Dissertation Defense

# Sustainable Metropolitan Growth Strategies: Exploring the Role of the Built Environment

August 2010

MI DIAO
Ph.D. Candidate
Department of Urban Studies and Planning
Massachusetts Institute of Technology



## Outline



- Research Context
  - Sustainable metropolitan development and smart growth strategies
- The Built Environment
  - Datasets
  - Built-Environment Indicators
  - Factor analysis
- Three Essays
  - Built-environment effect on household vehicle usage (Essay 1)
  - Built-environment effect on residential property values (Essays 2 and 3)
- Summary
  - Implications for growth management
  - Implications for urban modeling

## Abbreviations

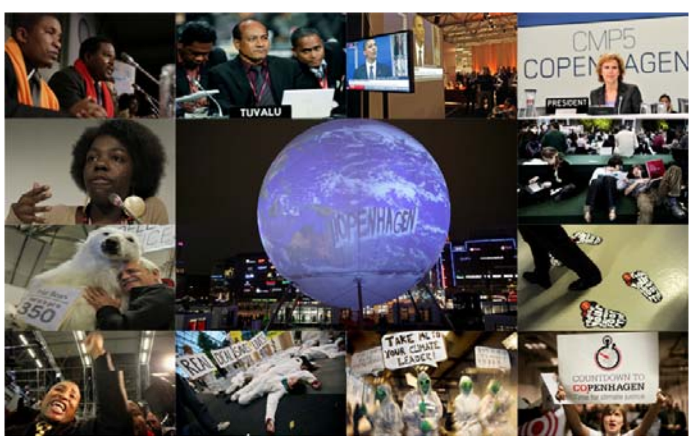


- GHG: Greenhouse Gas
- BE: Built Environment
- DEM: Demographic
- VMT: Vehicle Miles Travelled
- MAUP: Modifiable Areal Unit Problem
- WTP: Willingness-to-Pay
- RMV: Registry of Motor Vehicles
- MassDOT: Massachusetts Department of Transportation

## Research Context

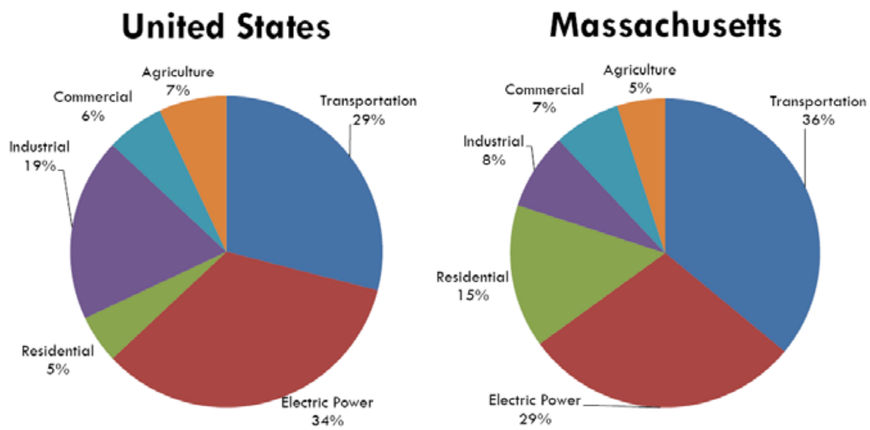


- Background: greenhouse gas (GHG) emission, global climate change, and sustainability



## Carbon Emissions by Sector

5



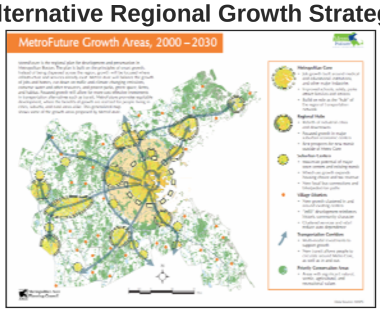

By Courtesy of Tim Reardon (Metropolitan Area Planning Council)
Source: MA Department of Transportation, from US Energy Information Administration

## Strategies

6

- Strategies to reduce transportation GHG
  - Technological – hybrids, biofuels, electric cars
  - Economic – taxes, congestion pricing etc.
  - **Physical** – smart growth (high density, mixed use, mass transit, job-housing balance, etc.)

**Alternative Regional Growth Strategy**




**New Urbanism Design**

## Research Summary



- Highlights of my dissertation
  - Newly available administrative datasets with spatial details
    - Safety inspection records from Registry of Motor Vehicles
    - Housing transaction records from Registry of Deeds
    - Assessing records from the Assessing Department
    - Built-environment data layers from MassGIS – the State's Office for Geographic and Environmental Information
  - Disaggregated built environment indicators
    - Utilizing GIS and database management system tools to compute a set of improved indicators to characterize the built environment at disaggregated level
  - Quantitative models
    - Integrating built-environment indicators into quantitative models to investigate the relationship between the built environment, household vehicle usage and residential property values.

## Dissertation Structure

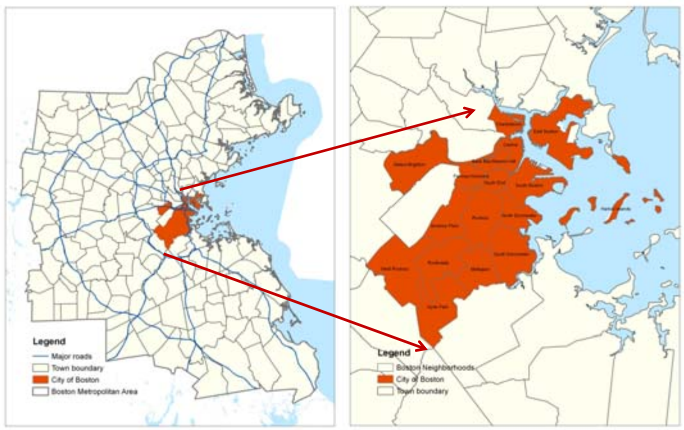


- One Major Theme:
  - Built environment
- Two Focus Areas
  - Transportation
  - Housing
- Three Research Essays
  - Vehicle Miles Traveled and the Built Environment: Evidence from Vehicle Safety Inspection Data in the Boston Metropolitan Area
  - Residential Property Values and the Built Environment: An Empirical Study in the Boston Metropolitan Area
  - Selectivity Effects, Spatial Autocorrelation, and Valuation of the Built Environment

## Study Area

9

**Boston Metropolitan Area** **City of Boston**




- 164 Municipalities

## Built-Environment Datasets

10

- Dun & Bradstreet business location database
  - Locations of business establishments
- Institutions and other destinations data from MassGIS
  - Locations of schools, hospitals, parks, etc.
- Census 2000
  - Spatial distribution of individuals and households
- Land use data from MassGIS
  - Land use classification
- Boston road inventory database from MassDOT
  - Number of lanes, central separation, curb presence, shoulder width, sidewalk width, speed limit, turn direction, etc.

## Spatial Units



- The basic study unit is 250x250m grid cells (119,834 grid cells in the metro area)
- For each grid cell, define a catchment area as the 3X3 nearest grid cells.
- Variable of interest is computed for the catchment area, and then assigned to the grid cell in the middle.

250X250m grid cells | Block groups | Census tracts

## Selection of BE Variables



### Distance to Destinations

**Non-Work**

- Distance to shopping mall
- Distance to grocery
- Distance to hardware store
- Distance to school
- Distance to restaurant
- Distance to church
- Distance to dentist
- Distance to gym

**Work**

- Job accessibility

### Transport Systems

- Pct of 4-way intersections
- Pct of roads w. access control
- Pct of roads over30mph limit
- Pct of roads with curbs
- Pct of roads with sidewalks
- Density of Roads
- Density of intersections
- Density of 3-way intersections
- Density of 4-way intersections
- Distance to subway stations
- Distance to comm. rail stations
- Distance to bus stops
- Distance to MBTA parking lots
- Distance to Highway Exits
- Average Road Width
- Average sidewalk width

### Other Variables

- Population density
- Land use mix

## Factor Analysis for BE Variables

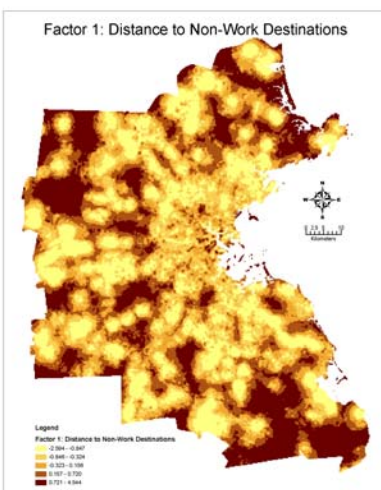


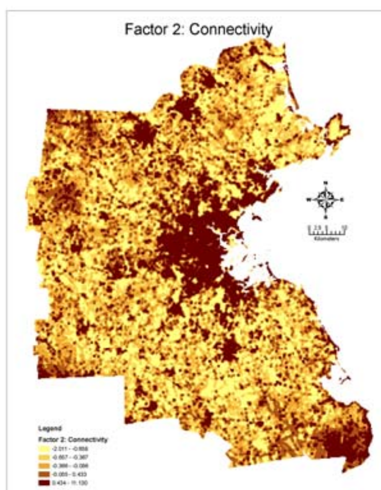
| | Variables | Factor 1 Dis. to non-work destinations | Factor 2 Connectivity | Factor 3 Inaccessi-bility to transit&job | Factor 4 Auto dominance | Factor 5 Walkability |
|---|---|---|---|---|---|---|
| 1 | Dis. to restaurant | 0.784 | | | | |
| 2 | Dis. to mall | 0.764 | | | | |
| 3 | Dis. to hardware store | 0.746 | | | | |
| 4 | Dis. to grocery | 0.733 | | | | |
| 5 | Dis. to dentist | 0.688 | | 0.398 | | |
| 6 | Dis. to gym | 0.676 | | | | |
| 7 | Dis. to church | 0.674 | | | | |
| 8 | Dis. to school | 0.645 | | | | |
| 9 | Land use mix | -0.480 | | | | |
| 10 | Den. of 4-way intersections | | 0.872 | | | |
| 11 | Intersection density | | 0.849 | | | |
| 12 | Den. of 3-way intersections | | 0.809 | | | |
| 13 | Population density | | 0.785 | | | |
| 14 | Road density | -0.353 | 0.765 | | | |
| 15 | Pct. of 4-way intersections | | 0.609 | | | |
| 16 | Dis. to MBTA bus stops | | | 0.833 | | |
| 17 | Dis. to comm. rail station | | | 0.810 | | |
| 18 | Dis. to subway stations | | | 0.801 | | |
| 19 | Dis. to MBTA parking lots | | | 0.775 | | |
| 20 | Job accessibility | | 0.486 | -0.636 | | |
| 21 | Pct. of road with access ctl. | | | | 0.910 | |
| 22 | Average road width | | | | 0.875 | |
| 23 | Pct. of road w. 30+ sp. limit | | | | 0.856 | |
| 24 | Dis. to highway exits | | | | -0.362 | |
| 25 | Pct. of road with sidewalks | | | | | 0.910 |
| 26 | Pct. of road with curbs | | | | | 0.908 |
| 27 | Average sidewalk width | | 0.583 | | | 0.602 |

* Principle component analysis using Varimax rotation

## Maps of BE Factors





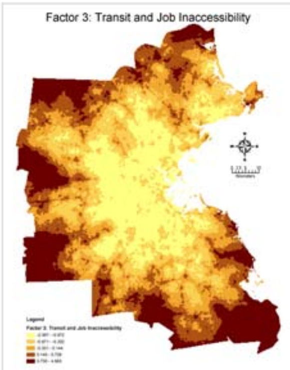



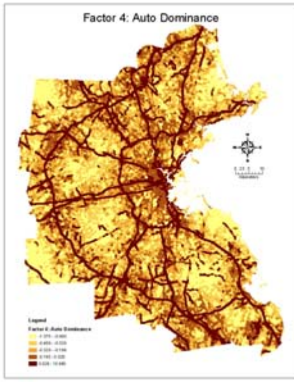

**Classification method: quantile.**

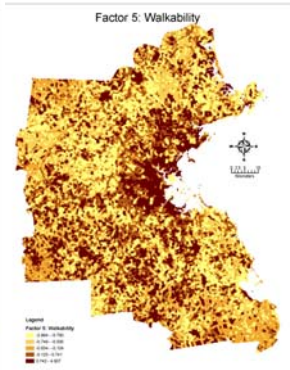
Darker = higher factor score

## Maps of BE Factors






**Classification method: quantile.**

Darker = higher factor score

## Essay One: Vehicle Miles Travelled and the Built Environment: Evidence from Vehicle Safety Inspection Data in the Boston Metropolitan Area

**Transportation**

## Research Objectives and Analytical Approaches



- Research objectives
  - Establish a baseline for transportation $CO_2$ emissions
  - Build models to explain VMT with geographic variables
    - evaluate significance of “smart growth” factors
    - quantify effectiveness of local and state level strategies to reduce $CO_2$ emissions

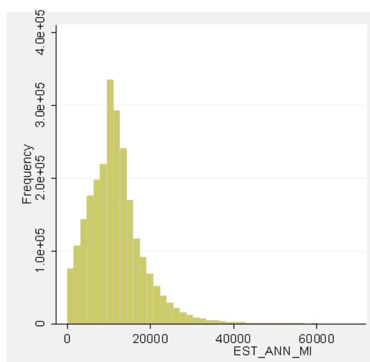
- Analytical approaches
  - Region-wide, cross-sectional
  - Unit of analysis: 250x250m grid cells
  - VMT = f ( BE, DEM )
  - Factor analysis to mitigate multicollinearity
  - Spatial econometric techniques to address spatial autocorrelation (Anselin, 1998)

## VMT Estimates from Safety Inspection Records



- Annual safety inspection records from Registry of Motor Vehicles (RMV), 2005-2007
  - All private passenger cars (2.5 million in Metro Boston)
  - Odometer mileages reported to RMV

- MassGIS estimates annual VMT using safety inspection records
  - Estimate annual mileage for 2.5m vehicles from safety inspection records
  - Geocode every vehicle to place of residence
  - Aggregate mileage to 250x250m grid cell



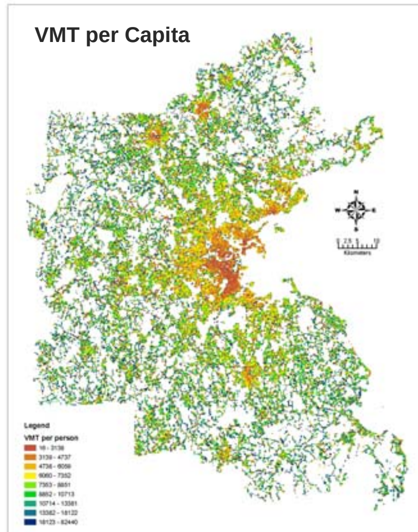

**Frequency Distribution of Annual VMT/Vehicle**

## VMT Maps (Grid Cell Level)



Spatial Pattern in Annual Vehicle Miles Traveled (VMT) by

250x250m grid cell




**Red** = low VMT

**Blue** = high VMT

## Factor Analysis -- Demographic Variables



- Using principle component analysis with Varimax rotation, 3 factors are extracted from 12 demographic variables, which explain 71.56% of variance in the original variables.
- Factor loadings:

| | | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|---|
| | | Wealth | Children | Working Status |
| 1 | Pct. of population below poverty level | -0.863 | | |
| 2 | Pct. of owner-occupied housing units | 0.818 | 0.386 | |
| 3 | Pct. of population with at least 13 years of schooling | 0.817 | | |
| 4 | Median household income | 0.812 | | |
| 5 | Pct. of population that is white | 0.796 | | |
| 6 | Per capita income | 0.707 | | |
| 7 | Unemployment rate | -0.613 | | |
| 8 | Pct. of households with less than 3 members | | -0.909 | |
| 9 | Pct. of population 3+ yrs that are enrolled in elem./high school | | 0.869 | |
| 10 | Pct. of population under 5 | | 0.728 | |
| 11 | Pct. of population 65 years old and over | | | -0.856 |
| 12 | Pct. of population 16 years old and over in labor force | 0.427 | | 0.793 |

## Model Specification



- OLS Model

$$VMT_i = \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i$$

- Spatial Lag Model

$$VMT_i = \rho W_{VMT_i} + \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i$$

- Spatial Error Model

$$VMT_i = \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i$$

$$\varepsilon_i = \lambda W_{\varepsilon_i} + \mu_i$$

Where
-- *VMT* is vehicle miles travelled per vehicle (per household, per capita) computed for the catchment area of each grid cell;
-- *BE* is built environment factor;
-- *DEM* is demographic factor.

## Model Estimation Summary Statistics



- Nine models are calibrated:
  - OLS model for VMT per vehicle
  - OLS model for VMT per household
  - OLS model for VMT per capita
  - Spatial lag model for VMT per vehicle
  - Spatial lag model for VMT per household
  - Spatial lag model for VMT per capita
  - Spatial error model for VMT per vehicle
  - Spatial error model for VMT per household
  - Spatial error model for VMT per capita
- Summary Statistics

| | VMT per Vehicle | | | VMT per Household | | | VMT per Capita | | |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | Spatial Lag | Spatial Err. | OLS | Spatial Lag | Spatial Err. | OLS | Spatial Lag | Spatial Err. |
| Observations | 52929 | 52929 | 52929 | 52929 | 52929 | 52929 | 52929 | 52929 | 52929 |
| R-squared | 0.527 | 0.789 | 0.810 | 0.418 | 0.626 | 0.631 | 0.342 | 0.566 | 0.573 |
| Log Likelihood | -451127 | -432073 | -429930 | -563448 | -553582 | -553497 | -505660 | -496458 | -496291 |
| Test | Statistic | p-value | | Statistic | p-value | | Statistic | p-value | |
| LM--Lag | 86355.0 | 0.00 | | 43966.2 | 0.00 | | 41094.4 | 0.00 | |
| LM--Error | 115402.4 | 0.00 | | 46425.7 | 0.00 | | 43147.3 | 0.00 | |
| Robust LM--Lag | 621.6 | 0.00 | | 619.4 | 0.00 | | 305.3 | 0.00 | |
| Robust LM--Err. | 29669.0 | 0.00 | | 3078.8 | 0.00 | | 2358.1 | 0.00 | |

## Estimation Results (Spatial Error Model)

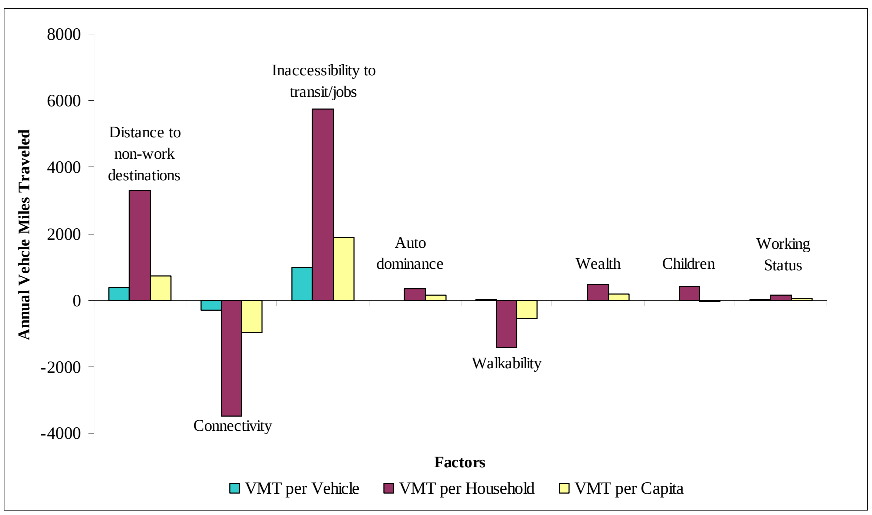


- Demographic factors
  - Wealth:
    - (+) for VMT/HH and VMT/Person
    - (-) for VMT/VIN
  - Children
    - (+) for VMT/HH
  - Working status
    - (+) for all VMT variables

| | VMT per Vehicle | | VMT per Household | | VMT per Capita | |
|---|---|---|---|---|---|---|
| | Coef. | t | Coef. | t | Coef. | t |
| DEM fac. 1: wealth | -26.9 | -2.0 * | 737.7 | 5.5 ** | 296.9 | 6.6 ** |
| DEM fac. 2: children | -9.1 | -1.0 | 545.5 | 5.9 ** | -45.9 | -1.5 |
| DEM fac. 3: working status | 29.6 | 4.4 ** | 160.3 | 2.3 * | 58.1 | 2.5 * |

## Built-Environment Effect on VMT





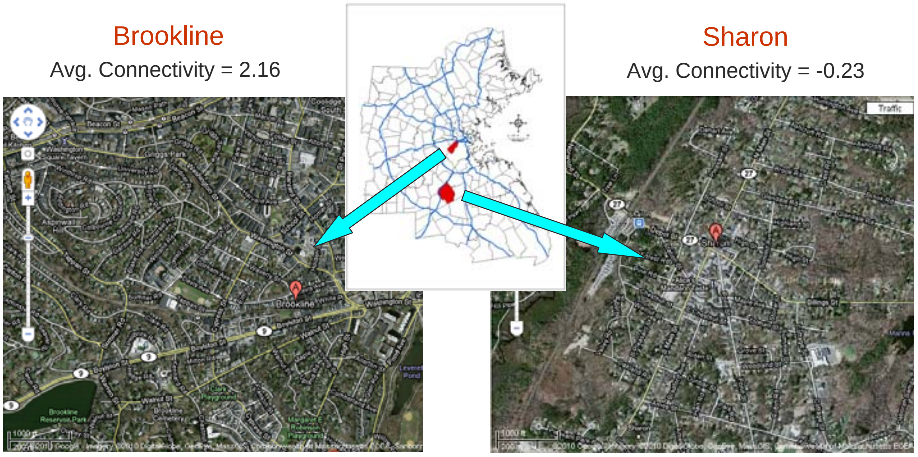

- **Change in VMT measures due to one standard deviation increase in factors**

## Example



**Brookline**
Avg. Connectivity = 2.16

**Sharon**
Avg. Connectivity = -0.23



**Std. dev. of the Connectivity factor= 1.172**
**1 std. dev. of Connectivity = -3480.5 mi (annual VMT per household)**
**Diff. between Sharon and Brookline is 2 std. dev., which is equivalent to 6,960 mi in VMT/HH.**

## Major Findings of Essay One



- Models explain 57.3-81.0% of VMT variation across grid cells
  - Built-environment factors
  - Demographic factors
- Significant built-environment effects
  - Increasing accessibility to work and non-work destinations, connectivity, and transit accessibility can significantly reduce VMT.
- Interesting demographic effects
  - Higher income → lower VMT per vehicles (but more vehicles)
  - Higher income → higher VMT per household
- Built environment factors have higher impacts on VMT than demographic factors

## Essay Two: Residential Property Values and the Built Environment: an Empirical Study in Boston Metropolitan Area

Housing



## Introduction

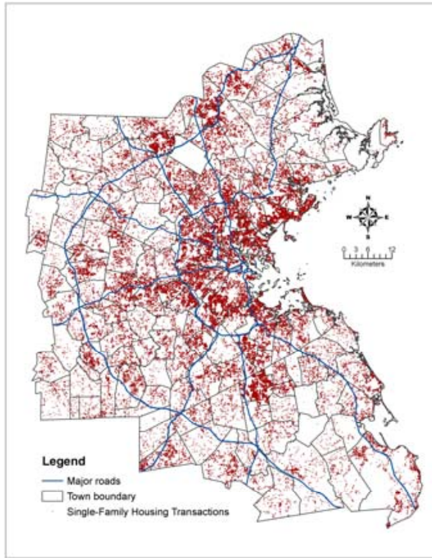


- Importance of understanding how housing market values the built environment:
  - To estimate housing price impact of land use change
  - To quantify the implicit tradeoffs associated with the positive (negative) impacts of smart growth policies on local neighborhoods.
  - To provide a potential financing mechanism for infrastructure investment through land value capture

## Major Datasets



- **Housing transaction records** from city and town assessors assembled by the Warren Group
- 92,774 single-family housing transactions in the Boston Metropolitan Area during 2004-2006 with detailed structural characteristics
- Same built-environment data as in Essay one.




## Model Specification



- Six models are estimated depending on the estimation method and the choice of factors:
  - Model 1: OLS model with built-environment variables
  - Model 2: OLS model with built-environment factors

$$Ln(y_{it}) = \sum \alpha_j X_{ijt} + \sum \beta_t D_{it} + \varepsilon_{it}$$

  - Model 3: Spatial lag model with built-environment variables
  - Model 4: Spatial lag model with built-environment factors

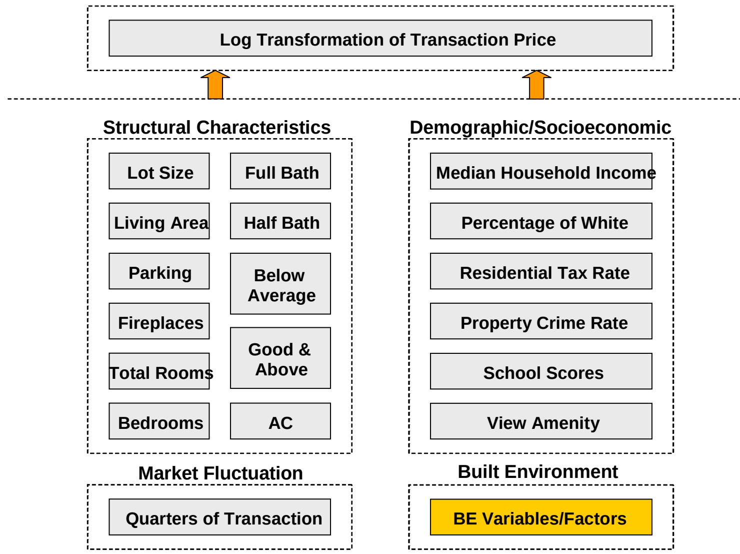
$$Ln(y_{it}) = \rho W_{Ln(y_{it})} + \sum \alpha_j X_{ijt} + \sum \beta_t D_{it} + \varepsilon_{it}$$

  - Model 5: Spatial error model with built-environment variables
  - Model 6: Spatial error model with built-environment factors

$$Ln(y_{it}) = \sum \alpha_j X_{ijt} + \sum \beta_t D_{it} + \varepsilon_{it}$$
$$\varepsilon_{it} = \lambda W_{\varepsilon_{it}} + \mu_{it}$$

# Variables in the Model






# Goodness-of-Fit Measures Comparison



- Spatial error models > spatial lag models > OLS models.
- Models with built-environment variables generally have slightly better fit statistics than corresponding models with built-environment factors, but the results are harder to interpret.

| Measures | Model (1) OLS+BE Var. | Model (2) OLS+BE Fac. | Model (3) Lag+BE Var. | Model (4) Lag+BE Fac. | Model (5) Err.+BE Var. | Model (6) Err.+ BE Fac. |
|---|---|---|---|---|---|---|
| $R^2$ | 0.750 | 0.733 | 0.751 | 0.735 | 0.794 | 0.797 |
| Log Likelihood | 5971.72 | 3008.82 | 6149.59 | 3238.25 | 13665.05 | 12797.12 |
| AIC | -11831.4 | -5949.64 | -12185.20 | -6406.50 | -27218.10 | -25526.20 |
| SC | -11302.9 | -5628.75 | -11647.20 | -6076.17 | -26689.57 | -25205.35 |

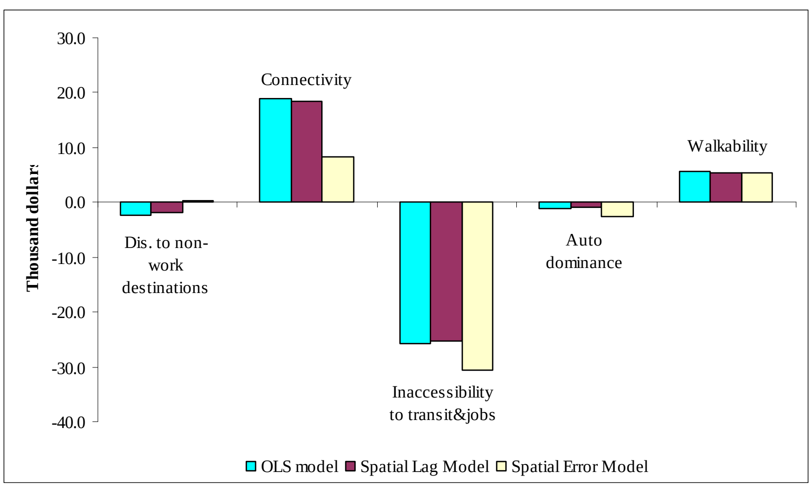
## Estimation Results



| Variables | (1) OLS+BE Var. Coeff. | (1) t | (2) OLS+BE Fac. Coeff. | (2) t | (3) Lag+BE Var. Coeff. | (3) t | (4) Lag+BE Fac. Coeff. | (4) t | (5) Err.+BE Var Coeff. | (5) t | (6) Err.+BE Fac. Coeff. | (6) t |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dis. to church (km) | 0.001 | 1.080 | | | 0.001 | 1.046 | | | 0.001 | 1.088 | | |
| Dis. to dentist (km) | **-0.004** | -7.590 ** | | | **-0.004** | -7.346 ** | | | **-0.004** | -5.304 ** | | |
| Dis. to grocery store (km) | **-0.002** | -1.990 * | | | -0.002 | -1.630 | | | 0.001 | 0.591 | | |
| Dis. to gym (km) | **-0.002** | -4.500 ** | | | **-0.002** | -4.448 ** | | | **-0.002** | -3.258 ** | | |
| Dis. to hardware store (km) | **0.009** | 9.020 ** | | | **0.008** | 8.963 ** | | | **0.009** | 5.840 ** | | |
| Dis. to shopping mall (km) | **0.004** | 5.480 ** | | | **0.004** | 5.813 ** | | | **0.006** | 4.750 ** | | |
| Dis. to restaurant (km) | 0.000 | -0.120 | | | 0.000 | -0.108 | | | 0.001 | 0.866 | | |
| Dis. to school (km) | **0.009** | 7.480 ** | | | **0.009** | 7.713 ** | | | **0.006** | 2.939 ** | | |
| Pct. of roads with access control | **0.074** | 5.490 ** | | | **0.076** | 5.752 ** | | | **0.073** | 3.964 ** | | |
| Pct. of roads with 30mph+ speed limit | 0.014 | 1.300 | | | 0.013 | 1.228 | | | -0.007 | -0.455 | | |
| Average road width (ft) | **-0.001** | -11.190 ** | | | **-0.001** | -11.127 ** | | | **-0.001** | -7.448 ** | | |
| Dis. to highway exit (km) | **-0.001** | -2.840 ** | | | **-0.001** | -2.458 * | | | -0.001 | -1.655 | | |
| Dis. to subway station (km) | **0.001** | 3.480 ** | | | **0.001** | 3.509 ** | | | 0.000 | 0.952 | | |
| Dis. to commuter rail station (km) | **-0.015** | -26.510 ** | | | **-0.015** | -26.335 ** | | | **-0.016** | -16.298 ** | | |
| Dis. to bus stop (km) | 0.000 | -0.330 | | | 0.000 | -0.530 | | | 0.000 | 0.594 | | |
| Dis. to MBTA parking lot (km) | **0.013** | 24.170 ** | | | **0.012** | 23.865 ** | | | **0.013** | 14.631 ** | | |
| Average sidewalk width (ft) | **-0.003** | -3.180 ** | | | **-0.002** | -2.649 ** | | | -0.002 | -1.712 | | |
| Pct. of roads with curbs | **-0.044** | -10.220 ** | | | **-0.045** | -10.453 ** | | | **-0.032** | -4.778 ** | | |
| Pct. of roads with sidewalks | **0.064** | 9.020 ** | | | **0.059** | 8.349 ** | | | **0.052** | 4.934 ** | | |
| Population density (10k/sq. km) | -0.002 | -0.170 | | | -0.002 | -0.230 | | | **-0.029** | -2.124 * | | |
| Land[illegible]ix | **-0.018** | -4.500 ** | | | **-0.016** | -4.059 ** | | | **-0.035** | -6.591 ** | | |
| Road[illegible]ty (km/sq. km) | **-0.003** | -9.050 ** | | | **-0.003** | -9.264 ** | | | **-0.003** | -7.927 ** | | |
| Inter[illegible] density (10/sq. km) | **-0.003** | -2.600 ** | | | **-0.003** | -2.815 ** | | | **-0.005** | -2.776 ** | | |
| Den.[illegible]ay intersections (10/sq.km) | **0.003** | 2.220 * | | | **0.003** | 2.351 * | | | 0.004 | 1.805 | | |
| Den.[illegible]ay intersections (10/sq.km) | **0.007** | 2.920 ** | | | **0.007** | 3.211 ** | | | **0.007** | 2.422 ** | | |
| Pct. [illegible]ay intersections | **-0.055** | -4.140 ** | | | **-0.056** | -4.247 ** | | | **-0.051** | -2.989 ** | | |
| Job [illegible]bility (k) | **0.009** | 66.480 ** | | | **0.009** | 65.823 ** | | | **0.010** | 42.899 ** | | |
| BE[illegible]s. to non-work destinations | | | **-0.008** | -[illegible] ** | | | **-0.007** | -5.647 ** | | | 0.001 | 0.543 |
| BE F[illegible] Connectivity | | | **0.036** | 4[illegible] ** | | | **0.035** | 46.637 ** | | | **0.016** | 12.210 ** |
| BE Fac.3: Inaccesibility to transit&jobs | | | **-0.070** | -77.10[illegible] ** | | | **-0.069** | -75.995 ** | | | **-0.084** | -42.916 ** |
| BE Fac.4: Auto dominance | | | **-0.005** | -3.720 ** | | | **-0.005** | -3.433 ** | | | **-0.012** | -5.470 ** |
| BE Fac.5: Walkability | | | **0.015** | 17.520 ** | | | **0.014** | 16.488 ** | | | **0.014** | 9.147 ** |
| LAMBDA | | | | | | | | | **0.495** | 105.93 ** | **0.637** | 177.428 ** |
| RHO | | | | | **0.011** | 18.802 ** | **0.013** | 21.340 ** | | | | |

## Built-Environment Effect on Property Values






- Change in property values due to one standard deviation increase in built-environment factors

## BE Effects in Sub-markets

35

- Rerun the spatial error model for two sub-markets
  - Properties within walking distance to subway station or bus stop
  - Properties beyond walking distance to subway station or bus stop
- Households living in neighborhoods with good transit accessibility pay higher premiums for smart-growth-type built environment characteristics

| | Observations within 800m of subway station / bus stop | | Observations beyond 800m of subway station / bus stop | |
|---|---|---|---|---|
| | Coeff. | t | Coeff. | t |
| BE Factor 1: Distance to non-work destinations | -0.007 | -0.727 | 0.000 | 0.149 |
| BE Factor 2: Connectivity | 0.017 | 4.529 ** | -0.008 | -4.070 ** |
| BE Factor 3: Inaccessibility to transit and jobs | -0.155 | -10.840 ** | -0.057 | -28.484 ** |
| BE Factor 4: Auto dominance | -0.001 | -0.092 | -0.015 | -6.696 ** |
| BE Factor 5: Walkability | 0.013 | 3.377 ** | 0.002 | 0.882 |
| LAMBDA | 0.824 | 152.357 ** | 0.517 | 106.862 ** |
| No. of observations | 28023 | | 64751 | |
| Pseudo R-squared | 0.833 | | 0.785 | |

** denote coefficient significant at the 0.05 level. ** denote coefficient significant at the 0.01 level.*

## Major Findings of Essay Two

36

- Variation of the built environment can be capitalized into property values
  - Property value is positively associated with accessibility to transit and jobs, connectivity, and walkability, and negatively related to auto dominance

**Smart-growth-type policies tend to be positively associated with property values.**

- The built-environment effects depend on neighborhood characteristics.

**Although smart-growth-type BE features do not have universal appeal, they no doubt satisfy an important market segment.**

## Essay Three: Selectivity, Spatial Autocorrelation, and Valuation of the Built Environment

Housing



## Background



- Essay 2 shows that the variation of built-environment factors can be capitalized into property values.
- Essay 3 further explores issues in tapping the property value effect of the built environment.
- Value Capture Program
  - Capturing the value-added effect of the public projects as a financing mechanism to support infrastructure investment and metropolitan planning.

# Value of Transit Accessibility



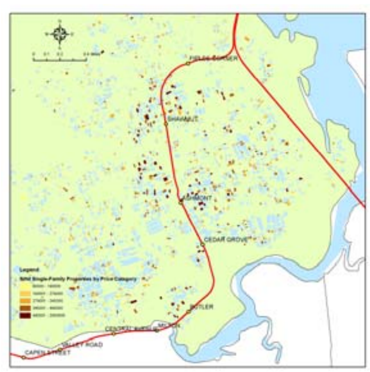
### Property value impacts of public transport proximity in North American cities

| Case/Location | Impact on | Impact | Source |
|---|---|---|---|
| Pennsylvania SEPTA rail | House prices | +3.8% | Voith 1991 |
| Buffalo, NY | House prices | +4-11% | Hess and Almeida 2006 |
| Miami | House prices | +5% | Gatzlaff and Smith 1993 |
| Portland Gresham | Residential Rent | >5% | Hass-Klau, Crampton, et al. 2004 |
| Boston | Residential property | +6.7% | APTA 2002, Armstrong 1994 |
| New Jersey SEPTA rail | House prices | +7.5-8% | Voith 1991 |
| New Jersey PACTO rail | House prices | +10% | Voith 1991 |
| Portland | House prices | +10% | Hass-Klau, Crampton, et al. 2004 |
| San Francisco Bay Area BART | Residential Rent | +10-15% | Cambridge systematics 1998, |
| Portland Metro Express | House prices | +10.5% | Al-Mosaind, Dueker, et al. 1993, Chen, Rufolo et al. 1998 |
| Santa Clara County | Residential Rent | +15% | Weinberger 2001 |
| San Francisco Bay Area BART | Residential Rent | +15-26% | Cervero 1996, Sedway Group 1999 |
| Chicago MTA | House prices | +20% | Gruen 1997 |
| Toronto Metro | House prices | +20% | Bajic 1983, Hack 2002 |
| Dallas DART | Property Values | +25% | Weinstein and Clower 1999, Kay and Haikalis 2000 |
| St. Louis | Property Values | +32% | Garrett 2004 |
| Santa Clara County | House prices | +45% | Cervero and Duncan 2002 |

Source: Modified from Martinez and Viegas (2009)

# Challenges in Valuation of BE





- Shortcomings of previous research
  - Selectivity
    - Use sample to infer population
  - Spatial autocorrelation
    - Violate assumptions of OLS
- My approach
  - Heckman 2-step procedure is applied to correct for sample selection bias
  - Spatial regression techniques are integrated into Heckman selection model to address spatial autocorrelation

## A Model of Housing Sale



Buyer ←  → Seller

$$\ln P_{it}^{b} = \sum \alpha_j^{b} X_{ijt} + \sum \beta_k^{b} Z_{ikt} + \sum \gamma_t D_{it} + \varepsilon_{it}^{b}$$

$$\ln P_{it}^{s} = \sum \alpha_j^{s} X_{ijt} + \sum \beta_k^{s} Z_{ikt} + \sum \gamma_t D_{it} + \varepsilon_{it}^{s}$$

Transaction: $P_{it} =$ Observed $P_{it}^{b} - P_{it}^{s} \geq 0$; Unobserved $P_{it}^{b} - P_{it}^{s} < 0$

$$E[\ln P_{it}] = \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + E[\varepsilon_{it} \mid P_{it}^{b} - P_{it}^{s} \geq 0]$$

- OLS regression using the observed transactions is a biased estimation, if

$$E[(\varepsilon_{it} \mid P_{it}^{b} - P_{it}^{s} \geq 0] <> 0$$

## Heckman Selection Models



- **Choice Model**: A binary probit model on the probability of housing sale in each quarter, calibrated with the full sample
  - $S_{it}^{*} = \ln P_{it}^{b} - \ln P_{it}^{s} = \sum (\alpha_j^{b} - \alpha_j^{s}) X_{ijt} + \sum (\beta_k^{b} - \beta_k^{s}) Z_{ikt} + (\varepsilon_{it}^{b} - \varepsilon_{it}^{s})$
    $= \sum \varpi_j X_{ijt} + \sum \mu_k Z_{ikt} + \nu_{it}$
  - One transaction is observed if S*>=0,
    $$\begin{cases} S_{it} = 1, \text{if } S_{it}^{*} \geq 0 \\ S_{it} = 0, \text{otherwise} \end{cases}$$
  - The choice model can be specified as
    $$\Pr[S_{it} = 1] = \Phi[\sum \varpi_j X_{ijt} + \sum \mu_k Z_{ikt}] \quad (1)$$
  - Inverse Mills Ratio
    $$InvMills_{it} = \phi(\sum \varpi_j X_{ijt} + \sum \mu_k Z_{ikt}) / \Phi(\sum \varpi_j X_{ijt} + \sum \mu_k Z_{ikt})$$
- **Price Model:** a hedonic price model estimated with the sold sample
  $$\ln P_{it} = \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + \eta_{it} InvMills_{it} + \varepsilon_{it} \quad (2)$$

## Spatial Regression



- Incorporate spatial regression techniques into the price model to address spatial autocorrelation

**Choice Model:**

$$\Pr[S_{it} = 1] = \Phi[\sum \varpi_j X_{ijt} + \sum \mu_k Z_{ikt}]$$

**Price Model:**

$$\ln P_{it} = \rho W_{\ln P} + \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + \eta_{it} InvMills_{it} + \varepsilon_{it}$$

**Or**

$$\ln P_{it} = \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + \eta_{it} InvMills_{it} + \varepsilon_{it}$$

$$\varepsilon_{it} = \lambda W_{\varepsilon} + \mu_{it}$$

## Major Datasets



- Housing Data

Housing Transactions

**Transaction Records from Suffolk County Registry of Deeds (1998-2007)**

- Date of sale
- Sale price
- Street address

Housing Stock

**Assessing Records from Assessing Department of Boston**

- Structural Characteristics
- Tax information
- Street address

Final dataset (1,198,031 observations)

- Every parcel of single-family houses multiplied by the number of years the house was included in the assessing data
  - Single-family housing transactions (10031)

- Built-Environment Data from MassGIS

# Variables in the Choice Model



Probability of Sale

**Structural Characteristics**

- Lot Size
- Full Bath
- Living Area
- Half Bath
- Year of Built
- A/C
- Fireplaces

**Built Environment**

- BE Variables

**Neighborhood Dummies**

- Neighborhood of Property

**Quarter Dummies**

- Q1, Q2, Q3

**Macroeconomic Variables**

- National GNP
- National Mortgage Rate
- Local Unemployment Rate

# Variables in the Price Model



Log Transformation of Transaction Price

**Structural Characteristics**

- Lot Size
- Full Bath
- Living Area
- Half Bath
- Year of Built
- AC
- Fireplaces

**Socioeconomic Variables**

- Median Household Income
- Percentage of White

**Time Dummies**

- Quarter of Transaction

**Built Environment**

- BE Variables

## Built-Environment Variables



- **Density**
  - Residential density — Population/Residential area
- **Land Use Mix**
  - Land use mix entropy — $-\sum_j P_j * \ln(P_j)/\ln(J)$
- **Street Network Layout**
  - Intersection density
- **Accessibility**
  - Job Accessibility — $A_i = \sum_j O_j f(C_{ij})$ $f(C_{ij}) = \exp(-\beta * C_{ij})$
  - Proximity to non-work destinations
  - Distance to CBD
  - Distance to highway exits
  - Presence of subway station within half mile
  - Distance to MBTA parking lot
  - Presence of commuter rail stations within half mile

## Estimation Results -- Choice Model (Probability of Sale)



| Variables | Coef. | t-Stat. |
|---|---|---|
| ***Structural Variables*** | | |
| ln(lot size) | -0.0598 | -5.75 *** |
| ln(gross area) | -0.1110 | -5.96 *** |
| Year built | -0.0002 | -3.54 *** |
| Number of floors | 0.0254 | 2.69 *** |
| Total number of rooms | -0.0065 | -2.17 ** |
| Number of fullbath | 0.0866 | 11.01 *** |
| Number of halfbath | 0.0253 | 3.46 *** |
| Presense of A/C | 0.1206 | 10.08 *** |
| Number of fireplaces | 0.0157 | 2.86 *** |
| ***Macroeconomic Variables*** | | |
| GNP | 0.0074 | 2.77 *** |
| Mortgage rate | -0.0683 | -7.04 *** |
| Unemployment rate | -0.0122 | -1.90 * |
| ***Built-Environment Variables*** | | |
| Population denisty (k/km$^2$) | 0.0047 | 2.11 ** |
| Land use mix | 0.0022 | 0.10 |
| Presense of subway station within half mile | 0.0163 | 1.43 |
| Distance to highway exits (km) | -0.0018 | -0.30 |
| Presense of commuter rail station within half mile | 0.0082 | 0.86 |
| Distance to MBTA parking lots (km) | -0.0025 | -0.32 |
| Distance to CBD (km) | 0.0201 | 2.23 ** |
| Job accessibility (k) | 0.0011 | 3.28 *** |
| Distance to non-work destinations (km) | -0.0734 | -2.56 *** |
| Intersection denisty (1/km$^2$) | -0.0003 | -1.58 |
| Observations | 1198031 | |
| LR chi-squre(40) | 1174.7600 (p=0.000) | |

## Comparison of BE Coefficients



- Coefficients of the inverse mills ratio are significantly different from 0
  - ➡ The existence of sample selection bias
- Both the spatial lag term and spatial error term are significant,
  - ➡ The existence of spatial autocorrelation

| | Hedonic price model | | Heckman selection model | | Heckman selection model with spatial lag | | Heckman selection model with spatial error | |
|---|---|---|---|---|---|---|---|---|
| | **Coef.** | **t-Stat.** | **Coef.** | **t-Stat.** | **Coef.** | **t-Stat.** | **Coef.** | **t-Stat.** |
| Population denisty (k/km$^2$) | 0.0180 | 11.61 *** | 0.0237 | 15.68 *** | 0.0145 | 9.89 *** | 0.0049 | 1.60 |
| Land use mix | 0.0179 | 1.09 | 0.0063 | 0.40 | 0.0069 | 0.46 | -0.0074 | -0.27 |
| Presence of subway sta. within half mile | 0.0570 | 6.48 *** | 0.0983 | 11.40 *** | 0.0539 | 6.47 *** | 0.0303 | 1.96 ** |
| Distance to highway exits (km) | -0.0095 | -3.47 *** | 0.0168 | 5.97 *** | 0.0115 | 4.28 *** | 0.0096 | 0.66 |
| Presence of commuter rail sta. within half mile | 0.0070 | 0.98 | 0.0136 | 1.98 ** | 0.0111 | 1.70 * | -0.0128 | -1.01 |
| Distance to MBTA parking lots (km) | -0.0838 | -19.97 *** | -0.0707 | -17.33 *** | -0.0384 | -9.39 *** | -0.0639 | -3.38 *** |
| Distance to CBD (km) | 0.0650 | 19.97 *** | 0.0654 | 20.84 *** | 0.0303 | 9.42 *** | 0.0086 | 0.68 |
| Job accessibility (k)[a] | 0.3570 | 27.19 *** | 0.3580 | 28.30 *** | 0.1872 | 13.78 *** | 0.1651 | 4.31 *** |
| Distance to non-work destinations (km) | 0.1276 | 6.08 *** | -0.0387 | -1.83 * | 0.0039 | 0.20 | -0.0862 | -1.71 * |
| Intersection denisty (1/km$^2$)[a] | -0.0158 | -1.41 | -0.0521 | -4.77 *** | -0.0179 | -1.71 * | 0.0046 | 0.21 |
| Inverse mills ratio | | | 1.9482 | 27.23 *** | 1.1316 | 15.09 *** | 1.0801 | 4.83 *** |
| Lambda | | | | | | | 0.8792 | 78.76 *** |
| W_ln_pirce | | | | | 0.3705 | 28.22 *** | | |
| R-square | 0.7541 | | 0.7711 | | 0.7913 | | 0.8091 | |

*, ** and *** denote significant at the 0.1, 0.05, and 0.01 level respectively. a Coefficient is x 10-2.

## Importance of Correction to WTP and Elasticity

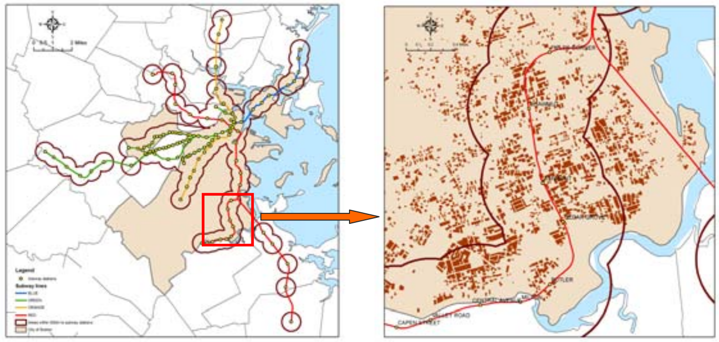


| | Hedonic price model | | Heckman selection model | | Heckman selection model with spatial lag | | Heckman selection model with spatial error | |
|---|---|---|---|---|---|---|---|---|
| | WTP(k$) | Elast. | WTP(k$) | Elast. | WTP(k$) | Elast. | WTP(k$) | Elast. |
| Population density (k/km$^2$) | 6.858 | 0.104 | 9.056 | 0.137 | 5.505 | 0.084 | 1.864 | 0.028 |
| Land use mix | 6.810 | 0.008 | 2.390 | 0.003 | 2.625 | 0.003 | -2.768 | -0.003 |
| Intersection density (1/km$^2$) | -0.060 | -0.018 | -0.197 | -0.061 | -0.068 | -0.021 | 0.017 | 0.005 |
| **Presence of subway sta. within half mile** | **22.163** | | **38.994** | | **20.905** | | **11.603** | |
| Distance to highway exits (km) | -3.577 | -0.031 | 6.406 | 0.055 | 4.376 | 0.038 | 3.633 | 0.031 |
| Presence of commuter rail sta. within half mile | 2.637 | | 5.161 | | 4.210 | | -4.806 | |
| **Distance to MBTA parking lots (km)** | **-30.346** | **-0.144** | **-25.757** | **-0.121** | **-14.214** | **-0.066** | **-23.388** | **-0.110** |
| Distance to CBD (km) | 25.344 | 0.524 | 25.523 | 0.528 | 11.626 | 0.245 | 3.263 | 0.069 |
| **Job accessibility (k)** | **1.350** | **1.647** | **1.354** | **1.652** | **0.708** | **0.864** | **0.624** | **0.762** |
| **Distance to non-work destinations (km)** | **51.391** | **0.128** | **-14.334** | **-0.039** | **1.492** | **0.004** | **-31.202** | **-0.087** |

- Willingness-to-pay: the change in property value due to one unit change in a built-environment variable
- Price elasticity: the percentage change of property value due to one percent increase of a built-environment variable
- The WTPs and price elasticities are computed for a property with an original price of $377.6k (the mean sale price of the sold sample)

## Value-Added Effect of Subway Stations in City of Boston





| | | | | | Hedonic Price Model | | Heckman Selection Model | | Heckman Selection Model + Spatial Lag | | Heckman Selection Model + Spatial Err. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Property Type | Total Value (Boston) | Total Property Tax (Boston) | Total Value within Buffer | Property Tax within Buffer | Value added of Subway Station | Property Tax Attr. to Subway | Value added of Subway Station | Property Tax Attr. to Subway | Value added of Subway Station | Property Tax Attr. to Subway | Value added of Subway Station | Property Tax Attr. to Subway |
| 1-Family | 10472.4 | 112.4 | 3574.8 | 38.4 | 209.8 | 2.3 | 369.2 | 4.0 | 197.9 | 2.2 | 109.9 | 1.2 |
| 2-Family* | 7092.3 | 76.1 | 2918.8 | 31.3 | 171.3 | 1.9 | 301.4 | 3.3 | 161.6 | 1.8 | 89.7 | 1.0 |
| 3-Family* | 6440.4 | 69.1 | 3584.1 | 38.5 | 210.4 | 2.3 | 370.1 | 4.1 | 198.4 | 2.2 | 110.1 | 1.2 |
| Condo.* | 15113.3 | 162.2 | 12502.4 | 134.2 | 733.8 | 8.0 | 1291.1 | 14.2 | 692.2 | 7.6 | 384.2 | 4.2 |
| Total | 39118.4 | 419.7 | 22580.1 | 242.3 | 1325.3 | 14.5 | 2331.8 | 25.6 | 1250.1 | 13.7 | 693.9 | 7.6 |

* Hypothetic values assuming the same coefficients as the 1-family housing. All numbers are in Million dollars.

## Major Findings of Essay Three



- The built environment has significant impacts on both the probability of sales and transaction prices of properties.
- It is important to account for sample selection bias and spatial autocorrelation when estimating the willingness-to-pay (WTP) for built-environment attributes.
  - The bias is 91%, if we compare the WTP for proximity to subway station estimated with Heckman selection model with spatial error and conventional hedonic price model.

## Summary: Implications for Growth Management



- **Essay 1**: Could the variation in vehicle miles travelled (VMT) be explained by the variation in the built environment?
  - Understand the environmental implications of alternative metropolitan growth scenarios
  - Evaluate the effectiveness of smart growth strategies
- **Essay 2**: Could the variation in the built environment be capitalized into property values?
  - Understand the local impact of smart growths strategies
- **Essay 3**: How can we assess the value-added effect of the built environment?
  - Design value-capture programs to support infrastructure investment and metropolitan planning

## Summary: Implications for Urban Modeling



- Administrative data with spatial information
  - GIS data layers: road networks, parcels, building footprints, etc
  - Transaction information: housing transactions, vehicle safety inspections, transit fare cards, utility records, cell phone use, etc.
- Calibrating urban models using administrative data
  - Pros
    - Low marginal cost
    - Broad temporal and spatial coverage
    - accuracy, automatic collection and central storage
  - Cons
    - Not primarily designed for modeling
    - Hard to cross-reference
    - Potential sample errors

**Administrative data are not substitute for survey data in urban modeling, but can reduce the dependence on surveys and complement their usage in metropolitan planning.**

## Questions and Comments



Q&A

Thank You!

diaomi@mit.edu