A New Approach to Synthesize Heterogeneous Agents and their Associations for Urban Microsimulations

Bilal Farooq
Michel Bierlaire, & Gunnar Flötteröd

URBANICS II

March 26–29, 2012

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Contents

- Introduction
- Comment on state-of-the-art
- Problem (re)statement
- Methodology
- Application
  - Implementation
  - Initial results
- Discussion and working direction

Introduction

- Urban Microsimulation: Forecasting behaviour using individual level models
  - Lack of individual level data for population
  - Synthesis of individual agents and their characteristics
- Initial work
  - Focused on synthesis of a small sub-set of characteristics
    - Usage: Activity-base travel demand models, etc.
- Frequently used approaches
  - Fitting based
    - Combinatorial optimization
    - Iterative proportional fitting

Comments on Existing Approaches

- Key issues with the existing approaches
  - Cloning of data rather than creation of a heterogeneous representative population
  - Over reliance on the accuracy of the microdata, without serious consideration to the sampling process and assumptions
  - Optimization resulting in one realization of synthetic population
  - Scalability issues

Problem re-statement

- We are interested in building a “joint distribution” of the population from which one or more realizations of synthetic population can be created, such that
  - Best representation of the real population
  - Synthetic population having a “continuous heterogeneity” rather than “discretized cloning”
  - Population synthesis process as a part of the microsimulation
  - Methodology does not need to know the data collection and aggregation process

Definition

- Synthetic Agents
  - Households, persons, families and their association
- Space/location
  - Zone: parcel/sector/commune, dissemination area/tract/sub-division, traffic analysis zone/planning district
- Time
  - Base year, *simulation year*
- Data
  - Aggregate: zonal level totals and cross-tables
  - Disaggregate: sample of individuals, usually without location [Public Use Microdata Sample (PUMS)]
    - Public Use Microdata Area (PUMA)

Methodology

- Synthetic Agents: Persons (X), Households (H)
  - X and H defined by their characteristics
- Associations (C)
- Obtain the posterior distribution for $\pi(X, H, C)$
  - Markov Chain Monte Carlo Simulation (MCMC)
    - Simulate the direct draws from the distribution
  - Sampling methods
    - Metropolis Hasting sampling
    - Gibbs sampling
  - Creating a pool of agents which is the un-normalized representation of $\pi(X, H, C)$
    - Synthetic population as a realization

Methodology

- Available Data
  - Data on persons characteristics Y_X
    - $\pi(X|Y_X)$
  - Data on households characteristics Y_H
    - $\pi(H|Y_H)$
  - Data on association characteristics Y_C
    - $\pi(C^i|Y_C,X^i,H^i)$

Methodology: Persons

- Persons synthesis ($X|Y_X$)
  - Method: Gibbs sampling
    - Conditionals for person characteristics known to certain extent
    - Run a MCMC simulation to generate the persons
    - Results in an infinite pool of feasible persons
    - Realization from this universe will result in the synthetic population of persons

Methodology: Persons

- X = {Age, Sex, Marital_Status, Dwell_Type}
- Data needed: Zone (sector, commune) level conditionals
  - E.g. ***P***(Age | Marital_status, Sex, Dwell_Type), ***P***(Marital_status | Age, Sex, Dwell_Type), ***P***(Sex | Age, Marital_status, Dwell_Type), ***P***(Dwell_Type | Age, Marital_status, Sex)
- Each iteration for X_t
  - Randomly pick a characteristic and realize its value from its conditional based on the other characteristics of X_{t-1}

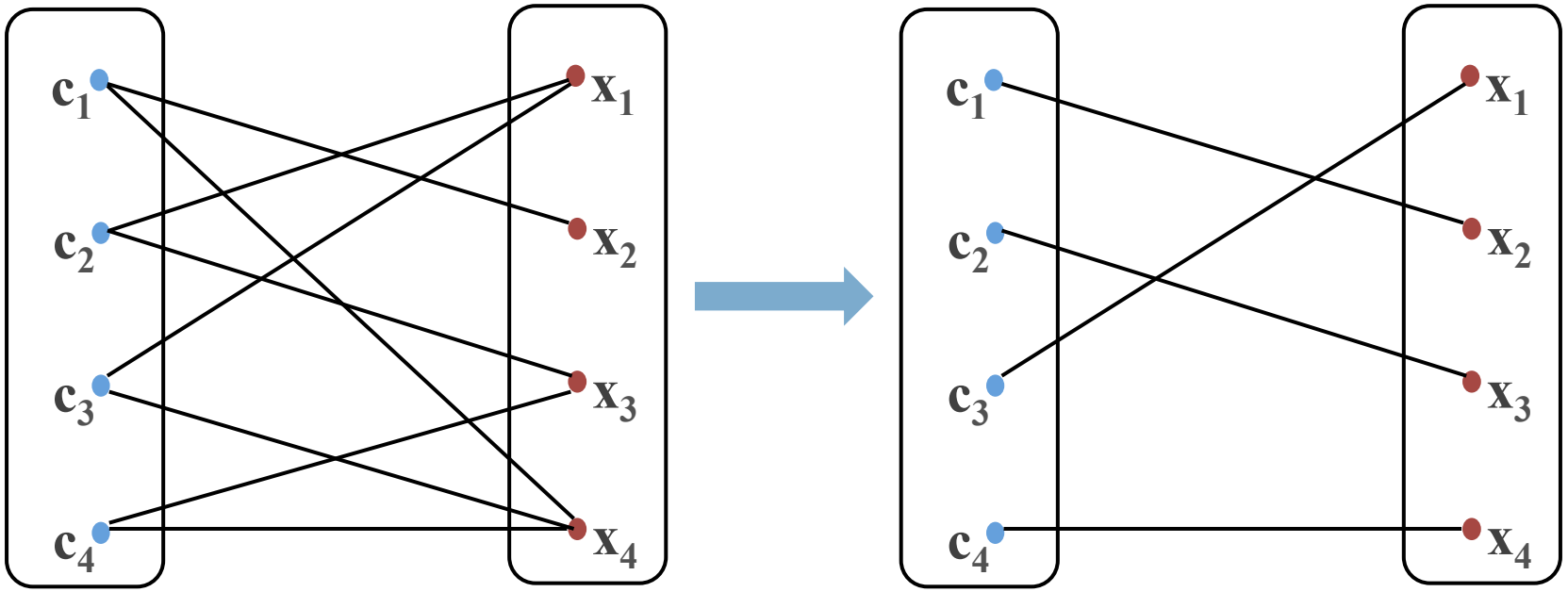
Methodology: Household

- Household synthesis (H)
  - Method: Gibbs sampling
  - Data needed: Zone (sector, commune) level conditionals
    - E.g. P(Hhld_Type | Veh_Count, Dwell_Tenure, Dwell_Type), P(Veh_Count | Hhld_Type, Dwell_Tenure, Dwell_Type), P(Dwell_Tenure | Hhld_Type, Veh_Count, Dwell_Type), P(Dwell_Type | Hhld_Type, Veh_Count, Dwell_Tenure)
  - Synthetic households
    - Resources
    - Encapsulation of positions
      - Realization as list of positions

Methodology: Associations

- Associations (C)
  - Matching Persons to Positions
    - Head of household, workers, children, adults
  - For each realization a distribution of association is computed that is based on the available microdata
    - Minimizing the difference in count with the microdata

Methodology: Association



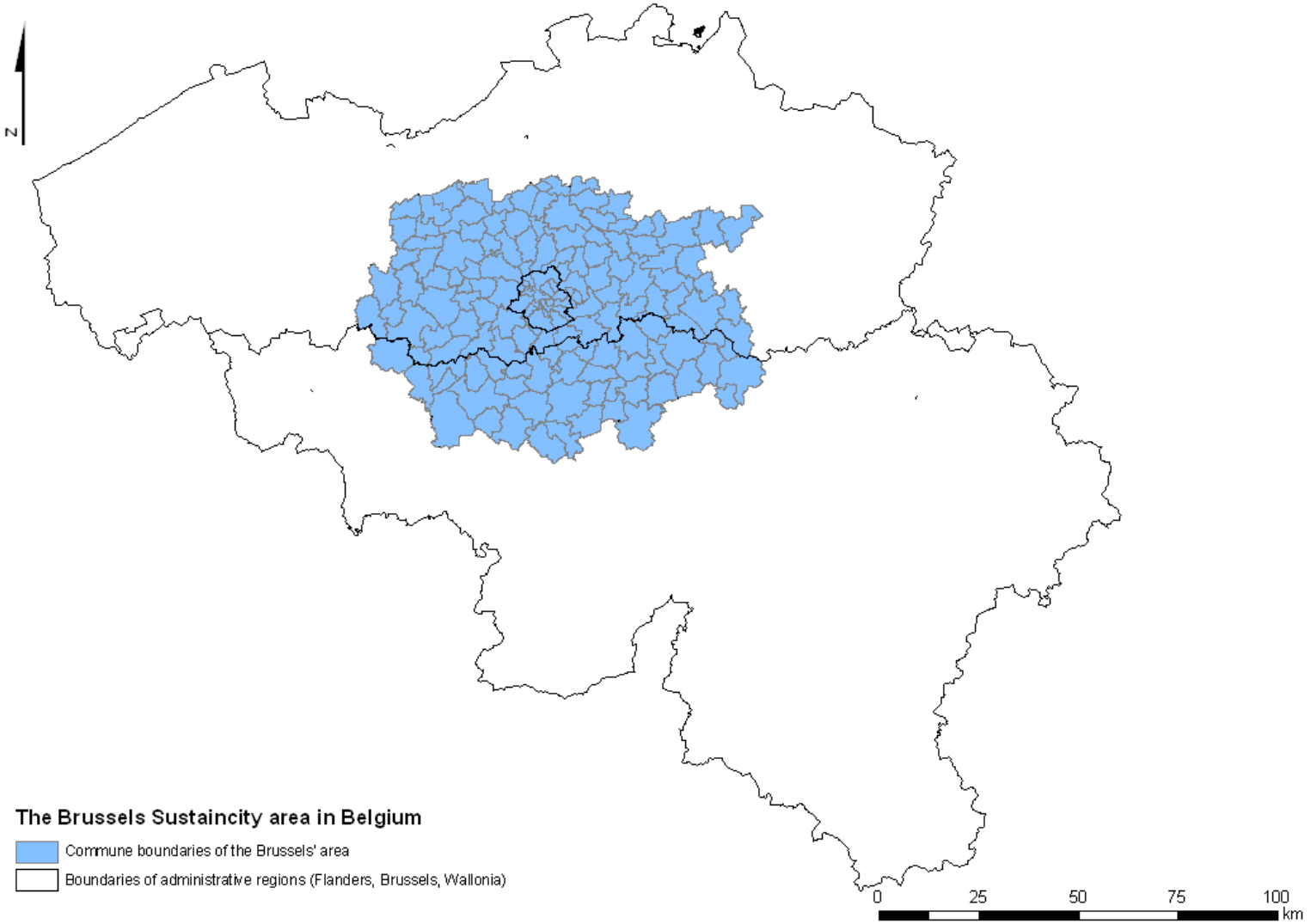
Methodology

- Method: Metropolis-Hasting sampling
  - State: A valid assignment
    - Examples: a two years old is not the head of household
  - Proposal matrix/function
    - Defined in terms of switching the association of two persons with each other (bi-directional)
  - Transition/proposal distribution
    - Acceptance rate (awarding good states and penalizing bad)
  - Initialized to certain random state of association
  - A realization from the distribution

Application: Greater Brussels Area

- SustainCity Project
  - European Union funded research project
  - 12 major European universities involved
  - Aims:
    - Address the modelling and computational issues of integrating modern mobility simulations with the latest microsimulation land use models
    - Demographics, environment, and multi-scale issues
  - Case studies
    - Paris
    - Zurich
    - Brussels

Application: Greater Brussels Area



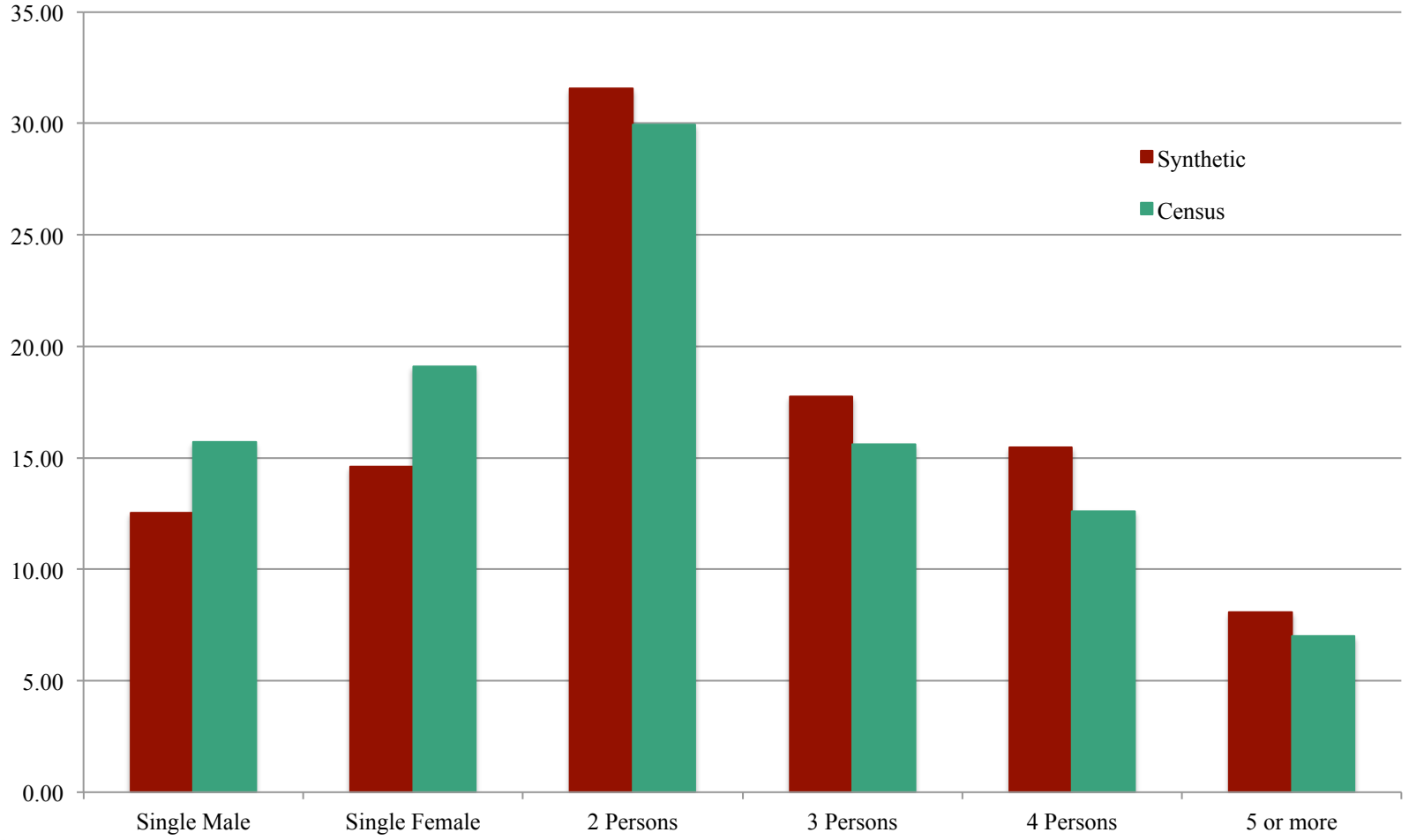
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

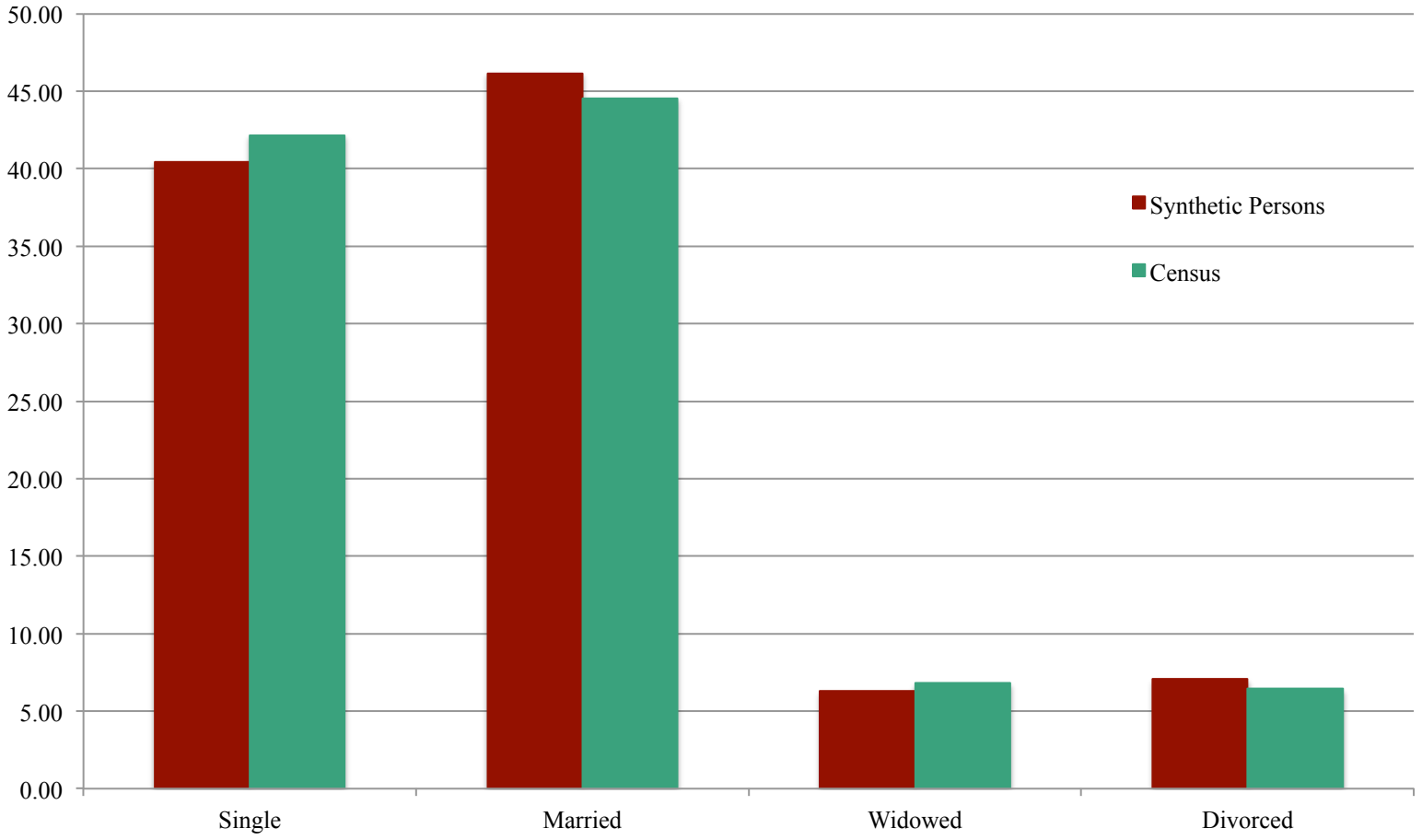
Application

- Synthetic population for Brussels region
  - Data sources (extremely limited)
    - Zonal conditionals of households and persons (Census 2001)
      - Incomplete conditionals
    - Travel survey of households and individuals (MOBEL 1999)
      - 3063 observations (0.2%)
  - Data Preparation
    - Aggregation
      - Spatial
      - Categorical
    - Model based conditionals (Logit)
      - Income and Education level
  - Pool of 100 million households
    - Realization: ~1.2 million

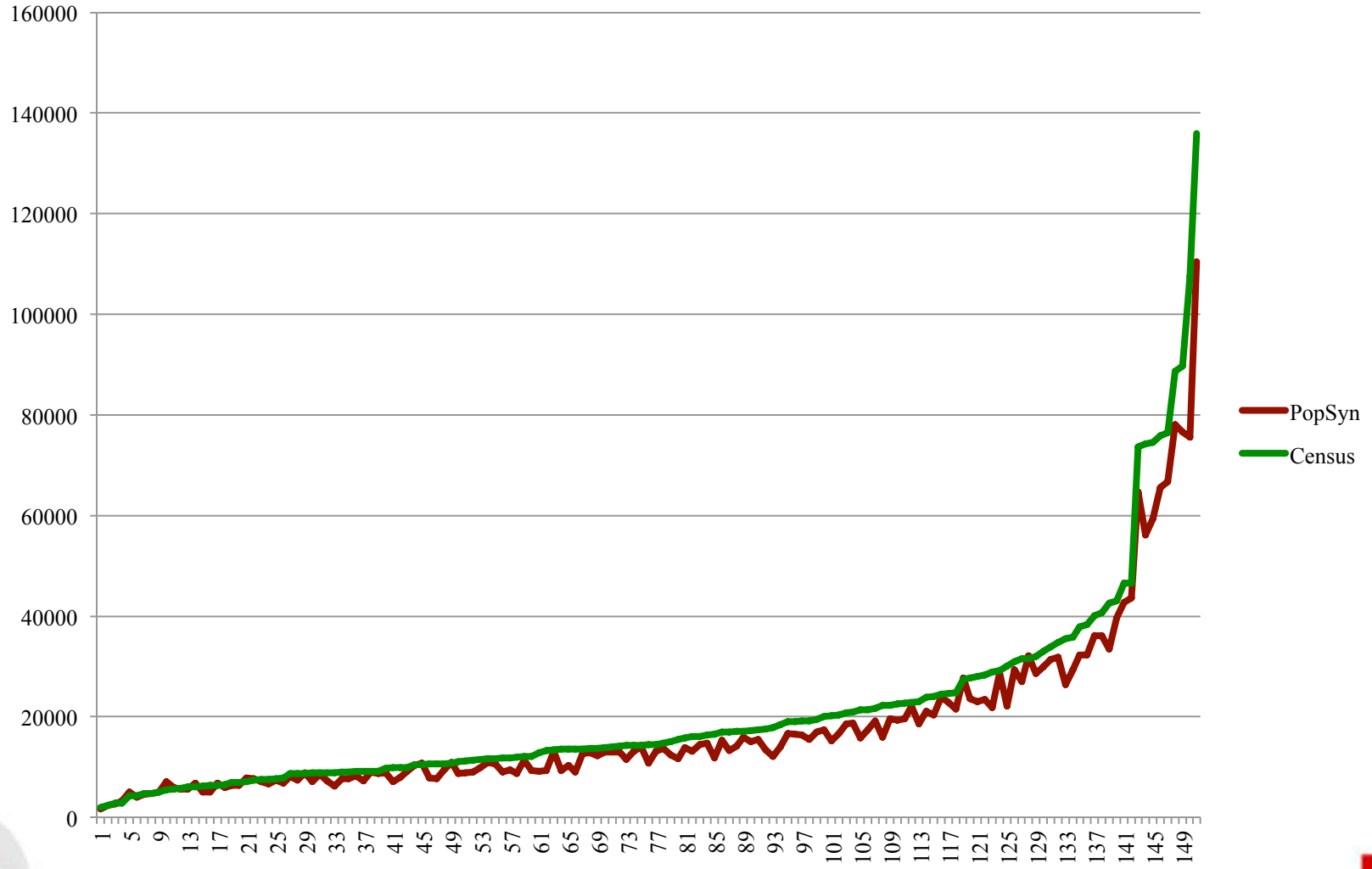
Household Size (Brussels, 2001)



Persons' Marital Status (Brussels, 2001)



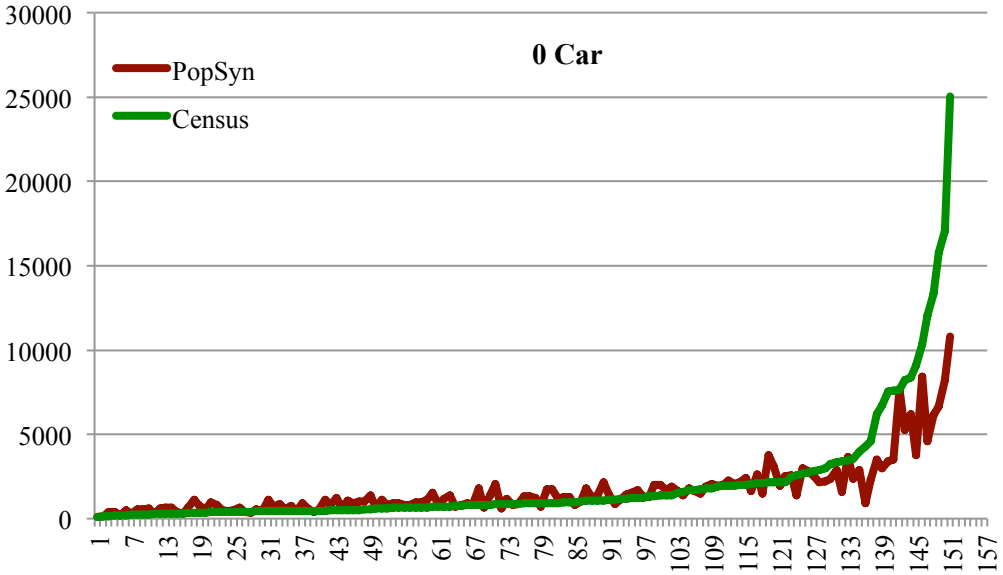
Number of Persons per Commune

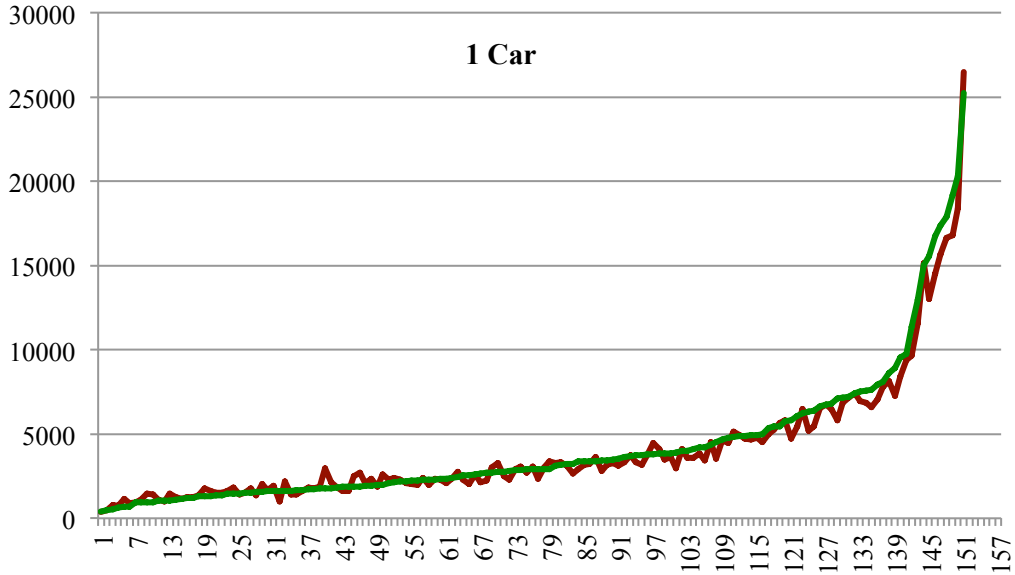


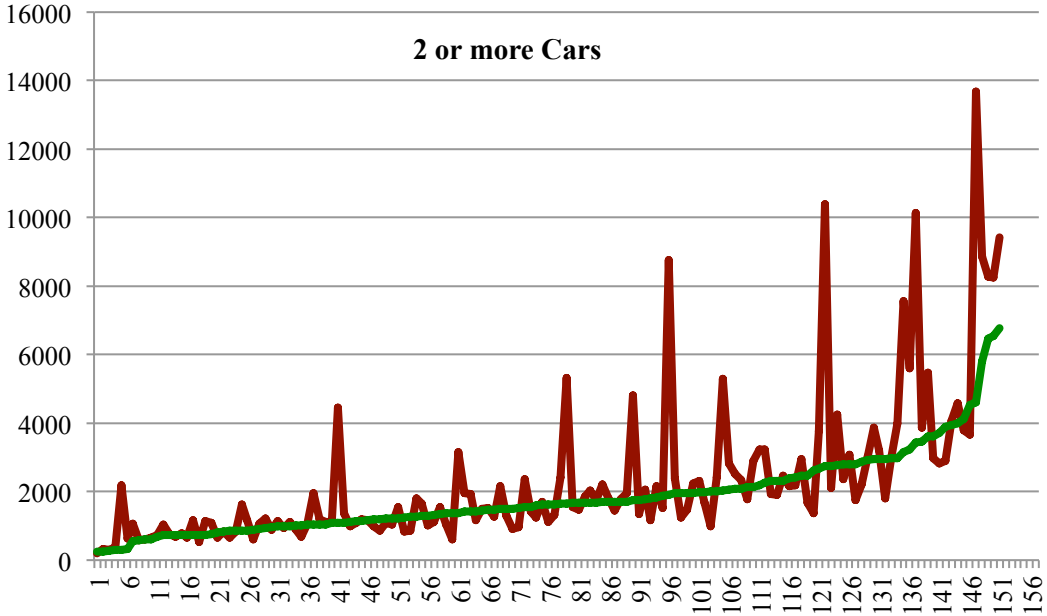
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

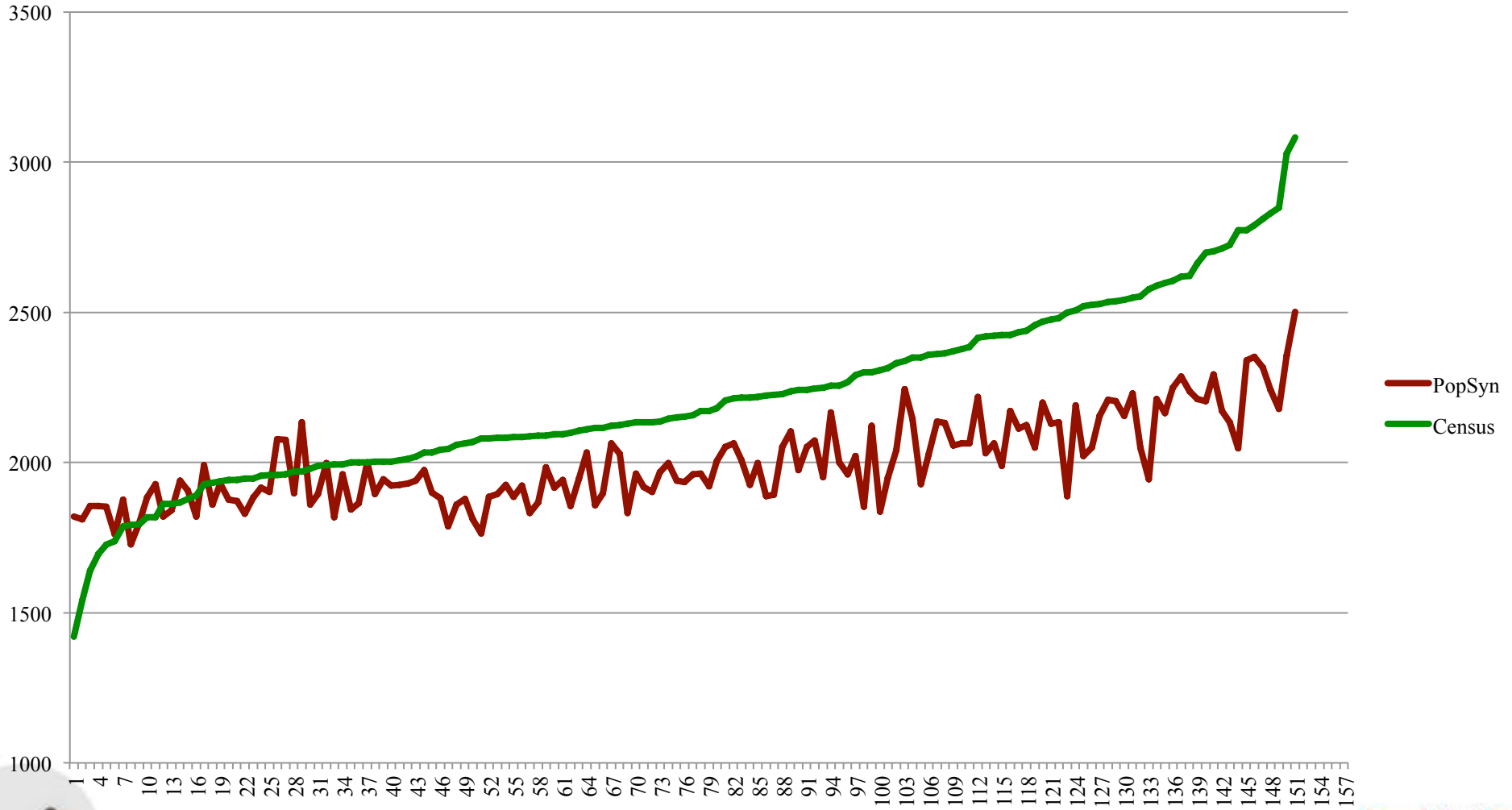
Number of Households by Cars per Commune







Average Income (Euro) by Commune



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Discussion

- From optimization to sampling from joint distribution
- Mix of sampling process can be utilized based on the situation
- Population synthesis as part of microsimulation
  - Sensitivity analysis
- Separation of data preparation from agent generation
  - Data, models, assumptions
- Works both for continuous and discrete or mixture of conditionals
- Computationally efficient and scalable
  - Clean and simple

Working Direction

- More detailed disaggregate level spatial and statistical analysis
- Using Swiss census
  - Compare the proposed methodology with other approaches
  - Study the effects of degree of incompleteness of distributions
  - Effect of sample sizes

Questions/Comments