András Borsos, Zsuzsanna Hosszú,
Bence Mérő, Nikolett Vágó, Zsolt Oláh

January 9th, 2024, Vienna
9th World Congress of the IMA

SIZE DOES MATTER - OPTIMAL SCALING IN ECONOMIC ABMs

MOTIVATION

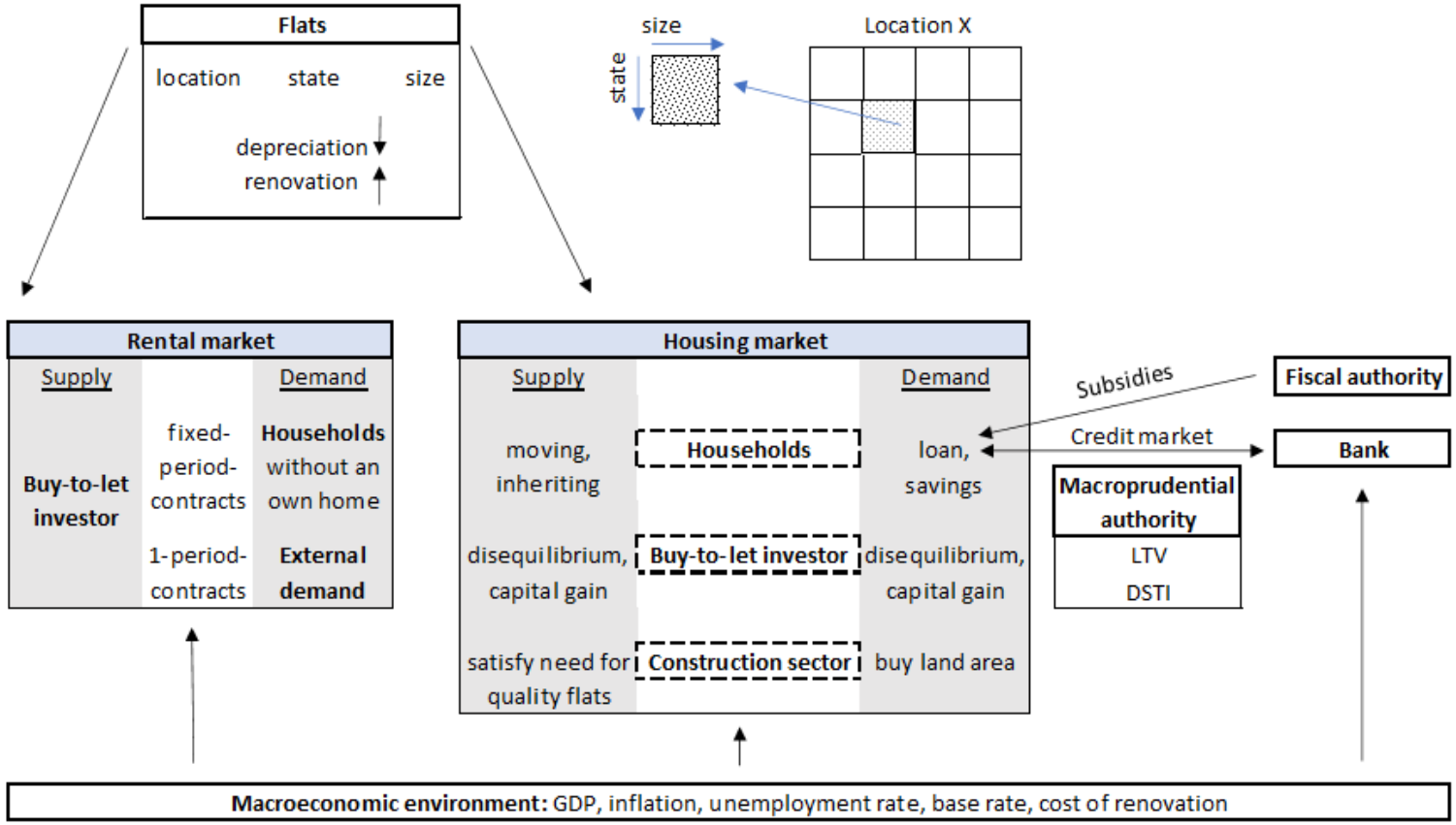
The model representation of complex economic systems **in full mapping** was traditionally considered
unsolvable analytically, infeasible computationally and **unrealistic due to data availability.**

Downscaling

(using a representative sample of agents)

- The typical simplification strategy in **computational economics.**
- The literature lacks both
  - the **theoretical** exploration of the consequences of downscaling and
  - the **empirical** assessment of the extent and severity of these problems.
- There is an increasing interest of scaling issues in other disciplines (i.e. transportation, epidemiology) as well.

THE MODEL IN A NUTSHELL



Mérő, B., Borsos, A., Hosszú, Z., Oláh, Z., & Vágó, N. (2023). A high-resolution, data-driven agent-based model of the housing market. Journal of Economic Dynamics and Control, 155, 104738.

https://github.com/mnbabm/housing-market-abm-public.git

EMPIRICAL DATA I

Households' income, demographics – Pension Contribution Dataset

- **Individuals' characteristics:**
  Birth date, Sex, ID for merging with loan data
- **Main variables:**
  - Individuals' anonym ID,
  - Gross income,
  - Employment type,
  - Weekly working hours,
  - 4-digit ISCO code,
  - Start and end date for income payment,
  - Location (NUTS3).

Loan contracts – Central Credit Registry

- **Unit of observation:**
  All home equity mortgage loan contract (at the end of 2017).
- **Main variables:**
  - Contracting date, Loan amount, Maturity
  - Principal outstanding, Current payment,
  - Market value, Provision,
  - County, Settlement type,
  - Non-performing status

EMPIRICAL DATA II

Housing stock, house price indices – housing market mediators and transaction data from the tax authority

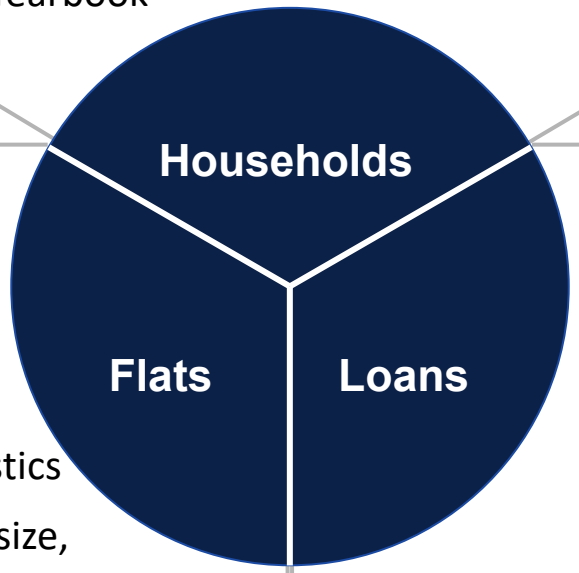
- **Unit of observation:**
  All housing market transactions (with tax burden).
- **Time period:**
  2014-2018.
- **Main variables (depending on the dataset):**
  - Price,
  - Time of transaction,
  - Size, room number,
  - Real estate type, State, Year of building, Heating type, Floor, Lift, Orientation, Garage, Balcony,
  - Location (holiday destination, settlement type, Distance from the capital/regional capital).

Interest rates of loans:

- **Unit of observation:**
  Loans with fixed rates at least for 10 years.
- **Main variables:**
  - Interest rate,
  - LTV,
  - Household income,
  - Age of debtors,
  - Length of interest rate fixation,
  - Time of the next interest rate adjustment

1:1 SCALE MAPPING

- **All of the 4 million Hungarian households**
- Occupational classification, income, social welfare benefits, education, age, sex, place of living, etc.
- Central Administration of National Pension Insurance
- Demographic Yearbook



- **cc. 200K flats (realtors)**
  + all transactions (NTA)
  + aggregated statistics of
  HCSO micro census → 4M flats
- Reconstructing the housing stock such that it mimics the agg. statistics
- 3 characteristics: neighborhood, size, quality attributes.

- **All 700K housing loan contracts**
- Central Credit Registry
- Start date, contracting value, maturity, principal outstanding, payment, interest rate, non-performing status, non-performing start date

ANALYTICAL STRATEGY

Variance arising from idiosyncratic shocks

- Downscaling → larger influence of the stochastic shocks → higher variance (Delli Gatti et al. (2011), Dosi et al. (2015))
  - 1,000,000 agents with 1:1 scaling: a shock of 1 agent affects 1/1,000,000 fraction of the population (vs 1:1000 scaling 1:1,000 fraction)
- The true extent of the impact cannot be always preserved by rescaling the shock (e.g. in disaggregated subpopulations or binary shocks)

Information loss due to insufficient interactions

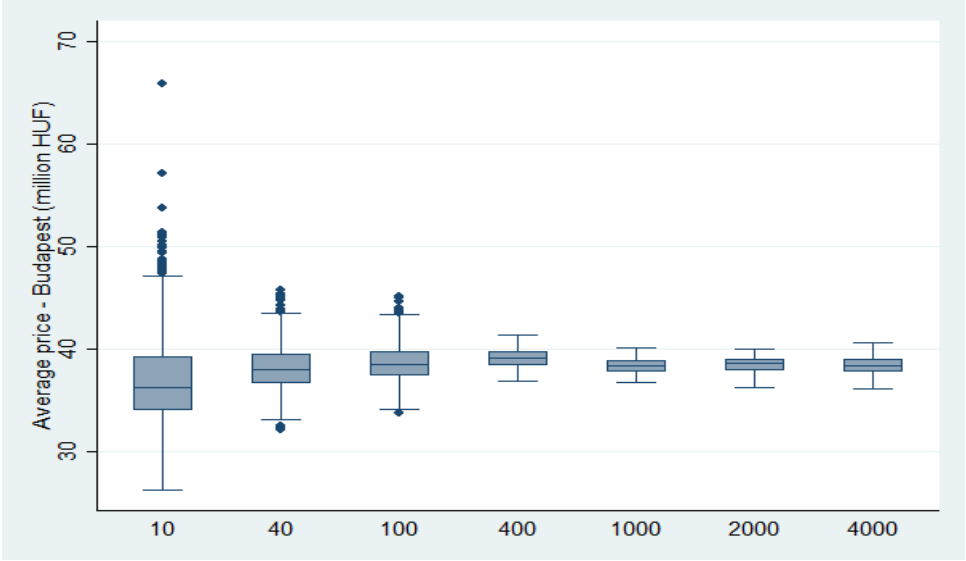
- Agents usually learn (partly) by obtaining new information from interactions on the market
  - Mérő et al. (2022): HHs estimate the price by observing similar sold flats
- Downscaling → fewer interactions → less information (than the empirically justifiable level) → distorted decisions → slower or not working adjustment to reach equilibrium

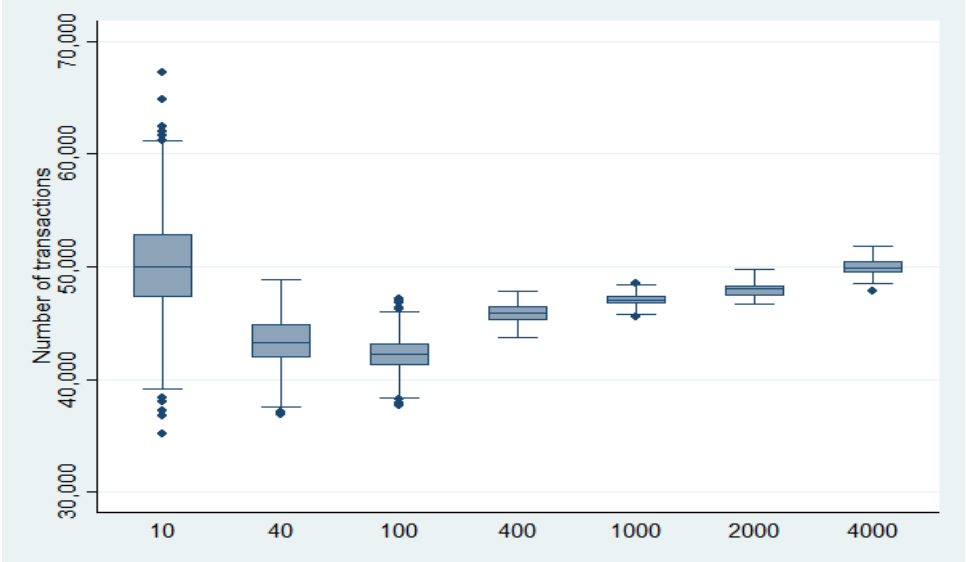
Simulations with various scaling ratios: 1:1, 1:2, 1:4, 1:10, 1:40, 1:100 and 1:400 (4M-10K HHs)
→ evaluate the runtime and the changes in the precision and the accuracy of the model outcomes

Precision | Runtime | Accuracy

RESULTS FROM THE MODEL: SCALING AND PRECISION

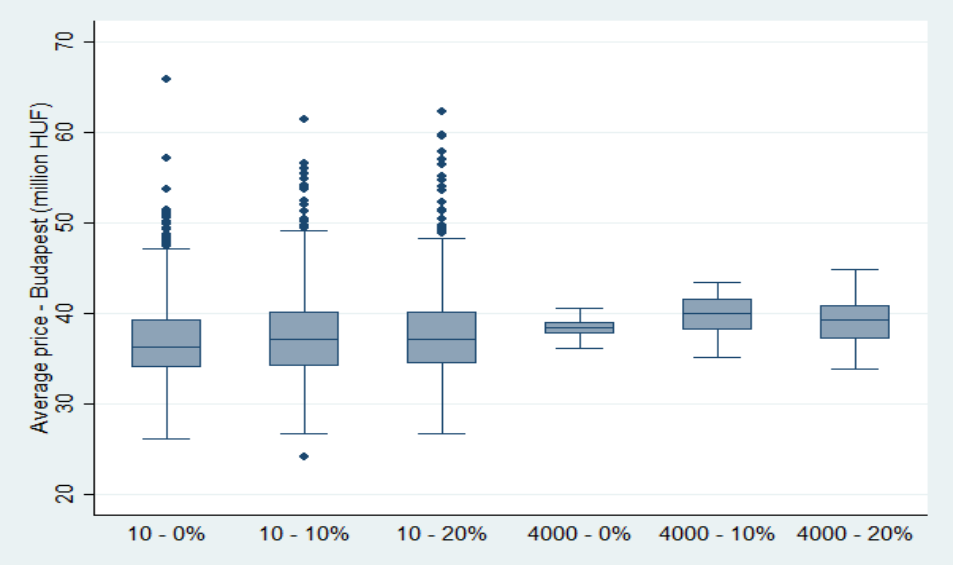
NON-LINEAR DETERIORATION OF PRECISION

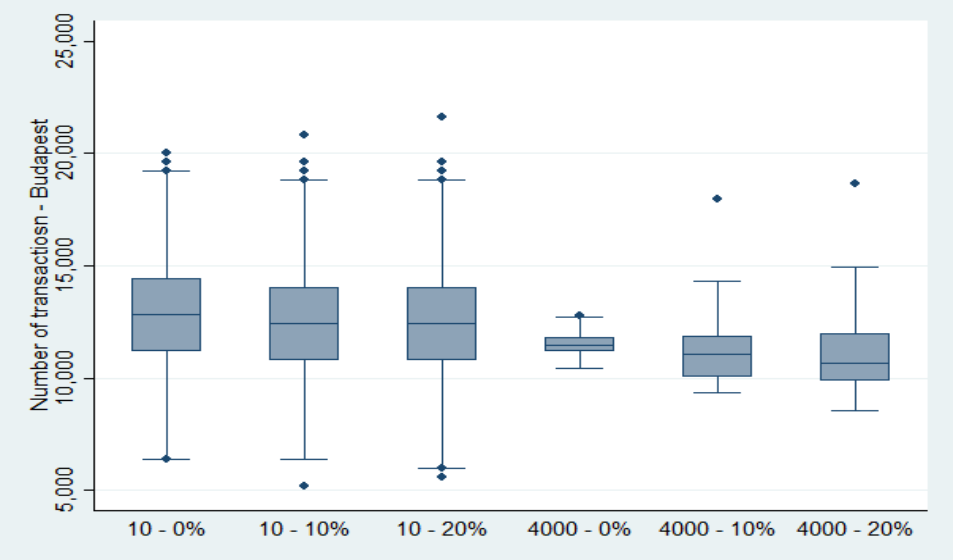




- Several runs for the different model size variants.
- We define **precision** as the total range of the most important output variables (house prices, number of transactions, credit market variables).
- Precision changes at the 25% scaling ratio in case of some variables.
- **Precision deteriorates substantially** below the 10% scaling ratios for all variables.
  - 6-10 times higher variance for 1:400 downscaling.

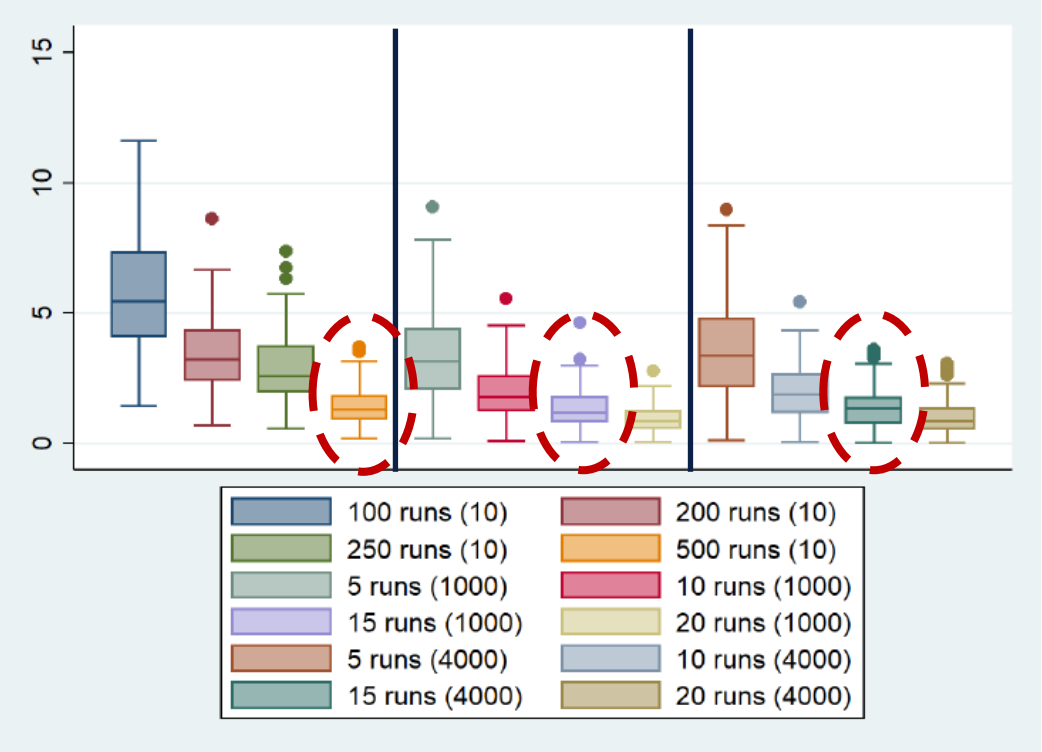
CONSEQUENCES OF HIGHER PRECISION: PARAMETER CALIBRATION COULD BE MORE PRECISE





- We randomly modified the calibrated parameters of the model with a random modifier factor from a (-10%; +10%) uniform distribution.
- We repeated this procedure many times and generated model outputs using each of these parameter sets.
- Then performed this procedure again using a (-20%; +20%) uniform distribution.
- Parameter perturbation increases the range of the larger model's variables but does not affect the outputs of the smaller model.
- It means that **the results of the larger model are influenced more strongly by the parameters and only to a lesser extent by the general model uncertainty.**
- So, **higher precision makes** the effect of the changes in the parameters clearer and **the calibration process more efficient**.

CONSEQUENCES OF HIGHER PRECISION: LESS SIMULATION RUNS ARE NEEDED

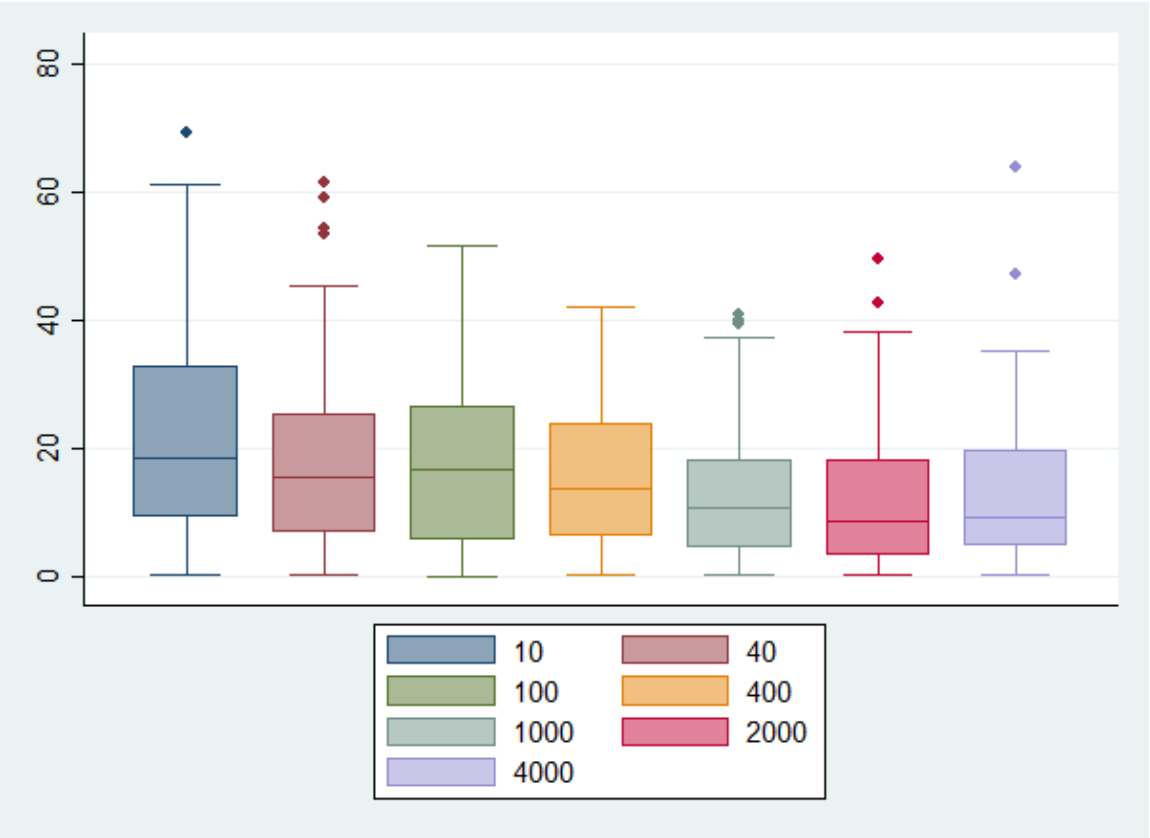


The distribution of deviations of the model outputs from the actual expected values in the cases of different model sizes and numbers of simulation runs

- The higher the precision is, the smaller is the number of runs needed to converge to the theoretical expected value of the variables.
- **Reaching an output with the same level of average deviation takes only 3-4 (12-16) times longer with 1,000.000 (4,000.000) agents** than with 10,000, while the running time of one simulation with the previous model takes 100 times longer than with the latter one.

RESULTS FROM THE MODEL: SCALING AND ACCURACY

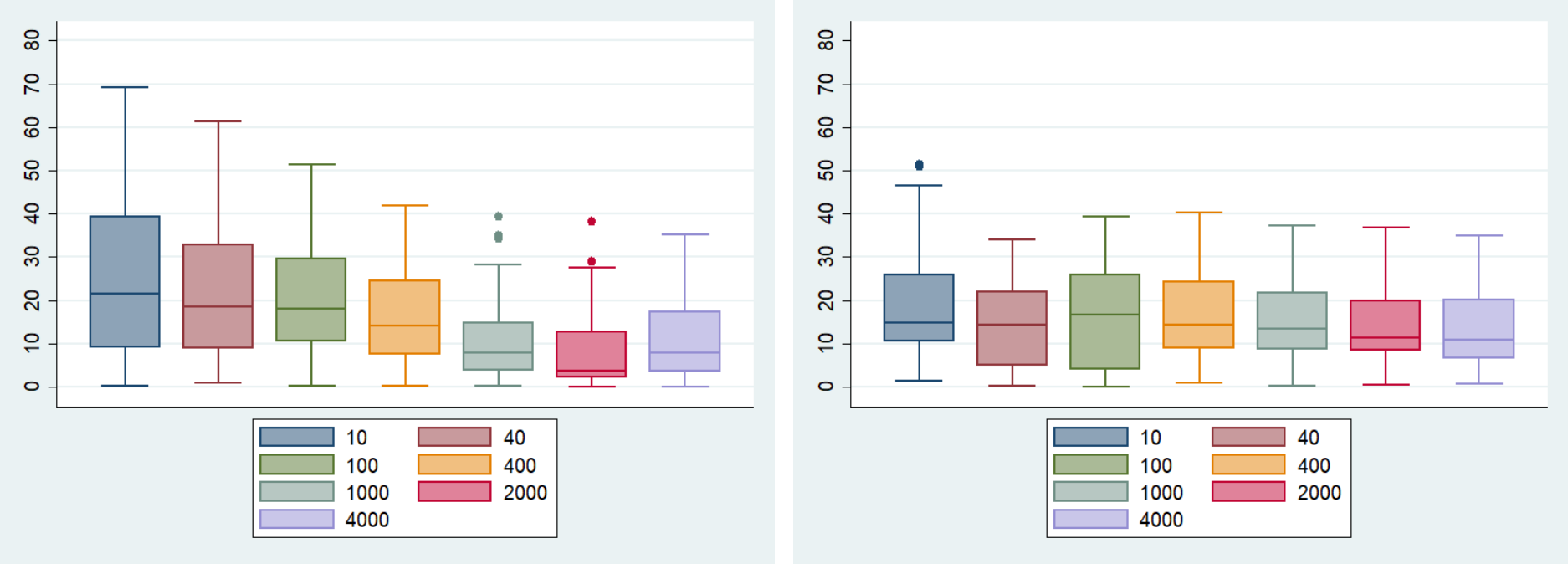
THE IMPACT OF SCALING ON ACCURACY



The deviation of model outputs from empirical data (percentage point)

- We define **accuracy as the goodness-of-fit** of the most important variables of the model.
- **Accuracy deteriorates** substantially **below the 25%** scaling factor.
- The average **deviation** of the smallest model is **twice as high** as the deviation of the model with 1 million households.

HETEROGENOUS EFFECTS OF SCALING: TYPES OF VARIABLES

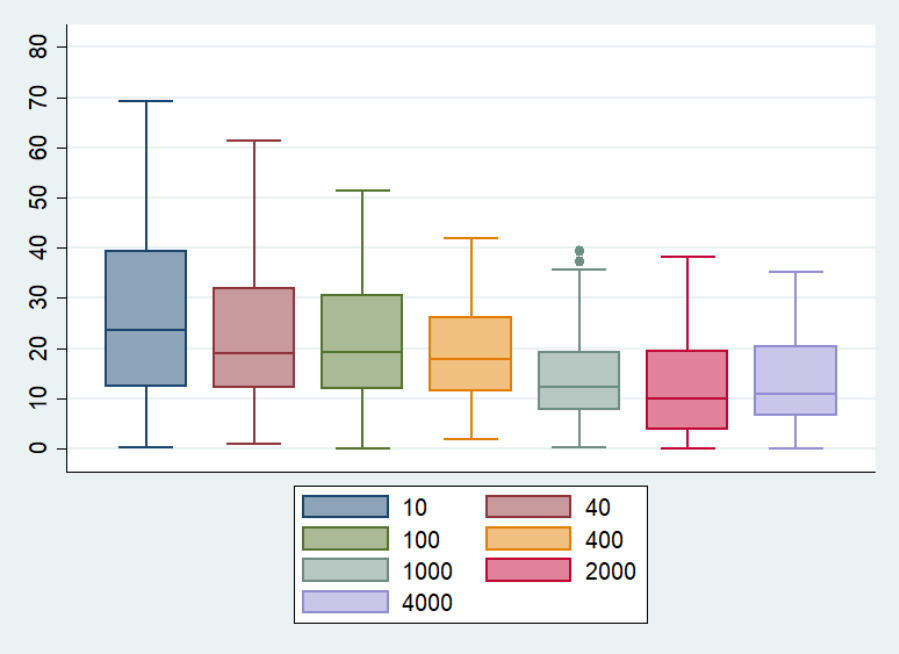


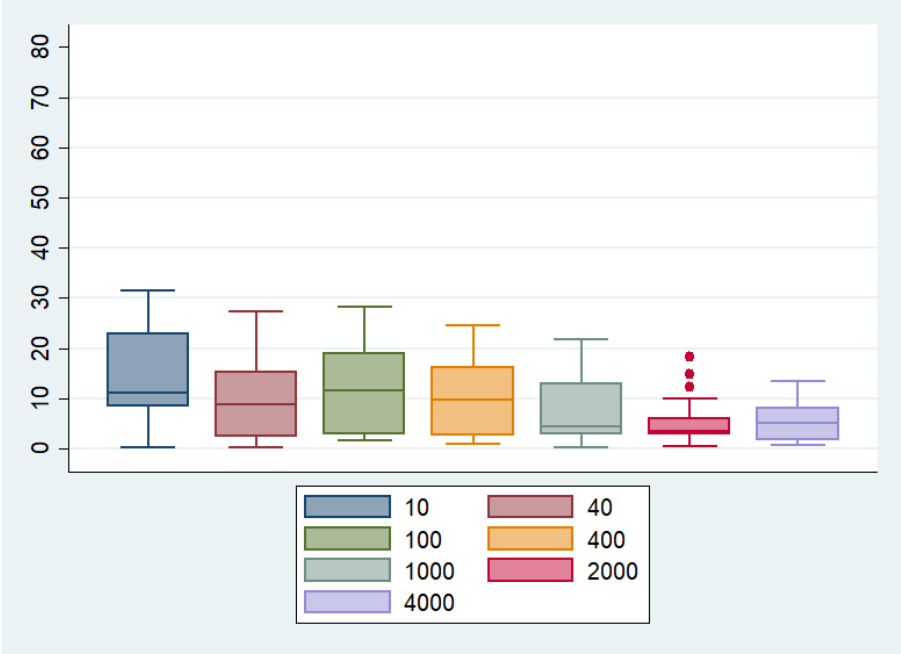
The deviations from the empirical data in the case of prices (left panel) and transaction numbers (right panel) with different model sizes.

Expectations based on the theory:

- Less transactions leads distortions in the pricing decisions → High bias in the price variables → Smaller spillover effects to other variables
- Transaction numbers: 5 percentage point difference between the 1:1 and the 1:400 model; House prices: 15 percentage points

HETEROGENOUS EFFECTS OF SCALING: AGGREGATION LEVEL



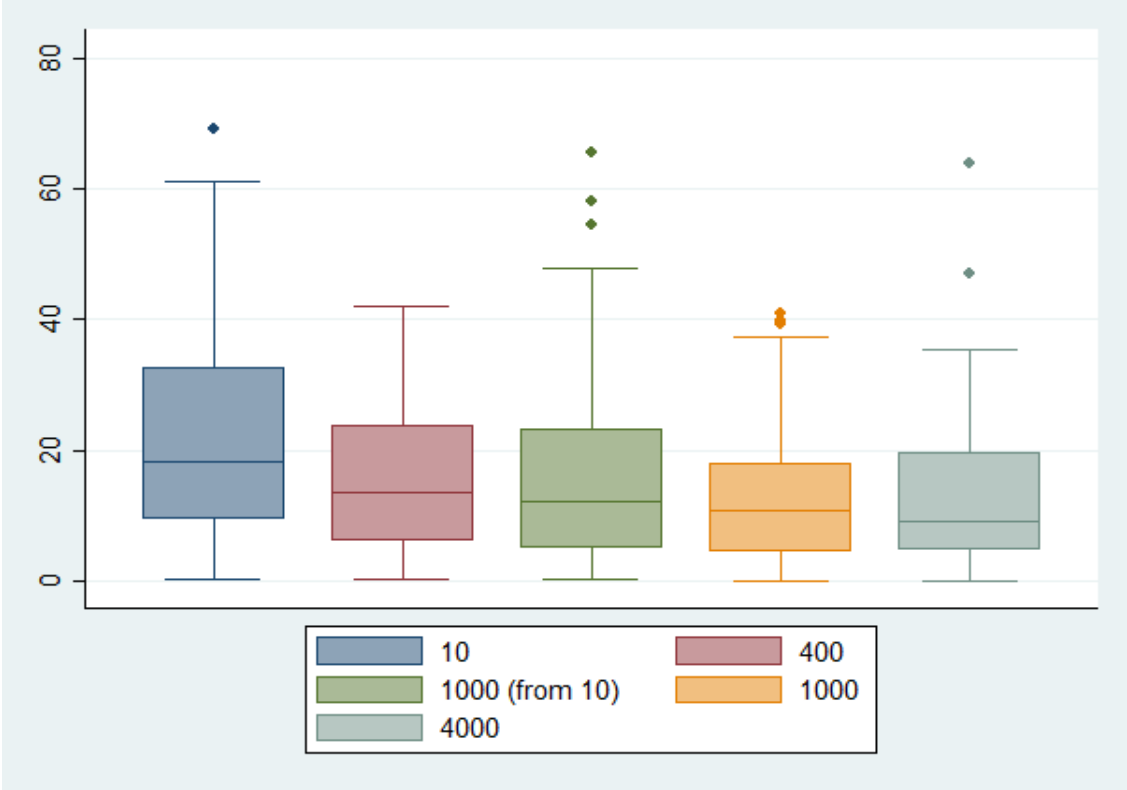


The deviations from the empirical data at regional (left panel) and aggregate level (right panel) with different model sizes.

Expectations based on the theory:

- Less transactions leads distortions in the pricing decisions → The smaller the market segment is, the higher the bias is → The deterioration gets even worse for disaggregated market segments
- Aggregates or bigger market segments: 6 percentage point difference between the 1:1 and the 1:400 model; smaller regions: 13 percentage points.

ACCURACY IN THE CASE OF SCALING UP A SAMPLE

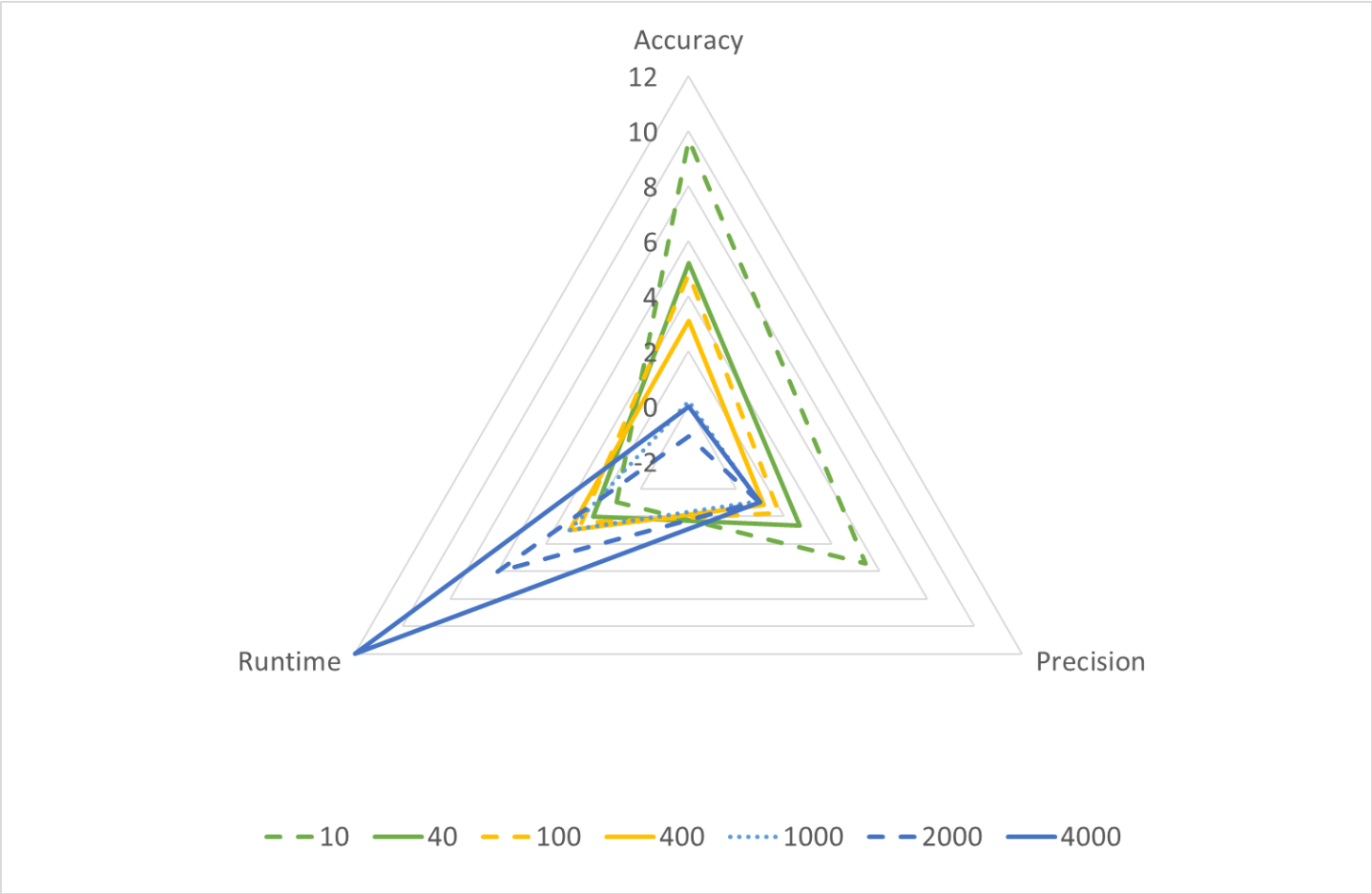


Deviation of the results of the downscaled (and up-scaled) model versions from the empirical data in percentage points

- We replicated 100 times each element of a 10K sample to generate an artificially inflated 1M size model.
- Accuracy:
  - the up-scaled version performs much better than the initial 10K model (by 7 % points in avg. deviation).
  - It is in between the 400K and actual 1M models.
- Precision:
  - the up-scaled version performs much better than the initial 10K model.
  - It is around the 400K model version's performance.

MAIN CONCLUSIONS

THERE ARE CLEAR TRADE-OFFS WHICH SHOULD BE ADDRESSED CAREFULLY



Downscaling to **25-50%** can be justified depending on the **run-time** and the required **disaggregation** level in the results.

MOST IMPORTANT RESULTS

1) **Precision deteriorates** considerably due to **downscaling**, and this effect is **non-linear**.

2) Although the **run-time** increases linearly with the number of agents, the run-time to evaluate a scenario **at a given level of precision** takes only 3-4 times longer with 100 times more agents.

3) The **accuracy** of the model also **deteriorates non-linearly** with downscaling.

4) The **decrease in accuracy** is higher in case of **smaller market segments and house prices**.

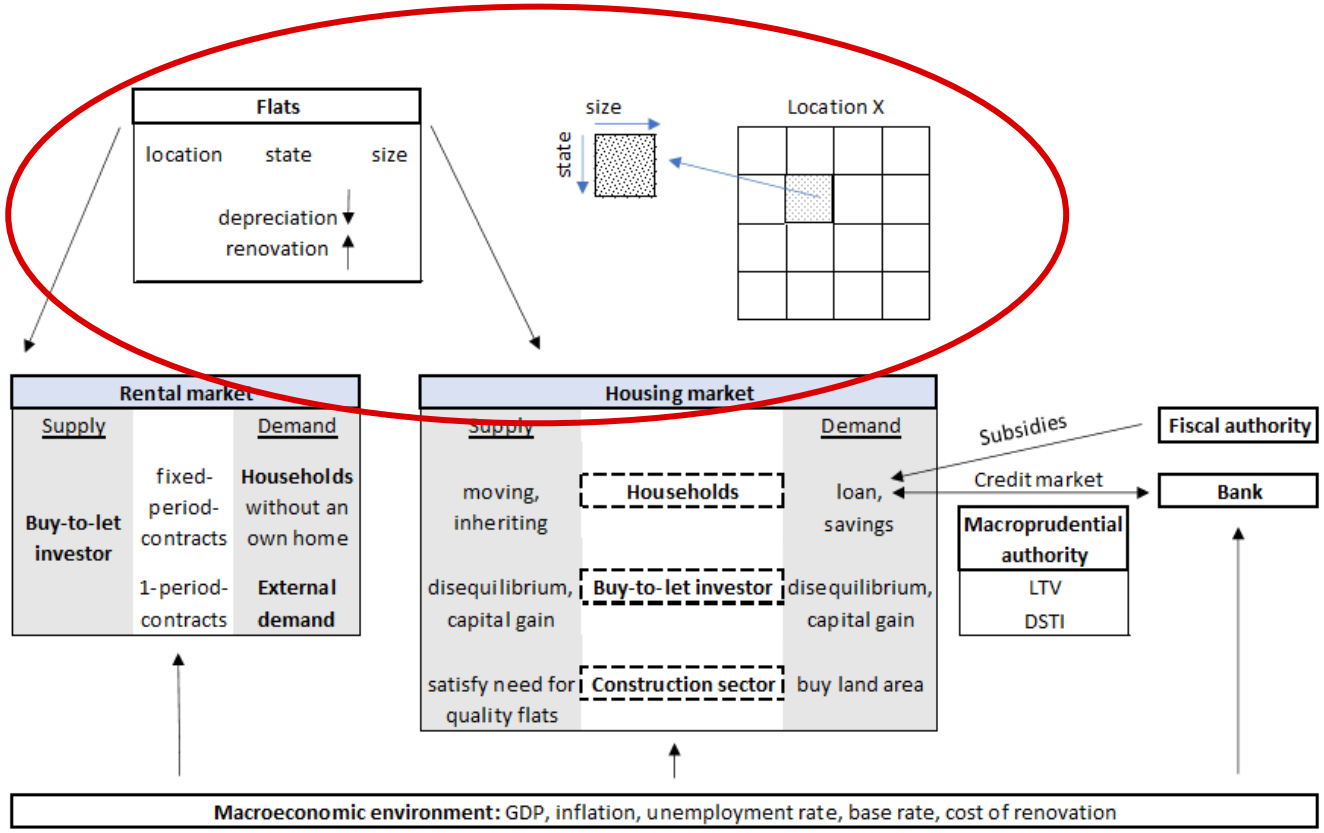
5) Most of the benefits of the large, empirically mapped model are also available by **scaling up** a representative **sample.**

THANK YOU FOR YOUR ATTENTION

merob@mnb.hu

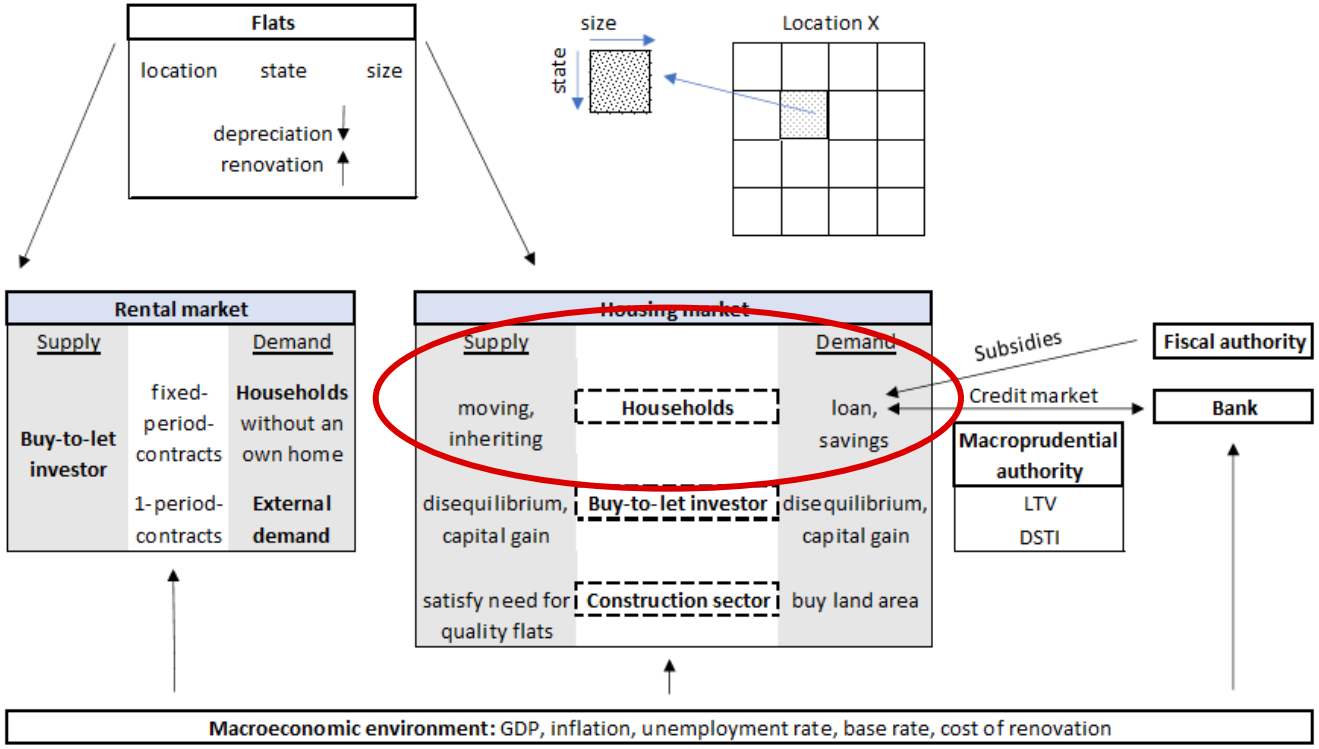
APPENDICES

THE MODEL IN A NUTSHELL 1/6 – FLATS



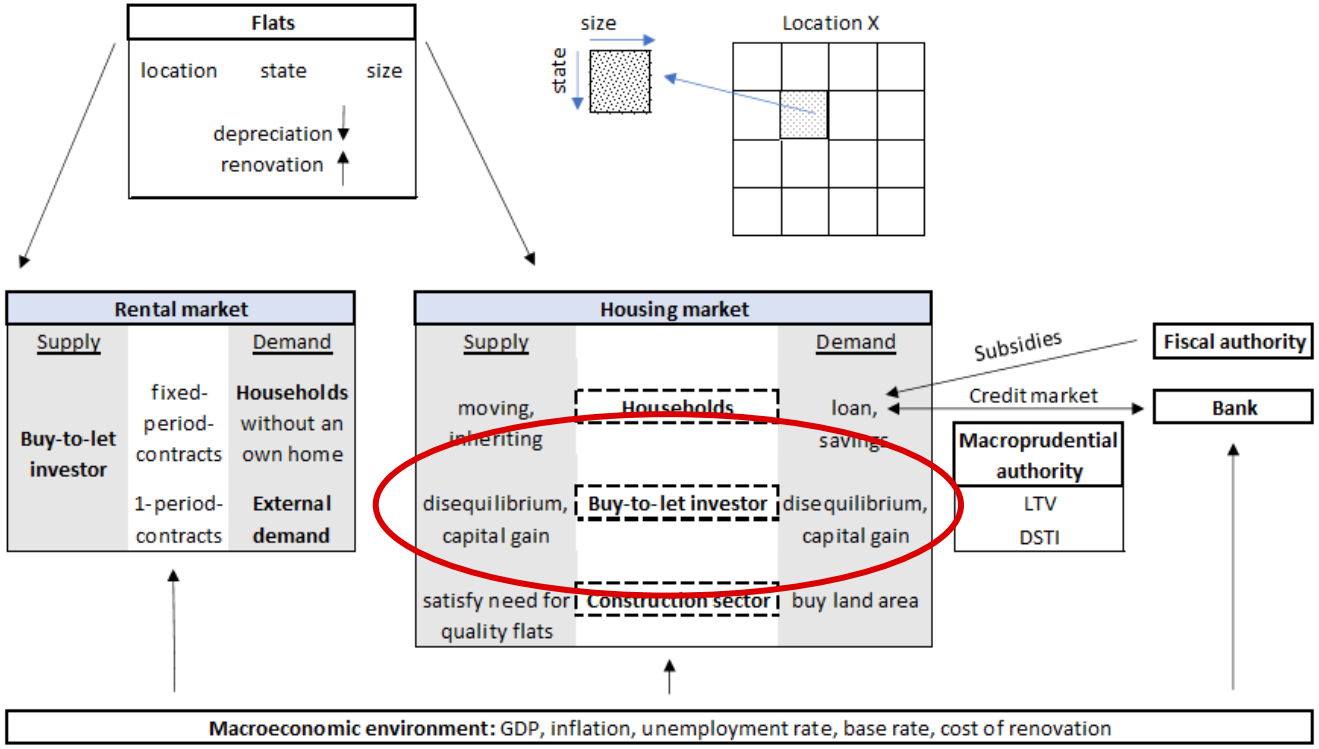
- Flats have three characteristics:
  - Size
  - Location (neighbourhood quality);
  - State (which is a composition of several measures which can change);
- Regarding the location of flats, we have divided the country into 124 actual, interpretable neighbourhoods, and estimated a cardinal quality value to each neighbourhood.
- Each month a flat's state depreciates, (but renovation can increase it).
- Within each location, flats are grouped into buckets: flats within a specific size and state interval constitute a bucket.

THE MODEL IN A NUTSHELL 2/6 – HOUSEHOLDS



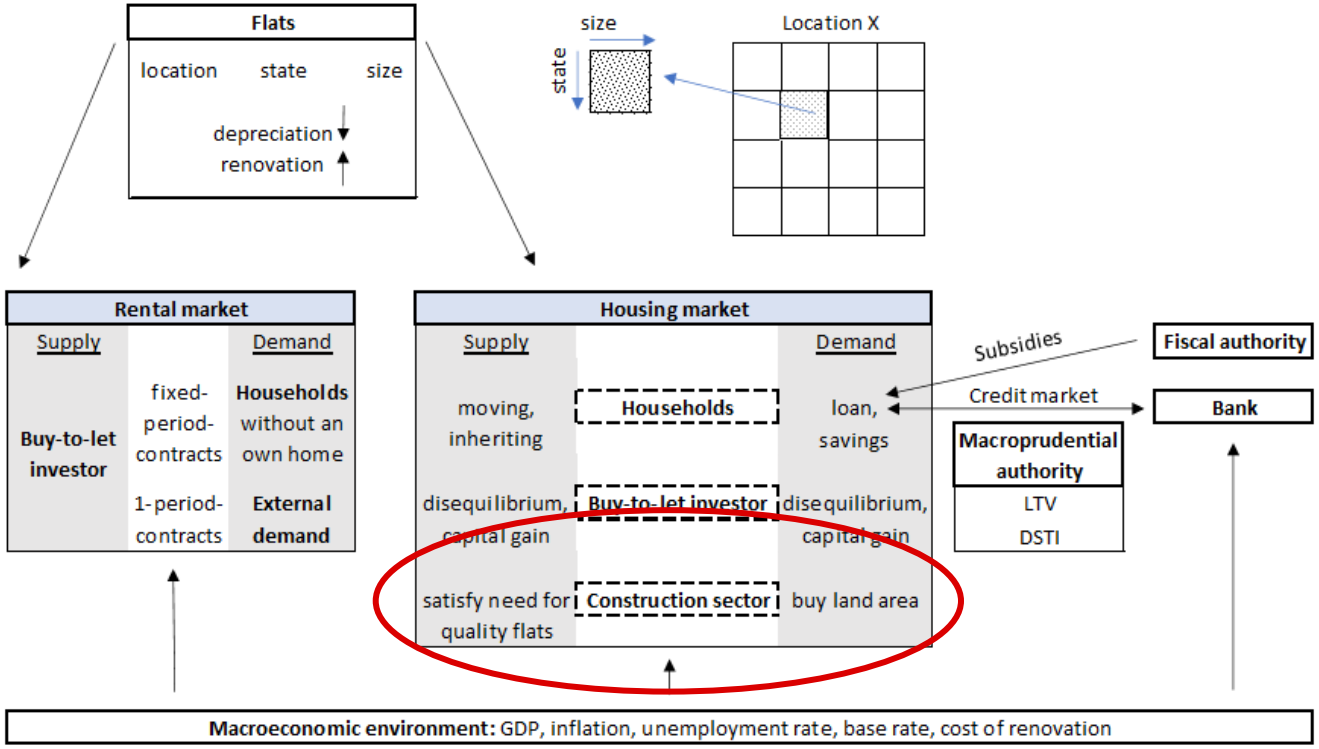
- We implemented demography in the model, including birth and death, according to empirical data.
- Households' income is determined in the model not only by their educational level and stochastic labour market shocks, but also by macroeconomic processes.
- Households can consume, accumulate savings, or buy a flat.
- They can take out a loan, but they have to meet regulatory requirements.
- Regular households may appear on the supply side when they inherit a flat and take it to the market.
- Each period, some households may decide to move if they find that they can achieve a significant increase in their consumer surplus by selling their home and moving to a new one.

THE MODEL IN A NUTSHELL 3/6 – INVESTORS



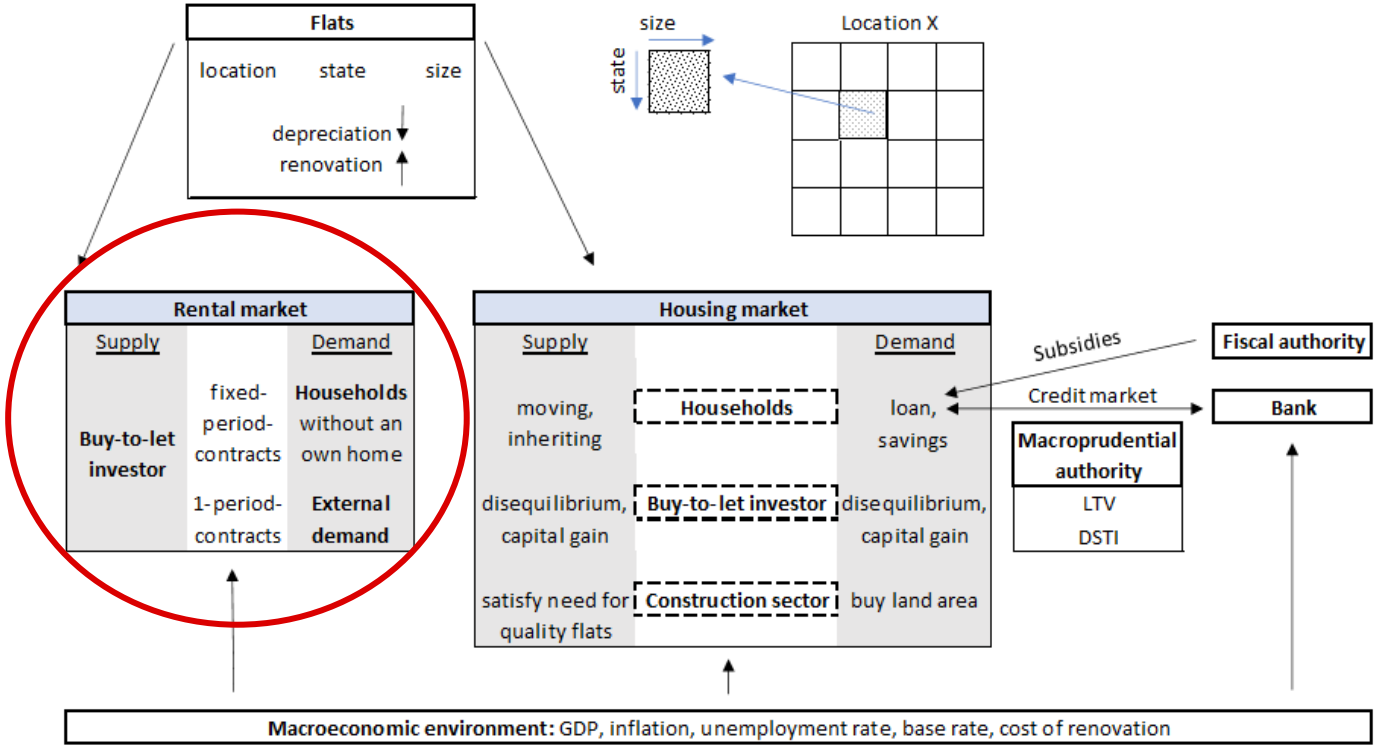
- **Representative professional investor + HHs**
- Demand is influenced by the obtainable **capital gain**, so their decisions are determined by
  - price changes and
  - vacancy rates.
- The supply and demand sides can be temporarily detached → **disequilibrium**
- The **endogenous change** in the **prices** and in the **rental markups** ensures the long term convergence to equilibrium

THE MODEL IN A NUTSHELL 4/6 – CONSTRUCTION



- The construction sector
  - is represented by a representative firm,
  - which estimates demand for newly built flats heterogeneously for neighborhoods and flat categories.
- Construction process:
  - The construction sector **builds high quality** flats.
  - It needs **land site** to build, so it buys the flats with the lowest unit price for sale in the neighborhood where it wants to build.
  - Construction takes **18 months**, but the construction firm can sell the flats even before they are finished.
  - Construction costs are proportional to the regional average salary (and higher than the renovation cost).

THE MODEL IN A NUTSHELL 5/6 – RENTAL MARKET



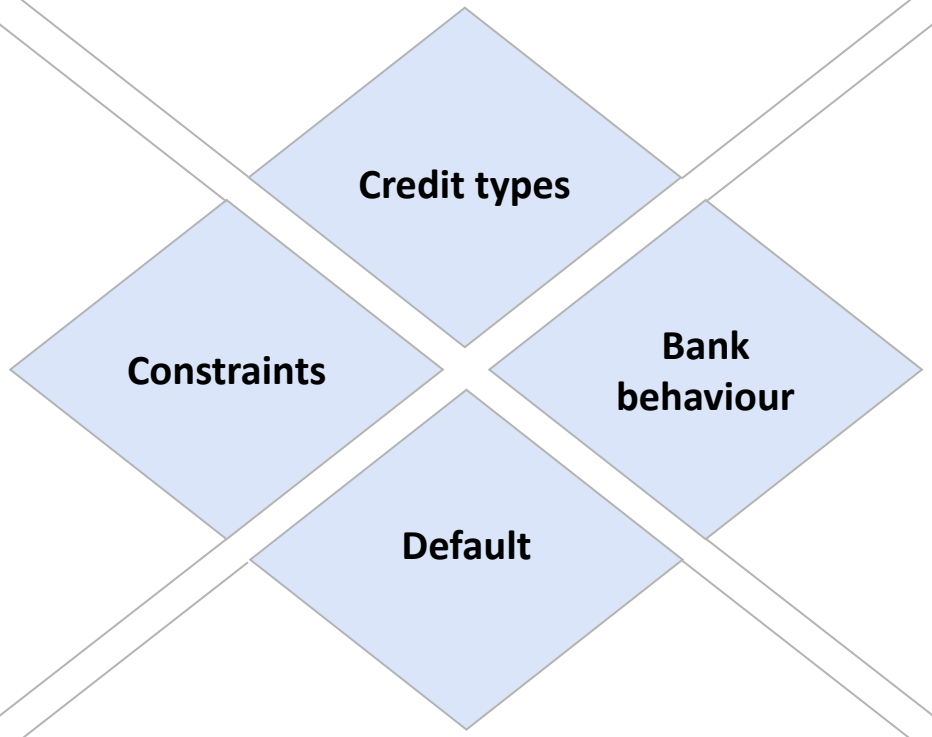
We distinguish between short-term (maximum one month) and long-term renting.

- Short-term
  - represents online market place platforms (e.g. Airbnb), which mostly serve the demand coming from tourism.
  - → time-dependent, exogenously given external demand for every for neighborhoods and flat categories based on empirical data
- Long term
  - If a household does not have an own home, it can go to the rental market.

THE MODEL IN A NUTSHELL 6/6 – CREDIT MARKET

- There are housing mortgage loans, bridge loans and also consumer loans for renovation.
- There are fix and variable rate loans as well.

- A household is eligible for a loan if:
  (1) it meets the LTV and DSTI rules;
  (2) its expected income covers the credit payments and a minimal consumption level;
  (3) and it did not have a defaulting loan in the past five years.
- There can be only one mortgage on one flat



- The bank increases and reduces the credit supply procyclically.
- The bank determines the credit margins with a regression model estimated on actual empirical data.

- In case of non-performance, households first try to reduce their consumption → the bank restructures the loan → finally the collateral will be liquidated.

HIGHLIGHTED FEATURES 1/4 – HOUSEHOLD'S BIDDING

In order to take into account the heterogeneity of agents' preferences, each household in the model has been assigned a reservation price function which was calibrated uniquely using a stochastic optimization procedure.

We investigated many forms (CES, Cobb-Douglas, exponential) and eventually choosed the function with the best goodness-of-fit:

$$U_h^l = (c\_m_h \times size_l^{\beta_h^m} \times (1 + c\_a_h) \times state_l^{\beta_h^a} + \frac{K_h}{(1 + e^{-neigborhood_l})^{1/\gamma_h}}) \times lifetime\_income$$

When bidding, all the potential buyers choose a „perfect" fictive flat. Then, the closer a flat's consumer surplus is to the consumer surplus of the ideal fictive flat, the higher the probability is for a household to bid on a flat.

The offer is the weighted average of the reservation price and the market price.

The highest bidder can buy the flat at the price of the second highest bid (in the case of multiple bidders) (→ Vickrey auction), or at the market price (if there is only one bidder).

HIGHLIGHTED FEATURES 2/4 – CHARACTERISTICS OF FLATS

Table 1: R^2 of the hedonic price regression with different sets of flat attributes

Flat characteristics	R^2
Original estimation	0.84
Without "state"	0.75
Without "size"	0.54
Without "neighborhood"	0.27

Table 2: R^2 of the hedonic price regression with different location variables

Location categories	R^2
None	0.27
Region	0.66
Counties	0.67
Neighborhoods	0.84

- In the model, every flat has three attributes: **size, state and location**.
- According to the results of our **hedonic price regressions**, all of these attributes are highly significant, and they increase the goodness-of-fit considerably.
- These results also suggest the outstanding importance of our **neighborhood quality** variable.
- We examined the neighborhood level location variable compared to more coarse-grained options in the price regression estimation

HIGHLIGHTED FEATURES 3/4 – CONSTRUCTION

In the existing housing market ABMs the construction sector is either not modelled at all, or it is included only to a limited extent. We elaborated it for 3 reasons:

Prices and the building up of bubbles on the housing market

- Cost shocks fundamentally determine the price of (not only newly built) houses.
- If the increase of prices is driven by excessive demand, it can be offset by the construction sector.
- The effectiveness of this mechanism is highly dependant on the time requirement of the constructions and the availability of the necessary capacities.

Demographic phenomena (reducing birth numbers or migration)

- These mechanisms have heterogeneous implications across regions.
- Thus, the assumption of time-invariant market conditions would lead to distorted results as the time span of the simulation expands.
- To ensure flexibility, one needs to elaborate the mechanisms responsible for the adaptability of the market.

Policy analyses

- Since 2016, vast amount of resources have been allocated on the subsidization of Hungarian families' home purchases.
- These policies affected immensely not only the housing loan markets, but also the housing market through several channels.
- To capture the implications, the detailed representation of the construction sector is inevitable in the model.

HIGHLIGHTED FEATURES 4/4 – SIZE OF THE MODEL

According to our knowledge, there is no other model in the literature, which analyses the housing market of a whole country by representing all the households and also the housing stock in its entirety.

It has at least 2 advantages:

- This way it is possible to analyse economic policies at more disaggregated levels: We can examine small groups of households, such as low income HHs, families with many children, first-time home buyers, etc.
- The model is not scale invariant. If we use fewer agents, the long term averages of the main time series generated by the model will change considerably.

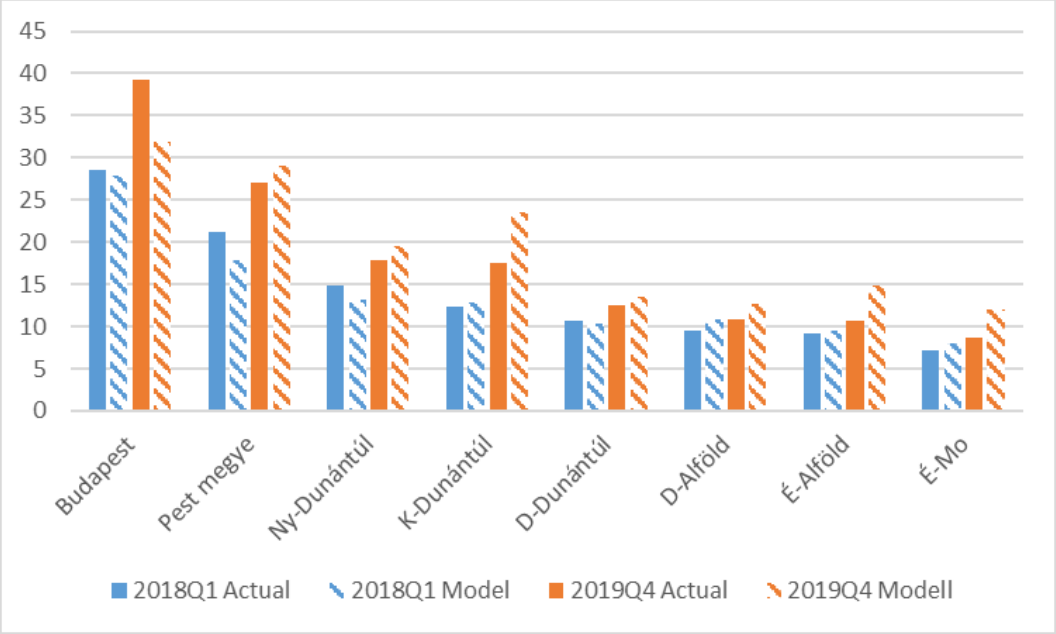
Table 4: Model simulation results for different model sizes

Number of individuals	500,000	50,000
Avg. house prices (million HUF)	21,1	15,2
House price autocorrelation	0,94	0,04
Avg. # of transactions per month	955	33
Transaction number autocorrelation	0,77	0,33
Ratio of transactions / individuals	0,19%	0,07%
Ratio of transactions / flats for sale	27,20%	1,10%

The main driver of this level of scale dependence is the agents' behaviour regarding the selection mechanism: Buyers would place a bid with a high probability only on flats with which they can obtain similar consumer surplus as that of the best fictive flat.

→ If there are fewer agents in the economy, there will be fewer flats on the market, which leads to lower probabilities for the households to find a suitable home, i.e. frictions on the market become more severe.

CALIBRATION



House prices in million HUF

Monthly averages	2018Q1-2019Q2	
	Actual	Model
Number of transactions on the housing market	15 148	14 752
Number of transactions of newly built flats	1 966	1 897

- Parameters are calibrated such that the dynamics of the observable variables match the empirical data:
  - average regional prices,
  - the number of transactions
  - newly built housing stock
- We used data from 2018/19.

VALIDATION

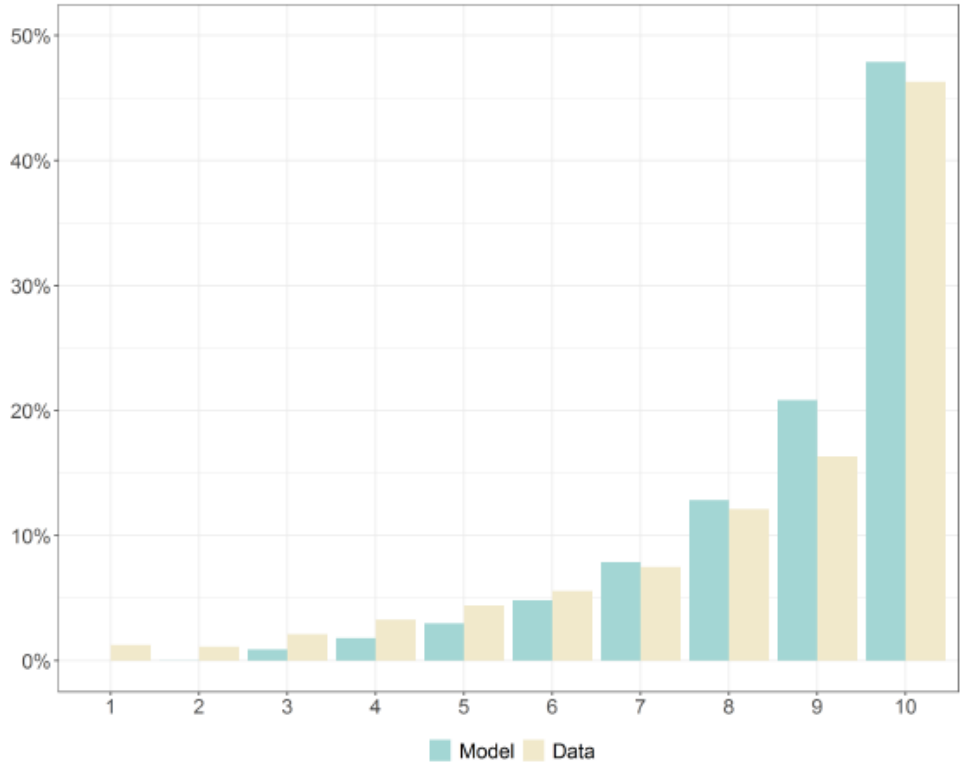


Figure 4: Distribution of the volume of newly issued housing loans in 2018 and 2019 based on the income deciles of the households. (Source: MNB.)

Yearly averages	New credit flow (billion HUF)		Number of contracts	
	Actual	Model	Actual	Model
2018-2020	895	1136	79744	75967

- We tested whether the variables of the model which were not calibrated follow the empirical data.
- We used mainly lending market variables:
  - Number of loan contracts,
  - New credit flow,
  - Distribution of loans based on income deciles, LTV and DSTI categories.
- But also some disaggregated housing market statistics:
  - # of transactions at the regional level
  - Average neighborhood quality of flats in transactions