Deep Learning for Corporate Bonds

Jetlir Duraj Oliver Giesecke[2]

[2]Hoover Institution, Stanford University

Frankfurt School of Finance & Management
AI Frontiers in Finance

March 11, 2025

This Paper

Asset pricing for U.S. corporate bonds

- U.S. corporate bonds are large asset class
- 11.2 trillion USD outstanding as of Q3 2024 (SIFMA)

⇒ Assemble unique dataset that contains bond returns, macro fundamentals, individual bond characteristics, issuer fundamentals, equity returns

Estimates SDF in tradable asset space via machine learning

- Deal with large amount of conditioning information (more than 240 time series!)
- Allow for flexible functional form of the SDF
- Contrast estimation using *minimization of mispricing loss* and *maximization of Sharpe ratio*

⇒ SDFs for both *bond portfolios* and *individual bonds*

Summary

Bond portfolios:

- Annual out-of-sample Sharpe ratios between 0.76 and 0.87
- Incorporation of macro time series ⇒ higher Sharpe ratios & lower maximum drawdown
- SDF portfolio shows statistically significant excess returns w.r.t. the S&P500, Fama French 3 and 5 factor models
- Little variation between mispricing-loss and Sharpe ratio maximization

Individual bond returns:

- Annual out of sample Sharpe ratios between .41 and 1.00
- Annual excess return of 12% (t-stat of 3.04) vis-a-vis the S&P500, 12% (t-stat of 2.50) against FF 5 factor model
- Mispricing-loss provides higher cross-sectional and time series predictability, but Sharpe ratio maximization yields better portfolio performance

Literature (Partial List)

- **Asset pricing for treasury securities**
  Fama and Bliss (1987), Cochrane and Piazzesi (2005), Cochrane and Piazessi (2008), Ludvigson and Ng (2009), Bianchi et al. (2021)
- **Asset pricing for corporate bonds**
  Merton (1974), Culp et al. (2018), Bai et al. (2019), Nozawa (2017), Bali et al. (2021), He et al. (2021)
- **Machine learning for asset pricing**
  Kelly et al. (2019), Gu et al. (2020), Freyberger et al. (2020), Gu et al. (2021), Bryzgalova et al. (2020), Chen et al. (2023), Bianchi et al. (2021)

Roadmap

Data Sources

- **Individual Bond Returns**
  - Panel of clean corporate bond prices between February 1973 and January 2020 (Elkamhi et al., 2021)
  - Four main data sources: Lehman Brothers Fixed Income Database, the Mergent FISD/NAIC Database, TRACE, and DataStream
- **Bond Portfolio Returns**
  - 40 bond portfolios constructed by Elkamhi et al. (2021)
  - Sorted based on credit spreads, intermediary capital exposure, credit rating, downside risk, maturity, idiosyncratic volatility, long-term reversals, etc.
- **Macroeconomic time series**
  - Monthly time series data obtained from the FREDMD database (McCracken and Ng, 2016)
  - Six additional time series from Welch and Goyal (2007)

Data Sources

- **Bond issuance and issuer matching**
  - Combination of Mergent FISD, S&P rating file, and Capital IQ identity file
  - Developed iterative name matching algorithm to establish ambiguous matches
- **Issuer fundamentals and equity characteristics**
  - Issuer fundamentals from Compustat
  - Includes total assets, cash position, cash flow information and profitability indicators, etc.
- **Market based characteristics**
  - Use established cross-walk between Compustat and CRSP
  - Share price, shares outstanding and derived variables as market capitalization and market-to-book ratios
  - CAPM and Fama-French 3-factor estimates on factor exposure, alpha, and idiosyncratic volatility from the WRDS Beta Suite

Roadmap

Estimation Approaches

Goal: Estimate the stochastic discount factor in the the space of tradable assets

We use two approaches as candidate objectives:

1. **Approach 1 - Minimizing mispricing-loss (MP-min):** The SDF is defined by the conditional moment restriction implied by no-arbitrage. We minimize the loss over all deviations from the moment condition.
2. **Approach 2 - Maximizing Sharpe ratio (SR-max):** Using the property that the projected SDF into the space of traded assets exhibits the highest Sharpe ratio, we maximize the Sharpe ratio of the SDF portfolio.

Approach 1: No Arbitrage and Mispricing Loss

No arbitrage implies a conditional moment restriction of the following form (Cochrane, 2009):

$$\mathbb{E}_t[M_{t+1} R^e_{t+1}] = 0, \tag{1}$$

M_{t+1} is the stochastic discount factor, R^e_{t+1} is the excess return of the asset over the risk-free rate, and $\mathbb{E}_t$ is the expectation with respect to the time t information set.

We consider a linear SDF with a single risk factor of the following form $M_{t+1} = 1 - \omega_t^T R^e_{t+1}$.

Empirical loss:

$$\frac{1}{N} \sum_{i=1}^{N} \frac{|T_i|}{T} \left\| \frac{1}{|T_i|} \sum_{t \in T_i} \left(1 - \omega(z_{t,\cdot}) \cdot R^e_{t+1}\right) R^e_{t+1,i} \right\|^2 \tag{2}$$

⇒ Estimate the weights ω_t, that minimize the mispricing loss.

⇒ Provides tradable SDF portfolio!

Approach 2: Maximizing Sharpe Ratio

Linearized Sharpe ratio:

- Fractional definition of the Sharpe ratio problematic
- Use a linear version of the Sharpe ratio during training
- The linear version of the Sharpe ratio still balances the first and the second moment of the return distribution

Concretely, we use the following linearized Sharpe ratio:

$$\begin{aligned} SR\left(R^e_{t+1}, \omega(z_{t,\cdot})\right) &= \text{t-series-mean}\left(\omega(z_{t,\cdot}) \cdot R^e_{t+1}\right) \\ &\quad - \gamma \times \text{t-series-stdev}\left(\omega(z_{t,\cdot}) \cdot R^e_{t+1}\right) \end{aligned} \tag{3}$$

- γ can be interpreted as a parameter of risk-aversion that controls the relative importance of the mean and the standard deviation of the distribution; similar to Guijarro-Ordonez et al. (2021)
- Use constant $\gamma = 0.01$

Roadmap

Bond Portfolio Model Results

Environment		Results Test Set						
Objective	Data	Sharpe	Sortino	Calmar	Max-DD	CAGR	Max- lev	Mean-lev
MP	Bondpf + Macro	0.25	14.09	0.03	0.06	0.02	1.15	0.44
SR	Bondpf + Macro	0.24	12.28	0.03	0.07	0.02	1.14	0.44
MP	Bondpf	0.22	5.10	0.03	0.17	0.05	2.70	0.99
SR	Bondpf	0.22	12.04	0.03	0.09	0.03	1.23	0.56

- The best model achieves a monthly Sharpe ratio of 0.25 (annual Sharpe ratio of 0.87).
- Little difference in Sharpe ratios using the Sharpe ratio maximization vis-a-vis the mispricing-loss.
- Omission of macro time series leads to overall worse performance $\Rightarrow$ lower Sharpe ratio, Sortino ratio, and lower statistical significance for the excess returns. Higher maximum drawdowns, and higher leverage

Bond Portfolio Model Results

Environment		Alphas in Test Set		
Objective	Data	S&P 500	FF 3 Factors	FF 5 Factors
MP	Bondpf + Macro	**0.001**	**0.001**	0.001
		(2.220)	(2.070)	(1.850)
SR	Bondpf + Macro	**0.001**	0.001	0.001
		(2.100)	(1.900)	(1.700)
MP	Bondpf	0.003	0.003	0.002
		(1.920)	(1.770)	(1.570)
SR	Bondpf	0.002	0.001	0.001
		(1.900)	(1.800)	(1.670)

- Significant excess returns vis-a-vis the FF3 factors, marginally significant against FF5 and significant returns against S&P500.
- Excess returns are of magnitude 0.1% at a monthly basis, or 1.2% annually.
- While similar magnitude, statistical significance lower when estimated with the linearized Sharpe ratio.
- Weaker results when omitting the macro time series as conditioning information from the model.

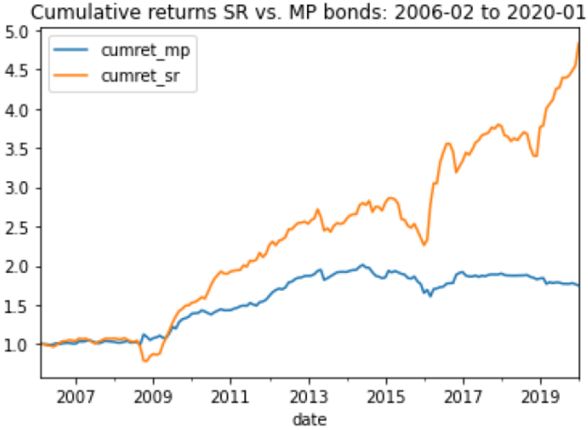
Roadmap

Individual Bond Results

Objective	Environment Data	Sharpe	Sortino	Calmar	Max-DD	CAGR	Max-lev	Mean-lev
		Results Test Set						
MP	Ind. + Bond + Macro	0.17	6.30	0.02	0.20	0.04	3.15	1.29
SR	Ind. + Bond + Macro	0.27	3.93	0.03	0.27	0.12	4.62	2.42
MP	Ind. + Bond	0.12	0.46	0.02	0.56	0.08	4.66	3.51
SR	Ind. + Bond	0.29	2.54	0.05	0.29	0.16	4.29	2.58
MP	Ind. + Bond + Macro - Fin	0.25	2.83	0.03	0.28	0.10	2.70	2.19
SR	Ind. + Bond + Macro - Fin	0.26	2.02	0.03	0.39	0.14	3.97	3.20

- The best performing model with individual bonds achieves a monthly Sharpe ratio of 0.29, or annual Sharpe ratio of 1.00.
- Maximizing the Sharpe ratio performs overall better than minimizing the mispricing loss in terms of Sharpe and Calmar ratios in the test set.
- Sharpe ratio maximization leads to higher CAGR, together with higher and more significant excess returns.
- Best model omits macro-ts but lower Sortino ratio and higher drawdowns than under inclusion.
- Results robust to the exclusion of financials and REITs.

Individual Bond Results



Individual Bond Results

	Environment	Alphas in Test Set			
Objective	Data	S&P 500	FF 3 Factors	FF 5 Factors	Bond Portfolios
MP	Ind. + Bond + Macro	0.003	0.003	0.003	0.001
		(1.870)	(1.920)	(1.820)	(0.380)
SR	Ind. + Bond + Macro	**0.007**	**0.006**	**0.006**	0.001
		(2.700)	(2.380)	(2.290)	(1.000)
MP	Ind. + Bond	0.001	0.002	0.002	-0.001
		(0.180)	(0.530)	(0.480)	(0.710)
SR	Ind. + Bond	**0.010**	**0.008**	**0.009**	**0.003**
		(3.040)	(2.780)	(2.500)	(3.280)
MP	Ind. + Bond + Macro - Fin	**0.005**	**0.005**	**0.005**	0.000
		(2.410)	(2.100)	(1.930)	(0.700)
SR	Ind. + Bond + Macro - Fin	**0.008**	**0.007**	**0.007**	0.001
		(2.470)	(2.140)	(1.960)	(0.700)

Specification with the highest Sharpe ratio yields:

- monthly excess returns of 1.0%, 12.0% annually, vis-a-vis the S&P 500 (t-statistic of 3.040)
- Against the FF5 benchmark: monthly excess return remains close to 12.0% (t-statistic of 2.500)

Individual Bond Results

Environment		Results Test Set	
Objective	Data	EV	XS-R^2
MP	Ind. + Bond + Macro	0.23	0.45
SR	Ind. + Bond + Macro	0.19	0.39
MP	Ind. + Bond	0.20	0.27
SR	Ind. + Bond	0.12	0.18
MP	Ind. + Bond + Macro - Fin	0.22	0.36
SR	Ind. + Bond + Macro - Fin	0.16	0.30

- SDF portfolio estimated using Sharpe ratio maximization performs typically better in terms of portfolio performance, but less well for predictability.
- Minimizing the mispricing loss leads often to substantial increases in the explained variation and the cross-sectional R^2.
- We do not observe this difference for bond portfolios.

⇒ Mispricing loss better for data with a lot of idiosyncratic variation

⇒ Sharpe ratio maximization shows better alignment with portfolio performance.

Conclusion

- We build on recent advances in data availability and machine learning to estimate asset pricing models for the U.S. corporate bonds market.
- Use two approaches: minimization of *mispricing loss* and maximization of the linearized *Sharpe ratio*.
- Sharpe ratio maximization better out-of-sample portfolio performance and excess returns.
- Mispricing loss yields higher predictive performance for samples with idiosyncratic variance.
- In comparison to conventional linear AP models, non-linear methods lead to smaller pricing errors, higher Sharpe ratios and more variance explainability.

References I

Bai, J., T. G. Bali, and Q. Wen (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics 131*(3), 619–642.

Bali, T. G., A. Subrahmanyam, and Q. Wen (2021). Long-term reversals in the corporate bond market. *Journal of Financial Economics 139*(2), 656–677.

Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond risk premiums with machine learning. *The Review of Financial Studies 34*(2), 1046–1089.

Bryzgalova, S., M. Pelger, and J. Zhu (2020). Forest through the trees: Building cross-sections of stock returns. *Available at SSRN 3493458*.

Chen, L., M. Pelger, and J. Zhu (2023). Deep learning in asset pricing. *Management Science*.

Cochrane, J. (2009). *Asset pricing: Revised edition*. Princeton university press.

References II

Cochrane, J. H. and M. Piazessi (2008). Decomposing the yield curve. *University of Chicago working paper, January 2008 version*.

Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American economic review 95*(1), 138–160.

Culp, C. L., Y. Nozawa, and P. Veronesi (2018). Option-based credit spreads. *American Economic Review 108*(2), 454–88.

Elkamhi, R., C. Jo, and Y. Nozawa (2021). A one-factor model of corporate bond premia1. *Working paper, January 2021 version*.

Fama, E. and R. R. Bliss (1987). The information in long-maturity forward rates. *American Economic Review 77*, 680–692.

Fama, E. and R. K. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. and R. K. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

References III

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1), 429–450.

Guijarro-Ordonez, J., M. Pelger, and G. Zanotti (2021). Deep learning statistical arbitrage. *arXiv preprint arXiv:2106.04028*.

He, X., G. Feng, J. Wang, and C. Wu (2021). Predicting individual corporate bond returns. *Available at SSRN 4374213*.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

References IV

Kingma, D. and J. Ba (2015). Adam: A method for stochastic optimization. *ICLR 2015*.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *The Review of Financial Studies 22*(12), 5027–5067.

McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance 29*(2), 449–470.

Nozawa, Y. (2017). What drives the cross-section of credit spreads?: A variance decomposition approach. *Journal of Finance 72*(5), 2045–2072.

Welch, I. and A. Goyal (2007). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies 21*(4), 1455–1508.

Bond Portfolios - Best Models for SDF estimation

Environment		**Best model**				
Objective	Data	Architecture	bs	lr	l2reg	l1reg
MP	Ind. + Bond + Macro	Model 1	256	0.005	0.001	none
SR	Ind. + Bond + Macro	Model 3	256	0.005	none	none
MP	Ind. + Bond	Model 2	256	0.001	0.001	none
SR	Ind. + Bond	Model 2	512	0.001	none	0.001

Notes: The best model describes the architecture, batch size (bs), learning rate (lr) and the l1 (l1reg) and l2 (l2reg) regularization for the model that obtained the highest Sharpe ratio in the validation set.

Back

Individual Bonds - Best Models for SDF estimation

Environment			Best model				
Objective	Data	Exclusions	Architecture	bs	lr	l2reg	l1reg
MP	Ind. + Bond + Macro		Model 1	1024	0.0001	none	0.001
SR	Ind. + Bond + Macro		Model 3	512	0.0001	none	0.001
MP	Ind. + Bond		Model 4	512	0.005	0.001	none
SR	Ind. + Bond		Model 4	2048	0.005	0.001	0.001
MP	Ind. + Bond + Macro	No financials	Model 1	1024	0.0001	none	0.001
SR	Ind. + Bond + Macro	No financials	Model 3	512	0.0001	none	0.001

Notes: The best model describes the architecture, batch size (bs), learning rate (lr) and the L^1 (l1reg) and L^2 (l2reg) regularization for the model that obtained the highest Sharpe ratio in the validation set.

Back

Sample - Bond Portfolios

	value
# of macro variables	126
# of Goyal-Welch Variables	6
# of median time series fundamentals	41
# of forward spreads	10
# return characteristics	9
# of portfolios	40
Sample start	October 1978
Sample end	December 2018

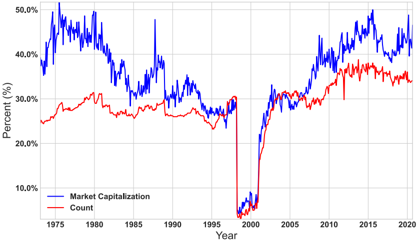
Table: Summary Statistics Bond Portfolio Dataset

Sample - Individual Bond Returns

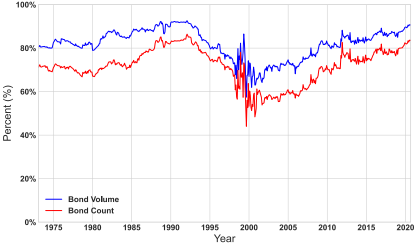
	value
# of firm fundamentals	41
# of macro variables	126
# of Goyal-Welch Variables	6
# of industry dummies	10
# of bond characteristics	9
# of forward spreads	10
# of median time series fundamentals	41
# securities	10,715
# issuers	1,922
Sample start	February 1973
Sample end	January 2020

Table: Summary Statistics Individual Bond Return Dataset

Individual Bond Returns - Sample Composition



(a) Sample as Share of S&P 500



(b) Fraction of the Sample in S&P 500

- Our sample accounts for 30% or more of the market capitalization, and approximately 30% of the number of firms in the S&P500, ex. '98-'01.
- 60%-80% of all bonds in our sample are issued by firms which are part of the S&P 500.

Sample Splitting

We divide our sample into training, validation, and test set.

Individual bond returns:

- We have a total of 560 dates (February 1973 - January 2020)
- first 236 dates as part of our training set ($\approx 40\%$)
- next 156 for the validation set ($\approx 30\%$)
- the remaining 168 dates for the test set (February 2006 to January 2020) ($\approx 30\%$)

Bond portfolios:

- Total of 437 dates (October 1978 - December 2018)
- first 183 dates in the training set ($\approx 40\%$)
- next 122 dates in the validation set ($\approx 30\%$)
- remaining 132 dates in the test set (August 2008 to December 2018) ($\approx 30\%$)

Data Management

Unbalanced longitudinal dataset

- Firms enter and exit public markets $\Rightarrow$ turnover for datasets with a long sample horizon
- Worse for bonds: bonds have a varying maturity between 3 to 7 years

Possible Workaround: Expand the sample with unknown values to create a balanced longitudinal dataset

- Increases the hardware requirements and the need for parallelization due to memory constraints $\Rightarrow$ noisier gradients

Our Workaround: Augment data with a double index: asset x time

- Processes the dates sequentially and calculates loss for data batch
- Able to estimate all models on single A100 GPU with 40GB memory
- Allows to use the batch size as a varying hyperparameter for training the neural network architectures

Model Architectures

Use machine learning to model portfolio weights:

$$\omega(z_{t,\cdot})$$

Four candidate architectures:

1. **Model 1 - Shallow FFN:** hidden units = [64,32,16,8]
2. **Model 2 - Deep FFN:** hidden units = [64,64,32,32,16,16,8,8]
3. **Model 3 - Deeper FFN:** hidden units = [64,64,64,32,32,32,16,16,16,8,8,8,4]
4. **Model 4 - Elastic Net:** hidden units = [1].

Other Optimization Considerations

- **Activation Function** We chose SiLU as activation function.
- **Optimizer** We use the Adam optimizer (Kingma and Ba (2015)), with learning rates of 0.0001, 0.001, and 0.005.
- **Batch size** We use batch sizes of 256, 512, 1024, and 2048 as hyperparameter.
- L^1**-regularization and** L^2**-regularization** with ["none", 0.001, 0.01, 0.1] as penalty hyperparameter.
- **Early stopping** with a stop after 8 periods of non-improvement in the validation phase.
- **Dropout** with probability of 0.01.

⇒ Model and hyper-parameter space create a large candidate space of **768 possible combinations.**

⇒ Search over in the validation phase to decide on the best architecture and hyperparameters.

Leverage Constraint

SDF portfolio weights and leverage:

- L^1-norm of SDF portfolio weights measure amount of leverage of strategy
- Without constraints, SDF portfolio assume large positive and negative values, implying large amount of leverage.

$\Rightarrow$ We impose additional leverage constraints via regularization.

Leverage constraint trade-offs:

- The L^1-norm measure the amount of leverage of a portfolio directly. L^1-norm typically implies much more concentrated portfolios, with some larger weights and lots of smaller weights.
- L^2-norm favors more diversified portfolio. However, re-balancing across positions imposes larger trading cost.

Leverage Constraint

The following inequality holds:

$$\|\omega\|_1 \leq \sqrt{n}\|\omega\|_2.$$

Hence, a bound of C on the L^2-norm can translate to a varying amount of leverage of the SDF on the date level, depending on how many dates are present in the batch.

Leverage constraint choice:

- Use an upper bound for the L^2-norm of $C_{port} = 1$ for bond portfolios. This leads to a maximum leverage in the range of 0.8 to 2.5.
- For individual bonds, we chose a bound that allows for maximum leverage between 2 and 7.

Model evaluation

Portfolio Performance:

- Sharpe ratio
- Sortino ratio
- Calmar ratio
- Cumulative annual growth rate ("CAGR")
- Maximum drawdown
- Mean-, and max-leverage

Excess Returns:

Evaluate portfolio returns against commonly used benchmarks in the asset pricing literature for equities:

- S&P 500 index,
- Fama-French 3 factor model (Fama and French (1993)),
- Fama-French 5 factor model (Fama and French (2015)).

Model evaluation

Out-of-sample predictability measures:

- Adopting average predictability over the cross-section, or over time from Chen, Pelger, and Zhu (2023)
- **Explained variation** is the R^2 of a cross-sectional regression on the loadings; this is a time-series R^2:

$$EV = 1 - \frac{\frac{1}{T}\sum_{t=1}^{T} \text{ cross-sec-mean } \left(\epsilon_t^2\right)}{\frac{1}{T}\sum_{t=1}^{T} \text{ cross-sec-mean } \left(\left(R_t^e\right)^2\right)}. \quad (4)$$

- **Cross-sectional $\mathbf{R^2}$** is the average pricing error, divided by the average return; this is a cross-sectional R^2:

$$\text{XS-}R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N} \frac{T_i}{T} \left(\text{t-series-mean } \left(\epsilon_i\right)\right)^2}{\frac{1}{N}\sum_{i=1}^{N} \frac{T_i}{T} \left(\text{t-series-mean } \left(R_i\right)\right)^2}. \quad (5)$$