Big data and a Spatial Hedonic Approach:
Addressing the land market information gap and estimating land prices’ determinants

Professor Cynthia Goytia and Guadalupe Dorna
Torcuato Di Tella University- CIPUV
Washington DC, June 2017
Introduction

Filling some gaps in the urban economics literature:

I. Lack of systematic and accessible information on land prices for metropolitan regions

- Big Data analytics to build CIPUV Land Price Atlas

II. Few studies on land price determination in contexts of rapid urban expansion

- → Hedonic price model to evaluate expectations of future development as determinants of pricing
Big Data Analytics used to collect, map, systematize and analyze land price data

- **Data collection:** Mainstream real-estate webpage scrapping → monthly parcel information (size, attributes, price, etc) and location
- **Mapping:** data is mapped based on published location
- **Systematizing data:** based on spatial location, lots are linked to further geo-referenced data (national census 2010, accessibility, zoning)
- **Analysis:** Visualize dynamics by differing aggregation levels

Available online at: http://atlas-cipuv.utdt.edu/#/
Accessibility indicators

• Total number of labour opportunities that can be reached within a given time.

• The higher the quantity of employment opportunities reachable within a realistic time extent, will increase the bid for plots at a given place and, consequently, equilibrium land values.

• Open Trip Planner Analyst (OTPA) -US (Levinson,1998), Buenos Aires (Peralta Quiros and Mehandaratha, 2014).

• Travel time from each origin-destination pair in the city, computed using the road network (OpenStreetMap) and transit attributes – (road network GIS layer and General Transit Feed Specification)
Atlas CIPUV de Precios del Suelo de la Región Metropolitana de Buenos Aires

<table>
<thead>
<tr>
<th>Media precio x m²</th>
<th>Junio 2016</th>
<th>Absoluta</th>
<th>Relativa</th>
<th>Localidades</th>
<th>lat</th>
<th>lng</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map</td>
<td>Satellite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15,954 TERRENO

Filtros

- Partidos
  - e.g.: Almirante Brown

Precio del terreno

- Desde 1 hasta 1,111,111 USD

Superficie

- Desde 100 hasta 50,000 m²

Precio por m²

- Desde 0 hasta 300 USD

Zonificación

Necesidades Básicas Insatisfechas

Cobertura de servicios básicos

Google Map data ©2016 Google Imagery ©2016 TerraMetrics | Terms of Use | Report a map error
Atlas CIPUV de Precios del Suelo de la Región Metropolitana de Buenos Aires

Filtros

- Partidos
- Precio de terreno
- Superficie
- Precio por m2
- Zonificación
- Necesidades Básicas Insatisfechas
- Cobertura de servicios básicos
- Densidad poblacional
- Acceso al trabajo en auto
- Acceso al trabajo en transporte público

21,860 TERRENOS
Información general

<table>
<thead>
<tr>
<th>Mínimo</th>
<th>Mediana</th>
<th>Máximo</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 (Almirante Brown)</td>
<td>208 (Esteban Echeverría)</td>
<td>1,875 (La Matanza)</td>
</tr>
</tbody>
</table>
CIPUV Land Price Atlas for MRBA
Spatial Hedonic Approach using the CIPUV Land Price Atlas

• MRBA: the largest urban agglomeration of Argentina

• Very fast urban expansion over the last decade (Angel, 2011)

→ interesting context to analyze price determinants (Capozza and Hesley, 1990)
Urban Expansion

Year 1990

Legend
- Limites Municipales
- Urbano
- Sub-Urbano

Source: Geo-Mex, IGME, INEGI, USAID, 1990

Urban Expansion

Year 2001

Legend
- Limites Municipales
- Urbana
- SubUrbana

Urban Expansion

Year 2010

Legend
- Límites Municipales
- Urbano
- Suburbano

Urban Expansion

1990-2000: 168 km² (16%)
2001-2010: 465 km² (37%)
Methodology overview

**Hedonic price method approach** → infer marginal valuation of attributes associated to the lot

**Sequential log-linear/log-log regression**

![Diagram](image)

- **Land price**
  - Structural characteristics of the parcel and neighbourhood
  - Accessibility
  - Expectations of future developments

Relevant in the context of fast growing urban areas
Methodology overview: Explanatory variables

Land characteristics:
- Plot’s area
- Population density
- Infrastructure coverage (water, gas and sanitation)

Structural characteristics of the parcel and neighbourhood

Accessibility
- Distance to train-metro stations, highways, CBD.
- Jobs accessed in 1 hour

Externalities, proxied by:
- Education level and distance to informal settlements

Land price

Expectations of future development

Expectations - variations over the last decade on:
- Infrastructure coverage
- Average adult educational levels
- UBN
- Population density in the vicinity of the lot

Relevant in the context of fast growing urban areas
Explanatory variables

• **Plot characteristics:**
  - plot’s area (linear and quadratic term),
  - population density measured as inhabitants per hectare of built-up land
  - infrastructure coverage index (access to basic public services - water, gas and sewage)

• **Accessibility:**
  - Distance to train stations, highways, Central Business District (CBD) and subways
  - Number of jobs attainable by car and by public transport within 1 hour of travel away from the lot

• **Externalities**: socio-economic characteristics of the neighboring area that exert prices boosts and hindrances, proxied by:
  - Average years of education of the adult population
  - distance to nearest informal settlements

• **Expectations**: variations over the last decade on:
  - infrastructure coverage,
  - adult educational levels,
  - unsatisfied basic needs
  - population density in the vicinity of the lot.
  - new developments between 1990-2000 and 2000-2010 within a 500 meter buffer in which the land is located

• **Zoning**
Estimated models

Model 1:

\[ \text{Ln}(\text{Price}) = \alpha + \text{Land Characteristics}' \beta_1 + \epsilon \]  
(1)

Model 2:

\[ \text{Ln}(\text{Price}) = \alpha + \text{Land Characteristics}' \beta_1 + \text{Controls}' \beta_c + \epsilon \]  
(2)

Model 3:

\[ \text{Ln}(\text{Price}) = \alpha + \text{Land Characteristics}' \beta_1 + \text{Accessibility}' \beta_2 + \text{Controls}' \beta_c + \epsilon \]  
(3)

Model 4:

\begin{align*}
\text{Ln}(\text{Price}) &= \alpha + \text{Land Characteristics}' \beta_1 + \text{Accessibility}' \beta_2 + \text{Externalities}' \beta_3 + \text{Controls}' \beta_c + \epsilon \\
&\quad + \text{Expectations}' \beta_4 + \epsilon
\end{align*}
(4)

Model 5:

\begin{align*}
\text{Ln}(\text{Price}) &= \alpha + \text{Land Characteristics}' \beta_1 + \text{Accessibility}' \beta_2 + \text{Externalities}' \beta_3 + \text{Expectations}' \beta_4 + \text{Controls}' \beta_c + \epsilon
\end{align*}
(5)

Model 6:

\begin{align*}
\text{Ln}(\text{Price}) &= \alpha + \text{Land Characteristics}' \beta_1 + \text{Accessibility}' \beta_2 + \text{Externalities}' \beta_3 + \text{Expectations}' \beta_4 + \text{Zoning}' \beta_5 + \text{Controls}' \beta_c + \\
&\quad + \epsilon
\end{align*}
(6)

Estimated by OLS. Categories of attributes added sequentially to assess robustness of results

Where

\( \text{Ln Price} \) is the (nx1 vector) of the natural logarithm of land price per sqm. for each of n observations

municipalities fixed effects amount for the control variables include in every specification of the model

\( \epsilon \) (nx1 vector) - the error term in each model's specification
Results: overall model’s prediction capability

Our HPM, in its more complete version, can explain about 77% of the variation in land prices in the MRBA.
Main results

• Some results we find are well documented in the literature; others are not.

• Most ubiquitous results found in the land price literature:
  • *accessibility and socio-economic characteristics of the neighborhood – mainly education -, acting as externalities, affect land prices.*
    • Positive relationship between the number of jobs attainable, particularly, those that can be reached by public transport.
    • And that a one-year increase in average years of education is associated with nearly a 10% increase in land prices.
Residential sub-urbanization and a trend to \textit{employment decentralization}.

Accessibility indicators to test: polycentric spatial structure vs monocentric framework.
Main results: accessibility

• Main results → both forces in place at the moment:
  • land prices positively related to the number of jobs attainable by public transport: net-effect of positive accessibility effects and negative congestion externalities
  • the monocentric approach increases model explanatory power (prices decrease $3 for a 10% increase in distance to CBD)
Implicit prices associated with Accessibility

Map shows log price differentials. It compares:
The price predicted by the model for each observation using all accessibility variables
to the price predicted for an ad-hoc observation with mean values for each explanatory variable corresponding to this group.

Accessibility values → highly correlated in space, explicit monocentric relationship

CBD concentrates both jobs and public transport networks
→ proximity to the city center adds value to the land, even after controlling for the rest of cofounders.
Main results: Expectations

• Area for further exploration: price dynamics in time series to understand the role of expectations. However, some hints with our HPM

• We find that in the context of the rapid urban growth of GBA, expectations are relevant price determinants to investigate.

  • Areas that have seen reduced shares of people living with unsatisfied basic needs exhibit higher land prices in comparison to those that have either stayed the same or have impoverished.
Implicit prices associated with Expectations

- Expectations seem to push prices up, controlling for the rest of covariates, as we move further away towards peri-urban area.

- Pattern not as linear as with accessibility but still clustered near the CBD (low values) and over the 3rd ring (high values)
Despite observed relationship, services become irrelevant in determining lot price as we include covariates.

“High correlation with accessibility”

“Context” $\rightarrow$ Expectations of future development: other things equal, many new urban developments that lack infrastructure, are expected to have it in the near future and this expectation might already be incorporated in price.
Implicit prices associated with Expectations and Zoning

Suggestive results for peri-urban areas: display the higher implicit values, after controlling for the rest of explanatory variables.

- The construction potential can explain price differentials (particularly in the City of BA)
- But still peri-urban areas exhibit higher than expected prices

One intuition:
Expectations of increased value for turning rural into urban land (gated urbanizations and lax regulatory environment allowing changes in land use)
Main results: Implicit prices associated with Expectations and Zoning

<table>
<thead>
<tr>
<th>Zoning</th>
<th>% change in land value</th>
<th>Implicit price in relation to average land price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designation (vs Rural)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other uses rural</td>
<td>305.93%</td>
<td>$121*</td>
</tr>
<tr>
<td>Lot in GC area</td>
<td>223.49%</td>
<td>$88</td>
</tr>
<tr>
<td>Centrality and commercial areas</td>
<td>228.38%</td>
<td>$90*</td>
</tr>
<tr>
<td>Nonresidential peri-urban area</td>
<td>277.73%</td>
<td>$110*</td>
</tr>
<tr>
<td>High density residential</td>
<td>216.77%</td>
<td>$86</td>
</tr>
<tr>
<td>Low density residential in urban area</td>
<td>206.49%</td>
<td>$82</td>
</tr>
<tr>
<td>Medium density residential in urban area</td>
<td>220.59%</td>
<td>$87</td>
</tr>
<tr>
<td>Mixed residential in urban area</td>
<td>168.32%</td>
<td>$66</td>
</tr>
<tr>
<td>Residential in peri-urban area</td>
<td>222.84%</td>
<td>$88*</td>
</tr>
<tr>
<td>Industrial</td>
<td>150.93%</td>
<td>$60</td>
</tr>
<tr>
<td>Other categories lots</td>
<td>255.73%</td>
<td>$101*</td>
</tr>
<tr>
<td>GC zoned rural</td>
<td>148.18%</td>
<td>$58</td>
</tr>
<tr>
<td>GC zonificado GC</td>
<td>218.36%</td>
<td>$86</td>
</tr>
<tr>
<td>GC consolidated in urban area</td>
<td>147.44%</td>
<td>$58</td>
</tr>
<tr>
<td>GC semi consolidated</td>
<td>183.77%</td>
<td>$73</td>
</tr>
<tr>
<td>GC in recovery area</td>
<td>156.51%</td>
<td>$62</td>
</tr>
</tbody>
</table>

Changes in zoning type with respect to rural.

They all generate an increase in price.

Only 5 of the 16 alternative zonings—other than rural—show significant coefficients in a positive direction.
### Main results: Implicit prices

<table>
<thead>
<tr>
<th>% change in land value</th>
<th>Implicit price in relation to average land price</th>
<th>Variation equivalent in units (with respect to average)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% increase</td>
<td>Average price: $519</td>
<td></td>
</tr>
<tr>
<td>Distance to CBD</td>
<td>0.58% decrease in land prices</td>
<td>-$3</td>
</tr>
<tr>
<td>Distance to subways</td>
<td>1.94% decrease in land prices</td>
<td>-$10</td>
</tr>
<tr>
<td>1 percentage point increase</td>
<td>Average price: $519</td>
<td></td>
</tr>
<tr>
<td>Ratio of jobs attainable by public transport to those reachable by car</td>
<td>1.11% increase in land prices</td>
<td>$6</td>
</tr>
<tr>
<td>Average years of education</td>
<td>9.64% increase in land prices</td>
<td>$50</td>
</tr>
<tr>
<td>10% increase</td>
<td>Average price: $519</td>
<td></td>
</tr>
<tr>
<td>Distance to shantytown</td>
<td>0.44% increase in land prices</td>
<td>$2</td>
</tr>
</tbody>
</table>

#### Change in average years of education

- If education in 2001 < 7 years
  - 12.64% increase in land prices
  - $66
- If education in 2001 >= 7 years and <= 10 years
  - 9.43% decrease in land prices
  - -$49
- If education in 2001 > 10 years
  - 14.79% decrease in land prices
  - -$77

#### Designation (vs no change in UBN*)

- Average price of parcels w/ no change in UBN: $864

- Change in population density
  - 7.79% increase in land prices
  - $85

#### Designation (vs no change in density)

- Average price of parcels w/ no change in density: $1087

<table>
<thead>
<tr>
<th>Other rural uses</th>
<th>Average price of rural parcels: $39</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centrality and commercial areas</td>
<td>312.38% increase in land prices $90</td>
</tr>
<tr>
<td>Nonresidential peri-urban area</td>
<td>277.73% increase in land prices $110</td>
</tr>
<tr>
<td>Residential in peri-urban area</td>
<td>222.84% increase in land prices $88</td>
</tr>
<tr>
<td>Other zoning categories</td>
<td>255.73% increase in land prices $101</td>
</tr>
</tbody>
</table>
Results: overall model’s prediction capability - Hot Spot Analysis

- HPM model unable to capture the high prices observed in the northern region of the city of Buenos Aires and its northern neighbors (the model predicts prices that are nearly half of those observed)

- Instead the model predicts higher prices than those observed in the southern region of the city of Buenos Aires and its immediate southern neighbor as well as in the Northern outskirts
Results: overall model’s prediction capability - Hot Spot Analysis

Possible ways to correct this spatial correlation of errors:

- INCORPORATED IN OUR ANALYS - Spatial Autoregressive Model (SAR)
- NOT INCORPORATED IN OUR ANALYS - Geographically weighted regression (GWR) → coefficients can only be interpreted locally so no overall conclusion can be drawn upon explanatory variables
- FOR FURTHER EXPLORATION - differential constructability. Include interaction term between zoning and being in the City.
Further analysis: spatial specification of our model and robustness checks

• The spatial agglomeration of residuals suggests that there are some characteristics that could be spatially related, that need to be accounted for.

• The spatial specification is a simple Spatial Autoregressive Model (SAR). It includes as an explanatory variable the spatially lagged value of land price (captured in coefficient Lambda)
  • Though statistically significant, its magnitude is very small implying that, though positive, the explanatory power that neighboring prices have over the lot price is small when compared to the rest of covariates.

• We modify the spatial buffer we use to compute neighboring characteristics.
  • coefficients are robust to these modifications: i.e. they preserve both magnitude and statistical significance.

• The SAR model, however, seems to exhibit lower standard errors and in consequence some of the coefficients that were previously deemed not statistically significant now become relevant.

• This result should be taken cautiously since one drawback of the SAR specification is that does not allow for error clustering (clustered standard errors instead of robust estimators) in the same fashion we have included in all preceding model specifications.
Summing up

• Big Data can bridge a gap in land prices data

• Urban growth expectations on metro areas. Urban land as a real option, the value of which is determined by the highest and best use rather than the current use (Capozza and Helsley, 1990)

• Interesting results related to the value that both (changes in) zoning in peri-urban areas and the relative level of urban development play in land price determination.
Thank you
cgoytia@utdt.edu