Real Estate Price Indices

New Methods, New Markets, New Uses

Marc Francke, David Geltner and Alex Van de Minne
m.k.francke@uva.nl dgeltner@mit.edu
avdminne@mit.edu

Massachusetts Institute of Technology
Center for Real Estate

University of Amsterdam
Amsterdam Business School

May, 2016

2016 Hoyt

Outline

The Repeat Sales Model

- The existing literature on real estate index construction has been developed initially around the hedonic (Rosen, 1974) and the repeated sale methodology (Bailey et al., 1963).
- Let X (Z) be (un-)observed variables and μ_t is the cumulative (log) price index. With the RS you can replace the (un-)observed characteristics with a dummy per property, denoted δ_i

$$\log P_{it} = p_{it} = \mu_t + \underbrace{X_{it}\beta + Z_{it}\alpha}_{\delta_i} + \epsilon_{it} \tag{1}$$

- The main advantages of the RS over the Hedonic models are (therefore) as follows;
  - The RS is less affected by specification errors.
  - The RS is less affected by missing characteristics in the Data

The Repeat Sales Model, *Continued*

- The main disadvantages of the RS model, over the hedonic, is (therefore) as follows:
  - The assumption is that characteristics do not change over time. The question is if this matters?
  - The coefficient for δ_i can only be identified if you have *at least* two observations for every δ, i.e. repeat sales, thus data is lost.
  - This can also allow for selection bias in the data.
- Estimating every single δ_i can seem cumbersome, however this is easily circumvented by estimating the model in 'differences'.

$$p_{it} - p_{is} = \mu_{it} - \mu_{is} + \underbrace{\delta_{it} - \delta_{is}}_{=0} + \epsilon_{it} - \epsilon_{is} \tag{2}$$

$$p_{it} - p_{is} = \mu_{it} - \mu_{is} + \epsilon_{it} - \epsilon_{is} \tag{3}$$

- Where *s* is the time of buy, compared to time of sell *t*.

Application of RS indices

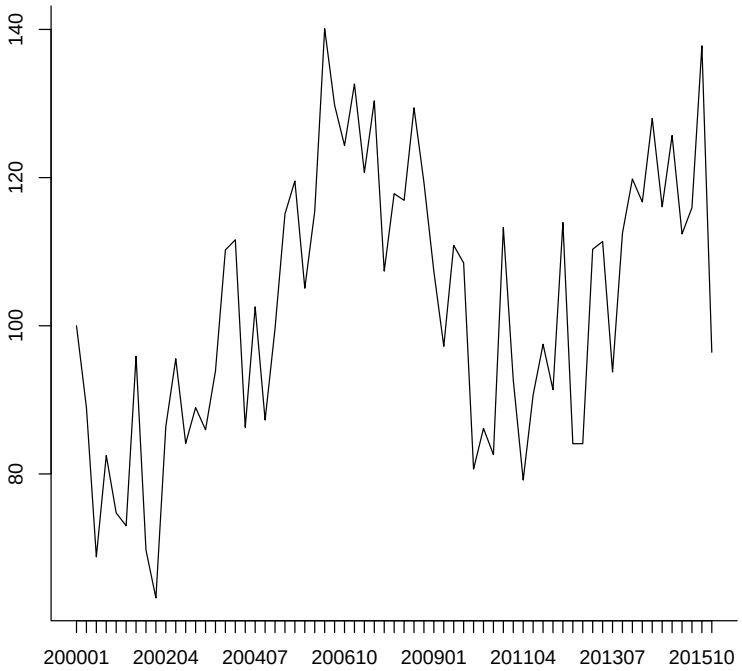


Figure: Commercial in Philly/Baltimore

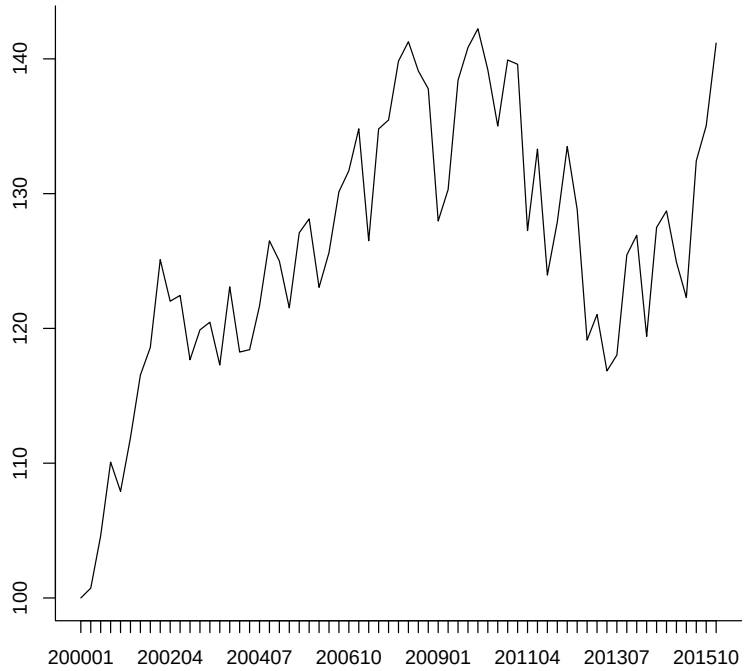


Figure: Housing in Amsterdam West

Outline

Even more granular indices

- Nice to have indices on an aggregate level, however we are interested in more granular indices.
- In the West of Amsterdam the following breakdown would be interesting:

Table: Number of pairs in Amsterdam West

Zip code	Single-family	Apartment	Total
1063	61	95	156
1064	75	137	212
1065	27	37	64
1066	148	124	272
1068	5	206	211
Total	316	599	915

Even more granular indices, *continued*

- For the Philadelphia and Baltimore region we could use the following breakdown;

Table: Number of pairs Philly/Baltimore

Metro	Multi-family	Industrial	Office	Retail	Total
Baltimore	99	111	60	44	314
Philadelphia	77	88	122	45	332
Rest	37	44	42	26	149
Total	213	243	224	115	795

- Estimating indices with this little observations is impossible . . .

Application of HRS indices

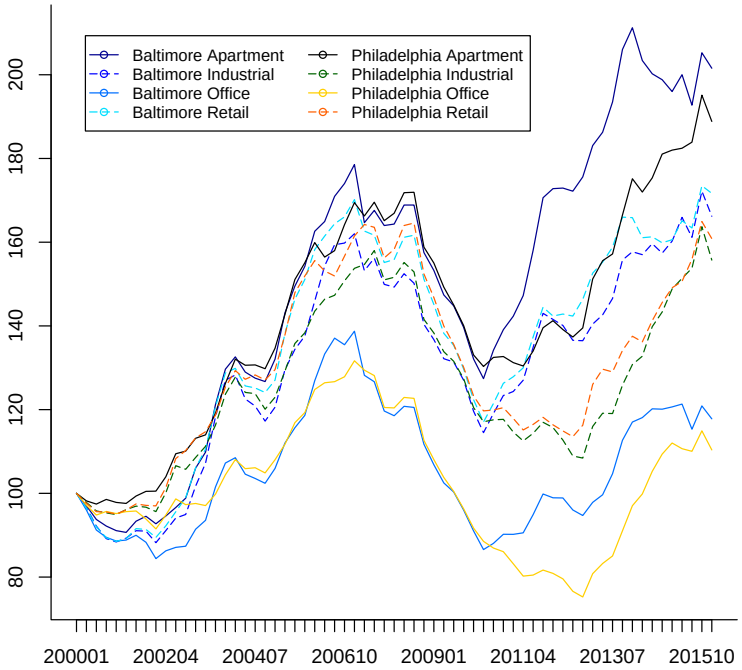


Figure: Commercial in Philly/Baltimore

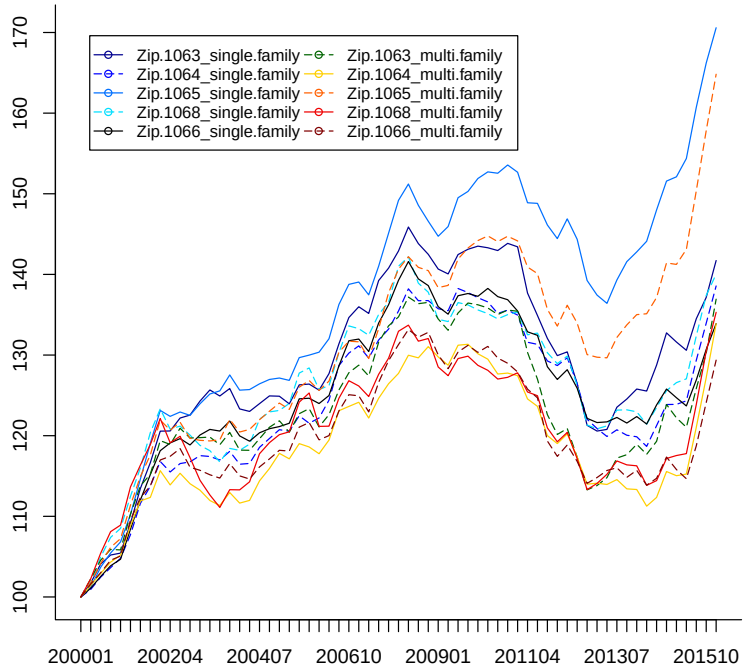
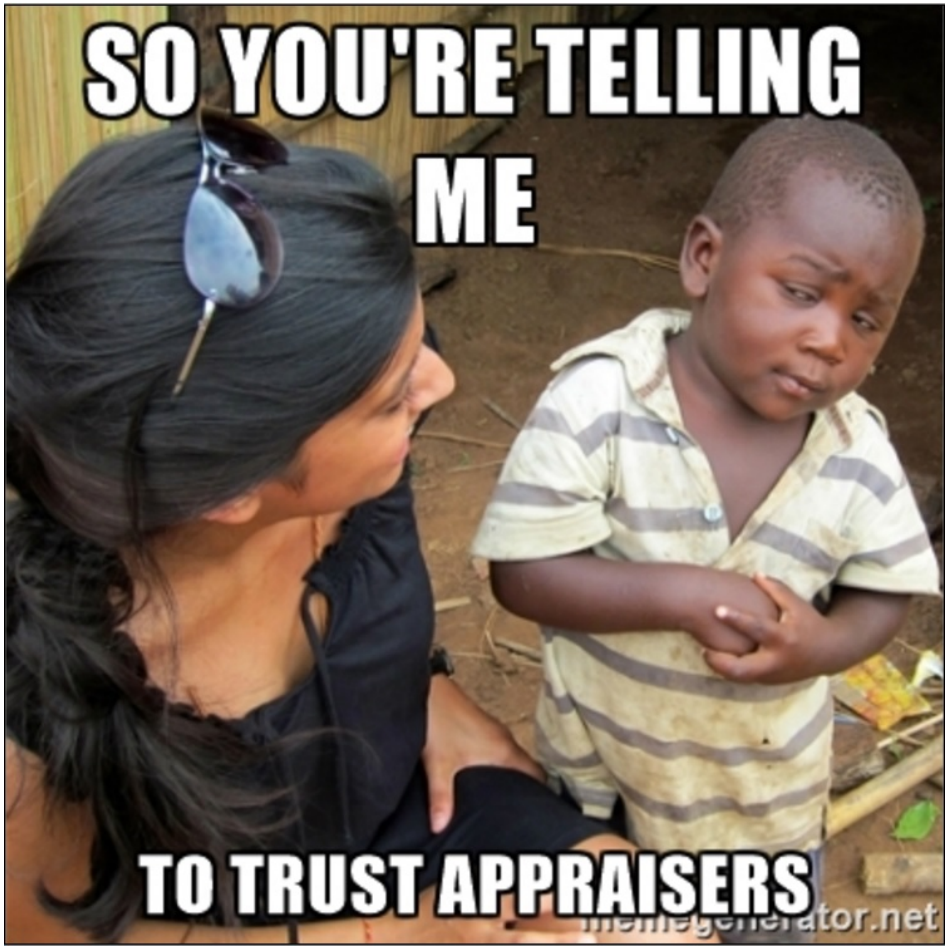


Figure: Housing in Amsterdam West

Bayesian Inference

- Why is Bayes Theorem so important for people working in *science*:
  - Test are not the event.
  - Test are flawed.
  - Test give us test-probabilities, not the real probabilities.
  - False positives skew results.
- Bayes‘ theorem converts the results from your test into the real probability of the event, using prior knowledge of the real probabilities.
- Ergo, we cannot trust the data *per se* (sorry Bob)! The way data came out the way it is, is also uncertain.
- Another analogy is that we should accept we are in *social* science and not *exact* science. Your opinion matters!

Bayesian Inference



Model I: The Hierarchical Repeat Sales Model

- . . . don't worry, we have machine learning!
- Here we introduce the 'standard' Hierarchical Repeat Sales Model;

$$p_{it} - p_{is} = \mu_t - \mu_s + \sum_{j=1}^{k} d_i^j(\lambda_t^j - \lambda_s^j) + \varepsilon_{it} - \varepsilon_{is}, \quad \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2),$$

$$\begin{aligned}
\mu_{t+1} &= \mu_t + \kappa_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2),\\
\kappa_{t+1} &= \kappa_t + \vartheta_t, & \vartheta_t &\sim N(0, \sigma_\zeta^2),\\
\lambda_{t+1}^j &= \lambda_t^j + \varsigma_t^j, & \varsigma_t^j &\sim N(0, \sigma_{\varsigma^j}^2 I_{n_j}).
\end{aligned}$$

- Where μ_t is the common trend and λ are cluster trends, in this case for location and property type.

Model II: Stochastic Volatility and Time-Varying Noise

- In this model we used a different structure on the time-series.

$$p_{it} - p_{is} = \mu_t - \mu_s + \epsilon_{it} - \epsilon_{is}, \qquad \epsilon_{it} \sim T(0, \sigma^2_{t,\epsilon}, \nu), \tag{4}$$

$$\mu_{t+1} = \mu_t + \rho(\mu_t - \mu_{t-1}) + \eta_t, \qquad \eta_t \sim N\left(0, \frac{\sigma^2_{t,\eta}}{1-\rho^2}\right), \tag{5}$$

$$h^k_{t+1} = h^k_t + \omega^k_t, \qquad \omega^k_t \sim N(0, \sigma^2_{\omega^k}). \tag{6}$$

- We use an AR component in the state vector and make both signal and noise time-varying.
  - When tail risk is high, investors will be willing to pay a higher price for a tail risk hedging asset.
  - Noise is time-varying by construct in a RS framework.
  - During market crises the shortage of arbitrage capital leaves the yields to move more freely relative to the yield-curve.
  - Less comparable assets, can result in prices ‘moving together’.

Outline

Information in Thin Markets

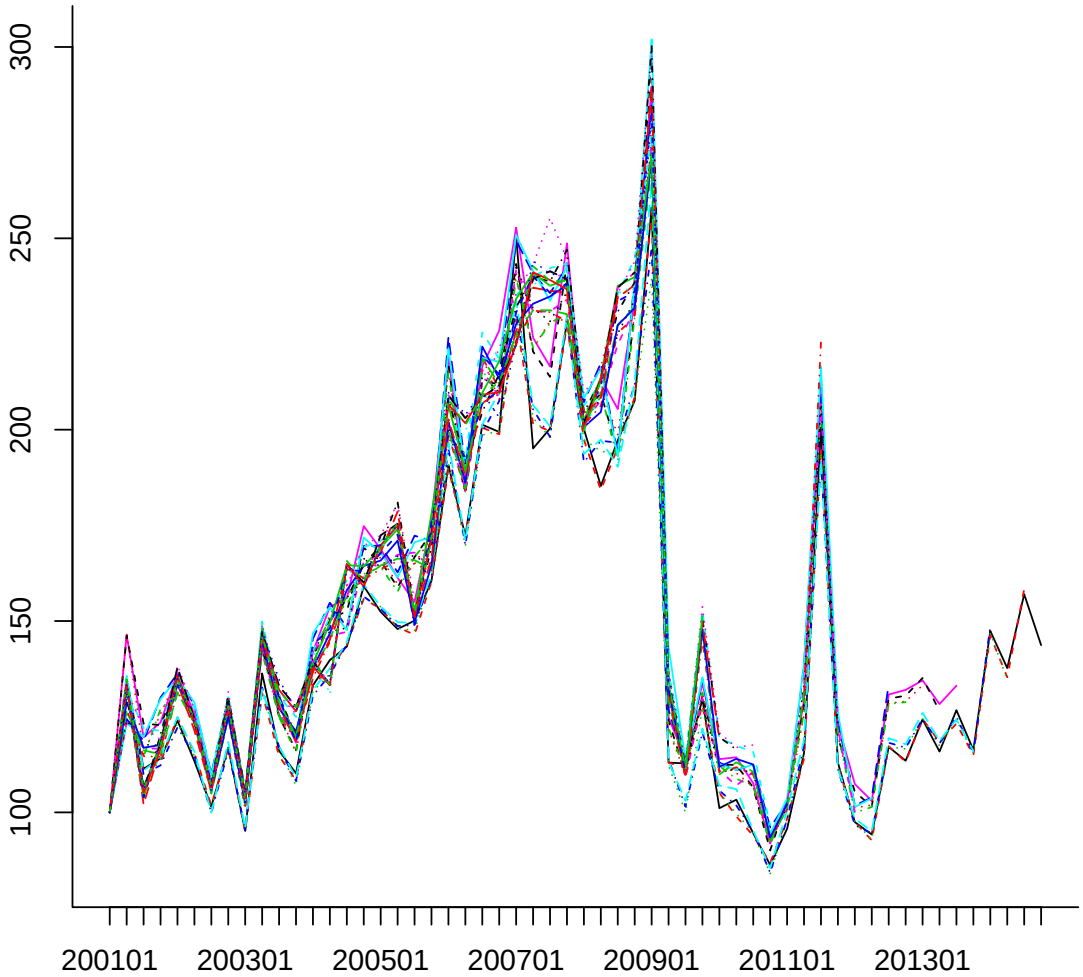
- Increasing attention in pricing literature has been given to thin markets. But why?
- Granular indices give a better representation of the space and asset market the properties are traded in and are thus more informative.
- Think of the following parties:
  - Mortgage underwriters who have evaluate risk/return in certain market segments.
  - Investors who would like to know the performance of their real estate portfolio.
  - (Local) Governments, for construction of GDP or even tax-purposes.
  - Derivative trading.
- For whatever reason you are interested in granular markets, the main challenge is to produce ‘stable indices’.

Outline

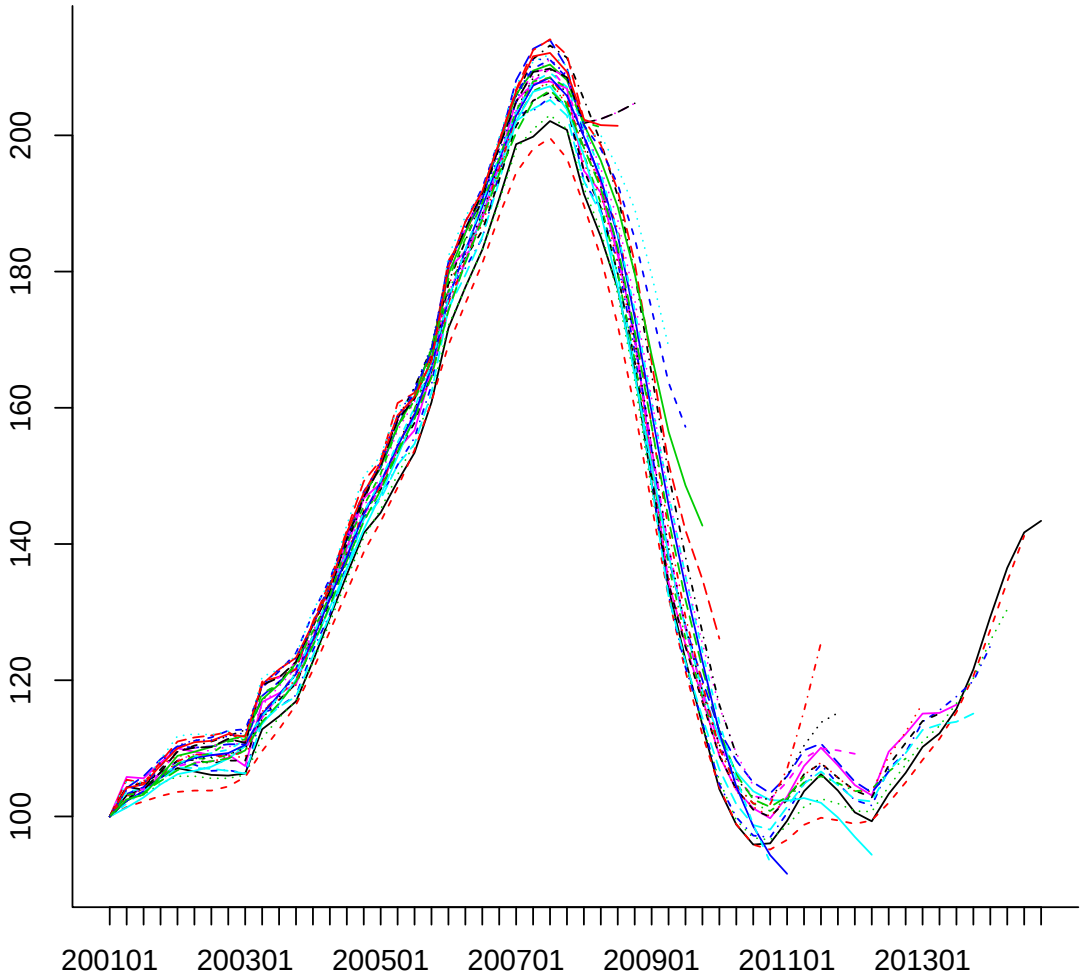
What Revisions?

- Revisions can occur for several reasons:
  - GDP, unemployment, etc. is revised simply because it takes time to collect *all* the data. Thus, an estimate is given first, which is subsequently revised as more data comes in.
  - The S&P 500 is revised as the composition of the S&P can change over time.
  - In real estate the RS methodology is also a culprit.
- In the RS framework, new pairs are observed, instead of new transactions. Thus, a new pair can affect the *entire* index.
  - Small side note: you actually can include single sales in a RS using Random Effects in a Bayesian framework.
- Also, RS indices tend to 'walk down'.
- Next slides provide the revisions for RS and STAR_TSN applied to **Offices** in Los Angeles (LAX) **Inland Empire** with 273 pairs, or 5 per quarter.

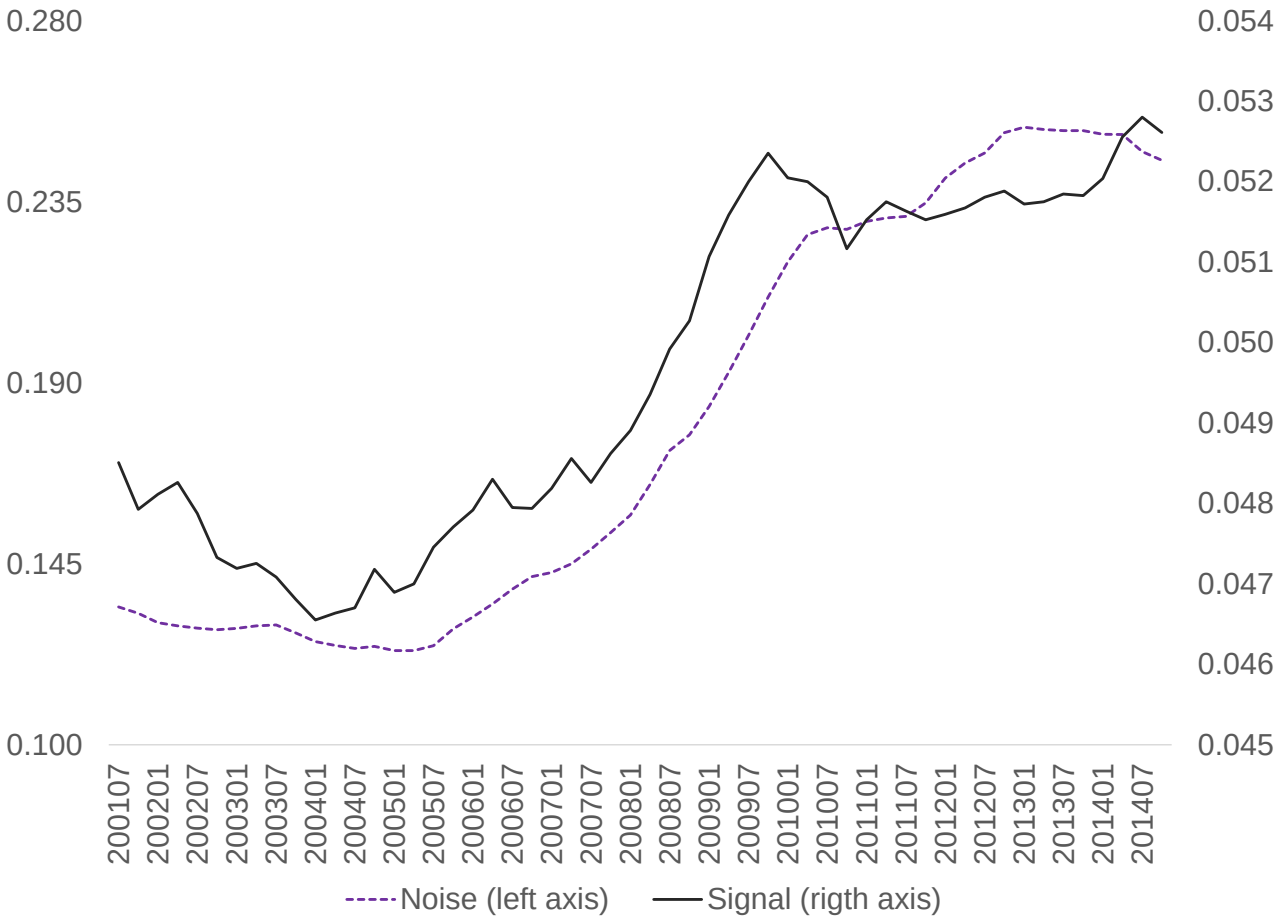
Example, LAX Inland Empire, Repeat Sales



Example, LAX Inland Empire, STAR_TVN

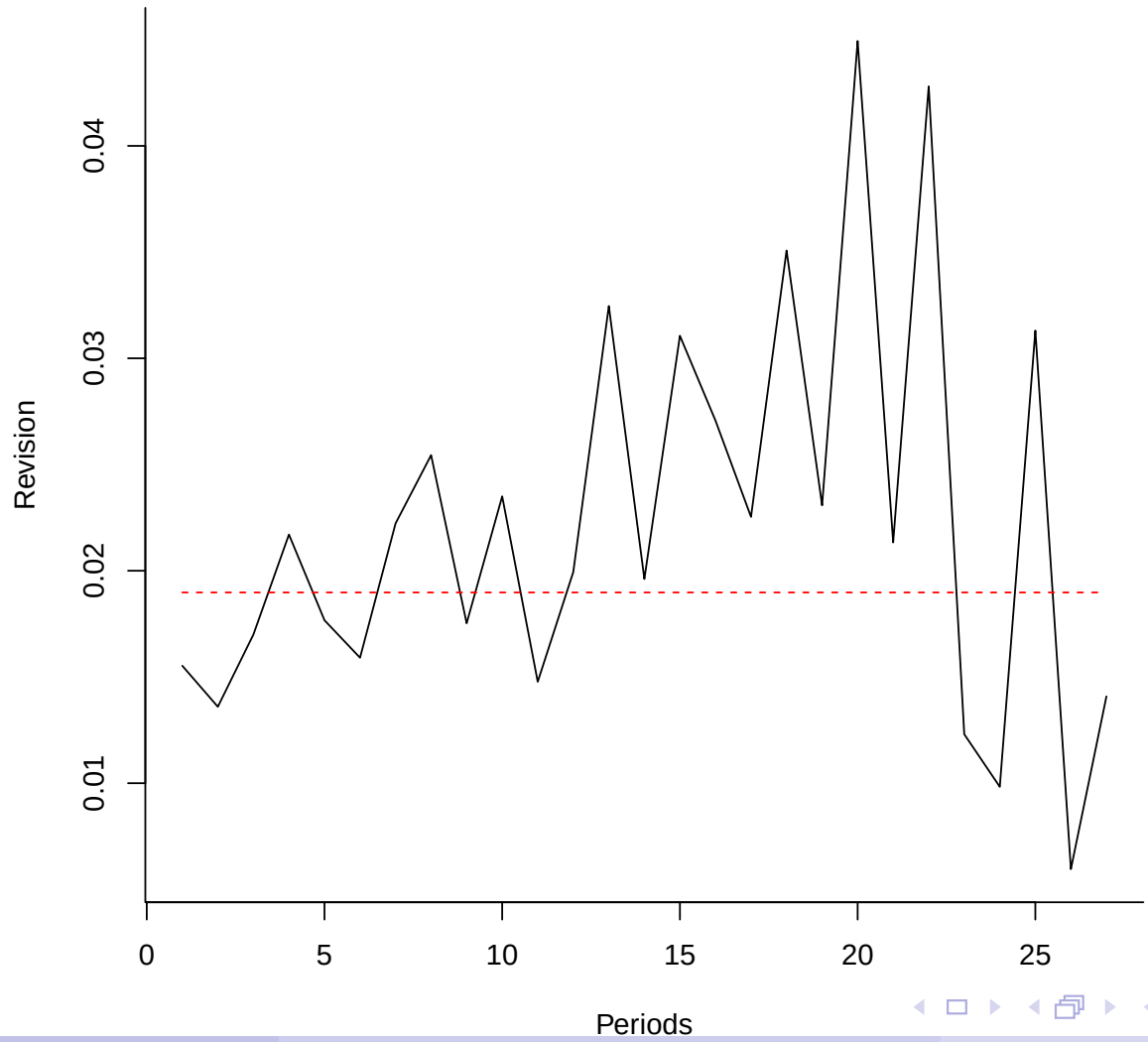


Signal-to-Noise

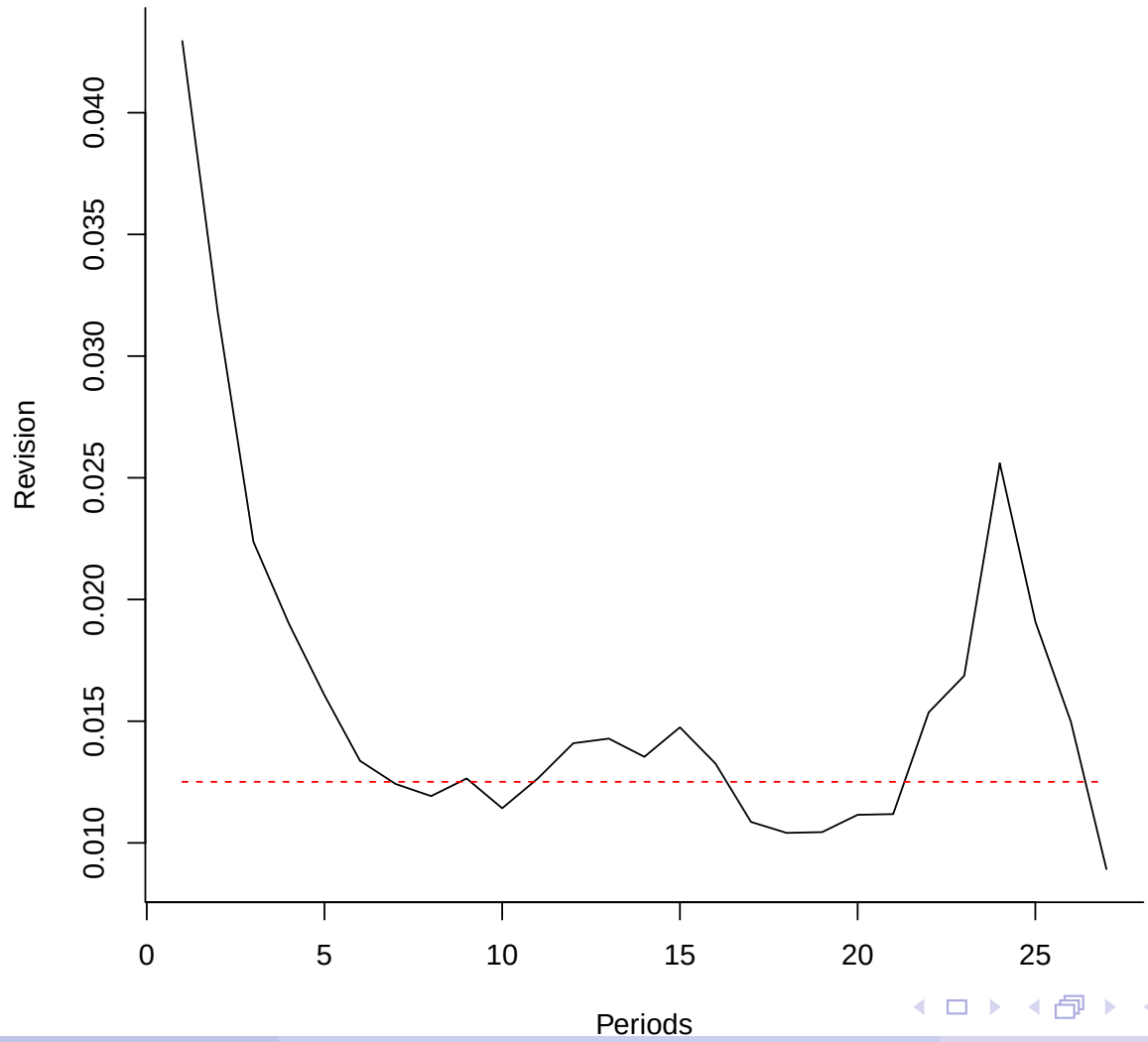


Outline

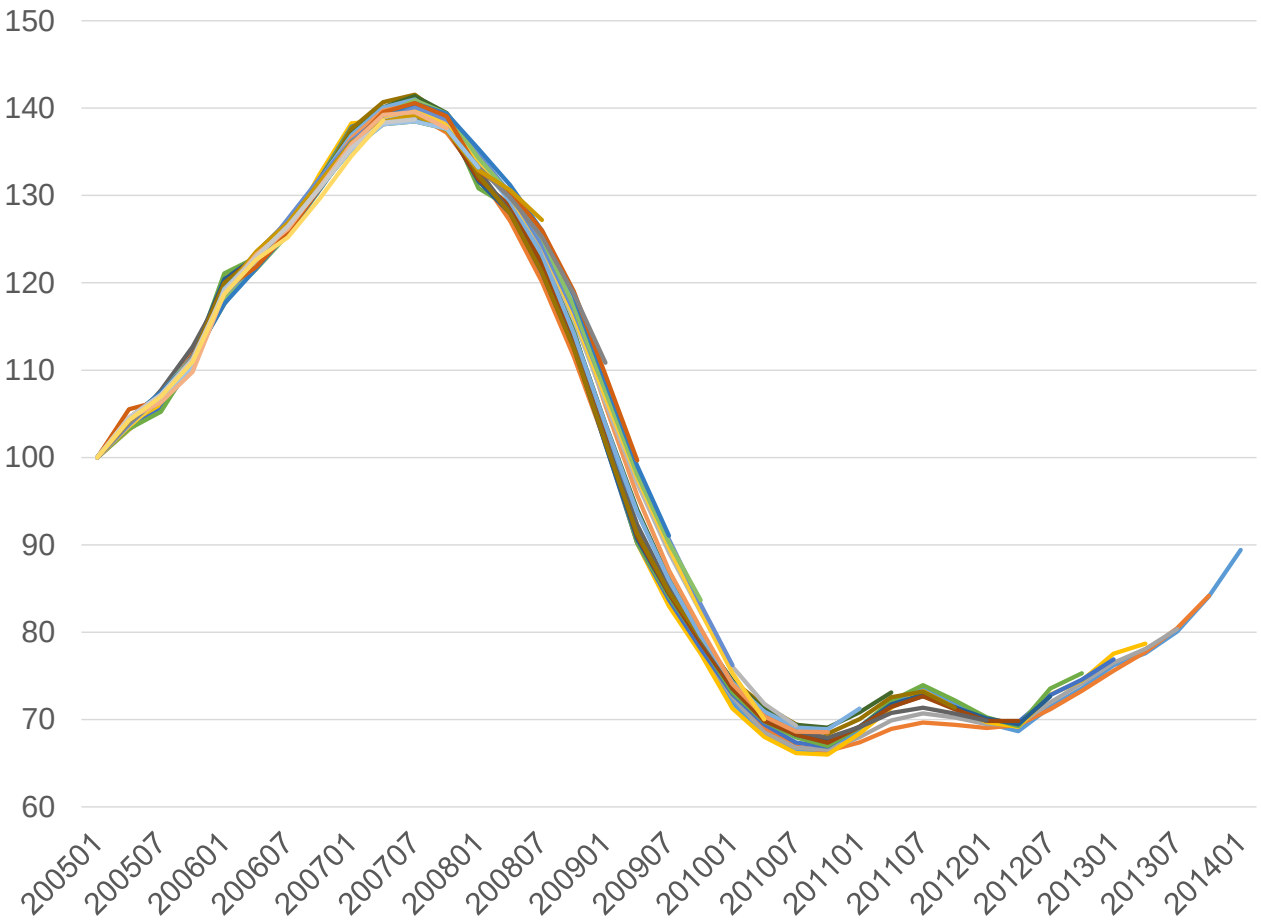
Distribution of Revisions in the RS



Distribution of Revisions in the STAR_TVN



Tradable Index?



Outline

Future Improvements

- In short, going Bayesian works.
- There are, however, many options I did not present. The possibilities are endless! Few examples:
  - Filtering the data prior to analysis improves the revisions by a considerable amount.
    - ★ Estimate polynomial first through the data first.
    - ★ Delete all observation with a residual larger than 0.5 in absolute values.
  - The state vector can contain a huge amount of information, unrelated (directly) to price discovery.
    - ★ Liquidity (trading volume) works wonders!
    - ★ Other economic indicators can be used as well, think of (local) GDP, unemployment, etc. As long as it looks 'cointegrated' with prices.
- For the derivatives market, contracts must allow for revisions for at least 3 periods (in thin markets).
- After those 3 periods we feel very comfortable freezing the index.