

Is the monocentric urban economic model still empirically relevant? Assessing urban econometric predictions in 192 cities on all five continents.

C. Liotta^{1, 2} Q. Lepeitit¹ V. Viguié¹

¹CIREN

Ecole des Ponts ParisTech, Paris, France

²TU Berlin

October 11, 2021

The Standard Urban Model (SUM)

- Initially developed by Alonso (1964), Muth (1969) and Mills (1967).

The Standard Urban Model (SUM)

- Initially developed by Alonso (1964), Muth (1969) and Mills (1967).
- Model households' location and transportation choices based on land-use constraints, transportation costs and employments locations.

The Standard Urban Model (SUM)

- Initially developed by Alonso (1964), Muth (1969) and Mills (1967).
- Model households' location and transportation choices based on land-use constraints, transportation costs and employments locations.
- Widely used for instance to forecast urban sprawl or analyze transportation and land-use policies (OECD, 2018, 2020).

The Standard Urban Model (SUM)

- Initially developed by Alonso (1964), Muth (1969) and Mills (1967).
- Model households' location and transportation choices based on land-use constraints, transportation costs and employments locations.
- Widely used for instance to forecast urban sprawl or analyze transportation and land-use policies (OECD, 2018, 2020).
- Based on two main mechanisms and one hypothesis:
 - **Real estate pricing model.**
 - **Housing production model.**
 - In its most standard version, **monocentricity hypothesis.**

The Standard Urban Model (SUM)

- Initially developed by Alonso (1964), Muth (1969) and Mills (1967).
- Model households' location and transportation choices based on land-use constraints, transportation costs and employments locations.
- Widely used for instance to forecast urban sprawl or analyze transportation and land-use policies (OECD, 2018, 2020).
- Based on two main mechanisms and one hypothesis:
 - **Real estate pricing model.**
 - **Housing production model.**
 - In its most standard version, **monocentricity hypothesis.**
- Main conclusions:
 - **Negative density gradient** from the city center to the periphery.
 - **Urbanized areas' sizes** depend positively on income and population and negatively on transportation costs and agricultural land prices.

Research question

Is the monocentric SUM still empirically relevant?

Research question

Is the monocentric SUM still empirically relevant?

Main objectives

- Assessing the potential of data availability increase to work on urban structures.
- Investigating the empirical relevance of the key stylized facts predicted by the SUM on a large number of cities and with a wide geographical coverage.
- Understanding discrepancies between the SUM and actual urban forms.

- Many studies investigate whether cities exhibit the **negative density gradient** expected from the theory:
 - Clark (1951): 19 cities in 9 countries.
 - Bertaud and Malpezzi (2014): 57 cities of different regions / levels of development.
 - Lemoy and Caruso (2017): 300 European cities.

- Many studies investigate whether cities exhibit the **negative density gradient** expected from the theory:
 - Clark (1951): 19 cities in 9 countries.
 - Bertaud and Malpezzi (2014): 57 cities of different regions / levels of development.
 - Lemoy and Caruso (2017): 300 European cities.
- Several papers also checked the aggregated predictions of the SUM on the **size of urbanized areas**:
 - Paulsen (2012): 329 cities in the United States.
 - Oueslati et al. (2015): 282 European cities.
 - Schmidt et al. (2020): 92 cities in Germany.
 - Jedwab et al. (2021): 1010 cities in developing countries.

- Few studies assess the **underlying mechanisms of the model**:
 - Housing production model: McMillen (2006) on Chicago.
 - Real estate pricing model: Yinger (1979), McDonal and Bowman (1979), Ahlfeldt (2011),...
 - **None of the two mechanisms have been assessed on a large number of cities.**

- **Geographical knowledge gap:**
 - Most studies assessing the empirical validity of the key stylized facts of the SUM analyze European / North American cities.

- **Geographical knowledge gap:**

- Most studies assessing the empirical validity of the key stylized facts of the SUM analyze European / North American cities.
- Cities with most future relevance in terms of total urban population and growth dynamics are systematically underrepresented:
 - Africa has the fastest growing cities (UN, 2018).
 - Asian cities have the largest share of the global urban population (UN, 2018).

- **Geographical knowledge gap:**
 - Most studies assessing the empirical validity of the key stylized facts of the SUM analyze European / North American cities.
 - Cities with most future relevance in terms of total urban population and growth dynamics are systematically underrepresented:
 - Africa has the fastest growing cities (UN, 2018).
 - Asian cities have the largest share of the global urban population (UN, 2018).
- **Knowledge gap on the underlying mechanisms of the model,** whereas they remain widely used to understand the impact of transport and land-use policies (OECD, 2020).

- The monocentric SUM is a simple and old model, but remains a useful analytical tool, consistent with many city structures across the world.

- The monocentric SUM is a simple and old model, but remains a useful analytical tool, consistent with many city structures across the world.
- It however much better describes aggregated variables on cities (e.g. urban area) than the precise internal structure of the cities.

- The monocentric SUM is a simple and old model, but remains a useful analytical tool, consistent with many city structures across the world.
- It however much better describes aggregated variables on cities (e.g. urban area) than the precise internal structure of the cities.
- High income inequalities in cities, informal housing, and strong locational amenities are important factors explaining why the model fails in practice to capture some city structures.

- 1 Introduction
- 2 Data
- 3 Do global conclusions of the model hold?
 - Urbanized areas
 - Density gradients
- 4 Do internal mechanisms of the model hold?
 - Real estate pricing model
 - Housing production model
 - Robustness checks
- 5 Conclusion

- 1 Introduction
- 2 Data
- 3 Do global conclusions of the model hold?
 - Urbanized areas
 - Density gradients
- 4 Do internal mechanisms of the model hold?
 - Real estate pricing model
 - Housing production model
 - Robustness checks
- 5 Conclusion

- Africa
- Asia
- Europe
- North America
- Oceania
- South America

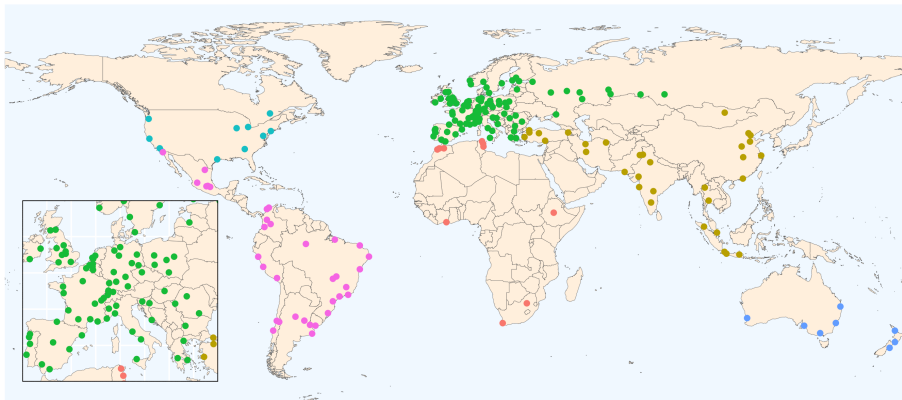


Figure: Sample of cities. *Total number of cities: 192.*

- **Land cover:** European Space Agency (ESA) data (300m resolution).

Spatially-explicit data

- **Land cover:** European Space Agency (ESA) data (300m resolution).
- **Population density:** GHSL data (250m resolution).

- **Land cover:** European Space Agency (ESA) data (300m resolution).
- **Population density:** GHSL data (250m resolution).
- **Real estate data** obtained from web scrapping of real estate websites, chosen following four criteria:
 - having a nation-wide coverage to ensure consistency in a given country.
 - geolocalizing the dwellings.
 - having values for both rents and dwelling sizes.
 - targeting local people, not expatriates.

- **Land cover:** European Space Agency (ESA) data (300m resolution).
- **Population density:** GHSL data (250m resolution).
- **Real estate data** obtained from web scrapping of real estate websites, chosen following four criteria:
 - having a nation-wide coverage to ensure consistency in a given country.
 - geolocalizing the dwellings.
 - having values for both rents and dwelling sizes.
 - targeting local people, not expatriates.
- **Transportation distances and durations** to the city center obtained from the Google Maps and Baidu Maps APIs.
 - We collected both driving and public transportation data, at typical afternoon rush hours.
 - City center: compromise between 5 criteria: geographical center, historical center, public transports hub, official CBD, and city hall.

- **Land cover:** European Space Agency (ESA) data (300m resolution).
- **Population density:** GHSL data (250m resolution).
- **Real estate data** obtained from web scrapping of real estate websites, chosen following four criteria:
 - having a nation-wide coverage to ensure consistency in a given country.
 - geolocalizing the dwellings.
 - having values for both rents and dwelling sizes.
 - targeting local people, not expatriates.
- **Transportation distances and durations** to the city center obtained from the Google Maps and Baidu Maps APIs.
 - We collected both driving and public transportation data, at typical afternoon rush hours.
 - City center: compromise between 5 criteria: geographical center, historical center, public transports hub, official CBD, and city hall.
- Aggregated over a 1km resolution grid encompassing the urban area.

Data at the city scale (1)

- **Urbanized area:** ESA data, summed over the grid.

Data at the city scale (1)

- **Urbanized area:** ESA data, summed over the grid.
- **Population:** GHSL data, also summed over the grid for consistency.

Data at the city scale (1)

- **Urbanized area:** ESA data, summed over the grid.
- **Population:** GHSL data, also summed over the grid for consistency.
- **Income:** GDP per capita at the country level from the World Bank.

Data at the city scale (1)

- **Urbanized area:** ESA data, summed over the grid.
- **Population:** GHSL data, also summed over the grid for consistency.
- **Income:** GDP per capita at the country level from the World Bank.
- **Agricultural rent:** agricultural GDP divided by the total agricultural area in the country (FAO data).

Data at the city scale (1)

- **Urbanized area:** ESA data, summed over the grid.
- **Population:** GHSL data, also summed over the grid for consistency.
- **Income:** GDP per capita at the country level from the World Bank.
- **Agricultural rent:** agricultural GDP divided by the total agricultural area in the country (FAO data).
- **Monocentricity index:** qualitative index that shows to which extent the geographical center, the historical center, the public transport hub, the official CBD and the town hall are at the same place.

Data at the city scale (1)

- **Urbanized area:** ESA data, summed over the grid.
- **Population:** GHSL data, also summed over the grid for consistency.
- **Income:** GDP per capita at the country level from the World Bank.
- **Agricultural rent:** agricultural GDP divided by the total agricultural area in the country (FAO data).
- **Monocentricity index:** qualitative index that shows to which extent the geographical center, the historical center, the public transport hub, the official CBD and the town hall are at the same place.
- **Fuel cost:** World Bank data on gasoline prices multiplied by IEA data on fuel efficiency.

- **Commuting speed:** density-weighted average speed to go to the city center at night rush hour using the fastest transportation mode, using Google/Baidu data.

- **Commuting speed:** density-weighted average speed to go to the city center at night rush hour using the fastest transportation mode, using Google/Baidu data.
- **Rent data quality:**
 - real estate data market cover.
 - real estate data spatial cover.

Data at the city scale (2)

- **Commuting speed:** density-weighted average speed to go to the city center at night rush hour using the fastest transportation mode, using Google/Baidu data.
- **Rent data quality:**
 - real estate data market cover.
 - real estate data spatial cover.
- **Informal housing data** and **Gini indices** from the World Bank.

- **Commuting speed:** density-weighted average speed to go to the city center at night rush hour using the fastest transportation mode, using Google/Baidu data.
- **Rent data quality:**
 - real estate data market cover.
 - real estate data spatial cover.
- **Informal housing data** and **Gini indices** from the World Bank.
- **Planned cities:** dummy based on Bertaud and Malpezzi (2014) for the cities in their dataset and on our judgement for the remaining cities.

- 1 Introduction
- 2 Data
- 3 Do global conclusions of the model hold?
 - Urbanized areas
 - Density gradients
- 4 Do internal mechanisms of the model hold?
 - Real estate pricing model
 - Housing production model
 - Robustness checks
- 5 Conclusion

Urbanized areas

	Urbanized area	ln(urbanized area)
population	0.000*** (0.000)	
ln(population)		0.856*** (0.039)
income	0.014*** (0.004)	
ln(income)		0.409*** (0.037)
land price	-169.932*** (34.984)	
ln(land prices)		-0.218*** (0.037)
fuel cost	-13199.329*** (3740.833)	
ln(fuel cost)		-0.160 (0.109)
commuting speed	39.789*** (7.215)	
ln(commuting speed)		0.448*** (0.103)
monocentricity	201.458* (107.085)	0.141* (0.082)
constant	-700.851** (350.325)	-13.349*** (0.644)
Observations	192	192
R^2	0.736	0.859
Adjusted R^2	0.728	0.854

Urbanized areas

	Urbanized area	ln(urbanized area)
population	0.000*** (0.000)	
ln(population)		0.856*** (0.039)
income	0.014*** (0.004)	
ln(income)		0.409*** (0.037)
land price	-169.932*** (34.984)	
ln(land prices)		-0.218*** (0.037)
fuel cost	-13199.329*** (3740.833)	
ln(fuel cost)		-0.160 (0.109)
commuting speed	39.789*** (7.215)	
ln(commuting speed)		0.448*** (0.103)
monocentricity	201.458* (107.085)	0.141* (0.082)
constant	-700.851** (350.325)	-13.349*** (0.644)
Observations	192	192
R^2	0.736	0.859
Adjusted R^2	0.728	0.854

Urbanized areas

	Urbanized area	ln(urbanized area)
population	0.000*** (0.000)	
ln(population)		0.856*** (0.039)
income	0.014*** (0.004)	
ln(income)		0.409*** (0.037)
land price	-169.932*** (34.984)	
ln(land prices)		-0.218*** (0.037)
fuel cost	-13199.329*** (3740.833)	
ln(fuel cost)		-0.160 (0.109)
commuting speed	39.789*** (7.215)	
ln(commuting speed)		0.448*** (0.103)
monocentricity	201.458* (107.085)	0.141* (0.082)
constant	-700.851** (350.325)	-13.349*** (0.644)
Observations	192	192
R^2	0.736	0.859
Adjusted R^2	0.728	0.854

Urbanized areas

	Urbanized area	ln(urbanized area)
population	0.000*** (0.000)	
ln(population)		0.856*** (0.039)
income	0.014*** (0.004)	
ln(income)		0.409*** (0.037)
land price	-169.932*** (34.984)	
ln(land prices)		-0.218*** (0.037)
fuel cost	-13199.329*** (3740.833)	
ln(fuel cost)		-0.160 (0.109)
commuting speed	39.789*** (7.215)	
ln(commuting speed)		0.448*** (0.103)
monocentricity	201.458* (107.085)	0.141* (0.082)
constant	-700.851** (350.325)	-13.349*** (0.644)
Observations	192	192
R^2	0.736	0.859
Adjusted R^2	0.728	0.854

Urbanized areas

	Urbanized area	ln(urbanized area)
population	0.000*** (0.000)	
ln(population)		0.856*** (0.039)
income	0.014*** (0.004)	
ln(income)		0.409*** (0.037)
land price	-169.932*** (34.984)	
ln(land prices)		-0.218*** (0.037)
fuel cost	-13199.329*** (3740.833)	
ln(fuel cost)		-0.160 (0.109)
commuting speed	39.789*** (7.215)	
ln(commuting speed)		0.448*** (0.103)
monocentricity	201.458* (107.085)	0.141* (0.082)
constant	-700.851** (350.325)	-13.349*** (0.644)
Observations	192	192
R^2	0.736	0.859
Adjusted R^2	0.728	0.854

Urbanized areas

	Urbanized area	ln(urbanized area)
population	0.000*** (0.000)	
ln(population)		0.856*** (0.039)
income	0.014*** (0.004)	
ln(income)		0.409*** (0.037)
land price	-169.932*** (34.984)	
ln(land prices)		-0.218*** (0.037)
fuel cost	-13199.329*** (3740.833)	
ln(fuel cost)		-0.160 (0.109)
commuting speed	39.789*** (7.215)	
ln(commuting speed)		0.448*** (0.103)
monocentricity	201.458* (107.085)	0.141* (0.082)
constant	-700.851** (350.325)	-13.349*** (0.644)
Observations	192	192
R^2	0.736	0.859
Adjusted R^2	0.728	0.854

- 1 Introduction
- 2 Data
- 3 Do global conclusions of the model hold?
 - Urbanized areas
 - Density gradients
- 4 Do internal mechanisms of the model hold?
 - Real estate pricing model
 - Housing production model
 - Robustness checks
- 5 Conclusion

$$\ln(n_i) = \alpha_1 + \beta_1 \ln(Y - T_i) + \beta_2 \ln(L_i) - \beta_3 \ln(q_i) + \epsilon_i$$

- We regress density n_i on income net of transportation costs $Y - T_i$. Our main parameter of interest is β_1 .
- Based on Qiang et al. (2020), we use commuting times and distances from Google Maps' data.
- Two additional contributions:
 - two transportation modes.
 - we account from both opportunity and monetary transportation costs.
- We control by land cover constraints L_i and dwelling sizes q_i , as predicted by the SUM.

Density gradients

Predictions of the SUM in terms of density gradient also hold on our sample of cities.

min.	0.029
10th	0.191
25th	0.252
50th	0.362
75th	0.477
90th	0.585
max.	0.734

Table: R2 distribution

Density gradients

Predictions of the SUM in terms of density gradient also hold on our sample of cities.

min.	0.029
10th	0.191
25th	0.252
50th	0.362
75th	0.477
90th	0.585
max.	0.734

Table: R2 distribution

Positive and significant (5%)	179 cities
Negative and significant (5%)	2 cities
Non significant (5%)	10 cities

Table: Parameter β_1

- 1 Introduction
- 2 Data
- 3 Do global conclusions of the model hold?
 - Urbanized areas
 - Density gradients
- 4 Do internal mechanisms of the model hold?
 - Real estate pricing model
 - Housing production model
 - Robustness checks
- 5 Conclusion

Do transportation costs impact rents?

$$\ln(R_i) = \alpha_2 + \beta_4 \ln(Y - T_i) + \epsilon_i$$

- We want to know how rents R_i depend on transportation costs.
- Our parameter of interest is β_4 ; it is expected to be positive and significant (households are willing to pay lower rents if transportation costs are high).

Predictions of the SUM hold in a majority of cities.

Positive and significant (5%)	159 cities
Negative and significant (5%)	2 cities
Non significant (5%)	30 cities

Table: Parameter β_4

Real estate pricing model

Predictions of the SUM hold in a majority of cities.

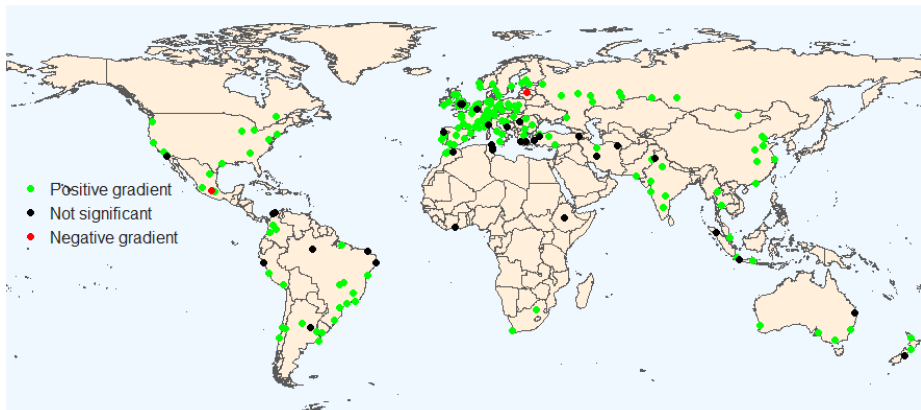


Figure: Estimates of the impact of income net of transportation costs on rents.

Real estate pricing model

	β_4 not significant
gini index	0.002 (0.004)
informal housing	0.006** (0.003)
planned city	0.011 (0.055)
monocentricity	0.022 (0.068)
coastal city	0.129** (0.056)
population	-0.000*** (0.000)
income	-0.000 (0.000)
market data cover	0.000 (0.000)
spatial data cover	-0.618* (0.316)
constant	0.103 (0.213)
Observations	190
R^2	0.186
Adjusted R^2	0.145

Real estate pricing model

	β_4 not significant
gini index	0.002 (0.004)
informal housing	0.006** (0.003)
planned city	0.011 (0.055)
monocentricity	0.022 (0.068)
coastal city	0.129** (0.056)
population	-0.000*** (0.000)
income	-0.000 (0.000)
market data cover	0.000 (0.000)
spatial data cover	-0.618* (0.316)
constant	0.103 (0.213)
Observations	190
R^2	0.186
Adjusted R^2	0.145

- 1 Introduction
- 2 Data
- 3 Do global conclusions of the model hold?
 - Urbanized areas
 - Density gradients
- 4 Do internal mechanisms of the model hold?
 - Real estate pricing model
 - **Housing production model**
 - Robustness checks
- 5 Conclusion

Do rents impact population densities?

$$\ln(n_i) = \alpha_3 + \beta_5 \ln(R_i) + \beta_6 \ln(L_i) - \beta_7 \ln(q_i) + \epsilon_i$$

- We want to know how housing production, and thus population density n_i depend on rents R_i .
- Our parameter of interest is β_5 ; it is expected to be positive and significant (developers are willing to build more dwellings if rents are high).

Predictions of the SUM hold in slightly more than half of the cities.

Positive and significant (5%)	106 cities
Negative and significant (5%)	15 cities
Non significant (5%)	70 cities

Table: Parameter β_4

Housing production model

Predictions of the SUM hold in slightly more than half of the cities.

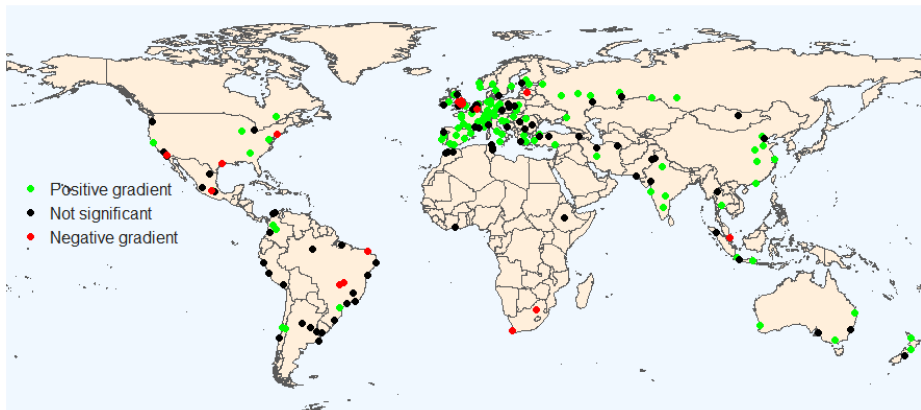


Figure: Estimates of the impact of rents on housing.

Housing production model

	(β_5 negative)	(β_5 not significant)
gini index	0.1759**	0.0340
	0.078	0.030
monocentricity	0.6093	0.1292
	0.939	0.459
informal housing	-0.0473	0.0488**
	0.067	0.023
spatial data cover	7.2976*	-8.2646**
	4.151	3.509
market data cover	-0.0001	-1.15e-06
	0.000	1.83e-06
planned city	0.1476	-0.1583
	0.675	0.352
coastal city	-0.3465	1.5579***
	0.647	0.413
population	-5.901e-08	-1.183e-07***
	8.03e-08	4.39e-08
income	1.142e-06	-2.354e-05*
	1.83e-05	1.36e-05
constant	-10.7329***	-1.0499
	3.533	1.630

Observations

192

Pseudo R^2

0.2469

Housing production model

	(β_5 negative)	(β_5 not significant)
gini index	0.1759**	0.0340
	0.078	0.030
monocentricity	0.6093	0.1292
	0.939	0.459
informal housing	-0.0473	0.0488**
	0.067	0.023
spatial data cover	7.2976*	-8.2646**
	4.151	3.509
market data cover	-0.0001	-1.15e-06
	0.000	1.83e-06
planned city	0.1476	-0.1583
	0.675	0.352
coastal city	-0.3465	1.5579***
	0.647	0.413
population	-5.901e-08	-1.183e-07***
	8.03e-08	4.39e-08
income	1.142e-06	-2.354e-05*
	1.83e-05	1.36e-05
constant	-10.7329***	-1.0499
	3.533	1.630

Observations

192

Pseudo R^2

0.2469

- 1 Introduction
- 2 Data
- 3 Do global conclusions of the model hold?
 - Urbanized areas
 - Density gradients
- 4 Do internal mechanisms of the model hold?
 - Real estate pricing model
 - Housing production model
 - Robustness checks
- 5 Conclusion

In this study, our main robustness concerns are about the quality of real estate data. We run 8 robustness checks:

- Aggregating rents data:
 - By taking the median instead of the mean in each grid cell.
 - By regressing rents on dwelling sizes in each grid cell.
- Excluding outliers using:
 - The boxplot method.
 - The percentile method.
 - The hampel method.
- Excluding the grid cells:
 - Where we have less than 4 data points.
 - Where we have less than 10 data points.
 - Where less than 50% of the area is urbanized.

- 1 Introduction
- 2 Data
- 3 Do global conclusions of the model hold?
 - Urbanized areas
 - Density gradients
- 4 Do internal mechanisms of the model hold?
 - Real estate pricing model
 - Housing production model
 - Robustness checks
- 5 Conclusion

Limitations of the study

3 main limitations:

- Despite the increasing availability of data on city structures, gathering data on some cities, e.g. African cities, remains difficult.

3 main limitations:

- Despite the increasing availability of data on city structures, gathering data on some cities, e.g. African cities, remains difficult.
- We could only study the most simple version of the SUM (monocentric / one income class), because of data availability on income distribution and employment sub-centers.

3 main limitations:

- Despite the increasing availability of data on city structures, gathering data on some cities, e.g. African cities, remains difficult.
- We could only study the most simple version of the SUM (monocentric / one income class), because of data availability on income distribution and employment sub-centers.
- It was not possible to run panel analysis because of real estate and transportation data, gathered at one point in time.

- Monocentric SUM is a simple and old model, but remains a useful analytical tool, coherent with many city structures across the world.

Conclusion

- Monocentric SUM is a simple and old model, but remains a useful analytical tool, coherent with many city structures across the world.
- Its conclusions on urbanized surface and on density profile seem consistent with a large majority of the cities in our sample.

Conclusion

- Monocentric SUM is a simple and old model, but remains a useful analytical tool, coherent with many city structures across the world.
- Its conclusions on urbanized surface and on density profile seem consistent with a large majority of the cities in our sample.
- Its conclusions on the internal structure of cities are less empirically relevant, especially for the housing production model.

- Monocentric SUM is a simple and old model, but remains a useful analytical tool, coherent with many city structures across the world.
- Its conclusions on urbanized surface and on density profile seem consistent with a large majority of the cities in our sample.
- Its conclusions on the internal structure of cities are less empirically relevant, especially for the housing production model.
- One reason is the inertia of cities evolutions: densities and built environment are evolving slowly, whereas rents and land prices can change quickly.

- Monocentric SUM is a simple and old model, but remains a useful analytical tool, coherent with many city structures across the world.
- Its conclusions on urbanized surface and on density profile seem consistent with a large majority of the cities in our sample.
- Its conclusions on the internal structure of cities are less empirically relevant, especially for the housing production model.
- One reason is the inertia of cities evolutions: densities and built environment are evolving slowly, whereas rents and land prices can change quickly.
- High income inequalities in cities, informal housing, and strong locational amenities are important factors explaining why the model fails in practice to capture some city structures.

Thank you for you attention!

charlotte.liotta@gmail.com