Big Data and AI in Econometrics, Finance, and Statistics Conference Review

LaTeX by Junfan Zhu
Stevanovich Center for Financial Mathematics
University of Chicago
October 3-5, 2024

October 7, 2024

Contents

1 Program Agenda

These are conference notes from prominent economists, statisticians, and leading academic researchers across various disciplines who were invited to the Big Data and AI in Econometrics, Finance, & Statistics (BDAI) Conference, held at the University of Chicago's Stevanovich Center for Financial Mathematics from October 3-5, 2024. My selected takeaways primarily focus on Financial Mathematics and Econometrics.

Disclaimer: The rights to the content presented in this material are reserved by the respective presenters, their collaborators, and the BDAI Conference hosted by the Stevanovich Center for Financial Mathematics at the University of Chicago. A detailed list of their work is referenced below, following the chronological order of the conference. [15] [9] [6] [4] [5] [16] [14] [19] [10] [7] [8] [11] [17] [18] [2] [12] [3] [1] [13]

References

[1] Yacine Ait-Sahalia. Asset pricing in an economy with changing sentiment and price feedback. *Princeton*.

[2] Denis Chetverikov. Estimation of risk premia with many factors. *UCLA*.

[3] Francis X. Diebold. Machine learning and the yield curve: tree-based macroeconomic regime switching. *UPenn*.

[4] Chao Gao. Are adaptive robust confidence intervals possible? *UChicago*.

[5] Fang Han. Chattejee's rank correlation: what is new? *UWashington*.

[6] Ruimeng Hu. Deep reinforcement learning for games with controlled jump-diffusion dynamics. *UC Santa Barbara*.

[7] Bryan Kelly. Arbitrage pricing theory or '*AI* pricing theory'? the surprising dominance of large factor models. *Yale*.

[8] Mladen Kolar. Confidence sets for causal discovery. *USC*.

[9] Zongming Ma. Multimodal data integration and cross-modal querying via orchestrated approximate message passing. *Yale*.

[10] Theodor Misiakiewicz. Deterministic equivalents and scaling laws for random feature regression. *Yale*.

[11] Andrea Montanari. Deterministic equivalents and scaling laws for random feature regression. *Stanford*.

[12] Whitney Newey. Automatic debiased machine learning via riesz regressions. *MIT*.

[13] Lan Zhang Per Mykland. Estimating the volatility of drift. *UChicago, UIC*.

[14] Bodhi Sen. Multivariate, heteroscedastic empirical bayes via nonparametric maximum likelihood. *Columbia*.

[15] Katja Smetanina. Perceived shocks and impulse responses. *UChicago*.

[16] Pragya Sur. Generalization error of min-norm interpolators in transfer learning. *Harvard*.

[17] Chenhao Tan. Towards human-centered ai: predicting fatigue and generating hypothesis with *LLM*s. *UChicago*.

[18] Dacheng Xiu. On the theory of autoencoders. *UChicago*.

[19] Wenxin Zhou. Nonparametric expected shortfall regression with tail-robustness. *UIC*.

2 Machine Learning and the Yield Curve: Based Macroeconomic Regime Switching [3]

Presenter: Francis X. Diebold, UPenn.

Diebold first introduced the Dynamic Nelson-Siegel (DNS) Model (2006), the famous model for yield curve in macroeconomics. The idea of this talk is to interpret abrupt shifts or regime switching by Yields-Macro DNS model, which facilitates investigation of macro spanning by allowing for yields-macro interaction.

Macro spanning means that all current and past macro relevant for future macro is contained in the current yield curve. Hence yields may Granger-cause macro, but macro should not Granger-cause yields.

Yields-Macro DNS with Regime Switching Setup

$$\begin{pmatrix} y_t \\ m_t \end{pmatrix} = \begin{pmatrix} \Lambda & 0 \\ 0 & I_3 \end{pmatrix} F_t + \begin{pmatrix} \Lambda & 0 \\ 0 & I_3 \end{pmatrix} \mu_{z_t} + \epsilon_t \tag{1}$$

$$F_{f_t} = A_{z_{t-1}} F_t + \eta_t \tag{2}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} Q & 0 \\ 0 & H_{z_t} \end{pmatrix}\right) \tag{3}$$

where State F_{f_t} is a 6-vector, 3 centered yield factors and 3 centered macro indicators (Inflation, capacity utilization, fed funds rate).

They used a Tree-Based Estimation approach that evaluates DNS model's marginal likelihood and select the highest. Technically, using Bayesian estimation conditional on exogenously-known regimes that requires only Kalman filter/smoothers + MCMC. Practically, they adopt a sequential greedy fashion.

Empirics used 1971 - 2022 monthly US TBond yields with 10 macroeconomic variables including CPI, unemployment rate, 3-month TBill Secondary Market Rate, etc.

Results show the comparison among 3 regimes under Yields-Only model with clear economic interpretations.

Summary

- Bond yield curve dynamics may switch with the macroeconomy.
- Implemented both yields-only and yield-macro models.
- Macro spanning may be regime-specific.

3 Sentiment-Based Asset Pricing [1]

Presenter: Yacine Ait-Sahalia, Princeton.

This paper brings up a new continuous-time equilibrium model (CCAPM) with a stochastic sentiment process η_t. The idea is that sentiment incorporates price feedback, where investor gets more optimistic when prices go up, and vice versa. Also, sentiment can jump when it gets more out of line with fundamentals. Equilibrium outcome yields testable closed-form solutions of the Asset price dynamics, Excess return and risk-free rate under sentiment affects, and Conditional variance of returns.

I think it's quite fascinating that they derived a tractable equilibrium framework that allows sentiment to drive investor beliefs and where past price changes feed back into investor sentiment. Specifically, there're two novel features that are important assumptions for the paper:

1. Investor is subject to sudden large jumps in sentiment. Investor is not aware that their beliefs are not grounded in reality. They only care about current sentiment and do not account for potential future changes.

2. The sentiment is subject to price feedback effects. Asset prices affect the investor's sentiment: higher (lower) prices relative to past prices increase sentiment. This generates positive time-series momentum.

So that this paper can explain in reality the facts like

- Boom and bust cycles due to feedback effects between sentiment and stock prices.
- The equilibrium stock price process that rapidly build-up and sudden crashes, even though the fundamentals driving the stock price process (dividends) might not jump, like a Minsky moment.

Framework of the model is based on Lucas-type endowment economy with one representative investor, who has to choose optimal consumption and can invest in two assets, standard except for jumps in dividends. With this setting, we could characterize risky stock price process (with Brownian motion volatility and Poisson process jump) and risk-free process in the equilibrium.

With this setting, construct Investor Sentiment with a Feedback Channel.
The sentiment process $\eta_t \in [-1, 1]$ follows a jump-diffusion process of the form:

$$d\eta_t = k(m_t - \eta_t)dt + \sigma_\eta \sqrt{1 - \eta_t^2} dW_t^\eta - (1 + \eta_{t-})dJ_t^{\eta-} + (1 - \eta_{t-})dJ_t^{\eta+},$$

where:

$$\text{negative jumps: } J_t^{\eta-} = \sum_{i=1}^{N_t^{\eta-}} \left(1 - e^{-Y_i^{\eta-}}\right) - \lambda_t^{\eta-}\phi_{\eta-}t, \quad \phi_{\eta-} = \mathbb{E}\left[1 - e^{-Y_i^{\eta-}}\right]$$

$$\text{positive jumps: } J_t^{\eta+} = \sum_{i=1}^{N_t^{\eta+}} \left(1 - e^{-Y_i^{\eta+}}\right) - \lambda_t^{\eta+}\phi_{\eta+t} \quad \phi_{\eta+} = \mathbb{E}\left[1 - e^{-Y_i^{\eta+}}\right]$$

$N^{\eta-}$ and $N^{\eta+}$ are Poisson processes with time-varying intensities $\lambda_t^{\eta-}$ (decreasing in η) and $\lambda_t^{\eta+}$ (increasing in η).

$$W_t^\eta = \rho W_t^D + \sqrt{1-\rho^2} W_t^i \text{ is a Brownian motion.}$$

Extreme optimism (pessimism) corresponds to the case when $\eta_t \to 1(-1)$, and neutral when $\eta_t = 0$.

Asset price feedback effect captured by the stochastic mean reversion level m_t:

$$m_t = \tanh(qM_t), \quad M_t = \log\left(\frac{I^{-1}\int_{t-1}^t S_u du}{L^{-1}\int_{t-L}^t S_u du}\right)$$

where $q > 0$ and $L > 0$ are parameters.

A moody investor can be modeled under Subjective Measure $\mathbb{P}^\eta$ using the Radon-Nikodym derivative, where the investor's belief of the dividend process is guided by the sentiment process η_t that's disconnected from the reality.

Taking into account investor's preferences and time inconsistency, the presenter further illustrated how they assume investor only considers the current value of sentiment, not the projected evolution of the sentiment. Investor is also not aware that the belief is not aligned with reality of dividend process because investor is not aware of the existence of sentiment, namely, they believe dividend rate will grow with drift forever in the future. This time-inconsistency that investor's belief might change in the future needs to be modeled in the sentiment dynamics and equilibrium.

Proposition 1: Equilibrium asset prices and & risk free rate
In an endowment economy populated by an investor with subjective sentiment-dependent preferences, the equilibrium asset price is given by

$$S_t = \frac{D_t}{k_t}, \text{ where } k_t = A - B\eta_t > 0$$
$$A = \beta - \frac{\psi - 1}{\psi}\left(\mu_D - \frac{\gamma\sigma_D^2}{2} - \lambda_D\zeta_D(\gamma)\right), B = \frac{\psi-1}{\psi}\sigma_D\theta$$

The resulting equilibrium risk-free rate is

$$r_{f,t} = \underbrace{\beta + \frac{1}{\psi}\left(\mu_D - (1+\psi)\frac{\gamma\sigma_D^2}{2}\right)}_{\text{standard diffusive model}} + \underbrace{\frac{1}{\psi}\eta_t\theta_{\sigma_D}}_{\text{misjudgement}} \underbrace{-\lambda_D\left(\xi_D(\gamma) - \frac{\psi-1}{\psi}\zeta_D(\gamma)\right)}_{\text{dividend jump risk}}$$

Proposition 2: Equity Risk Premium under sentiment swings
The instantaneous excess return $\mu_{S,t}^E := \frac{1}{dt}\mathbb{E}_t\left[\frac{dS_t + D_t dt}{S_t}\right] - r_{f,t}$ is given by

$$\mu_{S,t}^E = \underbrace{\gamma\sigma_D^2 + \lambda_D\xi_D(\gamma)}_{\text{Standard model}} - \underbrace{\eta_t\theta\sigma_D}_{\text{misjudgment}} + \underbrace{\left(\frac{B}{A - B\eta_t}\right)^2\sigma_\eta^2\left(1-\eta_t^2\right)}_{P/D\text{-Sentiment risk premium}} + \underbrace{\frac{B\rho\sigma_D\sigma_\eta\sqrt{1-\eta_t^2}}{A - B\eta_t}}_{\text{Cov. risk premium}}$$
$$+ \frac{B}{A - B\eta_t}\left\{\underbrace{\kappa\left(m_t - \eta_t\right)}_{\text{Feedback Effect}} + \underbrace{B\left(\lambda_t^{\eta-}\Psi_t^- + \lambda_t^{\eta+}\Psi_t^+\right)}_{\text{Sentiment jump risk premium}}\right\}$$

where the negative and positive sentiment jump terms are given by

$$\Psi_t^- = \int_0^\infty \frac{\left(1+\eta_t\right)^2\left(1-e^{-y}\right)^2}{A + B\left[1 - \left(1+\eta_t\right)e^{-y}\right]}\nu_-^\eta(dy), \quad \psi_t^+ = \int_0^\infty \frac{\left(1-\eta_t\right)^2\left(1-e^{-y}\right)^2}{A + B\left[\left(1-\eta_t\right)e^{-y} - 1\right]}\nu_+^\eta(dy)$$

Misjudgement term is reflecting an investor is overly pessimistic or optimistic, leading to higher or lower equity risk premium. Feedback effect increases (decreases) equity risk premium provided that $m_t > \eta_t$, high past price appreciation, i.e. m_t increase can generate positive momentum.

Empirics are done by S&P 500 log returns including dividends daily from 1962 to 2023. Sentiment Data comes from Investor Sentiment Survery Index by American Association of Individual Investors(AAII), authors construct a bullish-bearish spread measure as a proxy of investor sentiment, from 2000 daily.

Results:

- Positive association between the equilibrium risk-free rate and investor sentiment.
- Negative relationship between excess return and sentiment driven by general equilibrium effect.
  - → Since this is an endowment economy with a sole representative agent, prices need to adjust such that the investor is willing to hold all the asset supply.
  - → For markets to clear properly, if the investor becomes very optimistic, the implied excess return must fall otherwise the agent would want to short sell the risk free asset to increase his/her holdings in the risky asset.

Summary

- Develops a flexible yet tractable framework where the representative investor is subject to mood swings.
- Introduces a novel sentiment process which allows for price feedback effects (can generate boom and bust cycles without jumps in the stock price fundamentals).
- Empirically validates the model's features and shows that feedback effects as well as jumps in sentiment are important to capture stylized facts of asset markets.
- The model's predictions for the excess return, dividend yield, realized volatility, and the risk-free rate are largely supported by empirical results.

Further steps:

- Relationship between investor sentiment and higher order moments.
- Closed-form expressions for higher moments of excess return as a function of sentiment.
- Full estimation of the model to recover model parameters and analyze effect of sentiment changes on excess returns and the risk-free rate in more detail.

4 Volatility of Drift [13]

Presenter: Per Mykland & Lan Zhang, UChicago & UIC

Price drift is stock price movement in high frequency data. High frequency data's main features are: takes up to millisecond updates, almost continuous observation, observation times can be irregular, and has microstructure noise.

Start with log price process $\{x_t\}$ following semi-martingale,

$$dx_t = \theta_t dt + \sigma_t dW_t,$$

with drift θ_t, and volatility σ_t. The paper focuses on the drift over an interval $\int_0^T \theta_t dt$ and analyze multi-period extensions with intervals $(T_0, T_1], (T_1, T_2], \ldots$ The idea is to estimate the volatility $[\theta, \theta]_T - [\theta, \theta]_0$, or approximation of Average Realized Volatility (ARV) of the drift θ:

$$\text{ARV} = \frac{1}{\mathcal{T}} \frac{1}{L} \sum_i \left(\theta_{T_{i+L}} - \theta_{T_i}\right)^2$$

Consider tapered QV in our estimator:

$$\widehat{\text{ARV}} = \frac{1}{\mathcal{T}\lambda\delta} \left(QV_{B,k,L}(X) - QV_{B,k+L,0}^{\text{tapered}}(X)\right)$$

$\widehat{\text{ARV}}$ depends on k (short lag), L (long lag), and B. Asymptotic Variance of $\widehat{ARV}$ − ARV Martingale = martingale + edge effects = $O_p(T^{-1/2})$ for fixed δ, λ.

The Dog(s) that Didn't Bark in the Night

Problems:

- Provides no "discipline" for L and k.
- Average integrated volatility

$$IV = \frac{1}{\mathcal{T}} \left(\langle\theta, \theta\rangle_{T_{B-k-L}} - \langle\theta, \theta\rangle_{T_k}\right)$$

- Edge I: Difference between IV and $\frac{1}{\mathcal{T}} \left(\langle\theta, \theta\rangle_{\mathcal{T}} - \langle\theta, \theta\rangle_0\right)$; either additional asymptotic variance term, or adjust summation limits to cover whole interval.
- Edge II: Edge effect in $\widehat{ARV} - \text{ARV}$: bias and variance terms.

Derive the CLT for the Average Realized Volatility of Drift, consider a common sense penalty for large λ, even when θ is a not a martingale. There's an Uncertainty principle that if L is large, one can estimate volatility of drift, but cannot precisely locate it in time.

- Asymptotic Bias:

$$\underbrace{\mathcal{T}^{-1}\left(\frac{1}{4}\langle\theta, \theta\rangle'_{T_B} - \frac{1}{4}\langle\theta, \theta\rangle'_0 + \langle M, \theta\rangle'_{T_B} + \langle M, \theta]'_0\right\rangle}_{\text{from Edge II term, from } \widehat{\text{ARV}} - \text{ARV}}$$
$$- \underbrace{\mathcal{T}^{-1}\left(\delta\langle\theta, \theta\rangle'_0 + (\delta + \lambda)\langle\theta, \theta\rangle'_{\mathcal{T}}\right)}_{\text{Edge I term}}$$

- Asymptotic Variance

$$\underbrace{\mathcal{T}^{-2}\left(16\lambda^{-2}\delta^{-1}\int_0^{\mathcal{T}} \left(\langle M, M\rangle'_t\right)^2 dt + 8\lambda^{-1}\int_0^{\mathcal{T}} \left(\langle M, M\rangle'_t\langle\theta, \theta\rangle'_t + \left(\langle M, \theta\rangle'_t\right)^2\right) dt\right)}_{\text{martingale part of } \widehat{\text{ARV}} - \text{ARV}}$$
$$+ \underbrace{\mathcal{T}^{-2}\lambda\frac{4}{3}\int_0^{\mathcal{T}} \left(\langle\theta, \theta\rangle'_t\right)^2}_{\text{ARV} - \text{IV} = \mathcal{T}^{-2}\lambda\eta_B^2} + \underbrace{\frac{1}{4}\mathcal{T}^{-2}\lambda^{-1}\left(\langle\sigma^2, \sigma^2\rangle'_0\right) + \langle\sigma^2, \sigma^2\rangle'_{\mathcal{T}_B}\Big)}_{\text{from Edge II}} = O_p\left(\mathcal{T}^{-1}\right)$$

Summary

- Attempt to extract price drift, in the form of its variation
- Incorporating High Frequency technique (in-fill asymptotic) in drift analysis allows better estimation of drift variation
- Empirically, can detect drift change over months, even weeks in crisis time
- In higher dimension: Principal Component Analysis

5 On the Theory of Deep Autoencoders [18]

Presenter: Dacheng Xiu, Chicago Booth and NBER.

This paper adopted AutoEncoder (AE) due to their close connection with linear factor models and their capability in conducting nonlinear dimension reduction. Presenters throwed 3 questions:

- Can AEs capture the "commonalities" in the inputs, a procedure often referred to as denoising, and if so, what are the statistical error bounds?
- How do AEs' architecture parameters, such as depth, width, and the number of neurons, impact their statistical performance?
- Can AEs recover the hidden low-dimensional representations in a nonlinear factor model?

Start with nonlinear factor model by Yalcin and Amemiya (2001).

$$X_{it} = X_{it}^{*} + U_{it} = \phi(F_{it}) + U_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

- F_t^*, a K-dimensional vector, represents latent factors.
- $\phi_i^* : \mathbb{R}^K \to \mathbb{R}$, an unknown function, whose functional form can vary across i.
- U_{it} accounts for idiosyncratic noise.

The function f of a Deep Neural Networks (DNNs) with architecture parameters (d, w) can be expressed as:

$$f : \mathbb{R}^{n_0} \to \mathbb{R}^{n_{d+1}}, \quad \times \to f(x) = W_d \sigma_{v_d} W_{d-1} \sigma_{v_{d-1}} \cdots W_1 \sigma_{v_1} W_0 \times$$

- W_i represents the weight matrix and v_i the shift (bias) vector at layer i.
- d denotes depth of the DNN, whereas w is its width. n_0 and n_{d+1} represent the dimensions of the input and output.
- $\sigma_x : \mathbb{R}^r \to \mathbb{R}^r$ is the shifted ReLU activation function:

$$\sigma_x \begin{pmatrix} y_1 \\ \vdots \\ y_r \end{pmatrix} = \begin{pmatrix} \max(y_1 - x_1, 0) \\ \vdots \\ \max(y_r - x_r, 0) \end{pmatrix}$$

where $x = (x_1, \ldots, x_r) \in \mathbb{R}^r$

Define function class of DNNs

$$\mathcal{F}_{n_0}^{n_{d+1}}(d, w, C, B) := \left\{ f \text{ of the form (1)} : \|f\|_\infty \leq B, \max_{j=0,\ldots,d} \left\|W_j\right\|_\infty, \max_{j=1,\ldots,d} \left\|v_j\right\|_\infty \leq C \right\}$$

AE class can be defined as follows:

$$\mathcal{F}_{\mathrm{AE}}^{K_1} := \left\{ (\rho, \varphi_1, \ldots, \varphi_N) : \rho \in \mathcal{F}_{K_0}^{K_1}\left(d_1, w_1, T^{5\beta+5}, B\right), \varphi_i \in \mathcal{F}_{K_1}^{1}\left(d_2, w_2, T^{5\beta+5}, B\right) \right\}$$

K_1 is the pre-selected number of neurons in the AE's bottleneck layer. ρ and φ_i are DNNs.

In training AEs, the paper used stochastic gradient descent with adaptive learning rates (e.g., RMSprop, Adam, ...). The key tuning parameter is K_1, which is closely tied to the architecture of the *AE*. With examniations of error decomposition, denoising performance, overparameterized encoder and factor pervasiveness, the authors proposed properly parameterized encoder:

$$\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\hat{X}_{it} - \varphi_i^{\star}\left(F_t^{\star}\right) \right)^2 \lesssim \left(N^{-1} K_1 + T^{-\frac{2\beta}{2\beta+K}} + L^{-1} \right) \log^4(T),$$

where $\hat{X}_{it} := \hat{\varphi}_i\left(\hat{\rho}\left(X_t\right)\right), c$ and C are constants independent of N, T.

Authors used Monte Carlo simulations to compare the performance of different architecture of AEs and PCA as benchmark. They also showed applications in cross-sectional asset pricing and denoising measurement error.

Nonlinear Asset Pricing Model

- According to arbitrage pricing theory, asset returns follow a linear factor model:

$$R_{it} = \lambda_i F_t^\star + U_{it}$$

- Borri et al. (2024) proposed that excess returns can be modeled as:

$$R_{it} = \varphi\left(\Lambda_i^\top F_t^\star\right) + U_{it}$$

  where $\Lambda_i, F_t^\star \in \mathbb{R}^K$ with $K = 1$, and they fit a polynomial function for φ.

- In contrast, this paper employed AEs to estimate this model. This is achieved by constraining the parameters after the first layer of the decoder to be identical across all outputs and removing the bias in the first layer, ensuring that the output takes the form $\varphi\left(\Lambda_i^\top F_t^\star\right)$.

  AEs can handle cases where $K > 1$ and estimate more complex models beyond polynomial functions.

Summary

- AEs hold promise for nonlinear dimension reduction.
- Non-asymptotic analysis to AEs' denoising errors and factor learning errors, which achieves the optimal rate for nonparametric regression under mild conditions.
- Through simulations and empirical illustrations, AEs provide superior performance compared with PCA under nonlinear factor models.

6 Arbitrage Pricing Theory or AI Pricing Theory? The Surprising Dominance of Large Factor Models [7]

Presenter: Bryan Kelly, Yale.

Authors showed that out-of-sample univariate timing strategy performance generally increasing in model complexity (# of parameters). Bigger models are better. ML in cross-sectional Asset Pricing derives increasing SDF performance with higher portfolio Sharpe ratio and smaller pricing errors.

SDF representable as managed portfolios: $M_{t+1}^{\star}=1-\sum_{i=1}^{n} w\left(X_{t}\right)^{\prime} R_{i, t+1}$, s.t. $E_{t}\left[M_{t+1}^{\star} R_{i, t+1}\right]=0 \forall i$

- Cross-sectional asset pricing is about $w_{t}=w\left(X_{t}\right)$
- Fundamental challenge in cross-sectional asset pricing: w must be estimated
  This is a high-dimensional (complex) problem
- Standard approach: Restrict w's functional form
  - E.g., Fama-French: $w_{i, t}=b_{0}+b_{1}$ Size $_{i, t}+b_{2}$ Value $_{i, t}$ (Brandt et al. 2007 generalize)
  - Reduces parameters, implies factor model: $M_{t+1}=1-b_{0} M K T-b_{1} S M B-b_{2} H M L$
  - Shrinking the cross-section (Kozak et al., 2020): use a few PCs of anomaly factors

In Machine Learning perspective, rather than restricting $w\left(X_{t}\right)$, this paper:

- Expand parameterization, saturate with conditioning information
- E.g. approximation via neural network: $w\left(X_{i, t}\right) \approx \lambda^{\prime} S_{i, t}$, where $P \times 1$ vector $S_{i, t}$ is known nonlinear function of original predictors $X_{i, t}$
- Implies that empirical SDF is a high-dimensional factor model with factors F_{t+1}

$$M_{t+1}^{\star} \approx M_{t+1}=1-\lambda^{\prime} \underbrace{S_{t}^{\prime} R_{t+1}}_{=F_{t+1} \in \mathbb{R}^{P \times 1}}=1-\lambda^{\prime} F_{t+1}$$

Model

- n assets with returns R_{t+1}
- Empirical SDF $M_{t+1}=1-\lambda^{\prime} S_{t}^{\prime} R_{t+1}$. Think of S_{t} as "generated features" in neural net with input X_{t}. $P \times 1$ vector of instruments, S_{t} (i.e., P factors F_{t+1})
- Ridge-penalized objective: $\hat{\lambda}(z)=\left(z l+\frac{1}{T} \sum_{t} F_{t} F_{t}^{\prime}\right)^{-1} \frac{1}{T} \sum_{t} F_{t}$
- Goal: Characterize out-of-sample behaviors, contrast simple (small P) models vs. complex models
- Tools: Joint limits as numbers of observations and parameters are large, T, $P \rightarrow \infty$, RMT

Empirics: Monthly return of US stocks from CRSP 1963-2021. Conditioning info $\left(X_{i, t}\right)$: 130 stock characteristics from Jensen, Kelly, and Pedersen (2022). Out-of-sample performance metrics are SDF Sharpe ratio and Mean squared pricing errors (nonlinear factors as test assets).

This paper adopted Random Fourier Features in empirical model: $M_{t+1}=1-\lambda^{\prime} S_{t}^{\prime} R_{t+1}$
Let $X_{i, t}$ be 130×1 predictors. RFF converts $X_{i, t}$ into

$$S_{\ell, i, t}=\left[\sin \left(\gamma_{\ell}^{\prime} X_{i, t}\right), \cos \left(\gamma_{\ell}^{\prime} X_{i, t}\right)\right], \quad \gamma_{\ell} \sim \operatorname{iidN}(0, \gamma I)$$

$S_{\ell, i, t}$: Random lin-combo of $X_{i, t}$ fed through non-linear activation
RFF is a two-layer neural network with fixed weights (γ) in the first layer and optimized weights (λ) in the second layer.

7 Nonparametric Expected Shortfall Regression with Tail-robustness [19]

Presenter: Wenxin Zhou, UIC

Given $\tau \in (0,1)$, the standard Quantile Regression estimator is

$$\widehat{\beta}(\tau) = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(Y_i - X_i^T \beta\right)$$

where $\rho_\tau(u) = u\{\tau - \mathbb{1}(u<0)\}$

Positive Semidefinite Kernel. Fix a compact space $\mathcal{X}$, such as $[0,1]^p$. $K(\cdot,\cdot) : X \times X \to \mathbb{R}$ is symmetric and PSD : for any $n \geq 1$ and $\{x_i\}_{i=1}^n \subseteq \mathcal{X}, \left(K_{ij} = K\left(x_i, x_j\right)\right)_{1 \leq i,j \leq n}$ is positive semidefinite.

Reproducing Kernel Hilbert Space. RKHS $\mathcal{H} \subseteq L_2(\mathcal{X})$ associated to a PSD kernel K with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the unique Hilbert space satisfying

- For any $x \in \mathcal{X}$, the function $K(\cdot, x) : \mathcal{X} \to \mathbb{R}$ belongs to $\mathcal{H}$
- Kernel reproducing property: $\langle f, K(\cdot, x)\rangle_{\mathcal{H}} = f(x), \quad \forall f \in \mathcal{H}$

Kernel Trick. Fitting linear models depends on inner product $\langle x, x' \rangle$ in $\mathbb{R}^d$

Mercer's theorem. There exist eigenvalues $\mu_j \geq 0$ and eigenfunctions ϕ_j-an orthonormal basis of $L_2(\mathcal{X}, \mathbb{P})$ - of the integral operator

$$T_K(f)(x) = \int_{\mathcal{X}} K\left(x, x'\right) f\left(x'\right) \mathrm{d}\mathbb{P}\left(x'\right)$$

such that $T_K\left(\phi_j\right) = \mu_j \phi_j$ and $K\left(x, x'\right) = \sum_{j=1}^{\infty} \mu_j \phi_j(x) \phi_j\left(x'\right)$

Kernel Ridge Regression. Given $\{(Y_i, X_i)\}_{i=1}^n$ with $X_i \in \mathcal{X}, Y_i \in \mathbb{R}$,

$$\hat{f} \in \operatorname{argmin}_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \ell\left(Y_i - f\left(X_i\right)\right) + \lambda \|f\|_{\mathcal{H}}^2$$

The regularization parameter λ controls bias/variance trade-off:

$\lambda \uparrow \Rightarrow$ encourages $\hat{f}$ to have smaller $\|\hat{f}\|_{\mathcal{P}}$, or smoother

$\lambda \downarrow \Rightarrow \hat{f}$ fits the data better (less bias) at the cost of smoothness

Kernel Ridge Quantile Regression. In the first step, estimate the conditional quantile function f_0 using kernel ridge regression:

$$\hat{f} = \hat{f}_n\left(\lambda_q\right) = \operatorname{argmin}_{f \in \mathcal{H}} \left\{\frac{1}{n} \sum_{i=1}^{n} \rho_\tau\left(Y_i - f\left(X_i\right)\right) + \lambda_q \|f\|_{\mathcal{H}}^2\right\},$$

where λ_q is a regularization parameter

By the representer theorem (Kimeldorf & Wahba, 1971), the KRR estimator can be expressed as $\hat{f}(\cdot) = \sum_{j=1}^{n} \hat{\alpha}_j K\left(\cdot, X_j\right)$, then we get

$$\min_{\alpha \in \mathbb{R}^n} \left\{\frac{1}{n} \sum_{i=1}^{n} \rho_\tau\left(Y_i - \sum_{j=1}^{n} \alpha_j K\left(X_i, X_j\right)\right) + \lambda_q \alpha^T \mathbf{K} \alpha\right.$$

where $\mathbf{K} = \left(K\left(X_i, X_j\right)\right)_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$ is the kernel matrix.

Kernel Ridge ES Regression. Construct surrogate response variables $\hat{Z}_i = Z_i(\hat{f})$ by plugging-in the Qt-KKR estimate:

$$Z_i(f) = \tau^{-1} \min\left\{Y_i - f\left(X_i\right), 0\right\} + f\left(X_i\right)$$

Then, a two-step ES-KRR estimator $\hat{g}$ is defined as

$$\min_{\mathbf{a} \in \mathbb{R}^n} \left[\frac{1}{n} \sum_{i=1}^{n} \left\{\hat{Z}_i - \sum_{j=1}^{n} \alpha_j K\left(X_i, X_j\right)\right\}^2 + \lambda_e \alpha^T \mathbf{K} \alpha\right]$$

8 Perceived Shocks and Impulse Responses [15]

Presenter: Katja Smetanina, Chicago Booth.

Idea of this paper is to fit a time-varying factor model and recover latent factors (shocks) and loadings (impulse responses) at every point in time for a panel of expectation revisions for one variable across different forecast horizons and over time.

Construct the forecast revisions as difference in forecasts of the same variable made in two consecutive months: $X_{ht} = \widehat{Y}_{h|t} - \widehat{Y}_{h|t-1}$, where panel data of forecast revisions is X_{ht} for a term structure of horizons $h = 1, \ldots, H$ and times $t = 1, \ldots, T$. To model the term structure of forecast revisions as a factor model, allowing for time-varying loadings:

$$X_{ht} = \lambda'_{ht} F_t + e_{ht}$$

where F_t is a vector of $r < N$ latent factors, e_{ht} is noise, F_t are the perceived shocks, λ_{ht} are the perceived impulse responses to each shock at each horizon h and at each point in time t.

Heteroskedastic noise makes PCA estimator invalid in finite H (Bai and Wang, 2016), this paper used a PCA method that works in finite H and that allows for heteroskedastic noise.

Rational expectations and Vector Moving Average model

- Suppose agents use a correctly specified structural VMA models (Plagborg-Moller, 2019) to forecast one of the variables in the system: $Y_t = \Theta(L)\varepsilon_t$, where ε_t is a vector of structural shocks and $\Theta(L)$ the lag polynomial.
- The forecast revision between times $t-1$ and t is

$$X_{ht} = \underbrace{\theta'_h}_{\lambda'_{ht}} \underbrace{\varepsilon_t}_{F_t}$$

- Implies a factor structure with no noise, where perceived IRF = true IRF and perceived shocks = structural shocks

Rational expectations and factor model

- E.g., affine term structure or Nelson-Siegel models for interest rates (typically assuming three dynamic factors β_t) implies that for one maturity:

$$Y_t = \gamma'\beta_t + v_t$$
$$\beta_t = \Phi\beta_{t-1} + \varepsilon_t$$

- The forecast revision between times $t-1$ and t is

$$X_{ht} = \underbrace{\gamma'\phi^h}_{\lambda'_{ht}} \underbrace{\varepsilon_t}_{F_t}$$

- Implies a factor structure with no noise, where perceived IRF = true IRF and perceived shocks = surprises from the state equation

Information rigidities

- E.g., noisy information model of Coibion and Gorodnichenko (2015), where target variable is an AR(1).$Y_t = \rho Y_{t-1} + \varepsilon_t$, with ε_t Gaussian white noise.
- Agent observes a noisy signal $Z_t = Y_t + v_t$, with v_t Gaussian white noise independent of ε_t

- Agent forecasts using the Kalman filter:

$$\widehat{Y}_{t+h|t} = \rho^h \widehat{Y}_{t|t}$$
$$\widehat{Y}_{t|t} = GZ_t + (1-G)\widehat{Y}_{t|t-1}$$

  where the Kalman gain G captures the degree of information rigidity

- Forecast revision is

$$X_{ht} = \widehat{Y}_{t+h|t} - \widehat{Y}_{t+h|t-1} = \underbrace{\rho^h}_{\lambda_{ht}} \underbrace{\left(\widehat{Y}_{t|t} - \widehat{Y}_{t|t-1}\right)}_{F_t}$$

- The (single) factor is $F_t = \widehat{Y}_{t|t} - \widehat{Y}_{t|t-1} = G\left(Z_t - \widehat{Y}_{t|t-1}\right)$, the surprise from the Kalman filter updating equation

- Implies factor structure with no noise, where perceived IRF = true IRF and perceived shock = filtered shock

Finite-sample method with const loadings: HeteroPCA

Zhang, Cai and Wu (2022) proposed HeteroPCA, for estimating factor models under heteroskedasticity in finite H and T. Heteroskedastic noise means that eigenvectors of sample covariance of X_t can be very different from F_t. Issue is with the diagonal (because of independent noise assumption). Solution is to iteratively impute the diagonal of the sample covariance by the diagonal of its low-rank approximation to improve accuracy.

Time-varying HeteroPCA

- This paper extends this finite-sample procedure to allow for time-varying loadings.

- Model time variation in loadings as deterministic functions of time, then estimate nonparametrically:$\lambda_{ht} = \lambda_h(t/T), \quad t = 1, \ldots, T$ where, for each $h, \lambda_h(\cdot) : [0,1] \to \mathbb{R}$ is an unknown piece-wise smooth function of the rescaled time t/T.

Data used for this paper is CPI inflation expectations for a term structure of forecast horizons that are produced monthly. Two datasets are merged: Blue Chip (BC) and Cleveland Fed (CF) to construct a term structure of expectations (not revisions).

Summary

- Time-varying impulse response shows clear evidence of time variation in loadings.

- Agents react to one shock, highly correlated with inflation surprises.

- Secular decrease in the perceived persistence of the shock.

- Recent movements in long-run expectations due to large perceived shocks, not deanchoring.

9 Deep Reinforcement Learning for Games with Controlled Jump-diffusion Dynamics [6]

Presenter: Ruimeng Hu, UCSB.

Use Lévy models to model jump diffusions of varied sizes and frequencies, authors developed efficient ML algorithms for solving control and portfolio games and derived semi-explicit solutions.

Start with Stochastic Control with Jumps:

- State process X_t follows a controlled Itô-Lévy process:

$$\mathrm{d}X_t = b\left(X_{t-}, u_t\right)\mathrm{d}t + \sigma\left(X_{t-}, u_t\right)\mathrm{d}W_t + \int_{\mathbb{R}^d} G\left(X_{t-}, z, u_t\right)\tilde{N}(\mathrm{~d}t, \mathrm{~d}z)$$

- Components:
  - u_t : control process
  - W_t : standard Brownian motion
  - N : Poisson random measure with the Lévy measure ν
  - $\tilde{N}(\mathrm{~d}t, \mathrm{~d}z) = N(\mathrm{~d}t, \mathrm{~d}z) - \nu(\mathrm{d}z)$: compensated Poisson random measure

Optimal Value Function:

$$v(t, x) = \sup_{u \in U} J^u(t, x), \quad J^u(t, x) = \mathbb{E}\left[\int_t^T f\left(s, X_s, u_s\right)\mathrm{d}s + g\left(X_T\right) \mid X_t = x\right]$$

Solves the partial-integro differential equation (PIDE):

$$\frac{\partial v}{\partial t}(t, x) + \sup_{u \in U}\left[\mathcal{L}^u v(t, x) + {}^v f(t, x, u)\right] = 0, \quad v(T, x) = g(x)$$

$\mathcal{L}^u$ is the generator of X_t

$$\begin{aligned}\mathcal{L}^u v(t, x) = & b(x, u) \cdot \nabla v(t, x) + \frac{1}{2}\operatorname{Tr}\left[\sigma(x, u)\sigma^T(x, u)\mathrm{H}(v(t, x))\right] \\ & + \int_{\mathbb{R}^d}(v(t, x + G(x, z, u)) - v(t, x) - G(x, z, u) \cdot \nabla v(t, x))\nu(\mathrm{d}z)\end{aligned}$$

Key Challenges

- Non-locality of PIDE: The presence of jumps introduces partial-integro differential equations (PIDE), the term $\int_{\mathbb{R}^d}$ is hard to evaluate numerically
- High-Dimensionality
- Full Nonlinearity: The presence of u in G makes no explicit expression of u^*

This paper used Deep Reinforcement Learning Approach to tackle above challenges. They introduced Actor-critic type framework to solve both the optimal value function $v(t, x)$ and the optimal control $u^*(t, x)$ simultaneously, e.g., solve the value function by policy evaluation, the control (actor) by policy improvement, and then parameterize both by neural nets. Advantages of doing so are that it's good to have accurate solution for the control policy, and it's suitable for fully nonlinear problems (no explicit u^* required), and easy to update via SGD.

Objective: Compute the value function $J^u(t, x)$ for a given policy $u(t, x)$

In Temporal Difference Learning, incremental learning procedure is driven by the error between the current value estimate and the actual observed reward plus the future value, updates occur after every state: $X_{t_1} \rightarrow X_{t_2} \cdots X_{t_n} \rightarrow X_{t_{n+1}} \cdots$

Bellman equation:

$$J^u\left(t_n, X_{t_n}\right)=\mathbb{E}\left[R_{t_n t_{n+1}}+J^u\left(t_{n+1}, X_{t_{n+1}}\right) \mid X_{t_n}\right]$$
$$R_{t_n t_{n+1}}=\int_{t_n}^{t_{n+1}} f\left(t, X_t, u_t\right) \mathrm{d} t$$

TD Update Rule: parameterize J^u by a Neural Net $\mathcal{N}_v$

$$\text { Loss }=\left(R_{t_n t_{n+1}}+\mathcal{N}_v\left(t_{n+1}, X_{t_{n+1}}\right)-\mathcal{N}_v\left(t_n, X_{t_n}\right)\right)^2$$

To improve the TD Update, this paper updates rule for $\mathcal{N}_v$ includes additional martingale terms, and defined alternative reward $\widetilde{R}_{t_n t_{n+1}}$:

$$\begin{aligned}
\widetilde{R}_{t_n t_{n+1}}= & \int_{t_n}^{t_{n+1}} f\left(t, X_t, u\left(t, X_t\right)\right) \mathrm{d} t-\int_{t_n}^{t_{n+1}}\left(\sigma\left(X_t, u\left(t, X_t\right)\right)^T \nabla \mathcal{N}_v\left(t, X_t\right)\right)^T \mathrm{~d} W_t \\
& -\int_{t_n}^{t_{n+1}} \int_{R^d}\left[\mathcal{N}_v\left(t, X_t+G\left(X_t, z_v u\left(t, X_t\right)\right)\right)-\mathcal{N}_v\left(t, X_t, u\left(t, X_t\right)\right)\right] \widetilde{N}(\mathrm{~d} t, \mathrm{~d} z)
\end{aligned}$$

To improve the current policy u to achieve higher $J^u(t, x)$, parameterize u by $\mathcal{N}_\pi$, and improve $\mathcal{N}_\pi$ by GD of

$$J^u(t, x)=\mathbb{E}^{t, x}\left[\int_0^T f\left(s, X_s, u\left(s, X_s\right)\right) \mathrm{d} s+g\left(X_T\right)\right]$$

State Process:

$$\mathrm{d} X_t=b\left(X_{t-}, u_t\right) \mathrm{d} t+\sigma\left(X_{t-}, u_t\right) \mathrm{d} W_t+\sum_{k=1}^m z_k\left(X_{t-}, u_t\right) \mathrm{d} M_t^k$$

where $M_t^k=N_t^k-\lambda_k t$ is the compensated Poisson process. It shows better interpretability and traceability, with applications in portfolio optimization and option pricing.

Merton's problem. Consider investing between a bond S_t^0 and a stock S_t^1 :

$$\begin{aligned}
\mathrm{d} S_t^0 & =r S_t^0 \mathrm{~d} t \\
\mathrm{~d} S_t^1 & =S_{t-}^1\left(\mu \mathrm{~d} t+\sigma \mathrm{d} B_t+z \mathrm{~d} M_t\right)
\end{aligned}$$

The wealth process is

$$\mathrm{d} X_t=\left(r+u_t(\mu-r)\right) X_{t-} \mathrm{d} t+\sigma u_t X_{t-} \mathrm{d} B_t+z u_t X_{t-} \mathrm{d} M_t$$

and the agent aims to maximize $v(t, x)=\sup_u \mathbb{E}\left[g\left(X_T\right) \mid X_t=x\right], \quad$ utility func. g

Stochastic LQR problem. The state process X_t satisfies

$$\mathrm{d} X_t=u_t \mathrm{~d} t+\sigma \mathrm{d} B_t+z \mathrm{~d} M_t$$

where $M_t=\left(M_t^1, \cdots, M_t^d\right), M_t^i=N_t^i-\lambda_i t$ and N_t^i denotes a Poisson process with intensity λ_i. The agent aims to solve

$$v(t, x)=\inf_u \mathbb{E}^{t, x}\left[\int_t^T\left(q\left\|u_s\right\|_r^2+b\left\|X_s\right\|^2\right) \mathrm{d} s+a\left\|X_T\right\|^2\right]$$

Stochastic Games with Jumps. State Process Dynamics: $X_t=\left(X_t^1, X_t^2, \ldots, X_t^n\right)^T$ satisfy:

$$\mathrm{d} X_t=b\left(X_{t-}, \pi_t\right) \mathrm{d} t+\sigma\left(X_{t-}, \pi_t\right) \mathrm{d} W_t+\int_{\mathbb{R}^d} G\left(X_{t-}, z, \pi_t\right) \tilde{N}(\mathrm{~d} t, \mathrm{~d} z)$$

where $\pi_t=\left(\pi_t^1, \ldots, \pi_t^n\right)$ is the control vector, and $b(\cdot), \sigma(\cdot), G(\cdot)$ describe the drift, diffusion, and jump terms, respectively.

Multi-agent portfolio game

Game Objective: Agent i aims to maximize the expected utility:

$$v^i(t,x)=\sup_{\pi^i} J_i^\pi(t,x)=\sup_{\pi^i} \mathbb{E}^{t,x_s}\left[\int_t^T f_i\left(s,X_s,\pi_s\right)\mathrm{d}s+g_i\left(X_T\right)\right]$$

- The stock price and agent i 's wealth process:

$$\begin{aligned} \mathrm{d}S_t^i &= S_{t-}^i\left(\mu_i\ \mathrm{d}t+\nu_i\ \mathrm{d}W_t^i+\sigma_i\ \mathrm{d}B_t+\alpha_i\ \mathrm{d}M_t^i+\beta_i\ \mathrm{d}M_t^0\right), \quad 1\le i\le n \\ \mathrm{d}X_t^i &= \frac{\pi_t^i}{S_{t-}^i}\ \mathrm{d}S_t^i=\pi_t^i\left(\mu_i\ \mathrm{d}t+\nu_i\ \mathrm{d}W_t^i+\sigma_i\ \mathrm{d}B_t+\alpha_i\ \mathrm{d}M_t^i+\beta_i\ \mathrm{d}M_t^0\right)\end{aligned}$$

- Agent i aims to maximize

$$J_i\left(\pi^1,\cdots,\pi^n\right)=\mathbb{E}\left[U_i\left(X_T^1,\cdots,X_T^n\right)\right]$$

For the exp case, agent i 's utility function U_i depends on the arithmetic average of others':

$$U_i\left(x_1,x_2,\cdots,x_n\right)=-e^{-\frac{1}{\delta_i}\left(x_i-\theta_i\frac{1}{n}\sum_{k=1}^n x_k\right)}$$

For power and log cases, U_i takes into account the geometric average:

$$U_i\left(x_1,x_2,\cdots,x_n\right)=\frac{1}{p_i}\left(\frac{x_i}{\left(\prod_{k=1}^n x_k\right)^{\frac{\theta_i}{n}}}\right)^{p_i}, \log\frac{x_i}{\left(\prod_{k=1}^n x_k\right)^{\frac{\theta_i}{n}}}$$

References

[1] Yacine Ait-Sahalia. Asset pricing in an economy with changing sentiment and price feedback. *Princeton*.

[2] Denis Chetverikov. Estimation of risk premia with many factors. *UCLA*.

[3] Francis X. Diebold. Machine learning and the yield curve: tree-based macroeconomic regime switching. *UPenn*.

[4] Chao Gao. Are adaptive robust confidence intervals possible? *UChicago*.

[5] Fang Han. Chattejee's rank correlation: what is new? *UWashington*.

[6] Ruimeng Hu. Deep reinforcement learning for games with controlled jump-diffusion dynamics. *UC Santa Barbara*.

[7] Bryan Kelly. Arbitrage pricing theory or '*AI* pricing theory'? the surprising dominance of large factor models. *Yale*.

[8] Mladen Kolar. Confidence sets for causal discovery. *USC*.

[9] Zongming Ma. Multimodal data integration and cross-modal querying via orchestrated approximate message passing. *Yale*.

[10] Theodor Misiakiewicz. Deterministic equivalents and scaling laws for random feature regression. *Yale*.

[11] Andrea Montanari. Deterministic equivalents and scaling laws for random feature regression. *Stanford*.

[12] Whitney Newey. Automatic debiased machine learning via riesz regressions. *MIT*.

[13] Lan Zhang Per Mykland. Estimating the volatility of drift. *UChicago, UIC*.

[14] Bodhi Sen. Multivariate, heteroscedastic empirical bayes via nonparametric maximum likelihood. *Columbia*.

[15] Katja Smetanina. Perceived shocks and impulse responses. *UChicago*.

[16] Pragya Sur. Generalization error of min-norm interpolators in transfer learning. *Harvard*.

[17] Chenhao Tan. Towards human-centered ai: predicting fatigue and generating hypothesis with *LLM*s. *UChicago*.

[18] Dacheng Xiu. On the theory of autoencoders. *UChicago*.

[19] Wenxin Zhou. Nonparametric expected shortfall regression with tail-robustness. *UIC*.