Simulation based Approach for Agents Synthesis in Large-Scale Urban Systems Modelling

Bilal Farooq, Michel Bierlaire, Ricardo Hurtubia & Gunnar Flötteröd

IATBR 2012

TRANSP-OR

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Contents

- Motivation
- Problem (re)statement
- Methodology
- Experiments with a real population
- Application in case of very limited data
- Discussion and concluding remarks

Motivation

- 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 representativeness and accuracy of the microdata for retrieving conditional distribution
    - While fitting focused on matching marginals only
  - 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

Methodology

- Retrieving the underlying joint distribution of the Population using **Simulation** techniques
  - Markov Chain Monte Carlo processes
    - Simulate the direct draws from the distribution
  - Sampling methods
    - Metropolis-Hasting sampling
    - Gibbs sampling
  - Creating an infinite pool of agents which is the un-normalized representation of the underlying distribution
    - Synthetic population as a realization

Methodology

- Retrieving the underlying joint distribution of the Population using **Simulation** techniques
  - Markov Chain Monte Carlo processes
    - Simulate the direct draws from the distribution
  - Sampling methods
    - Metropolis-Hasting sampling
    - **Gibbs sampling**
  - Creating an infinite pool of agents which is the un-normalized representation of the underlying distribution
    - Synthetic population as a realization

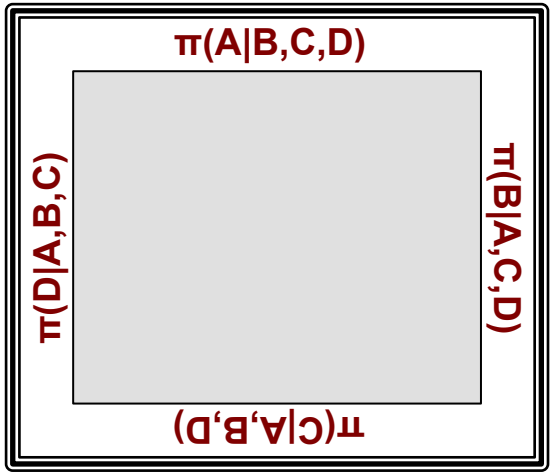
Gibbs sampling

π(A,B,C,D)??

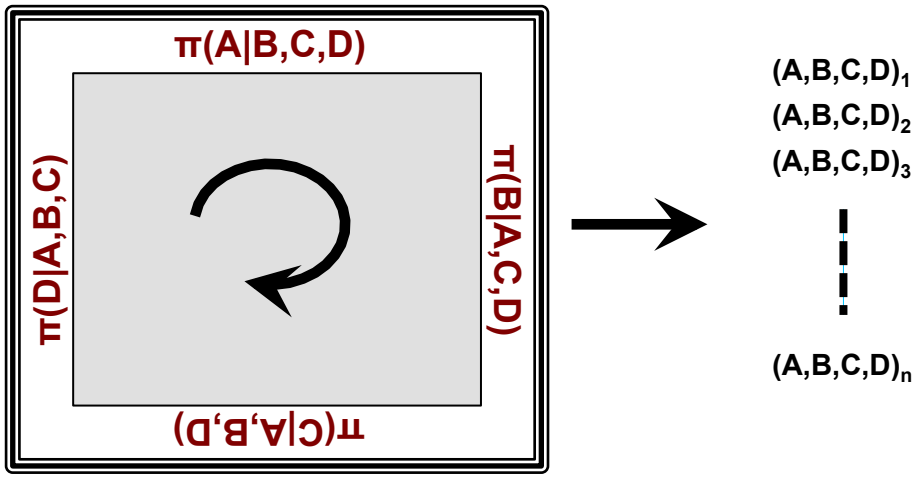
π(A|B,C,D)
π(B|A,C,D)
π(C|A,B,D)
π(D|A,B,C)

Gibbs sampling

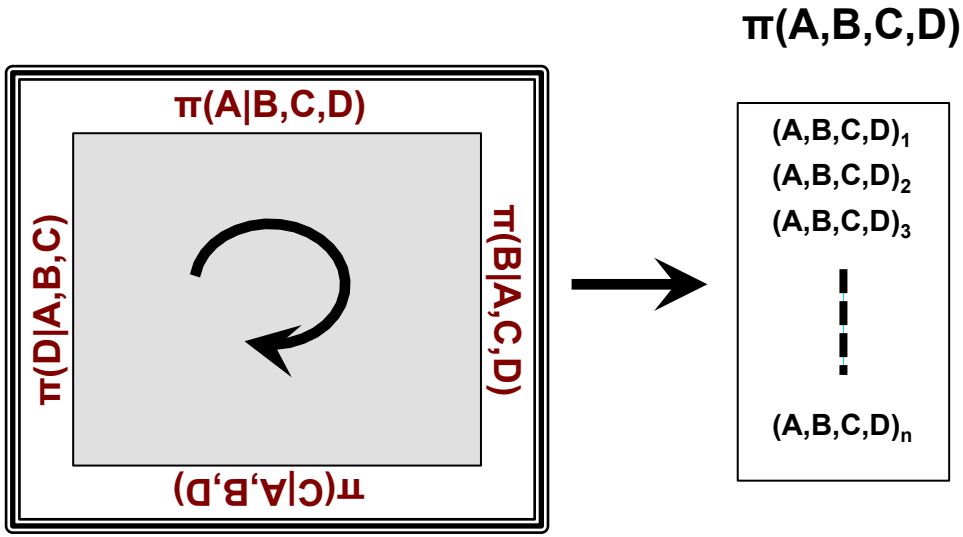
π(A,B,C,D)??



Gibbs sampling



Gibbs sampling



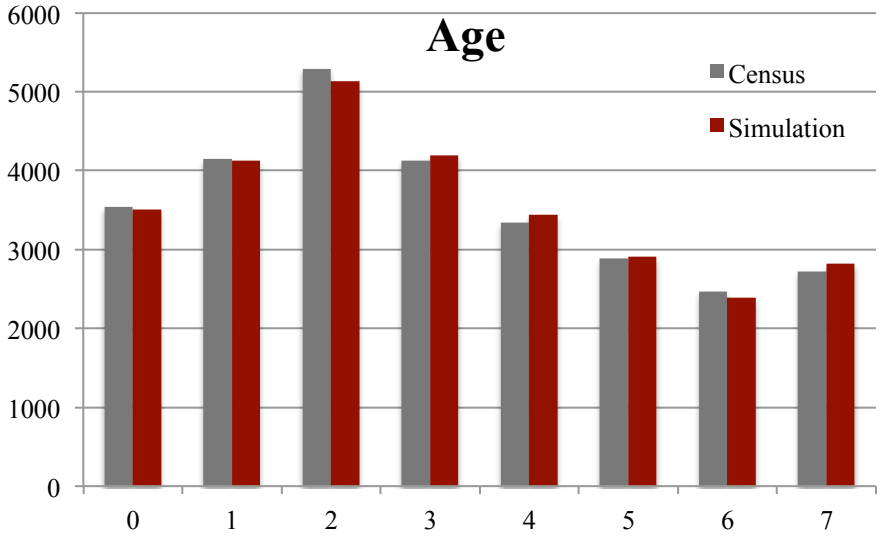
Experiments with Real Population

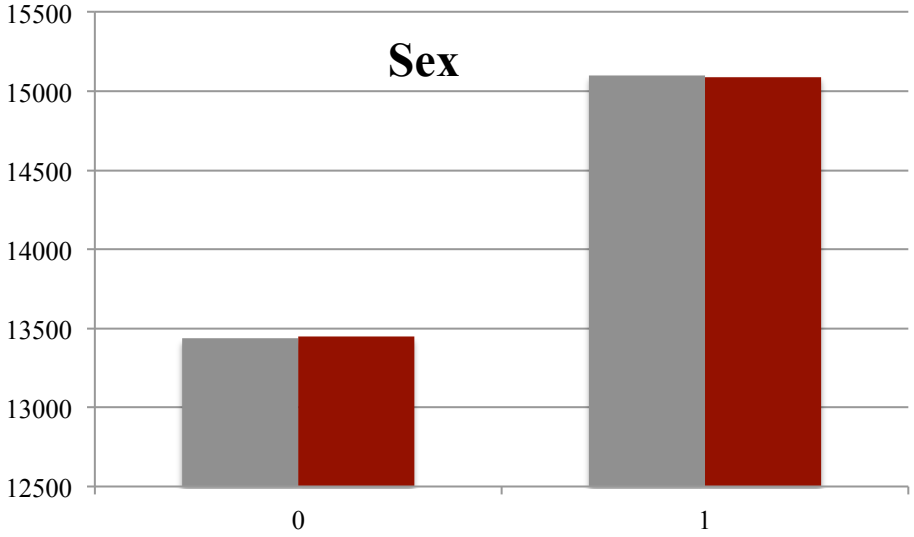
- Swiss census 2000
  - Person and household attributes (Except for Income)
- Postal code in Lausanne
  - CH-1004
  - 28,533 persons
- Four *Person* attributes
  - Age *(<15, 15-24, 25-34,...,>74)*
  - Sex *(Female, Male)*
  - Household size *(1, 2, 3, 4, 5, 6 or more)*
  - Education level *(none, primary, secondary, university/college)*

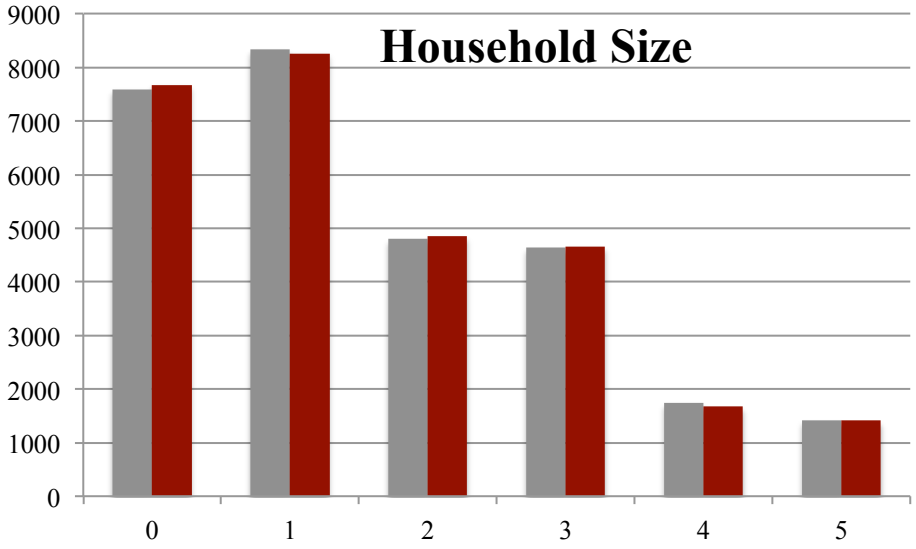
Experiments with Real Population

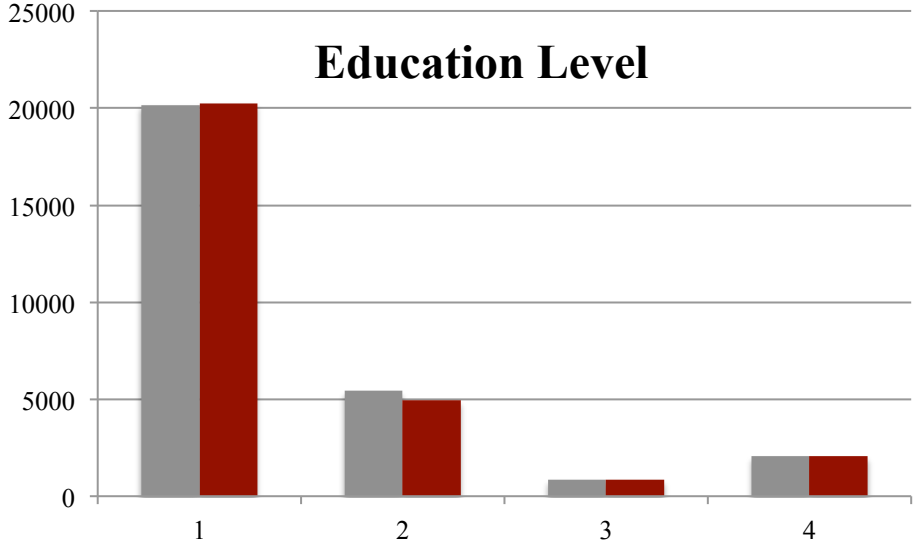
- Comparisons with other methods
  - Criteria
    - Not just marginals retrieval, but also the joint distribution
  - Simulation
    - Effect of incompleteness of conditionals
  - IPF
    - Effect of change in sample size

Simulation based Marginals

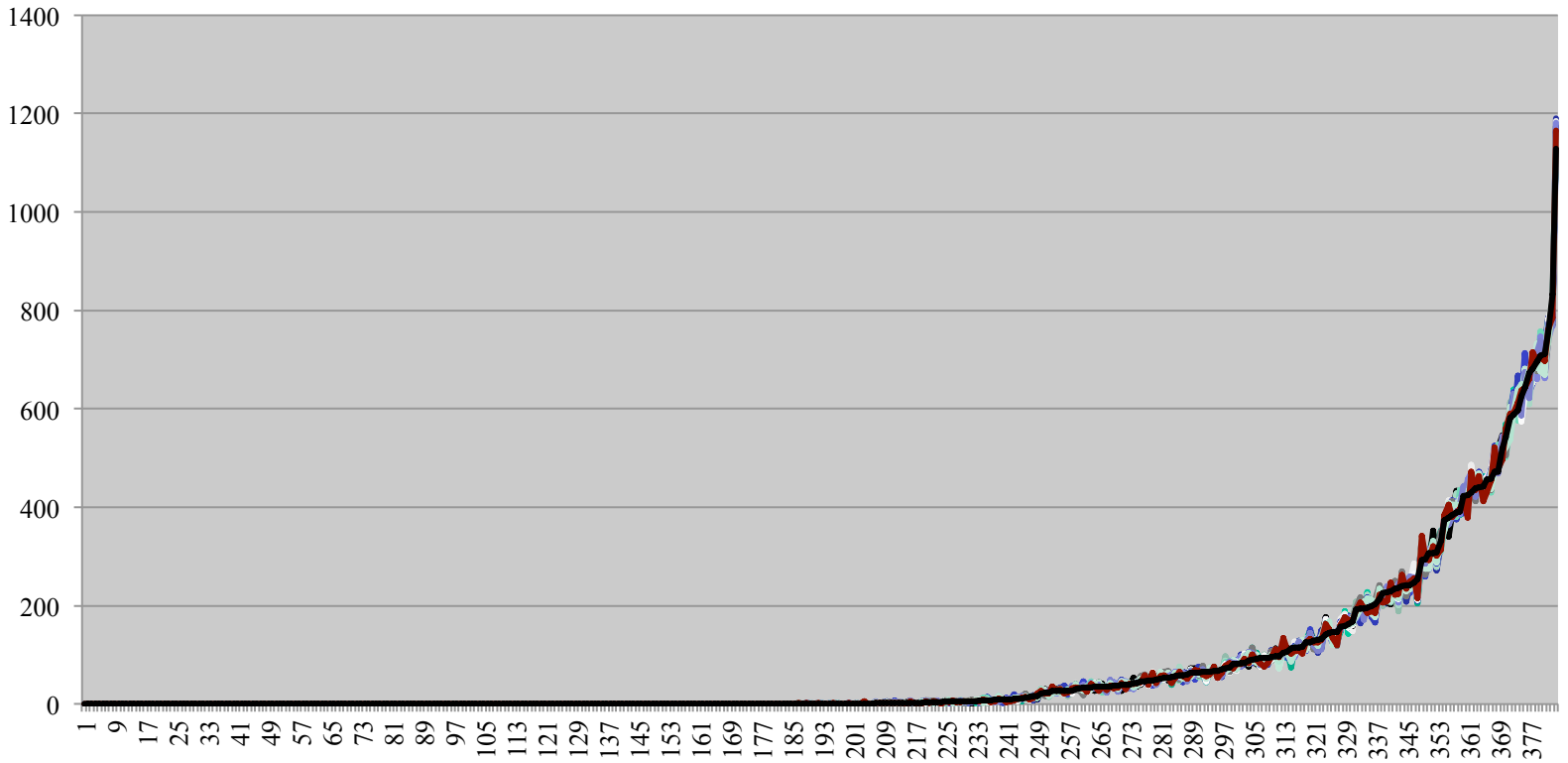








Simulation based Synthesis with Full Conditionals

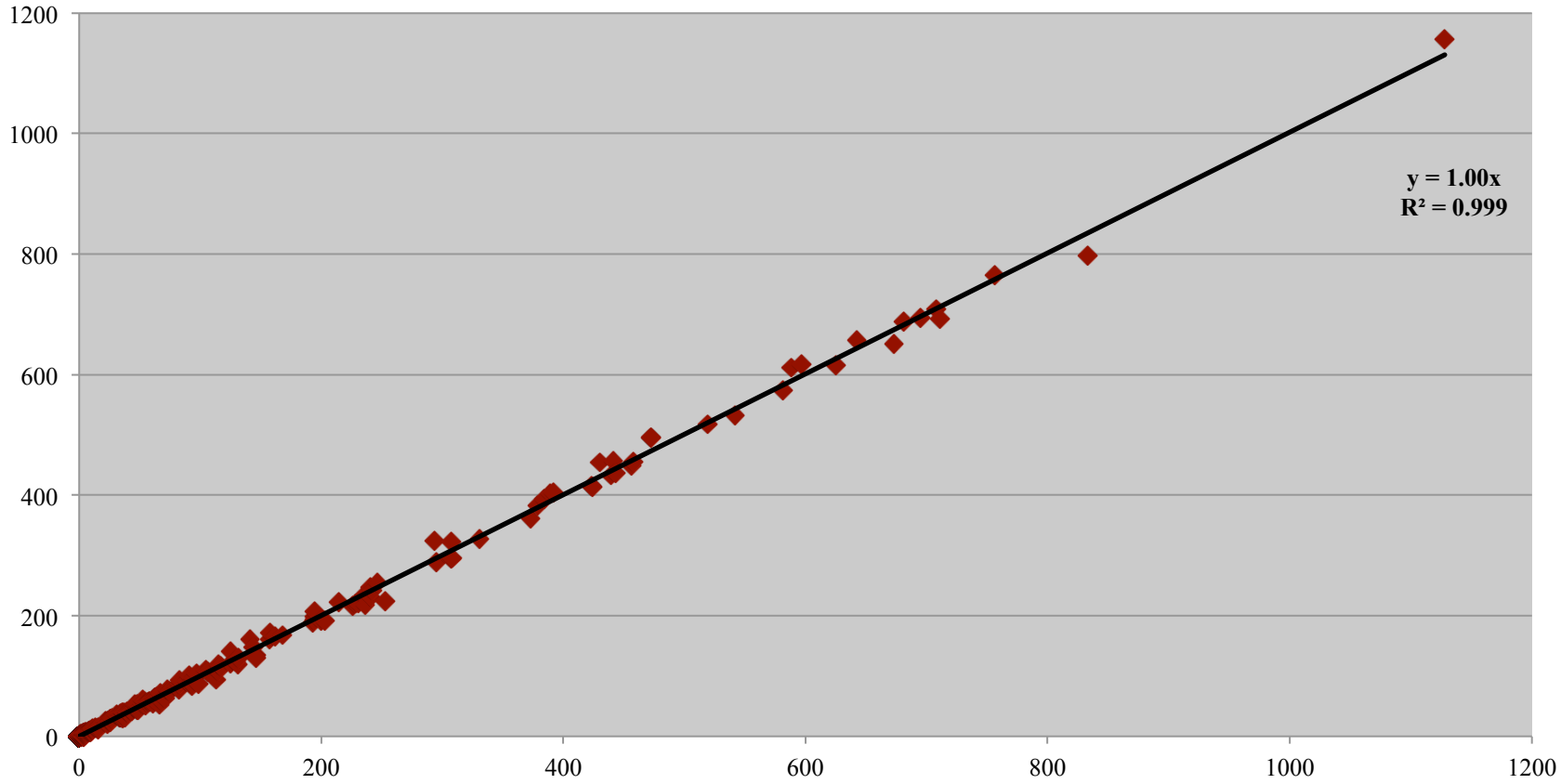


Joint distribution for Age(8), Sex(2), Household size(6), and Education level(4)
(With 20 Simulation Runs)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Simulation based Synthesis with Full Conditionals

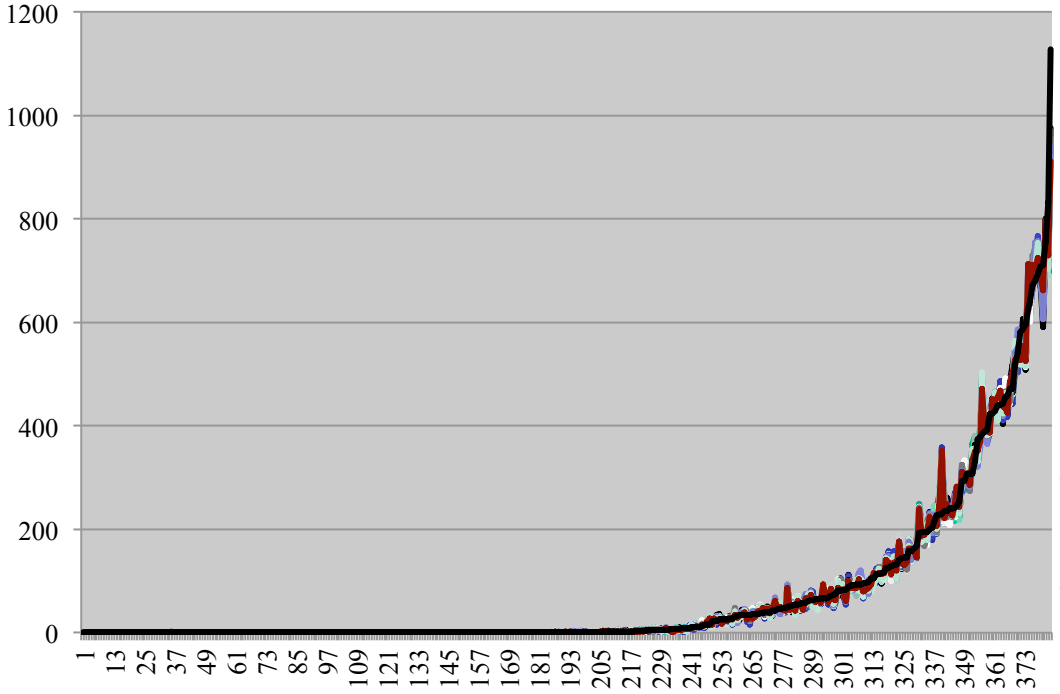


Fit of the joint distribution with respect to census data
(Averaged out of 20 Simulation Runs)

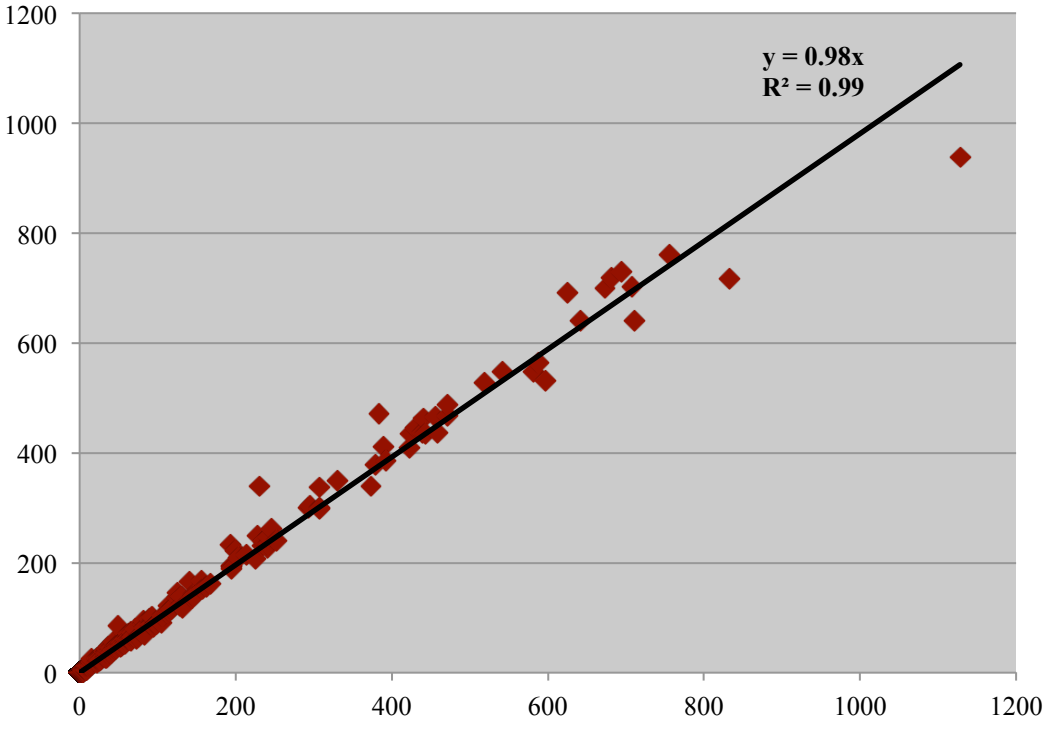
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

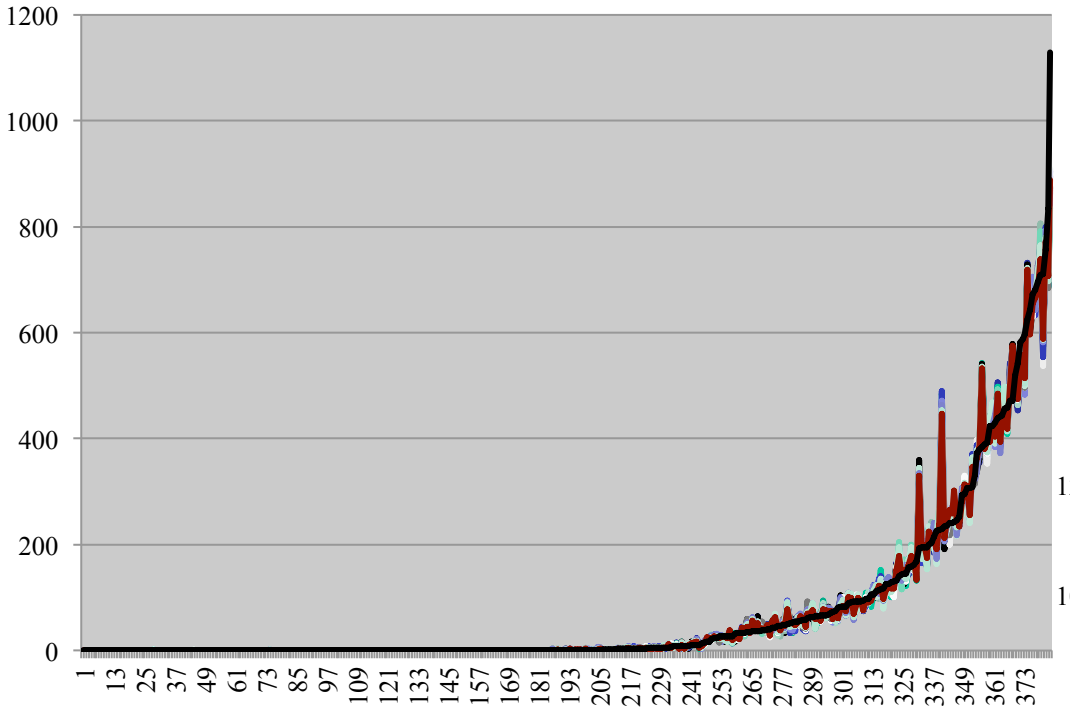
Simulation Results (Partial Conditionals)



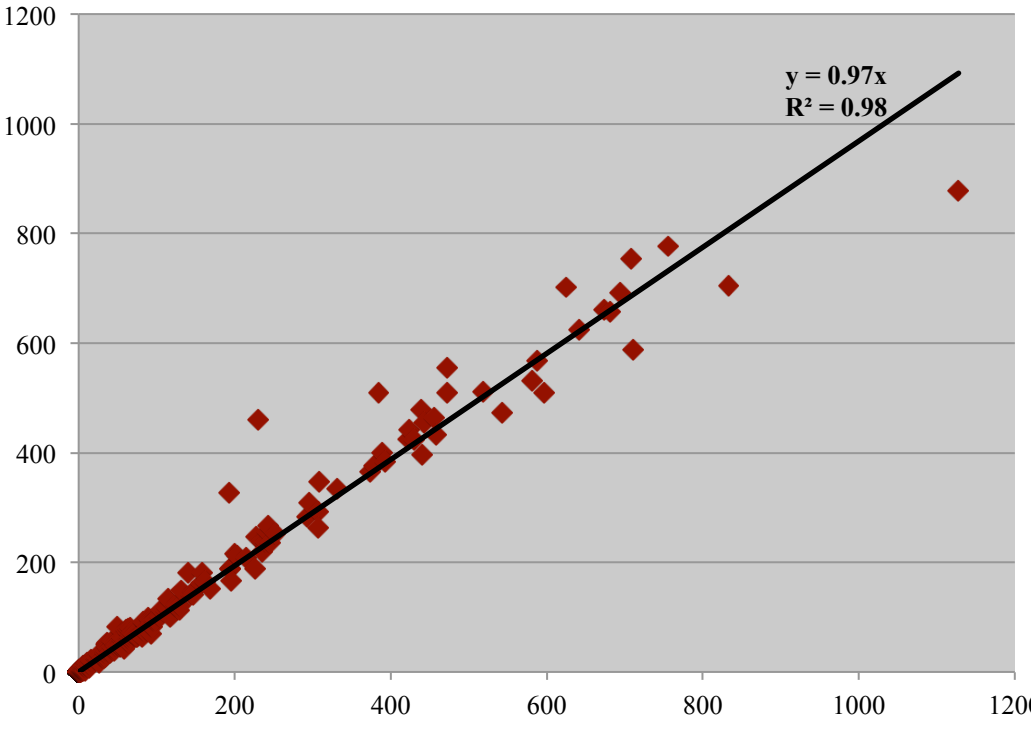
π(Age|*,Hhld_Size,Edu_Lvl)
π(Sex|Age,Hhld_Size,Edu_Lvl)
π(Hhld_Size|Age, Sex,Edu_Lvl)
π(Hhld_Size|Age, Sex,Edu_Lvl)



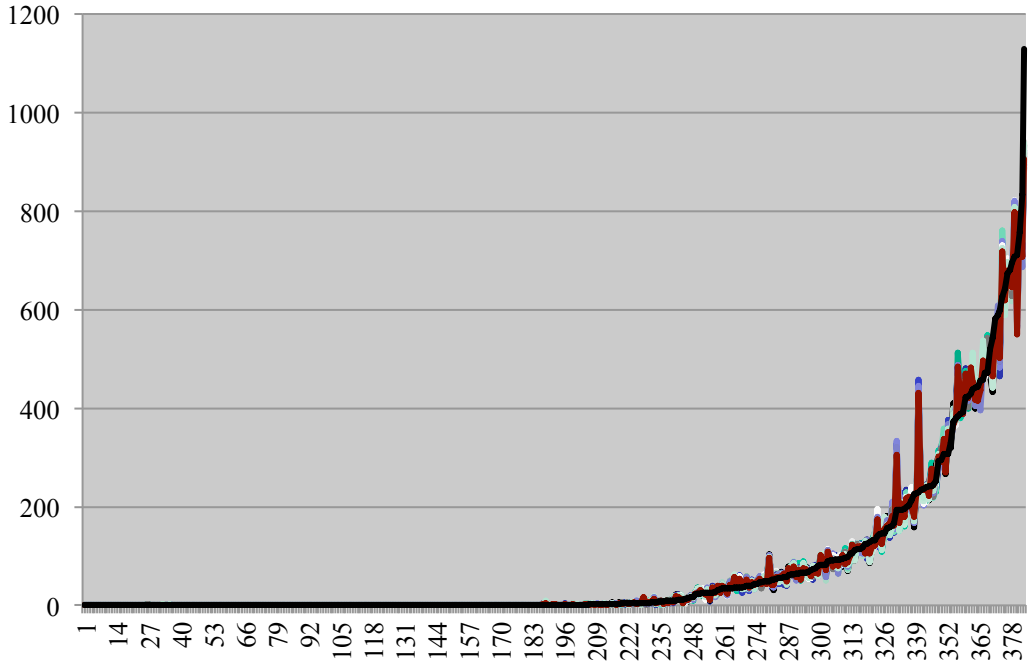
Simulation Results (Partial Conditionals)



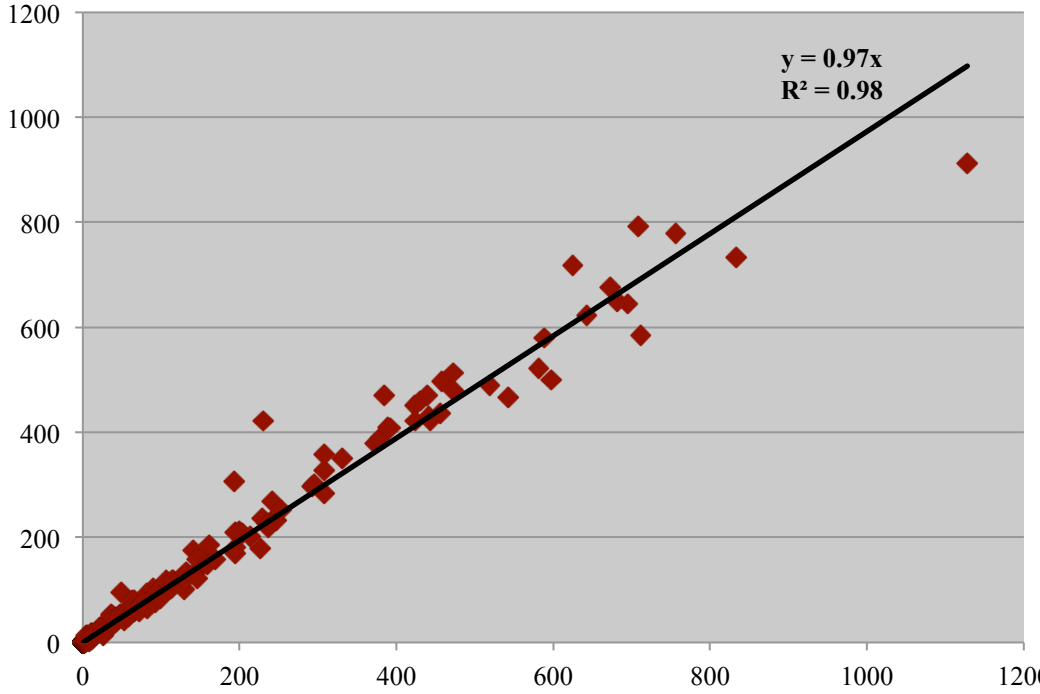
π(Age|*,Hhld_Size,Edu_Lvl)
π(Sex|Age,Hhld_Size,Edu_Lvl)
π(Hhld_Size|Age,*,Edu_Lvl)
π(Edu_Lvl|Age, Sex, Hhld_Size)



Simulation Results (Partial Conditionals)



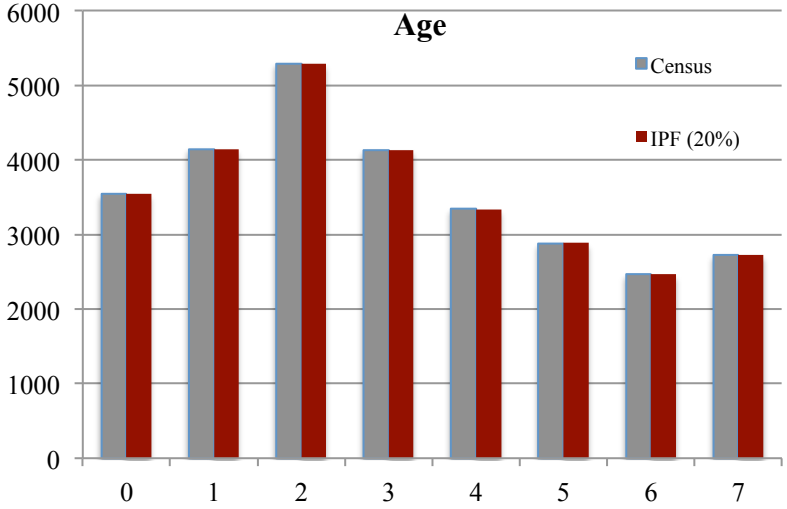
π(Age|*,Hhld_Size,Edu_Lvl)
π(Sex|Age,Hhld_Size,Edu_Lvl)
π(Hhld_Size|Age,*,Edu_Lvl)
π(Edu_Lvl|Age, *, Hhld_Size)

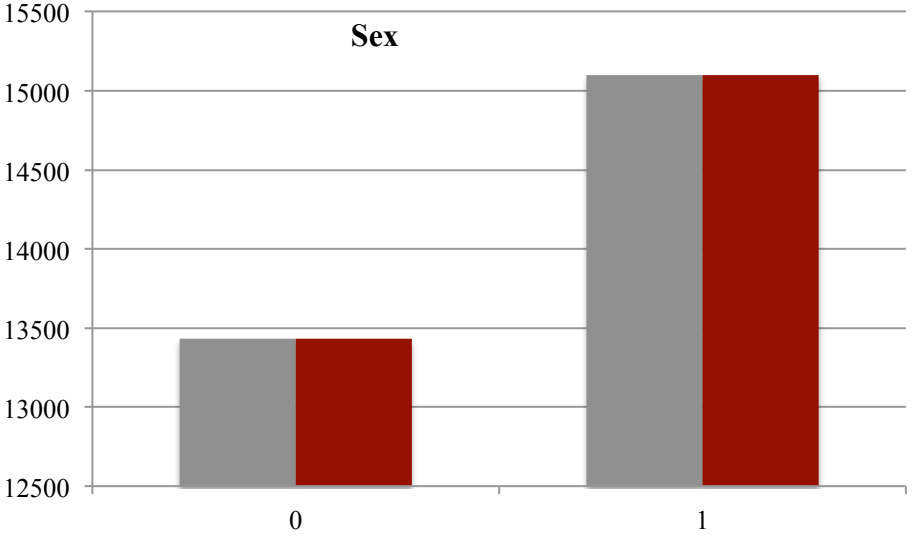


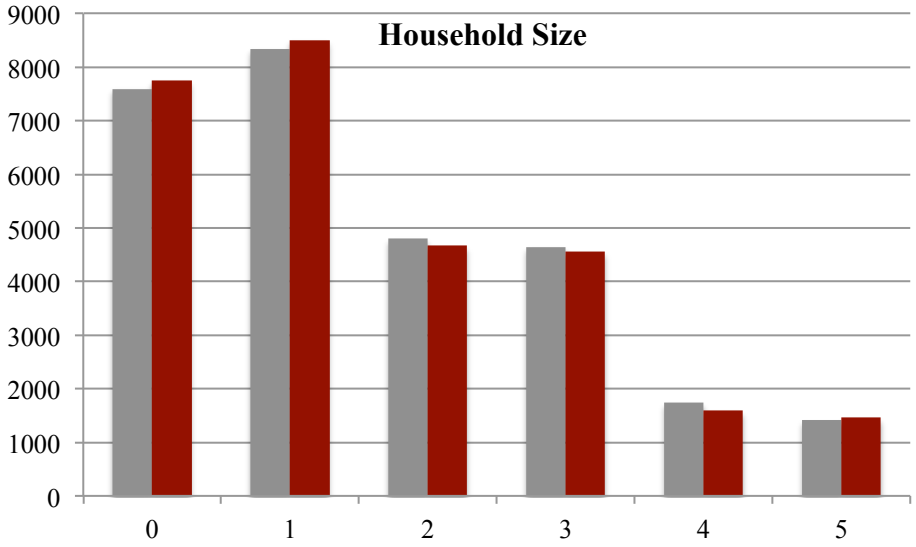
Iterative Proportional Fitting

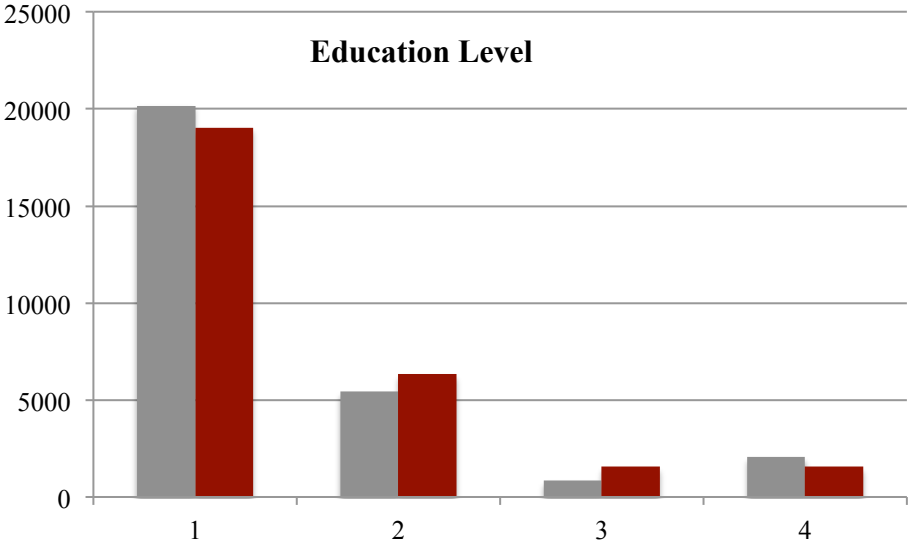
- Two steps
  - Fitting process
  - Cloning processes
    - MC process for the fractions
- Same conditionals
  - Converted to Marginals
- Additionally: randomly selected sample
  - Size: 20% to 3%

IPF Marginals

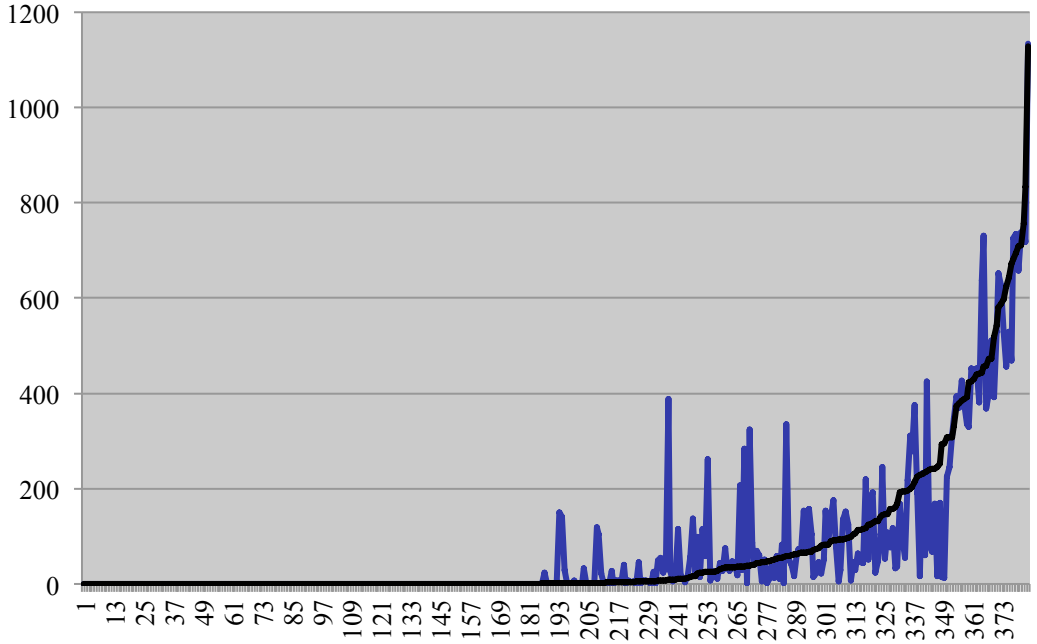


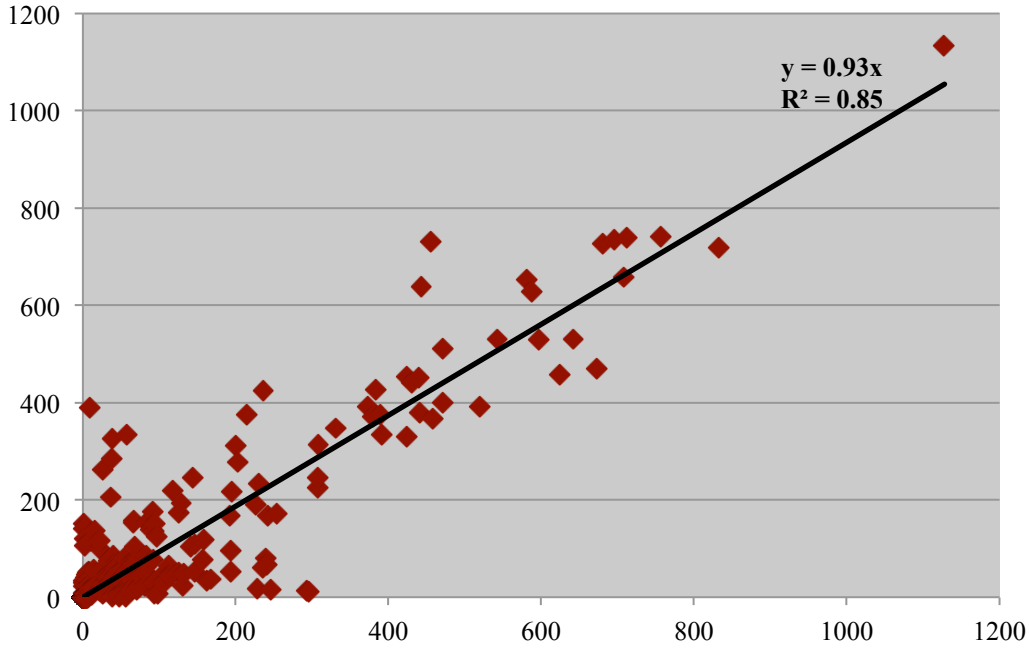




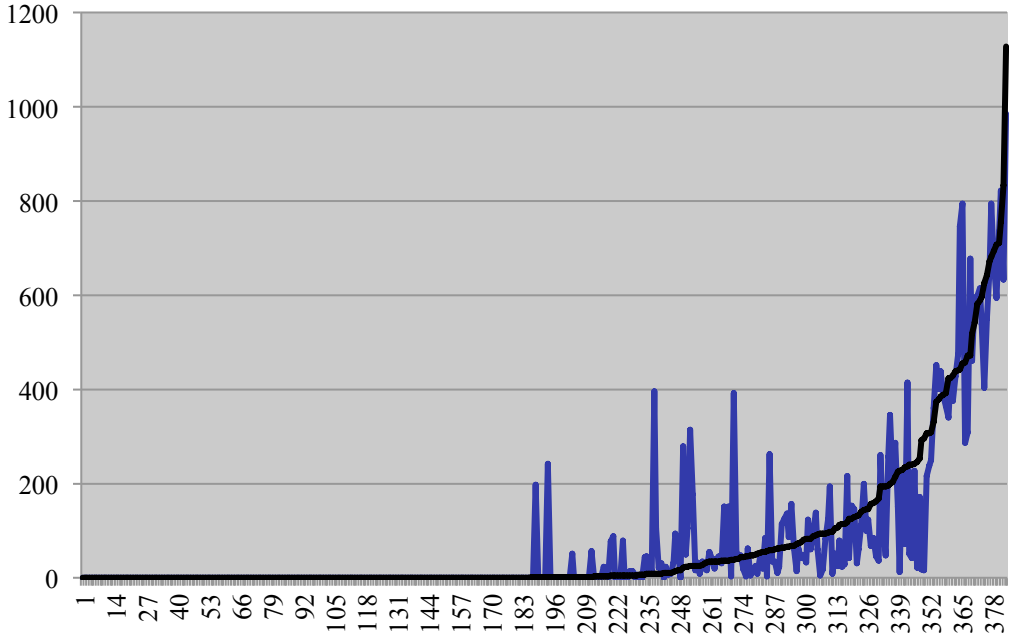


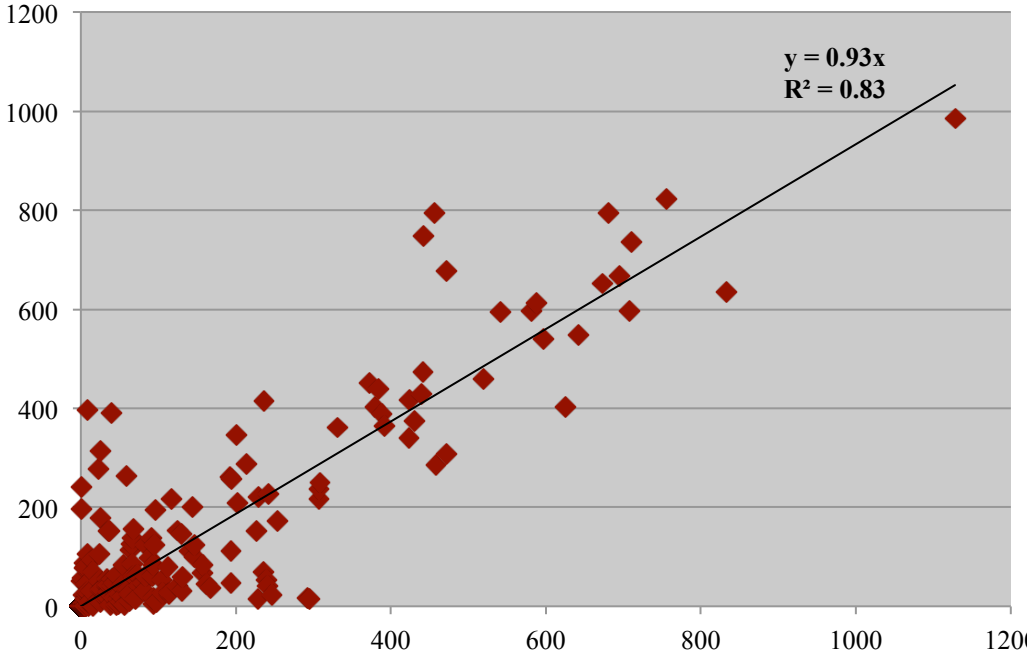
Iterative Proportional Fitting (Full Marginals and 20% Sample)



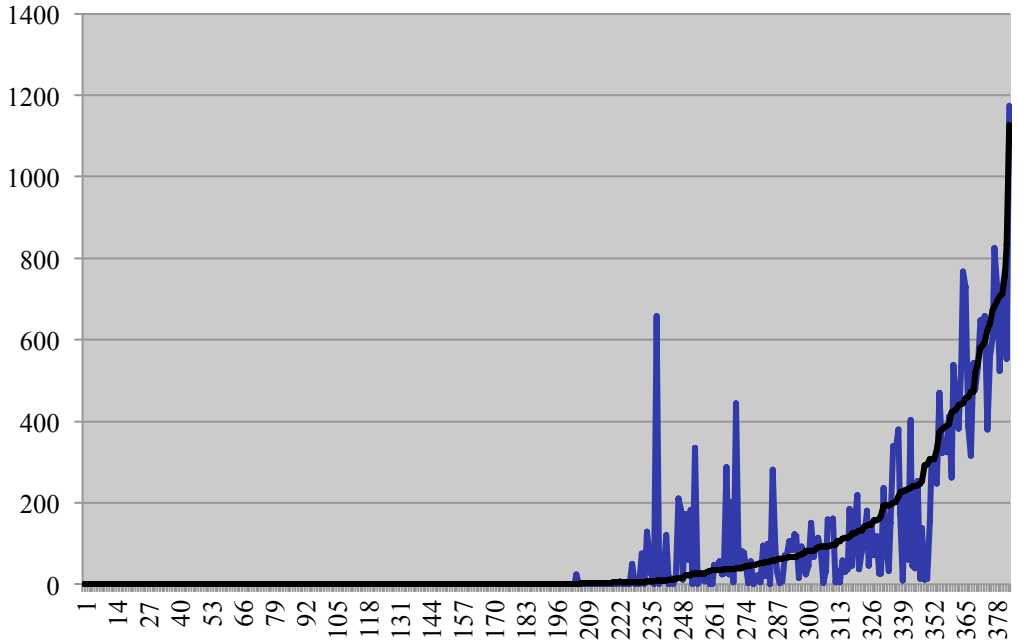


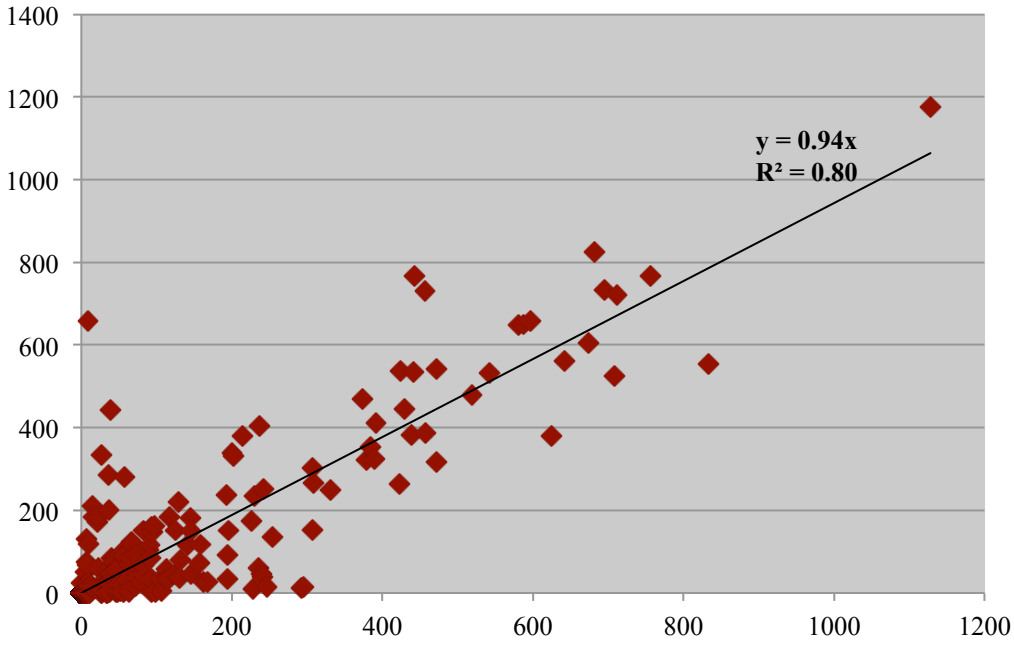
Iterative Proportional Fitting (Full Marginals and 10% Sample)



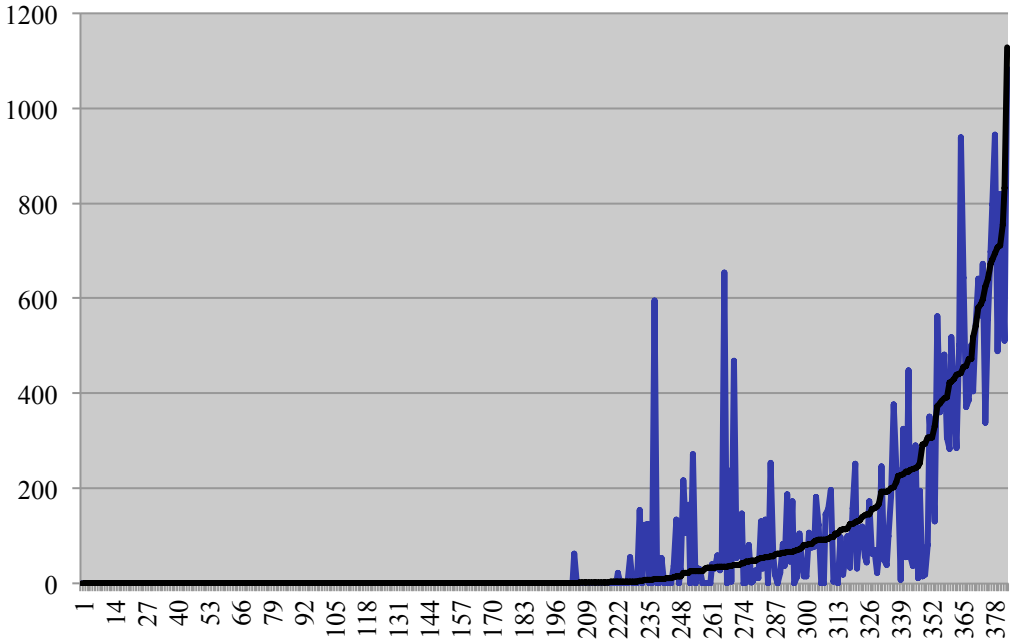


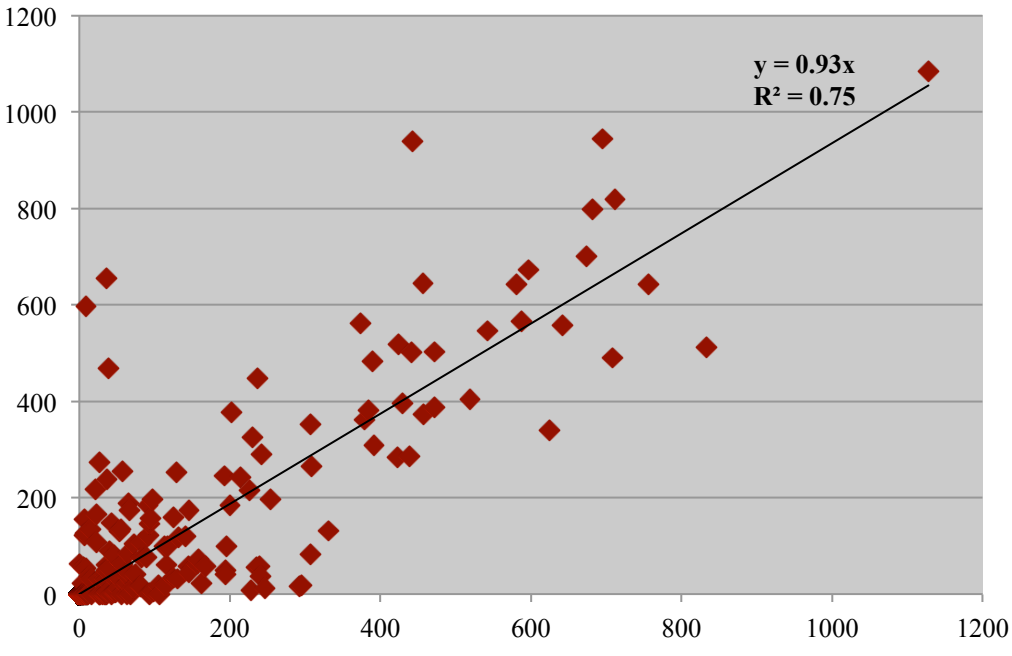
Iterative Proportional Fitting (Full Marginals and 5% Sample)



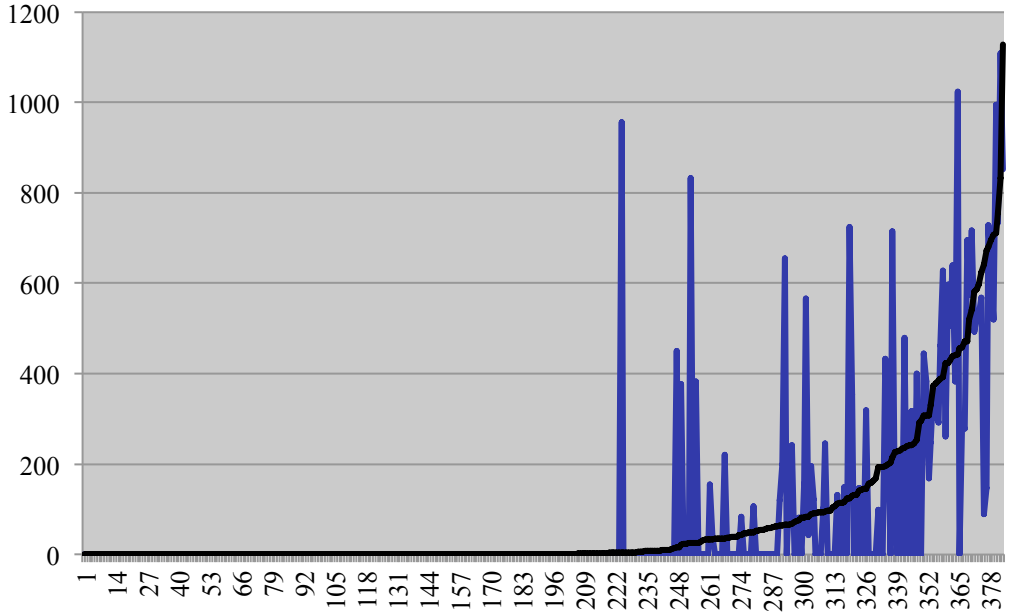


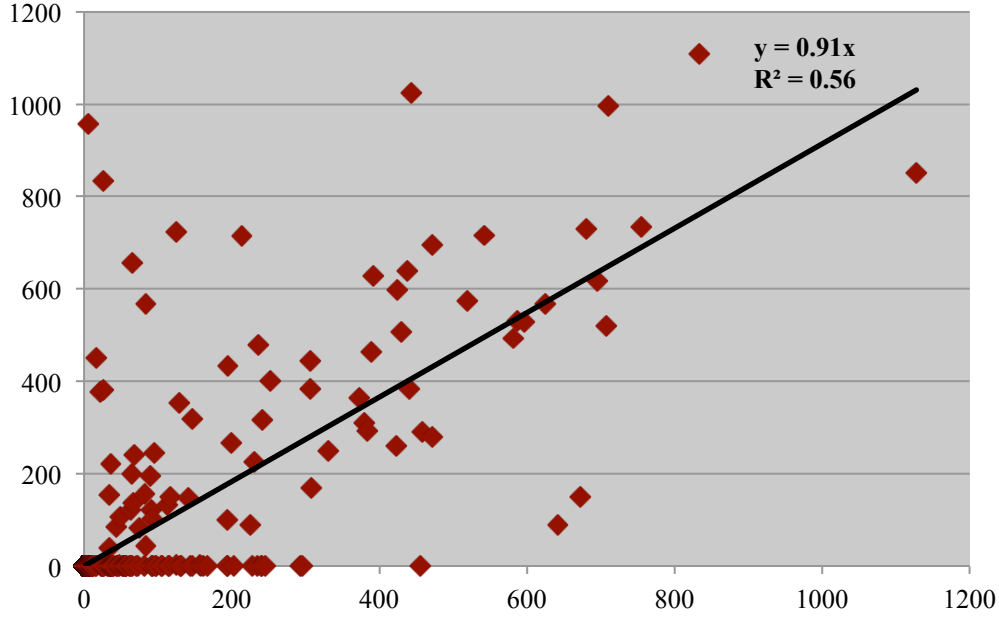
Iterative Proportional Fitting (Full Marginals and 3% Sample)





Iterative Proportional Fitting (Full Marginals and 1% Sample)





Goodness of fit test

- Standardized Root Mean Square Error

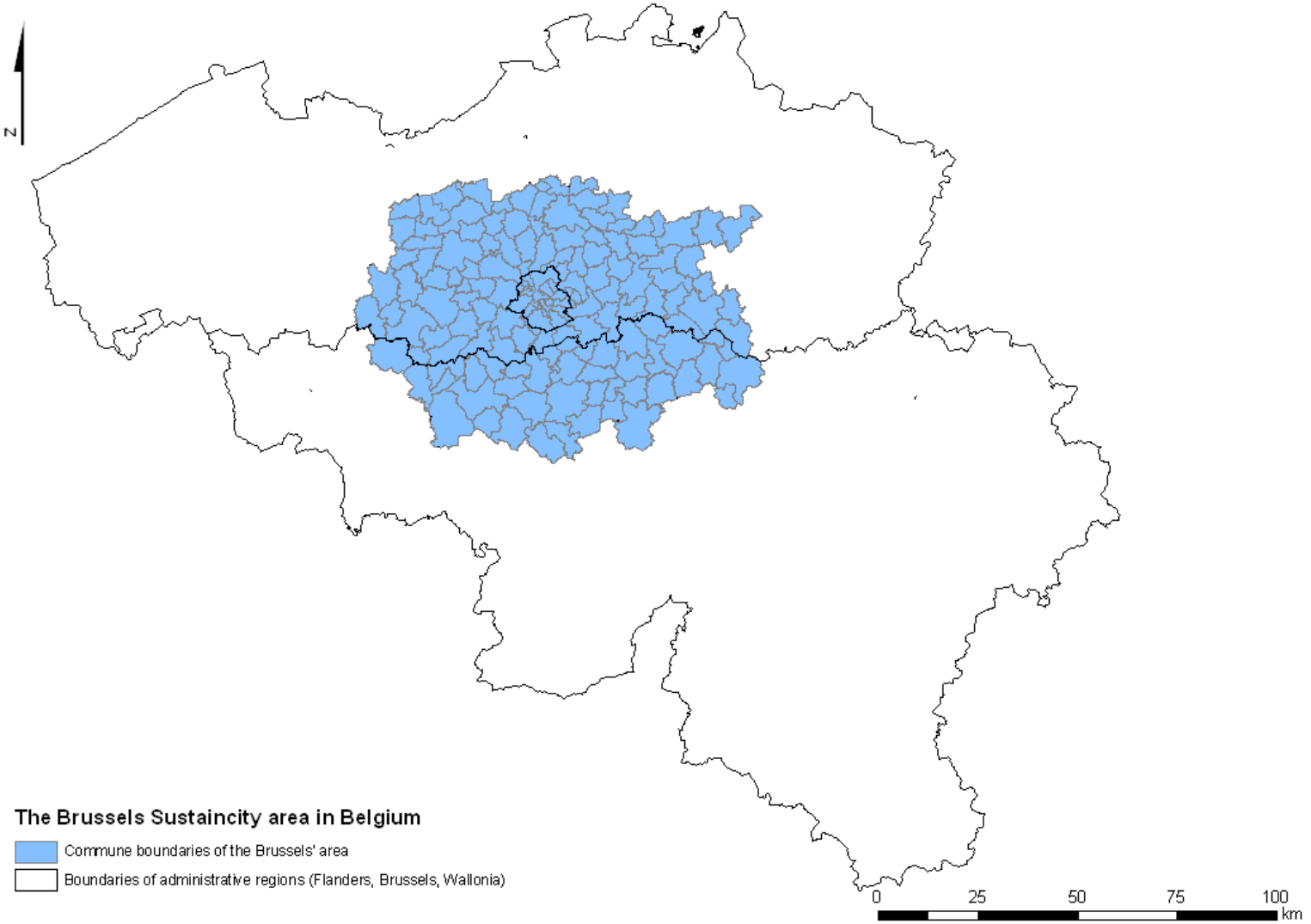
	Simulation			
	Full Cond	Partial_1	Partial_2	Partial_3
SMRSE	0.13	0.24	0.34	0.35

	IPF				
	20% Sample	10%	5%	3%	1%
SMRSE	0.853	0.928	1.02	1.16	1.73

Case of Very Limited Data

- 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



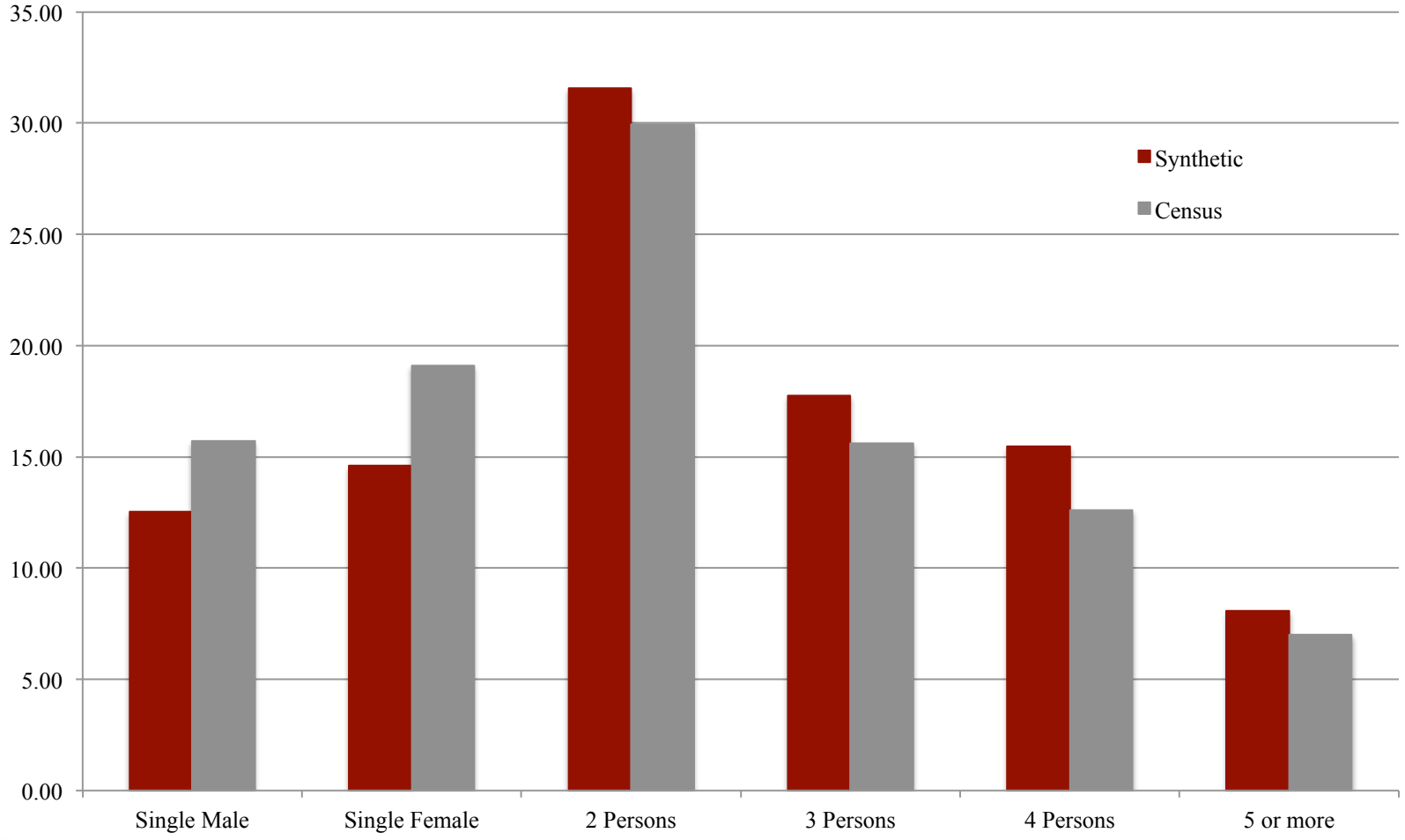
TRANSP-OR

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

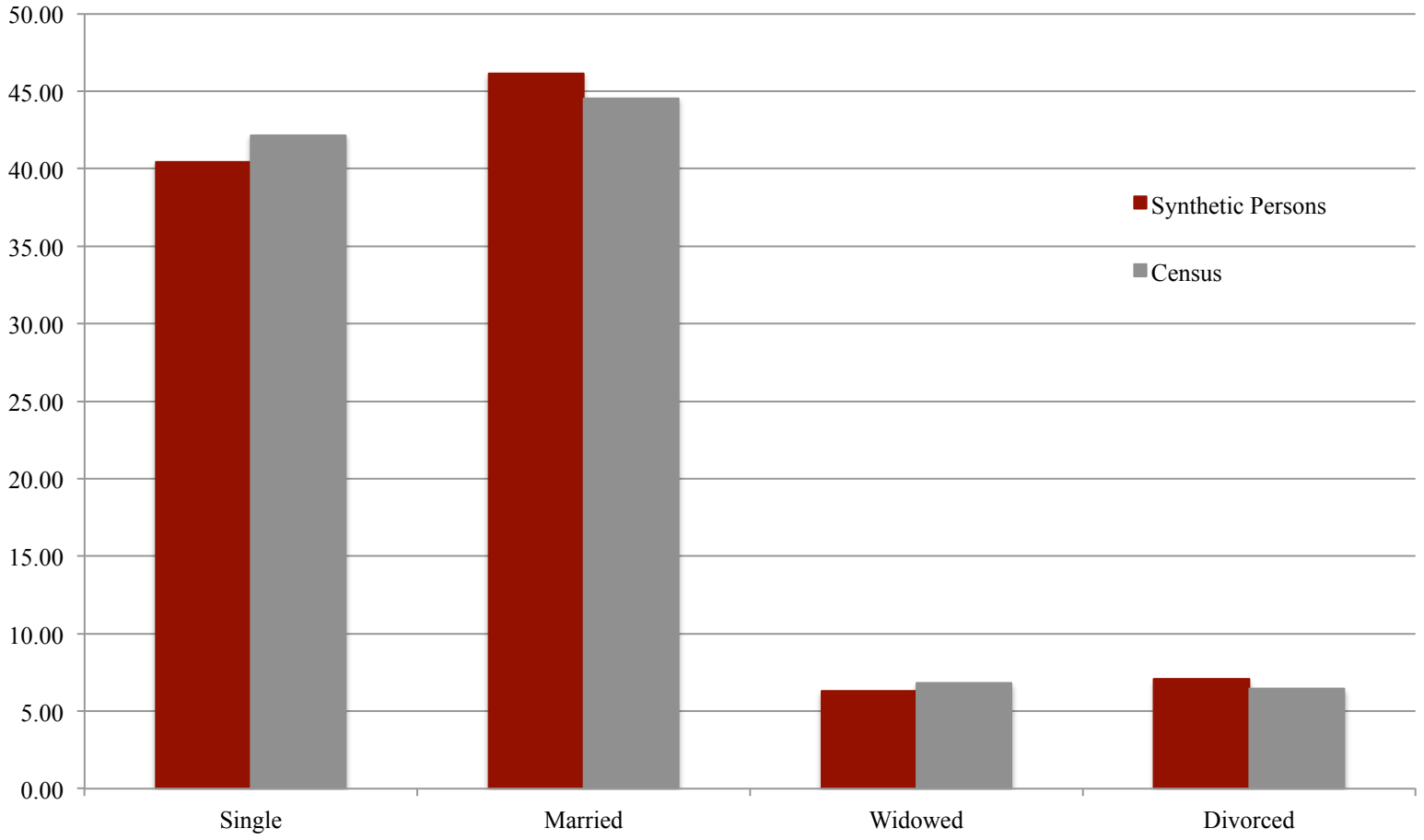
Household Size (Brussels, 2001)



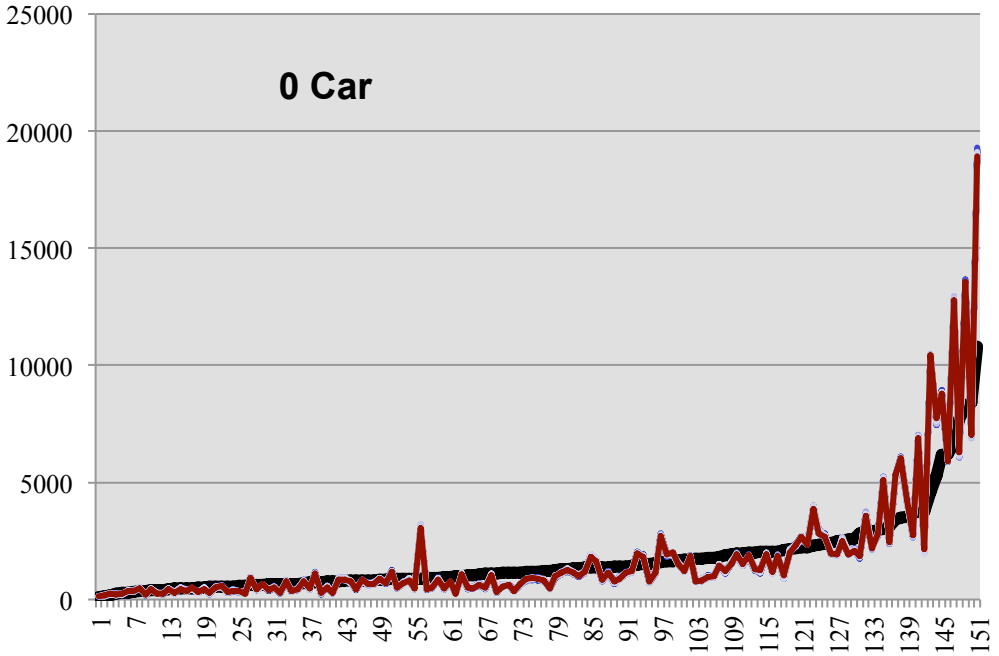
TRANSP-OR

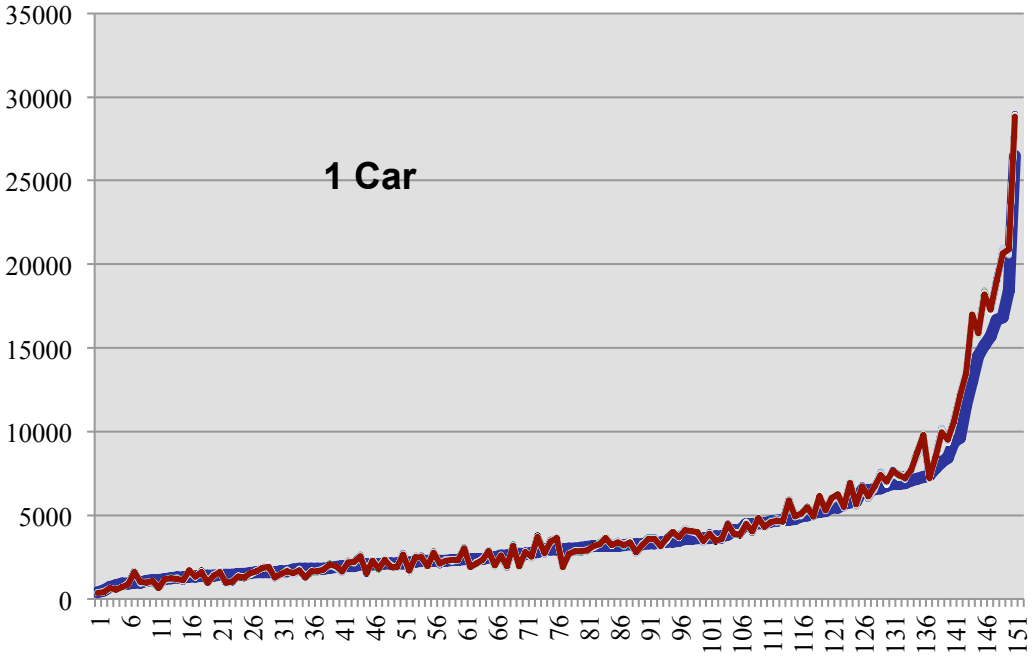
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Persons' Marital Status (Brussels, 2001)

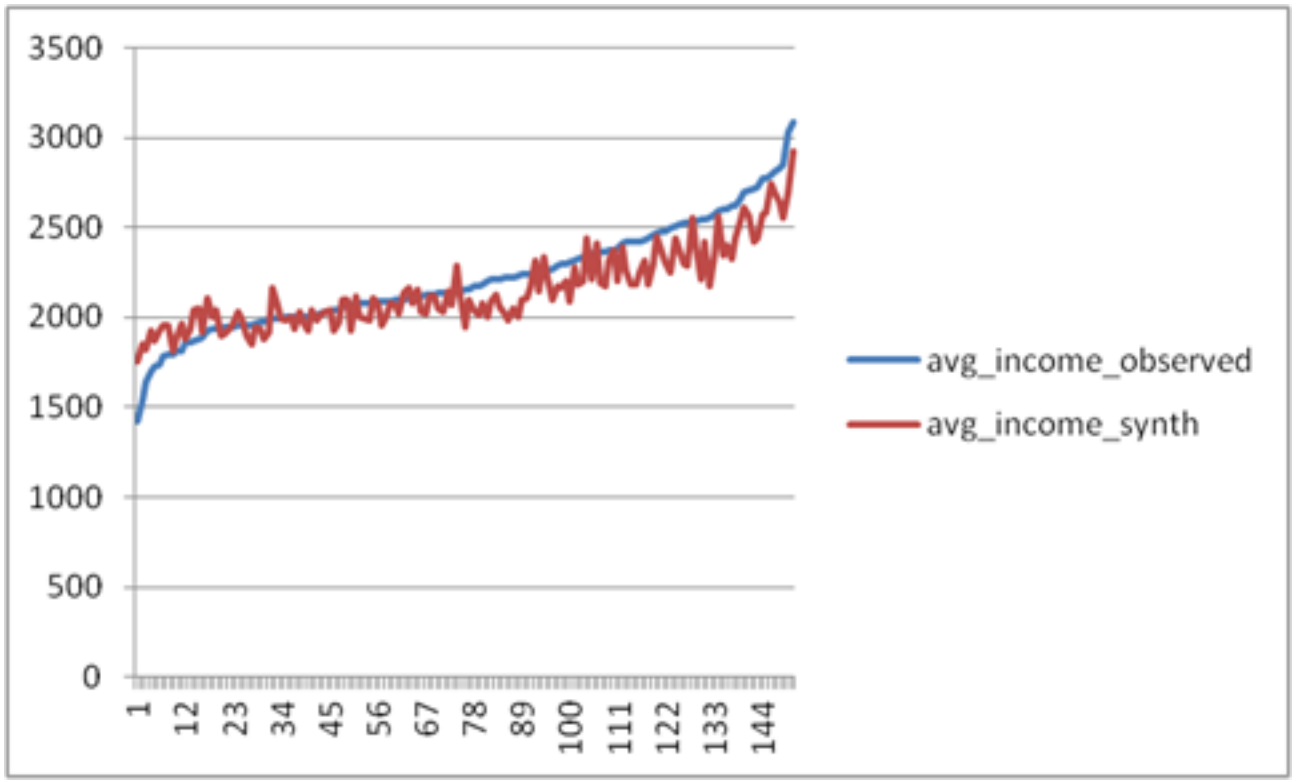


Cars in the Hhld per Commune (50 runs)





Average Income per Commune



Discussion and Conclusions

- From single solution optimization problem to sampling from joint distribution
  - Output of Land Use and Transport models
    - Integration over all possible populations
- Focus on reproducing not just marginals, but the whole joint distribution
- Heterogeneous not cloned population
- Population synthesis as part of microsimulation
  - Sensitivity analysis in a coherent way

Discussion and Conclusions

- Separation of data preparation from agent generation
  - Data, models, assumptions
- Mix of sampling process can be utilized based on the situation
- Works both for continuous and discrete or mixture of conditionals
- Computationally efficient and scalable
  - Clean and simple

Thanks!

Questions?