However, by implementing a probabilistic sensitivity analysis using Gaussian process emulators, we are able to develop a more systematic understanding of the impact of each key input parameter on the final output variance. In turn, we are able to insulate ourselves somewhat from the analysability and tractability concerns so often cited by those skeptical of simulation approaches.

Thus, we can begin to see how model-based demography can take shape, and take advantage of the empirical richness of demography and the investigation of complexity facilitated by agent-based modelling approaches. However, the model demonstrated here remains relatively simple, and while we were able to generate useful illustrative results from this effort, establishing model-based demography as a substantive paradigm within the discipline will require testing these principles and methods with more robust simulations.

In the next chapter, we will present another simulation study which takes the simulation aspects to a higher level of detail, while still remaining highly tractable and analysable using the same tools outlined here. From there we will take stock of the lessons learned from these early explorations in model-based demography, and outline how future investigations within this paradigm might proceed.

Acknowledgements This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/H021698/1 "Care Life Cycle", funded within the "Complexity Science in the Real World" theme. Contributions from Jason Noble and Viet Dung Cao are gratefully acknowledged.

References

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harbeger, J., Chakravarty, S., Hammond, R., Parker, J., & Parker, M. (2002). Population growth and collapse in a multi–agent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences, 99*(3), 7275–7279.

Bernardi, L. (2003). Channels of social influence on reproduction. *Population Research and Policy Review, 22*(5–6), 527–555.

Bernardi, L., Keim, S., & von der Lippe, H. (2007). Social influences on fertility: A comparative mixed methods study in Eastern and Western Germany. *Journal of Mixed Methods Research, 1*, 23–47.

Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view* (Springer series on demographic methods and population analysis, Vol. 24). Dordrecht: Springer.

Bijak, J., Hilton, J., Silverman, E., & Cao, V. (2013). Reforging the wedding ring: Exploring a semi-artificial model of population for the united kingdom with Gaussian process emulators. *Demographic Research, 29*(27), 729–766.

Billari, F. (2006). Bridging the gap between micro-demography and macro-demography. In G. Caselli, J. Vallin, & G. Wunsch (Eds.), *Demography: Analysis and synthesis* (pp. 695–707). New York: Elsevier.

Billari, F., & Prskawetz, A. (2003). *Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour*. Heidelberg: Physica Verlag.

Billari, F., Fent, T., Prskawetz, A., & Scheffran, J. (2006). *Agent-based computational modelling: Applications in demography, social, economic and environmental sciences*. Heidelberg: Physica Verlag.

Billari, F., Aparicio Diaz, B., Fent, T., & Prskawetz, A. (2007). The "Wedding Ring": An agent-based marriage model based on social interaction. *Demographic Research, 17*(3), 59–82.

Bohler, C, & Fratczak, E. (2007). Learning from others and receiving support: The impact of personal networks on fertility intentions in Poland. *European Societies, 9*(3), 359–382.

Courgeau, D. (2012). *Probability and social science: Methodological relationships between the two approaches* (Methodos series, Vol. 10). Dordrecht: Springer.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation, 11*(4), 12. http://jasss.soc.surrey.ac.uk/11/4/12.html.

Eurostat. (2011). Eurostat statistics database: Domain population and social conditions. http://epp.eurostat.ec.europa.eu. Accessed 27 Oct 2011.

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the social scientist*. Maidenhead: Open University Press.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*(5), 725–740.

Grim, P., Thomas, S. B., Fisher, S., Reade, C., Singer, D. J., Garza, M. A., Frye, C. S., & Chatman, J. (2012). Polarization and belief dynamics in the black and white communities: An agent-based network model from the data. In *Artificial life XIII: Proceedings of the thirteenth international conference* (pp. 186–193). Cambridge: MIT Press.

Hills, T., & Todd, P. (2008). Population heterogeneity and individual differences in an assortative agent-based marriage and divorce model (MADAM) using search with relaxing expectations. *Journal of Artificial Societies and Social Simulation, 11*(4), 5.

Human Mortality Database. (2011). Human mortality database. http://www.mortality.org/cgi-bin/hmd. Accessed 26 Jul 2011.

Janssen, M. A. (2009). Understanding artificial Anasazi. *Journal of Artificial Societies and Social Simulation, 12*(4), 13.

Kennedy, M. (2004). Description of the Gaussian process model used in GEM-SA. Software manual. http://ctcd.group.shef.ac.uk/gem.html. Accessed 23 May 2012.

Kennedy, M., & O'Hagan, T. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B, 63*(3), 425–464.

Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review, 7*(4), 579–593.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association, 87*(419), 659–671.

Noble, J., Silverman, E., Bijak, J., Rossiter, S., Evandrou, M., Bullock, S., Vlachantoni, A., & Falkingham, J. (2012). Linked lives: The utility of an agent-based approach to modelling partnership and household formation in the context of social care. In C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, & J. M. Uhrmacher (Eds.), *Proceedings of the 2012 winter simulation conference*. IEEE.

Oakley, J., & O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika, 89*(4), 769–784.

Oakley, J., & O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: A Bayesian approach. *Journal of the Royal Statistical Society, Series B, 66*(3), 751–769.

Office for National Statistics. (1998). *Birth statistics* (Series FM1 (27)). London: Office for National Statistics.

Silverman, E., & Bullock, S. (2004). Empiricism and artificial life. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. A. Watson (Eds.), *Artificial life IX: Proceedings of the ninth international conference on the simulation and synthesis of living systems*. MIT Press/Bradford Books.

Silverman, E., Bijak, J., & Noble, J. (2011). Feeding the beast: Can computational demographic models free us from the tyranny of data? In *Advances in artificial life: ECAL 2011* (pp. 747–754). Cambridge, MA: MIT Press.

Silverman, E., Bijak, J., Hilton, J., Cao, V., & Noble, J. (2013). When demography met social simulation: A tale of two modelling approaches. *Journal of Artificial Societies and Social Simulation, 16*(4), article 9.

Todd, P. M., Billari, F. C., & Simao, J. (2005). Aggregate age-at-marriage patterns from individual mate-search heuristics. *Demography, 42*(3), 559–574.

Vlachantoni, A., Shaw, R., Willis, R., Evandrou, M., & Luff, J. F. R. (2011). Measuring unmet need for social care amongst older people. *Population Trends, 145*, 60–76.

Willekens, F. (2005). Biographic forecasting: Bridging the micro–macro gap in population forecasting. *New Zealand Population Review, 31*(1), 77–124.

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting, 25*(4), 813–825.