Towards Data-Driven Agent-Based Simulation: 
A Case Study of West Asian Urban Dynamics 
(Extended abstract)

Stefano Picascia\textsuperscript{1,2}, Ali Termos\textsuperscript{3}, and Neil Yorke-Smith\textsuperscript{3}

\textsuperscript{1} Centre for Policy Modelling, Manchester Metropolitan University, UK.  
\textsuperscript{2} Laboratorio Dati Economici Storici Territoriali, University of Siena, Italy.  
\textsuperscript{3} Olayan School of Business, American University of Beirut, Lebanon.

\texttt{stefano@cfpm.org} 
\texttt{at36,nysmith}@aub.edu.lb

Abstract. This extended abstract discusses the challenges of obtaining data for agent-based simulation in one case. Our motivation is to explore the effects of significant migrations upon urban dynamics. To date, we have produced a theoretically-grounded spatial agent-based model, and validated it with data in urban contexts without migration. Now, in contexts with migration, upon moving from an abstracted model to a data-grounded model, we encounter a number of challenges in extending and validating the model. The majority of these challenges concern data.

1 Introduction

Rapid international migration of significant populations is generating profound implications for countries in West Asia, Europe, and other regions. Our motivation is to develop an agent-based model to capture the existence of such migrant and refugee flows, and to explore their effects on urban dynamics: housing prices, segregation, slums, gentrification, and the economic cycle.

Our previous work [2,1] built an agent-based simulation of urban dynamics founded on Rent-Gap Theory and validated it (in the absence of refugees) on historical data from two UK cities. Turning to metropolitan Beirut, a capital city in West Asia that has experienced significant migration in the last five years, we extended the model to account for migration, although with the city modelled in abstract form [3].

In now moving from an abstracted model to a data-grounded model, we encounter a number of challenges in calibrating and validating the model. The majority of these challenges concern data. We discuss them in this short paper.

The simulation, developed in NetLogo with GIS extensions, requires four types of data: geographic, demographic, economic, and socio-physical. First, official geographic data in Lebanon is considered a matter of national security. Second, no nationwide census has been taken in the country for 85 years, let alone official enumeration of refugees. Third, whereas the Central Bank will
only in 2017 begin recording housing prices, we require a monthly index by
neighbourhood. Fourth, we require data about housing quality and density.

The relevance of our work to the ABMUS workshop is as a case study of
the inter-disciplinary and multi-source effort to obtain and integrate data [4].
While open-source GIS data proves adequate, and while there are a basket of
population estimates that can be weighed, to obtain economic data we developed
and validated a multi-variate time series econometric analysis. We are currently
performing fieldwork to obtain socio-physical housing information.

2 Challenges of Data

We briefly discuss our current pragmatic process for obtaining data of each type,
in order of escalating challenge.

**Geographic data.** While official geographic data is not available, we can use
freely-available GIS data from OpenStreetMap. Fig. 1 shows a visualization of
the simulation model using geographical data.

![Fig. 1: Simulation visualization of metropolitan Beirut. The Mediterranean Sea lies to the north and west. Neighbourhood locations (squares) are coloured according to price. City quarters are marked by narrow lines. Agents are seen as circles on locations; the type and colour of circle indicates ethnicity and income of the agent.](image-url)
Demographic data. With demographics of Lebanon being a delicate matter for sectarian and political reasons, there are no official figures. Estimates circa 2010 put the Lebanese population living in the country at around 4 million. By 2015, the UN had recorded more than 1 million official refugees from Syria, while including undocumented refugees and those who have emigrated through Lebanon, the estimates are as high as 2 million. We use a basket of estimates and perform projection onto the city quarters.

Economic data: Housing price index. Due to the lack of a housing price index in Lebanon or historical real estate prices for residential dwellings in Beirut, we developed a two-stage least squares model using the economic theory of demand and supply. In the first stage we estimate the following structural model in which the demand and supply sides are given respectively by:

\[ q_t = \alpha_0 + \alpha_1 y_t + \alpha_2 p_t + \alpha_3 r_t + u_t \]  
\[ q_t = \beta_0 + \beta_1 c_t + \beta_2 p_t + \beta_3 r_t + v_t \]  

where \( q_t \) is the housing permits in square metres, i.e., representing the quantity of construction, \( y_t \) is income as proxied by the Coincident Index, \( c_t \) is the cost of construction, \( r_t \) is the long-term lending interest rate, and \( p_t \) is a price index for real estate. For the lack of any official real estate price index, we devised a crude price index as: the monthly ratio of the dollar value of all real estate transactions in time \( t \) divided by the number of real estate transactions in time \( t \). Note \( p_t \) includes non-residential and residential transactions.

In equilibrium, by setting equations (1) and (2) equal, we obtain the reduced form equations for \( p^*_t \) and \( q_t \) as follows:

\[ p^*_t = \lambda_0 + \lambda_1 y_t + \lambda_2 p_t + \lambda_3 r_t + \epsilon_t \]  
\[ q_t = \gamma_0 + \gamma_1 c_t + \gamma_2 p_t + \gamma_3 r_t + \nu_t \]  

In the second stage, we estimate the structural equations (1) and (2) using the predicted value derived in the first stage, i.e., \( p^*_t \) by equation (3).

Fig. 2 shows (light line) the average real estate sales value, i.e., \( p_t \), and (dark line) the index for price of residential real estate in Beirut, i.e., \( p^*_t \). It can be seen that our estimate tracks the data quite robustly and tends to smooth out excess volatility. We are working on a projection to each city quarter.

Socio-physical data: Housing quality. The most difficult data to obtain concerns property maintenance levels. Our current plan is to use a combination of a field survey of city quarters, construction inspection reports, and expert opinion about the quarters at aggregate level.

3 Summary

This extended abstract summarized the challenges of obtaining data for a grounded agent-based simulation of urban dynamics. Processes we are currently using comprise open-source data, estimates from published studies, statistical regression from available economic indicators, expert opinion, and field surveys.
Fig. 2: Actual crude real estate prices (light line), and housing price index (dark line).

Acknowledgments. We thank the ABMUS reviewers, and J. Bechara, B. Edmonds, S. Eid, M. Fawaz, L. Halabi, and A. Heppenstall. This work received support from the National Council for Scientific Research in Lebanon (CNRS) under grant number 103290, and the University Research Board of the American University of Beirut under grant number 103183.

References