How to Price a House

An Interpretable Bayesian Approach

Dustin Lennon

dustin@inferentialist.com

Inferentialist Consulting
Seattle, WA

April 9, 2014

Introduction

- Project to tie up loose ends / came out of interview prep for Climate Corp
- Disclaimer: two week sprint, not a dissertation
- An easier version of a more involved spatio-temporal model for zipcode aggregation

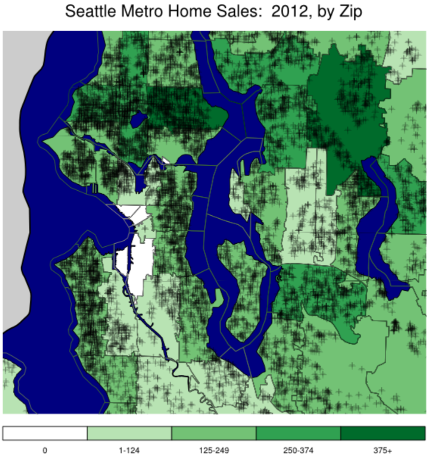
Outline

Housing Market

A few Wikipedia Facts

- Outstanding U.S. residential mortgages: $10.6 trillion as of midyear 2008
- By August 2008, 9.2% of all U.S. mortgages outstanding were either delinquent or in foreclosure

Housing Market

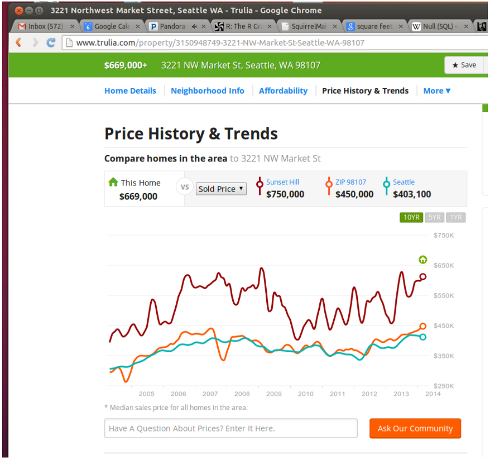


A Valuation Problem?

- Subprime loans, yes, but was there also a systemic failure in estimating home values?

Temporal Instability

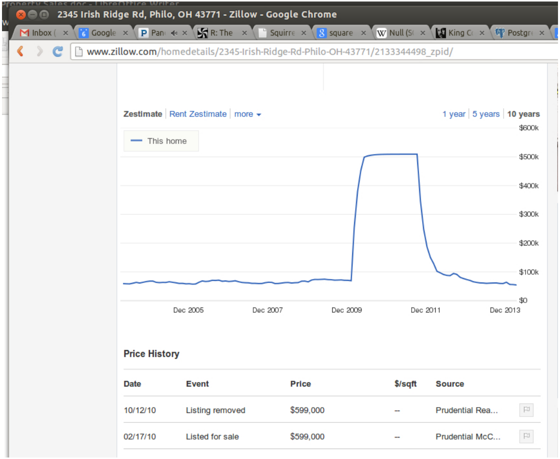
Trulia



Seasonality, perhaps. But a sliding median approach breaks down as the window size goes to zero.

page accessed on 6/4/2014

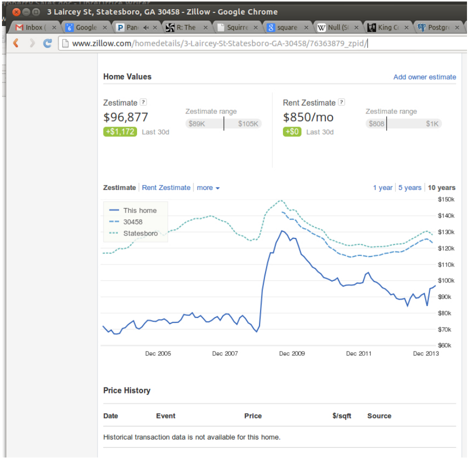
Overfitting
Zestimates



The time series appears to “chase” the listing data, stays elevated for a time, then abruptly returns to baseline.

page accessed on 6/4/2014

Spatial Instability
Zestimates



The time series appears to adjust to the recently added zipcode level information, perhaps indicating some spatial instability when adjusting to new data.

page accessed on 6/4/2014

Ad-hoc Analysis

- Limiting case failures
- Lack of regularization / prior information
- Uninterpretable models

Outline

Hedonic Model I

- Decompose home value into constituent parts

$$Z_i = x_i^t \beta + a_i Y(s_i) + \delta_i,$$

Z_i	price paid for the i^{th} home
x_i	covariates associated with β [e.g., square footage]
β	coefficients fixed across space [e.g., build cost per square foot]
a_i	lot size
$Y(s)$	unit cost of land
s_i	location
δ_i	difference between the “true” value and the price paid

Hedonic Model II

Data Model

$$[Z|\beta, Y] \sim N\left(\begin{bmatrix} X & A \end{bmatrix} \begin{bmatrix} \beta \\ Y \end{bmatrix}, \Delta\right)$$

$$\Delta = \text{diag}\left([\sigma^2 z_1^2, \ldots, \sigma^2 z_n^2]\right)$$

Process Model

$$[\beta, Y] = [\beta][Y]$$

$$[\beta] \sim N(\nu, \Phi)$$

$$\Phi = \text{diag}\left([\phi_1, \ldots, \phi_k]\right)$$

$$[Y] \sim N(\tau \mathbf{1}, \Sigma)$$

$$\Sigma = \Sigma(\theta)$$

Hedonic Model III

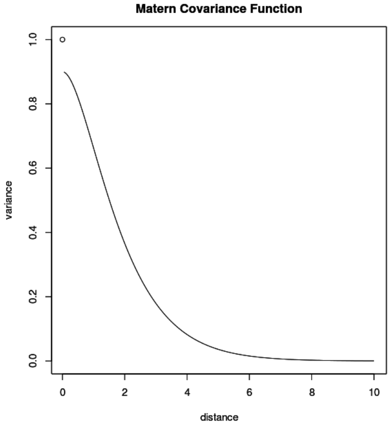
σ^2	interpretable as coefficient of variation
$\Sigma(\theta)$	defines the covariance structure of the land value term

In particular, $\Sigma(\theta)$ is specified through an isotropic, Matern covariance function:

$$\begin{aligned}\Sigma_{ij}(\theta) &\equiv C\left(d_{ij}; \theta_1, \theta_2, \sigma_0^2, \sigma_1^2\right) \\ &= \sigma_0^2 I_0\left(d_{ij}\right) + \sigma_1^2\left(2^{\theta_2 - 1}\Gamma(\theta_2)\right)^{-1}\left(\frac{d_{ij}}{\theta_1}\right)^{\theta_2} K_{\theta_2}\left(\frac{d_{ij}}{\theta_1}\right)\end{aligned}$$

and d_{ij} is the Euclidean distance between s_i and s_j.

Hedonic Model III



General Model Formulation

Hierarchical Formulation

$$[Z|G] \sim N(MG, \Delta)$$
$$[G] \sim N(\mu, \Omega)$$

Joint Distribution

$$[Z, G] \sim N\left\{ \begin{pmatrix} M\mu \\ \mu \end{pmatrix}, \begin{bmatrix} \Delta + M\Omega M^t & M\Omega \\ \Omega M^t & \Omega \end{bmatrix} \right\}$$

Posterior Distribution

$$[G|Z] \sim N\left(\breve{\mu}, \breve{\Omega}\right)$$
$$\breve{\mu} \equiv \mu + \Omega M^t \left(\Delta + M\Omega M^t\right)^{-1} (Z - M\mu)$$
$$\breve{\Omega} \equiv \Omega - \Omega M^t \left(\Delta + M\Omega M^t\right)^{-1} M\Omega$$

Fitting the Model

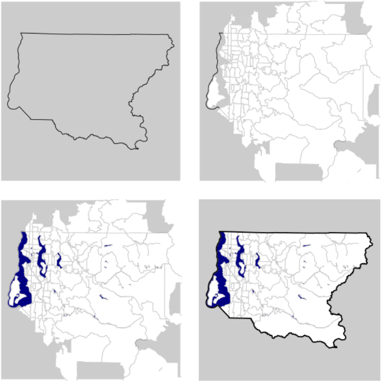
- Inference is on posterior distribution $[G|Z;\Theta]$
- Specialize general case to hedonic model
- EM Algorithm to obtain $\hat{\Theta}$. Iterate until convergence:
  - update $\breve{\mu}$, $\breve{\Omega}$
  - minimize $-2\mathbb{E}\left[\log\left[Z,G\right]|Z;\Theta\right]$

$$
\begin{aligned}
-2\mathbb{E}\left[\log[Z,G]|Z;\Theta\right] = \; & \operatorname{logdet}\Delta + \operatorname{logdet}\Omega + Z^t\Delta^{-1}Z + \mu^t\Omega^{-1}\mu \\
& - 2\left[Z^t\Delta^{-1}M + \mu^t\Omega^{-1}\right]\breve{\mu} \\
& + \breve{\mu}^t\left[M^t\Delta^{-1}M + \Omega^{-1}\right]\breve{\mu} \\
& + \operatorname{tr}\left[\left(M^t\Delta^{-1}M + \Omega^{-1}\right)\breve{\Omega}\right]
\end{aligned}
$$

Outline

Data: Maps

TIGER/Line Shapefile Data



Data: Home Sales

King County Department of Assessments

- Table Joins:
  - Real Property Sales (non-flagged 2012 records)
    - Exempt From Excite Tax
    - Related Party, Friend, or Neighbor
    - Quit Claim Deed
    - Multi-Parcel Sale
  - Residential Buildings
  - Parcel Information
- Outlier Filtering:
  - Sale Price: $100k to $5m
  - Lot Size $\leq$ 1.03 acres
  - No properties with multiple sale records in 2012
- 11,812 homes

Data: Geocoding
Yahoo

- 2012: KC records have UID, street address, no lat/long
- 2014: Sporadic lat/long (Seattle, not Tacoma)
- Yahoo geocoder: bash script, 500k lookups over two weeks

```
curl -s "http://where.yahooapis.com/geocode?&q=${addr},+${zip}&flags=C&appid=..."
```

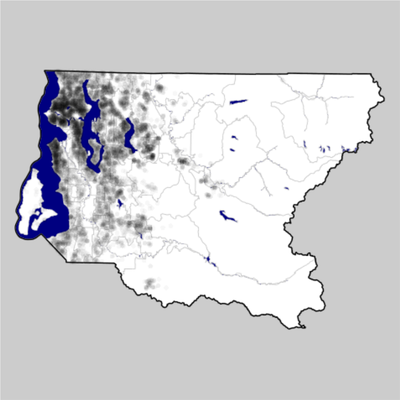
Scalability and Sampling I

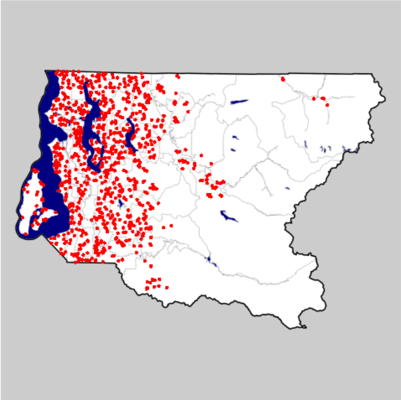
Recall the objective function to be optimized on each iteration of EM algorithm:

$$
\begin{aligned}
-2\mathbb{E}\left[\log[Z,G]|Z;\Theta\right] = \operatorname{logdet}\Delta + \operatorname{logdet}\Omega + Z^t\Delta^{-1}Z + \mu^t\Omega^{-1}\mu \\
- 2\left[Z^t\Delta^{-1}M + \mu^t\Omega^{-1}\right]\breve{\mu} \\
+ \breve{\mu}^t\left[M^t\Delta^{-1}M + \Omega^{-1}\right]\breve{\mu} \\
+ \operatorname{tr}\left[\left(M^t\Delta^{-1}M + \Omega^{-1}\right)\breve{\Omega}\right]
\end{aligned}
$$

- Naive approach with dense matrices:
  - extremely memory intensive
  - $O(n^3)$ cost to compute inverse
- Solution: sample, weighted by inverse local density

Scalability and Sampling II





Model Output

Coefficients

σ^2	0.22^2	coefficient of variation [active constraint]
ν_1, ϕ_1	$(139.51, 57.4^2)$	build cost per square foot (living)
ν_2, ϕ_2	$(0.00, 35.9^2)$	build cost per square foot (basement)
ν_3, ϕ_3	$(0.00, 14.6^2)$	build cost per square foot (garage)
τ	7.19	lot size cost per square foot
θ_1	2000	matern “spread” parameter [active constraint]
θ_2	3.00	matern “shape” parameter [active constraint]
σ_0^2	0.1	matern “nugget” effect [active constraint]
σ_1^2	73.00	matern “variance”

Model Output

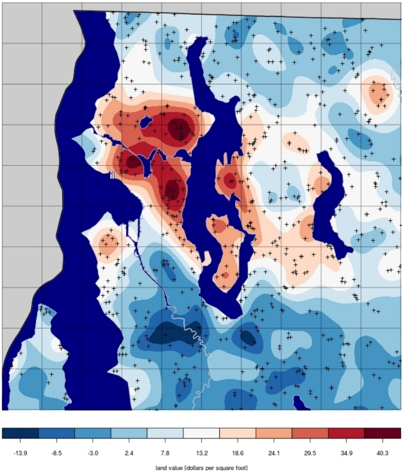
Heatmaps

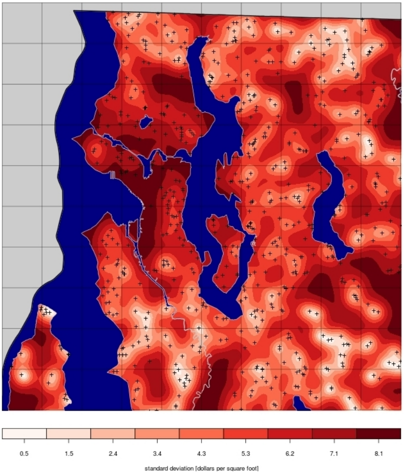
- Need predictive distribution $[y_0|Z]$:

$$\begin{aligned}\mathbb{E}[y_0|Z] &= \mathbb{E}[\mathbb{E}(y_0|Y,Z)|Z] \\ &= \mathbb{E}[\mathbb{E}(y_0|Y)|Z]\end{aligned}$$

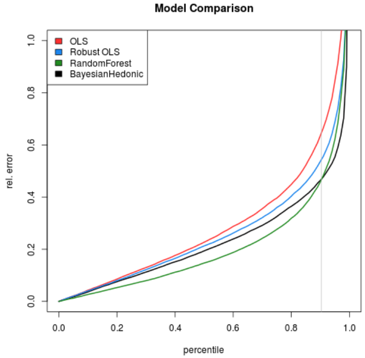
$$\begin{aligned}\mathrm{Var}[y_0|Z] &= \mathrm{Var}[\mathbb{E}(y_0|Y,Z)|Z] + \mathbb{E}[\mathrm{Var}(y_0|Y,Z)|Z] \\ &= \mathrm{Var}[\mathbb{E}(y_0|Y)|Z] + \mathbb{E}[\mathrm{Var}(y_0|Y)|Z]\end{aligned}$$

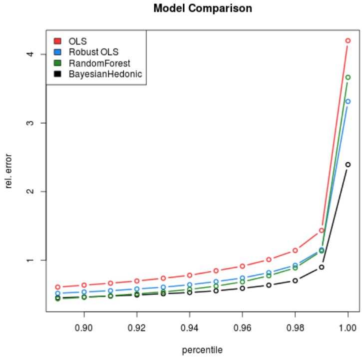
- $[y_0|Y]$ is immediate: extend $\Sigma(\theta)$


-13.9
-8.5
-3.0
2.4
7.8
13.2
18.6
24.1
29.5
34.9
40.3
land value [dollars per square foot]


0.5
1.5
2.4
3.4
4.3
5.3
6.2
7.1
8.1
standard deviation [dollars per square foot]

Model Comparison





Model Validation

- Not a predictive model; attempts to characterize variation
- Out of sample coverage of 95% confidence intervals:

Process	86.7%
Process + Proxy	92.0%
Process + Data	97.2%

- Conclusion: the typical variability in a home's sale price is inherently large

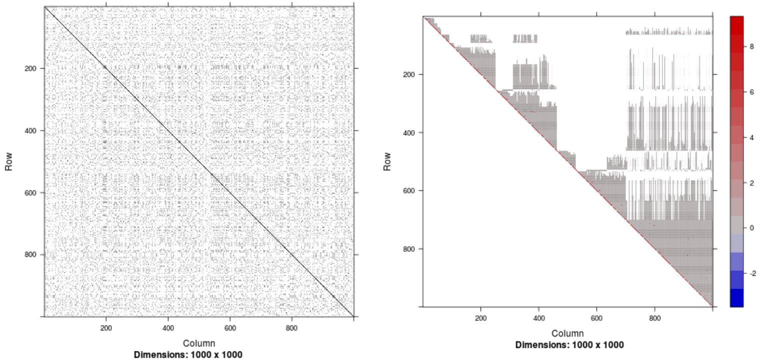
Outline

Scalability I

- Goal: linear algebra operations to evaluate objective function, gradient should be:
  - sparse matrices
  - low rank perturbations to sparse matrices
  - arbitrarily close to sparse matrices under reasonable parameter choices
- Larger sample sizes require sparse representation
- Specializing the general model:
  - M is sparse; Ω decomposes into a diagonal and the Matern matrix, $\Sigma(\theta)$.
  - For θ_1 small and θ_2 bounded, $\Sigma(\theta)$ is arbitrarily close to a sparse matrix
  - For θ_1 and θ_2 bounded, $\Sigma(\theta)$ is well conditioned; relative to underlying Euclidean distances

Scalability I

For $\theta_1 = 500$:



Scalability II

More on θ_1

- $\hat{\theta}_1$ is an active constraint, at the upper bound
- reflects a “desire” to increase spatial scale of correlation; smoother surface
- Conclusion: the upper bound enforced on θ_1 should be interpreted as a model complexity parameter
  - keeping θ_1 small increases sparsity of $\Sigma(\theta)$ and decreases scale of spatial correlation effect
  - choose upper bound via cross validation

Inner Optimization

- EM algorithm requires an inner optimization
- Dynamically adjust the convergence tolerance (optim/factr) in early iterations for speed

Outline

Summary

- The Hedonic Bayesian model needs very few parameters to describe a complex spatial field.
- The model does a good job describing the variability inherent in the data.

- Future Work
  - Experimentation with smaller σ^2; cross validation of θ_1 upper bound
  - Increase scalability through a more thorough approach to sparsity