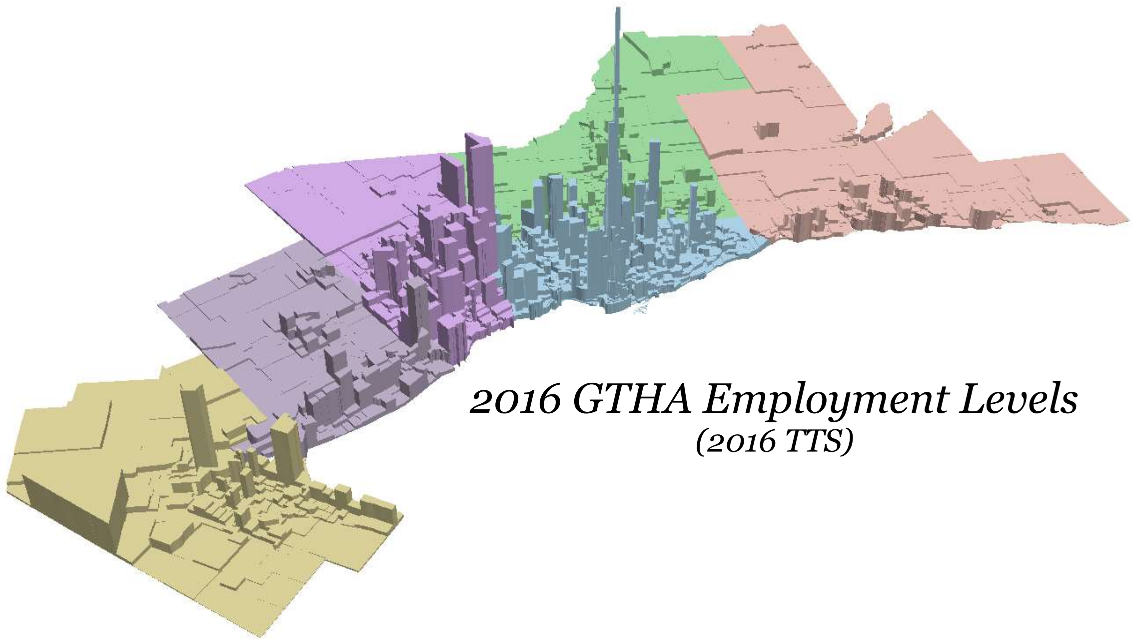

MODELLING DEMOGRAPHIC & SOCIO-ECONOMIC EVOLUTION: ISSUES, OPTIONS & PROPOSITIONS FOR MODEL IMPROVEMENT

A Discussion Paper prepared for Metrolinx.

By:
Eric J. Miller, Ph.D.
In collaboration with IBI Group

March, 2020. Revised August, 2020

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

UTTRI

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

This paper discusses a range of issues and options for modelling demographic and socio-economic evolution in the Greater Toronto-Hamilton Area (GTHA) and, by extension the Greater Golden Horseshoe (GGH). It is intended to serve as the basis for further discussion with Metrolinx concerning desirable "next steps" in improving the inputs into regional travel demand forecasting model systems for the GTHA and GGH.

Section 2 provides a motivating discussion of the importance of population and employment forecasts in the travel demand modelling process. It introduces and discusses a number of key issues associated with current methods and options for short- and longer-run improvements to these methods. In particular, three key forecasting problems are introduced: forecasting population demographics, forecasting employment, and forecasting the workers' commuting patterns in terms of their place of residence – place of work (PoRPoW) linkages. In all cases, these forecasts must be made at the spatial level of detail of individual traffic analysis zones (TAZs), with the attributes of persons, households and jobs defined that are required for input into the travel demand model system.

Section 2 also differentiates between two primary approaches to this forecasting problem. The first is the approach currently used by Metrolinx and other regional agencies. It involves directly forecasting a future year, end state distribution of the variables of interest. This is labelled a "static" forecast, since it is based on strong static equilibrium assumptions; that is, it assumes that the future conditions can be statistically forecasted as a path-independent state based on assumed probability distributions that characterize this end state. The second approach is labelled a "dynamic" forecast, since it generates a future year end state by explicitly evolving the system state over time (typically year by year) from a known base year, initial state. It thus treats future states as path-dependent outcomes of interacting events over time.

A basic proposition of this paper is that, in the short run, many opportunities exist to improve and elaborate current static methods. But, in parallel to these short-run improvements, R&D effort should also be put into developing more dynamic modelling methods, given their potential to eventually replace the static methods with significantly improved models. This parallel, "twin-track" approach is argued to represent the best approach to minimizing overall risk, to generate a steady stream of incremental improvements in the region's operational models, while at the same time developing more advanced methods that can eventually replace current methods once they are proven "to be ready for prime time".

Given the overview discussion in Section 2, Sections 3 and 4 address in greater detail issues and options for improving current static models for the three primary processes mentioned above: population and employment forecasting (Section 3) and modelling PoRPoW linkages (Section 4). Section 5 and the appendices discuss dynamic modelling issues and options. Section 6 of the paper summarizes key points raised in the paper and recommendations for "next steps".

2. MOTIVATION

2.1 Problem Definition

Regional travel demand forecasting model systems depend critically on future year forecasts of population and employment for every traffic zone in the model system's study region. These forecasts of total zonal population and employment are typically developed independently from the travel demand model system, and enter the system as exogenous inputs for a given model system run scenario. If the projected magnitudes of population and/or employment and/or their projected spatial distributions are significantly in error then the corresponding travel demand forecasts will also contain significant errors. Indeed, the single biggest source of error in many travel demand forecasts can often be attributed to errors in population and/or employment forecasts.[1]

This dependency on the population and employment inputs becomes even greater in modern, best practice model systems, such as both the Greater Golden Horseshoe Model V4 (GGHM4)[2] and GTAModel V4[3], the Greater Toronto-Hamilton Area's (GTHA) two primary travel demand model systems.[4] These not only require population and employment totals by zone, but also require these totals to be disaggregated by a variety of attributes. In the case of population, these can include both person-level attributes (age, occupation, employment status, possibly gender, personal income, possession of a driver's license and/or transit pass, etc.) and household-level attributes (allocation of persons to households, number of household personal-use vehicles, household income, etc.). Note that these attributes can be classified as falling into three broad categories: *demographic* (age, gender), *socio-economic* (employment status (full- or part-time), occupation, income) and *mobility "tools"* (driver's license, transit pass, household personal-use vehicles).

In the case of employment, jobs similarly need to be categorized consistently with the attributes used to characterize workers (e.g., occupation, employment status and possibly income). Note that, unlike persons, this does not necessarily mean that individual jobs need to be instantiated as separate objects within the model system data structure; it is often sufficient to simply generate the number of jobs by job category (e.g., number of jobs in a given occupation – employment status category).

Note that within this paper, *population* is consistently used to refer to aggregate totals (by zone, etc.). *Persons* refer to individual people within the aggregate population; these individuals possess attributes (age, etc.). Every person resides within a unique *household*. Individual workers

[1] This is not to say that travel demand forecasts can not be of use in the face of errors in the population and employment forecasts, but, clearly, the better these forecasts are, the better the starting point for the travel demand forecast.

[2] Note that, as the name implies, GGHM4 models the entire GGH. For simplicity of discussion, this proposal focusses on the GTHA as the study area, but issues apply equally to modelling the full GGH.

[3] Travel Modelling Group (2015) *GTAModel V4.0*, Toronto: University of Toronto Transportation Research Institute, Toronto, ON, Canada.

[4] MTO and Metrolinx both use GGHM4, while the Region of York is developing a somewhat similar model system based on CT-RAMP. The Cities of Toronto, Mississauga, Brampton and Vaughan and the Regions of Durham, Halton and Peel are either using GTAModel V4 or in the process of adopting it at the time of this paper's preparation.

aggregate up to the *employed labour force* (ELF) within a zone or region. Similarly, *employment* (EMP) consistently refers to aggregate totals, which are the sum of individual *jobs*, which possess attributes (occupation class, etc.).

Some of these attributes in some cases might be determined as part of the exogenous process used to generate the zonal totals.[5] More typically, these attributes need to be *synthesized* as a first step within the overall travel demand model system. For example, both GGHM4 and GTAModel model systems can be characterized at a very high level as consisting of the same overall structure shown in Figure 1. In this high-level representation of model system architecture, the population and employment synthesis tasks occur within the first level of the model system, which in Figure 1 is labelled "Agent Synthesis", thereby generalizing the concept of the synthesis task from just the task of synthesizing persons (the focus of much of the literature and modelling effort) to include job synthesis and (as discussed immediately below) the relationships among agents that are required as inputs to the actual daily activity/travel model that lies at the core of the overall model system.

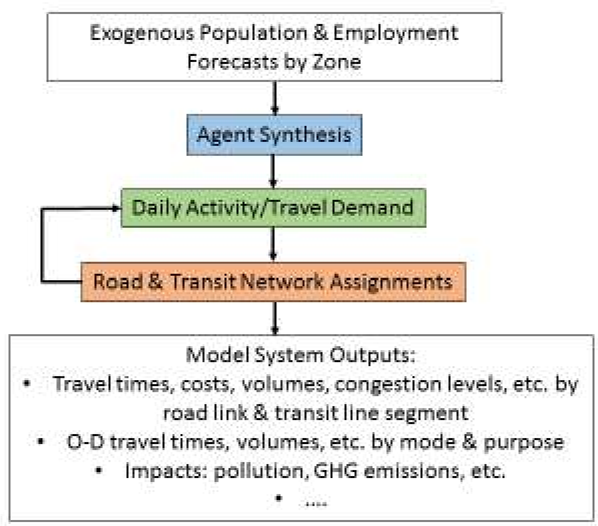

Figure 1: High-Level View of GTHA Travel Demand Model System's Architecture

In addition to synthesizing person and job attributes, the "upper level" of both GGHM4 and GTAModel also model the "long-term"[6] *linkages* between where workers live (place of residence) and work (place of work)[7]. That is, it is assumed that these Place of Residence – Place

[5] The issue of determining aggregate population and employment totals by zone (and, possibly, some population and/or employment attributes) is returned to below.
[6] As opposed to the "short-run", activity/travel patterns generated for a typical fall weekday by the main travel demand component of the model system.
[7] Not all workers have a typical (fixed) place of work; other workers may work at home (WAH) rather than at an out-of-home location. Both of these cases need to be dealt with at this point in the model system as well.

of Work (PoRPoW) linkages are the outcome of longer-run labour market participation processes that are best modelled as occurring outside of the activity/travel model *per se*, which deals with forecasting the trip-making behaviour of the GTHA population for a "typical" fall weekday, given the previously determined PoRPoW linkages (i.e., given that each worker's workplace is known). Note that a similar need exists to determine students' Place of Residence – Place of School (PoRPoS) linkages, which then similarly condition student trip-making behaviour within the activity/travel demand component of the model system.

Thus, the Agent Synthesis box in Figure 1 can be expanded, as illustrated in Figure 2, to show the key modelling tasks needed to generate the complete set of person, household and job attributes and relationships required by the activity/travel demand core of the model system. For the sake of illustration, Figure 2 depicts the current GTAModel V4.1 structure for these model components, but GGHMv4 undertakes similar tasks in a roughly similar manner.

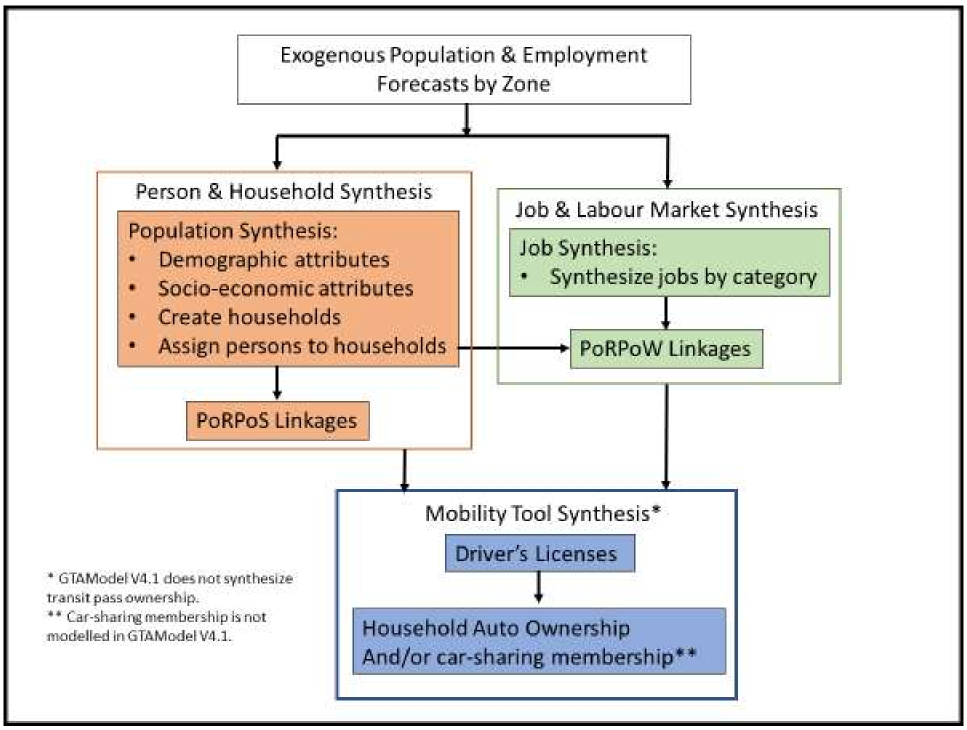

Figure 2: Typical Agent Synthesis Model Components

2.2 Issues in Agent Synthesis

Among undoubtedly others, five major points can be noted concerning the agent synthesis process as depicted in Figure 2. These are discussed in the following sub-sections.

2.2.1 Static versus Dynamic Modelling

The agent synthesis problem as presented in Section 2.1 is posed in a very static manner. That is, a future-year, "end-state" list of agents is generated "from scratch" from a projected total population of agents that have been generated by some other procedure for the forecast year. Many methods for doing this synthesis exist, but they all involve making assumptions concerning the joint correlation structure among the population attributes being synthesized (such as correlations among person age, employment status and income, as one possible example).[8]

This standard approach is a very practical one since it is consistent with the static, end-state nature of the travel demand models that it is designed to serve, as well as requires relatively limited data and computing effort to implement. It does not, of course, correspond to the actual process by which the attributes of a future population are determined. The world actually moves into the future in a continuously dynamic, evolutionary way. People are born, age and eventually die. People move into and out of the urban area. Households evolve over time as a result of these processes, but also through marriage, divorce, non-family household formation/dissolution processes, adult children moving out to form their own households, etc. The future state is clearly actually a path-dependent one, which depends both on the initial base year conditions, but also on the dynamics of demographic and socio-economic processes along the way.

Thus, an alternative approach to generating future-year population and employment disaggregated by their attributes, is to start with a base year disaggregate population (which may well be generated by a static synthesis procedure applied to available base year aggregate data), and/or employment and then dynamically evolve this population/employment, year-by-year, from the base year to the desired forecast year. Such a dynamic approach is currently generally not adopted in operational practice, except in a few advanced integrated transport – land-use model systems which may dynamically evolve population. Numerous reasons for this current state exist. The approach involves additional modelling complexity, data requirements and computational burden relative to simpler static methods. Concerns may also exist with respect to stability of the method's outputs – will forecast error bands actually grow over time resulting in greater uncertainty concerning future end states than in a static case? In particular, will a dynamic forecast be consistent with macro expectations concerning population and employment totals, which most planning agencies typically wish to impose on the forecasting exercise? These and other issues are discussed further in Section 5 and Appendix II.

The same comments hold for the PoRPoW model. Both GGHM4 and GTAModel have static gravity/entropy type models that compute probabilities of workers living in zone i working in zone j. But, as with demographics, the PoRPoW linkages that exist at any point in time actually evolve one worker and one job at a time as people enter and leave the labour force, are hired and fired, and as jobs are created and eliminated.[9] Thus, the alternative to the current static approach would be a dynamic evolution of PoRPoW linkages over time. This is discussed further in Section 4 and Appendix I.

2.2.2 Consistency between Aggregate & Disaggregate Methods

A potential issue exists concerning linkages between how population totals by traffic zone are generated versus how individual persons and households are then synthesized. To the extent that some population attributes may be used/assumed in estimating the zone totals, the question arises as to the internal/logical consistency of the two procedures. For example, it may be that a cohort-

---

[8] This paper does not review specific methods for <u>population</u> synthesis. Detailed reviews of population synthesis methods can be found in (Hammadi, 2020; Müller and Axhausen, 2010). Job synthesis is discussed in Section 3.

[9] Arguably PoRPoS linkages are perhaps less sensitive to such evolutionary dynamics, and, hence, can be more readily treated in a static fashion, as is currently done. Hence, this paper focuses on the PoRPoW problem.

survival method is used to evolve the population totals, which generates not just total population, but also population by age (and possibly gender as well). This age distribution, however, is typically "thrown away" when just the population totals are passed to the population synthesizer, which then goes through a new (and possibly quite different) process for attaching ages (and possibly gender) to the agents being synthesized. A similar comment holds for employment, which may be generated on the basis of projected floorspace, economic growth, etc., which may well have implications for job occupation type, income, etc. But, again, this information may well be "lost in translation", with only the resulting total employment being passed to the travel demand model, which will impose its own assumptions about employment type distributions to generate disaggregated jobs.

2.2.3 In- & Out-Migration

Hidden in Figure 2 and the discussion to this point is the issue of in- and out-migration. In-migration, in particular, is a major challenge to deal with in the GTHA, given the massive numbers of persons and households that move into the region each year. In-migrants, in general, can be assumed to have somewhat different demographics than residents, as well as different labour force participation characteristics. But if in-migrants are not identifiable within the population, then the synthesis and PoRPoW models cannot take these differences into consideration in their calculations.[10]

2.2.4 Modelling Employment & Employed Labour Force

In general, the modelling of employment and employed labour force is arguably less well developed than the modelling of population demographics. Issues include how to:

- Ensure consistency in how ELF and EMP are forecasted, with often different procedures and assumptions being used for the two variables. ELF and EMP, however, are clearly causally linked, since one can't have workers without jobs and vice versa.[11]
- Categorize job/worker types, given the enormous heterogeneity of job types in a modern economy and the diversity of categorization schemes used in various employment-related datasets. Differences in census and TTS representations are particularly important to resolve.
- Incorporate income within our models (personal versus household incomes; whether to attach income to the worker or the job or both).
- Model PoRPoW linkages (discussed in detail in Section 4).
- Deal with the impact of continuing evolution of the economy, technology, etc. on employment numbers, locations and types.
- Model changes in time in the propensity to work at home (WAH), rather than at an out-of-home location.

---

[10] While the Census distinguishes between newcomers and more settled residents, the TTS does not, which makes an investigation into the differences very challenging. Indeed, as a more general issue, Canada has few if any true panel studies that would allow researchers to directly study household mobility (to derive a series of bottom-up choice models) while the US has the Survey of Income and Program Participation (SIPP) and the Panel Study of Income Dynamics (PSID).

[11] Ignoring job vacancies and unemployed workers. Current models always ignore these and assume equality between workers and jobs.

- Deal with jobs (such as construction workers, many truck drivers, etc.) that do not have fixed workplaces.[12]

Also, as with demographics, employment forecasting can, in principle, be undertaken as a static, "leap ahead" to a forecast year (the typical current case) or by simulating changes in employment, labour force and PoRPoW linkages over time on a year-by-year basis.

2.2.5 Modelling Housing Markets & Land Use Evolution

Finally, any discussion of population and/or employment dynamics inevitably raises the question of residential location dynamics as well; i.e., the demographics of a given zone will change over time not just due to "natural" demographic evolution (births, deaths, etc.), but by households moving into or out of the zone. Thus, housing market dynamics also ideally should factor into forecasting future year population and employment distributions. Packaging demographic, housing and "firmographic" (employment) dynamics together leads into the field of integrated transportation – land use models, or, more generally, integrated urban models (such as sketched in Figure 3), which, it can be argued, are the theoretically preferred approach to modelling future urban system states (Miller, 2009).

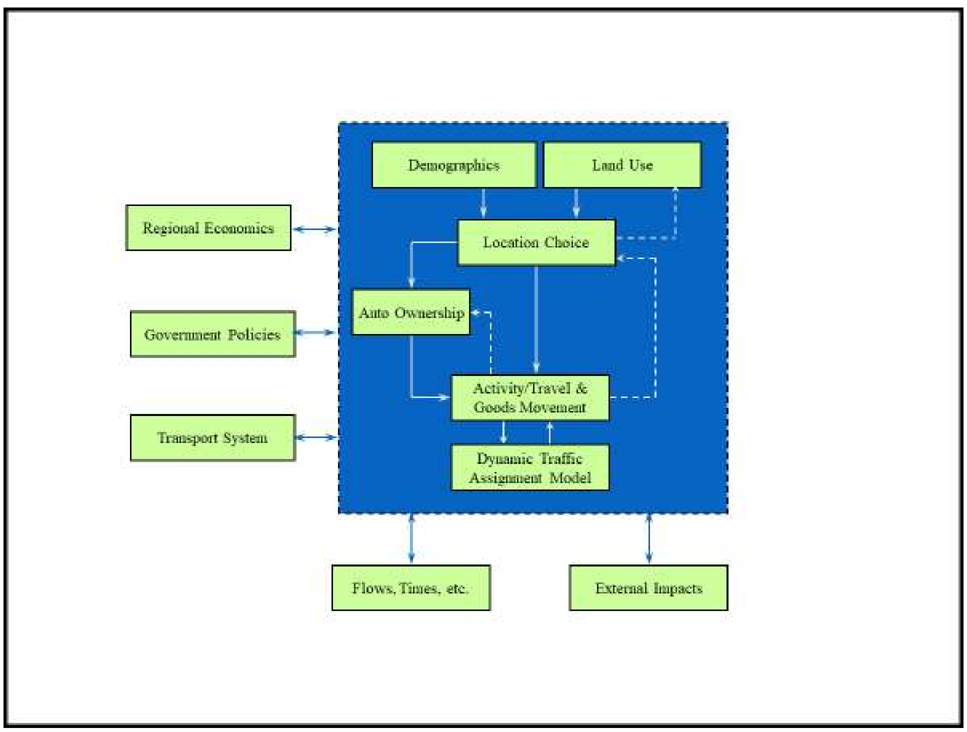

Figure 3: Abstract Representation of an Integrated Urban Model System

2.3 Paper Outline

Given this overview discussion the remainder of this paper investigates in greater detail many of the issues raised above. Section 3 explores issues, needs and options with respect to improved population and employment forecasting. Section 4 discusses PoRPoW modelling. Section 5 investigates approaches for moving beyond static to dynamic forecasting of demographic and

[12] The emergence of the "gig economy" clearly exacerbates this problem.

socio-economic evolution, including a very brief discussion of integrated urban modelling issues and options. Section 6 then concludes the paper with a brief summary and some suggested next steps.

3. POPULATION & EMPLOYMENT FORECASTING

3.1 Introduction

This section explores in greater detail population and employment forecasting issues and options for improvement. Section 3.2 describes the current method used to support the GGHM4 model system. Building on this discussion Sections 3.3 recommends steps for improving the current procedure with a more integrated approach to the problem. Section 3.4 then presents additional discussion of the employment forecasting problem and suggestions for developing improved methods.

3.2 Current GGHM4 Method

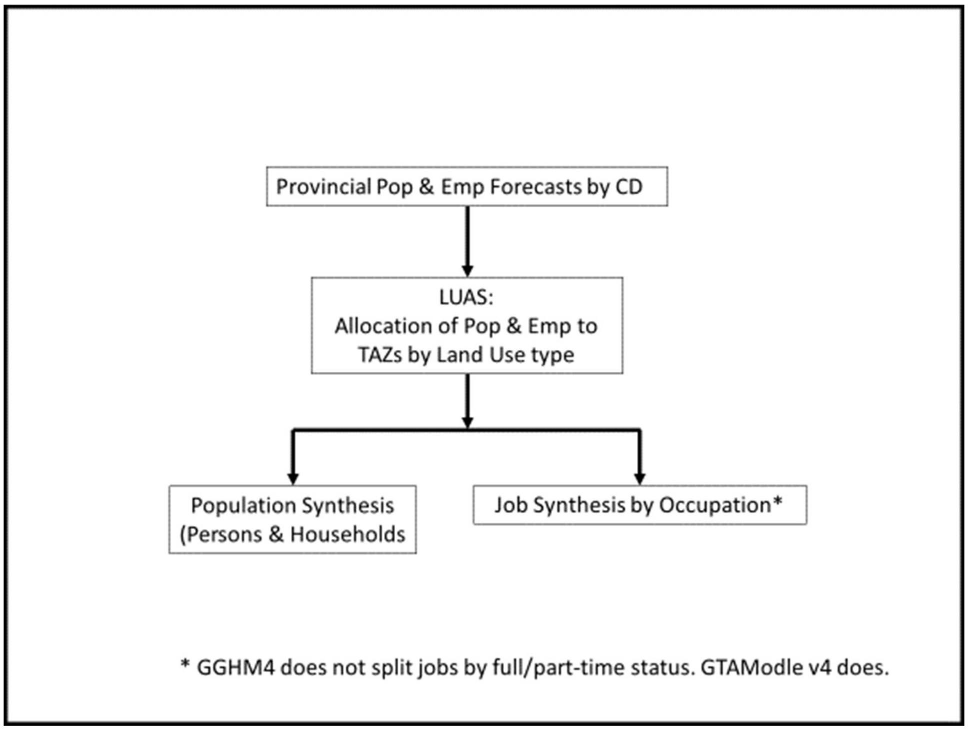

Figure 4: GGHM4 Employment Input Forecasting Process

Figure 4 describes the current process for population and employment forecasting used by MTO and Metrolinx for generating inputs for GGHM4, which is typical of current practice. It consists of four steps:

- The Province of Ontario produces forecasts of population and employment by Census Division (CD) based on provincial Growth Plan policy and projections.
- A "Land Use Allocation System" (LUAS) is used to allocate the CD population and employment totals by major land use types, first to the CSD level and subsequently to GGHM4 traffic analysis zones (TAZs) (IBI, 2017)

- A formal population synthesis procedure is used to synthesize individual persons and households from TAZ population totals. A commercial software, PopSyn3, has been used for this purpose by both GGHM4 and GTAModel, but at time of writing this paper, a new, more advanced population synthesis procedure is under development by Metrolinx that, it is hoped, will replace PopSyn3 in both model systems.

Total employment in each traffic zone is disaggregated into categories defined by occupation group (and, in the case of GTAModel V4. employment status -- full- and part-time). Base-year observed frequency distributions are used for this purpose, representing a very simplified "synthesis" process. Note that individual jobs are not instantiated, rather just the total number of jobs in each category are generated. This procedure is being updated by Metrolinx at the time of writing this paper. The updated procedure will generate TAZ-level employment distributions for input into both GGHM4 and GTAModel. (which use different occupation categorizations).

3.2.1 The LUAS Procedure

The LUAS procedure is noteworthy in that it takes into consideration several built land use and transportation accessibility factors in the allocation of people and jobs to traffic zones. These factors include:

- Provincial Growth Plan designations.
- Existing land use.
- Market forces, as defined by the professional judgement of real estate development experts.
- Accessibility to: major transit stops and stations; regional commuter rail stations; freeway ramps; and freight railway lines.

A weighted score combining these factors is constructed for each traffic zone and total CD population and employment are allocated to individual CSDs and, subsequently, traffic zones within each CSD based on their scores relative the total scores of all zones in the CSD (i.e., the relative attractiveness of each zone). Population is allocated to two land-use types: "ground-related" and "apartment" population. Employment is allocated to four land-use types: "major office", "employment land", "population-related" and "rural employment.

Four key issues can be identified with the structure of the current procedure:

1. Lack of influence over the CD control totals. These obviously drive the overall modelling process, but are generally not influenced by transportation factors in any way. This means that these totals may be inconsistent with the capability of actual future transportation system options to support them. It also means that major changes in the transportation system do not result in changes in these forecasts. It is also often mandated, as a matter of government policy, to test alternative scenarios for these CD totals.[13]
2. The occasional dependency on outside experts to generate the input to the LUAS model. While professional judgement is always a useful input into planning analysis, this dependency makes the overall process slow and cumbersome to use, particularly when testing scenarios that depart significantly from baseline forecasts.

---

[13] It is well-known, for example, that the City of Toronto fundamentally disagrees with Growth Plan projections for employment within the City and uses their own projections in much of their transportation modelling work. This leads to inconsistencies between City and Provincial travel demand forecasts that cannot always be resolved.

3. There is no connection between LUAS and either of the population or employment synthesis procedures.
4. In particular, the land use categorization of employment, while sensible in terms of the LUAS procedure itself, requires the ability to map this categorization into the occupation and employment status categorization required by the travel demand models.

3.3 Integrating LUAS & Agent Synthesis

The current two-stage process of generating population and employment totals by TAZ in LUAS and then synthesizing individual agents (persons, households and jobs) in a second step represents an arbitrary modelling structure that could be replaced by a single, integrated procedure that both allocates CD totals to TAZs and synthesizes the agents and their attributes located within each TAZ. Advantages of such an integrated procedure include:

- Taking advantage of the full set of explanatory variables available within the two current procedures. Currently variables used in LUAS play no role in the agent synthesis process, but might well be useful in improving this process. Similarly, information concerning agent attribute distributions might help improve the spatial allocation process. For example, information concerning household sizes might help inform population allocation to zones that is more consistent with both the zones' built form and demographics.
- Creating a smoother, more efficient and automated workflow that would reduce the potential for errors in the calculations and the burden on modellers, as well make it easier to produce alternative population and employment scenarios for use in modelling.

A key question in upgrading the current LUAS procedure into an integrated process is the extent to which the expert opinions of developers used within the current procedure can be replaced by a more systematic, quantitative model of "development potential". Some literature exists on the development of "development potential scores", although use of such models appears to be limited in practice. An important first step towards the development of an improved procedure should be to review the state of the art in this area and develop a model for GTHA/GGH application and test it against the professional judgements currently used. This ideally should involve the use of time-series land use (and other relevant data) both to identify key, measurable factors that influence land development and to test the predictive capabilities of these measures over time. It should also use current professional judgement scores to help "train" the model.

Additional tasks in developing an integrated procedure would involve:

- Designing the integrated model structure. This might still involve a "bi-level" structure in which population and employment are allocated to zones in one step and synthesis occurs in the next "lower-level" steps, but with all calculations occurring within an integrated software implementation. Or a single-step procedure may be conceivable.
- Operationalizing the new procedure within an efficient software implementation. Ideally, this should be an extension of the population synthesis software currently under development.
- Testing the new procedure against the current procedure to assess the extent to which it provides improved results.

3.4 Employment Forecasting

Despite its critical importance within the urban travel demand modelling process, employment forecasting and job synthesis are generally not as well developed in operational modelling applications as their population counterparts. Several reasons for this exist, including:[14]

- The enormous heterogeneity that exists within a modern urban economy. Categorizing employment into a manageable number of categories (inevitably a prerequisite for any practical modelling) is a non-trivial task, and will still result in considerable heterogeneity within each category.
- The need to deal with both industry type (which tends to determine the locations of business establishments[15] and their employees) on the one hand, and occupation type (which is needed to characterize workers' travel and labour force participation behaviour) on the other adds complexity to the problem.
- Business establishment / employment locations are tied very directly to land use / built form decisions about where to locate office buildings, retail floorspace, factories, warehouses, etc. These built form development decisions may be made by individual firms (especially large ones), but more generally are made by the development industry, which responds to its own imperatives, as well as in many cases by government. Thus, employment location is generally the outcome of two processes: (1) very long-term decisions concerning land development, and (2) decisions by firms concerning where to locate, given the built form (and available space) at the time these decisions are being made.
- An urban economy is a highly dynamic system. Industries and firms are constantly evolving. In particular, the occupational distribution within the economy changes over time, as does the balance between full- and part-time employment, with both these job/worker attributes being very important in forecasting PoRPoW linkages and work trip commuting behaviour.
- The propensity for employees to work at home and/or not have a fixed place of work changes over time. Thus, establishing what jobs are still located at a fixed, traditional workplace (office, store, etc.) and which are not needs to be addressed.
- Aggregate employment levels are typically forecasted by economists or urban planners who are not directly involved in urban travel demand modelling, often for other purposes. As a result, these forecasts are not sensitive to transportation accessibility or other "feedback effects", may be based on undocumented assumptions, and may not otherwise fully meet travel demand modelling needs. This also means that travel demand forecasters are often dependent on other groups (other government departments/agencies and/or private consultants) to produce their employment inputs, with, as just noted, little or no control over the methodology and assumptions used.

Other potential issues with current procedures include:

---

[14] Equivalent issues of heterogeneity, dependence on land development processes, etc. also exist in forecasting population spatial distributions over time, and these are briefly raised in Section 5. But we generally have much more experience in dealing with these issues when modelling population relative to employment modelling.

[15] A *business establishment* is a physical location at which the production of goods and services occurs and where jobs are housed. A *firm* is an economic agent that produces goods and services. Each firm will operate within one or more business establishments ("virtual", "numbered companies" aside).

- Household income is currently determined within the population synthesis process. No connection with workers' jobs exists, and no representation of job salaries/wages[16] exist at the employment end. Household income, however, derives directly from its workers' wages.[17] Thus, one could imagine a model structure in which household income derives from a labour market process (aka PoRPoW), as discussed further in Section 4. Forecasting future year income is, of course, a major task in itself. Any assumptions made within the travel demand modelling process should be tied to Ministry of Finance data/methods, given that they are presumably the experts in this area.
- Similarly, the number of workers by occupation and employment status by residential zone (ELF) is determined within population synthesis, while the number of jobs by occupation and employment status by workplace (EMP) is independently generated within the job "synthesis" process. There is no guarantee of consistency between these two processes. In current travel demand models, this inevitable inconsistency is rectified through ad hoc "balancing" procedures that scale worker and/or job totals to be equal.
- People who work at home (WAH), either full-time or part-time, are also generated as part of the population synthesis process. It is not clear the extent to which this synthesis is consistent with the accounting for overall employment, and the allocation of this employment to traffic zones.
- The probabilities used to split workers and jobs by occupation, employment status, WAH, etc. are typically derived from base year observed rates. Even in a static forecasting system such as represented by Figures 1, 2 and 4, these probabilities should reflect as best as possible projections of how categorization rates are likely to change from the base year to the forecast year. For example, how will full-time vs part-time employment evolve? How will the propensity to work at home change?[18] What will the shifts in occupation distributions be?

Given these considerations, suggestions for moving towards improved employment forecasting for travel demand modelling purposes include the following.

1. As discussed in Section 3.2.2, develop a single, integrated procedure that combines the current LUAS and "job synthesis" procedure into a single model.
2. Explore in detail trends in GTHA/GGH ELF and EMP by occupation, employment status, income and WAH propensity. Trends in the number of workers per household would also be useful to explore. Such an analysis could provide a much sounder basis for projecting the future year joint distributions of these key attributes for both worker and job synthesis.
3. Methods for the joint synthesis of workers (and their attributes) and jobs (and their attributes) should be investigated. Such a joint method will greatly enhance the internal/logical consistency of the overall model system and would, almost certainly improve the accuracy/robustness of the forecasts. It would, however, probably require a

[16] In this paper we treat "wages" and "salaries" as synonyms.
[17] Non-employment income from investments, etc. aside, which we ignore for the current discussion, but can be very significant in certain household types.
[18] It is very interesting to speculate how the current mandated working at home for a vast segment of the population that is in force at the time of writing of this paper will affect long-term behaviour once the restrictions are eventually removed.

rethinking of the person/household synthesis process versus how these persons map int workers and job.

4. As noted above, income is treated very simplistically and incompletely in the current procedure. Income, however, is extremely critical to both the determination of PoRPoW linkages (workers look for high-paying jobs; where workers and their households live depends on where they can afford to live) and travel behaviour. As is discussed further in the next section, much more attention should be paid to how income is represented and modelled within our model systems.
5. Given the importance of employment forecasts to travel demand analysis, it would be useful in developing the integrated procedure described above that has a user interface that allows alternative scenarios to be readily and robustly generated by changing basic inputs (CSD-level totals), occupation distributions, WAH rates, etc.[19]

4. MODELLING PORPOW LINKAGES

4.1 Model Specification

The GTAModel V4.1 PoRPoW model is representative of current GTHA modelling practice.[20] For a given occupation group and employment status (full- or part-time) category,[21] a doubly-constrained gravity/entropy model is used to predict the probability of a worker living in residential zone i being employed in employment zone j. The model is defined by the following equations.

$$T_{ij} = \frac{ELF_i B_j EMP_j e^{\alpha f_{ij}}}{\sum_{j'}^{N} B_{j'} EMP_{j'} e^{\alpha f_{ij'}}} \qquad [1]$$

$$B_j = 1/\sum_{i'}^{N} A_{i'} ELF_{i'} e^{\alpha f_{ij'}} \qquad [2]$$

$$A_i = 1/\sum_{j'}^{N} B_{j'} EMP_{j'} e^{\alpha f_{ij'}} \qquad [3]$$

$$f_{ij} = log \sum_m e^{\beta^T X_{ijm}} \qquad [4]$$

$$P_{j|i} = \frac{T_{ij}}{ELF_i} = \frac{B_j EMP_j e^{\alpha f_{ij}}}{\sum_{j'}^{N} B_{j'} EMP_{j'} e^{\alpha f_{ij}}} \qquad [5]$$

Where:

ELF_i = Employed labour force (number of workers) living in zone i
EMP_j = Employment (number of jobs) located in zone j

---

[19] A similar comment holds for the population synthesis process.

[20] Note that the GGHM4 PoRPoW model details differ somewhat from the GTAModel model shown here for the sake of illustration.

[21] Subscripts denoting a worker's occupation and employment status are not included in these equations in order to keep the notation as simple as possible. The equations and discussion in this section, therefore, apply to a single occupation-employment status group. The same model structure would then be applied to all other such groups. Each group will generally have its own set of parameters.

N = Number of zones
f_{ij} = Impedance function for travel from zone i to zone j
β = Column vector of parameters
X_{ijm} = Column vector of explanatory variables characterizing the systematic utility of travel by mode m from zone i to zone j: travel times, etc. (based on morning peak-period travel conditions)
A_i , B_j = "Balancing factors" that ensure that equation [1] satisfies the "row and column" constraints:

$$\sum_j^N T_{ij} = ELF_i \quad \forall\, i = 1, N \qquad [6]$$

$$\sum_i^N T_{ij} = EMP_j \quad \forall\, j = 1, N \qquad [7]$$

$P_{j|i}$ = Conditional probability that a worker living in zone i works in zone j

Each worker in zone i is assigned a specific work zone based on a Monte Carlo draw for the probability distribution defined by equation [5].

4.2 Issues in PoRPoW Modelling

A number of comments can be made concerning the strengths and weaknesses of this modelling approach. These are discussed in the following sub-sections.

4.2.1 Doubly-Constrained Formulation

First, the doubly-constrained (DC) formulation is a very important feature of the model, compared to so-called "work location choice" (WLC) models that are typical of US practice. The DC approach insures that there is a one-to-one matching between workers and jobs, something that is not achieved in a singly-constrained WLC model.[22] That is, every worker is constrained to be allocated a job, but the number of workers assigned to jobs in any given zone can either exceed or be less than the number of jobs in the zone. This introduces an internal inconsistency within the WLC model that is troublesome and that has no compensating theoretical or practical advantage.

The WLC approach is also grounded on a persistent assumption among many modellers that its logit model formulation is somehow superior to the DC formulation. But, as Anas (1983) demonstrated long ago, "logit" and "entropy" models are mathematically identical if consistently specified. Further, the entropy formulation is based on Information Theory, which is just as robust and justifiable a starting point for developing this model as Random Utility Theory. That is, a properly specified entropy model is guaranteed to generate the statistically least-biased, most-likely estimate of PoRPoW linkage probabilities possible (Shannon, 1948; Wilson, 1970; Webber, 1977). Given the largely statistical (as opposed to behavioural) nature of these models, this is a very desirable property.

This is not to say that the Random Utility interpretation is not also useful. The nice implication of Anas' findings is that we can use both Random Utility and Information Theory to provide insights into the model specification, as appropriate.

---

[22] Note that no unemployed workers or vacant jobs exist in this model. Dealing with unemployment and vacant jobs is well beyond the scope of current PoRPoW modelling capabilities and would require the sort of dynamic labour market model that is described in Appendix I.

The B_j balancing factors in the DC model also have a very useful interpretation as measures of the competition among workers for jobs. If a large number of workers are attracted to a particular employment zone (e.g., a large zone that is close to large numbers of resident workers), then the likelihood of any one worker getting a job at this location will be reduced. This is generated in the model by such zones having balancing factor values less than 1.0 to reduce the allocation of workers to the zone relative to what an unconstrained assignment would generate. At the same time, in order for all workers to find a job in this competitive environment, less attractive zones will have balancing factors greater than 1.0 in value to "induce" workers to travel further to smaller employment zones. This role played by balancing factors in representing competitive effects (albeit in a simplistic and "reduced form"[23] manner) can perhaps be more easily seen by rewriting equation [1] as:

$$T_{ij} = ELF_i \left\{ \frac{e^{log(B_j)+\log(EMP_j)+\alpha f_{ij}}}{\sum_{j'}^{N} e^{log(B_{j'})+\log(EMP_{j'})+\alpha f_{ij'}}} \right\} \qquad [8]$$

In equation [8], if $B_j < 1.0$ then $log(B_j) < 0.0$, thereby reducing the "utility" of zone j for workers seeking employment in this zone, while if $B_j > 1.0$ then $log(B_j) > 0.0$, which increases the employment zone's attractiveness. These terms serve a purpose similar to alternative-specific constants (ASCs) in mode choice models (i.e., to make the model balance to observed control totals). But, unlike ASCs, these are not fixed parameters in the model, but dynamically adjust in application to reflect the relative competitiveness of employment zones in the forecast year.

Equation [8] also makes clear the correspondence of the DC with Random Utility formulation, since the expression in the brace brackets is easily recognizable as a standard multinomial logit model for the conditional probability $P_{j|i}$.

4.2.2 Static Equilibrium Modelling

A second important point is that the DC formulation can be interpreted as a static equilibrium model of the labour market, in which each worker has been allocated to a job so that the market is in a stable or equilibrium state. This can be thought of as a Nash Equilibrium since it is the result of a competitive (non-cooperative) "game" among workers competing for a fixed number and distribution of jobs. This static nature of the model needs to be acknowledged as a major assumption. The alternative to this would be to simulate the dynamic evolution of the labour market over time as workers enter and leave the labour force, jobs are created and terminated, etc. Section 5 below discusses the dynamic simulation approach further, and Appendix I presents a prototype of one such dynamic labour market model developed at the University of Toronto to illustrate both the potential but also the complexity of this approach.

---

[23] I.e., without explicitly modelling the competition among workers for specific jobs.

4.2.3 Impedance Function Specification

A third major point concerning this model is the use of an implicit logit mode choice model to define the impedance function, f_{ij}, as the logsum (expected utility) of this mode choice model.[24] This mode choice model, however, is not the mode choice model actually used to predict travel modes in the model system. This use of a simplified mode choice model to define the PoRPoW impedance function is employed in both GGHM4 and GTAModel, and is, in fact, common practice in many model systems. This approach can be justified on a number of grounds, including:

- The PoRPoW linkage problem is a longer-term one relative to the day-to-day choice of mode given fixed home and work locations, so it may well be the case that a somewhat different evaluation of modal accessibility enters into this decision.
- In particular, at the time of choosing a work location, the worker is likely to have more limited knowledge about modal alternatives and service levels than once s/he is actually travelling to the chosen workplace and experiencing the full range of alternatives and their characteristics on a daily basis.
- As a practical matter, the computational burden of evaluating the full mode choice model for every possible workplace for every worker would be very expensive indeed. This is particularly the case in a tour-based model system, in which the choice of mode for a given trip depends not just on that trip but, potentially, on all the other trips in the tour. This adds essentially an unmanageable amount of complexity to the problem.

Given both these "behavioural" and practical considerations, it is expected that future versions of this type of PoRPoW model will continue to use some variation on this approach.[25] Possible improvements in specification of the impedance function are always conceivable, but the basic approach is unlikely to change appreciably. Inclusion of worker-specific demographic and socio-economic attributes, for example, may well be well worthwhile to investigate in terms of improvement in model performance. This would create worker-specific impedance terms, and, hence worker-specific location choice probabilities. This would increase computational complexity and burden, but not necessarily excessively, but should be feasible to do given the overall microsimulation framework of the model system.

4.2.4 Auto Ownership & Workplace Location Choice

Fourth, both GGHM4 and GTAModel V4.1 allocate workers to workplaces prior to determining household auto ownership levels, which are subsequently determined given these assigned workplaces. GTAModel V4.0, on the other hand, determined household auto ownership levels first, and workplace choice probabilities where then conditional upon these known household auto ownership levels.[26] Arguments in favour of both model structures can be made. In V4.1, GTAModel adopted the GGHMV4 structure, partially to achieve greater consistency between the structure of the two model systems, but also to facilitate the development of an endogenous household auto ownership choice model, which is, in some ways at least, easier to do in this

[24] Note that this implicitly treats the PoRPoW linkage and work trip mode choice as a nested logit type problem.
[25] Indeed, even in a dynamic model, such as the one sketched in Appendix I, this use of a simplified (but multi-modal) impedance measure is likely to continue.
[26] This was done by adding a third dimension to the entropy model in which the number of workers residing in each zone were further divided into discrete "auto sufficiency" categories, and workers in each category has separate utility functions.

structure. While sticking with the current model workflow, it may be worthwhile to investigate means for introducing some "feedback" between auto ownership and workplace location, since auto ownership definitely changes workers' accessibilities to competing workplaces (the main argument for the V4.0 model structure).

4.2.5 "Moving Beyond Gravity": Model Specification

Fifth, regardless of whether one labels equation [1] (or [8]) an "entropy" model or a "random utility model", in terms of its specification, it is a pure "gravity" model in that work location probabilities are a function of employment location "size" (number of jobs) and "how far away" (impedance) they are from workers' residences. An important question is whether this model specification can be improved by introducing additional variables into the model.[27]

In the prototype labour market simulation model described in Appendix I a number of socio-economic variables are included, such as years of work experience, worker education level, etc. While such detailed variables might conceivably be feasible in a detailed simulation model, in which they would be generated as part of the system evolution (e.g., by tracking workers as they gain experience over time), in the static formulation under discussion in this section, it is not practical to consider the use of such variables, as useful as they might appear to be. Having already controlled for occupation and employment status by modelling each occupation-status category separately, the only demographic/socio-economic variables typically available within an operational model system such as GGHM4 and GTAModel are: worker age and gender, household attributes (size, number of children, number of workers)[28], and household income.

As discussed in Section 4.2.3, it may well be useful to include age and (possibly) gender in the impedance function specification, but it is not clear that adding these variables into the location choice "utility" function per se is likely to be useful. While younger people may be more or less competitive for certain jobs than workers in other age groups, for example, it is not clear that one can capture such detailed effects in a static, relatively aggregate[29] PoRPoW linkage model such as being discussed herein. Further age (and possible gender) effects are taken into consideration in determining workers' occupation and employment status allocations. In addition, further disaggregating current occupation-status categories into, say, age category sub-groups would generate a proliferation of models and parameters that is very unlikely to generate significantly improved overall model fit.

Similarly, it may be that factors such as household size, number of children, and/or the "status" of a worker within the household (i.e., is the worker the "primary" worker in the household or a "secondary" worker) may affect work location choice. There is, for example, some evidence that "secondary" workers tend to have shorter average commuting distances than "primary" workers. But it is not clear the extent to which this is due to the "secondary" status of the worker or the

---

[27] One might also ask whether the gravity/entropy/logit "classical spatial interaction" (SI) functional form is the best one available. Over the years other model formulations have been posited ("competing opportunities", "intervening opportunities", as well as, more recently coming out of the "big data / machine learning" world, "radiation" models, etc.). In general, such models have not been found to fit better than SI models, are often more difficult to calibrate, and, typically, lack theoretical foundation. As noted above, SI models have a very solid foundation in Information Theory that should not be discarded lightly, in addition to their ties to Random Utility Theory.
[28] Household auto ownership has already been discussed in Section 4.2.4.
[29] In the sense of a small number of heterogeneous occupation categories.

occupation type and/or employment status (full- or part-time) that "secondary" workers take on – both of which are accounted for in the current model structure.[30] Thus, again, it is not clear that, for example, further categorizing workers by such household level variables is an attractive proposition.[31]

Income is a potentially interesting variable, but it would, at a minimum, likely require restructuring current model systems. In economics-based labour market modelling it is a given that wages[32] are endogenous to the worker-job matching process. Workers in "high demand" occupations can command higher wages than those for whom there is little demand for their skills. It is often the case that workers needing to commute long distances (such as commuting into a city's central area) may need to be paid more to compensate them for these long commutes, relative to those who are able to work closer to home. It is also important to note that employment income[33] is fundamentally an attribute of the job, which a worker then "inherits" when s/he takes that job. It is arguable that job-based income could enter the PoRPoW model by modifying the attractiveness of job locations based on the wages offered at these locations.

A <u>very</u> simplistic illustration[34] of this is to modify equation [8] to read:

$$T_{ij} = ELF_i \left\{ \frac{e^{log(B_j)+\log(EMP_j)+\alpha f_{ij}+\gamma W_j}}{\sum_{j'}^{N} e^{log(B_{j'})+\log(EMP_{j'})+ \alpha f_{ij} +\gamma W_{j'}}} \right\} \qquad [9]$$

Where W_j is the average wage offered in zone j. This would increase the attractiveness of zones with higher wages. It would also increase the competition for these job among the workers. Household income would then be the sum of the incomes of the workers within each household.[35]

Another variable that might be considered for inclusion in the model is K-factors. Like mode choice alternative-specific constants, these terms capture systematic "biases/preferences" for particular residence-workplaces that are not captured by the other systematic variables in the model. K-factors have "a bad name" in spatial distribution models as "fudge factors", but judicious use of such terms is as justified as ASCs are in mode choice models. And given the paucity of explanatory variables typically available, their use in operational models is almost always necessary. Also like ASCs, however, their use assumes that the "biases" that they are capturing persist into the future. Given that urban structures (including PoRPoW linkages)

---

[30] Indeed, "primary" and "secondary" designations are often based on some combination of occupation, employment status and/or worker income level.

[31] The alternative to categorizing workers by either personal or household attributes is to include these attributes directly in the location "utility" function. This is a non-trivial thing to do in an "unlabelled" choice model such as this one, since such attributes need to enter in an alternative-specific way, since, for a given person and household these attributes do not vary across locations – a classic technical issue in discrete choice models.

[32] Herein we treat "wages" and "salaries" as being equivalent terms.

[33] As opposed to non-employment income from investments and other financial activities. The existence of non-employment income obviously complicates the analysis.

[34] More complicated (and probably more appropriate) formulations can be imagined.

[35] Again, how to deal with the question of non-employment income is not discussed herein, but is a potentially non-trivial issue, at least for "high-income" households, households containing retirees, etc.

typically change slowly over time, this is often not an indefensible assumption, but it is always one that should be carefully evaluated.

A final possibility for model specification is to identify key major employment zones, such as the Toronto Central Area and the Pearson Airport employment zone, to have different impedance function parameters relative to other, more minor employment areas. Support of this approach derives from Central Place Theory (CPT), which argue that a hierarchy of centres exists in an urban region, with "higher order" centres having much larger "ranges" of influence than "lower order" centres; i.e., their commuter sheds tend to be much larger (workers are willing to travel much farther to access jobs in the higher order centres). Conversely, smaller, more dispersed employment locations tend to draw their workers from more local commuter sheds. In various earlier versions of GTAModel, such spatial segmentation has often been found to improve model performance.

The issue of spatial segmentation may be particularly interesting to explore with respect to modelling PoRPoW linkages for workers and employment zones in portions of the GGH lying outside of the GTHA. Regions such as Niagara and Waterloo have local economies, with much of their labour force working within their home region. These regions, however neighbour GTHA municipalities and do have interactions with them (in both directions). It is possible that some form of "bilevel" model structure, which sorts out "local" versus "inter-regional" interactions might prove use to capture the extent of local "self-containment" versus longer-distance linkages.

4.3 Towards Improved PoRPoW Models: Possible Next Steps

Summarizing the Section 4.2 discussion, recommended avenues for possible short-run improvements to current PoRPoW models include investigations into:

- Improved impedance function specifications.
- Revisiting mechanisms for introducing auto ownership effects within PoRPoW calculations.
- How to incorporate wages/income into the model.
- Improved, systematic use of K-factors and/or spatial segmentation of parameters to improve model fit, including examination of GGH-level interactions and segmentations.

In addition, it would be well worthwhile to investigate the temporal trends in PoRPoW patterns over time in the region. TTS provides us with at least a 20-year[36] time-series database within which we can explore how commuting patterns have evolved as the region has grown significantly. A comprehensive, detailed examination of these trends has never been undertaken,[37] but it should provide significant insights for building more robust PoRPoW

[36] 1996-2016. 1991 is a much smaller sample and so is less reliable for comparative purposes. 1986 lacks complete information concerning some key variables – notably the place of work is not known for workers who did not travel to work on the survey day. It also becomes increasingly difficult to construct transit service levels as one moves back in time, since only AM-Peak networks are generally available for these very early years.

[37] A few, partial caveats to this very strong statement exist. Elmi, et al. (1999) investigated the temporal stability of work trip distribution models between 1964 (using MTARTS data) and 1986. Fox, et al. (2012) examined the temporal stability of nested logit models of the joint choice work location and mode choice models. In both cases, the emphasis was on testing the temporal transferability of model parameters rather than a more detailed investigation of commuting patterns per se.

models. Recent detailed exploration of work activity generation rates has identified remarkable stability in these rates over the 1996-2016 time period (Ozonder and Miller, 2020). It would be very illuminating to similarly explore how the structure of commuting has evolved within the region.

Finally, it is recommended that, in parallel to these short-run investigations/improvements, that R&D effort should be put into developing an evolutionary model of labour market dynamics, such as is exemplified by the prototype model presented in Appendix I.

5. TOWARDS DYNAMIC DEMOGRAPHIC & SOCIO-ECONOMIC FORECASTING

In the previous sections of this paper issues and options for improving the "static" forecasting of employment and PoRPoW linkages for a future year end state have been discussed. At several points it is observed that the alternative to this static approach is to dynamically evolve population and employment from a "known" base year over time, year-by-year, to generate the desired future year end state as the emergent outcome of a path-dependent set of processes. Potential advantages of such a dynamic, simulation process include:

- If conducted as a microsimulation, in which individual agents (persons, households, firms) are tracked over time, full heterogeneity and correlation in population[38] attributes are maintained over time.
- In particular, consistency between ELF and EMP is maintained as both co-evolve.
- Similarly, household income is generated consistently as the outcome of the household's workers engaging in a labour market process.
- Attributes of the residential population in each TAZ that are critical to forecasting travel demand change smoothly and consistently over time from base year conditions as a function of incremental changes in demographics (babies are born, people age, etc.), labour market participation (people enter/leave the labour market, change jobs, etc.) and residential mobility (households move from zone-to-zone; in- and out-migration).
- Data for intermediate years are simulated naturally and incrementally as part of the evolution of the system state to the final forecast year, thereby enabling the consistent analysis of travel demand and transportation system performance over time, and not just for the future end state.
- Multiple evolutionary paths and associated future end states can be readily generated to provide a range of possible futures in order to test the robustness of alternative transportation policies and their benefits and costs across scenarios and assumptions, rather than "locking in" policy analysis and decision-making on a single (and perhaps rather arbitrary) "point estimate" of the future. These alternative paths/states can be generated either through different assumptions concerning key inputs (birth rates, in-migration rates, economic growth rates, etc.), or generating multiple runs (replications) for the same set of inputs (through the use of different random number streams in the stochastic components of the model), or (generally speaking preferably) both.

[38] Here, "population" is being used in the most general sense to refer to whatever is being simulated – persons, jobs, PoRPoW linkages, etc.

Potential disadvantages associated with a dynamic, microsimulation approach, on the other hand, include (Lee, 1973, 1994; Timmermans, 2003):

- Model complexity.
- Data requirements.
- Computational burden.
- Technical capabilities of planning agencies to develop, maintain and use complex models.

These are very important, non-trivial considerations, especially for operational planning agencies with constrained resources and having good reasons to be relatively risk-averse in terms of adopting "excessively cutting-edge" methods. But, as Miller (2009) argues, the current state of data availability, computing power, modelling methodology, base theory and experience with large-scale (i.e., region-wide) microsimulation modelling all combine to make large-scale urban evolutionary simulation modelling a practical possibility.

The discussion of evolutionary urban simulation models is usually tied to the case of large-scale "land use" model systems that attempt to provide a comprehensive solution to modelling the full range of urban spatial dynamics, as sketched in Figure 3 above. Two or the most common examples of such model systems in operational use are UrbanSim[39] and PECAS[40]. A full review of these model systems is beyond the scope of this paper, but a few key observations relevant to the current discussion include:

- Neither of these systems (or other typical "land use") model systems) are true demographic simulators. They tend to work primarily with households, rather than persons, and do little, if any, demographic updating. Households, for example, may be characterized by income class and, perhaps, number of workers/persons.
- Similarly, these model systems generally allocate employment (rather than firms or business establishments) to parcels/zones, based on the building supply located at these points in space.
- The model systems focus to a large extent on housing market demand/supply and, to a somewhat lesser extent on employment location processes. PoRPoW linkages are not dealt with at all within UrbanSim. PECAS has a much more formal representation of the urban economy, both in terms of inter/intra-industry economic interactions and industry-worker labour markets.
- These systems are largely quite monolithic, requiring implementation of the entire package, as well as not being very flexible in terms of modifications or extensions to the core software.

Four primary processes are under discussion in this paper:

- Demographic evolution of the population.
- Firmographic evolution of employment.
- Evolution of the labour market (PoRPoW linkages).
- Housing market dynamics (residential location/relocation of households).

[39] https://urbansim.com/
[40] http://hbaspecto.com/products/pecas/

As noted above, these four processes are typically bundled together (to a greater or lesser extent) within a large, monolithic "land use" model. While obvious interconnections and interdependencies exist among these four processes, it is also the case that each can be "disentangled" to a fair degree from the others and modelled as a separate process. Such a modular approach has several potential advantages, including:

- It reduces risk, since it does not require investing significant resources into a single, very large effort that will succeed or fail as a whole. Instead, individual, incremental improvements can be made and built upon over time.
- It enables prioritization of effort to first deal with the element or elements that are deemed to be most critical to improving the overall modelling process.
- At the same time, it permits parallel development efforts to be undertaken, if desired, each one of which requires less resources, involves less risk and should be able, by focussing effort on a single objective (rather than multiple objectives, all of which need to be met for a successful conclusion to be achieved) to be achieved more quickly.

Of the four processes listed above, firmographic modelling is the least developed. Examples include (Maoh, 2005; Moeckel, 2005; Elgar, 2007; Farooq, et al., 2013; Mostafa, 2017), all of which are essentially research efforts and/or partial in their treatment of the full range of firmographic processes. Given this, it is likely that employment forecasting (at, say, the CSD level and above) will remain the domain of economic forecasters for at least the short- and medium term. Travel demand modellers, however, should work with these forecasters to ensure that the forecasts:

- Are generated on a year-by-year basis to support dynamic modelling of labour markets and travel demand.
- Incorporate and "make visible" to users of the forecasts whatever disaggregation of employment is possible with respect to industry, occupation, employment status and income that is possible to achieve. That is, to the extent possible, CSD-level "control totals" for employment categorized by one or more of these key attributes would greatly assist in improving the consistency and robustness of subsequent allocations of employment to TAZs (as described in Section 3).

Microsimulation models of the other three processes (demography, labour markets and housing markets) are much better developed. Although not yet typically implemented operationally within planning agencies, the potential for doing so is significant, and the associated risks are not excessive. As a specific (but not unique) example, work over the past 15 years at the University of Toronto has developed "operational prototype" dynamic, microsimulation models of the labour market, demographic evolution and the housing market for the GTHA. These prototype models are described in detail in Appendices I, II and III, respectively. Similar to the way in which UofT research on the agent- and activity-based model activity/travel model TASHA (Travel/Activity Scheduler of Household Agents, Miller and Roorda, 2003) has led to the successful operational implementation of the GTAModel V4 travel demand model system by many GTHA agencies, it would be possible with a reasonable R&D effort to implement one or more of these evolutionary systems for operational use by GTHA/GGH agencies.

6. SUMMARY & POSSIBLE NEXT STEPS

This paper discusses a range of issues and options for modelling demographic and socio-economic evolution in the Greater Toronto-Hamilton Area (GTHA) and, by extension the Greater Golden Horseshoe (GGH). This discussion provides a basis for further discussion with Metrolinx concerning desirable "next steps" in improving the inputs into regional travel demand forecasting model systems for the GTHA and GGH. These include:

Develop an Integrated Zone Allocation & Agent Synthesis Procedure

1. Develop a single, integrated procedure that combines the current LUAS procedure for allocating population and employment to TAZs and agent (persons, households and jobs) synthesis procedure into a single procedure. This should include exploring ways in which improved model specification and structure can be achieved.
2. Explore the extent to which the expert opinions of developers can be replaced by a more systematic, quantitative model of "development potential". The state of the art in this area should be reviewed as the starting point to develop and test a model for GTHA/GGH application. This ideally should involve the use of time-series land use (and other relevant data) both to identify key, measurable factors that influence land development and to test the predictive capabilities of these measures over time.

Employment Forecasting Improvements:

1. Explore in detail trends in GTHA/GGH ELF and EMP by occupation, employment status, income and WAH propensity. Trends in the number of workers per household would also be useful to explore. Such an analysis could provide a much sounder basis for projecting the future year joint distributions of these key attributes for both worker and job synthesis.
2. Methods for the joint synthesis workers (and their attributes) and jobs (and their attributes) should be investigated.
3. Investigate method for incorporating income within the model system.
4. Develop a user interface that allows alternative scenarios to be readily and robustly generated by changing basic inputs (CSD-level totals), occupation distributions, WAH rates, etc.

PoRPoW Modelling Improvement:

Investigate:

1. Improved impedance function specifications.
2. Revisiting mechanisms for introducing auto ownership effects within PoRPoW calculations.
3. How to incorporate wages/income into the model.
4. Improved, systematic use of K-factors and/or spatial segmentation of parameters to improve model fit, including examination of GGH-level interactions and segmentations.
5. Temporal trends in PoRPoW patterns over time in the region.

Dynamic Simulation Modelling:
R&D should be undertaken in parallel to the recommended short-run static model improvements to develop and test operational microsimulation models of the dynamic evolution of population demographics, labour markets and housing markets.

REFERENCES

Anas, A. (1983) "Discrete choice theory, information theory, and the multinomial logit and gravity models", *Transportation Research B* 17, 13-23.

Elgar, I. (2007) *Modelling Office Mobility and Loation Decisions in a Microsimulation Environment,* Ph.D. thesis, Toronto: Department of Civil & Mineral Engineering, University of Toronto Ph.D. thesis, Toronto: Department of Civil Engineering, University of Toronto.

Elmi, A., D.A. Badoe, and E.J. Miller (1999) "Transferability Analysis of Work-Trip Distribution Models", *Transportation Research Record, Journal of the Transportation Research Board*, No. 1676, 169-176.

Farooq, B., E.J. Miller and M. Haider (2013) "A Multidimensional Decision Modelling Framework for Built Space Supply", *Journal of Transport and Land Use*, 6:3, 61-74.

Fox, J., E.J. Miller, A. Daly and S. Hess (2012_ "Temporal Transferability of Models of Mode-Destination Choice of the Greater Toronto and Hamilton Area", presented at the 13th Conference of the International Association for Travel Behaviour Research, Toronto, July 15-20.

Hammadi, A. (2020) *Modelling Transportation System Impacts by Housing Supply Dynamics*, Ph.D. thesis, Toronto: Department of Civil & Mineral Engineering, University of Toronto.

IBI Group (2017) *Traffic zone Population and Employment Forecasts for the Greater Golden Horseshoe*, report prepared for Metrolinx, Toronto.

Lee, D.B. (1973) "Requiem for large scale models", *Journal of the American Institute of Planners*, 39, 163-178.

Lee, D.B. (1994) "Retrospective on large scale urban models", *Journal of the American Planning Association*, 60, 35-40.

Maoh, H. (2005) *Modeling Firm Demography in Urban Areas with an Application to Hamilton, Ontario: Towards an Agent-Based Microsimulation Model*, PhD thesis, Hamilton: School of Geography and Earth Sciences, McMaster University.

Miller, E.J. (2009) "Integrated Urban Models: Theoretical Prospects", invited resource paper, Chapter 14 in Kitamura, R., T. Yoshii and T. Yamamoto (eds.), *The Expanding Sphere of Travel Behaviour Research: Selected Papers from the 11th International Conference on Travel Behaviour Research*, Bingley, UK: Emerald, 351-384.

Miller, E.J. and M.J. Roorda (2003) "A Prototype Model of Household Activity/Travel Scheduling", *Transportation Research Record, Journal of the Transportation Research Board*, No. 1831, 114-121.

Moeckel, R. (2005) "Microsimulation of firm location decisions", *9th International Conference on Computers in Urban Planning and Urban Management*, London.

Mostafa, T.S. (2017) *A Microsimulation Platform of Firm Evolution Prcesses*, PhD thesis, Toronto: Department of Civil & Mineral Engineering, University of Toronto.

Müller, K., and Axhausen, K. W. (2010) *Population synthesis for microsimulation: State of the art. Institute for Transport Planning and Systems*, ETH.

Ozonder, G. and E.J. Miller (2020) "Longitudinal Analysis of Activity Generation in the Greater Toronto and Hamilton Area", T*ransportation*, DOI: 10.1007/s11116-020-10089-w

Shannon, C.E. (1948) "A Mathematical Theory of Information", *Bell System Technical Journal*, 27:379-423, 623-656.

Timmermans, H. (2003) "The saga of integrated land use-transport modeling: how many more dreams before we wake up?", Conference keynote paper, 10th International Conference on Travel Behaviour Research, Lucerne.

Travel Modelling Group (2015) *GTAModel V4.0*, Toronto: University of Toronto Transportation Research Institute, Toronto, ON, Canada.

Webber, M.J. (1977) "Pedagogy Again: What is Entropy?" *Annals of the Association of American Geographers,* 67(2), 254-266.

Wilson, A.G. (1970) "A Statistical Theory of Spatial Distribution Models", Chapter 3 in R. Quandt (ed) *The Demand for Travel: Theory and Measurement*, Lexington, Mass: Lexington Books, 55-82.

APPENDIX I
A PROTOTYPE MODEL OF LABOUR MARKET DYNAMICS[41]

1. Introduction

This appendix introduces a new agent-based microsimulation (ABM) model of urban labour markets, in which workers actively seeking employment in each time period are matched with vacant jobs. The model is designed to operate within the ILUTE (Integrated Land Use, Transportation, Environment) urban simulation model system (Salvini and Miller, 2005; Miller, et al., 2011) under development for the Greater Toronto-Hamilton Area (GTHA). As discussed in the Section 2 literature review, relatively little effort has gone into the development of ABM labour market models, despite the critical importance of modelling place of residence – place of work linkages within integrated models of urban spatial processes.

In the current model application, 1986 is taken as the base year, with 20-year simulations being run (1986-2006) to test the model's performance within a known historical time-period. Section 3 of the paper describes the data used to construct and test the model. Section 4 presents the overall model structure and specification. Section 5 demonstrates the model's performance when applied to a historical validation test period. Additional detail concerning all aspects of this modelling exercise is available in Harmon (2013). Section 6 concludes the paper with a brief discussion of next steps in the model's development and application.

2. Literature review

Agent-based microsimulation models of labour force markets are relatively rare. The Integrated Land Use Modelling and Transportation System Simulation (ILUMASS) (Wagner and Wegener, 2007) includes a firmography and employment simulation model which models the lifecycles of individual firms, but not individual jobs. From a workforce point of view, the number of employees existing at each modelled firm is a product of a Markov-based firmological model. Rather than microsimulating the job market, ILUMASS assigns synthesized employees to firms as they grow without regard for employment demand or wage negotiation.

Barlet, et al. (2009), on the other hand, models individual workers' employment careers, without regard to which firm they may belong. It also does not differentiate among job types. Like ILUMASS, this model assigns each employee a discrete skill level. Workers in higher skill levels may take jobs in lower rungs, but the opposite case is not allowed. Additionally, demographic attributes such as age and worker efficiency influence the job transition process. An efficiency measure relates the employee's attributes to the output of their work. The variable essentially measures the "profitability" of an employee from the perspective of the firm, and plays a role in hiring/firing decisions and wage negotiations.

[41] This appendix is based on Harmon, A. and E.J. Miller (2019) "Microsimulating Labour Market Job-Worker Matching", *Journal of Ambient Intelligence and Humanized Computing*, 11(3) 993-1006.

SAGE (Simulating Social Policy in an Ageing Society) links an agent-based demographics engine with a dynamic labour force model (Zaidi, et al., 2001). Each individual is updated annually to simulate life path transitions which may include attributes such as health, education and the existence of "personal support networks" such as a family structure. These attributes are then used to determine the type of employment the agent seeks via regression models, which are estimated separately for each gender and qualification level ("advanced", and "non-advanced"). Employment is categorized by level (full-time or part-time), type (employee or self-employed), and location (sector, industry, and occupation). Wages are then calculated based on a combination of demographic and job attributes.

LABORSim models the Italian job market, with a focus on educational and job demand interactions and employment transition levels (Leombruni and Richiardo, 2006). An interesting feature in LABORSim is the flexibility that is allowed around education and labour market transitions. Unlike most other systems, educational participation is not modelled as a static process based solely on age; rather, different educational pathways are allowed for each agent. Provided that individuals are of legal age, they may choose to both work and attend school at the same time. Because these pathways are not static, persons are also allowed to "drop out" of both the labour and education modules at any time. LABORSim also explicitly considers immigration and emigration processes within the labour market.

A common feature among most labour market models is the lack of a true job matching model. In most cases, job supply is either assumed to be filled by a synthesized population, or jobs are simply created "on demand" to fulfill some external quota. The Agent-Based Model of Origin Destination Estimation (ABODE) is a notable exception Tilahun and Levinson (2013). It splits jobs into specific skill levels and factors in travel distances as a basis for matching workers with jobs, with this match being dependent on both employee and firm characteristics, specifically the spatial location of each agent and the successful matching of skill levels to the job being offered. This creates a more constrained environment for the job matching process and allows wage determination and job demand to be a function of both the quality of the workforce and the spatial characteristics of the built environment. Additionally, both the prospective employee and the employer can adjust their job searching habits depending on the state of the labour market. In cases where the employee either cannot find a job, or where the employer cannot find suitable candidates to fill a vacancy, both will adjust the "intensity" of their job search accordingly, which dictates how many job applications are sent or received per time slice. Employees also have the option of expanding their search by applying to positions outside of their prescribed skill level. In the end, both the employer and employee have separate criteria that define the utility of the potential applicant and job. The worker, who is concerned with the offered wage and commuting time, compares alternatives and accepts a job that meets a minimum wage threshold. The firm, on the other hand, weighs potential applicants by skills, education and experience. Employees set their minimum accepted wage based on past experiences – whether they are weighing the job against a position they already occupy, or making additional sacrifices in cases where they have been unemployed for an extended period of time.

In addition to the models described above, several other integrated urban microsimulation systems explicitly deal with labour modelling, although all have significant exogenous components. UrbanSim, for example, contains an Economic Transition Model that models both job demand and job supply exogenously, and then maps these jobs to a well-formed spatial location engine (Waddell, et al., 2003). DESTINIE2, which models the French labour market and pension system, models career trajectories using first-order Markovian processes but does not explicitly model labour supply or sector-based employment changes Blanchet, et al., 2009). DYNAMOD-2, on the other hand, does model employment according to industrial sector, occupation, and job type (full-time or part-time); but places caps on the job supply side according to exogenous unemployment and job participation data (King, et al., 1999).

3. Data

A full description of the data sources and attributes used in this research is provided in Harmon (2013). Major datasets employed include the following:

- *Census of Canada, 1986:* Individuals, families and households in the current version of ILUTE are seeded in the base year using publicly available microdata taken from the 1986 Canadian Long-Form Census (Statistics Canada, 1986). The data contains a wide breadth of information on demographics, education and labour force metrics and was used to estimate several of the models outlined in this paper.
- *Labour Force Survey, 1987-2006:* To supplement the Census data, which is only conducted every five years, the Canadian Labour Force Survey (Statistics Canada, 1987-2006) tracks employment levels annually by both occupation and industry.
- *Canadian Business Patterns, June 2012:* Statistics Canada's semi-annual Canadian Business Patterns release (Statistics Canada, 2012a), which contains information on Canadian businesses, including their employee counts, NAISCS code, and latitude-longitude coordinates, is used to attach spatial locations to jobs. To maintain privacy, latitude/longitude coordinates were first taken from the June 2012 data release, converted into the standard X/Y coordinate system used in ILUTE, and then rounded off to the nearest 10 km block.
- *Survey of Labour and Income Dynamics (SLID):* Is an annual survey conducted by Statistics Canada designed to measure economic shifts and its overall effects on the Canadian workforce (Statistics Canada, 1996).

4. An ABM labour market model

4.1 Introduction

The labour market model presented in this paper (hereafter "the model") assumes that a fully synthesized population (consisting of persons within households, by residential location), as well as a fully synthesized set of jobs (by industry, occupation and location), exists at time t in the simulation. Population demographics are updated in each simulation time-step (Chingcuanco and Miller, 2018), as are the residential locations of all households (Rosenfield, et al., 2013). Job supply ideally should be updated by a firmographic model that evolves firms and their attributes (notably, their employment levels by type) over time. Such a firmographic model does not yet exist for the GTHA, so the evolution of jobs over time is directly simulated in this model, as described in Section 4.2.

The labour market model is patterned after the ILUTE residential housing market model (Rosenfield, et al., 2013; Farooq and Miller, 2012), in which workers seeking jobs in each time step are matched with vacant jobs within a Monte Carlo simulation framework. This process can be viewed as a multi-agent game in which workers must search for job vacancies and decide whether to accept or reject any job offers that they receive, and employers must decide to which applicants they will make job offers. Wages (salaries) are endogenously determined through the offer/accept process.

Section 4.2 describes the job supply component of the labour market, while Section 4.3 describes the job demand component. Section 4.4 brings the demand and supply components together within a market clearing process which matches workers to jobs. This market process, in turn, involves processes for firms to evaluate workers with respect to their suitability for a given job, workers to assess jobs (specifically the wage offers associated with each job), selection of a pool of applicants for each job, the generation of job offers to these applicants (along with the determination of the wage being offered), and job offer accept/reject decision-making. Finally, Section 4.5 further discusses the market clearing model in terms of the key feedback mechanisms at play with the process – a critical strength of the ABM framework.

4.2 Job supply process (employer's perspective)

Harmon (2013) describes in detail the attributes of the *Job* class within ILUTE and the synthesis of a complete list of individual jobs for the GTHA for the 1986 base year. These synthesized jobs are stored in lists that are segregated by four primary criteria: industry (16 SIC categories), occupation (10 custom-defined categories), type (full- or part-time) and "experience" (desired worker experience level; 5 categories: 0-2, 3-6, 7-11, 12-20, >20 years), resulting in 1,600 distinct job lists. These are stored in a multi-dimensional *Jobs* array, where each element in the array points to a single list which contains zero or more Job IDs. For example, *Jobs[0,0,0,0]* would return a complete list of full-time management jobs in the agriculture sector with specified experience levels of zero to two years. In addition, hash sets are associated with each of the 1,600 job lists in order to distinguish a sub-set of jobs that are vacant. This is achieved using a similar multi-dimensional structure, where each element in the *VacantJobs* array points to a distinct hash set. *VacantJobs[0,0,0,0]*, for example, would contain a collection of Job IDs representing a vacant sub-set of the *Jobs[0,0,0,0]* master list. Historical data on job vacancies was not available for this study and so an initial, uniform 2% 1986 vacancy was assumed to initialize the simulation, and 2% is maintained as a minimum vacancy rate throughout the simulation runs.

The synthesized base year set of jobs is updated for each one-year time step by adding or deleting jobs within each job category, based on a set of macro-economic regression models that predict the year-over-year percentage change in jobs (positive or negative) for each of the 160 industry-occupation job categories. Explanatory variables in the model are inflation-adjusted real Canadian GDP and the Ontario average unemployment rate, with the models being estimated using Labour Force Survey data for the period 1987-2006. The predicted percentage change in each category is then applied equally across all experience levels for this category. If new jobs are to be created then these are randomly generated. Full-time workers are assumed to work 40 hours/week. Part-time workers are randomly assigned a number of hours per week using a simple normally-distributed probability model based on 1986 Census data. If jobs are to be deleted, then these are randomly selected from the existing set of jobs. Job deletions trigger "layoffs" of the workers holding these jobs, with these workers needing to decide whether to seek a new job or leave the labour force (see Section 4.3). Note that this process does not allow for the simultaneous deletion and addition of jobs within a given job category within any one time step.

Given the current lack of a full firmographic model in ILIUTE, firmographic attributes needed by the wage model are instead synthesized and attached as attributes for each job. Three cross-tabulation probability models are used (see Harmon (2013) for details):

- Firm size classification model (6 categories: <20, 20-99, 100-499, 500-999, ≥1000 employees).
- Multiple firm location model (more than one establishment location per firm; Boolean: yes/no).
- Collective bargaining agreement model (Boolean: yes/no).

A Firm Location Model is tasked with assigning each job in the ILUTE system a unique X and Y coordinate based on the observed spatial distribution of jobs across the GTHA. The Canadian Business Patterns database, which contains a registry of businesses in the area along with their associated NAISCS code, employee count and latitude/longitude coordinates, is utilized (Statistics Canada, 2012b). In order to protect the privacy of the specific businesses contained in the survey, each latitude/longitude pairing is first converted into ILUTE's coordinate system, and then rounded to the nearest 10 km. This system essentially divides the GTHA study area into a 10x10km rectangular grid. By weighting the entries of the database by their employee counts, and then filtering them by industrial sector, a cross-tabulation table was created where each of the sixteen industries is represented on the columns, and each cell within the spatial grid represented on the rows.

4.3 Job demand process (worker's perspective)

Harmon (2013) describes in detail the *Person* class within ILUTE, in particular the person attributes of relevance to labour market participation and job choice. Corresponding to the 1,600 distinct job types discussed above, a multi-dimensional array called *JobSeekers* contains 1,600 distinct lists containing zero or more person IDs of potential applicants to each job bucket. Each list, by definition, contains a collection of individuals possessing a *LabourForceStatus* attribute of *Unemployed*. The individual's industry, occupation, job type and experience attributes determine which of the 1,600 *JobSeeker* lists they belong to. For example, *JobSeeker[0,0,0,0]* would contain a list of unemployed person IDs who are seeking a full-time management position in the agriculture industry, specifically targeting jobs seeking candidates with zero to two years of experience.

These unemployment lists are seeded from Census microdata in the 1986 base year. Deaths and out-migration of persons in the labour force result in additions of vacated jobs to the *VacantJobs* list (for employed persons) and deletions from the *JobSeekers* list (for unemployed persons in the labour force). The workforce updating procedure is performed on an annual basis and is repeated for all individuals 15 years or older. A labour force transition model developed by Hain (2010) determines entries and exits to/from the labour market and "voluntary" transitions of employed workers to unemployed status (non-voluntary "layoffs" are the outcome of the job deletion process described above). Note that workers quitting jobs create job vacancies, which are added to the *VacantJobs* list.

A "job bucket selection model" is used to assign industry-occupation categories to new workers (tenure and experience attributes are assumed to be zero). Similarly, as agents age and gain experience, they may become eligible for different job buckets through time, or may simply want to switch occupations in cases where jobs are hard to come by. In all cases, a procedure is required that assigns individuals to an appropriate bucket based on their demographic, educational and job history attributes. This procedure consists of the following three sub-models:

- *Worker industrial selection model:* This model assigns persons to one of the sixteen industrial sectors of employment. A simple cross-tabulation of industrial sector participation rates as a function largely on the educational attributes of the agent (a multinomial logit (MNL) specification failed to produce acceptable results and so the simpler fixed-rates model was adopted).
- *Worker occupational selection model:* A MNL model (0.406 ρ^2) assigns occupations to agents based on their demographic, educational and industrial selection attributes.
- *Worker job type selection model:* A binominal logit model (0.430 ρ^2) determines choice of full- or part-time employment status. Demographic variables (including head-of-household status, age and highest level of schooling) dominate the model, indicating the important role that "life stages" play in the acceptance of part-time employment.

4.4 Job matching

This section describes the process used to match vacant jobs with unemployed workers within each time step. ILUTE's agent-based nature allows for the simulation of individual applications to individual jobs, and a subsequent wage determination based on specific applicant attributes and market conditions. This section outlines the process, starting with a description of how the firm evaluates potential applicants, followed by a similar description of how workers view job offers. The application collection and job acceptance processes are then detailed, followed by a discussion of the model's market feedback mechanisms.

4.4.1 Applicant utility from the firm's perspective

A common challenge in modelling job markets is a general lack of detailed information concerning firm behaviour, notably concerning employee recruiting practices. However, while the *reasons* why firms decide to hire a specific applicant are largely unobserved, the *value* that a company places on the hired applicant is readily available in the form of income surveys (Marks and Harold, 2011). Specifically, assuming that most firms are profit maximizers, the wages paid to each employee represents an observable measure of the firm's assessment of the employee's value. Herein this is labelled the employee's "utility" to the employer. Therefore, when a firm is considering several applicants for a job posting, the utility of each individual may be viewed in terms of the "theoretical" amount of money s/he would normally be paid. This "theoretical" amount essentially corresponds to the outputs of the ILUTE Wage Model, which produces the average hourly wage of an individual based on a regression model of observed survey data (Hain. 2010). From the firm's point of view, the utility of each applicant is therefore defined by the outcome of the Wage Model. For a given position, the job/firm components of the wage model are constant and so only the person-specific attributes with the model of each applicant are relevant. While the outputs of the Wage Model are regression-based and therefore deterministic, market-based random components are later introduced in the job acceptance algorithm, meaning the final offered wage to an applicant will deviate from the static outcome of the Wage Model.

4.4.2 Job utility from the worker's perspective

Workers are assumed to be price-takers in this model. That is, if a job offer is made with a given wage, the worker must decide to accept or reject this offer as-is; negotiating over the wage is not allowed. Further, the worker can only process one job offer at a time and can only consider a subsequent offer if the first offer has been rejected. Given these assumptions, the offer accept/reject decision is based the concept of a *MinimumWageAccepted* variable. This variable defines a wage floor for each potential applicant, and can loosely be described as a "utility threshold" for each job offer. It is assumed that once an applicant receives an offer exceeding this threshold, the job is automatically accepted. The calculation of the *MinimumWageAccepted* variable is based on a combination of current market conditions, job history and location choice. The base value for this variable is represented by either a 5% premium over the applicant's previous job, or, in the case of new workforce participants, the outcome of the Wage Model. This approach corresponds to findings in the literature, which indicate previous job holders are more likely to demand higher salaries while new entrants are more likely to simply accept the first wage that they are offered (O'Shea and Bush, 2002).

Once the base value has been determined, the spatial attributes of the job are taken into consideration. The coordinates of the job are compared to the coordinates of the applicant's dwelling, and a round-trip Manhattan distance is calculated. Total round-trip travel time is then determined based on actual observed AM peak travel speeds within the GTHA of 50.6 km/hr (Metrolinx, 2008). This travel time is then converted into an actual cost by using the US Department of Transportation's guideline which stipulates that business-related travel costs are equal to 100% of the person's wage per unit of time (Belenky, 2011). This cost is then divided by 8 in the case of a full-time job, or by 5 in the case of a part time job, which is designed to convert the travel expense into its equivalent impact on the hourly wage (using the assumption that full-time employees work an average of 8 hours a day, compared to 5 hours for their part-time counterparts). This result is then added to the *MinimumWageAccepted* value, which effectively raises the employee utility threshold in proportion to the travel costs incurred at a potential job.

On top of these adjustments, a final subtraction from the *MinimumWageAccepted* variable is made based on the current conditions of the employment market, specifically the *JobApplicationAttempts* attribute of the applicant. This attribute effectively records the number of failed previous application attempts of the person, and a high average value of this attribute across all applicants indicates a very competitive job market. Therefore, for each failed application attempt, the applicant will subtract 1% from their *MinimumWageAccepted* attribute to reflect a lower wage threshold in these highly competitive circumstances. The maximum discount allowed is 12%.

4.4.3 Job application process

A random vacant job is first selected from the *VacantJobs* list, and then 12 job candidates are randomly selected from the *JobSeeker* pool, which simulates the worker application process. In cases of weaker job demand, it is possible that the *JobSeeker* pool for a given job contains less than the requisite twelve applicants. In this case, the company will start to consider less qualified candidates by attempting to select additional applications from lower experience buckets, where up to two "step-downs" are allowed. For example, if a firm is having difficulty finding enough applicants for a job stipulating *SevenToElevenYears* of experience, it will first attempt to pull additional candidates from the *ThreeToSixYears* bucket, and failing that, will finally resort to the *ZeroToTwoYears* bucket. Obviously, for jobs stipulating an experience range of *ZeroToTwoYears*, no "step-downs" are possible, and these firms must instead proceed with the matching process with the applications they have on hand.

The applicant pool, which now contains twelve or fewer candidates, is then passed to the job acceptance algorithm (described in detail in the next section), which effectively attempts to offer the job to one of the top three candidates. If this algorithm fails to assign the job, the *FailedRecruitingAttempts* job attribute is incremented. Each increment of the *FailedRecruitingAttempts* attribute will cause the firm to increase their offered wage by 1% up to a maximum possible premium of 12% in future job matching attempts. The application acceptance procedure is iterated until either all job seekers have been exhausted, or until the 2% vacancy rate ceiling for that particular job bucket has been reached. This process is then repeated for each of the 1,600 vacant job buckets.

4.4.4 Job acceptance & wage determination process

The job acceptance algorithm takes the currently selected vacant job and application pool as inputs and attempts to hire the best available candidate for each job. The process effectively sorts the applicant pool by ranking the available candidates by utility. For a randomly selected job, a wage is offered to the best available candidate, which is based on a combination of overall market conditions and the quality of the selected applicant pool. If this offer is rejected, the algorithm is then repeated up to two more times; thus, a job may be offered to up to three candidates in the pool. If these three candidates choose to reject the wage offer, the recruiting attempt is deemed a failure, and the next job in the vacancy pool is randomly chosen. Provided the vacancy rate limit is not reached and eligible candidates remain, the job may later be randomly chosen from the pool again, and subsequent recruiting efforts can be made.

The wage offered to the current best candidate is actually a reflection of the utility of the second-best candidate on the list; i.e., a Vickery Auction process is assumed Vickery (1961). However, a variable premium is added to the wage offer based on the quality of the other candidates. This value, $\Delta V(best)$, represents the percent improvement of utility the top candidate offers when compared to the average of all other applicants in the current pool. The $\Delta V(best)$ percentage is then added to the Wage Model result of the second-best candidate to arrive at the base offered wage amount. This base amount is then further increased by a *MarketAdjustmentFactor*, which increases by 1% for each increment of the job's *FailedRecruitingAttempts* attribute to a maximum of 12%. The resulting amount, deemed the *FinalOffer*, is then forwarded to the best candidate on the list. The agent then compares the *FinalOffer* amount to their *MinimumWageAccepted* attribute, and accepts the job provided if it meets/exceeds the threshold. All other candidates, meanwhile, are said to have failed the current job application process, and as such, their *JobApplicationAttempts* attribute is incremented by 1 and they are returned to the main *JobSeeker* pool where they will be eligible to apply for the next randomly selected job in the bucket.

If the best candidate chooses to reject the employment offer, they receive an increment in their *JobApplicationAttempts* attribute and are automatically removed from the current applicant pool. They will, however, still remain on the master *JobSeeker* list, where they are eligible to apply for subsequent jobs. Meanwhile, the job acceptance algorithm will repeat the above process up to two more times by offering the second and third-best candidates the position before finally exiting and incrementing the *FailedRecruitingAttempts* attribute.

Because the offered wage calculation reflects the second-best candidate in the applicant pool, a special case must be considered when only a single candidate is applying to the job. Obviously, this case represents a *very* weak job demand market, and as such, the firm will attempt to offer the candidate their maximum possible wage in anticipation of no other valid candidates applying to the job in subsequent recruiting attempts. This maximum is once again defined as a 12% *MarketAdjustmentFactor* premium and is applied directly to the outcome of the Wage Model.

4.5 Market feedback mechanisms

While the Job Matching Model described above considers spatial, job history, demographic and educational attributes of applicants, it is important to emphasize the endogenous market feedback mechanisms that are occurring throughout the simulation, as these would normally not exist within an aggregate model. Both workers and firms within the system have an idea of current market conditions based on their agent-specific *JobApplicationAttempts* and *FailedRecruitingAttempts* attributes. In the case of workers, a high *JobApplicationAttempts* attribute would signal a highly competitive applicant pool, while from the firm's point of view, a high *FailedRecruitingAttempts* attribute would signal a weak overall labour market. Both entities adjust their behaviour according to these conditions, and these dynamic adjustment mechanisms within the simulation make for a more intelligent model which is able to react to a wider variety of model states. These feedback mechanisms include:

- *Experience bucket spreading:* As outlined above, in cases where not enough potential applicants exist in a desired experience category, the firm will expand its search to lower experience bucket rungs – a practice which has a direct effect on the eventual offered wage. Clearly, these cases of "experience bucket spreading" occur in situations where the labour market is weak and may result in highly divergent applicant pools. Specifically, as applicants from up to three different experience bucket ranges may exist within the same pool, the difference in worker utility between the best and worst candidate may be very high. Fortunately, the job acceptance algorithm has been set up such that the hiring firm takes these conditions into account, as the offered wage is directly proportional to the $\Delta V(best)$ value. For example, in cases where the best candidate has a significantly higher utility than other applicants in the pool, the $\Delta V(best)$ value will be correspondingly higher, and the firm will therefore increase their initial wage offer accordingly.
- *Market adjustment factor:* The market adjustment factor is applied to both the applicant and the firm in a symmetrical fashion and plays an instrumental part in preventing the overall Job Market Model from diverging over time. The adjustment factors are based on the *JobApplicationAttempts* and *FailedRecruitingAttempts* attributes, and give both the firm and worker a myopic view of current market conditions based on their own job matching histories. For each increment of the corresponding attribute, the worker will decrease his/her minimum accepted wage by 1% while the firm will likewise increase

their offered wage by the same amount. Both agents are limited to a maximum 12% adjustment. By including these feedback adjustments, the chances of "employment stalemates", whereby workers are constantly demanding high wages while firms are correspondingly offering lower wages, are reduced. Further, the symmetric nature of the factor, where wage expectation adjustments occur on the same scale and maximum adjustments are identical for both parties, ensure that neither one has the "upper-hand" in the wage determination calculation.

- *Worker occupation switching:* Although the experience bucket spreading and market adjustment factors ensure offered wages are a continual reflection of market conditions, situations are bound to arise where there are simply too many individuals in a given *JobSeeker* pool compared to the number of available vacancies. In these cases, it is feasible for lower quality candidates to remain in the unemployment pool indefinitely – a situation which would be untenable in the real world. Because of this prospect, an occupation switching algorithm was introduced which gives workers with a *JobApplicationAttempts* attribute exceeding twelve the option of moving to a new occupation bucket within their industry. This effectively allows agents who have already set their *MinimumAcceptedWage* to the lowest possible value without success the ability to move to a new career that may offer better job prospects. The algorithm is based heavily on the Worker Occupational Selection Model, but also draws on real-time job market conditions. The agent first calls the Worker Occupational Selection Model in an iterative fashion until three distinct occupation choices are determined. The probability of selecting a particular occupation is then weighted by taking the ratio of the number of current vacancies in the bucket to the number of potential applicants in the *JobSeeker* list. It should be noted that the agent's previous occupation type is still "fair game" in this algorithm, which may result in some individuals remaining in their current occupation for the time being. This method essentially allows lower-utility agents to switch from a highly competitive job seeking environment to one where jobs are more "in demand". The implementation of this algorithm ensures that the workforce is able to respond to the varying job supply levels of certain occupations over time. It is run on an annual basis after the Job Matching Model has concluded and all desired positions have been filled.

5. Results

Only a few results from applying the model to a historical simulation of the 1986-2006 time period for the GTHA study region are shown in this paper (for more detailed results, see Harmon (2013)). Figure 1 compares the 1986 base year synthesized jobs versus observed Census data, demonstrating an excellent fit across the key job categories. Figures 2 and 3 compare the simulated employment counts by industry and occupation, respectively, over the simulation period versus observed Labour Force Survey data. The results show good fit to observed employment levels across all industrial and occupational classifications. As expected, the accuracy with each job category is directly related to the magnitude of its job count. Actual employment counts for lower job categories, such as Industry 2 (other primary industries) and Occupation 9 ("unique" jobs such as farming), exhibit more stochastic behaviour to the smaller sample sizes. Because the Job Supply Model is regression-based, the resulting simulation tends to "smooth out" these buckets which results in a slightly less accurate prediction.

The differing curve shapes of each job bucket support the need to microsimulate individual occupations and industries, rather than continue with the largely aggregate labour market models employed by most urban microsimulation implementations in use today. The employment counts displayed above are shown over a twenty-year period (1986-2006), and include several periods of macroeconomic growth and decline. Clearly, each and every individual job bucket modelled within the ILUTE environment reacts to these shocks in a unique manner.

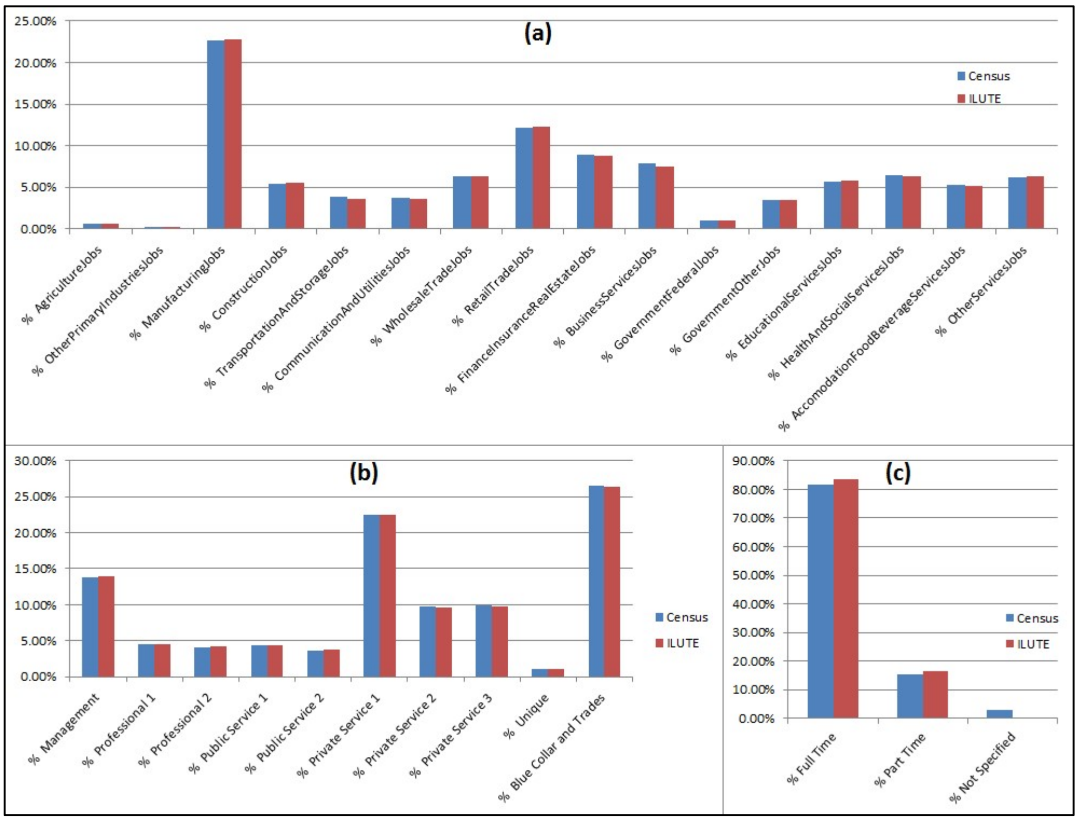

Figure 1. Synthesized jobs versus observed Census microdata by (a) industry, (b) occupation and (c) job type (Source: Harmon, 2013)

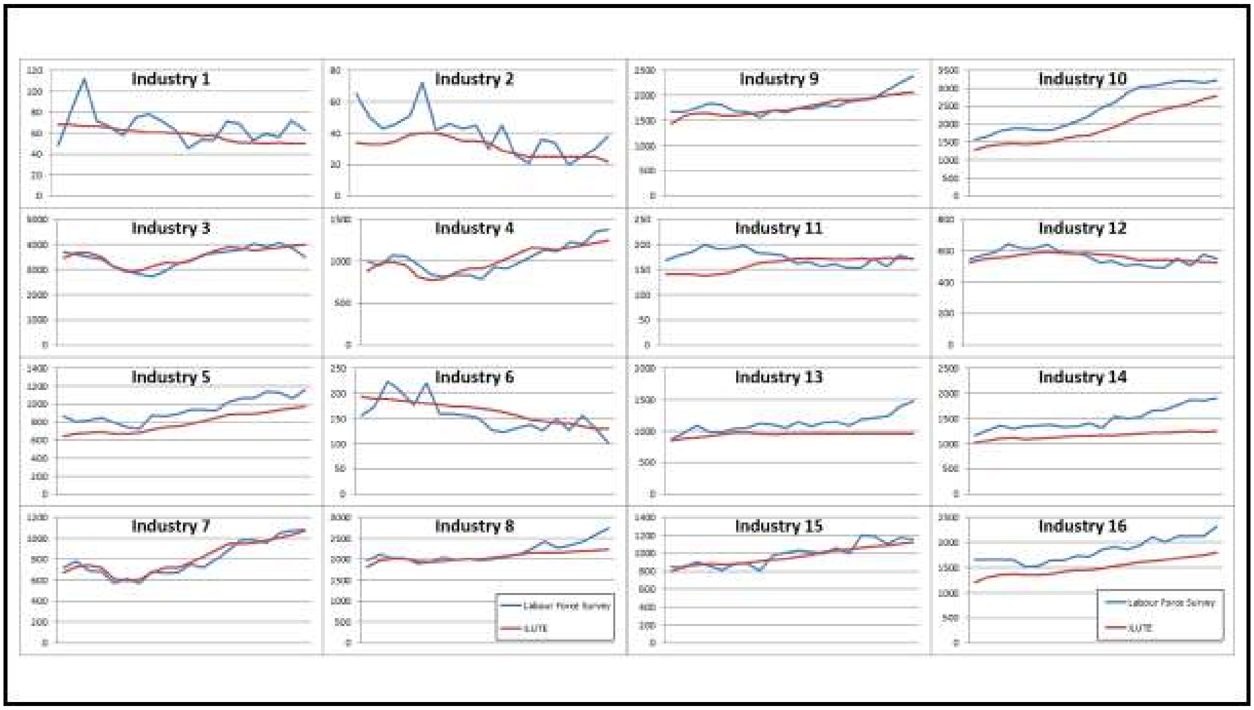

Figure 2. ILUTE simulated employment vs. observed Labour Force Survey data (1987-2006) by industry (Source: Harmon, 2013)

6. Summary & future work

The preliminary results presented in this appendix show that accurate labour market simulation is possible with agent-based components on both the job supply and demand sides. On the supply side, individual jobs are treated as agents and reflect specific industrial- and occupation-based growth patterns based on macroeconomic data. Conversely, the job demand side allows the labour force to be modelled on a per-applicant basis, which considers their personal histories, including demographic and educational attributes. The resulting model is able to capture the agent-based market interactions that would occur in the real world, and provides a much greater level of complexity than most of the aggregate-based job market implementations currently in use.

Although the method presented in this paper has demonstrated that an agent-based job market simulation using currently available data is indeed possible, much research still needs to be done to further understand these market interactions. Specifically, whereas plenty of data are available on the employment details of individuals, very little is known about the *failed* attempts of job seekers and the reasons behind such events. In a similar fashion, the internal behaviours of firms, including the exact criteria behind their hiring decisions, are largely unknown. As discussed in more detail in Harmon (2013), virtually all components of the current prototype model could be improved with access to improved data, as well as further experimentation with alternatives to the many assumptions embedded in the current model.

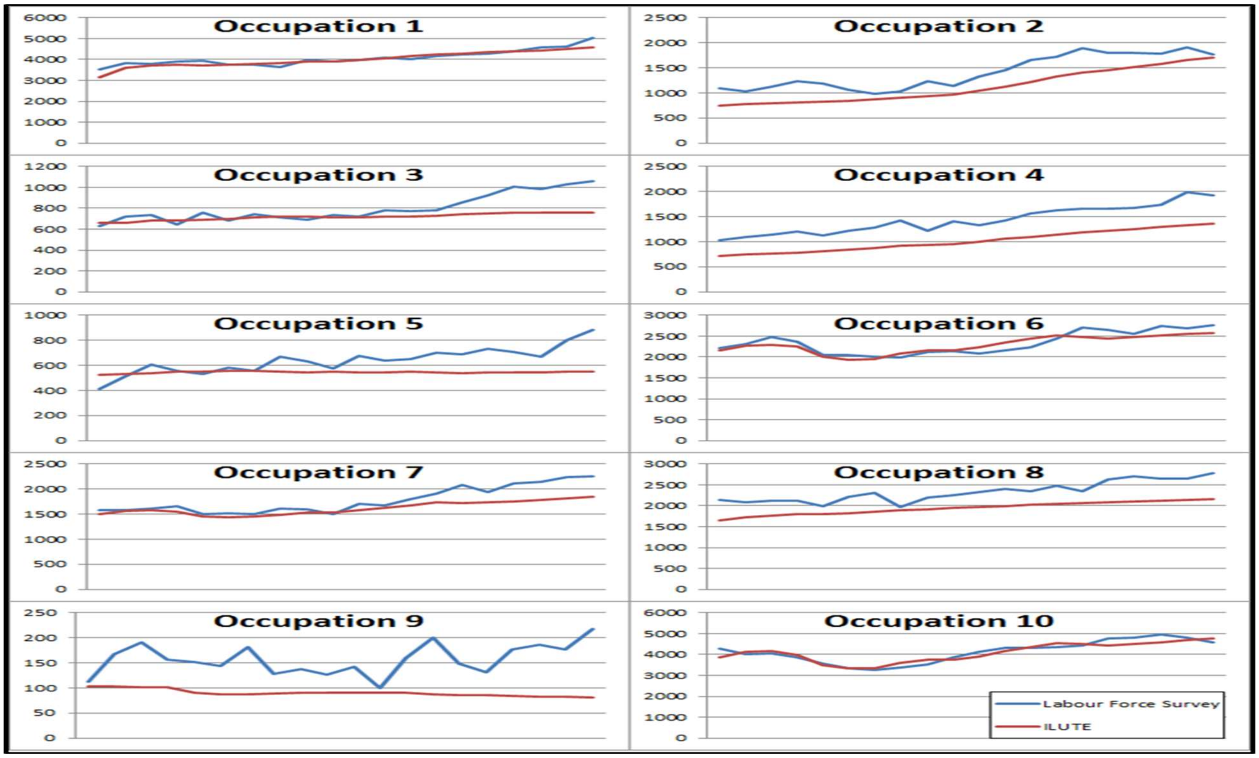

Figure 3. ILUTE simulated employment vs. observed Labour Force Survey data (1987-2006) by occupation (Source: Harmon, 2013)

References

Barlet M, Blanchet D, Le Barbanchon T (2009) Microsimulating supply/demand interactions on a labour market: a prototype. In Zaidi A., Harding A, Williamson P. (eds.) New Frontiers in Microsimulation Modelling, Chapter 7. Ashgate: Aldershot.

Belenky, P. (2011) The value of travel time savings: departmental guidance for conducting economic evaluations, revision 2. Washington: U.S. Department of Transportation. Retrieved Aug 28, 2013 from the U.S. Department of Transportation: http://www.dot.gov/sites/dot.dev/files/docs/Value%20of%20Travel%20Time%20 Memorandum.pdf.

Blanchet D, Crenner E, Le Minez S (2009) The new version of the DESTINIE microsimulation model, increased flexibility and adaptation to users' needs. Retrieved Sept 3, 2013, from Les Journées de Méthodologie Statistique: http://jms.insee.fr/files/documents/2009/93_2-JMS2009_S13-2_BLANCHET-ACTE.PDF

Chingcuanco F. Miller EJ (2018) The ILUTE demographic microsimulation model for the greater Toronto-Hamilton area: current operational status and historical validation. In Thill JC, Dragicevic S (eds), GeoComputational Analysis and Modeling of Regional Science, Advances in Geographic Information Science. Springer International Publishing, 2018:139-159, DOI 10.1007/978-3-319-59511-5_10.

Farooq B, Miller EJ (2012) Towards integrated land use and transportation: a dynamic disequilibrium based microsimulation framework for built space markets. Transportation Research A, 46(7):1030-1053.

Hain, M (2010) Labour market model of the Greater Toronto and Hamilton area for integration within the integrated land use, transportation, environment modelling system. MASc Thesis, University of Toronto.

Harmon A. (2013) A microsimulated industrial and occupation-based labour market model for use in the Integrated Land Use, Transportation, Environment (ILUTE) modelling system, MASc thesis. Toronto: Department of Civil Engineering, University of Toronto.

King A, Bækgaard H, Robinson M (1999) DYNAMOD-2: an overview. National Centre for Economic Modelling, Technical Paper no. 19. University of Canberra.

Leombruni R, Richiardi MG (2006) LABORsim: An agent-based microsimulation of labour supply: an application to Italy. Computational economics 27(1):63–88.

Marks M, Harold C (2011) Who asks and who receives in salary negotiation. Journal of Organizational Behavior. 32:371–394.

Metrolinx (2008) Costs of road congestion in the greater Toronto and Hamilton area: impact and cost benefit analysis of the Metrolinx draft regional transportation plan. Retrieved August 28, 2013 from Metrolinx: http://www.metrolinx.com/en/regionalplanning/costsofcongestion/ISP_08-015_Cost_of_Congestion_report_1128081.pdf.

Miller EJ, Farooq B, Chingcuanco F and Wang D (2011) Historical validation of an integrated transport – land use model system. Transportation Research Record, Journal of the Transportation Research Board 2255: 91-99.

O'Shea PG, Bush DF (2002) Negotiation for starting salary: Antecedents and outcomes among recent college graduates. Journal of Business and Psychology. 16:365–382.

Rosenfield AF, Chingcuanco F, Miller EJ (2013) Agent-based housing microsimulation for integrated land use, transportation, environment model system. Procedia Computer Science, 19:841-846.

Salvini PA, Miller EJ (2005) ILUTE: an operational prototype of a comprehensive microsimulation model of urban systems. Networks and Spatial Economics 5: 217-234.

Statistics Canada (1986) Census of Canada, 1986, individuals file. (subset compiled from public-use microdata file). Retrieved Feb 4, 2013 from: http://sda.chass.utoronto.ca.myaccess.library.utoronto.ca/sdaweb/html/canpumf.htm.

Statistics Canada (1987-2006) Labour Force Survey, 1987-2006 summary table. Retrieved May 12, 2013 from: http://sda.chass.utoronto.ca.myaccess.library.utoronto.ca/sdaweb/html/lfs.htm.

Statistics Canada (1996) Survey of labour and income dynamics, 1996 person file, summary table. Retrieved Aug 15, 2013 from: http://sda.chass.utoronto.ca.myaccess.library.utoronto.ca/cgi-bin/sda/hsda?harcsda+slid96pr.

Statistics Canada (2012a) Table 2820086: Labour force survey estimates (LFS), supplementary unemployment rates by sex and age group. CANSIM (database) version last updated November 2012. Accessed January 8, 2013.

Statistics Canada (2012b) Canadian business patterns, June 2012. Statistics Canada catalogue number 551-0002. Ottawa, Ontario. Accessed December 14, 2012.

Tilahun N, Levinson D (2013) An agent-based model of origin destination estimation (ABODE). Journal of Transport and Land Use, 6:73-88.

Vickery W (1961) Counterspeculation, Auctions, and Competitive Sealed Tenders. The Journal of Finance. 16(1):8–37. doi:10.1111/j.1540-6261.1961.tb02789.x.

Waddell P, Borning A, Noth M, Freier N, Becke M, Ulfarsson G (2003) Microsimulation of urban development and location choices: design and implementation of UrbanSim. Networks and Spatial Economics, 3:43-67.

Wagner P., Wegener M (2007) Urban land use transport and environment models: experiences with an integrated microscopic approach. disP 170:45–56.

Zaidi A, Rake K (2001) Dynamic microsimulation models: A review and some lessons for SAGE, SAGE discussion paper 2. ESRC SAGE Research Group, London School of Economics, London.

APPENDIX II
AN EVOLUTIONARY DEMOGRAPHIC MODEL (EDM) FOR THE GTHA[42]

1. Introduction

This appendix presents Evolutionary Demographic Model (EDM) which updates the residential population demographics within the ILUTE model system. ILUTE is an agent-based microsimulation model that dynamically evolves the urban spatial form, economic structure, demographics and travel behavior over time for the Greater Toronto-Hamilton Area (GTHA). It has been designed to be a credible, policy-sensitive decision support tool for transportation and land use planning [1,2,3,4].

In particular, the appendix provides a comprehensive description of the EDM, as well as presents some historical validation tests. It has undergone significant development and has reached a state of maturity where a 100% synthetic GTHA population of persons, families and households has been tested against a twenty-year historical (1986-2006) period.

A microsimulation approach is highly desirable for demographic modelling in order to enhance behavioral fidelity and reduce aggregation bias [5]. It can easily be argued that the relatively limited impact that disaggregate mode choice models have had on travel demand modeling, for example, can be rooted in the difficulty of projecting the required population socio-demographic attributes [10].

The rest of this appendix is organized as follows. Section 2 briefly reviews demographic microsimulation. Section 3 gives a high-level description of the EDM. In particular, the section describes its design and implementation, the data sources used, the demographic attributes generated and maintained throughout the simulation, and the demographic processes modeled. Section 4 then gives a detailed description of each of the EDM processes being modeled. Section 5 concludes the appendix by presenting and discussing the results from the full population twenty-year validation runs and touches on the model's computational performance. Finally, a conclusion follows as well as an outline of future research directions for the EDM.

2. Literature Review

Microsimulation is a general method to exercise a disaggregate model over time [5]. It is used to analyze complex and/or dynamic systems with many elements that interact with each other. For this type of system, a closed-form analytical expression is often not available due to the complex nature of its processes. In this case, computer-based simulations offer the best alternative to make intelligent predictions by evolving the system through time.

---

[42] This appendix is based on Chingcuanco, F. and E.J. Miller (2018) "The ILUTE Demographic Microsimulation Model for the Greater Toronto-Hamilton Area: Current Operational Status and Historical Validation", J.C. Thill & S. Dragicevic (eds), *GeoComputational Analysis and Modeling of Regional Science, Advances in Geographic Information Science*, Springer International Publishing.

2.1 Demographic Microsimulation Mechanics

A number of demographic microsimulation models have been built in order to analyze issues such as retirement, population projection, labour supply, and other matters related to household life-cycle changes. Comprehensive reviews of existing models can be found in Morand et al. [14] and Ravulaparthy and Goulias [15], who collectively examine sixteen models built for different regions. For most of these models, the demographic events represented can be categorized as: population changes (in- and out-migration, birth and death); household formation (marriage/cohabitation, divorce/separation, children leaving homes); and the education, health and work status of the population.

Microsimulation models typically have one of two starting points: a cross-section of the population or a birth cohort. In both cases, the initial step is to define the agents (e.g., households) where a starting point could be a snapshot of the population of interest, such as disaggregate records from a census [16]. However, such data are often not available due to privacy and cost concerns. One way around this is to use different sources of publicly available aggregate data to synthesize a base population. Once a synthetic population has been created, the microsimulation engine acts on the agents in the simulation. The occurrences of demographic events (ageing, marriage, etc.) are evaluated for each member, and their attributes (age, marital status, etc.) are updated once these events have been identified. The goal is to maintain the representativeness of the base sample throughout the simulation.

2.2 Demographic Microsimulation Typology

Microsimulation models can differ in the way they execute events over time (continuous vs. discrete) and how they manage relationships among population members (open vs. closed models) [14]. For continuous time models, the durations of all possible state transition events are generated for each member of the population. The first event to occur is executed, and this procedure is repeated using the first event as the starting point. In contrast, discrete time models treat time periods one after the other, "stepping through" time in the classic sense. These models execute all possible state transition events that are realized at every time step. While continuous time models may have some theoretical advantages, they are often more complex to implement and less transparent than their discrete time counterparts.

In a closed demographic model, the simulation usually starts with a sample of the population, which includes links between population members (e.g., family ties). Members can enter/exit the population through birth/death and in-/out-migration events. Throughout the simulation, the relationships among the members are tracked and the changes are propagated throughout their social networks. For instance, if agents X and Y get married, new links are formed between them. Both agents are full population members being simulated. In contrast, open models do not maintain associations. Using the same example, if agent X gets married, a new spouse will be attached to agent X as an attribute. The new spouse is not a full population member being explicitly simulated, but only exists to properly model agent X's marriage and life path.

2.3 ILUTE Demographic Microsimulation

The ILUTE EDM is a closed and discrete time demographic model. Being closed, social networks are maintained throughout the simulation, which can be useful for modeling social travel behavior [17]. In addition, the spatial distribution of these social networks (e.g., where one's parents live) arguably also serve as "anchor points" that characterize household residential search behavior [18]. With respect to its treatment of time, ILUTE uses a modified discrete time approach that supports multiple temporal scales. This allows models with different time periods to be integrated into the model system in a simple and transparent manner.

3. Overview of the ILUTE Demographic Updating Module

Given a synthetic base population, EDM updates these agents' attributes at each time step. New agents are introduced through birth and in-migration, while agents exit through death and out-migration events. Unions between agents are formed through a marriage market, while a divorce model dissolves existing marriages. Transitions to new households are also triggered by a move-out model. In addition, each person's driver's license ownership and education level are managed.

3.1 Demographic Attributes

Population members in ILUTE are represented by household, family, and person agents. Households are defined as one or more persons living within the same dwelling unit. They can consist of any combination of individuals and families. Families are defined either as husband-wife couples with or without children, or single parents living with children. Links between these members are explicitly maintained throughout the simulation, which allows family relationships to be tracked over time. Note that all families and individuals must belong to a household.

Table 4.1 lists the demographic attributes of the person class. All these attributes are maintained and/or modeled for all agents across the entire simulation. Persons have an exclusive association with a family or a household. Hence, when either the FamilyId or HouseholdId is non-zero, the other by definition is zero. Agents maintain family relationships through identifiers (e.g., SpouseId). Sex, MaritalStatus and EducationLevel are enumeration types, which are defined data types of named constants. There is also a flag to signify driver's license ownership.
Both family and household classes have member lists: households have a list of families and individuals and families have a list of members. Like person agents, families maintain associations with their households through a household ID. Similarly, households have unique dwelling IDs, which imply a one-to-one mapping between households and dwelling units.

Table 4.1 Person class demographic attributes in defined in EDM

Attribute	Type	Attribute	Type	Attribute	Type
MyID	int	ExSpouseIdList	List<int>	EducationLevel	none
HouseholdId	int	ChildIdList	List<int>		kindergarten
FamilyId	int	SiblingIdList	List<int>		elementary
MotherId	int	DriversLicense	bool		highschool
FatherId	int	MaritalStatus	single		college
SpouseId	int		married		undergrad
Age	short		divorced		graduate
Sex	male/female		widowed		

3.2 EDM Processes

EDM is executed in yearly time-steps. A bottom-up approach is employed in which the demographic evolution emerges through the sequential updating of each person. The whole model can be broken down into a sequence of three main parts. First, demographic events are evaluated for each agent in the simulation. The process takes a list of agents, uses their attributes to compute transition probabilities (e.g., age), evaluates these events (e.g., death), and adds the agents to respective lists (e.g., list of deceased agents). After all the possible state transitions have been determined for the entire population, all the realized events are processed to reflect their changes. For instance, the family relationships of deceased agents are managed (e.g., the spouse is widowed). A cleanup process is executed to delete or convert invalid families and households after they have been updated.

Table 4.2 lists the demographic processes modeled as well as the factors that drive them. Depending on data availability, these models range from simple empirical probabilities (birth) to more advanced methods such as hazard (divorce) and logit (education) models. They can also either be static or dynamic. The letters under the "Data Code" column match the data sources found in Table 4.3, which describe the geographic levels of the data. Model outcomes are conditioned on the agent's current state. For instance, the likelihood of a birth event is a function of a female's age, marital status and current year of the simulation. Each of these models is described in further detail in the next section.

In addition to parallelization concerns, EDM has also been designed for modularity. This allows components to be easily replaced. For instance, a hazard divorce model was at one point implemented in place of an older rate-based one, with very minor code modifications.

Table 4.2 Demographic processes modeled in EDM and a summary of factors that drive their transition probabilities

Process	Factors	Temporal	Type	Data Code
Birth	age; marital status	Dynamic	Rate-based	A, B, C, G, I
Death	age; marital status; gender	Dynamic	Rate-based	A, B, C, G, I
Marriage	age; marital status; gender	Dynamic	Rate-based and Logit	D, E, H
Divorce	ages; marital status; years of birth	Static	Hazard	J
Move Out	school/job changes; gender	Static	Hazard	K
Driver's License	age; gender; geographic location	Dynamic	Rate-based	L
Education Level		Dynamic	Logit	Under development

Out-Migration	Dynamic	Rate-based	F, G
In-Migration	Dynamic	Rate-based	F, G

3.3 Data Sources

A list of the data sources used by the EDM models are found in Table 4.3. The “Data Code” column matches that of Table 4.2 to map the respective data sources to the EDM processes they drive. Except for the Ontario birth and death registries (item I), all the data are publicly available. Data sources B to H are available through the Canadian socioeconomic information management system, which is a database maintained by Statistics Canada. Sources A, J and K are housed under the Computing in the Humanities and Social Sciences (CHASS) data center while source L is provided by the Data Management Group (DMG). Both CHASS and DMG are University of Toronto data centers.

Table 4.3 Data sources for the EDM models by level of aggregation

Data Code	Data Source and Description	Sources
A	Public Use Microdata Files, by census metropolitan area (1986, 1991, 1996, 2001, 2006)	[33]
B	Estimates of population by sex and age group, by census division (1986-2002) [051-0016]	[34]
C	Estimates of population by sex and age group, by census division (1996-2006) [051-0052]	[34]
D	Marriages by marital status and age of groom and bride, Canada (2000-2002) [101-1005]	[34]
E	Estimates of population by marital status, age group and sex, Canada, provinces and territories (1986-2006) [051-0010]	[34]
F	Total population, census divisions and census metropolitan areas (1986-2006) [051-0034]	[34]
G	Components of population growth, by census division (1986-2006) [051-0035]	[34]
H	Estimates of births, deaths and marriages, Canada, provinces and territories (1986-2006) [053-0001]	[34]
I	Ontario births and deaths registry, by municipality (1986, 1991, 1996)	[34]
J	General Social Survey on the family, Canada (1995)	[35]
K	General Social Survey on family transitions, Canada (2006)	[35]

L	Transportation Tomorrow Survey, by wards (1986, 1991, 1996, 2001, 2006)	[36]

Data with varying spatial and temporal levels are used, and the best available proxy data are employed when needed. Many of the empirical probabilities employed combine various data sources to get a comprehensive cross-section across the required socio-demographic dimensions (e.g., age groups, gender, and marital status) and time periods.

4. Descriptions of Individual EDM Processes

This section gives detailed descriptions of the individual models used that drive EDM. It also contains estimation results as well as rate calculations that explain how the agents in ILUTE make demographic decisions, and how the changes from these events are propagated throughout the simulation. EDM manages demographic relationships and seeks to maintain reasonable population, family and household counts throughout the simulation. These variables serve as important inputs to other ILUTE components. For instance, the new households that result from marriages, births and divorces are key drivers to ILUTE's residential housing market.

EDM is initialized with a set of agents/objects which are synthesized from base year Census (and perhaps other) data. A 100% population of persons, families, households and dwelling units for each census tract in the study area has been constructed for 1986 using a modified IPF procedure [19] that:

- Simultaneously generates these four objects in a fully consistent manner.
- Permits a large number of object attributes to be included in the synthesis.
- Is computationally efficient.
- Makes full use of multiple multivariate tables of observed data.
- Is extendable to include additional elements (e.g., household auto ownership, which is not yet included in the synthesis procedure).

For model testing purposes, either the full 100% population can be used, or a smaller subset, randomly drawn from the full population, can be used to speed up run times, with all other model elements and processes (e.g., building supply, etc.) being appropriately scaled.

The education model is not discussed in this appendix.

4.1 Marriage

Marriages in ILUTE are broken into three main steps: 1) a marriage event occurs in which potential marriage candidates join a marriage market; 2) a marriage market is executed where potential grooms and brides are paired off; and 3) the family relationships and attributes of the new couple are processed (e.g. setting husband-wife relationships, transferring existing children, etc.). Marriage events, which trigger an individual to join the marriage market, are driven by rates. These rates are calculated through empirical probabilities for population cross-sections across 13 age groups, 3 marital statuses (single, divorced, widowed), 20 time periods (1987-2006), and gender.

After a marriage event, the individuals join a marriage market in which they are paired with other potential brides and grooms. The matching is executed under a utility maximization framework. A potential bride or groom is randomly chosen from the marriage pool. A choice set is generated for this candidate by drawing agents of the opposite gender from the pool. The candidate's utilities for these matches are calculated based on [37]. These utilities are based on the potential couple's incomes, education levels, and the male/female ratios in their respective geographic areas. These utilities are converted to choice probabilities via a multinomial logit formulation, and a match is made through simulation. A logit formulation is used in order to introduce some stochasticity in the matching. The marriage market is discussed further in [2]. After the marriage market is cleared, the family relationships of the new couple are updated. Depending on the situations of the individual newlyweds, this could include forming new households or merging existing ones, as well as transferring any children over. Newlyweds with new households enter the housing market. Note that at the start of the ILUTE simulation, marriage durations for the base population are estimated from census data using a regression model. This is critical for the operation of the divorce module, which is discussed in the subsection below.

Note that the marriage module intends to include common law unions between males and females. The authors chose to continue denoting these events as a "marriage" to follow convention (e.g., "marriage market"), as well as to be consistent with prior ILUTE work. The authors also caution that there is some inconsistency with this intention and the data, as the marriage process data (e.g., for matching males and females) only account for officially recognized marriages. More importantly, the current marriage module can only handle heterosexual unions. Same-sex unions are not explicitly modeled, but are implicitly accounted for in the non-marital household formation process briefly described in subsection 5.5.

4.2 Divorce

The ILUTE divorce process evaluates whether a divorce event occurs for existing husband-wife couples in the simulation. The agents' attributes (e.g. marital status) and family relationships are also updated. A spouse is moved out and a new household is created for this agent. Custody is also handled according to Ontario aggregate rates (59% for mother single custody and 33% for joint custody). A proportional hazards regression model (Table 5.1) was estimated using the 1995 and 2001 General Social Surveys on the family (source J in Table 4.3) to model divorce decisions. Due to a lack of data, the divorce model does not include a temporal component, i.e., the same regression is applied for all divorces across the 20-year simulation.

Table 5.1 Proportional-hazards regression results for the divorce model

Variable	Coef.	Exp(Coef.)	S.E.	t-Stat	Pr(>|Z|)
hPreviousDivorce	0.675	1.964	0.108	6.243	0.000
wPreviousDivorce	0.571	1.770	0.124	4.610	0.000
hAgeSquaredFrom25	-0.001	0.999	0.001	-2.898	0.004
wAgeSquaredFrom25	0.002	1.002	0.001	2.983	0.003
withIn5Yrs	-0.119	0.888	0.073	-1.622	0.105
marriedAfter1960s	0.664	1.943	0.109	6.120	0.000
wMarriedBefore1950s	-0.473	0.623	0.325	-1.454	0.146
hMarriedBefore1950s	-0.595	0.551	0.323	-1.842	0.065
hBornBefore1945	-0.232	0.793	0.096	-2.404	0.016

wBornBefore1945	-0.391	0.676	0.108	-3.632	0.000
hBornAfter1959	0.129	1.138	0.146	0.885	0.376
wBornAfter1959	0.270	1.310	0.119	2.265	0.023
Number of Observations	25,262				
Number of Events	5,012				
Likelihood Ratio Test	9,341	on 14 df	p = 0		
Wald Test	11,467	on 14 df	p = 0		
Score (Logrank) Test	29,660	on 14 df	p = 0		

4.3 Birth

The birth process handles all birth related events in ILUTE, including evaluating the birth event, updating the attributes of the mother, creating the new born baby, and managing family relationships (e.g. adding parent-child links, creation of a new family, etc.). The birth rates are calculated through empirical probabilities for population cross-sections across 7 age groups, 4 marital statuses (single, married, divorced, widowed) and 20 time periods (1987-2006). If the new mother is married or already has children, then the new baby is simply added to the mother's existing family and household. Otherwise, a new family and household are created, and the agents enter the housing market.

4.4 Death

The death process handles all death related events in ILUTE, including evaluating the death event, removing the deceased from the simulation, and managing family relationships (e.g. making the spouse a widow, making the children orphans or finding new guardians, exiting the housing market if active, etc.). The death rates are calculated through empirical probabilities for population cross-sections across 24 age groups, 4 marital statuses (single, married divorced, widowed), 20 time periods (1987-2006) and gender. When the household head agent of a non-individual household dies, a new agent is designated.

4.5 Moving Out

A move out process is used to transition young adults into moving out from their families into their own households. New households are created for transitioning agents and they enter the housing market. A proportional hazards regression model was estimated using the 2006 General Social Surveys on family transitions (Source K in Table 4.3). Table 5.2 displays the estimation results. Similar to divorce, the move out model does not include a temporal component. A complementary household formation process is used to create and maintain non-family households with more than one individual (e.g., student roommates, friends sharing an apartment, etc.).

4.6 Driver's License

The driver's license process has two functions: grant drivers' licenses to eligible candidates; and revoke these licenses when drivers get too old to drive. The Transportation Tomorrow Survey (Source L in Table 4.3) was primarily used to calculate the driver's license acquisition and revocation rates, which are taken for cross-sections across 3 levels of aggregation of the 46 TTS planning districts in the GTHA, 80 valid ages (16-95), 20 time periods (1987-2006) and gender.

Table 5.2 Proportional-hazards regression results for the move out model

Variable	Coef.	Exp(Coef.)	S.E.	t-Stat	Pr(>|z|)
LiveParents15	-0.127	0.880	0.002	-54.750	<2e-16
School	1.620	5.053	0.002	817.020	<2e-16
Job	1.259	3.520	0.002	505.970	<2e-16
Male	-0.141	0.869	0.002	-80.400	<2e-16
Number of Observations	1,497				
Number of Events	906				
Likelihood Ratio Test	748,022	on 4 df	p = 0		
Wald Test	721,136	on 4 df	p = 0		
Score (Logrank) Test	834,458	on 4 df	p = 0		

4.7 Out-Migration

The out-migration process manages all out-migration related events in ILUTE. Out-migration numbers for the GTHA census divisions (Toronto, Durham, Peel, York, Halton and Hamilton) were taken from Statistics Canada. These values were divided by the corresponding GTHA census division populations to obtain the out-migration rates for 6 census divisions and 20 years. Out-migration events are handled in the same manner as death, though out-migrating household heads have the decision to out-migrate their entire families with them. At present, there is a 75% chance of this happening, and if this event is true, the family members are simply added to the out-migration persons list.

4.8 In-Migration

The in-migration process manages all in-migration related events in ILUTE. Unlike out-migration, in-migration does not require calculating in-migration rates. Instead, actual in-migration numbers are used to synthesize in-migrant agents for each year. The attributes of the in-migrating agents (e.g., age, gender, household status, etc.) are determined from the data. Note that these in-migration numbers are scaled down by a factor that corresponds to ILUTE's base population size, and these factors were calibrated to get the observed total population numbers per year.

When new agents are immigrated in, their corresponding families and households are also built. A process that builds familial relationships across a batch of agents, which is also used by ILUTE's population synthesizer, is executed. There may be some advantages of synthesizing in-migrant agents directly from data distributions instead of randomly drawing from the observed data. This alternative is intended to be explored.

5. Simulation Results

This section presents a twenty-year (1986-2006) simulation run for a fully synthesized population against historical data for the GTHA. The simulation starts with over 6.5 million agents (4.1 million persons, 1.1 million families, 1.4 million households), and the overall number of agents grow past 10 million after a twenty-year run. On a computer with an i7-2600 processor (3.4 GHz, 4 cores) with 16 GB of RAM running on a 64-bit Windows 7 operating system, the simulation takes just under 10 minutes to complete, including 2.5 minutes to load a base population and form the initial relationships among the agents.

While the figures below aggregate the simulation outputs for the entire region, each simulation process follows the geographic level of detail afforded by the data, as defined in Tables 4.2 and 4.3. Furthermore, note that the empirical rates to drive these models are known ex-post, as the objective of this entire section is to illustrate the performance of running the full EDM. That is, the focus is to demonstrate a valid model system, and less on building accurate individual models (e.g., in-migration forecasts).

Figures 6.1 and 6.2 compare the 1986 and 2006 age distribution of males and females in ILUTE with historical data. Each of the four sets of bar graphs sum to 100%. For the most part, the simulation produces the correct age distributions by gender after 20 years. Although the simulation under-predicts females greater than 75 and over-predicts 10 to 19-year-olds, the errors are relatively small (in the order of 1% absolute error per age group). Figures 6.3 and 6.4 add another dimension by plotting the 2006 distribution of males and females by age and marital status for ILUTE and their corresponding historical values. The areas under each set of marital status curves sum to 100%. Note the presence of two axes (married and singles on the left, widowed and divorced on the right) and the scale difference between the male and female widowed and divorced axes. Again, the distributions are generally tracked quite well after the twenty-year simulation. The under- and over-predictions of females illustrated in Figure 6.2 are revealed in Figure 6.3 to correspond to widowed and single agents.

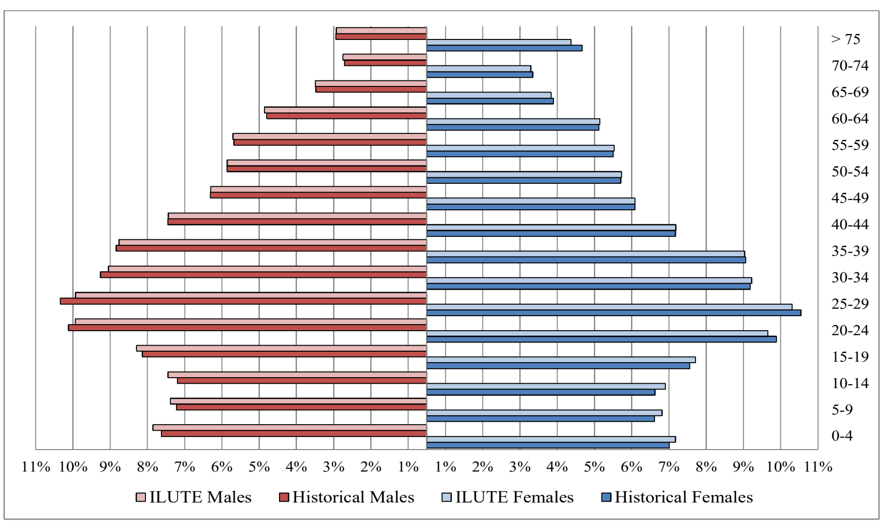

Fig. 6.1 1986 ILUTE vs. historical age distributions for males and females

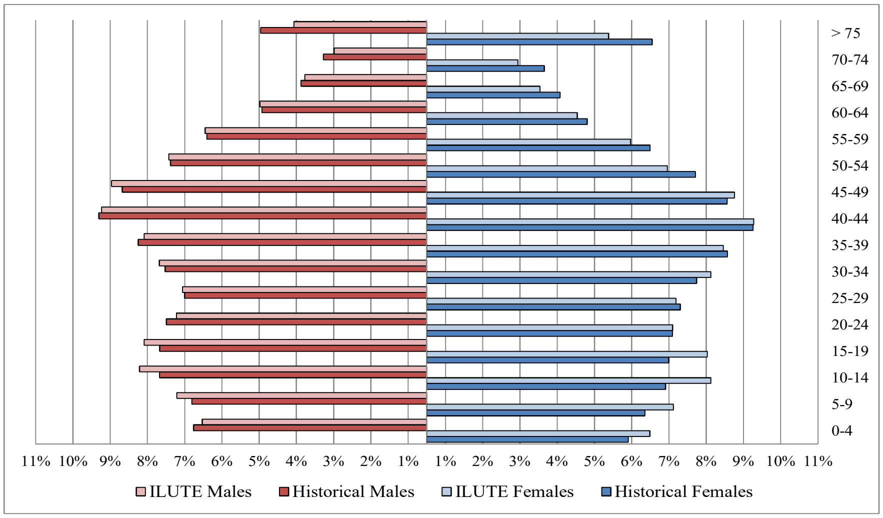

Fig. 6.2 2006 ILUTE vs. historical age distributions for males and females

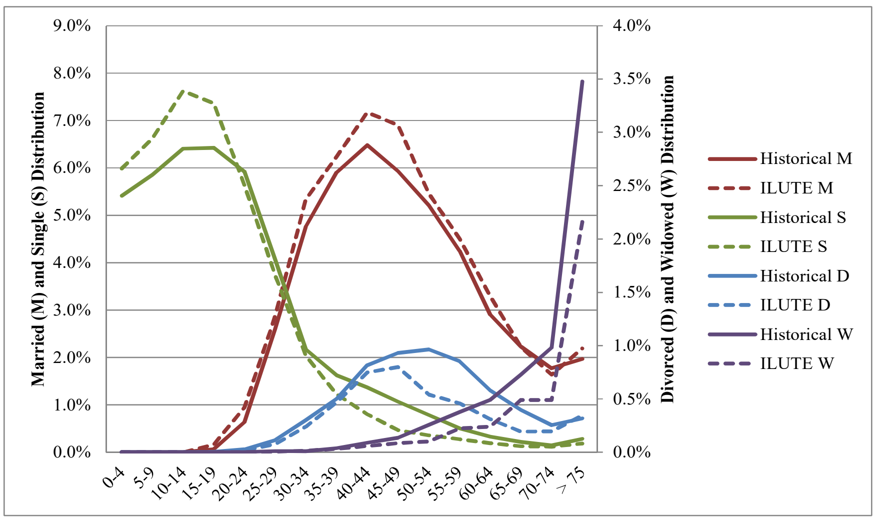

Fig. 6.3 2006 ILUTE vs. historical female population by marital status and age

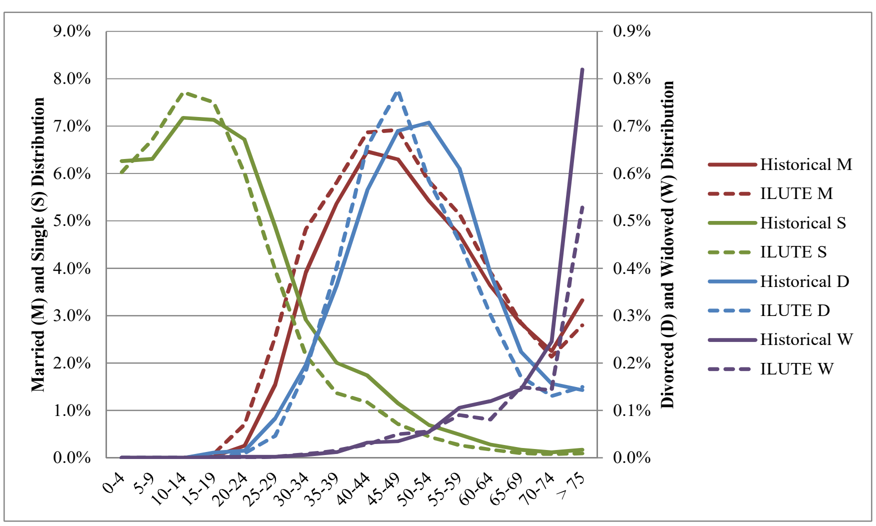

Fig. 6.4 2006 ILUTE vs. historical male population by marital status and age

Besides maintaining proper marital status and age distributions, EDM also seeks to preserve the correct distribution of household types. Table 6.1 presents simulated vs. historical household type distributions for four years (2006 data are not available). ILUTE tends to produce too many single individuals and too few single families as the simulation progresses. This discrepancy can be attributed to multiple factors, including: birth and marriage rates being too low, divorce and move out rates being too high, and the out-migration model's insensitivity to socio-demographic factors. The overproduction of female widows (Figure 6.3) can also be related to this issue.

Table 6.1 ILUTE vs. historical household type distributions

		Single Individual	Multiple Individual	Single Family	Single Family and Individuals	Multiple Family
Census	1986	20.8%	2.8%	74.0%	2.2%	0.1%
	1991	21.4%	3.7%	71.6%	3.1%	0.2%
	1996	22.0%	3.0%	72.6%	2.2%	0.2%
	2001	22.2%	2.9%	72.6%	2.1%	0.2%
ILUTE	1986	21.1%	3.3%	74.1%	1.0%	0.5%
	1991	23.3%	2.8%	71.8%	1.8%	0.4%
	1996	25.3%	2.4%	70.3%	1.7%	0.3%
	2001	27.3%	2.2%	68.7%	1.5%	0.3%

Figure 6.5 plots the birth, death and out-migration rates (left axis) as well as the absolute population levels (right axis) for ILUTE and the corresponding historical benchmarks. The birth and death rates seem to perform quite well (with a slight under-prediction of deaths), but the out-migration rates start off a bit too high. While the model corrects itself as the simulation progresses, this may be due to the population levels increasing faster than they should have. That is, a larger denominator results in lower out-migration rates. Absolute population levels are also plotted on the same figure for comparison. While ILUTE starts with about 300,000 less persons in 1986, the rate of growth seems to match the observed values quite well. The delta in the base population numbers is an issue with the population synthesis and is currently being investigated.

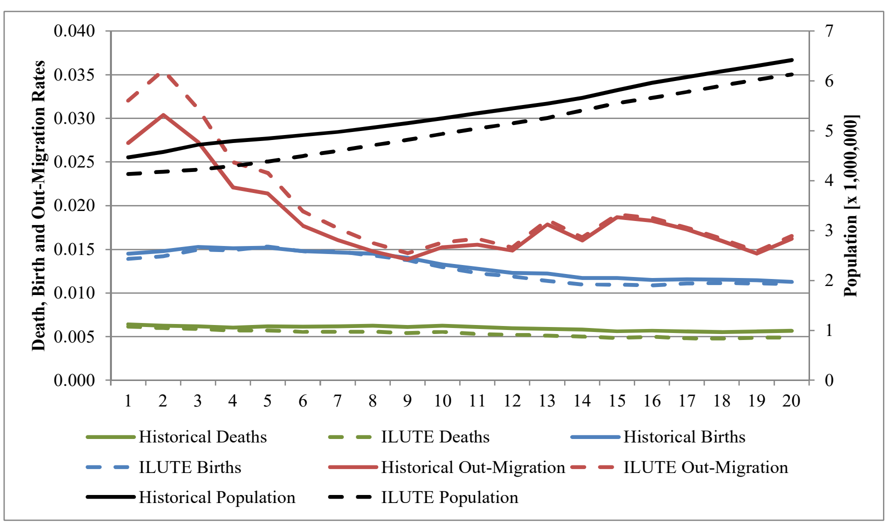

Fig. 6.5 ILUTE vs. historical birth, death and out-migration rates and total population levels

Following this, Figures 6.6 and 6.7 demonstrate how social networks are built and maintained throughout the simulation. At the very start of the simulation, only synthesized families have relationships among each other (e.g., parent and child association). As the simulation progresses, agents start to build secondary associations (e.g., grandparent-grandchildren links) through intermediate agents (i.e., the parent). Histories are recorded as shown by the growing ex-spouse list. Note that the percentage of agents that have relationships plateau out due to agents exiting the simulation. Figure 6.7 depicts the population's growing social network connectivity throughout the 20 years. As mentioned earlier, these social connections can help predict the spatial choices of people (e.g., residential location choice, destination choice, etc.) and is beneficial to be maintained.

Figure 6.8 compares the distribution of divorces in ILUTE from 1986 to 2006 to historical values, which illustrates the utility of tracking agent histories across the simulation. The marriage date of each agent couple is maintained, and this is used in evaluating divorce decisions. While the plot demonstrates a well performing divorce model, it also reaffirms the performance of the marriage model. For example, since the divorce model uses a hazard function with age-related covariates, agents would have to get married at the right age and find partners with the appropriate age differences to get the correct divorce distributions shown. Preliminary results of EDM's marriage market can be found in [1] and [2].

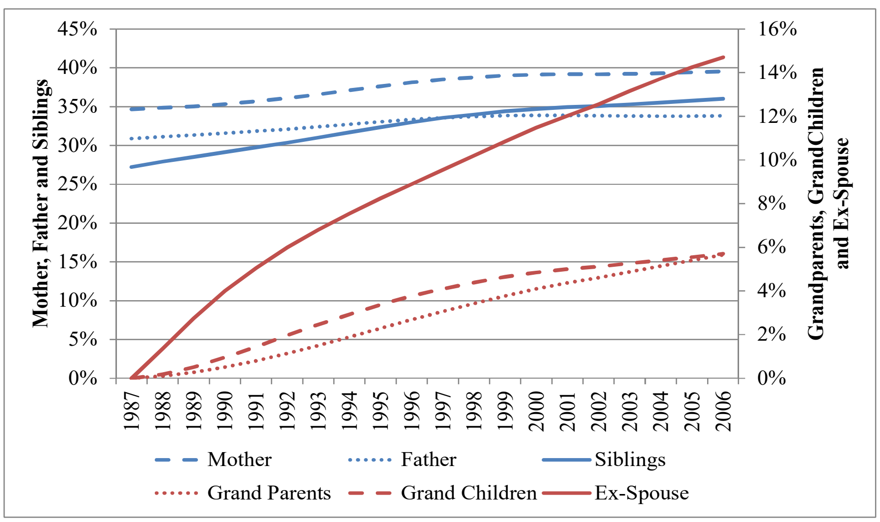

Fig. 6.6 Percent of ILUTE population with a particular relationship over time

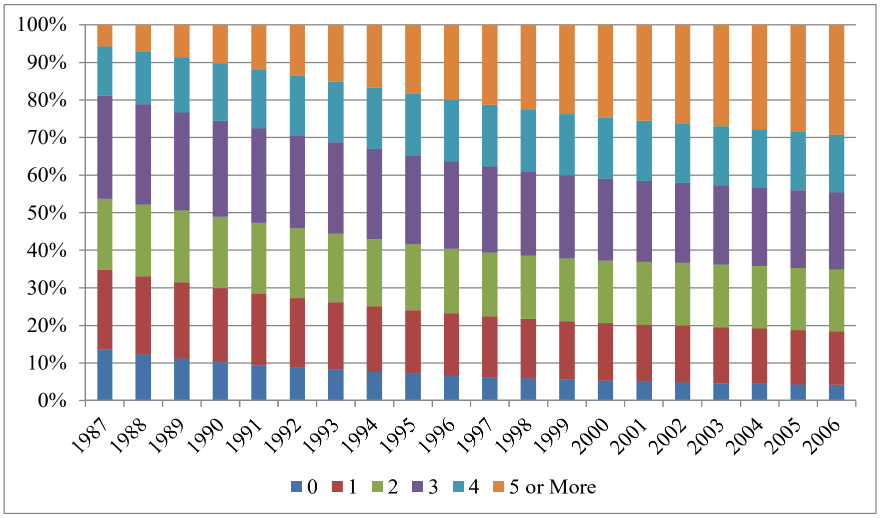

Fig. 6.7 Percent of population with social connections over time

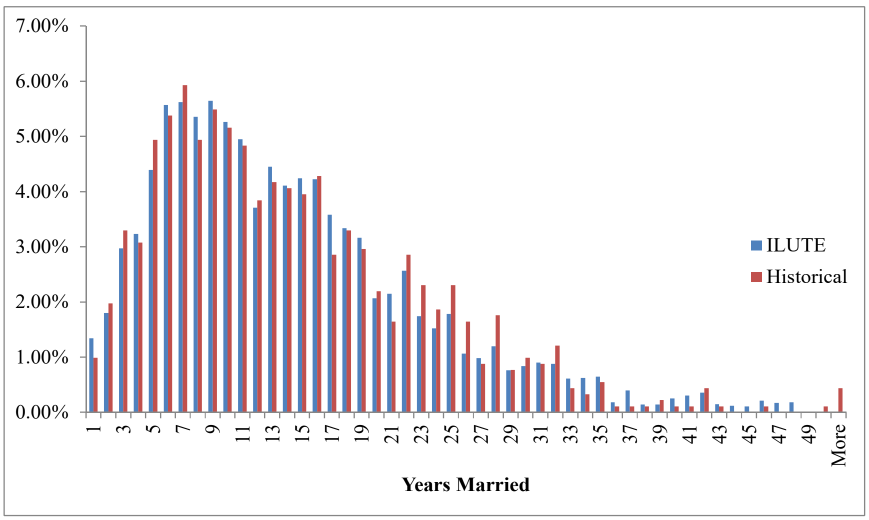

Fig. 6.8 1986-2006 ILUTE vs. historical divorces by years married

6. Discussion and Future Directions

This appendix presents the operational status of the ILUTE Demographic Updating Module (EDM). The performance of EDM is then compared against historical observations across multiple dimensions. In general, EDM exhibits a strong performance, and the authors have confidence that it can maintain the validity of inputs to the other behavioral models in ILUTE. Note that multiple simulation runs have also been conducted to explore the uncertainty of the outputs, which are important for validating microsimulation models. The results (not shown here) are distributed very tightly around the single-run outputs presented in this paper, which is reasonable given that relatively simple demographic models are used throughout. This also suggests that clear demographic patterns could emerge across millions of simulated agents, despite their heterogeneity.

As discussed previously, ex-post values are used in building the individual models. A focus going forward is to conduct demographic forecasting exercises. While finding new independent sources of data would be helpful, it is also possible to estimate the models for half the simulation period (1986-1996) and evaluate its performance going forward (1997-2006). Some components of EDM are still under development (education and driver's license), and this is also the focus of current research. Future research steps include integrating EDM with models of labor force participation and automobile ownership, which require operational and validated education and driver's license sub-models.

References

1. Miller, E.J., B. Farooq, F. Chingcuanco and D. Wang. Historical Validation of an Integrated Transport – Land Use Model System. In Transportation Research Record: Journal of the Transportation Research Board, No. 2255, 2011, pp. 91-99.
2. Farooq, B., E.J. Miller, F. Chingcuanco and M. Giroux-Cook. Microsimulation Framework for Urban Price-Taker Markets. In Journal of Transportation and Land Use, Vol. 6, No. 1, 2013, pp. 41-51.
3. Salvini, P.A. and E.J. Miller. ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. In Network and Spatial Economics, Vol. 5, No. 2, 2005, pp. 217-234.
4. Miller, E.J., J.D. Hunt, J.E. Abraham and P.A. Salvini. Microsimulating urban systems. In Computers, Environment and Urban Systems, Vol. 28, 2004, pp. 9-44.
5. Miller, E.J. Microsimulation. In Transportation Systems Planning: Methods and Applications. Goulias, K. ed. CRC Press LLC: Boca Raton, 2003.
6. Eluru, N., A.R. Pinjari, J.Y. Guo, I.N. Sener, S. Srinivasan, R. Copperman and C.R. Bhat. Population Updating System Structures and Models Embedded in the Comprehensive Econometric Microsimulator for Urban Systems. In Transportation Research Record: Journal of the Transportation Research Board, No. 2076, 2008, pp. 171-182.
7. Miller, E.J. and M.J. Roorda. A Prototype Model of Household Activity/Travel Scheduling. In Transportation Research Record: Journal of the Transportation Re-search Board, No. 1831, 2003, pp. 114-121.
8. Goulias, K. and R. Kitamura. Travel Demand Forecasting with Dynamic Microsimulation. In Transportation Research Record, No. 1357, 1992, pp. 8-17.
9. Chingcuanco, F. and E.J. Miller. A microsimulation model of urban energy use: Modelling residential space heating demand in ILUTE. In Computers, Environment and Urban Systems, Vol. 36, No. 2, 2012, pp. 186-194.
10. Miller, E.J., D.S. Kriger, and J.D. Hunt. Integrated Urban Models for Simulation of Transit and Land Use Policies: Final Report. TCRP Web Document 9, 1998.
11. Orcutt, G.H. A new type of socio-economic system. In Review of Economics and Statistics, Vol. 39, No. 2, 1957, pp. 116-123.
12. Mitton, L., H. Sutherland, and M. Weeks. Introduction. In Microsimulation Modelling for Policy Analysis. Mitton, L., Sutherland, H. and Weeks, M. Eds. Cambridge University Press: Cambridge, 2000.
13. Joachim, M. Microsimulation - A Survey of Methods and Applications for Analyzing Economic and Social Policy. FFB discussion paper no. 9, Universität Lüneburg, Germany, 1994.
14. Morand, E., L. Toulemon, S. Pennec, R. Baggio and F. Billari. Demographic modelling: the state of the art. SustainCity Working Paper 2.1a, Paris, 2010.
15. Ravulaparthy, S. and K. Goulias. Forecasting with Dynamic Microsimulation: Design, Implementation, and Demonstration. UCTC-FR-2011-07, University of California Transportation Center: Santa Barbara, 2011.

16. Müller, K. and K.W. Axhausen. Population synthesis for microsimulation: State of the Art. Presented at the Swiss Transport Research Conference, Ascona, 2010.
17. Carrasco, J.A. and E.J. Miller. How Far and with Whom Do People Socialize? Empirical Evidence About Distance Between Social Network Members. In Transportation Research Record: Journal of the Transportation Research Board, No. 2076, 2008, pp. 114-122.
18. Huff, O.J. Geographic Regularities in Residential Search Behaviour. In Annals of the Association of American Geographers, Vol. 76, No. 2, 1986, pp. 208-227.
19. Pritchard. D.R. and E.J. Miller. Advances in Population Synthesis: fitting many attributes per agent and fitting to household and person margins simultaneously. In Transportation, Vol. 39, No. 3, 2012, pp. 685-704.
20. Hain, M. Labour Market Model of the Greater Toronto and Hamilton Area for Integration within the Integrated Land Use, Transportation, Environment Modelling System. Master's thesis, University of Toronto, 2010.
21. Harmon, A. A Microsimulated Industrial and Occupation-Based Labour Market Model for Use in the Integrated Land Use, Transportation, Environment (ILUTE) Modelling System. Master's thesis, University of Toronto, 2013.
22. Habib, M.A. and E.J. Miller. Influence of Transportation Access and Market Dynamics on Property Values: Multilevel Spatio-Temporal Models of Housing Price. In Transportation Research Record: Journal of the Transportation Research Board, No. 2076, 2008, pp. 188-191.
23. Habib, M.A. and E.J. Miller. Reference-Dependent Residential Location Choice Model within a Relocation Context. In Transportation Research Record: Journal of the Transportation Research Board, No. 2133, 2009, pp. 56-63.
24. Haider, M. and E.J. Miller. Modeling Location Choices of Housing Builders in the Greater Toronto Area, Canada. In Transportation Research Record: Journal of the Transportation Research Board, No. 1898, 2004, pp. 148-156.
25. Farooq, B. and E.J. Miller. Towards Integrated Land Use and Transportation: A Dynamic Disequilibrium based Microsimulation Framework for Built Space Markets. In Transportation Research Part A: Policy and Practice, Vol. 46, No. 7, 2012, pp. 1030-1053.
26. Rosenfield, A., F. Chingcuanco and E.J. Miller. Agent-based Housing Market Microsimulation for Integrated Land Use, Transportation, Environment Model System. Presented at the 2nd International Workshop on Agent-Based Mobility, Traffic and Transportation Models, Methodologies and Applications (ABMTRANS'13), Halifax, 2013.
27. Mohammadian, A. and E.J. Miller. Dynamic Modeling of Household Automobile Transactions. In Transportation Research Record: Journal of the Transportation Research Board, Vol. 1831, 2003, pp. 98-105.
28. Duivestein, J. Household Vehicle Fleet Decision-Making for an Integrated Land Use, Transportation and Environment Model. Master's thesis, University of Toronto, 2013.
29. Hatzopoulou, M., E.J. Miller and B. Santos. Integrating Vehicle Emission Modeling with Activity-based Travel Demand Modeling: A Case Study of the Greater Toronto Area (GTA). In Transportation Research Record: Journal of the Transportation Research Board, Vol. 2011, 2007, pp. 29-39.
30. Hao, J., M. Hatzopoulou and E.J. Miller. Integrating an Activity-Based Travel Demand Model with Dynamic Traffic Assignment and Emission Models: An Implementation in the Greater Toronto Area. In Transportation Research Record: Journal of the Transportation Research Board, Vol. 2176, 2010, pp. 1-13.

31. Mostafa, T. and M.J. Roorda. A Framework and Analysis of Firm Evolution Processes. Presented at the METRANS International Urban Freight Conference (I-NUF) Long Beach, 2013.
32. Roorda, M.J., R.A. Cavalcante, S. McCabe and H. Kwan. A conceptual framework for agent-based modelling of logistics services. Transportation Research Part E: Logistics and Transportation Review, Vol. 46, No. 1, 2010, pp. 18-31.
33. Statistics Canada. Public Use Microdata Files. http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=11-625-XWE&1ang=eng. Accessed: June 28, 2012.
34. Statistics Canada. Canadian socio-economic information management system. http://www5.statcan.gc.ca/cansim/home-accueil?lang=eng. Accessed: June 28, 2012.
35. Statistics Canada. The General Social Survey. http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=89F0115X&CHROPG=1&lang=eng. Accessed: June 28, 2012.
36. Data Management Group. Transportation Tomorrow Survey. http://www.dmg.utoronto.ca/transportationtomorrowsurvey/index.html. Accessed: June 28, 2012.
37. Choo, E. and S. Seitz. The Collective Marriage Matching Model: Identification, Estimation and Testing. In Eugene Choo, Matthew Shum (ed.) Structural Econometric Models (Advances in Econometrics, Volume 31), Emerald Group Publishing Limited, 2013, pp.291-336.

APPENDIX III
A HOUSING MARKET SIMULATION MODEL[43]

1. Introduction

The ILUTE (Integrated Land Use, Transportation, Environment) model system is an agent-based microsimulation model that evolves an urban region's spatial form, demographics, travel behavior and economic structure over time. An operational prototype of ILUTE has been developed for the Greater Toronto-Hamilton Area (GTHA), simulating the evolution of a synthesized population over a twenty-year timespan (1986-2006).

The model system uses a dynamic population of agents (individuals, households, firms, etc.) that are endogenously evolved as the simulation progresses. Recently, there has been much progress towards developing the ILUTE Demographic Updating Module (I-DUM), which updates socio-demographic attributes throughout the simulation. I-DUM has received comprehensive testing, including a historical validation over a twenty-year period [1]. Current efforts have focused on improving ILUTE's land use modeling component, specifically the owner-occupied housing market which consists of models for: households' residential mobility decisions, location choices and valuations [2] [3]; the endogenous supply of housing by type and location [4]; and the endogenous determination of sale prices and rents [5].

This paper presents the Housing Market Evolutionary System (HoMES), an updated implementation of ILUTE's housing market module. It is a complete microsimulation of urban housing market dynamics, with new models for market demand, price formation, as well as a bid-auction process on a fully disaggregate level. A framework of the new implementation is discussed in comparison to previous modeling efforts, and preliminary results are presented.

2. HoMES Overview

2.1. Highlights

HoMES retains some of the desirable features of an initial version of the housing market implemented in Farooq and Miller [5]. These include:

- Life-cycle changes (e.g., arrival of a new baby, job changes, etc.) trigger the decision of households to become active in the housing market.
- Asking prices and new supply numbers are sensitive to previous sales prices and volumes.
- The market has a disaggregate representation of buyers and sellers, which are defined to be utility/profit maximizing agents with limited information.
- Prices are formed endogenously through a non-cooperative market clearing solution.
- The transaction price for each dwelling is directly determined by its bidders but also influenced by the prices and conditions across the entire market.
- The market clearing does not impose equilibrium and instead operates on constant disequilibrium that arguably better captures the stochasticity and path dependency of real housing markets.

In comparison to the previous model system, HoMES offers the following new features and advantages:

---

[43] This appendix is based on Rosenfield, A. F. Chingcuanco and E.J. Miller (2013) "Agent-Based Housing Microsimulation for Integrated Land Use, Transportation, Environment Model System", *Procedia Computer Science*, 19, 841-846.

- The market is cleared by simulating individual bids through households' willingness to pay for dwellings in their choice sets.
- The relatively strict assumptions of the previous formulation are relaxed. For instance, the old clearing mechanism fails when a dwelling being cleared has either only one or a multitude of bidders.
- There is better integration among the housing market sub-models, such as households' mobility triggers affecting their choice set formations and tenure decisions.
- A new asking price model was estimated that is more sensitive to macro-economic conditions, in addition to its dependence on previous transaction prices.
- HoMES' implementation has been parallelized, which brings significant computational performance improvements.

2.2. Module Description

Various processes in ILUTE employ either a price-taking or price-formation market framework as a means of matching supply and demand agents [5]. The labor, marriage, and rental housing markets are all examples of price-taking markets, whereby prices are set prior to the market clearing process. In the case of marriage, no 'price' exists, but the process can otherwise be modeled through the concept of supply and demand agents [6].

The owner-occupied housing market is, conversely, a price-formation market in which buyers bid the amounts they are willing to pay and sellers choose among the bids they receive. The housing market is updated in monthly time-steps, where individual dwelling units are listed on or withdrawn from the market and are cleared in an auction-type approach. However, ILUTE's housing market remains in a perpetual state of disequilibrium as there is no requirement for all supply and demand agents to clear at each step. Rather, homes may remain vacant should an acceptable bid not be received, and potential buyers may remain active in the market for several months at a time. Market dynamics are thus affected by the micro-level decisions of agents, as well as by the macroscopic impacts of excesses or deficiencies in either supply or demand.

Market entry for demand agents (i.e. households) arises from (a) the decision of an existing household to seek relocation, or (b) the formation of a new household seeking residence. Existing households may choose to enter the market by means of the *Residential Mobility Model*, while new households are formed as a result of the Demographic Updating Module. Either process results in a household entering the market demand pool. A binary choice model then determines whether the household will search in the rental or owner-occupied market.

Supply agents, namely dwellings, also enter the market through two different means: (a) the dwelling of an actively searching household may be put up on the market once that household decides to begin its search, or (b) a new dwelling may be constructed and listed on the market by its developer.

Following market entry, the clearing process takes place whereby:

1. Sellers determine their asking prices based on their perceived value of their dwellings.
2. Buyers form their choice sets of potential dwellings.
3. Auctions take place for the active dwellings in the current time-step.

Agents may exit the market by either completing a successful transaction, or withdrawing from the auction. In either case, the market agent will then assign the appropriate linkages and update the bidder and seller pools. An overview of market processes is shown in Figure 1.

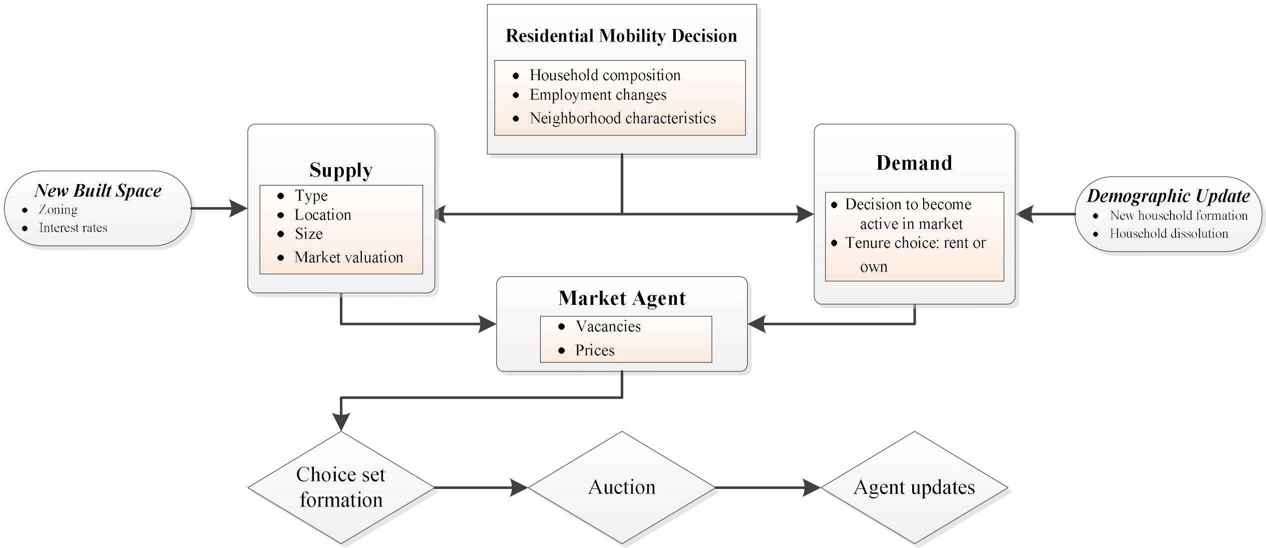

Figure 1. Framework of housing market supply, demand and clearing mechanisms.

3. Module Components

3.1. Market Entry

The determinants of residential mobility can be interpreted through the concept of residential *stressors* as originally introduced in Rossi's 1955 classical study, *Why Families Move* [7]. Various demographic and economic factors play a role in determining a household's relative satisfaction for its current dwelling against alternatives; this difference in perceived satisfaction is defined as stress. Residential stress evolves over the course of a household's tenure, with changing family composition and surrounding economic conditions resulting in push or pull forces towards moving.

Life-cycle events occurring through I-DUM such as marriage, childbirth and divorce all contribute to increased likelihood of residential mobility and are thus important factors to be considered in a mobility model. A mixed-logit model developed by Habib [3] quantifies the effect of various residential stressors on a household's residential mobility decision. These stressors include: changes in employment (e.g., gains/losses/changes of jobs), changes in family composition (e.g., childbirth, moving out, aging), duration in current dwelling, and spatiotemporal economic data.

If a household decides to enter the housing market, the factors that trigger its mobility later become determinants of its choice set of dwelling alternatives. This forms an important linkage between the decision to become mobile and the ensuing location choice set. This is of particular significance since the tenure decision (rental or owner-occupied housing), which directly impacts the supply and demand numbers for the owner-occupied market, has been shown to be dependent on a household's life-cycle phase, as well as the macroeconomic climate in which it resides [8]. Furthermore, this decision is also found to be highly correlated to the previous tenure of a household: renters frequently choose to become homeowners, while the inverse is relatively uncommon. Other significant explanatory factors include household size and gross income, both of which are positively correlated to a higher propensity of homeownership.

3.2. Property Valuation

Upon market entry, each seller determines a price at which to list its house, with the asking price serving two purposes: firstly, it narrows the set of potential bidders, who have preconceived financial search constraints; secondly, it acts as a benchmark for dwelling utility calculations which form the basis of the market clearing mechanism. The asking price model has been updated to better incorporate the set of factors that have been found to affect list prices. These include physical dwelling attributes such as size and structural type. Also included are several macroeconomic indicators, namely unemployment, fuel prices and mortgage rates, all of which were found to be strong predictors of future housing market performance.

The most powerful predictor of future prices, however, is through historical trends. The new valuation model accounts for the results of previous endogenous housing market interactions, thus forming a longitudinal link between each month's market activity. Aggregate market trends such as supply-demand imbalances will affect prices, as will an individual dwelling's market history. For example, a home that has lingered on the market for several months will tend to drop its asking price. By integrating the dwelling valuation model within the ILUTE system, the interdependencies between economic growth and housing market performance can be captured while allowing for exogenous macroeconomic indicators to influence entire market dynamics.

3.3. Location Choice

While the seller seeks to maximize profit, the homebuyer is seeking to maximize his or her utility attained by purchasing a new home. Although the buyer is seeking an optimal solution (i.e., the "perfect home"), various constraints preclude the searcher from considering all possible dwelling alternatives. Thus, the household must confine its search to a limited number of dwelling alternatives. This choice set forms the sample space for each potential transaction. In this sense, buyers and sellers have a myopic approach to the market whereby only limited information about each agent is known to others. Prospective buyers only have detailed information on dwellings in their choice set, and sellers are similarly unaware of the actions of other sellers. As such, it is imperative that the choice set formation model accurately represent the preferences of each household as all subsequent market clearing processes depend on an appropriate selection of homes.

Of all the mechanics of the housing market, the choice set problem is possibly the most difficult to model, owing to the inherent subjectivity associated with how households choose potential homes. Young [9] proposed an *elimination-by-aspects* (EBA) approach in which the choice set is confined through a decision tree, where a household's most important attributes are first used to eliminate alternatives, followed by successively less important traits. This is in contrast to traditional residential location models, which assume a trade-off between the agent's satisfaction for one attribute and its satisfaction for another [10]. An EBA approach simplifies the choice set restriction while more closely capturing the decision process of individuals who seldom consider all attributes of an alternative simultaneously when confining their search [11].

Because only completed transactions are reported by most real estate agencies, location choice data is limited to households' final destinations rather than the alternatives considered during the search process. Modelers must therefore rely on other data sources such as retrospective preference surveys, which seek to uncover what criteria were important in a household's search, and how these criteria were prioritized. In analyzing residential search behavior, the 1998 Residential Search Survey (RSS) is used as a primary dataset. The survey reveals the spatial and temporal trends of search processes in the GTHA, with a comparison of stated and revealed preferences for homebuyers and renters. More interested readers are directed to Pushkar [12].

The HoMES choice set formation is based on an EBA algorithm, using search preferences revealed by the RSS. Up to twelve dwellings are considered each month, reflecting observed search tendencies. Dwelling alternatives are first filtered by tenure, with the assumption that dwelling tenure remains fixed and no tenure conversion or subletting occurs. Alternatives are then narrowed down by structural type (detached, semi-detached, row house, low/high rise apartment) using a multinomial logit with household income, size, and previous dwelling characteristics as predictor variables.

Following choice set restriction by structural types, the searcher will choose to only consider dwellings whose size matches the needs of the household. Such a restriction links to the residential mobility model (expanding or contracting households looking to upsize or downsize, respectively). The final restriction of alternatives is by asking price, which scopes down the list of potential dwellings significantly. Historical patterns provide a range of typical price-to-income ratios for homebuyers. All of these restrictions result in a choice set small enough to be manageable for the potential homebuyer.

3.4. Auction Process

The final step in the HoMES model is market clearing. The previous version, as described in Farooq and Miller [5], expresses a buyer's bid for a house through the probability that the house is sold to this buyer at a given price. A price is determined through a search procedure that clears this dwelling unit. Unfortunately, due to this probabilistic formulation, the previous version has some limitations. For instance, the market fails to clear whenever a bidder only has one house in his or her choice set since that probability will always be one at any price. To overcome this limitation, the new version works on the basis of individual bids based on households' willingness-to-pay (WTP) for dwellings in their choice sets.

Under a random utility framework [13], the attractiveness of dwelling j to buyer i can be expressed as a utility composed of three terms: a price utility term (V_{ij}^r), a non-price utility term (V_{ij}^R) based on dwelling characteristics, and an error term (ε_{ij}) accounting for differences between observed and predicted choice behavior. Dwelling utility U_{ij} is the sum of these terms:

$$U_{ij} = V_{ij}^r + V_{ij}^R + \varepsilon_{ij} \tag{1}$$

where $V_{ij}^r = \gamma R_{ij}$, where $R_{ij} = \log(price_{ij})$, weighted by a parameter γ;
$V_{ij}^R = \boldsymbol{\beta} \cdot \boldsymbol{X}_{ij}$, where $\boldsymbol{X}_{ij}$ is a vector of dwelling attributes weighted by a vector of parameters $\boldsymbol{\beta}$; and
ε_{ij} is a Type I Extreme Value distributed error term.

The total utility is therefore:

$$U_{ij} = \gamma R_{ij} + \boldsymbol{\beta} \cdot \boldsymbol{X}_{ij} + \varepsilon_{ij} \tag{2}$$

Each active household i has a choice set of dwellings (C_i), and each active dwelling has a set of bidders (B_j). In the previous market clearing formulation [5], a micro-equilibrium constraint was imposed whereby a dwelling's selling price is varied until certain conditions are met (probability sums across buyers and sellers both equal one). This algorithm was found to break down when bidder sets are too small or too large. If a dwelling has very few bidders, the transaction price may be forced to drop significantly to achieve a unit sum of probabilities. Conversely, for a dwelling with many interested homebuyers, the price may be inflated beyond reasonable values. Furthermore, such an algorithm only determines a transaction price, with the winning bidder chosen randomly from amongst the bidder set.

A new clearing mechanism has been implemented that, rather than imposing a micro-equilibrium, auctions off each dwelling using bidders' WTP. Bidders first determine their non-price utility for each dwelling in their choice set:

$$\widetilde{U}_{ij} = \boldsymbol{\beta} \cdot \boldsymbol{X}_{ij} + \varepsilon_{ij} \tag{3}$$

$\boldsymbol{\beta} \cdot \boldsymbol{X}_{ij}$ is evaluated based on the attributes of the dwelling unit and the parameters found in Habib [14]. The random error term ε_{ij} is simulated by drawing from a Type 1 Extreme Value distribution. Then, for each active dwelling j with bidder set B_j, members of the bidder set determine the highest utility they can achieve for all homes in their choice set *other* than the current dwelling j:

$$U_{ij_{max}} = \max_{j' \neq j \in C_i} (\gamma Ask_{j'} + \widetilde{U}_{ij'}) \tag{4}$$

where $Ask_{j'}$ is the dwelling's asking price. $U_{ij_{max}}$ gives the maximum utility bidder i can obtain from the other alternatives in its choice set and therefore sets bidder i's WTP for all other dwellings in its choice set. Herein lies the importance of the asking prices, as they reflect the market's valuation of each home and are a necessary means of comparison between choice set alternatives.

Household i, with a maximum utility of $U_{ij_{max}}$, can bid a certain amount on dwelling j to achieve that same utility. Substituting equations (3) and (4) into (2), the price the bidder is willing to pay for this utility is therefore

$$R_{ij} = \left(U_{ij_{max}} - \widetilde{U}_{ij}\right)/\gamma \tag{5}$$

and reflects the bidder's relative preference to all other dwellings in its choice set.

Once all bids have been tendered, the seller evaluates its options and may choose to either sell to the highest bidder, or reject all offers if none are deemed acceptable. If the highest bidder's offer is accepted, the dwelling is transacted in the manner of a Vickrey auction whereby the transaction price is equal to the second highest bid (plus a dollar) which most closely reflects the true market value of the home [15].

This bid-auction process results in a more realistic simulation of market transactions with respect to asking and transaction prices as well as the duration of agents' market activity. It further improves the dependence of market clearing on macroeconomic and land use trends as manifested in supply-demand interactions.

4. Model Performance and Validation

The HoMES module of ILUTE has been implemented in the C# .NET framework. Recent efforts to parallelize many of the computationally expensive algorithms have resulted in starkly improved performance; run-time for a twenty-year simulation of the GTHA's 1.5 million households has dropped from the order of days to approximately one hour as a result of streamlined and parallelized code.

Efforts are underway to comprehensively test and validate HoMES using data from the Toronto Real Estate Board (TREB) and the Canadian Mortgage and Housing Corporation (CMHC). Using a 100% population sample, the model's average asking and selling prices follow historical trends (see Figure 2a), though ILUTE selling prices tend to be slightly higher than asking, contrary to conventional market dynamics. Figure 2b compares cross-sectional distributions of transaction prices in 1987, and is indicative of systematic over-prediction in that year. Supply of new detached houses, illustrated in Figure 2c, generally exhibits strong correspondence with CMHC data. Results for other types of housing show similar temporal trends. While preliminary validation results are promising, challenges remain to attribute model inaccuracies to the appropriate sub-component. For example, skewed distributions of transaction prices may be a product of the bid-auction process, asking price formation, improper residential mobility determination or even population demographics. Such interdependencies are reflective of the integrated modeling paradigm and present challenges in model calibration.

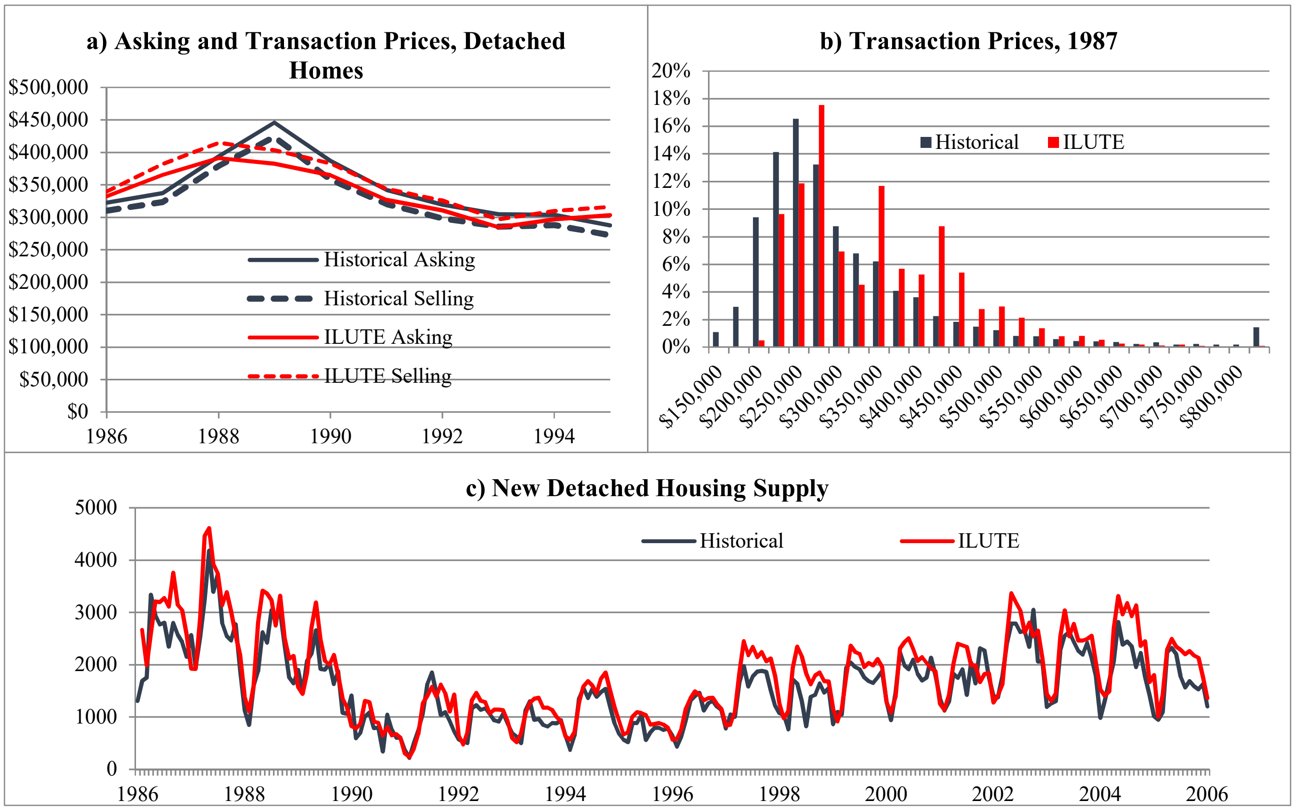

Figure 2. HoMES validation against TREB and CMHC data.

5. Conclusion

This paper has presented an agent-based implementation of urban housing market dynamics as part of the *Integrated Land Use, Transportation, Environment* model system. HoMES introduces some key features that offer marked improvement over previous models. These include: better integration among housing market sub-models, such as predictors of residential mobility becoming determinants of choice set formation; explicit modeling of the residential tenure decision; an improved asking price model more sensitive to endogenous micro and macro-economic factors; a willingness-to-pay framework of market clearing using a bid-auction process; and finally, a streamlined technical implementation with performance capabilities to rapidly execute full-population simulations.

Future work remains to be done to further incorporate the spatial attributes of location choice into the land use model. Furthermore, completed implementation of a rental housing market simulation would complete the HoMES framework. Commercial and industrial land use models are also a key necessity in modeling firmographic and labor force dynamics, and remain as priorities for the ILUTE modeling team.

References

[1] Chingcuanco, Franco and Miller, Eric J. Demographic Microsimulation Model for Integrated Land Use, Transportation, and Environment Model System. In 92nd Annual Meeting of the Transportation Research Board (Washington, D.C. 2013).

[2] Habib, M. A. and Miller, E. J. Influence of Transportation Access and Market Dynamics on Property Values: Multilevel Spatio-Temporal Models of Housing Price. Transportation Research Record: Journal of the Transportation Research Board (2008), 2076: 188-191.

[3] Habib, M. A. and Miller, E. J. Reference-Dependent Residential Location Choice Model within a Relocation Context. Transportation Research Record, Journal of the Transportation Research Board (2009), 2133: 56-63.

[4] Haider, Murtaza and Miller, Eric J. Modelling location choices of housing builders in the Greater Toronto Area. Transportation Research Record, 1898 (2004), 148-156.

[5] Farooq, Bilal and Miller, Eric J. Towards integrated land use and transportation: A dynamic disequilibrium based microsimulation. Transportation Research Part A: Policy and Practice, 46, 7 (2012), 1030-1053.

[6] Miller, Eric J., Chingcuanco, Franco, Farooq, Bilal, Habib, K.M.N., Habib, M.A., and McElroy, D. Development of an operational integrated urban model system, Volume IV: Demographic and labour market updating. Urban Transportation Research and Advancement Centre, University of Toronto (2008).

[7] Rossi, P. Why families move: A study in the social psychology of urban residential mobility. Free Press, Glencoe, Illinois, 1955.

[8] Henderson, J. V. and Ioannides, Y. M. A model of housing tenure choice. The American Economic Review, 73, 1 (1983), 98-113.

[9] Young, William. A non-tradeoff decision making model of residential location choice. Transportation Research, 18A (1984), 1-11.

[10] Lerman, S. R. A disaggregate behavioural model of urban mobility decisions. PhD Thesis, Cambridge, 1975.

[11] Michelson, W. Environmental choice, human behaviour, and residential satisfaction. Oxford University Press, New York, 1977.

[12] Pushkar, A. O. Modelling household residential search processes: methodology and preliminary results of an original survey. MASc Thesis, Toronto, 1998.

[13] McFadden, D. Conditional logit analysis of qualitative choice behaviour. Frontiers in Econometrics (1974), 105-142.

[14] Habib, M. A. Microsimulating residential mobility and location choice processes within an integrated land use and transportation modelling system. PhD Thesis, Toronto, 2009.

[15] Vickrey, William. Counterspeculation, Auctions, and Competitive Sealed Tenders. The Journal of Finance, 16, 1 (1961), 8-37.