My Problem: Portfolio Selection

- Modern finance treats financial assets as being identified by measurable characteristic sets.
  - Size
  - Book-to-Market Ratio
  - Recent price behavior
- Although it has the same name and ticker symbol, AT&T stock in 2025 is a very different investment from AT&T stock in 1960.
- Use the set of measurable characteristics to characterize multivariate return density (e.g., rather than / in addition to historical mean and variance).
- **Question:** Do the characteristics—measurable at time t—contain information for *beating the market*?

Brandt, Santa-Clara, Valkanov (2009)

- Develop a brilliant algorithm to translate this perspective into optimal portfolio construction.
  - Maximize average utility of monthly returns in-sample over a small set of parameters tilt portfolio weights away from value-weighting at time t as function of characteristics at $t-1$.
  - Start with value-weighted (market) portfolio, augment with 1 "tilt" parameter (θ_j) for every characteristic, $j = 1, \ldots, K$.
  - Since this algorithm does not work with the underlying dgp hope is that it solves the well-known problem of overfitting—estimation risk that plagues in-sample optimization. For example, traditional Markowitz, Mean-Variance portfolio selection.
  - It is more general than the Mean-Variance approach as higher-order moments of the return distribution affect general CRRA utility.

Lamoureux and Zhang (2024)

- Examine overfitting in this setting:
  - estimation risk lives in portfolio variance.
  - A relatively risk averse investor's optimal portfolio does not exhibit overfitting (e.g., Power Utility with RRA coefficient ≥ 6).
  - But it is penurious for more risk-tolerant investors (e.g., log utility, Power Utility with RRA < 4).
- We address in a way common in Machine Learning: Bootstrap to construct sampling distributions and out-of-sample validation.
- We select optimal characteristic set and tuning parameter using Max-Min.
- We use the curvature of the loss function brought to the data as a tuning parameter: An power utility investor with CRRA $= 2$ should use a CRRA $= 3$ *in-sample* to mitigate estimation risk.
- Find that the gains from the characteristic-based strategy vanish in the 21st Century.

Current Project

- I start with several things I don't like about my 2024 paper:
  - Approach violates the Likelihood Principal. (Data I might have seen but did not matters to my decision.)
  - It is not a coherent decision rule for several reasons. (Bernardo and Smith, 2000)
  - Out-of-sample validation destroys immediacy – very costly in a changing environment.
  - The bootstrap requires certain properties of the dgp, which is undesirable since we want to be agnostic in this regard.
  - I know my utility function–why am I using Max-Min? I want to integrate signal extraction and portfolio selection. The distinction between ambiguity aversion and my utility function is artificial.

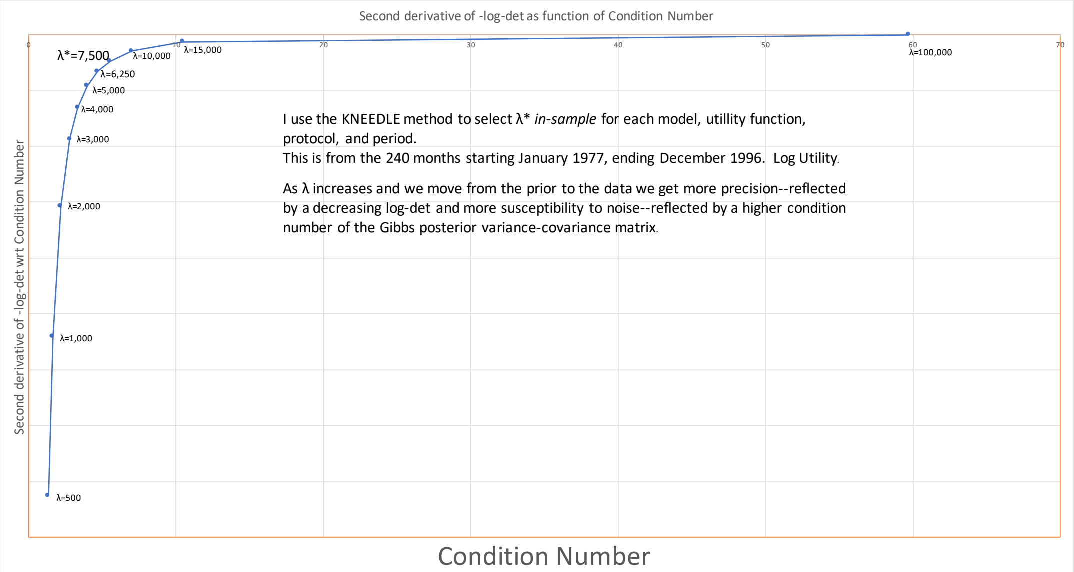
Generalized Bayesian Inference: Gibbs Posteriors

- Form prior on what I don't know: $\pi(\theta)$. $\pi(\theta) \sim \mathcal{N}(0, 1)$.
- Gibbs Posterior: $P_\lambda(\theta|\mathrm{data}) \sim \exp\left(\lambda \cdot U(\theta, \mathrm{data})\right)\pi(\theta)$.
  - Since $\exp(U)$ is not a density, λ scales it to the prior; λ is weight on data relative to prior.
  - Minimizes Kullback-Liebler distance to prior (coherence).
- Implementation: I use a Random Walk Metropolis-Hastings within Gibbs algorithm to generate the Gibbs posterior.
  - Numerically intensive
  - Proposal density: Stable Paretian
  - Excellent convergence properties
  - More numerically robust than optimization (esp. for log utility)

λ

- Intuition:
  - Low values of λ relative to prior's scale will result in Gibbs posterior that looks like the prior.
  - Very high values of λ relative to prior's scale will result in Gibbs posterior that looks like optimizing Utility in sample.
- Rather than use out-of-sample validation to choose λ^*, evaluate the effects of λ on the Gibbs posterior variance-covariance matrix. As λ increases:
  - Tightens precision: log-det decreases
  - Increases sensitivity to noise: Condition Number increases (ratio of largest eigenvalue to smallest)

λ^*: Inflection Point

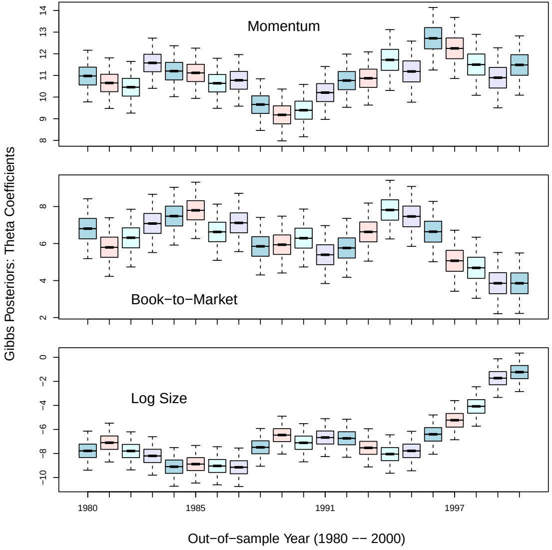


λ^*

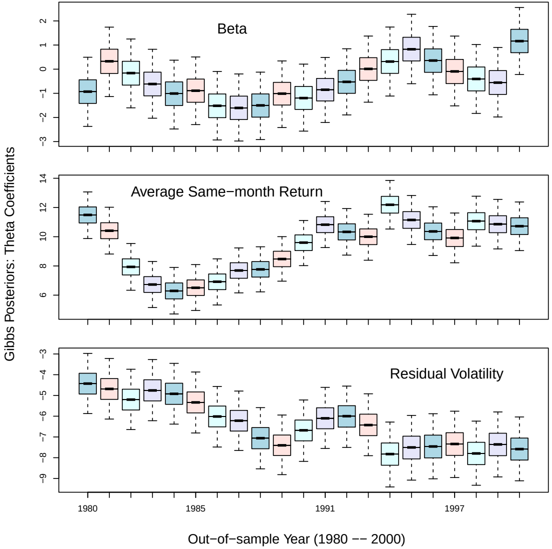
- Generally decreasing in risk aversion.
- For log utility tends to decrease in time.
- Example: log utility: λ^* drops over 18% in 20th Century and another 45% by 2015.
- But it's still not enough. As in Lamoureux and Zhang, better off with market portfolio in this century.
- Confronted with the portfolio decision: Use the posterior $E_{\lambda^*}(\theta|\mathrm{data})$ rolling annually.

Monthly Certainty Equivalent Returns			
Period	Opt. Portfolio	VWI	EWI
1980–2000	6.07%	1.27%	1.29%
2001–2024	0.20%	0.68%	0.84%

Gibbs posteriors on $(\theta|\lambda^*)$



Gibbs posteriors on $(\theta|\lambda^*)$



Key references on Generalized Bayesian Inference and Gibbs Posteriors

- Bissiri, P.G., C.C. Holmes, and S.G. Walker, 2016, A general framework for updating belief distributions, *Journal of the Royal Statistical Society Series B*, **78**, 1103–1130.
- Zhang, T, 2006, From ϵ-entropy to KL-entropy: Analysis of minimum information complexity density estimation, *Annals of Statistics*, **34**, 2180–2210.
- Zhang, T, 2006, Information theoretical upper and lower bounds for statistical estimation, *IEEE Transactions on Information Theory*, **52**, 1307–1321.