AI and Machine Learning in Real Estate Forecasting

Timothy H. Savage, Ph.D. CRE

NYU Schack Institute of Real Estate

Faculty Director of the CREFC Center for Real Estate Finance

ARES 38th Annual Conference

WARNING: IMPORTANT CONTRIBUTORS ARE OMITTED FOR NARRATIVE

Key Points of This Talk

- Econometrics has been a core of economics for at least a century.
- Economists have learned the power and limitations of algorithms.
  - The Lucas critique and Goodhart's law.
  - Creation of panel data and the use of natural experiments.
  - The gold standard (and challenges) of randomized control trials.
- The age of digital exhaust emerges with the smart phone.
- The "AI gang" emerges to claim the machine knows best.
- Contours of the future.
  - Counterfactuals.
  - MCMC simulation and probabilistic inference.

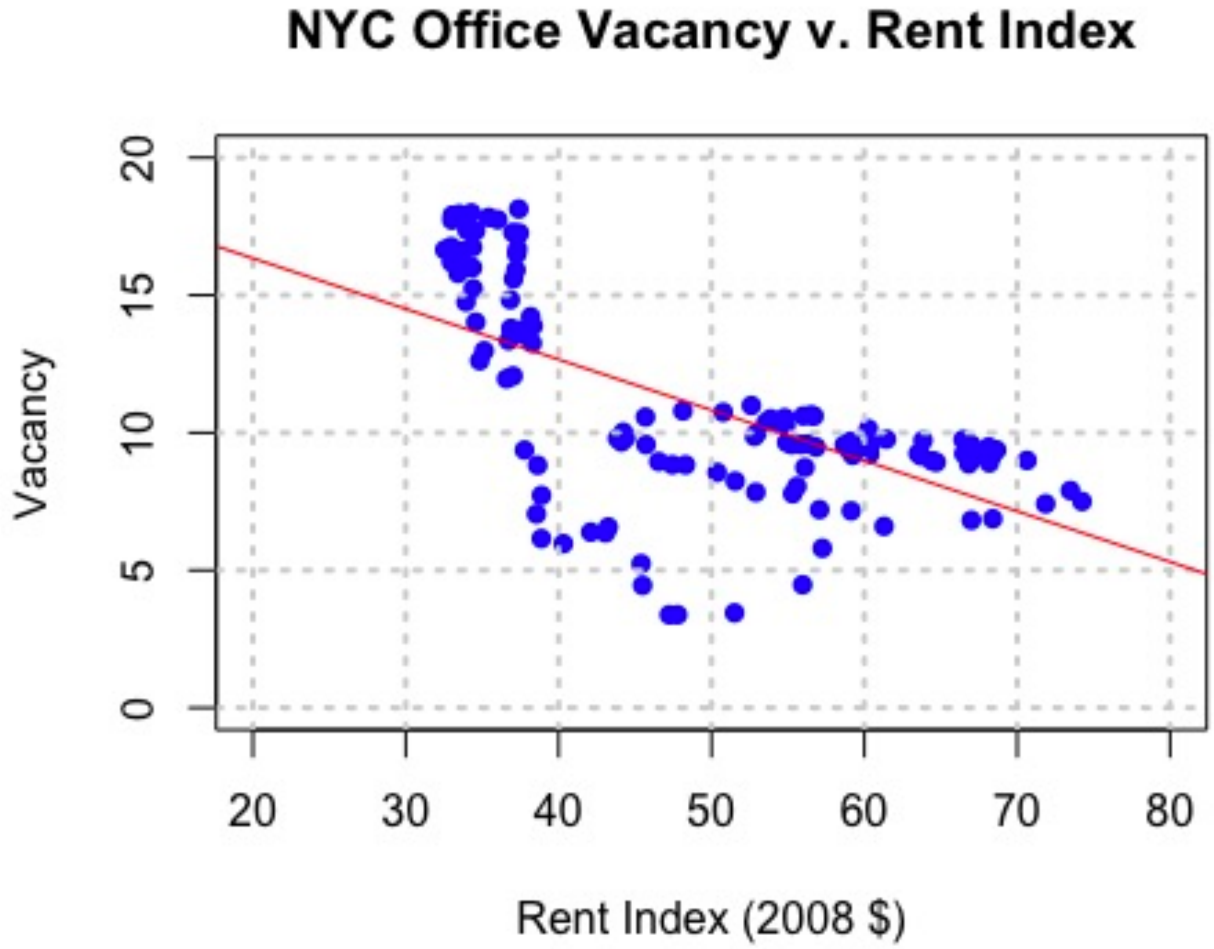
A ***Very*** Terse History

- Bernoulli advances probability as frequency (urn draws).
- Gauss advances OLS and the normal distribution.
- Quetlet advances “the common man” with anthropometry.
- Marshall advances the use of mathematical modeling in economics.
- Markov challenges Bernoulli.
- Fisher, Neyman and Pearson (FNP) advance frequentism as NHST.
- The Great Depression advances the need for national accounts.
- Keynes advances the distinction between micro and macro.
- Cowles advances macro-econometrics to address fluctuations.

A ***Very*** Terse History of Cowles

- GNP = C + G + I + X – M
  - C = C(interest rates, income, G and other stuff)
  - G = G_{bar}
  - I = I(interest rates, income, G and other stuff)
  - X = X(interest rates, income, G and other stuff)
  - M = M(interest rates, income, G and other stuff)
- NIPA develops more granular measures of C and I, in particular.
- Cowles adds more equations.
- Given computational power at the time, a herculean undertaking.

Stark Relief: Raise Rents to Lower Vacancy!



Applying NHST under the FNP Paradigm

- For every $10 per sqft increase in rent, vacancy falls by 184 bps.
  - $p < 0.01$.
- Cowles falters in the 1970’s in the face of stagflation.
  - Goodhart and Lucas.
- Macro goes narrow with Bayesian DSGE.
- Micro goes wide with panel data, natural experiments and RCTs.

Response 1.0: Goodhart

- Goodhart "… asserts that any economic relation tends to break down when used for policy purposes." (Wickens [2008].)
- Proposed relationships, economic or otherwise, are not structural in nature.
- Instead they are derived from fundamental behavioral relationships (structural).

Response 1.1: Lucas

- Lucas (1976) notes that individual decision rules, affected by policy, are driven by “deep structural parameters.”
- Decision rules and, therefore, decisions are contingent on the state of the system as it is.
- Change the system through policy, change the decision rule.
- Such changes may not be captured in non-structural models.

Econometrics Learns

- The emergence of
  - Identification and instrumental variables. (Many.)
  - VARs and VARIMAs to forecast using OOS to evaluate. (Hamilton and Sargent.)
  - Natural experiments to exploit induced-randomization. (Levitt.)
  - RCTs to introduce actual randomization for intervention evaluation. (Duflo.)
- Substantial progress in evaluating the "impact of interventions."
  - Small development grants on economic development.
  - Retirement plans on after-tax savings.
  - Health insurance on health outcomes.
  - Minimum wage laws on employment levels.
  - Opportunity zones on economic development. (Not until RCA gives me the data.)
  - COVID on just about anything. (A few years from now.)

The “Big Data” Revolution Emerges

- What happens when
  - The marginal cost of data storage go to zero?
  - Clustered (parallel) computing emerges?
  - Smart phones create the digital exhaust exhaust of human activity?
  - Open-source computing engines like R and Python emerge?
  - Machines are successfully taught to read numbers and letters?
- Computer scientists ditch “statistical learning” for “AI.”
  - Seemingly unaware of Cowles, Lucas and Goodhart.
  - The fact that we humans choose the algorithms to be used (GIGO).
  - Deep learning is “artificial intelligence” because it approximates any functional form.
- Deep learning is not AI, and outside of image recognition has limited use.

Savage and Vo (2017)

Table 1: CAPM with AAPL Daily Returns			
Model	**Run Time[1]**	**Average MSE[2]**	**Smallest MSE[2]**
OLS	0.009	0.264	0.209
MLP	8.557	0.274	0.217
LSTM[3]	12.170	0.436	0.355
LSTM[4]	19.064	0.384	0.304

1: Average seconds per replication
2: 10^{-3}
3: 30-day window
4: 90-day window

Table 2: FF5F with AMZN Daily Returns			
Model	**Run Time[1]**	**Average MSE[2]**	**Smallest MSE[2]**
OLS	0.024	0.109	0.099
MLP	13.424	0.111	0.100
LSTM[3]	14.093	0.142	0.128
LSTM[4]	22.236	0.133	0.116

1: Average seconds per replication
2: 10^{-2}
3: 30-day window
4: 90-day window

Response 2.0: Pearl’s Ladder of Causation

- **Association**: What if I see something?
  - What does a symptom tell me about a disease?
  - What does a survey tell me about political attitudes?
- **Intervention**: What if I do something?
  - What if I take this medicine, will my disease be cured?
  - What if the government subsidizes education, will wages rise?
- **Counterfactual**: What if I acted differently?
  - Was is it the medicine that cured my disease?
  - If Hurricane Sandy had not hit NYC, would traffic patterns be the same?

Savage and Vo (2015)

Big data typically arises non-experimentally. It is the digital exhaust of human activity. As a result, correlation cannot be interpreted as a causal treatment effect. Using data on taxi rides in New York City, an increasingly popular big dataset, we present an approach that permits causal interpretation. Using a random forest learner and all taxi journeys in the year prior to Hurricane Sandy, we develop Markov transition probabilities that capture the likelihood of a particular drop off location conditional on the pickup location and other features of the journey. The fitted forest is then used to simulate an alternative set of transition probabilities using randomization over the pickup location together with random sampling with replacement of the other features of the journey. These simulated transition probabilities can be compared to observed transition probabilities with differences being equivalent to a treatment effect. Our use of the NYC taxi data serves as a proof of concept. This methodology can easily be extended to a number of different contexts.

Response 2.1: Bayes, Price and Leplace

- What is probability?
  - “There is a catastrophic error in the logic of the standard statistical methods in almost all of the sciences. As science has become increasingly data-driven, this foundational crack has spread into a reproducibility crisis threatening to bring down entire disciplines of research. Outside of science, we see the same fallacies in medicine, law, and public policy, often with disastrous consequences.” [Clayton (2021)].
- Probability is not the outcome of independent Bernoulli urn draws.
  - Such a process measures frequency not probability.
  - Linear regression captures sampling frequencies and not inferential probability.
- Probability is
  - Deductive reasoning with uncertainty (degree of belief) properly expressed as a PDF.
  - The plausibility of a hypothesis (conjecture) given information (the data).
- Leplace’s equation of inverse probability: $P(A \mid B) = P(A) * P(B \mid A) / P(B)$

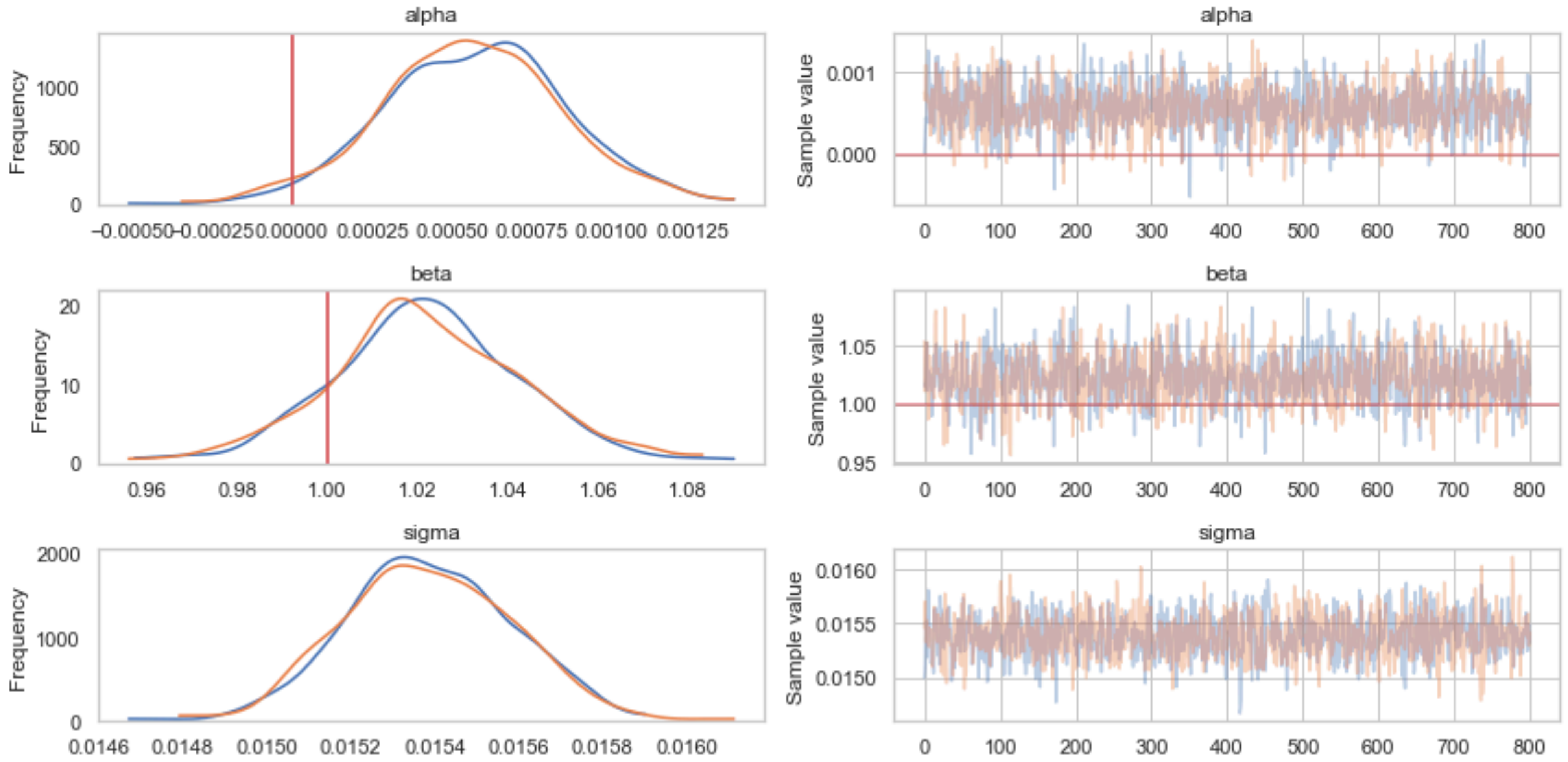
Bayes-Price-Laplace Inference

- “The now-familiar Bayes-Price-Laplace formula states that the posterior probability distribution over a parameter of interest is proportional to its prior probability distribution, the researcher’s degree of belief, times the likelihood function describing the data. On paper, this is a logically-consistent approach to learning and hypothesis testing. But mathematically, closed-form solutions are limited, and only recent advances in computational power have made it tractable for general application.” [Savage (2019)].
- “The physicist Richard Feynman once said that, ‘I can live with doubt and uncertainty... I have approximate answers and possible beliefs and different degrees of certainty about different things, and I’m not absolutely sure of anything...’ Arguably, this is not a statement that sees probability as the outcome of history of coin flips. Economics and finance researchers study constrained decision-making under uncertainty. Real estate development and investing, given its vast array of risk, is a case study for living with doubt and uncertainty. For every real estate success, there may be a large corpus of failure, where probability as the outcome of a history provides little guidance. As a result, a Bayesian approach, either philosophical or practical, might be a tonic.” [Savage (2019)].

But Bayesian Priors Are Subjective

- True, but they are transparent and stated in the terms of probability.
- Besides
  - Why would I ignore prior information, properly expressed as degree of belief?
  - IID is as subjective but not transparent. (And is foundational to frequentism.)
  - FNP paradigm of NHST relies on “repeated samples” that do not exist.
  - FNP paradigm was developed to reflect their views on eugenics.
  - Modern computing power makes large-scale MCMC simulation trivial.
  - The data dominate: Bayesian CLT.
  - Posteriors can be used as priors with new information.

Savage, REAL-CG.3135, Class 14



P(I Am Wrong) in the Absence of Urn Draws?

- P(It Is Not AI) $\cong$ P(the sun will come out tomorrow) < 1.
- AI must include ideas on probabilistic modeling and causation.
- For CRE forecasting, new approaches will be deployed.
  - Bayesian ARIMAs.
  - Unsupervised learning.
  - Advanced dimensionality reduction.
- For this, CRE data must be **more**
  - **Accurate**.
  - **Granular**.
  - **Timely**.

Thank You

My NYU Schack course has much more information and background.

Link: https://github.com/thsavage/Real-Estate-Data-Analytics