# Transaction Costs and the Stochastic Discount Factor*

Daniele Bianchi† Teng Jiao† Hao Ma†

**Abstract**

When investors face transaction costs, assets are priced by their covariance with both the state of the economy and trading frictions, fundamentally reshaping the stochastic discount factor (SDF). Using flexible non-linear SDF specifications, we incorporate transaction costs directly into estimation rather than as post-optimization adjustments, substantially outperforming frictionless models on a net-of-cost basis. This operates through endogenous reallocation of the tangency portfolio weights: transaction cost-aware SDFs are more diversified with lower turnover, de-emphasizing high-turnover characteristics while stable fundamental signals gain prominence. These effects persist across market regimes, model specifications, leverage constraints, and alternative cost definitions, establishing transaction costs as structural determinants of equilibrium asset prices, not merely implementation constraints.

**Keywords:** Asset Pricing, Transaction Costs, Deep Learning, Risk Premiums, Stock Characteristics.

**JEL Classification:** G11, G12, G14, C45, C58.

*This version: December 1, 2025. We are thankful to Markus Pelger, Guanhao (Gavin) Feng, Tom Schmitz, Fabio Trojani, and Yu Zheng for their helpful comments and suggestions. We are also grateful to the participants of the 3rd FinEML at Erasmus University Rotterdam and of seminars at Queen Mary, University of London, and the Free University of Bozen-Bolzano for their helpful comments and suggestions.

†School of Economics and Finance, Queen Mary University of London, UK. E-mail: d.bianchi@qmul.ac.uk.

†School of Economics and Finance, Queen Mary University of London, UK. E-mail: t.jiao@qmul.ac.uk.

†School of Economics and Finance, Queen Mary University of London, UK. E-mail: hao.ma@qmul.ac.uk.

# 1 Introduction

Despite decades of research identifying hundreds of return predictors, translating this knowledge into profitable investment strategies remains challenging due to implementation frictions. While sophisticated models may appear to generate substantial profits in theory, their performance often deteriorates dramatically when confronted with realistic economic restrictions (Avramov et al., 2023). This creates a fundamental tension between the theoretical promise of advanced asset pricing models and their practical implementability (Jensen et al., 2024).

This disconnect between theory and practice is particularly problematic for asset pricing, where the stochastic discount factor (SDF) is often operationalized via an affine transformation of the tangency portfolio. The assumption of frictionless trading creates a wedge between theoretical SDFs and their implementable counterparts (Korsaye et al., 2021). The result is that models may identify seemingly significant risk factors that fail in practice due to the very trading costs they ignore (Detzel et al., 2023).

We show that transaction costs should not be treated as post-optimization adjustments, but rather as structural determinants of the SDF itself. When investors face trading costs, only portfolios with sufficiently large risk-adjusted returns relative to their implementation costs can survive in equilibrium. This creates an endogenous reallocation that fundamentally reshapes the SDF portfolio, with high-turnover characteristics becoming penalized while more stable fundamentals gain prominence. Transaction costs thus alter the pricing kernel itself, which reflects not merely risk preferences, but the economic constraints imposed by costly arbitrage (Gârleanu and Pedersen, 2013).

Building on the adversarial neural network framework of Chen et al. (2024), we incorporate stock-specific transaction costs that vary across stocks and time into SDF estimation. This approach employs two competing networks linked by no-arbitrage conditions – an SDF network that explicitly accounts for trading frictions and an "instrument" network that identifies the

most challenging test portfolios to be priced by the SDF.[1] Both networks process a comprehensive set of 73 stock characteristics and a handful of latent macroeconomic factors extracted from 122 economic indicators using long short-term memory (LSTM) networks (Hochreiter and Schmidhuber, 1997).

Using data on nearly 10,000 stocks from 1972 to 2023, we find that transaction cost-aware SDFs deliver economically and statistically significant improvements over frictionless specifications, especially when models are evaluated net of trading costs. Our approach generates Sharpe ratios 40% to 77% higher than conventional models – a difference that translates to significant economic value for investors. These improvements are significant at a conventional 5% confidence threshold and reflect not just better risk-adjusted returns, but fundamentally different portfolio composition: transaction cost-aware SDFs naturally produce more implementable strategies with lower turnover and more diversified positions.

The performance advantage extends beyond portfolio implementation to asset pricing fundamentals. Our approach substantially improves explained variation by 56% (from 6.3% to 9.8%) and cross-sectional $R^2$ by 54% (from 14.5% to 22.3%) for our baseline specification. These improvements come from systematic reductions in exposure to high-turnover signals (e.g., short-term reversal, return skewness) and increased weight on stable fundamental measures (e.g., return on equity, operating leverage).

Critically, transaction cost-aware SDFs retain their performance edge across all market regimes – during both tight and loose financial conditions, high and low volatility periods, and when excluding microcap stocks. This suggests that the outperformance over the testing sample arises from improved handling of trading frictions, not merely from better treatment of the smallest, most illiquid securities. The results are also robust across different leverage

[1] This adversarial approach addresses two fundamental challenges in asset pricing tests. First, it solves the joint-hypothesis problem by simultaneously learning both the SDF and the optimal test assets, rather than relying on pre-specified portfolios that may be "too easy to price" (Lewellen et al., 2010). Second, it creates an evolutionary "arms race": the adversarial network continuously constructs portfolios that maximize pricing errors (searching for the hardest counterexamples to the current SDF), while the SDF network adapts to price these challenging portfolios correctly. This dynamic process drives convergence toward an SDF that remains robust across different market conditions. This iterative process, consistent with the intuition in Hansen and Jagannathan (1997), drives convergence toward an SDF that remains robust to different market conditions.

constraints, testing periods, alternative sets of firm characteristics, and varying definitions of transaction costs (such as size-related costs as in Brandt et al., 2009).

It is essential to recognize that our objective is not to compare the asset pricing performance of complex, non-linear machine learning models against simpler linear specifications (e.g., Kelly et al., 2019; Kozak et al., 2020; Lettau and Pelger, 2020). Rather, our objective is to show the economic significance of transaction costs as a crucial input in the SDF estimation process. To establish this point more firmly, we leverage the flexibility of neural networks to impose a strict linear parameterization of the pricing kernel, while conditioning the SDF estimates on the same set of firm characteristics and macroeconomic indicators used in our baseline non-linear estimates.

The results reveal that, under the linear specification, both the explained variation and the cross-sectional $R^2$ decrease by approximately 50% compared to the non-linear SDF specification. Nevertheless, a transaction cost-aware SDF still consistently outperforms its frictionless counterpart after accounting for trading costs, reinforcing the importance of trading frictions regardless of model complexity.

Our work contributes to two important strands of literature. First, we advance the growing literature exploring machine learning methods in cross-sectional asset pricing (Gu et al., 2020; Bryzgalova et al., 2023; Fan et al., 2022; Chen et al., 2024; Jensen et al., 2024; Feng et al., 2024; Shen and Xiu, 2025; Kelly et al., 2025). While we build upon the adversarial neural network framework of Chen et al. (2024), our approach differs fundamentally by treating transaction costs as an integral component of the tangency portfolio and thus of the SDF itself.

More importantly, our analysis contributes to the substantial literature on the asset pricing implications of trading frictions (Vayanos, 1998; Balduzzi and Lynch, 1999; Novy-Marx and Velikov, 2016; DeMiguel et al., 2020; Korsaye et al., 2021; Barroso and Detzel, 2021; Avramov et al., 2023). We embed the key insight from this literature directly into the structural estimation of the SDF. Rather than treating trading frictions as a post-hoc implementation detail, we make them an integral component of the pricing kernel. This approach provides a struc-

tural solution to the fundamental tension between machine learning's theoretical promise and practical implementability, demonstrating that transaction costs materially affect the fundamental properties of the SDF, determining which sophisticated pricing relationships can exist in equilibrium.

## 2 Asset Pricing With Transaction Costs

We assume the existence of a strictly positive SDF $M_{t+1}$ such that for any asset return in excess of the risk-free rate $r^e_{i,t+1} = r_{i,t+1} - r^f_{t+1}$, the no-arbitrage condition $\mathbb{E}_t[M_{t+1} r^e_{i,t+1}] = 0$ holds. Without loss of generality, any valid minimum variance SDF can be represented as an affine function of a portfolio of tradable assets (e.g., Hansen and Jagannathan, 1991; Back, 2010),

$$M_{t+1} = 1 - \sum_{i=1}^{N} \omega_{i,t} r^e_{i,t+1} \tag{1}$$

where $\omega_{i,t}$ denotes the tangency portfolio weight on asset $i$ at time $t$. This formulation aligns with the fundamental principle that assets are priced based on their covariance with the state of the economy, as proxied by the returns of the tangency portfolio.

In frictionless markets, asset prices depend solely on the covariance with the gross tangency portfolio returns. However, constructing the tangency portfolio is arguably not free of costs. The core economic insight underlying our framework is that transaction costs fundamentally alter the pricing kernel by introducing a penalty on the extent of intertemporal portfolio rebalancing. The latter is directly linked to stock-specific trading costs. Consider the transaction cost-aware SDF:

$$M^{TC}_{t+1} = \underbrace{1 - \sum_{i=1}^{N} \omega_{i,t}\, r^e_{i,t+1}}_{M_{t+1}} + \underbrace{\sum_{i=1}^{N} c_{i,t}\, |\Delta\omega_{i,t}|}_{\text{TC component}} \tag{2}$$

where $\Delta\omega_{i,t} = \omega_{i,t} - \omega_{i,t-1}(1 + r_{i,t})$ represents the trading required in asset $i$, $c_{i,t} \in \mathbb{R}^+$ represents the proportional transaction cost, and $r_{i,t}$ is the raw return of asset $i$.[2] The component

[2]The term $\omega_{i,t-1}(1+r_{i,t})$ accounts for the natural evolution of position sizes due to price movements before

$\sum_{i=1}^{N} c_{i,t} |\Delta\omega_{i,t}|$ captures aggregate trading costs. The positive sign ensures that higher trading frictions $c_{i,t}$ increase the discount factor, reflecting the fact that costly rebalancing states require higher risk premiums to compensate investors for deteriorating investment opportunities.

It is important to highlight that our framework is agnostic about the definition of transaction costs. We treat $c_{i,t} \forall i, t$ as given parameters determined by microstructure factors outside the marginal investor's control (e.g., Gârleanu and Pedersen, 2013; Korsaye et al., 2021). Since these costs represent structural market features rather than choice variables, the specific functional form (e.g., linear vs. quadratic vs) is inconsequential for our main contribution, provided transaction costs are non-zero and affect the equilibrium tangency portfolio (DeMiguel et al., 2020).[3] Put differently, our contribution lies not in modeling transaction costs per se, but in showing how incorporating any realistic cost structure directly into the pricing kernel – rather than as post-optimization adjustments – fundamentally reshapes our understanding of the SDF properties.

**Asset pricing implications.** The transaction cost-aware SDF (henceforth defined TC-aware SDF) in Eq.(2) incorporates the notion that trading frictions are not only economic costs the marginal investor pays when constructing the optimal tangency portfolio, but they fundamentally alter expected returns. Assets are priced not just by their covariance with the state of the economy but also by their covariance with trading frictions, akin to a liquidity beta of the form, $\beta_i^{LIQ} = \frac{\text{Cov}(r^e_{i,t+1}, \sum_{j=1}^{N} c_{j,t}|\Delta\omega_{j,t}|)}{\text{Var}(\sum_{j=1}^{N} c_{j,t}|\Delta\omega_{j,t}|)}$.

Assets that require more trading (higher $|\Delta\omega_{i,t}|$) when transaction costs ($c_{i,t}$) are high will command a higher expected return as compensation for the systematic risk of deteriorating investment opportunities. This transforms transaction costs from a purely implementation consideration into a fundamental asset pricing component which relates to the market microstructure (Acharya and Pedersen, 2005).

any trading occurs (e.g., DeMiguel et al., 2009, 2020).

[3]Bid–ask spreads do adjust with market-wide order flow and volatility (Glosten and Milgrom, 1985; Kyle, 1985), but they are set by liquidity providers and, from the point of view of the representative investor are taken as given.

Thus, accounting for transaction costs generates a feedback mechanism between portfolio optimization and asset pricing. Transaction costs alter optimal portfolio weights by penalizing high-turnover characteristics, which changes the SDF through both components in Equation (2). The modified SDF $M_{t+1}^{TC}$ then affects equilibrium expected returns, creating endogenous selection toward implementable factors and away from costly trading strategies (Jensen et al., 2024).

## 2.1 Model Structure and Estimation

Our TC-aware SDF is estimated based on a min-max objective function of the form,

$$\min_{\omega} \max_{g} \frac{1}{N} \sum_{j=1}^{N} \left\| \mathbb{E}\left[ \left( 1 - \sum_{i=1}^{N} \omega_{i,t}\, r_{i,t+1}^{e} + \sum_{i=1}^{N} c_{i,t}\, |\Delta\omega_{i,t}| \right) \cdot r_{j,t+1}^{e} \cdot g_{j,t} \right] \right\|^2 \tag{3}$$

where $g_{j,t} \in \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^d$ are the instruments used to identify the $d$ moment conditions. The weights $\omega_{i,t}$ and the instrument $g_{j,t}$ play distinct economic roles in capturing the feedback mechanism between transaction costs and asset prices: the weights act as the "generator" that constructs the SDF explicitly penalizing portfolio turnover, while the instruments act as the "discriminator" that creates the most challenging test assets for the TC-aware SDF to price.[4]

We assume that the weights on the tangency portfolio are an unknown function of a wide range of $p$ stock-level characteristics $I_{i,t}$ and $q$ macroeconomic indicators $I_t$,

$$\omega_{i,t} = \omega(I_t, I_{i,t}; \vartheta_\omega) = \kappa \cdot (\sigma\left(f(I_t, I_{i,t}; \vartheta_\omega)\right) - 0.5) \tag{4}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function. Re-centering the output by subtracting 0.5 allows the portfolio to take both long and short positions, whereas scaling by factor $\kappa$ limits weights to the range [-$\kappa$/2, $\kappa$/2]. In the empirical application, we examine $\kappa \in \{0.05, 0.10, 0.20\}$,

[4]Choosing the conditioning function $g_{j,t}$ is akin to finding the optimal instruments in a conventional Generalized Method of Moments (GMM) estimation (e.g., Hansen and Jagannathan, 1991). In this respect, the maximization step in Eq. (3) aims to expose the weaknesses of the SDF by identifying the most challenging test portfolios—those with the largest pricing errors—whereas the minimization step aims to learn the portfolio weights that minimize pricing errors on these test portfolios.

corresponding to maximum position sizes – or leverage constraints – of $\pm 2.5\%$, $\pm 5\%$, and $\pm 10\%$, respectively.

For a neural network with $L$ hidden layers, the function $f(I_t, I_{i,t}; \vartheta_\omega)$ can be expressed as:

$$f(I_t, I_{i,t}; \vartheta_\omega) = W_\omega^{(L+1)} h_\omega^{(L)} + b_\omega^{(L+1)} \tag{5}$$

where the hidden layer activations $h_\omega^{(l)}$ for $l \in \{1, 2, ..., L\}$ are computed recursively as:

$$h_\omega^{(1)} = \text{ReLU}(W_\omega^{(1)}[I_t, I_{i,t}] + b_\omega^{(1)}) \tag{6}$$

$$h_\omega^{(l)} = \text{ReLU}(W_\omega^{(l)} h_\omega^{(l-1)} + b_\omega^{(l)}) \quad \text{for} \quad l \in \{2, 3, ..., L\} \tag{7}$$

Here, $\vartheta_\omega = \{W_\omega^{(1)}, W_\omega^{(2)}, ..., W_\omega^{(L+1)}, b_\omega^{(1)}, b_\omega^{(2)}, ..., b_\omega^{(L+1)}\}$ contains all trainable parameters and bias terms across the $L+1$ layers. This multi-layer architecture allows the model to capture the complex feedback mechanisms between transaction costs and asset prices discussed earlier.

Similarly, we parametrize the instrument function $g_{j,t} = g(I_t, I_{j,t}; \vartheta_g)$ with a multi-layer architecture where $\vartheta_g$ contains all trainable parameters and bias terms for the moment functions. The output dimension of $g(I_t, I_{j,t}; \vartheta_g)$ corresponds to the $d$ unconditional moment conditions used in our SDF estimation.

**Generative Adversarial Network (GAN).** We follow Chen et al. (2024) and adopt a Generative Adversarial Network (GAN) approach in which two neural networks compete under the min–max loss function in Eq.(2). For a random subset (mini-batch) $\mathcal{B}$ of size $B$ from our data, the empirical loss function is defined as,

$$\mathcal{L}_{\mathcal{B}}^{\text{GAN-TC}}(\vartheta_\omega, \vartheta_g) = \frac{1}{B} \sum_{(t,i)\in\mathcal{B}} \frac{1}{N} \sum_{j=1}^{N} \left\| M_{t+1}^{TC}(\vartheta_\omega) \cdot r_{j,t+1}^{e} \cdot g(I_t, I_{j,t}; \vartheta_g) \right\|^2 \tag{8}$$

Here the term $M_{t+1}^{TC}(\vartheta_\omega)$ is the TC-aware SDF defined as:

$$M_{t+1}^{TC}(\vartheta_\omega) = 1 - \sum_{i=1}^{N} \omega(I_t, I_{i,t}; \vartheta_\omega)\, r_{i,t+1}^{e} + \sum_{i=1}^{N} c_{i,t}\, |\Delta\omega_{i,t}(\vartheta_\omega)| \tag{9}$$

with the cost penalty term $\Delta\omega_{i,t}(\vartheta_\omega) = \omega(I_t, I_{i,t}; \vartheta_\omega) - \omega(I_{t-1}, I_{i,t-1}; \vartheta_\omega)(1 + r_{i,t})$.

The adversarial estimation consists of two main steps that capture the constant feedback between the tangency portfolio and asset prices. First, we perform multiple updates to the moment network parameters $\vartheta_g$ by maximizing the empirical loss function $\mathcal{L}_{\mathcal{B}}^{\text{GAN-TC}}(\vartheta_\omega^*, \vartheta_g)$ while fixing $\vartheta_\omega^*$. This allows the instruments in the moment conditions to effectively identify assets whose prices are most difficult to price. Then, we update the SDF parameters $\vartheta_\omega$ by minimizing the empirical loss function $\mathcal{L}_{\mathcal{B}}^{\text{GAN-TC}}(\vartheta_\omega, \vartheta_g^*)$ while fixing the previously obtained $\vartheta_g^*$. This step adjusts portfolio weights to account for transaction costs in response to identified pricing errors. Algorithm 1 summarizes the iterative estimation procedure.

In our baseline implementation, we set $n_{\text{mom}} = 64$ and $n_{\text{sdf}} = 256$ epochs for the instrument and SDF networks, respectively. All gradient-based updates employ Adam with a learning rate of $\eta_g = \eta_\omega = 0.001$. To initialize the parameters $\vartheta_\omega^{\text{init}}$ and $\vartheta_g^{\text{init}}$ that serve as inputs to Algorithm 1, we implement an upstream pretraining procedure that provides identical initial conditions for both frictionless and TC-aware SDFs. Appendix A.1 provides a more detailed description of this pretraining routine.

**Processing macroeconomic indicators.** To incorporate time-varying macroeconomic conditions into our SDF estimation, we process a large panel of macroeconomic indicators using recurrent neural networks that capture time-series temporal dependencies. Let $I_s \in \mathbb{R}^q$ denote the $q$-dimensional vector of macroeconomic indicators observed at time $s$. We follow Chen et al. (2024) and process the cumulative sequence $(I_1, \ldots, I_s, \ldots, I_T)^\top$ with *two* parallel LSTM networks that are tailored to the distinct roles within our adversarial framework.

The first network, LSTM$^\omega$, has a compact hidden dimension of 4 and supplies the tangency portfolio weight generator with a state vector $h_t^\omega \in \mathbb{R}^4$. This parsimonious representation

**Algorithm 1:** TC-Aware SDF Estimation with GAN

**Input :** Pretrained parameters $\vartheta_\omega^{\text{init}}$, $\vartheta_g^{\text{init}}$; hyperparameters: total cycles $K$, moment network epochs $n_{\text{disc}}$, SDF network epochs $n_{\text{sdf}}$, learning rates $\eta_g, \eta_\omega$

**Initialize:** $\vartheta_\omega \leftarrow \vartheta_\omega^{\text{init}}$, $\vartheta_g \leftarrow \vartheta_g^{\text{init}}$;

**for** *cycle* $k = 1$ **to** $K$ **do**

  **if** $k > 1$ **then**

    `/* Step 1: Update moment network (discriminator) */`

    **for** *epoch* $= 1$ **to** $n_{mom}$ **do**

      **foreach** *mini-batch* $\mathcal{B}$ **do**

        **Fix** $\vartheta_\omega$ at current value;

        Compute loss $\mathcal{L}_{\mathcal{B}}^{\text{GAN-TC}}(\vartheta_\omega, \vartheta_g)$;

        Update $\vartheta_g \leftarrow \vartheta_g + \eta_g \nabla_{\vartheta_g} \mathcal{L}_{\mathcal{B}}^{\text{GAN-TC}}$ (gradient ascent);

  `/* Step 2: Update SDF network (generator) */`

  **for** *epoch* $= 1$ **to** $n_{sdf}$ **do**

    **foreach** *mini-batch* $\mathcal{B}$ **do**

      **Fix** $\vartheta_g$ at current value;

      Compute loss $\mathcal{L}_{\mathcal{B}}^{\text{GAN-TC}}(\vartheta_\omega, \vartheta_g)$;

      Update $\vartheta_\omega \leftarrow \vartheta_\omega - \eta_\omega \nabla_{\vartheta_\omega} \mathcal{L}_{\mathcal{B}}^{\text{GAN-TC}}$ (gradient descent);

  Evaluate model on validation set;

  Store $\vartheta_\omega^{(k)} \leftarrow \vartheta_\omega$;

**Output:** Final parameters $\widehat{\vartheta}_\omega = \vartheta_\omega^{(K)}$ and corresponding $M_{t+1}^{\text{TC}}(\widehat{\vartheta}_\omega)$

focuses on extracting the essential macroeconomic factors most relevant for the SDF construction, consistent with asset pricing models that attribute time-varying expected returns to a small set of macroeconomic state variables (e.g., Ludvigson and Ng, 2009).

The second network, LSTM$^g$, generates a richer 32-dimensional representation ($h_t^g$) that enables the model to identify more subtle conditional pricing errors across different macroeconomic regimes. This higher-dimensional encoding allows for greater flexibility in detecting state-dependent mispricing, consistent with the GAN discriminator's role in identifying challenging test assets. Both LSTMs share the same functional form. Specifically, for $x \in \{\omega, g\}$,

the LSTM update equations are:

$$i_s^x = \sigma\big(W_i^x I_s + U_i^x h_{s-1}^x + b_i^x\big), \qquad f_s^x = \sigma\big(W_f^x I_s + U_f^x h_{s-1}^x + b_f^x\big), \tag{10}$$

$$o_s^x = \sigma\big(W_o^x I_s + U_o^x h_{s-1}^x + b_o^x\big), \qquad \tilde{c}_s^x = \tanh\big(W_c^x I_s + U_c^x h_{s-1}^x + b_c^x\big), \tag{11}$$

$$c_s^x = f_s^x \odot c_{s-1}^x \; + \; i_s^x \odot \tilde{c}_s^x, \qquad h_s^x = o_s^x \odot \tanh(c_s^x), \qquad s = 1, \ldots, t, \tag{12}$$

where $i_s^x$, $f_s^x$, and $o_s^x$ represent the input, forget, and output gates; $\tilde{c}_s^x$ is the candidate cell state (the new information to be added to the memory); $c_s^x$ is the resulting cell state after combining past memory and the candidate input; and $h_s^x$ is the hidden state that serves as our macroeconomic encoding.

The resulting state vectors $h_t^\omega$ and $h_t^g$ are integrated into Eq.(4) and $g_{j,t} = g(I_t, I_{j,t}; \vartheta_g)$ as,

$$\omega_{i,t} = \kappa\big(\sigma(f(h_t^\omega, I_{i,t}; \vartheta_\omega)) - 0.5\big) \quad \text{and} \quad g_{j,t} = g(h_t^g, I_{j,t}; \vartheta_g) \tag{13}$$

The parameter sets $\vartheta^\omega_{\text{LSTM}}$ and $\vartheta^g_{\text{LSTM}}$ are estimated *jointly* with the SDF and moment conditions parameters $\vartheta_\omega$ and $\vartheta_g$ through the same min-max objective in Equation (3), ensuring that the macroeconomic state encoding is optimized for the asset pricing task.

It is worth noticing that LSTM networks offer several advantages over traditional principal component analysis (PCA) or linear factor models for encoding macroeconomic information. While PCA can efficiently reduce dimensionality, it assumes linear relationships and static patterns in the data, limiting its ability to capture the complex, time-varying relationships between macroeconomic conditions and asset prices. In contrast, LSTMs explicitly model sequential dependencies through their gating mechanisms (Hochreiter and Schmidhuber, 1997). This capability is particularly valuable for identifying regime shifts, structural breaks, and remembering or forgetting information as economic conditions evolve selectively. Importantly, the LSTM's ability to extract rich temporal features enhances both the tangency portfolio's ability to harness time-varying risk premia and the moment conditions' capacity to identify pricing errors that vary with macroeconomic conditions.

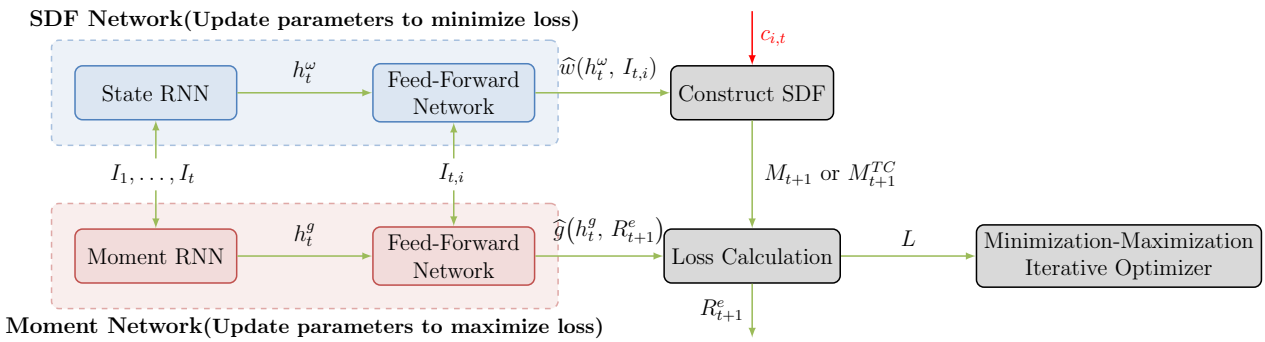



Figure 1: **GAN architecture for TC-aware SDF estimation**. This figure illustrates the adversarial training framework used to estimate transaction cost-aware stochastic discount factors.

Figure 1 visualizes our generative adversarial network architecture for estimating transaction cost-aware SDFs. The SDF network (top, blue) takes macroeconomic indicators $I_1, \ldots, I_t$ as input to an LSTM, which generates hidden states $h_t^\omega$ that are combined with firm-level characteristics $I_{t,i}$ to produce portfolio weights $\widehat{w}(h_t, I_{t,i})$. Transaction costs $c_{i,t}$ directly enter the SDF network construction. The instrument network (bottom, red) conditions on a similar LSTM to generate the macroeconomic states $h_t^g$, which are then added to individual stock characteristics.

**Hyperparameters tuning.** We note that our primary research objective is to investigate the impact of transaction costs on the SDF construction and performance. We do not address computational considerations of different cross-validation techniques in this paper. For this reason, we opt to adopt the optimal network configurations and hyperparameter choices from Chen et al. (2024). They perform an extensive search for optimal hyperparameters across 384 distinct network configurations, fitting the GAN model for each combination and selecting the top four hyperparameter configurations based on their performance on the validation dataset. By holding the network architecture and hyperparameters constant, we can compare the frictionless SDF and its TC-aware counterpart on a level playing field. Table A.1 in Appendix A.1 summarizes the hyperparameters and network architecture choices employed in the main empirical analysis.

## 2.2 Alternative Model Structures

In addition to the GAN structure, we also consider a simplified Feed-Forward Neural (FFN) network specification that eliminates the adversarial component when estimating the SDF. Unlike the GAN approach, the FFN model directly minimizes the aggregate pricing error,

$$\min_{\vartheta_\omega} \frac{1}{N}\sum_{j=1}^{N}\left\|\mathbb{E}\Big[\big(1-\sum_{i=1}^{N}\omega_{i,t}\,r^e_{i,t+1}+\sum_{i=1}^{N}c_{i,t}\,|\Delta\omega_{i,t}|\big)\cdot r^e_{j,t+1}\Big]\right\|^2 \tag{14}$$

This simplification removes the adaptive moment conditions generated by the adversarial network, effectively setting $g(I_t, I_{j,t}; \vartheta_g) \equiv 1$ for all assets and time periods. For a mini-batch $\mathcal{B}$ of size $B$, the empirical loss becomes:

$$\mathcal{L}_{\mathcal{B}}^{\text{FFN-TC}}(\vartheta_\omega) = \frac{1}{B}\sum_{(t,i)\in\mathcal{B}}\frac{1}{N}\sum_{j=1}^{N}\left\|M^{TC}_{t+1}(\vartheta_\omega)\cdot r^e_{j,t+1}\right\|^2 \tag{15}$$

where the TC-aware SDF $M^{TC}_{t+1}(\vartheta_\omega)$ maintains the same form as in Equation (9). Algorithm 2 summarizes the FFN training procedure, highlighting its streamlined nature compared to the GAN counterpart in Algorithm 1. The absence of the adversarial component significantly simplifies the optimization process, requiring only a single network update per epoch.

**Algorithm 2:** TC-Aware SDF Estimation with FFN

**Input :** Pretrained parameters $\vartheta_\omega^{\text{init}}$; hyperparameters: total epochs $E$, learning rate $\eta_\omega$
**Initialize:** $\vartheta_\omega \leftarrow \vartheta_\omega^{\text{init}}$;
**for** *epoch* $e = 1$ **to** $E$ **do**
  **foreach** *mini-batch* $\mathcal{B}$ **do**
    Compute loss $\mathcal{L}_{\mathcal{B}}^{\text{FFN-TC}}(\vartheta_\omega)$ according to Eq. (15);
    Update $\vartheta_\omega \leftarrow \vartheta_\omega - \eta_\omega \nabla_{\vartheta_\omega}\mathcal{L}_{\mathcal{B}}^{\text{FFN-TC}}$;
  Evaluate model on validation set;
  Store $\vartheta_\omega^{(e)} \leftarrow \vartheta_\omega$;
**Output:** Final parameters $\widehat{\vartheta}_\omega = \vartheta_\omega^{(E)}$ and corresponding $M^{\text{TC}}_{t+1}(\widehat{\vartheta}_\omega)$

The FFN maintains the same tangency portfolio network structure as the GAN model, including sigmoid activation for output weights and ReLU activations in the hidden layers,

so that differences in performance can be attributed specifically to the absence of adversarial training rather than architectural variations.

# 3 Data

We obtain stock characteristics from the Open Asset Pricing database.[5] The initial dataset includes 161 characteristics which have been identified by Chen and Zimmermann (2021) as "clear predictors". Throughout our analysis, we maintain the original sign convention for each stock characteristic.

We balance three key aspects of the dataset composition: the cross-sectional breadth (number of characteristics, $p$), the temporal depth (sample length in months, $T$), and the coverage of firms ($N_t$). Since $N_t$ varies over time due to firm entries and exits, we first identify a plausible trade-off between $p$ and $T$. For each candidate start date, we compute the number of characteristics with non-missing values. The left panel of Figure 2 reports the number of characteristic-month $(p, T)$ pair observations for different choices of starting date, holding the end date fixed at November 2023. The objective is to identify sample periods that yield the largest possible total number of characteristic-month observations $p \times T$, and among these, prioritize configurations that offer a richer set of characteristics $p$.

**Sample selection.** We select January 1972 (red vertical line in the left panel) as the sample start date, since it provides the largest set of firm characteristics continuously available, i.e., $p = 129$, for a moderately long time-series, i.e., $T = 624$ months. This sample period strikes a balance between cross-sectional breadth and time-series length.

We exclude characteristics available for fewer than 800 stocks in any month or those with maximum cross-sectional coverage below 2,500 stocks. This filtering reduces our characteristic set from 129 to 73.[6] Table B in Appendix B provides a description of each characteristic.

[5] We use version 1.4.1, released in October 2024.

[6] This is nearly twice the size of Chen et al. (2024) (46 characteristics), more than double that of Kelly et al. (2019) (36 characteristics), and on par with Freyberger et al. (2020) (62 characteristics).

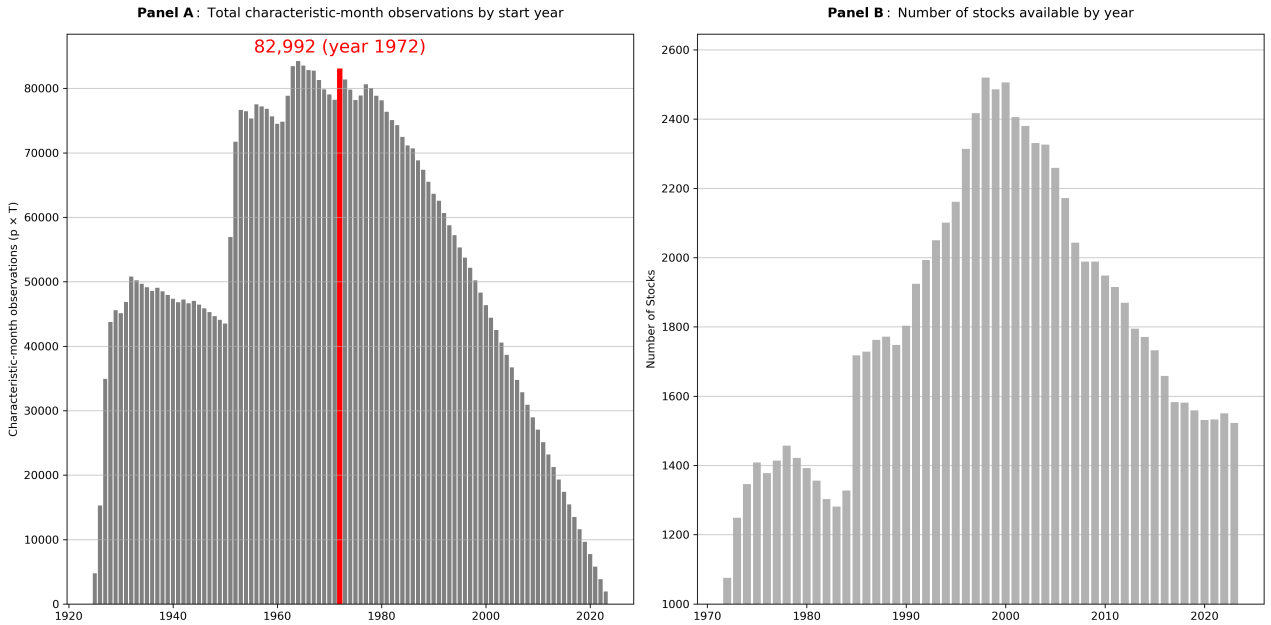



Figure 2: **Sample selection**. The left panel reports the number of characteristic-month $(p, T)$ pair observations for different choices of starting date, holding the end date fixed at November 2023. The right panel reports the cross-section of stocks available over the sample from January 1972 to November 2023.

To construct our final dataset, we merge the 73 characteristics with monthly stock return data from CRSP. We are limited to stocks for which all selected characteristics are available in any given month (e.g., Kelly et al., 2019; Chen et al., 2024). The sample includes all common stocks (share codes 10 and 11) listed on NYSE, AMEX, or NASDAQ from January 1972 to November 2023. We incorporate delisting returns to address survivorship bias as in Shumway and Warther (1999). The excess return for each stock is computed by subtracting the one-month Treasury bill yield from the realized return. The right panel of Figure 2 shows the number of stocks available each year based on the selected characteristics. The resulting dataset comprises an unbalanced panel of 9,599 unique stocks, with a monthly cross-section ranging from 949 to 2,567 stocks (average: 1,805 stocks).[7]

Each month, firm-level characteristics are transformed into cross-sectional percentile ranks and subsequently linearly mapped into the range $[-0.5, 0.5]$ (e.g., Kelly et al., 2019; Kozak et al., 2020). This rank normalization procedure standardizes the scale of input variables while

[7]This is on par with (Chen et al., 2024, see Section 4.1.), which includes around 10,000 unique stocks with a monthly cross-section ranging from 400 to 2,800.

preserving their ordinal structure. To mitigate look-ahead bias, all firm-level characteristics and macroeconomic variables are lagged by one month prior to their use in the estimation.

It is important to acknowledge that excluding stocks with missing characteristics may impact the performance of the tangency portfolio. However, this methodological choice likely leads to conservative estimates rather than overstatement. As shown by Freyberger et al. (2024) and Bryzgalova et al. (2025), missingness in financial characteristics is not at random, but predominantly affects stocks with small market capitalization. This attribute is strongly associated with lower liquidity and thus higher transaction costs as discussed in the next paragraph (Avramov et al., 2023). Since our TC-aware SDF explicitly penalizes assets with higher trading frictions (see Eq.(2)), excluding these stocks systematically removes from our sample the very securities that would exhibit the largest effects on the SDF. Therefore, our estimates represent, if anything, a lower bound on the asset pricing implications of transaction costs.

**Transaction costs.** Following Bessembinder and Venkataraman (2010), we use the half bid-ask spread as our proxy for stock-level transaction costs. For each asset $i$ at time $t$, we compute the spread using the Corwin and Schultz (2012) approximation and scale it by the stock price. The left panel of Figure 3 reveals a significant time-series variation of half bid-ask spreads (in %) over the sample period, with pronounced spikes during periods of market stress such as the early 1990s recession, the 2000-2002 tech bubble collapse, the 2008 financial crisis, and the 2020 COVID-19 market disruption. The substantial gap between the 90th percentile and the median indicates considerable right-skewness in the cross-sectional distribution, with the lowest-liquid stocks facing dramatically larger trading costs.

The right panel documents the inverse relationship between market capitalization and transaction costs. Stocks in the lowest market capitalization quartile exhibit transaction costs approximately 3-4 times larger than those in the highest quartile, with this size-based differential becoming particularly pronounced in the early part of the sample. This pattern aligns with the intuition in Hasbrouck (2009), who posit that transaction costs decrease systematically with firm size – a relationship that has significant implications for the SDF estimation.

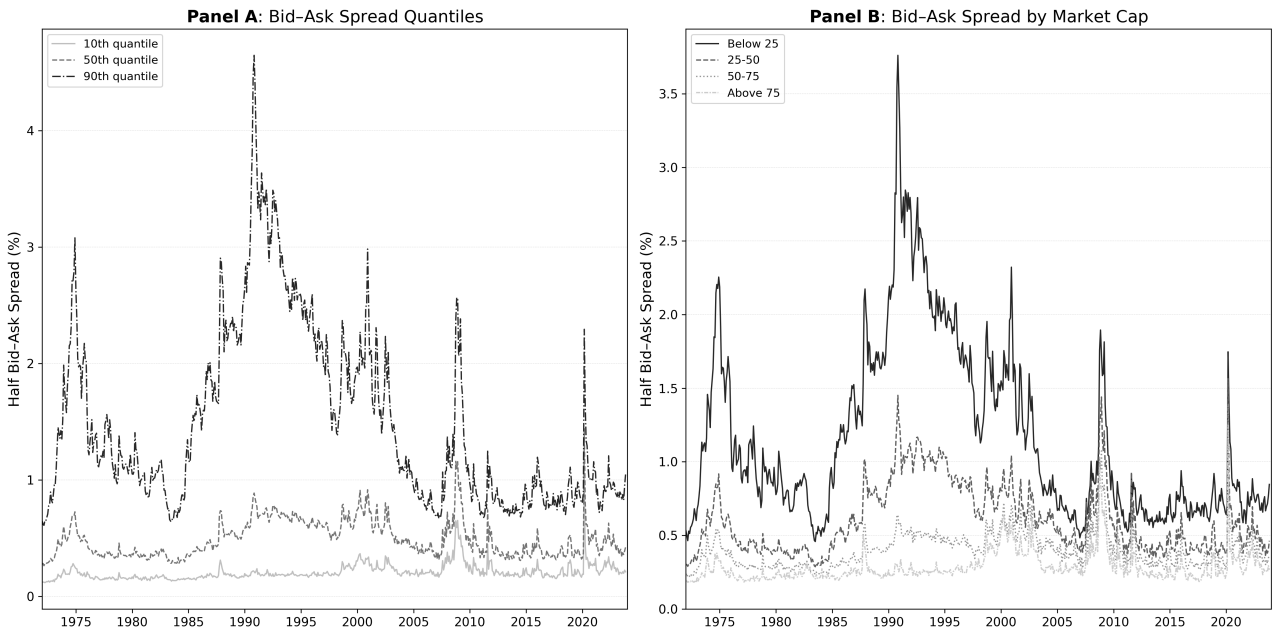



Figure 3: **Sample variation of bid-ask spreads**. The left panel reports the 10th, 50th, and 90th percentiles of the cross-sectional distribution of effective bid-ask spreads (in %) over the out-of-sample period. The right panel reports the average bid-ask spread (in %) for stocks sorted by market capitalization. The full sample period spans from January 1972 to November 2023.

**Macroeconomic variables.** We incorporate 122 macroeconomic variables from the FRED-MD database as documented by McCracken and Ng (2016). These series capture a broad spectrum of economic conditions, including real activity, labor markets, inflation, interest rates, and financial indicators. All series are transformed using the recommended procedures to ensure stationarity. Table B in Appendix B provides a detailed summary of each macroeconomic indicator.

# 4 Empirical Results for U.S. Equities

Our main sample period is from January 1972 to November 2023, totaling 52 years. As customary in machine learning applications, we divide the data into 20 years of training (January 1972 - December 1991), 5 years of validation (January 1992 - December 1996), and 27 years of testing sample (January 1997 - November 2023).

## 4.1 SDF Portfolio Composition

To investigate how transaction costs reshape the SDF portfolio, we first compare the weights $\omega_{i,t}$ of the TC-aware vs frictionless SDF. Figure 4 shows ratios of key portfolio statistics, where values above 1.0 indicate that frictionless portfolios exhibit higher values than TC-aware portfolios. We consider the estimates from both the GAN and the FFN architectures. For brevity, the results focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. Appendix D.1 reports the results for the alternative constraints of $\pm 10\%$ and $\pm 2.5\%$.

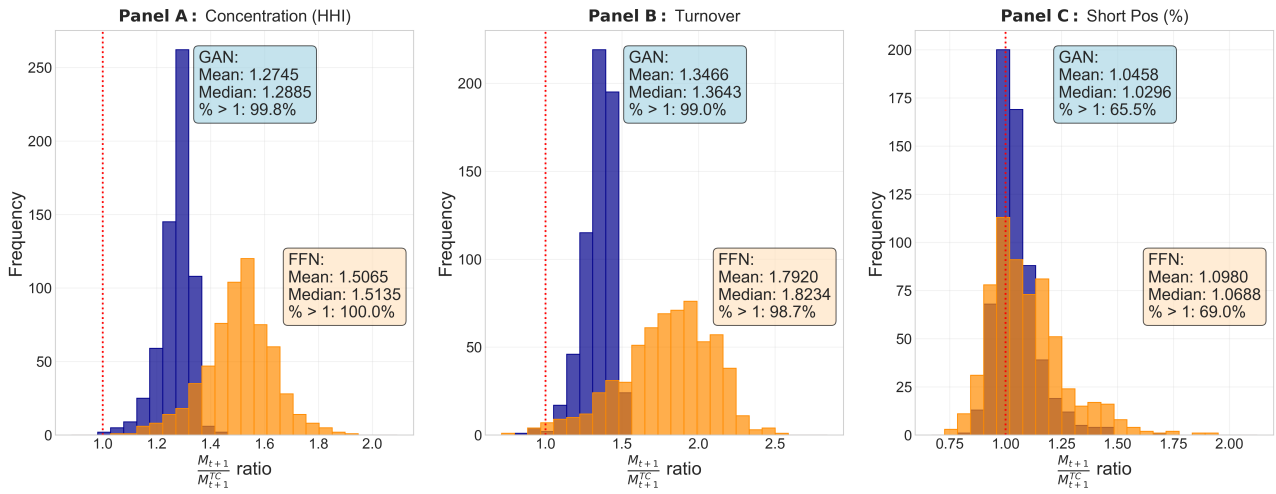



Figure 4: **SDF portfolio composition**. This figure reports the ratio between the frictionless SDF ($M_{t+1}$) and the TC-aware SDF ($M^{TC}_{t+1}$) of three main descriptive statistics. The left panel reports the ratio of the HHI indexes, the middle panel the ratio of the aggregate turnover, and the right panel the ratio of the fraction of short sales in the tangency portfolio. We report the results for both the FFN (orange bars) and the GAN (blue bars). The results are for the out-of-sample period from January 1997 to November 2023.

The results provide a first insight into the effect of transaction costs on the SDF portfolio composition. First, transaction costs promote diversification. The Herfindahl-Hirschman Indexes (HHI) ratios of 1.27 (GAN) and 1.51 (FFN) indicate that frictionless portfolios are substantially more concentrated than their TC-aware counterparts.[8] DeMiguel et al. (2020) provides some earlier intuition for this results: broader diversification creates natural hedging opportunities where position adjustments in different securities offset each other, reducing aggregate rebalancing costs. The larger magnitude of the ratio in FFN architectures suggests

[8]The HHI is calculated as $\text{HHI}_t = \sum_{i=1}^{N_t} \widehat{\omega}_{it}^2$, where $\widehat{\omega}_{it}$ represents the portfolio weight of asset $i$ at time $t$.

that adversarial training provides more nuanced effects on diversification.

Second, transaction costs reduce turnover.[9] Turnover ratios of 1.35 (GAN) and 1.79 (FFN) suggest that TC-aware tangency portfolios are rebalanced much less compared to the frictionless tangency implementation. This reflects the implicit incentive of the cost penalty to rebalance only when the expected benefits exceed the implementation costs, thereby discouraging unnecessary trading. Interestingly, transaction costs have minimal impact on short-selling. Short-selling ratios near 1.0 (1.05 for GAN and 1.10 for FFN) indicate that both approaches utilize short-selling to a similar degree. This suggests that the benefits of short positions for hedging generally outweigh their additional costs.

The patterns in Figure 4 reflect transaction costs operating at both intensive and extensive margins: discouraging excessive trading in individual securities while encouraging broader diversification to minimize implementation costs. The result is SDF portfolios that are more diversified, stable, and implementable than their frictionless counterparts.

Figure D.1 in Appendix D.1 shows the robustness of these findings across different leverage constraints. Under tighter constraints ($\pm 2.5\%$), diversification effects are strongest, with HHI ratios of 1.37 (GAN) and 1.69 (FFN). Under milder constraints ($\pm 10\%$), the turnover reduction intensifies as GAN turnover ratios increase from 1.25 to 1.44, indicating that frictionless portfolios trade even more excessively when given greater flexibility.

The difference between tighter and milder portfolio leverage reveals an important economic mechanism: transaction cost awareness provides the greatest value precisely when it is most needed. Under tight leverage constraints, portfolio restrictions naturally limit concentration, so transaction costs primarily improve diversification. Under milder constraints, investment restrictions are less binding, so transaction costs serve as an endogenous disciplinary mechanism that prevents excessive trading when investors have maximum flexibility to engage in costly rebalancing.

[9] Aggregate turnover at time $t$ is calculated as $TO_t = \sum_{i=1}^{N_t} |\Delta\widehat{\omega}_{i,t}|$ with $\Delta\widehat{\omega}_{i,t} = \widehat{\omega}_{i,t} - \widehat{\omega}_{i,t-1}(1 + r_{i,t})$ representing the rebalancing in asset $i$ from time $t-1$ to time $t$.

**SDF characteristics.** We now examine how transaction cost awareness translates into fundamental properties of the SDF portfolios. Specifically, we analyze the tangency portfolio characteristics calculated as:

$$\widehat{x}_t^j = \left| \frac{1}{N_t} \sum_{i=1}^{N_t} \widehat{\omega}_{i,t} I_{i,t}^j \right|, \qquad j = 1, \ldots, k \tag{16}$$

where $N_t$ denotes the number of stocks in the portfolio at time $t$, $\widehat{\omega}_{i,t} = \omega(I_t, I_{i,t}; \widehat{\vartheta}_\omega)$ is the estimated weight assigned to stock $i$, and $I_{i,t}^j$ is the value of characteristic $j$ for stock $i$. We take absolute values to focus on the magnitude of the portfolio characteristic exposure regardless of its direction. This analysis can be interpreted as a simple, non-parametric surrogate model that provides an intuitive interpretation of how the SDF portfolio weights translate into tangible characteristics (e.g., Bianchi and Venturi, 2024).

Figure 5 reports the difference $\Delta\widehat{x}^j = \text{med}(\widehat{x}_t^j) - \text{med}(\widehat{x}_t^{TC,j})$ where $\text{med}(\widehat{x}_t^j)$ ($\text{med}(\widehat{x}_t^{TC,j})$) is the time-series median of the frictionless (TC-aware) SDF portfolio characteristic. A positive value of $\Delta\widehat{x}^j$ means that the frictionless SDF portfolio is more exposed than its TC-aware counterpart to a given characteristic $j$ over the testing period. For brevity, we focus on the baseline leverage constraints of $\pm 5\%$. Appendix D.1 reports the difference in portfolio characteristics under the $\pm 10\%$ and $\pm 2.5\%$ position constraints.

The results reveal how transaction costs fundamentally reshape investment focus by reducing exposures to momentum-related characteristics such as short-term reversals and residual momentum, as well as skewness-related characteristics. Conversely, TC-aware SDFs exhibit higher exposures to more fundamental characteristics, including earnings quality, leverage, taxable income, and profitability. The magnitude of these effects varies meaningfully across model specifications. FFN models (left panel) show more pronounced shifts across most characteristics. In contrast, GAN models (right panel) demonstrate more moderate adjustments $\Delta\widehat{x}^j$, suggesting that the adversarial estimate enables more nuanced price impacts, reflecting the SDF composition in Figure 4.

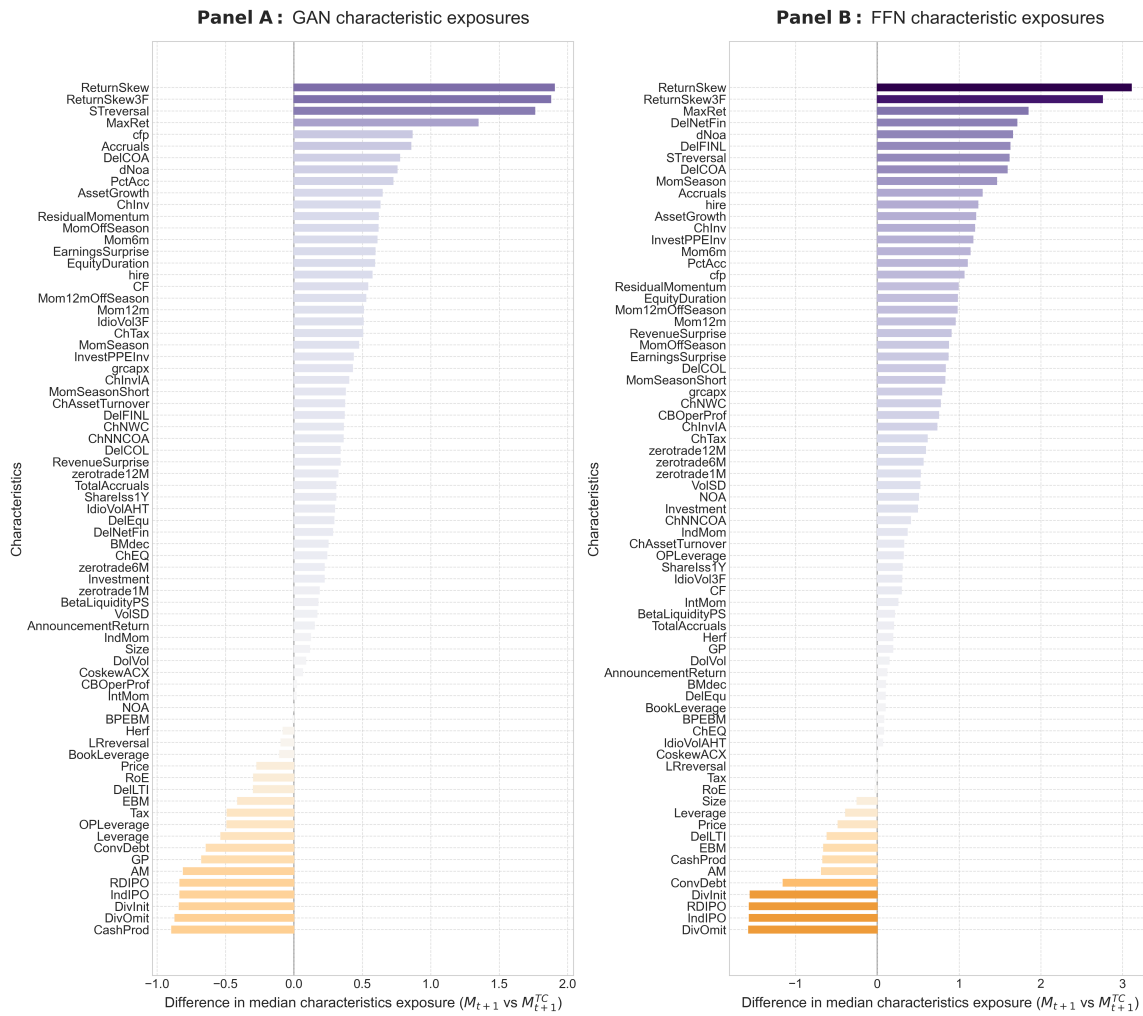



Figure 5: **SDF portfolio characteristics**. This figure reports the difference in the time-series median SDF portfolio characteristics between frictionless and TC-aware SDFs. The left panel presents the results for the GAN estimates, while the right panel displays the results for the FFN estimates. The out-of-sample period is from January 1997 to November 2023.

Figure D.2 shows that under milder constraints (±10%), portfolio characteristic exposures are substantially larger, as frictionless models pursue aggressive, high-turnover strategies that are instead disciplined by transaction costs. Under tighter constraints (±2.5%), differences become smaller since investment restrictions naturally act as a portfolio regularization mechanism.

The results in Figure 5 suggest that the cost of frequent portfolio adjustments required to capture short-term price trends and lottery-like payoffs may imply a more decisive reallocation of the SDF composition when transaction costs are considered. To test this intuition more

directly, we next examine which specific characteristics are most affected by transaction costs.

Taking stock of the results in Figure 5, our hypothesis is that the high transaction costs associated with frequent portfolio adjustments required to capture high-turnover characteristics such as short-term price trends and lottery-like payoffs should make these characteristics less attractive, leading to reduced exposure in TC-aware SDFs, while more stable characteristics should gain prominence.

To test this, we measure the extent to which a given characteristic cross-sectional ranking changes over time. For each characteristic $j$, we compute the Average Rank Change (ARC):

$$\text{ARC}_j = \frac{1}{T-1} \sum_{t=2}^{T} \left[ \frac{1}{|S_t^{(j)} \cap S_{t-1}^{(j)}|} \sum_{i \in S_t^{(j)} \cap S_{t-1}^{(j)}} \left| k_{i,t}^{(j)} - k_{i,t-1}^{(j)} \right| \right] \tag{17}$$

where $S_t^{(j)}$ is the set of firms for which characteristic $j$ is available at time $t$, and $k_{i,t}^{(j)} \in [-0.5, 0.5]$ is the cross-sectionally rank-standardized value of characteristic $j$ for firm $i$ at time $t$. We restrict to firms present in both consecutive periods ($S_t^{(j)} \cap S_{t-1}^{(j)}$) to avoid bias from entries and exits. Higher ARC values indicate more volatile rankings and thus higher implementation costs.

Figure 6 shows the relationship between characteristic turnover ($ARC_j$, y-axis) and its differential $\Delta\widehat{x}^j$ (x-axis). We focus in the main text on the baseline constraints $\omega_{it} \in \pm 5\%$ and leave the results for the alternative constraints of $\pm 10\%$ and $\pm 2.5\%$ to Appendix D.1.

The results confirm our central hypothesis: there is a strong negative correlation between characteristic turnover and its TC-aware SDF exposure. High-turnover characteristics such as short-term reversal (`STreversal`), return skewness (`ReturnSkew`), and momentum signals (`Mom6m`) cluster in the upper-right quadrant. The economic mechanism behind these results is clear: characteristics that require frequent rebalancing become less relevant for asset pricing when implementation costs are factored in. Conversely, more stable fundamental characteristics like return-on-equity (`ROE`), book leverage (`BookLeverage`), and operating leverage

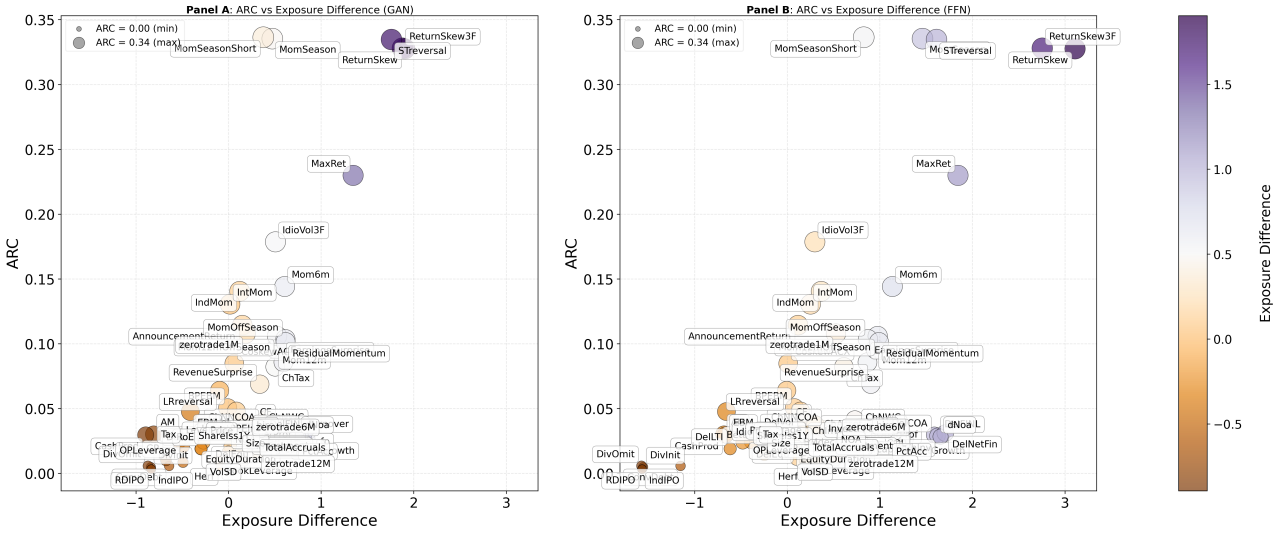



Figure 6: **Characteristics turnover vs SDF characteristic exposures**. This figure illustrates the relationship between characteristic turnover (ARC, y-axis) and the change in SDF exposure between frictionless and TC-aware SDFs (x-axis). The left panel presents the results for the FFN estimates, while the right panel displays the results for the GAN estimates. The results are for the out-of-sample period from January 1997 to November 2023.

(`OPLeverage`) gain relevance to explain the variation in stock returns. This pattern holds regardless of the leverage constraints (see Figure D.3) and applies to both GAN and FFN architectures.

Appendix D.1 reports additional results based on an alternative measure of characteristic turnover and the managed-portfolio implementation cost. Figure D.4 replaces the absolute rank difference with the Spearman rank correlation (SRC) between consecutive months as an alternative measure of characteristic stability. The results are qualitatively similar: high-turnover characteristics such as `ReturnSkew`, `STreversal`, and `Max` remain those most penalized in the pricing kernel, while more stable characteristics maintain their relative importance. Figure D.5 provides direct evidence by building on the definition of portfolio implementation costs from Section C. The results confirm that the more expensive it is to trade a given characteristic, the larger the effect of transaction costs on its SDF exposure.

**DeepSHAP attribution and characteristics importance.** To provide further evidence about which characteristics drive SDF portfolios with and without transaction costs, we employ

DeepLIFT (Shrikumar et al., 2017). This neural network attribution method decomposes each portfolio weight into contributions from individual characteristics. Unlike gradient-based approaches, which can suffer from saturation issues in deep networks, DeepLIFT compares each neuron's activation to a baseline reference and propagates attribution scores backward through the network.

The method builds on Shapley value decomposition from cooperative game theory (Shapley et al., 1953), which treats portfolio construction as a game where each characteristic contributes to the final weights. For characteristic $j$, the Shapley value measures its average marginal contribution across all possible combinations of other characteristics. However, exact computation requires $2^{73}$ model evaluations for our SDF implementations, a number computationally infeasible even with modern hardware.

We therefore implement DeepSHAP (Lundberg and Lee, 2017), which efficiently approximates expected Shapley values by averaging attributions across multiple baseline instances.[10] We define a baseline instance as the cross-sectional median of each characteristic, representing a market-neutral firm profile. For each stock, DeepSHAP quantifies the extent to which each characteristic's deviation from this median contributes to the final tangency portfolio weight. We use 300 randomly selected firms each month (approximately 10% of our universe) as the background empirical distribution to ensure robust attributions.

We apply DeepSHAP to the SDF network, which processes 77 inputs: four LSTM-generated macro factors plus 73 characteristics. While computing attributions for all inputs, we focus our analysis on stock characteristics. This approach provides time-varying feature importance scores that capture how the relative importance of different characteristics evolves with market conditions and transaction cost considerations. The four macro factors serve as contextual inputs during the attribution computation but are not included in our reported results. A detailed discussion on the technical implementation is provided in Appendix A.2.

[10]The latest version of the DeepLIFT algorithm (DeepSHAP) is available at `https://shap.readthedocs.io/en/latest/generated/shap.DeepExplainer.html#shap.DeepExplainer`. We thank Scott Lundberg for providing the Python package.

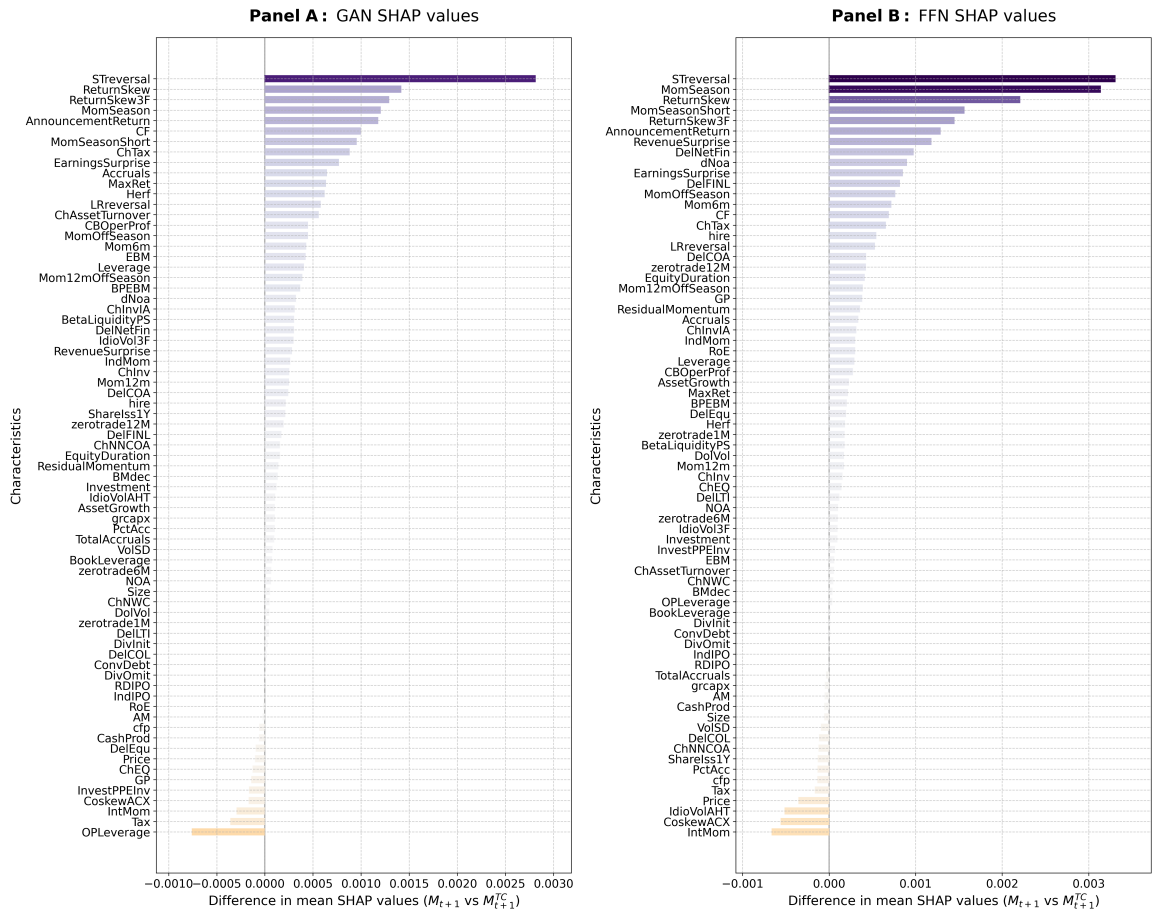



Figure 7: **Differentials in SDF characteristic SHAP attributions**. This figure reports the difference in the mean absolute SHAP values between the frictionless SDF and the TC-aware SDF for the GAN (left panel) and the FFN (right panel). Positive values indicate characteristics that receive less importance in transaction cost-aware models. The results pertain to ±5% leverage constraints over January 1997 to November 2023.

Figure 7 reports the difference in mean absolute SHAP values between frictionless and TC-aware SDFs. Positive values suggest that the frictionless SDF attributes greater relevance to characteristic $j$ compared to the TC-aware SDF, on average across the testing period.

Consistent with our earlier evidence, the SHAP values indicate a marked reduction in the importance attributed to high-turnover characteristics for TC-aware SDFs. Short-term reversal (`STreversal`) shows the largest reduction in both GAN and FFN estimates, followed by other high-turnover signals including momentum-based characteristics (`ReturnSkew`, `MomSeason`), earnings announcement effects (`AnnouncementReturn`, `EarningsSurprise`), and various liquidity-related measures.

A lower reliance on more volatile signals is compensated by marginally increasing the importance of more stable characteristics. Taxable income (`Tax`) and operating leverage (`OPLeverage`) show small increases in importance. These characteristics change slowly over time, making them more cost-effective sources of risk-adjusted returns.

Comparing the GAN and FFN panels reveals interesting architectural differences. The FFN model shows larger adjustments for individual characteristics but concentrates these changes in fewer variables. In contrast, the GAN model displays more distributed adjustments across a broader set of characteristics, potentially reflecting its adversarial training process that encourages robustness over the testing period.

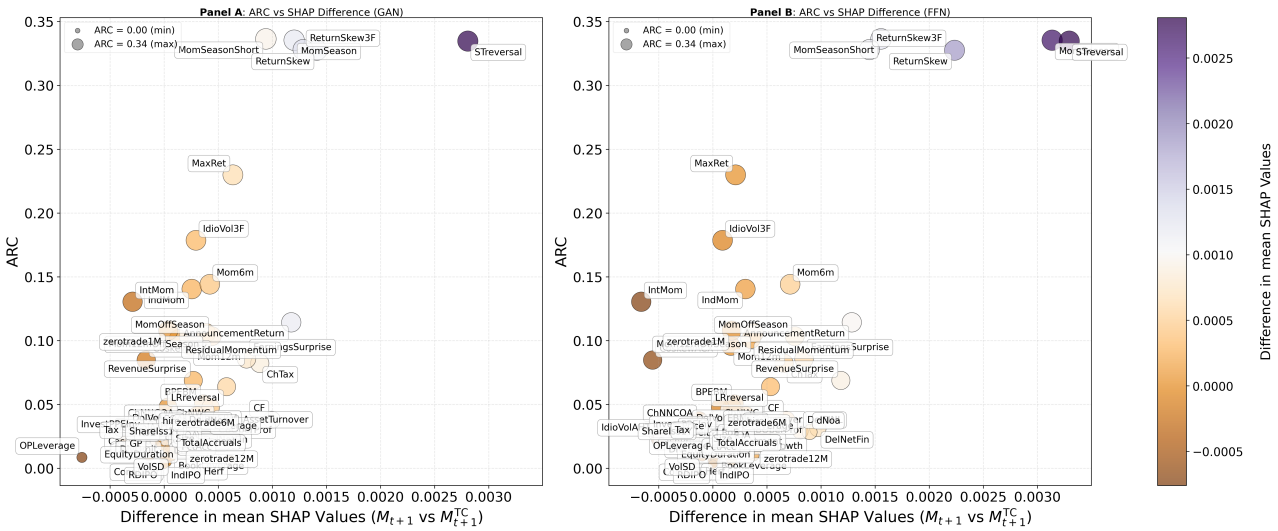



Figure 8: **Characteristics turnover vs Difference SHAP Attributions**. This figure shows the relationship between characteristic turnover (ARC, y-axis) and the difference in the mean SHAP attributions (x-axis). The left panel reports the results for the FFN estimates, whereas the right panel reports the results for the GAN estimates. The results are for the out-of-sample period from January 1997 to November 2023.

Similar to Figure 6, in Figure 8 we illustrate the relationship between characteristic turnover ($ARC_j$, y-axis) and the difference in mean SHAP attributions across characteristics (x-axis). We focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The results confirm a strong negative correlation between characteristic turnover and the SHAP importance attributions for the TC-aware SDF compared to its frictionless counterpart.

## 4.2 Asset Pricing Performance

**SDF portfolio returns.** Table 1 compares monthly Sharpe ratios across SDF specifications. We report both raw performance and returns net of transaction costs for GAN and FFN architectures under different leverage constraints. For completeness, the results span training, validation, and testing periods. We highlight in red the Sharpe ratio differentials – TC-aware minus frictionless SDF – which are statistically significant at a 5% confidence level according to a bootstrap test as in Ledoit and Wolf (2008). Appendix C.1 describes the bootstrap testing implementation details.

| | | | $\omega_{it} \in (-5\%, 5\%)$ | | | | $\omega_{it} \in (-10\%, 10\%)$ | | | | $\omega_{it} \in (-2.5\%, 2.5\%)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Raw returns | | Net returns | | Raw returns | | Net returns | | Raw returns | | Net returns | |
| | | Mkt | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| GAN | Train | 0.088 | 1.921 | 1.441 | 1.297 | 1.169 | 2.017 | 1.550 | 1.179 | 1.267 | 1.997 | 1.846 | 1.543 | 1.643 |
| | Valid | 0.345 | 1.178 | 1.043 | 0.201 | **0.353** | 1.309 | 1.145 | 0.082 | **0.417** | 1.032 | 0.909 | 0.219 | **0.400** |
| | Test | 0.143 | 0.348 | 0.317 | 0.064 | **0.113** | 0.352 | 0.334 | 0.047 | **0.143** | 0.286 | 0.231 | 0.063 | 0.084 |
| FFN | Train | 0.088 | 1.547 | 1.279 | 1.065 | 1.132 | 1.706 | 1.252 | 1.020 | 1.067 | 1.468 | 1.267 | 1.139 | 1.134 |
| | Valid | 0.345 | 1.216 | 0.933 | 0.216 | **0.521** | 1.299 | 1.032 | 0.091 | **0.492** | 0.999 | 0.702 | 0.256 | 0.352 |
| | Test | 0.143 | 0.315 | 0.253 | 0.051 | **0.118** | 0.321 | 0.256 | 0.032 | **0.107** | 0.317 | 0.266 | 0.087 | **0.137** |

Table 1: **Sharpe ratios**. This table reports the Sharpe ratios of tangency portfolios across training, validation, and testing periods for both GAN and feed-forward neural network (FFN) architectures. We present both raw performance and returns net of transaction costs for the standard SDF ($M_{t+1}$) and transaction-cost-aware SDF ($M^{TC}_{t+1}$). Results are shown for three different leverage constraints: ±5%, ±10%, and ±2.5%. We highlight in **red** the Sharpe ratio differentials (TC-aware minus frictionless SDF), which are statistically significant at a 5% confidence level. P-values are calculated based on the robust bootstrap test of Ledoit and Wolf (2008). Raw market Sharpe ratios (Mkt) are provided for reference.

The key finding is that while frictionless models appear on par on a gross basis, they become substantially inferior once transaction costs are considered, the scenario most relevant for implementation. This performance gap is statistically significant at a conventional 5% threshold. The TC-aware GAN achieves a test-period monthly Sharpe ratio of 0.113 vs 0.064 for the frictionless version, representing a 77% improvement under 5% leverage constraints. This advantage becomes even more pronounced under 10% leverage constraints, where TC-aware GAN delivers a monthly Sharpe ratio of 0.143 compared to 0.047 for the frictionless SDF, a 204% improvement.

The comparison between GAN and FFN models reveals interesting leverage-dependent patterns. At higher leverage (10%), GAN outperforms FFN among TC-aware models (0.143 vs 0.107 Sharpe ratio). However, at lower leverage (2.5%), FFN achieves superior results (0.137 vs 0.084). This suggests that GAN provides greater benefits when portfolio weights are less constrained, potentially because the adversarial training can better exploit the additional flexibility in position sizing.

Importantly, both GAN and FFN architectures exhibit clear overfitting, with Sharpe ratios declining substantially from training to testing periods. Despite this challenge, TC-aware SDFs maintain positive out-of-sample Sharpe ratios and consistently outperform their frictionless counterparts.

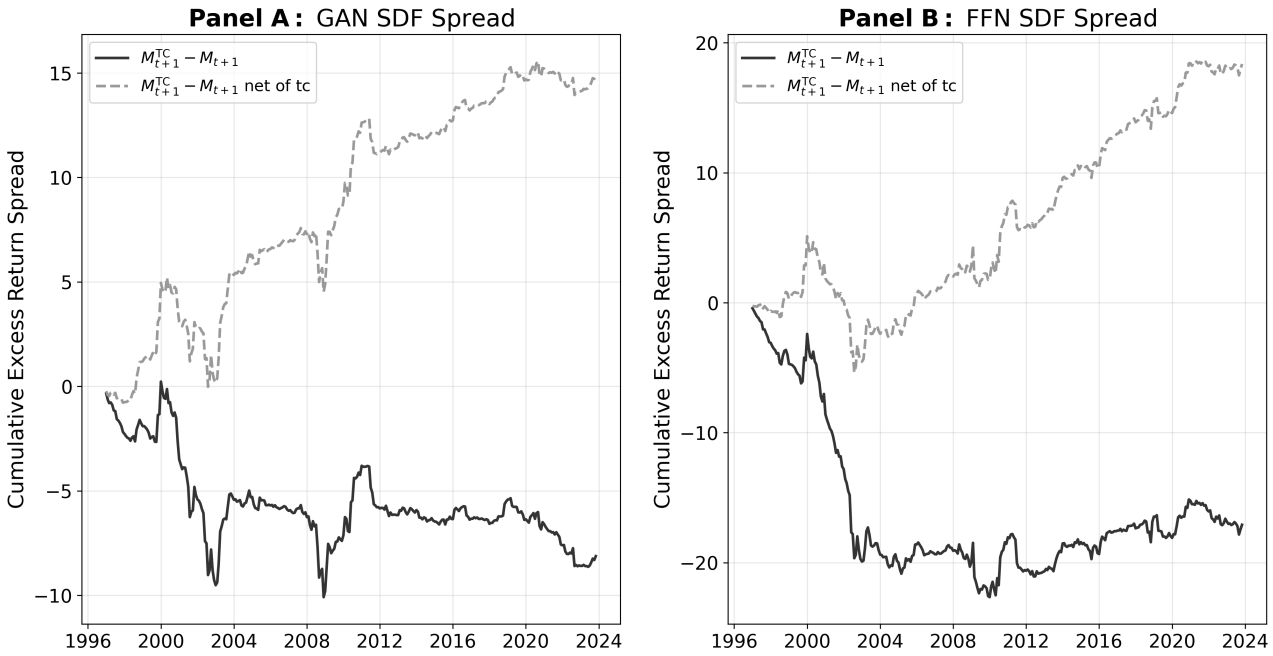



Figure 9: **Cumulative SDF returns spread**. This figure shows cumulative log return spread between $M^{TC}_{t+1}$ and $M_{t+1}$ over the out-of-sample period (January 1997 to November 2023). The left panel reports the results from the GAN and the right panel reports the results for the FFN architectures.

Figure 9 illustrates the cumulative return spread between the TC-aware and the frictionless SDFs over the testing period. TC-aware SDFs underperform on a gross basis (black solid lines), but this relationship reverses after transaction costs (dashed gray lines). For the GAN architecture (left panel), the net performance of $M^{TC}_{t+1}$ significantly exceeds that of $M_{t+1}$ over

the out-of-sample period. This performance gap is even more pronounced for the FFN model (right panel), where the return spread in favor of the frictionless SDF is almost 25% larger than that of GAN by the end of the testing period.

Appendix D.2 reports the results for the alternative leverage constraints. Figure D.6 shows that the performance gap in favour of the TC-aware SDF increases for less constrained allocations ($\pm 10\%$) compared to more restrictive constraints ($\pm 2.5\%$). This confirms the intuition that the transaction cost penalty becomes more binding when tangency portfolio allocations are left unconstrained a priori.

**Explaining the returns variation.** We consider two standard metrics that assess the SDF's ability to explain return patterns. Table 2 reports the explained variation (EV) and the cross-sectional $R^2$ (XS-$R^2$) across SDF specifications.

Explained variation measures the fraction of total return variation captured by the SDF, while cross-sectional $R^2$ measures the model's ability to explain differences in average returns across assets. Both metrics require estimates of each asset's exposure (beta) to the SDF factor. Following Chen et al. (2024), we obtain these betas by fitting an auxiliary neural network to predict the conditional covariance between each asset and the tangency portfolio return. Specifically, the SDF beta $\beta_{i,t}$ is proportional to $\mathbb{E}[F_{t+1} r^e_{i,t+1}]$ where $F_{t+1} = \sum_{i=1}^{N_t} \omega_{i,t} r^e_{i,t+1}$ is the SDF (or tangency) portfolio return. Appendix A.3 provides a detailed description of the technical details for the estimate of $\widehat{\beta}_{i,t}$.

The results reveal a consistent pattern: TC-aware specifications substantially outperform frictionless SDFs. For explained variation, the TC-aware GAN achieves 9.8% compared to 6.3% for the frictionless version (a 56% improvement) under a $\pm 5\%$ leverage constraint. This advantage becomes even more pronounced under tighter constraints ($\pm 2.5\%$, with the TC-aware GAN reaching 11.7% versus 6.5% for the frictionless GAN (an 80% improvement).

The cross-sectional $R^2$ results reinforce these findings, with TC-aware SDFs consistently achieving 20-40% higher values across all specifications. For instance, TC-aware GAN achieves

| | | Explained Variation | | | | XS-$R^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | GAN | | FFN | | GAN | | FFN | |
| | | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| $\omega_{it} \in (-5\%, 5\%)$ | Train | 0.124 | 0.190 | 0.151 | 0.204 | 0.134 | 0.240 | 0.150 | 0.233 |
| | Valid | 0.036 | 0.052 | 0.044 | 0.056 | 0.147 | 0.195 | 0.158 | 0.193 |
| | Test | 0.063 | 0.098 | 0.081 | 0.109 | 0.145 | 0.223 | 0.172 | 0.230 |
| $\omega_{it} \in (-10\%, 10\%)$ | Train | 0.133 | 0.186 | 0.129 | 0.205 | 0.124 | 0.202 | 0.156 | 0.255 |
| | Valid | 0.038 | 0.051 | 0.038 | 0.056 | 0.148 | 0.187 | 0.157 | 0.202 |
| | Test | 0.067 | 0.096 | 0.068 | 0.109 | 0.147 | 0.209 | 0.163 | 0.237 |
| $\omega_{it} \in (-2.5\%, 2.5\%)$ | Train | 0.128 | 0.215 | 0.154 | 0.203 | 0.122 | 0.261 | 0.139 | 0.241 |
| | Valid | 0.036 | 0.057 | 0.043 | 0.055 | 0.143 | 0.189 | 0.152 | 0.192 |
| | Test | 0.065 | 0.117 | 0.081 | 0.107 | 0.143 | 0.241 | 0.165 | 0.228 |

Table 2: **Explained variation and cross-sectional $R^2$.** This table reports the explained variation (EV) and cross-sectional $R^2$ for different SDF specifications across training, validation, and testing periods. Results are shown for both GAN and feed-forward neural network (FFN) architectures, comparing the standard SDF ($M_{t+1}$) and transaction-cost-aware SDF ($M^{TC}_{t+1}$) under ±5%, ±10%, and ±2.5% leverage constraints.

22.3% cross-sectional $R^2$ compared to 14.5% for the frictionless model under 5% leverage constraints. Comparing models, FFN architectures generally achieve slightly higher explained variation than GAN (10.9% versus 9.8% with 5% leverage constraints in the test period). However, these differences are modest.

**Cross-sectional predictability.** The no-arbitrage condition in Eq.(1) implies that stocks with higher SDF betas ($\beta_{i,t}$) should have higher expected returns, as investors require compensation for bearing systematic risk. We test this relationship by sorting stocks into decile portfolios based on their estimated risk loadings $\beta_{i,t}$.

Figure D.7 in Appendix D.2 demonstrates a clear separation between extreme deciles (Decile 1: lowest beta, Decile 10: highest beta). This confirms that stocks with higher systematic risk exposure indeed earn higher returns, validating the fundamental risk-return relationship underlying our SDF framework. More importantly, the figure reveals how transaction costs affect this relationship. For high-beta portfolios (Decile 10), performance remains comparable between frictionless and TC-aware SDFs (solid vs dashed lines). However, for low-beta portfolios (Decile 1), transaction cost-awareness delivers notably better performance than a frictionless

SDF. This pattern suggests that embedding transaction costs into the SDF estimate particularly penalizes investing in low-beta stocks. These are likely smaller, less liquid stocks that are expensive to trade.

This differential impact reinforces our main finding: transaction costs don't just reduce returns uniformly, but systematically affect cross-sectional predictability in ways that improve the implementability of the overall tangency portfolio.

Figure D.7 also shows that the beneficial effect of transaction costs on low-beta portfolios (the improvement in dashed vs. solid lines for Decile 1) is most pronounced under looser leverage constraints ($\pm 10\%$), where there is more scope for the frictionless model to make costly trading that the TC-aware model avoids. This reinforces our earlier finding that transaction costs provide the greatest economic value precisely when investors have maximum flexibility – acting as an endogenous disciplinary mechanism that prevents excessive trading in hard-to-implement portfolios.

Table 3 reports the time-series pricing performance of these beta-sorted portfolios. For each decile, we compute average excess returns and pricing errors (alphas) relative to the CAPM and Fama and French (2015) five-factor models.[11]

The results reveal that both frictionless and TC-aware SDFs successfully generate monotonic return spreads across beta-sorted portfolios. The long-short portfolio (Decile 10-1) delivers substantial risk-adjusted returns: 27.8% annually for the frictionless GAN and 22.0% for the TC-aware SDF. While the raw spread is smaller for TC-aware models, this reduction reflects their more implementable nature rather than inferior cross-sectional predictability.

The GRS test statistics (Gibbons et al., 1989), which jointly test whether all decile portfolios have zero pricing errors, strongly reject the null hypothesis of no joint pricing error for TC-aware models for both the GAN (t-stat=8.92 under the CAPM and t-stat=9.15 under the Fama and French (2015) five-factor model) and the FFN specification (t-stat=7.89 under the

[11] If our SDF successfully captures systematic risk, we should observe: (1) a monotonic relationship between estimated betas and average returns, and (2) small, statistically insignificant pricing errors when portfolios are regressed against standard factor models.

GAN SDF Estimates

| | Frictionless SDF | | | | | TC-aware SDF | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Average | Mkt-Rf | | FF5 | | Average | Mkt-Rf | | FF5 | |
| Decile | | $\alpha$ | t-stat | $\alpha$ | t-stat | | $\alpha$ | t-stat | $\alpha$ | t-stat |
| 1 | 0.004 | -0.028 | -0.470 | -0.006 | -0.102 | 0.021 | -0.006 | -0.101 | 0.012 | 0.185 |
| 2 | 0.072 | 0.052 | 1.034 | 0.068 | 1.283 | 0.089 | 0.071 | 1.524 | 0.088 | 1.747 |
| 3 | 0.091 | 0.077 | 1.798 | 0.092 | 2.007 | 0.091 | 0.076 | 1.826 | 0.089 | 2.031 |
| 4 | 0.105 | 0.093 | 2.303 | 0.107 | 2.466 | 0.117 | 0.105 | 2.569 | 0.120 | 2.768 |
| 5 | 0.122 | 0.112 | 2.945 | 0.128 | 3.168 | 0.119 | 0.107 | 2.687 | 0.121 | 2.793 |
| 6 | 0.132 | 0.124 | 3.229 | 0.138 | 3.334 | 0.134 | 0.124 | 3.123 | 0.138 | 3.230 |
| 7 | 0.141 | 0.132 | 3.302 | 0.145 | 3.390 | 0.149 | 0.139 | 3.445 | 0.151 | 3.486 |
| 8 | 0.163 | 0.152 | 3.619 | 0.168 | 3.672 | 0.142 | 0.130 | 3.024 | 0.145 | 3.206 |
| 9 | 0.188 | 0.177 | 3.849 | 0.189 | 3.872 | 0.197 | 0.184 | 4.053 | 0.203 | 4.257 |
| 10 | 0.281 | 0.264 | 4.784 | 0.290 | 5.015 | 0.240 | 0.226 | 4.251 | 0.252 | 4.491 |
| 10-1 | 0.278 | 0.291 | 8.417 | 0.297 | 8.319 | 0.220 | 0.232 | 7.693 | 0.241 | 7.389 |
| GRS test | | t-stat | p-value | t-stat | p-value | | t-stat | p-value | t-stat | p-value |
| | | 10.259 | 0.000 | 10.631 | 0.000 | | 8.924 | 0.000 | 9.150 | 0.000 |

FFN SDF Estimates

| | Frictionless SDF | | | | | TC-aware SDF | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Average | Mkt-Rf | | FF5 | | Average | Mkt-Rf | | FF5 | |
| Decile | | $\alpha$ | t-stat | $\alpha$ | t-stat | | $\alpha$ | t-stat | $\alpha$ | t-stat |
| 1 | 0.009 | -0.021 | -0.358 | -0.002 | -0.038 | 0.027 | -0.001 | -0.017 | 0.016 | 0.248 |
| 2 | 0.076 | 0.055 | 1.133 | 0.071 | 1.363 | 0.079 | 0.060 | 1.358 | 0.078 | 1.636 |
| 3 | 0.074 | 0.061 | 1.492 | 0.075 | 1.714 | 0.090 | 0.076 | 1.849 | 0.090 | 2.022 |
| 4 | 0.112 | 0.100 | 2.547 | 0.115 | 2.779 | 0.108 | 0.097 | 2.498 | 0.111 | 2.660 |
| 5 | 0.122 | 0.111 | 2.833 | 0.123 | 2.922 | 0.114 | 0.103 | 2.616 | 0.114 | 2.678 |
| 6 | 0.144 | 0.136 | 3.481 | 0.150 | 3.524 | 0.124 | 0.114 | 2.932 | 0.129 | 3.081 |
| 7 | 0.140 | 0.129 | 3.217 | 0.144 | 3.369 | 0.154 | 0.142 | 3.507 | 0.154 | 3.561 |
| 8 | 0.159 | 0.147 | 3.456 | 0.162 | 3.517 | 0.157 | 0.147 | 3.360 | 0.162 | 3.508 |
| 9 | 0.174 | 0.163 | 3.483 | 0.183 | 3.682 | 0.188 | 0.173 | 3.683 | 0.193 | 3.909 |
| 10 | 0.289 | 0.272 | 4.844 | 0.297 | 5.086 | 0.258 | 0.244 | 4.307 | 0.273 | 4.582 |
| 10-1 | 0.281 | 0.293 | 8.597 | 0.300 | 8.542 | 0.231 | 0.245 | 8.000 | 0.257 | 8.104 |
| GRS test | | t-stat | p-value | t-stat | p-value | | t-stat | p-value | t-stat | p-value |
| | | 12.580 | 0.000 | 12.057 | 0.000 | | 7.891 | 0.000 | 8.224 | 0.000 |

Table 3: **Pricing errors for beta-sorted portfolios**. This table reports average excess returns and time-series pricing errors for decile portfolios sorted on estimated SDF betas ($\hat{\beta}_{i,t}$) using GAN (Panel A) and FFN (Panel B) estimates. Pricing errors ($\alpha$) are computed using the CAPM and Fama-French 5-factor models. All returns are annualized. The GRS test examines the null hypothesis that all decile portfolios are correctly priced. The sample period is January 1997 to December 2023. Within each decile, stocks are equally weighted.

CAPM and 8.22 under the five-factor model).

## 4.3 Mean-Variance Efficiency and Observable Risk Factors

To further evaluate the asset pricing implications of transaction costs, we compare the mean-variance efficiency of our SDF portfolios with that of popular observable risk factors and characteristic-managed portfolios. Specifically, we follow Detzel et al. (2023) and test to what extent our SDF portfolio expands the mean-variance frontier both in gross terms and net of

transaction costs.

For a model with $K$ factors, we construct a $2K$-dimensional investment opportunity set consisting of both long and short positions for each factor. Net returns incorporate costs asymmetrically: $f_{kt}^{\text{net}} = f_{kt} - TC_{kt}^{f}$ for long positions and $f_{kt}^{S,\text{net}} = -f_{kt} - TC_{kt}^{f}$ for short positions, where $TC_{kt}^{f}$ represents the one-way transaction cost (Barroso and Detzel, 2021). A detailed description of how transaction costs are calculated for each anomaly is available in Appendix C. Each model is then compared based on the maximum squared Sharpe ratio attainable by solving:

$$SR^2(f^{\text{net}}) = \max_{\vartheta \in \mathbb{R}^{2K}} \left[ \frac{\vartheta' \mu}{\sqrt{\vartheta' \Sigma \vartheta}} \right]^2 \tag{18}$$

subject to $\vartheta_k \geq 0$ and $\sum_k \vartheta_k = 1$, where $\mu = E(f^{\text{net}})$ and $\Sigma = \text{cov}(f^{\text{net}})$. The non-negativity constraints ensure that we can only take long positions in the individual factors or portfolios, while the sum constraint normalizes the portfolio weights. Unlike the frictionless case with a closed-form solution $\vartheta^* \propto \Sigma^{-1}\mu$, the inclusion of transaction costs renders the optimization non-differentiable, which necessitates numerical optimization.

We report three performance metrics. The *in-sample* $SR^2$ represents the maximum achievable Sharpe ratio using the information within the training period (January 1972 - December 1991). This captures the in-sample mean-variance efficiency of an SDF portfolio. The *ex-ante* $SR^2$ uses the validation sample (January 1992 - December 1996) to estimate optimal weights $\vartheta^*$, which are then projected onto the testing period (January 1997 - November 2023). This provides the most economically relevant comparison as it mirrors more realistic implementation scenarios without look-ahead bias. Finally, the *ex-post* $SR^2$ re-optimizes $\vartheta^*$ using the information over the testing period, establishing the theoretical upper bound of achievable performance.

We first compare our frictionless and TC-aware SDF portfolios with the market, size, and value risk factors (Fama and French, 1993, FF3), plus the investment and profitability risk factors (Fama and French, 2015, FF5), and the cross-sectional momentum factor (Jegadeesh and Titman, 1993; Carhart, 1997, FF5 + MOM). With "gross" we indicate raw returns without

transaction costs, while with "net" we indicate returns after deducting transaction costs.

Table 4 reports the results for value-weighted and equal-weighted implementations of the risk factors. In-sample performance shows very high squared Sharpe ratios (up to 3.186), but the substantial gap with out-of-sample performance indicates considerable overfitting. Despite achieving lower gross performance, TC-aware SDFs consistently outperform their frictionless counterparts when evaluated after transaction costs.

The *ex-post* $SR^2$ provides an upper bound on achievable performance, yet even in these favorable conditions, TC-aware SDFs maintain their advantage over frictionless SDFs. For instance, the FF5 + $M_{t+1}^{TC}$ (GAN) achieves a monthly *ex-post* $SR^2$ of 0.199 compared to 0.255 for the frictionless version on a gross basis, but the ranking reverses when transaction costs are considered (0.057 vs 0.054 for net factor returns).

The systematic performance reversals when considering transaction costs – reversals that hold across all evaluation periods, architectures, and factor specifications – confirm that traditional gross-return comparisons may favor unimplementable high-turnover strategies (see Figure 5 to Figure 8).

**Pairwise comparison with characteristic-managed portfolios.** To further evaluate the extent to which our TC-aware SDF estimates encapsulate key asset pricing information, we implement a pairwise comparison with managed portfolios based on the set of tradable stock characteristics employed in our analysis. Specifically, we construct augmented mean-variance portfolios of the form $\{F_t, C_{j,t}\}$, where $F_t$ represents the SDF factor (either frictionless or TC-aware) and $C_{j,t}$ represents the double-sorted long-short portfolio based on size and characteristic $j$. Following Barroso and Detzel (2021), we evaluate each characteristic-managed portfolio using both gross and net returns. The incremental contribution of each characteristic is measured by:

$$\Delta SR_j^2 = SR^2(\{F_t, C_{j,t}\}) - SR^2(\{F_t\}) \tag{19}$$

| Model | Value-weight portfolios | | | | | | Equal-weight portfolios | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Gross | | | Net | | | Gross | | | Net | | |
| | IS $SR^2$ | OOS $SR^2$ | | IS $SR^2$ | OOS $SR^2$ | | IS $SR^2$ | OOS $SR^2$ | | IS $SR^2$ | OOS $SR^2$ | |
| | | ExAnte | ExPost | | ExAnte | ExPost | | ExAnte | ExPost | | ExAnte | ExPost |
| FF3 | 0.058 | 0.031 | 0.032 | 0.027 | 0.022 | 0.025 | 0.092 | 0.046 | 0.049 | 0.019 | 0.022 | 0.025 |
| FF3 + $M_{t+1}$ $(FFN)$ | 2.212 | 0.156 | 0.182 | 0.670 | 0.029 | 0.034 | 2.208 | 0.184 | 0.219 | 0.663 | 0.030 | 0.035 |
| FF3 + $M^{TC}_{t+1}$ $(FFN)$ | 1.557 | 0.081 | 0.098 | 0.991 | 0.034 | 0.038 | 1.562 | 0.108 | 0.129 | 0.978 | 0.035 | 0.039 |
| FF3 + $M_{t+1}$ $(GAN)$ | 2.887 | 0.193 | 0.212 | 0.836 | 0.033 | 0.038 | 2.899 | 0.212 | 0.246 | 0.833 | 0.033 | 0.039 |
| FF3 + $M^{TC}_{t+1}$ $(GAN)$ | 1.907 | 0.129 | 0.152 | 0.895 | 0.038 | 0.043 | 1.921 | 0.148 | 0.099 | 0.883 | 0.037 | 0.045 |
| FF5 | 0.355 | 0.089 | 0.127 | 0.105 | 0.027 | 0.047 | 0.518 | 0.191 | 0.216 | 0.026 | 0.030 | 0.041 |
| FF5 + $M_{t+1}$ $(FFN)$ | 2.392 | 0.189 | 0.229 | 0.718 | 0.032 | 0.051 | 2.376 | 0.276 | 0.319 | 0.663 | 0.036 | 0.047 |
| FF5 + $M^{TC}_{t+1}$ $(FFN)$ | 1.768 | 0.122 | 0.161 | 1.053 | 0.035 | 0.054 | 1.820 | 0.216 | 0.257 | 0.981 | 0.038 | 0.050 |
| FF5 + $M_{t+1}$ $(GAN)$ | 3.100 | 0.226 | 0.255 | 0.890 | 0.035 | 0.054 | 3.084 | 0.312 | 0.344 | 0.833 | 0.040 | 0.050 |
| FF5 + $M^{TC}_{t+1}$ $(GAN)$ | 2.142 | 0.161 | 0.199 | 0.962 | 0.037 | 0.057 | 2.156 | 0.253 | 0.297 | 0.885 | 0.043 | 0.055 |
| FF5 + MOM | 0.413 | 0.094 | 0.132 | 0.110 | 0.027 | 0.047 | 0.619 | 0.187 | 0.220 | 0.026 | 0.030 | 0.041 |
| FF5 + MOM + $M_{t+1}$ $(FFN)$ | 2.469 | 0.150 | 0.230 | 0.718 | 0.032 | 0.051 | 2.401 | 0.260 | 0.321 | 0.663 | 0.036 | 0.047 |
| FF5 + MOM + $M^{TC}_{t+1}$ $(FFN)$ | 1.784 | 0.114 | 0.161 | 1.053 | 0.035 | 0.054 | 1.820 | 0.214 | 0.257 | 0.981 | 0.038 | 0.050 |
| FF5 + MOM + $M_{t+1}$ $(GAN)$ | 3.186 | 0.174 | 0.256 | 0.890 | 0.035 | 0.054 | 3.112 | 0.298 | 0.346 | 0.833 | 0.040 | 0.050 |
| FF5 + MOM + $M^{TC}_{t+1}$ $(GAN)$ | 2.157 | 0.134 | 0.199 | 0.962 | 0.037 | 0.057 | 2.156 | 0.119 | 0.297 | 0.885 | 0.043 | 0.055 |

Table 4: **SDF comparison with observable risk factors**. This table reports in-sample, ex-ante out-of-sample, and ex-post out-of-sample maximum squared Sharpe ratios for conventional factor models (FF3, FF5, FF5+MOM) augmented with frictionless SDFs ($M_{t+1}$) and transaction cost-aware SDFs ($M^{TC}_{t+1}$) estimated using GAN and FFN architectures. Results are presented for both gross returns (ignoring transaction costs) and net returns (accounting for transaction costs) under value-weighted and equal-weighted portfolio construction. The sample period spans 1972-2023 with training period 1972-1996 and test period 1997-2023.

Figure 10 reports the incremental squared Sharpe ratio contributions for both value-weighted (Panel A) and equal-weighted (Panel B) portfolios. For brevity, we focus on the baseline leverage constraints (±5%).

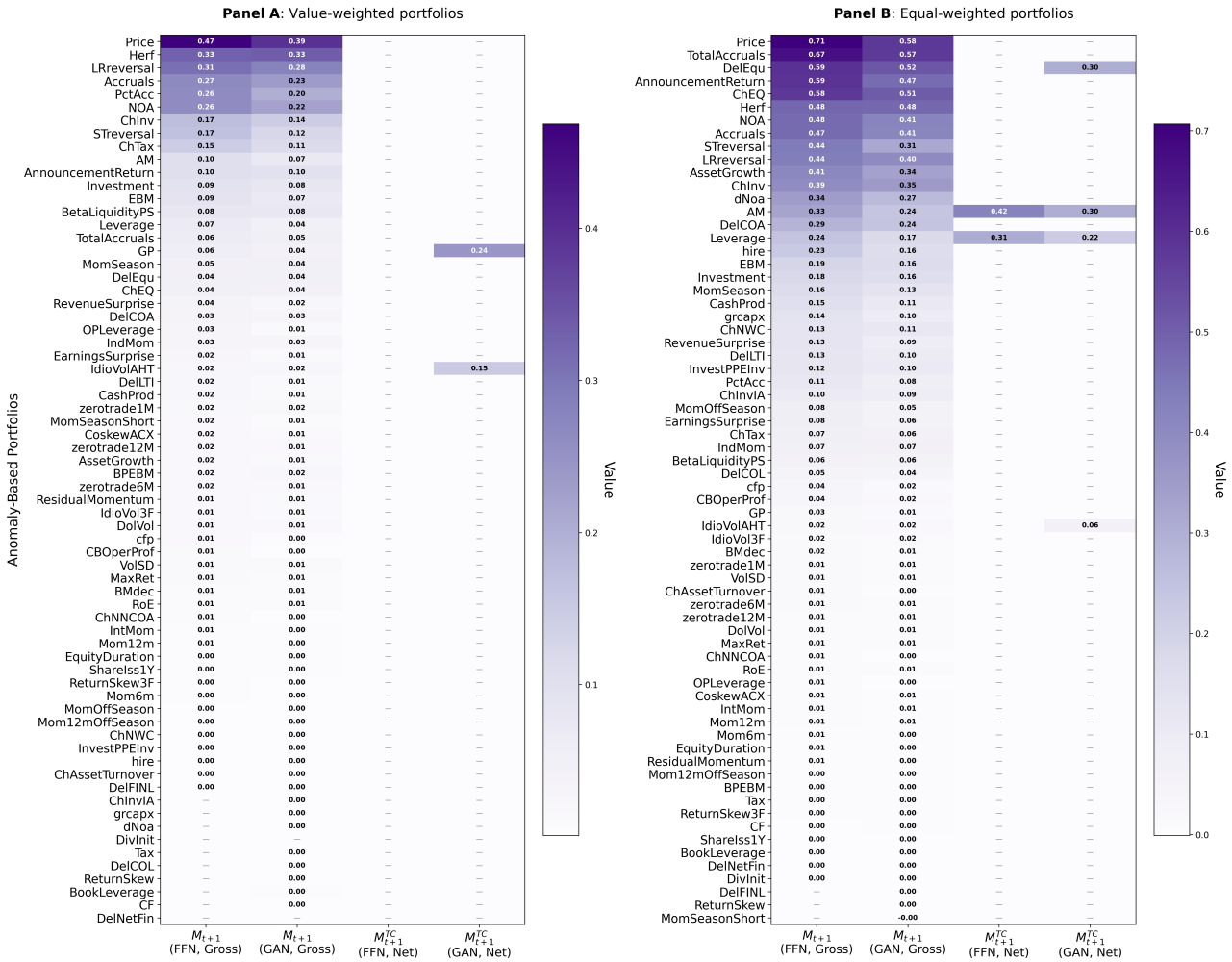



Figure 10: **SDF vis-a-vis portfolio anomalies**. This figure reports the incremental squared Sharpe ratio contributions ($\Delta SR^2_j$) when augmenting SDF factors with characteristic-based long-short portfolios. Panel A shows value-weighted portfolios, and Panel B shows equal-weighted portfolios. Results are presented for both gross and net returns over the out-of-sample period (January 1997 to November 2023).

The results reveal systematic differences between frictionless and TC-aware SDFs. Gross returns on high-turnover characteristics such as `Price`, `STReversal`, and `LtReversal` exhibit substantial incremental contributions to the frictionless SDF. However, these contributions diminish considerably when using our TC-aware SDFs and managed portfolio returns are net of transaction costs. Except for a few nuances, such as `AM` and `Leverage` for equal-weighted

portfolios, none of the net characteristic returns significantly expand the mean-variance frontier compared to our TC-aware SDF.

This analysis corroborates our core finding that incorporating transaction costs during SDF estimation, rather than treating them as post-optimization adjustments, results in pricing models that emphasize implementable risk factors while de-emphasizing those that appear valuable only under frictionless assumptions. Importantly, the systematic reduction in incremental contributions when moving from gross to net evaluation supports the evidence from Detzel et al. (2023) that transaction costs fundamentally alter which characteristics provide economically meaningful pricing information.

## 4.4 A Linear SDF Specification

To assess the contribution of non-linearity in the SDF specification, we examine a linear architecture that removes all nonlinear activations (ReLU functions) from the neural networks. This creates a purely linear mapping from characteristics to tangency portfolio weights, equivalent to traditional linear asset pricing models (e.g. Kozak et al., 2020), while still conditioning on our comprehensive set of stock characteristics and macroeconomic variables.

The tangency portfolio weight function maintains the same structure as Equation (4). However, the inner network $f^{\text{linear}}$ now implements a strictly linear transformation:

$$f^{\text{linear}}(I_t, I_{i,t}; \vartheta_\omega) = W_\omega^{(2)} h_\omega^{(1)} + b_\omega^{(2)} \qquad \text{with} \qquad h_\omega^{(1)} = W_\omega^{(1)}[I_t, I_{i,t}] + b_\omega^{(1)} \tag{20}$$

so that the hidden layer output $h_\omega^{(1)}$ is a pure linear transformation of the inputs. The linear specification transforms our neural networks into simple matrix multiplications: for a linear network with one hidden layer, this reduces to,

$$f^{\text{linear}}(I_t, I_{i,t}; \vartheta_\omega) = W_\omega^{\text{combined}}[I_t, I_{i,t}] + b_\omega^{\text{combined}} \tag{21}$$

where $W_\omega^{\text{combined}} = W_\omega^{(2)} W_\omega^{(1)}$ and $b_\omega^{\text{combined}} = W_\omega^{(2)} b_\omega^{(1)} + b_\omega^{(2)}$. Following the same lineariza-

tion principle applied to the portfolio weights, the adversarial instrument function is also restricted to be a linear in stock characteristics and macroeconomic indicators, such that $g^{\text{linear}}(I_t, I_{i,t}; \vartheta_g) = W_g^{(2)} h_g^{(1)} + b_g^{(2)}$ with $h_g^{(1)} = W_g^{(1)}[I_t, I_{i,t}] + b_g^{(1)}$. The algorithmic implementation for the linear specifications follows the same structure as their nonlinear counterparts (Algorithms 1 and 2), with the only difference being the removal of nonlinear activations from the network architecture. While this removes the model's ability to capture complex interactions between characteristics, it allows us to test whether our transaction cost findings depend on nonlinear relationships or represent a more fundamental phenomenon.

Sharpe Ratios

| | | GAN | | | | FFN | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Raw returns | | Net returns | | Raw returns | | Net returns | |
| | Mkt | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ |
| Train | 0.088 | 2.949 | 1.965 | 1.900 | 1.630 | 2.783 | 1.623 | 1.853 | 1.394 |
| Valid | 0.345 | 0.843 | 0.922 | -0.020 | **0.176** | 0.936 | 0.816 | 0.036 | **0.256** |
| Test | 0.143 | 0.256 | 0.268 | 0.028 | **0.084** | 0.281 | 0.301 | 0.051 | **0.141** |

Pricing Performance

| | | Explained Variation | | | | XS-$R^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | GAN | | FFN | | GAN | | FFN | |
| | Mkt | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ |
| Train | | 0.099 | 0.113 | 0.117 | 0.106 | 0.124 | 0.104 | 0.121 | 0.119 |
| Valid | | 0.028 | 0.030 | 0.032 | 0.029 | 0.125 | 0.116 | 0.132 | 0.129 |
| Test | | 0.047 | 0.051 | 0.056 | 0.050 | 0.116 | 0.109 | 0.127 | 0.121 |

Table 5: **Linear SDF results**. This table reports the asset pricing performance of a linear SDF specification across training, validation, and testing periods for both GAN and FFN architectures. Results are shown for the baseline leverage constraints of ±5%. The top panel reports the Sharpe ratios based on raw performance and returns net of transaction costs. We highlight in **red** the Sharpe ratio differentials (TC-aware minus frictionless SDF), which are statistically significant at a 5% confidence level. P-values are calculated based on the robust bootstrap test of Ledoit and Wolf (2008). Raw Market Sharpe ratios (Mkt) are provided for reference. The bottom panels report the explained variation and the cross-sectional $R^2$, respectively.

Table 5 reveals striking differences between linear and nonlinear specifications. Linear models exhibit severe overfitting, achieving extraordinarily high training Sharpe ratios (2.949 for GAN vs. 1.921 for nonlinear) that completely collapse out-of-sample (0.084 vs. 0.113).

Importantly for our main thesis, the benefits of incorporating transaction costs are still

significant in linear models. The latter shows a comparable relative Sharpe ratio improvement (0.084 vs. 0.028), which is also significant at 5% confidence level according to the Ledoit and Wolf (2008) bootstrap test. This confirms that transaction costs are not merely implementation constraints but fundamental forces that shape the economic significance of the SDF.

Interestingly, the asset pricing performance from linear SDFs is substantially lower compared to our baseline non-linear specifications. The explained variation advantage of TC-aware models drops from 56% in nonlinear specifications to just 9% in linear models. This demonstrates that while transaction costs matter even in simple linear settings, their full explanatory impact fully emerges when models can capture the complex, state-dependent relationships that characterize asset prices.[12]

**SDF portfolio composition and characteristics.** Appendix D.3 compares portfolio composition for linear SDFs estimated with and without transaction costs. The results in Figure D.8 reveal that linear models exhibit narrower responses to transaction costs.

Both GAN and FFN architectures show more modest changes in concentration and turnover compared to their nonlinear counterparts (see Figure 4). The HHI ratios remain close to unity, particularly for GAN estimation, indicating that linear SDFs can only marginally adjust portfolio weights in response to trading frictions. This stems from the inability of linear models to capture the complex interactions among characteristics and partly explains the lower explained variation and cross-sectional $R^2$ achieved by linear models. Nevertheless, transaction cost awareness does reduce the SDF portfolio turnover even in linear specifications.

The fundamental difference between linear and nonlinear responses emerges more clearly in the SDF characteristics. Figure D.9 shows that linear TC-aware SDFs exhibit diffuse reallocation across many characteristics compared to frictionless SDFs, while nonlinear SDFs (Figure 5) concentrate their adjustments on fewer characteristics. This pattern reflects the

[12]Linear models treat all characteristics as having constant marginal effects regardless of market conditions or characteristic levels. In contrast, nonlinear models can learn that the same characteristic might have amplified importance during certain market states or when combined with specific other characteristics, to the extent that trading costs make such complex strategies implementable.

superior ability of nonlinear models to capture characteristic interactions. While linear models treat each characteristic as an independent signal with constant marginal effects, nonlinear models can learn complex relationships, such as how profitability and leverage interact, or how dividend policy affects the value of other fundamental signals.

Overall, the evidence from Figure D.8 and Figure D.9 demonstrates that the choice between linear and nonlinear SDF specifications fundamentally affects how models adapt to trading frictions. Linear models offer only limited response to transaction costs, whereas nonlinear models can fully capture the interplay between trading costs and characteristic turnover, which ultimately affect the SDF implementability.

## 4.5 Limits to Arbitrage and Economic Restrictions

Avramov et al. (2023) show that SDF extracted from machine learning methods tend to excel when limits to arbitrage are high, thus questioning their implementability in realistic scenarios. To directly assess whether TC-aware SDFs address these implementation challenges, we conduct a series of subsample analyses. We note that the models are not re-estimated over different sub-samples but simply evaluated over different slices of the testing period based on the market regime indicator variable.[13]

We partition the main sample along two key time-series dimensions. First, we use the National Financial Conditions Index (NFCI) from the Federal Reserve Bank of Chicago to distinguish between tight and loose financial conditions. The NFCI synthesizes information from 105 financial indicators into a single measure, where positive values indicate tighter-than-average financial conditions, with reduced liquidity provision, and negative values indicate looser-than-average conditions, with milder funding constraints.[14] Second, we split the sample by VIX levels (above vs below the time-series median) to capture periods of high versus low

[13]Re-estimating the models based on different sample periods implies that SDF networks can mechanically adjust to different market conditions. For this reason, one can interpret our results as a lower bound on the ability of transaction-cost aware SDFs to adapt to market conditions.

[14]Specifically, for each month $t$ in our sample: Financial Condition$_t$ = Tight if NFCI$_t$ $>$ 0, and Financial Condition$_t$ = Loose if NFCI$_t$ $\leq$ 0.

Sharpe Ratio (Raw)

| | Tight Financial Conditions | | | | Loose Financial Conditions | | | | High VIX | | | | Low VIX | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| $\omega_{it} \in (-5\%, 5\%)$ | -0.104 | -0.129 | -0.165 | -0.181 | 0.409 | 0.381 | 0.388 | 0.353 | 0.364 | 0.330 | 0.311 | 0.225 | 0.364 | 0.336 | 0.390 | 0.391 |
| $\omega_{it} \in (-10\%, 10\%)$ | -0.142 | -0.074 | -0.172 | -0.133 | 0.412 | 0.398 | 0.381 | 0.318 | 0.350 | 0.338 | 0.311 | 0.236 | 0.441 | 0.397 | 0.418 | 0.361 |
| $\omega_{it} \in (-2.5\%, 2.5\%)$ | -0.112 | -0.157 | -0.110 | -0.052 | 0.359 | 0.298 | 0.387 | 0.317 | 0.313 | 0.254 | 0.335 | 0.254 | 0.266 | 0.214 | 0.324 | 0.351 |

Sharpe Ratio (Net)

| | Tight Financial Conditions | | | | Loose Financial Conditions | | | | High VIX | | | | Low VIX | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| $\omega_{it} \in (-5\%, 5\%)$ | -0.423 | -0.350 | -0.432 | -0.288 | 0.131 | **0.180** | 0.124 | **0.210** | 0.063 | **0.115** | 0.036 | **0.088** | 0.074 | **0.126** | 0.101 | **0.237** |
| $\omega_{it} \in (-10\%, 10\%)$ | -0.507 | -0.277 | -0.505 | **-0.286** | 0.115 | **0.207** | 0.099 | **0.169** | 0.033 | **0.137** | 0.011 | **0.082** | 0.099 | **0.188** | 0.097 | **0.198** |
| $\omega_{it} \in (-2.5\%, 2.5\%)$ | -0.315 | -0.292 | -0.340 | **-0.185** | 0.130 | 0.147 | 0.157 | 0.190 | 0.080 | 0.105 | 0.094 | 0.121 | 0.030 | 0.043 | 0.085 | **0.209** |

Explained Variation

| | Tight Financial Conditions | | | | Loose Financial Conditions | | | | High VIX | | | | Low VIX | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| $\omega_{it} \in (-5\%, 5\%)$ | 0.090 | 0.139 | 0.114 | 0.158 | 0.059 | 0.090 | 0.076 | 0.101 | 0.073 | 0.110 | 0.092 | 0.123 | 0.040 | 0.061 | 0.051 | 0.067 |
| $\omega_{it} \in (-10\%, 10\%)$ | 0.091 | 0.137 | 0.096 | 0.156 | 0.063 | 0.088 | 0.063 | 0.101 | 0.077 | 0.108 | 0.078 | 0.121 | 0.043 | 0.059 | 0.043 | 0.068 |
| $\omega_{it} \in (-2.5\%, 2.5\%)$ | 0.089 | 0.172 | 0.113 | 0.154 | 0.060 | 0.107 | 0.075 | 0.099 | 0.074 | 0.130 | 0.092 | 0.120 | 0.041 | 0.071 | 0.051 | 0.066 |

XS-$R^2$

| | Tight Financial Conditions | | | | Loose Financial Conditions | | | | High VIX | | | | Low VIX | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| $\omega_{it} \in (-5\%, 5\%)$ | 0.077 | 0.135 | 0.102 | 0.160 | 0.141 | 0.216 | 0.167 | 0.221 | 0.132 | 0.212 | 0.158 | 0.220 | 0.115 | 0.152 | 0.125 | 0.152 |
| $\omega_{it} \in (-10\%, 10\%)$ | 0.059 | 0.125 | 0.084 | 0.149 | 0.142 | 0.202 | 0.159 | 0.228 | 0.133 | 0.200 | 0.149 | 0.228 | 0.115 | 0.145 | 0.123 | 0.157 |
| $\omega_{it} \in (-2.5\%, 2.5\%)$ | 0.066 | 0.157 | 0.098 | 0.150 | 0.139 | 0.230 | 0.161 | 0.220 | 0.130 | 0.241 | 0.149 | 0.217 | 0.113 | 0.147 | 0.121 | 0.153 |

Table 6: **Limits to arbitrage and pricing performance**. This table reports Sharpe ratios (raw and net of transaction costs), explained variation, and cross-sectional $R^2$ for subsamples partitioned based on financial conditions using the National Financial Conditions Index (NFCI) from the Chicago Fed, and market uncertainty regimes as proxied by the VIX. Results are shown for both GAN and feed-forward neural network (FFN) architectures under three leverage constraint specifications. $M_{t+1}$ denotes the standard frictionless SDF, while $M^{TC}_{t+1}$ represents the transaction cost-aware SDF. We highlight in **red** the Sharpe ratio differentials (TC-aware minus frictionless SDF), which are statistically significant at a 5% confidence level. P-values are calculated based on the robust bootstrap test of Ledoit and Wolf (2008). The sample period spans from January 1972 to December 2023 for financial conditions analysis, and from January 1990 to December 2023 for VIX-based analysis. All results are based on out-of-sample test performance.

aggregate market uncertainty.[15]

The results in Table 6 reveal that transaction cost awareness provides benefits that are consistent across market and liquidity regimes, with the magnitude of improvements being largest precisely when frictions matter most. During tight financial conditions, the explained variation from TC-aware SDF improves by 54.4% (from 0.090 to 0.139) and cross-sectional $R^2$ by 75.3% (from 0.077 to 0.135), compared to 53.2% and 53.4% improvements, respectively, during loose conditions. Similarly, high-uncertainty periods show cross-sectional $R^2$ improvements of 60.6% (from 0.132 to 0.212) versus 32.2% (from 0.115 to 0.152) during low-uncertainty periods, demonstrating that the value of incorporating implementation constraints scales with the severity of market frictions.

Consistent with the intuition of Avramov et al. (2023), the results reveal that frictionless SDFs deteriorate most severely when frictions matter most, with explained variation falling to only 0.090 and 0.114 for GAN and FFN during tight financial conditions compared to 0.139 and 0.158 from TC-aware SDFs. The net Sharpe ratio results provide the most compelling evidence of this asymmetric benefit: during tight financial conditions, frictionless SDFs collapse with deeply negative net Sharpe ratios ($-0.315$ to $-0.507$), while transaction cost-aware models maintain substantially better and more stable performance ($-0.185$ to $-0.292$), representing statistically significant improvements of 77% to 204%. High-uncertainty periods amplify these benefits further, with transaction cost awareness delivering significant improvements ranging from 88% to 745% compared to 43% to 135% during calm periods. Notably, the variability in net Sharpe ratios across specifications is consistently lower for TC-aware models during stressed conditions, suggesting enhanced stability when traditional approaches become most unreliable.

These findings demonstrate that our TC-aware approach delivers superior robustness across aggregate liquidity and volatility regimes, with the robustness advantage being most pronounced during periods of market stress when effective risk management is most crucial. This

[15]For each month $t$: Uncertainty Regime$_t$ = High if VIX$_t$ > median(VIX), and Uncertainty Regime$_t$ = Low if VIX$_t \leq$ median(VIX).

pattern reflects the economic reality that embedding transaction costs in the SDF estimate is most critical precisely when market conditions render frictionless approaches economically unviable.

Next, we also examine the SDFs performance both including and excluding the bottom 25% of firms by market capitalization. Table 7 reveals that transaction cost awareness remains valuable even when focusing on larger, more liquid stocks, with the benefits extending beyond simply handling difficult-to-trade securities.

The substantial improvement in baseline performance when excluding microcaps – with explained variation increasing from 0.063 to 0.117 and cross-sectional $R^2$ rising from 0.145 to 0.196 – confirms that traditional frictionless approaches perform better when focusing on more liquid securities, consistent with Avramov et al. (2023) finding that machine learning profitability is concentrated in difficult-to-arbitrage stocks. Notably, however, the net Sharpe ratios for frictionless models remain quite low even after excluding microcaps (0.022 for GAN, 0.013 for FFN under 5% constraints).

Our TC-aware SDF approach demonstrates that this concentration need not be a limitation. Rather than simply excluding the securities where machine learning methods are most effective, we can explicitly model the trading frictions that make these securities challenging to exploit. The TC-aware SDF achieves remarkably similar XS-$R^2$ whether microcaps are included (0.223 for GAN and 0.230 for FFN) or excluded (0.226 for GAN and 0.237 for FFN), suggesting that explicit transaction cost modeling allows us to capture the cross-sectional variation across the full spectrum of stocks without sacrificing implementability.

Most importantly, transaction cost awareness continues to provide meaningful benefits even when focusing on the more liquid segment of the market. When excluding the bottom 25% of stocks by market capitalization, explained variation still improves by 25.6% (from 0.117 to 0.147) and cross-sectional $R^2$ by 15.3% (from 0.196 to 0.226). The net Sharpe ratio improvements are particularly striking: from near-zero or negative values for frictionless models to substantially positive performance for TC-aware versions (0.067 vs 0.022 for GAN, 0.082 vs

Full Cross-Section

| | Explained Variation | | | | XS-$R^2$ | | | | Sharpe Ratio (Raw) | | | | Sharpe Ratio (Net) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| $\omega_{it} \in (-5\%, 5\%)$ | 0.063 | 0.098 | 0.081 | 0.109 | 0.145 | 0.223 | 0.172 | 0.230 | 0.348 | 0.317 | 0.315 | 0.253 | 0.064 | **0.113** | 0.051 | **0.118** |
| $\omega_{it} \in (-10\%, 10\%)$ | 0.067 | 0.096 | 0.068 | 0.109 | 0.147 | 0.209 | 0.163 | 0.237 | 0.354 | 0.335 | 0.321 | 0.256 | 0.047 | **0.143** | 0.032 | **0.107** |
| $\omega_{it} \in (-2.5\%, 2.5\%)$ | 0.065 | 0.117 | 0.081 | 0.107 | 0.143 | 0.241 | 0.165 | 0.228 | 0.286 | 0.231 | 0.317 | 0.266 | 0.063 | 0.084 | 0.087 | **0.137** |

Exclude MicroCap (Bottom 25%)

| | Explained Variation | | | | XS-$R^2$ | | | | Sharpe Ratio (Raw) | | | | Sharpe Ratio (Net) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| $\omega_{it} \in (-5\%, 5\%)$ | 0.117 | 0.147 | 0.139 | 0.162 | 0.196 | 0.226 | 0.215 | 0.237 | 0.223 | 0.218 | 0.201 | 0.191 | 0.022 | **0.067** | 0.013 | **0.082** |
| $\omega_{it} \in (-10\%, 10\%)$ | 0.128 | 0.145 | 0.121 | 0.158 | 0.203 | 0.223 | 0.204 | 0.236 | 0.204 | 0.217 | 0.187 | 0.185 | -0.004 | **0.075** | -0.013 | **0.065** |
| $\omega_{it} \in (-2.5\%, 2.5\%)$ | 0.126 | 0.167 | 0.144 | 0.159 | 0.193 | 0.242 | 0.216 | 0.232 | 0.203 | 0.160 | 0.232 | 0.216 | 0.040 | 0.049 | 0.069 | 0.113 |

Table 7: **Performance metrics of transaction cost-aware SDFs**. This table presents a comprehensive analysis of stochastic discount factor (SDF) performance metrics across different portfolio weight constraints and sample compositions. The analysis compares traditional frictionless SDFs ($M_{t+1}$) with transaction cost-aware SDFs ($M^{TC}_{t+1}$) using both GAN and FFN architectures. The upper panel reports results for the full cross-section of stocks, while the lower panel excludes microcap stocks (bottom 25% by market capitalization). Performance is evaluated using explained variation, cross-sectional $R^2$, and Sharpe ratios for both raw and net returns. We highlight in **red** the Sharpe ratio differentials (TC-aware minus frictionless SDF), which are statistically significant at a 5% confidence level. P-values are calculated based on the robust bootstrap test of Ledoit and Wolf (2008). Portfolio weight constraints $\omega_{it}$ are tested at three levels: $\pm 2.5\%$, $\pm 5\%$, and $\pm 10\%$. The sample period covers January 1972 to December 2023, with all metrics computed using out-of-sample test data.

0.013 for FFN under 5% constraints). The persistence of benefits across both liquid and illiquid segments directly provides evidence that our approach captures the fundamental structural features of asset pricing, rather than merely leveraging microcap-specific frictions.

To establish that the transaction cost awareness benefits across market segments are not merely artifacts of neural network complexity, we also examine a linear specification of the SDF. Table D.4 in Appendix D.3 presents results where we constrain the neural networks to impose a strictly linear parameterization of the pricing kernel (see Section 4.4 for details).

The results confirm our main findings while revealing the complementary value of non-linear modeling. Under the linear constraint, both explained variation and cross-sectional $R^2$ decrease by approximately 50% compared to the non-linear specification (e.g., explained variation drops from 0.098 to 0.051 for the TC-aware GAN), highlighting the importance of capturing complex characteristic interactions. Nevertheless, the linear TC-aware SDF still consistently outperforms its frictionless counterpart across all metrics and market conditions. For instance, net Sharpe ratios improve from 0.028 to 0.084 in the full sample, and the benefits remain pronounced during tight financial conditions (from $-0.157$ to $-0.026$), reinforcing that the importance of transaction cost awareness transcends model complexity.

These results demonstrate that while non-linear specifications capture richer patterns in the data, the fundamental insight about the role of transaction costs over market segments remains valid even under restrictive linear assumptions.

# 5 Robustness and Additional Results

To assess the robustness of our main findings, we expand the main empirical results along two dimensions. First, we examine two alternative sample periods that exploit the tradeoff between time-series length and breadth of firm characteristics. Second, we consider an alternative definition of transaction costs. Throughout these robustness tests, we focus on the baseline leverage constraints of $\omega_{it} \in (-5\%, +5\%)$ for simplicity. As demonstrated in Appendix D, the

alternative leverage constraints of ±2.5% and ±10% yield qualitatively similar results and do not alter the main insights.

## 5.1 Alternative Data Configurations

Asset pricing performance can be sensitive to the trade-off between time-series length and cross-sectional breadth, especially when signals are weak (e.g., Liao et al., 2023; Shen and Xiu, 2025). To assess robustness, we examine two alternative sample configurations that vary this trade-off: a longer sample starting in January 1967 (56 years, but only 62 characteristics available) and a shorter sample starting in January 1986 (37 years, but 88 characteristics available). Both samples maintain our standard 20-year training period and 5-year validation period, with testing periods of January 1991-2023 and January 2011-2023, respectively.

| Jan 1967 to Dec 2023 | | GAN | | | | FFN | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Raw returns | | Net returns | | Raw returns | | Net returns | |
| | Mkt | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| Train | 0.072 | 1.763 | 1.204 | 1.452 | 1.015 | 1.171 | 1.041 | 0.956 | 0.919 |
| Valid | 0.135 | 0.752 | 0.602 | -0.149 | -0.047 | 0.654 | 0.470 | -0.078 | **-0.005** |
| Test | 0.159 | 0.232 | 0.221 | -0.073 | **-0.005** | 0.231 | 0.251 | -0.018 | **0.072** |

| Jan 1986 to Dec 2023 | | GAN | | | | FFN | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Raw returns | | Net returns | | Raw returns | | Net returns | |
| | Mkt | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| Train | 0.144 | 2.482 | 1.889 | 1.370 | 1.267 | 1.968 | 1.646 | 1.114 | 1.257 |
| Valid | 0.039 | 0.499 | 0.499 | 0.081 | 0.227 | 0.594 | 0.451 | 0.159 | 0.272 |
| Test | 0.239 | 0.407 | 0.385 | 0.167 | **0.236** | 0.404 | 0.340 | 0.170 | **0.242** |

Table 8: **Sharpe ratios for alternative samples**. This table reports the Sharpe ratios of tangency portfolios across different training, validation, and testing periods for both GAN and feed-forward neural network (FFN) architectures. We present both raw performance and returns net of transaction costs for the standard SDF ($M_{t+1}$) and transaction-cost-aware SDF ($M^{TC}_{t+1}$). Results are shown for the baseline leverage constraints of ±5%. We highlight in **red** the Sharpe ratio differentials (TC-aware minus frictionless SDF), which are statistically significant at a 5% confidence level. P-values are calculated based on the robust bootstrap test of Ledoit and Wolf (2008). Raw Market Sharpe ratios (Mkt) are provided for reference.

Table 8 reveals that longer time series do not necessarily improve out-of-sample performance when they come at the cost of fewer characteristics. However, different testing periods compli-

cate this interpretation, as market Sharpe ratios vary substantially. The superior performance in the shorter sample likely reflects the more favorable market environment during 2011-2023 rather than purely the benefit of additional characteristics.

Despite these different market conditions, TC-aware SDFs consistently outperform frictionless SDFs. The improvement in net returns is remarkably stable across different data configurations. Over the longer testing period (1991-2023 testing), where frictionless SDFs produce negative net Sharpe ratios (-0.073 for GAN, -0.018 for FFN), TC-aware SDFs maintain near-zero or positive performance (-0.005 for GAN, 0.072 for FFN). This performance spread is statistically significant at a 5% conventional threshold.

| Jan 1967 to Dec 2023 | Explained Variation | | | | XS-$R^2$ | | | |
|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ |
| Train | 0.188 | 0.229 | 0.243 | 0.252 | 0.295 | 0.323 | 0.334 | 0.338 |
| Validation | 0.124 | 0.148 | 0.157 | 0.162 | 0.083 | 0.102 | 0.109 | 0.111 |
| Test | 0.086 | 0.105 | 0.112 | 0.116 | 0.224 | 0.250 | 0.259 | 0.262 |

| Jan 1986 to Dec 2023 | Explained Variation | | | | XS-$R^2$ | | | |
|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ |
| Train | 0.068 | 0.067 | 0.081 | 0.086 | 0.189 | 0.116 | 0.218 | 0.213 |
| Validation | 0.097 | 0.076 | 0.120 | 0.128 | 0.107 | 0.038 | 0.142 | 0.142 |
| Test | 0.073 | 0.059 | 0.090 | 0.096 | 0.160 | 0.117 | 0.187 | 0.184 |

Table 9: **Pricing performance for alternative samples**. This table reports the explained variation (EV) and cross-sectional $R^2$ for different SDF specifications across different training, validation, and testing periods. Results are shown for both GAN and feed-forward neural network (FFN) architectures, comparing the standard SDF ($M_{t+1}$) and transaction-cost-aware SDF ($M_{t+1}^{TC}$) under the baseline leverage constraints of $\pm 5\%$.

Table 9 shows that transaction cost awareness generally improves both explained variation and cross-sectional $R^2$, though the magnitude varies across samples. In the longer sample, improvements are substantial: approximately 20% for explained variation and 10% for cross-sectional $R^2$. However, in the shorter sample, we occasionally observe deterioration in pricing metrics despite superior Sharpe ratios.

**SDF portfolio composition and characteristics.** Figure D.10 in Appendix D.4 shows that the effect of transaction costs on the SDF portfolio composition varies across sample periods. The most striking difference appears in portfolio concentration. Models tested on the longer sample (1991-2023) show more dramatic concentration in the frictionless SDF portfolio compared to its TC-aware counterpart: GAN architectures exhibit a 66% HHI increase compared to just 5% for the shorter sample. FFN models show more modest differences (28% vs 19% increase), suggesting they inherently favor concentrated positions regardless of historical context. Portfolio turnover is lower when transaction costs are considered for both samples, confirming that cost awareness consistently discourages excessive trading.

The economic intuition of why transaction costs affect the SDF portfolio composition less in the shorter sample is compelling: the pre-1986 period included substantially higher trading frictions, as shown in Figure 3. Thus, models trained on this older period learn to place larger penalties on high-turnover characteristics. In contrast, the post-1986 period, characterized by generally declining transaction costs, presents a milder penalty on tangency portfolio rebalancing.

This historical learning effect extends to SDF portfolio characteristics. Figure D.11 in Appendix D.4 shows the differential impact of transaction costs on the SDF characteristics. Models trained on the longer sample exhibit greater sensitivity to transaction costs, with pronounced aversion to high-turnover strategies. In contrast, models trained on the post-1986 sample display more nuanced adjustments, with moderate reductions in high-turnover characteristics and increased weights on smoother signals, such as dividend policy and IPO activity.

This evidence demonstrates, perhaps unsurprisingly, that the effect of transaction cost is not static but depends critically on the historical context upon which a model is trained. Models exposed to high-friction periods develop a stronger aversion to costly trading strategies.

## 5.2 Alternative Transaction Cost Measure

Our baseline estimates utilize effective bid-ask spreads as proxies for transaction costs, which vary substantially across stocks and over time. To ensure our findings are not artifacts of specific cost estimates or microstructure assumptions, we follow Brandt et al. (2009) and examine an alternative specification whereby transaction costs are approximated as $c_{i,t} = 0.006 - 0.0025 \times \text{me}i,t$, with $\text{me}i,t$ representing the market capitalization percentile of firm $i$ (normalized between 0 and 1).

Figure D.12 in Appendix D.5 reports the cross-sectional distribution of the Brandt et al. (2009) transaction costs (in %) over the testing period. This specification generates transaction costs ranging from 0.6% for the smallest firms to 0.35% for the largest firms, substantially lower than our baseline bid-ask spread estimates. The more conservative cost structure provides a stringent test of our main thesis: if transaction cost awareness only matters when costs are extremely high, we should observe minimal benefits under this milder specification.

Table 10 confirms that TC-aware SDFs maintain their performance advantage under the alternative cost specification, with results that are remarkably consistent with our baseline findings. Net Sharpe ratios improve substantially for FFN (from 0.146 to 0.210, a 44% increase) and remain stable for GAN (from 0.167 to 0.165), demonstrating robust implementability even under conservative cost assumptions. More remarkably, the asset pricing improvements are just as significant as under our baseline specification: explained variation increases by 44% for GAN (6.3% to 9.1%) and 40% for FFN (8.1% to 11.3%), while cross-sectional $R^2$ improves by 52% (14.5% to 22%) and 39% (17.2% to 23.9%), respectively.

These results provide further evidence that our findings reflect fundamental changes in SDF estimation rather than artifacts of inflated cost estimates. The persistence of both portfolio performance and asset pricing improvements suggests that the value of incorporating implementation constraints extends beyond simply avoiding high-cost trades to encompassing a more fundamental reorientation toward implementable tangency portfolios, as discussed in the next

| Sharpe Ratios | | GAN | | | | FFN | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Raw returns | | Net returns | | Raw returns | | Net returns | |
| | Mkt | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| Train | 0.088 | 1.921 | 1.554 | 1.548 | 1.293 | 1.547 | 1.447 | 1.243 | 1.256 |
| Validation | 0.345 | 1.178 | 1.130 | 0.811 | 0.808 | 1.216 | 1.108 | 0.836 | 0.854 |
| Test | 0.143 | 0.348 | 0.320 | 0.167 | 0.165 | 0.315 | 0.336 | 0.146 | **0.210** |

| Pricing Performance | Explained Variation | | | | XS-$R^2$ | | | |
|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ | $M_{t+1}$ | $M^{TC}_{t+1}$ |
| Train | 0.124 | 0.179 | 0.151 | 0.211 | 0.134 | 0.242 | 0.150 | 0.256 |
| Validation | 0.036 | 0.049 | 0.044 | 0.058 | 0.147 | 0.198 | 0.158 | 0.200 |
| Test | 0.063 | 0.091 | 0.081 | 0.113 | 0.145 | 0.220 | 0.172 | 0.239 |

Table 10: **Transaction costs as in Brandt et al. (2009)**. This table reports the asset pricing performance of a linear SDF specification across training, validation, and testing periods for both GAN and FFN architectures. Results are shown for the baseline leverage constraints of ±5%. The top panel reports the Sharpe ratios based on raw performance and returns net of transaction costs. Transaction costs are calculated as in Brandt et al. (2009). We highlight in **red** the Sharpe ratio differentials (TC-aware minus frictionless SDF), which are statistically significant at a 5% confidence level. P-values are calculated based on the robust bootstrap test of Ledoit and Wolf (2008). Raw Market Sharpe ratios (Mkt) are provided for reference. The middle and bottom panels report the explained variation and the cross-sectional $R^2$, respectively.

section.

**SDF portfolio composition and characteristics.** The portfolio composition effects under alternative transaction costs reveal consistent patterns with our baseline results, though with attenuated magnitudes reflecting the lower cost assumptions. Figure D.13 in Appendix D.5 shows that transaction costs still promote portfolio concentration (25% HHI increase for GAN, 12% for FFN) and reduce turnover (22%-37% higher turnover ratios for frictionless models).

More importantly, the reallocation pattern towards more stable characteristics remains qualitatively similar despite the lower cost levels. Figure D.14 in Appendix D.5 demonstrates that both architectures continue to reduce exposure to high-turnover characteristics – such as short-term reversal (`STreversal`), return skewness measures (`ReturnSkew`, `ReturnSkew3F`), and momentum signals (`Max`). The reallocation toward profitability measures (`CashProd`, `AM`), leverage indicators (`OPLeverage`, `ConvDebt`), and liquidity proxies remains economically mean-

ingful.

The magnitude of these adjustments is smaller due to the lower transaction cost estimates. When costs are moderate, models make measured adjustments that balance the benefits of various characteristics against their implementation costs. When costs are high (as in our baseline specification), the adjustments are correspondingly larger. This sensitivity to cost levels validates that our approach captures genuine economic trade-offs.

# 6 Conclusion

This paper examines the asset pricing implications of transaction costs from the perspective of a conventional no-arbitrage restriction. Rather than treating trading frictions as post-optimization adjustments, we argue that they should be incorporated structurally into asset pricing models during the estimation process. Using adversarial neural networks on nearly 10,000 stocks from 1972 to 2023, we find that transaction cost-aware SDFs deliver higher Sharpe ratios and better pricing performance than frictionless models when evaluated net of trading costs.

This improvement stems from endogenous reallocation: high-turnover characteristics see reduced influence while more stable fundamentals gain prominence. The resulting SDFs are more diversified, exhibit lower turnover, and emphasize characteristics with favorable return-to-turnover ratios, creating tangency portfolios that are still theoretically sound but practically more implementable.

The effect of transaction cost awareness persists across different model architectures, leverage constraints, sample periods, and cost specifications, demonstrating that incorporating implementation costs represents a fundamental improvement rather than a response to specific market conditions.

# References


Abdi, F. and Ranaldo, A. (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *The Review of Financial Studies*, 30(12):4437–4480.

Acharya, V. V. and Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of financial Economics*, 77(2):375–410.

Amihud, Y. and Mendelson, H. (1986). Liquidity and stock returns. *Financial Analysts Journal*, 42(3):43–48.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, pages 817–858.

Ardia, D., Guidotti, E., and Kroencke, T. A. (2024). Efficient estimation of bid–ask spreads from open, high, low, and close prices. *Journal of Financial Economics*, 161:103916.

Avramov, D., Cheng, S., and Metzker, L. (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science*, 69(5):2587–2619.

Back, K. (2010). Martingale pricing. *Annu. Rev. Financ. Econ.*, 2(1):235–250.

Baldi-Lanfranchi, F. (2023). Transaction-cost-aware factors. *Available at SSRN 4737166*.

Balduzzi, P. and Lynch, A. W. (1999). Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, 52(1):47–78.

Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of financial economics*, 99(2):427–446.

Barroso, P. and Detzel, A. (2021). Do limits to arbitrage explain the benefits of volatility-managed portfolios? *Journal of Financial Economics*, 140(3):744–767.

Bessembinder, H. and Venkataraman, K. (2010). Bid-ask spreads: Measuring trade execution costs in financial markets. *Encyclopedia of quantitative finance*, pages 184–190.

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

Bianchi, D. and Venturi, P. (2024). Weak signals, small bets: A portfolio perspective on firm characteristics. *Working Paper*.

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9):3411–3447.

Bryzgalova, S., Lerner, S., Lettau, M., and Pelger, M. (2025). Missing financial data. *The Review of Financial Studies*, 38(3):803–882.

Bryzgalova, S., Pelger, M., and Zhu, J. (2023). Forest through the trees: Building cross-sections of stock returns. *Review of Financial Studies, forthcoming*.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Chen, A. Y. and McCoy, J. (2024). Missing values handling for machine learning portfolios. *Journal of Financial Economics*, 155:103815.

Chen, A. Y. and Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review, Forthcoming*.

Chen, H., Lundberg, S., and Lee, S.-I. (2020). Explaining models by propagating shapley values of local components. In *Explainable AI in Healthcare and Medicine: Building a Culture of Transparency and Accountability*, pages 261–270. Springer.

Chen, L., Pelger, M., and Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2):714–750.

Cong, L. W., Tang, K., Wang, J., and Zhang, Y. (2021a). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai. *Available at SSRN 3554486*.

Cong, L. W., Tang, K., Wang, J., and Zhang, Y. (2021b). Deep sequence modeling: Development and applications in asset pricing. *arXiv preprint arXiv:2108.08999*.

Corwin, S. A. and Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The journal of finance*, 67(2):719–760.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J., and Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, 33(5):2180–2222.

Detzel, A., Novy-Marx, R., and Velikov, M. (2023). Model comparison with transaction costs. *The Journal of Finance*, 78(3):1743–1775.

Dubey, S. R., Singh, S. K., and Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 503:92–108.

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

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

Fan, J., Ke, Z. T., Liao, Y., and Neuhierl, A. (2022). Structural deep learning in conditional asset pricing. *Available at SSRN 4117882*.

Feng, G., He, J., Polson, N. G., and Xu, J. (2024). Deep learning in characteristics-sorted factor models. *Journal of Financial and Quantitative Analysis*, 59(7):3001–3036.

Freyberger, J., Höppner, B., Neuhierl, A., and Weber, M. (2024). Missing data in asset pricing panels. *The Review of Financial Studies*, page hhae003.

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

Gagliardini, P. and Ma, H. (2019). Extracting statistical factors when betas are time-varying. *Swiss Finance Institute Research Paper*, (19-65).

Gârleanu, N. and Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152.

Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100.

Goyenko, R., Kelly, B. T., Moskowitz, T. J., Su, Y., and Zhang, C. (2024). Trading volume alpha. *Available at SSRN 4802345*.

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

Hansen, L. P. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of political economy*, 99(2):225–262.

Hansen, L. P. and Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *The Journal of Finance*, 52(2):557–590.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3):1445–1477.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91.

Jensen, T. I., Kelly, B. T., Malamud, S., and Pedersen, L. H. (2024). Machine learning and the implementable efficient frontier. *Swiss Finance Institute Research Paper*, (22-63).

Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *Journal of Finance*, pages 889–908.

Kelly, B. T., Kuznetsov, B., Malamud, S., and Xu, T. A. (2025). Artificial intelligence asset pricing models. Technical report, National Bureau of Economic Research.

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

Korsaye, S. A., Quaini, A., and Trojani, F. (2021). Smart stochastic discount factors. *Swiss Finance Institute Research Paper*, (21-51).

Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.

Lettau, M. and Pelger, M. (2020). Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies*, 33(5):2274–2325.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial economics*, 96(2):175–194.

Li, J., Zhang, C., Zhou, J. T., Fu, H., Xia, S., and Hu, Q. (2021). Deep-lift: Deep label-specific feature learning for image annotation. *IEEE transactions on Cybernetics*, 52(8):7732–7741.

Liao, Y., Ma, X., Neuhierl, A., and Shi, Z. (2023). Economic forecasts using many noises. *Available at SSRN 4659309*.

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

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

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

Nagel, S. and Singleton, K. J. (2011). Estimation and evaluation of conditional asset pricing models. *The Journal of Finance*, 66(3):873–909.

Novy-Marx, R. and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147.

Pástor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3):642–685.

Romano, J. (1992). A circular block-resampling procedure for stationary data. *Exploring the Limits of Bootstrap*, 270:263.

Schneider, P. and Trojani, F. (2019). (almost) model-free recovery. *The Journal of Finance*, 74(1):323–370.

Shapley, L. S. et al. (1953). A value for n-person games.

Shen, Z. and Xiu, D. (2025). Can machines learn weak signals? Technical report, National Bureau of Economic Research.

Shrikumar, A., Greenside, P., and Kundaje, A. (2017). Learning important features through propagating activation differences. In *International conference on machine learning*, pages 3145–3153. PMlR.

Shumway, T. and Warther, V. A. (1999). The delisting bias in crsp's nasdaq data and its implications for the size effect. *The Journal of Finance*, 54(6):2361–2379.

Simon, F., Weibels, S., and Zimmermann, T. (2022). Deep parametric portfolio policies. *Available at SSRN 4150292*.

Vayanos, D. (1998). Transaction costs and asset prices: A dynamic equilibrium model. *The Review of Financial Studies*, 11(1):1–58.

# Supplementary Appendix for: Transaction Costs and the Stochastic Discount Factor

In this appendix, we provide a more detailed description of the deep learning implementations. The steps follow the approach of Chen et al. (2024) with the addition of the transaction cost penalty. We also provide a more detailed discussion of the DeepSHAP attribution approach as well as a description of the firm characteristics and the sample of stock returns used in the empirical analysis. Finally, we also provide additional empirical results. All explanations and additional results are referenced in the main text of the paper where appropriate.

## A Estimation and Implementation Details

### A.1 SDF Upstream Pre-Training

To initialize the parameters $\vartheta_\omega^{\text{init}}$ and $\vartheta_g^{\text{init}}$ that serve as inputs to Algorithm 1, we implement an upstream pretraining procedure that provides identical initial conditions for SDF with and without transaction cost awareness. We first train the SDF network under unconditional constraints, setting $g_{j,t} \equiv 1$ and ignoring transaction costs. Then, conditional on the trained SDF network parameters, we train the $g(\cdot;\vartheta_g)$ network to identify instruments that maximize pricing errors. Specifically, pretraining consists of two steps:

**Step 1 (SDF network pretraining):** For a random subset (mini-batch) $\mathcal{B}$ of size $B$ from our training data, the empirical loss is

$$\mathcal{L}_{\mathcal{B}}(\vartheta_\omega) = \frac{1}{B}\sum_{(t,i)\in\mathcal{B}} \frac{1}{N}\sum_{j=1}^{N} \left\| M_{t+1}(\vartheta_\omega)\, r^e_{j,t+1} \right\|^2, \tag{A.1}$$

where

$$M_{t+1}(\vartheta_\omega) = 1 - \sum_{i=1}^{N} \omega(I_t, I_{i,t};\vartheta_\omega)\, r^e_{i,t+1}.$$

During this step, we update *only* the SDF-portfolio parameters $\vartheta_\omega$; the moment network remains switched off $\left(g_{j,t} \equiv 1\right)$ and transaction costs are ignored. Gradient descent on $\mathcal{L}_{\mathcal{B}}(\vartheta_\omega)$ is performed for $n_{\text{sdf}}$ epochs using Adam with learning rate $10^{-4}$.

**Step 2 (Moment network update):** We freeze the SDF-portfolio parameters at their best Step 1 value $\vartheta_\omega^\star$ and update only the instrument network parameters $\vartheta_g$. We minimize the negative empirical loss:

$$\mathcal{L}_{\mathcal{B}}(\vartheta_g;\vartheta_\omega^\star) = \frac{1}{B}\sum_{(t,i)\in\mathcal{B}} \frac{1}{N}\sum_{j=1}^{N} \left\| M_{t+1}(\vartheta_\omega^\star)\, r^e_{j,t+1}\, g(I_t, I_{j,t};\vartheta_g) \right\|^2, \tag{A.2}$$

with

$$M_{t+1}(\vartheta_\omega^\star) = 1 - \sum_{i=1}^{N} \omega(I_t, I_{i,t}; \vartheta_\omega^\star)\, r_{i,t+1}^e.$$

Gradient ascent on $\mathcal{L}_\mathcal{B}(\vartheta_g; \vartheta_\omega^\star)$ is performed for $n_{\text{mom}}$ epochs using Adam, while all other parameters in the SDF network remain fixed. Algorithm 3 summarizes the pretraining steps.

**Algorithm 3:** Pretraining: Step 1 (SDF network) and Step 2 (Moment network)

**Input :** Random initial parameters $\vartheta_\omega, \vartheta_g$
**Output:** Pretrained $\vartheta_\omega^\star$ and $\vartheta_g^\star$
**Stage 1 (SDF pretraining)**;
**for** *epoch = 1 to* $n_{\text{sdf}}$ **do**
  **foreach** *mini-batch* $\mathcal{B}$ **do**
    `/* Update` $\omega$ `only;` $\vartheta_g$ `is frozen`
    (i) Compute the loss $\mathcal{L}_\mathcal{B}(\vartheta_\omega)$ from (A.1).;
    (ii) $\vartheta_\omega \leftarrow \vartheta_\omega - \eta_\omega \nabla_{\vartheta_\omega} \mathcal{L}_\mathcal{B}(\vartheta_\omega)$;
  Evaluate $\vartheta_\omega^\star$ based on the loss;
**Step 2 (Moments pretraining)**;
**for** *epoch = 1 to* $n_{\text{mom}}$ **do**
  **foreach** *mini-batch* $\mathcal{B}$ **do**
    `/* Update` $g$ `only;` $\vartheta_\omega = \vartheta_\omega^\star$ `is frozen`
    (i) Compute the loss $-\mathcal{L}_\mathcal{B}(\vartheta_g; \vartheta_\omega^\star)$ from (A.2).;
    (ii) $\vartheta_g \leftarrow \vartheta_g + \eta_g \nabla_{\vartheta_g} \mathcal{L}_\mathcal{B}$;
  Evaluate $\vartheta_g^\star$ based on the loss;

Upon completion of this pretraining procedure, we obtain $\vartheta_\omega^\star$ and $\vartheta_g^\star$, which serve as the initialization for the conditional SDF training described in Algorithm 1 in Section 2.1. These pretrained parameters ensure that all models start conditional training under identical initial conditions. In our empirical analysis, we set $n_{\text{unc}} = 256$ and $n_{\text{mom}} = 64$ as the number of epochs for Steps 1 and 2, respectively, and optimize using Adam with a learning rate of 0.001.

## Hyperparameters setting

The SDF network consists of two components: a Long Short-Term Memory (LSTM) cell that summarizes macroeconomic dynamics into four hidden states, and a feedforward network that takes these hidden states along with asset-specific information as inputs to generate portfolio weights. The feedforward component includes two hidden layers, each with 64 hidden units.

Similarly, the instrument network employs an LSTM cell to encode macroeconomic dynamics into 32 hidden states, followed by a feedforward network that combines these macroeconomic hidden states and asset characteristics into eight conditional moments. These moments capture the pricing errors of portfolios constructed based on asset characteristics or state-dependent payoffs under particular macroeconomic scenarios.

Chen et al. (2024) perform an extensive search for optimal hyperparameters across 384 distinct network configurations—fitting the GAN model for each combination and selecting the

best four hyperparameter configurations based on performance on the validation dataset. Given that our primary research objective is to investigate the incremental impact of transaction costs on model performance, we opt not to replicate this computationally intensive process and directly adopt the optimal network configuration suggested by Chen et al. (2024).

This choice allows us to maintain consistency across model specifications, isolating the influence of introducing transaction costs from other computational considerations. By holding the network architecture and hyperparameters constant, we can effectively gauge the effect that transaction costs have on the models' economic performance. Table A.1 summarizes the hyperparameters and network architecture choices employed in the main empirical analysis.

| Hyperparameters | Candidates in Chen et al. (2024) | Optimal |
|---|---|---|
| Layers in SDF network | 2, 3, or 4 | 2 |
| Hidden units in SDF network | 64 | 64 |
| Hidden states in SDF network | 4 or 8 | 4 |
| Hidden states in instrument network | 16 or 32 | 32 |
| Layers in instrument network | 0 or 1 | 0 |
| Hidden units in instrument network | 4, 8, 16, or 32 | 8 |
| Initial learning rate | 0.001, 0.0005, 0.0002, or 0.00001 | 0.001 |
| Dropout | 0.95 | 0.95 |

Table A.1: **Network architectures and hyperparameters.** The table reports the hyperparameters adopted from the optimal configuration in Chen et al. (2024). The dropout probability denotes the probability of keeping a node during training.

## A.2 DeepLIFT Attribution and SHAP Values

Consider our tangency portfolio weight network $\omega(I_t, I_{i,t}; \vartheta_\omega)$ that maps asset characteristics $\mathbf{x} = I_{i,t} \in \mathbb{R}^{73}$ to portfolio weight $\omega_i \in \mathbb{R}$. Let $\mathbf{x}^0 = (x_1^0, \ldots, x_{73}^0)$ denote the reference baseline input for each input feature and $\Delta x_j = x_j - x_j^0$ the difference-from-reference for feature $j$. DeepLIFT assigns attribution scores $C_{\Delta x_j, \Delta\omega}$ such that:

$$\Delta\omega = \sum_{j=1}^{73} C_{\Delta x_j, \Delta\omega} \cdot \Delta x_j \tag{A.3}$$

where $\Delta\omega = \omega(\mathbf{x}) - \omega(\mathbf{x}^0)$ is the difference in the SDF network output. Here, $C_{\Delta x_j, \Delta\omega}$ represents the contribution score that quantifies how much of the output difference $\Delta\omega$ can be attributed to the input difference $\Delta x_j$. Intuitively, if a stock has characteristics that differ from the median firm and receives a different portfolio weight, $C_{\Delta x_j, \Delta\omega}$ measures how much of that weight difference is "caused by" characteristic $j$.

The key innovation of DeepLIFT is computing these attribution scores by propagating contribution ratios backwards through the network, rather than using gradients, which can suffer from saturation issues in deep networks with ReLU activations (e.g., Dubey et al., 2022). The contribution scores $C_{\Delta x_j, \Delta\omega}$ are calculated by propagating contribution scores backwards through the network, layer by layer, using a process analogous to backpropagation.

For a feedforward network with layers $l = 0, 1, \ldots, L$, let $\mathbf{y}^{(l)} = (y_1^{(l)}, \ldots, y_{n_l}^{(l)})$ denote the activations at layer $l$, and $\mathbf{y}^{(l),0}$ the reference activations at layer $l$. The attribution from input $j$ to output is calculated through three steps:

**Step 1: Forward Pass.** For each layer $l = 0, 1, \ldots, L$:

- Compute activations for actual input: $\mathbf{y}^{(l)}$
- Compute activations for reference input: $\mathbf{y}^{(l),0}$
- Compute activation differences: $\Delta y_i^{(l)} = y_i^{(l)} - y_i^{(l),0}$

**Step 2: Backward Pass (Contribution Propagation).** Starting from the output layer and working backwards, we compute contribution scores for each connection. For ReLU activations $z_i^{(l)} = \max(0, y_i^{(l)})$, DeepLIFT uses the "rescale rule":[16]

$$C_{\Delta y_i^{(l)}, \Delta z_i^{(l)}} = \begin{cases} 1 & \text{if } y_i^{(l)} > 0 \text{ and } y_i^{(l),0} > 0 \\ 0 & \text{if } y_i^{(l)} \leq 0 \text{ and } y_i^{(l),0} \leq 0 \\ \frac{\Delta z_i^{(l)}}{\Delta y_i^{(l)}} & \text{if they cross the activation boundary} \end{cases} \tag{A.5}$$

This handles the nonlinearity by considering three cases: when both actual and reference activations are above the ReLU threshold (full contribution passes through), when both are below (no contribution passes through), and when they cross the boundary (partial contribution based on the actual change ratio).

**Step 3: Path Aggregation.** The final attribution from input $j$ to output is computed by following all possible paths through the network:

$$C_{\Delta x_j, \Delta\omega} = \sum_{\text{paths } P} \prod_{(k,i) \in P} C_{\Delta y_k, \Delta y_i} \tag{A.6}$$

where the sum is over all paths $P$ from input $j$ to the output through the network, and the product is over all edges $(k, i)$ in each path.

To understand the main intuition, consider a simplified case with input $\mathbf{x} = (x_1, x_2)$ representing (profitability, size) relative to a median firm baseline $\mathbf{x}^0 = (0, 0)$. Suppose $\Delta x_1 = 0.8$ (high profitability) and $\Delta x_2 = -0.3$ (small size). The hidden layer activation is $y_1^{(1)} = \text{ReLU}(W_{11} \cdot 0.8 + W_{12} \cdot (-0.3))$, and the output is $\omega = W_{21} \cdot y_1^{(1)} + \ldots$. If the network assigns a +2% weight difference ($\Delta\omega = 0.02$) to this stock relative to the median firm, DeepLIFT traces this difference back through the network. For instance, it might find $C_{\Delta x_1, \Delta\omega} = +0.025$ and $C_{\Delta x_2, \Delta\omega} = +0.067$, such that $\Delta\omega = 0.025 \times 0.8 + 0.067 \times (-0.3) = 0.02 - 0.02 = 0.02$.

**From DeepLIFT to DeepSHAP.** DeepSHAP approximates Shapley values by averaging DeepLIFT attributions over multiple reference points. Given a background distribution

[16] For a linear transformation $y_i^{(l+1)} = \sum_j W_{ij} y_j^{(l)} + b_i$, the contribution from neuron $j$ in layer $l$ to neuron $i$ in layer $l + 1$ is:

$$C_{\Delta y_j^{(l)}, \Delta y_i^{(l+1)}} = W_{ij} \frac{\Delta y_j^{(l)}}{\sum_k W_{ik} \Delta y_k^{(l)}} \tag{A.4}$$

when $\sum_k W_{ik} \Delta y_k^{(l)} \neq 0$. This rule distributes the contribution proportionally: neuron $j$'s contribution to neuron $i$ is proportional to its weight $W_{ij}$ and its share of the total input change to neuron $i$.

$\mathcal{D}_{\text{background}}$, the DeepSHAP attribution is:

$$\phi_j^{\text{DeepSHAP}} = \mathbb{E}_{\mathbf{x}^0 \sim \mathcal{D}_{\text{background}}}[C_{\Delta x_j, \Delta\omega} | \mathbf{x}^0] \tag{A.7}$$

This averages the contribution scores across different baseline scenarios, providing more robust attributions that approximate the expected Shapley values over the background distribution.

**Implementation in the SDF context.** In our SDF network application, we define the reference input as the cross-sectional median of each characteristic. Each month, we randomly sample 300 firms (approximately 10% of our universe) to form $\mathcal{D}_{\text{background}}$. This provides multiple baseline scenarios while maintaining computational efficiency. For each firm $i$ at time $t$, we compute:

1. Input differences: $\Delta x_{j,i,t} = x_{j,i,t} - \text{median}_k(x_{j,k,t})$
2. Output difference: $\Delta\omega_{i,t} = \omega_{i,t} - \omega_{\text{baseline},t}$
3. DeepSHAP attribution: $\phi_{j,i,t} = \mathbb{E}_{\mathbf{x}^0 \sim \mathcal{D}_t}[C_{\Delta x_{j,i,t}, \Delta\omega_{i,t}} | \mathbf{x}^0]$

We report mean absolute SHAP values across time and firms:

$$\bar{\phi}_j = \frac{1}{T \cdot N} \sum_{t=1}^{T} \sum_{i=1}^{N_t} |\phi_{j,i,t}| \tag{A.8}$$

The resulting SHAP values capture the causal flow of influence from each input characteristic to the final portfolio weights, accounting for: (1) the network weights along all paths from input to output, (2) how input changes are distributed among different neurons in the network, (3) the nonlinear activations (ReLUs) that may block or pass signals depending on activation levels, and (4) the interaction effects between characteristics as they flow through the hidden layers. This provides more meaningful attribution than gradient-based methods, which only capture local sensitivity and can be zero even when inputs are crucial for the final decision due to ReLU saturation. Algorithm 4 summarizes the computational procedure for calculating SHAP values using the DeepLIFT approach.

## A.3 SDF Risk Loadings (Betas)

Without lost of generality we assume that the SDF beta $\beta_{i,t}$ is proportional to $\mathbb{E}[F_{t+1} r^e_{i,t+1}]$ where $F_{t+1} = \sum_{i=1}^{N_t} \omega_{i,t} r^e_{i,t+1}$ is the tangency portfolio return. Following Chen et al. (2024), we estimate the SDF risk loadings $\beta_{i,t} = \mathbb{E}_t[r^e_{i,t+1} F_{t+1}]$ using a feedforward neural network that maps firm characteristics to expected factor loadings:

$$\beta_{i,t} = \mathbb{E}_t[r^e_{i,t+1} F_{t+1}] \approx f(I_{i,t}; \vartheta_\beta) \equiv \widehat{\beta}_{i,t} \tag{A.9}$$

where $F_{t+1} = \sum_{i=1}^{N_t} \widehat{\omega}_{i,t} r^e_{i,t+1}$ is the SDF portfolio return from a given model.

For each asset $i$ at time $t$, we compute the target variable:

$$y_{i,t+1} = r^e_{i,t+1} \cdot \widehat{F}_{t+1} \tag{A.10}$$

We then train a neural network to predict $y_{i,t+1}$ using firm characteristics $I_{i,t}$ as inputs. The network architecture consists of three hidden layers with 32, 16, and 8 units respectively, using

ReLU activations and 5% dropout. The network parameters $\vartheta_\beta$ are estimated by minimizing:

$$\mathcal{L}(\vartheta_\beta) = \frac{1}{|\mathcal{I}|} \sum_{(i,t)\in\mathcal{I}} \left( y_{i,t+1} - \widehat{\beta}_{i,t} \right)^2 \tag{A.11}$$

where $\mathcal{I}$ denotes the set of valid (asset, time) pairs. We use the Adam optimizer with learning rate $10^{-3}$ and train for 1,024 epochs with early stopping.

Once trained, the network provides estimated risk loadings:

$$\widehat{\beta}_{i,t} = f(I_{i,t}; \vartheta^*_\beta) \tag{A.12}$$

These $\widehat{\beta}_{i,t}$ values capture the expected covariation between each asset's excess return and the tangency portfolio return, conditional on observable firm characteristics.

# B Data Description

For the main empirical analysis we construct a monthly panel of U.S. common stocks listed on the NYSE, AMEX, and NASDAQ spanning January 1972 through November 2023. The dataset includes total returns (adjusted for stock splits and dividends), share prices, shares outstanding, exchange codes, and delisting information, all identified using CRSP's PERMNO. To address survivorship bias, we incorporate delisting returns following the procedure established by Shumway and Warther (1999). Excess returns are computed by subtracting the one-month Treasury bill yield, obtained from Kenneth French's data library, from each stock's return.

We incorporate a wide range of asset characteristics from Chen and Zimmermann (2021). A predictor is retained if it is continuously available from January 1972 to November 2023, has no month with fewer than 800 firms with non-missing values, and reaches a maximum monthly coverage above 2,000 firms. This filtering process yields a final set of 73 firm-level characteristics. Table B summarizes the resulting list of available characteristics.

**Algorithm 4:** DeepSHAP Attribution Computation for SDF Portfolios

**Input :** Trained SDF network $\omega(\cdot; \widehat{\vartheta}_\omega)$; firm characteristics $\mathbf{x}_{i,t}$ for firm $i$ at time $t$; background sample size $B = 300$

**Initialize:** $\boldsymbol{\phi}_{i,t} \leftarrow \mathbf{0} \in \mathbb{R}^{73}$;

Compute cross-sectional median baseline $\mathbf{x}_t^{\text{median}} \leftarrow \text{median}_k(\mathbf{x}_{k,t})$;

Sample background distribution $\mathcal{S}_t \leftarrow$ randomly select $B$ firms at time $t$;

`/* DeepSHAP computation over background distribution */`

**for** *each background firm* $k \in \mathcal{S}_t$ **do**

  Set reference point $\mathbf{x}^0 \leftarrow \mathbf{x}_{k,t}$;

  `/* Forward pass: compute activations and differences */`

  $\mathbf{y}^{(0)} \leftarrow \mathbf{x}_{i,t}$, $\mathbf{y}^{(0),0} \leftarrow \mathbf{x}^0$;

  **for** *layer* $l = 1$ **to** $L$ **do**

    Compute actual activations $\mathbf{y}^{(l)} \leftarrow f^{(l)}(\mathbf{y}^{(l-1)})$;

    Compute reference activations $\mathbf{y}^{(l),0} \leftarrow f^{(l)}(\mathbf{y}^{(l-1),0})$;

    Compute differences $\boldsymbol{\Delta y}^{(l)} \leftarrow \mathbf{y}^{(l)} - \mathbf{y}^{(l),0}$;

  `/* Backward pass: propagate contribution scores */`

  Initialize output layer contributions $\mathbf{C}^{(L)} \leftarrow \mathbf{1}$;

  **for** *layer* $l = L - 1$ **to** $0$ **do**

    **foreach** *connection from neuron* $j$ *in layer* $l$ *to neuron* $i$ *in layer* $l + 1$ **do**

      **if** *connection is linear* **then**

        $C_{\Delta y_j^{(l)}, \Delta y_i^{(l+1)}} \leftarrow W_{ij} \frac{\Delta y_j^{(l)}}{\sum_k W_{ik} \Delta y_k^{(l)}}$;

      **if** *connection involves ReLU activation* **then**

        $$C_{\Delta y_j^{(l)}, \Delta y_i^{(l+1)}} \leftarrow \begin{cases} 1 & \text{if } y_j^{(l)} > 0 \text{ and } y_j^{(l),0} > 0 \\ 0 & \text{if } y_j^{(l)} \leq 0 \text{ and } y_j^{(l),0} \leq 0 \\ \frac{\Delta y_j^{(l+1)}}{\Delta y_j^{(l)}} & \text{otherwise} \end{cases};$$

  `/* Chain rule: compute input-to-output attributions */`

  **for** *input feature* $j = 1$ **to** $73$ **do**

    Compute total contribution $C_{\Delta x_j, \Delta \omega} \leftarrow \sum_{\text{paths } P} \prod_{(m,n) \in P} C_{\Delta y_m, \Delta y_n}$;

    Accumulate SHAP score $\phi_{j,i,t} \leftarrow \phi_{j,i,t} + \frac{1}{B} \cdot C_{\Delta x_j, \Delta \omega}$;

`/* Aggregate across characteristics (excluding macro factors) */`

Extract firm characteristic attributions $\boldsymbol{\phi}_{i,t}^{\text{firm}} \leftarrow (\phi_{5,i,t}, \ldots, \phi_{73,i,t})$;

**Output:** DeepSHAP attribution scores $\boldsymbol{\phi}_{i,t}^{\text{firm}} \in \mathbb{R}^{73}$ for firm characteristics

Table B.2: **Summary of firm characteristics**. This table reports a summary of the firm-level characteristics $I_{i,t}$ used in the main empirical analysis as input for the SDF network and the instrument network.

| Acronym | Name | Definition | Reference |
|---|---|---|---|
| Accruals | Accruals | Net operating accruals scaled by average total assets | Sloan (1996) |
| AM | Assets to market | Total assets divided by market value of equity | Fama & French (1992) |
| AnnouncementReturn | Earnings announcement return | Sum of excess returns [-1,+2] around earnings announcements | Chan et al. (1996) |
| AssetGrowth | Asset growth | Annual growth rate of total assets | Cooper et al. (2008) |
| BetaLiquidityPS | PS liquidity beta | Monthly excess returns regressed on Pastor's liquidity innovations | Pastor & Stambaugh (2003) |
| BMdec | Book to market (Dec ME) | BM using most recent December market equity | Fama & French (1992) |
| BookLeverage | Book leverage | Total assets divided by equity plus deferred taxes and preferred stock | Fama & French (1992) |
| BPEBM | Leverage component of BM | Difference between book-to-price ratio and EBM | Penman et al. (2007) |
| CashProd | Cash productivity | Market equity minus assets, scaled by cash and short-term investments | Chandrashekar & Rao (2009) |
| CBOperProf | Cash-based operating prof. | Operating cash flow scaled by total assets | Ball et al. (2016) |
| CF | Cash flow to market | Net income plus depreciation divided by market equity | Lakonishok et al. (1994) |
| cfp | Operating CF to price | Operating cash flow scaled by market value of equity | Desai et al. (2004) |
| ChAssetTurnover | Change in asset turnover | Annual change in asset turnover | Soliman (2008) |
| ChEQ | Growth in book equity | Ratio of current-year to prior-year book equity | Lockwood & Prombutr (2010) |
| ChInv | Inventory growth | 12-month change in inventory divided by average total assets | Thomas & Zhang (2002) |
| ChinviA | Change in capex (ind adj) | Capex growth minus industry average, using 2-year average | Abarbanell & Bushee (1998) |
| ChNNCOA | Change in net noncurrent op assets | 12-month change in noncurrent operating assets | Soliman (2008) |
| ChNWC | Change in net working capital | 12-month change in net working capital | Soliman (2008) |
| ChTax | Change in taxes | Annual change in quarterly total taxes, scaled by previous year's assets | Thomas & Zhang (2011) |
| ConvDebt | Convertible debt indicator | Indicator = 1 if common shares reserved for convertible debt > 0 | Valta (2016) |
| CoskewACX | Coskewness (daily returns) | Stock's sensitivity to asymmetric, extreme market movements | Ang et al. (2006) |
| DelCOA | Change in current op assets | Change in current operating assets scaled by average total assets | Richardson et al. (2005) |
| DelCOL | Change in current op liabilities | Change in current operating liabilities scaled by average total assets | Richardson et al. (2005) |
| DelEqu | Change in equity to assets | Change in book equity scaled by average total assets | Richardson et al. (2005) |
| DelFINL | Change in financial liabilities | Change in financial liabilities scaled by average total assets | Richardson et al. (2005) |
| DelLTI | Change in LT investment | Change in investment and advances scaled by average total assets | Richardson et al. (2005) |
| DelNetFin | Change in net financial assets | Annual change in net investments scaled by average total assets | Richardson et al. (2005) |
| DivInit | Dividend initiation | 1 if dividends started in past 6 months after no dividends for 24 months | Michaely et al. (1995) |
| DivOmit | Dividend omission | 1 if consistent dividend payer missed scheduled dividend in past 2 months | Michaely et al. (1995) |
| dNOA | Change in net op assets | Annual growth in net operating assets scaled by previous year's assets | Hirshleifer et al. (2004) |
| DolVol | Past trading volume | Log of (trading volume × price) lagged by two months | Brennan et al. (1998) |
| EarningsSurprise | Earnings surprise | Current EPS growth minus 2-year average EPS growth | Foster et al. (1984) |
| EBM | Enterprise component of BM | Net financial equity to market-adjusted net financial equity | Penman et al. (2007) |
| EquityDuration | Equity duration | Calculated from projected ROE, growth, and cash flows over 10 years | Dechow et al. (2004) |
| GP | Gross profits to total assets | Gross margin (revenue minus COGS) scaled by total assets | Novy-Marx (2013) |
| grcapx | Change in capex (2 years) | 2-year growth in capital expenditures | Anderson & Garcia-Feijoo (2006) |
| Herf | Industry concentration | 3-year rolling average of industry revenue-based Herfindahl index | Hou & Robinson (2006) |
| hire | Employment growth | Yearly change in employee count scaled by 2-year average | Bazdresch et al. (2014) |
| IdioVol3F | Idiosyncratic risk (3-factor) | Value-weighted std dev of residuals from 3-factor regression | Ang et al. (2006) |
| IdioVolAHT | Idiosyncratic risk (AHT) | Standard deviation of CAPM regression residuals | Ali et al. (2003) |
| IndIPO | Initial public offerings | 1 if IPO occurred 6-36 months ago, 0 otherwise | Ritter (1991) |
| IndMom | Industry momentum | Market-value-weighted average of firm-level 6-month returns by industry | Grinblatt & Moskowitz (1999) |
| IntMom | Intermediate momentum | Stock return between months $t$-12 and $t$-6 | Novy-Marx (2012) |
| Investment | Investment to revenue | Capital investment-to-revenue ratio scaled by 36-month rolling average | Titman et al. (2004) |
| InvestPPEInv | Change in PPE and inv/assets | Annual change in PPE and inventory, scaled by previous year's assets | Lyandres et al. (2008) |
| Leverage | Market leverage | Total liabilities divided by market value of equity | Bhandari (1988) |
| LRreversal | Long-run reversal | Stock return between months $t$-36 and $t$-13 | De Bondt & Thaler (1985) |
| MaxRet | Maximum return over month | Maximum of daily returns over the previous month | Bali et al. (2010) |
| Mom12m | Momentum (12-month) | Stock return between months $t$-12 and $t$-1 | Jegadeesh & Titman (1993) |
| Mom12mOffSeason | Momentum w/o seasonal part | Average return in other months over the previous year | Heston & Sadka (2008) |
| Mom6m | Momentum (6-month) | Stock return between months $t$-6 and $t$-1 | Jegadeesh & Titman (1993) |
| MomOffSeason | Off-season LT reversal | Average return in other months over the preceding 2-5 years | Heston & Sadka (2008) |
| MomSeason | Return seasonality yrs 2-5 | Average return in same month over the preceding 2-5 years | Heston & Sadka (2008) |
| MomSeasonShort | Return seasonality last year | Average return in same month in the previous year | Heston & Sadka (2008) |
| NOA | Net operating assets | Net operating assets scaled by previous year's total assets | Hirshleifer et al. (2004) |
| OPLeverage | Operating leverage | Sum of administrative expenses and COGS scaled by total assets | Novy-Marx (2010) |
| PctAcc | Percent operating accruals | Accruals as income minus cash flow scaled by absolute earnings | Hafzalla et al. (2011) |
| Price | Price | Log of absolute value of price | Blume & Husic (1972) |
| RDIPO | IPO and no R&D spending | 1 if R&D expense = 0 and IndIPO = 1, 0 otherwise | Gou et al. (2006) |
| RoE | Return on equity | Net income divided by book equity | Haugen & Baker (1996) |
| ResidualMomentum | Momentum based on FF3 residuals | Past 11-month average of lagged residuals from rolling 36-month FF3 regressions | Blitz et al. (2011) |
| ReturnSkew | Return skewness | Skewness of daily returns over previous month | Bali et al. (2015) |
| ReturnSkew3F | Idiosyncratic skewness (3F) | Monthly skewness of daily idiosyncratic returns from FF regression residuals | Bali et al. (2015) |
| RevenueSurprise | Revenue surprise | Surprise in annual revenue-per-share growth vs past 2-year average | Jegadeesh & Livnat (2006) |
| Streversal | Short-term reversal | Stock return over the previous month | Jegadeesh (1989) |
| ShareIss1Y | Share issuance (1 year) | Growth in split-adjusted shares outstanding from months $t$-18 to $t$-6 | Pontiff & Woodgate (2008) |
| Size | Size | Log of monthly market capitalization | Banz (1981) |
| Tax | Taxable income to income | Ratio of actual taxes paid to statutory taxes due based on net income | Lev & Nissim (2004) |
| TotalAccruals | Total accruals | Net income minus total cash flows, scaled by lagged assets | Richardson et al. (2005) |

*(continued)*

*(continued)*

| Acronym | Name | Definition | Reference |
|---|---|---|---|
| VolSD | Volume variance | 36-month rolling standard deviation of monthly trading volume | Chordia et al. (2001) |
| zerotrade1M | Days with zero trades (1M) | Monthly measure combining zero-trade days and adjusted turnover ratio | Liu (2006) |
| zerotrade6M | Days with zero trades (6M) | 6-month average measure of zerotrade1M | Liu (2006) |
| zerotrade12M | Days with zero trades (12M) | 12-month average measure of zerotrade1M | Liu (2006) |

In addition to individual asset characteristics, we consider 122 macroeconomic variables from the FRED-MD database as documented by McCracken and Ng (2016). These series capture a broad spectrum of economic conditions, including real activity, labor markets, inflation, interest rates, and financial indicators. All series are transformed using the recommended procedures to ensure stationarity. These macroeconomic indicators are used in the main empirical analysis as input for the LSTM cells to extract the latent factors, which are then included as input in the SDF and the instrument networks (see Section 2.1 for a full description)

Table B.3: **Summary of macroeconomic indicators**: This table reports a summary of the 122 macroeconomic indicators used in the main empirical analysis.

| Variable Name | Description | Group | Source |
|---|---|---|---|
| RPI | Real Personal Income | Output and income | FRED-MD |
| W875RX1 | Real personal income ex transfer receipts | Output and income | FRED-MD |
| DPCERA3M086SBEA | Real personal consumption expenditures | Consumption, orders, and inventories | FRED-MD |
| CMRMTSPLx | Real Manu. and Trade Industries Sales | Consumption, orders, and inventories | FRED-MD |
| RETAILx | Retail and Food Services Sales | Consumption, orders, and inventories | FRED-MD |
| INDPRO | IP Index | Output and income | FRED-MD |
| IPFPNSS | IP: Final Products and Nonindustrial Supplies | Output and income | FRED-MD |
| IPFINAL | IP: Final Products (Market Group) | Output and income | FRED-MD |
| IPCONGD | IP: Consumer Goods | Output and income | FRED-MD |
| IPDCONGD | IP: Durable Consumer Goods | Output and income | FRED-MD |
| IPNCONGD | IP: Nondurable Consumer Goods | Output and income | FRED-MD |
| IPBUSEQ | IP: Business Equipment | Output and income | FRED-MD |
| IPMAT | IP: Materials | Output and income | FRED-MD |
| IPDMAT | IP: Durable Materials | Output and income | FRED-MD |
| IPNMAT | IP: Nondurable Materials | Output and income | FRED-MD |
| IPMANSICS | IP: Manufacturing (SIC) | Output and income | FRED-MD |
| IPB51222S | IP: Residential Utilities | Output and income | FRED-MD |
| IPFUELS | IP: Fuels | Output and income | FRED-MD |
| CUMFNS | Capacity Utilization: Manufacturing | Output and income | FRED-MD |
| HWI | Help-Wanted Index for United States | Labor Market | FRED-MD |
| HWIURATIO | Ratio of Help Wanted/No. Unemployed | Labor Market | FRED-MD |
| CLF16OV | Civilian Labor Force | Labor Market | FRED-MD |
| CE16OV | Civilian Employment | Labor Market | FRED-MD |
| UNRATE | Civilian Unemployment Rate | Labor Market | FRED-MD |
| UEMPMEAN | Average Duration of Unemployment (Weeks) | Labor Market | FRED-MD |
| UEMPLT5 | Civilians Unemployed - Less Than 5 Weeks | Labor Market | FRED-MD |
| UEMP5TO14 | Civilians Unemployed for 5-14 Weeks | Labor Market | FRED-MD |
| UEMP15OV | Civilians Unemployed - 15 Weeks & Over | Labor Market | FRED-MD |
| UEMP15T26 | Civilians Unemployed for 15-26 Weeks | Labor Market | FRED-MD |
| UEMP27OV | Civilians Unemployed for 27 Weeks and Over | Labor Market | FRED-MD |
| CLAIMSx | Initial Claims | Labor Market | FRED-MD |
| PAYEMS | All Employees: Total nonfarm | Labor Market | FRED-MD |
| USGOOD | All Employees: Goods-Producing Industries | Labor Market | FRED-MD |
| CES1021000001 | All Employees: Mining and Logging: Mining | Labor Market | FRED-MD |
| USCONS | All Employees: Construction | Labor Market | FRED-MD |
| MANEMP | All Employees: Manufacturing | Labor Market | FRED-MD |
| DMANEMP | All Employees: Durable goods | Labor Market | FRED-MD |
| NDMANEMP | All Employees: Nondurable goods | Labor Market | FRED-MD |
| SRVPRD | All Employees: Service-Providing Industries | Labor Market | FRED-MD |
| USTPU | All Employees: Trade, Transportation & Utilities | Labor Market | FRED-MD |
| USWTRADE | All Employees: Wholesale Trade | Labor Market | FRED-MD |
| USTRADE | All Employees: Retail Trade | Labor Market | FRED-MD |
| USFIRE | All Employees: Financial Activities | Labor Market | FRED-MD |
| USGOVT | All Employees: Government | Labor Market | FRED-MD |
| CES0600000007 | Avg Weekly Hours : Goods-Producing | Labor Market | FRED-MD |
| AWOTMAN | Avg Weekly Overtime Hours : Manufacturing | Labor Market | FRED-MD |
| AWHMAN | Avg Weekly Hours : Manufacturing | Labor Market | FRED-MD |
| HOUST | Housing Starts: Total New Privately Owned | Housing | FRED-MD |
| HOUSTNE | Housing Starts, Northeast | Housing | FRED-MD |
| HOUSTMW | Housing Starts, Midwest | Housing | FRED-MD |
| HOUSTS | Housing Starts, South | Housing | FRED-MD |
| HOUSTW | Housing Starts, West | Housing | FRED-MD |
| PERMIT | New Private Housing Permits (SAAR) | Housing | FRED-MD |
| PERMITNE | New Private Housing Permits, Northeast (SAAR) | Housing | FRED-MD |
| PERMITMW | New Private Housing Permits, Midwest (SAAR) | Housing | FRED-MD |
| PERMITS | New Private Housing Permits, South (SAAR) | Housing | FRED-MD |
| PERMITW | New Private Housing Permits, West (SAAR) | Housing | FRED-MD |
| AMDMNOx | New Orders for Durable Goods | Consumption, orders, and inventories | FRED-MD |
| AMDMUOx | Unfilled Orders for Durable Goods | Consumption, orders, and inventories | FRED-MD |
| BUSINVx | Total Business Inventories | Consumption, orders, and inventories | FRED-MD |
| ISRATIOx | Total Business: Inventories to Sales Ratio | Consumption, orders, and inventories | FRED-MD |
| M1SL | M1 Money Stock | Money and credit | FRED-MD |
| M2SL | M2 Money Stock | Money and credit | FRED-MD |
| M2REAL | Real M2 Money Stock | Money and credit | FRED-MD |
| BOGMBASE | Monetary Base | Money and credit | FRED-MD |
| TOTRESNS | Total Reserves of Depository Institutions | Money and credit | FRED-MD |
| NONBORRES | Reserves Of Depository Institutions | Money and credit | FRED-MD |
| BUSLOANS | Commercial and Industrial Loans | Money and credit | FRED-MD |
| REALLN | Real Estate Loans at All Commercial Banks | Money and credit | FRED-MD |
| NONREVSL | Total Nonrevolving Credit | Money and credit | FRED-MD |
| CONSPI | Nonrevolving consumer credit to Personal Income | Money and credit | FRED-MD |
| S&P 500 | S&P's Common Stock Price Index: Composite | Stock market | FRED-MD |
| S&P div yield | S&P's Composite Common Stock: Dividend Yield | Stock market | FRED-MD |

*(Continued on next page)*

*(Table continued from previous page)*

| Variable Name | Description | Group | Source |
|---|---|---|---|
| S&P PE ratio | S&P's Composite Common Stock: Price-Earnings Ratio | Stock market | FRED-MD |
| FEDFUNDS | Effective Federal Funds Rate | Interest and exchange rates | FRED-MD |
| CP3Mx | 3-Month AA Financial Commercial Paper Rate | Interest and exchange rates | FRED-MD |
| TB3MS | 3-Month Treasury Bill: | Interest and exchange rates | FRED-MD |
| TB6MS | 6-Month Treasury Bill: | Interest and exchange rates | FRED-MD |
| GS1 | 1-Year Treasury Rate | Interest and exchange rates | FRED-MD |
| GS5 | 5-Year Treasury Rate | Interest and exchange rates | FRED-MD |
| GS10 | 10-Year Treasury Rate | Interest and exchange rates | FRED-MD |
| AAA | Moody's Seasoned Aaa Corporate Bond Yield | Interest and exchange rates | FRED-MD |
| BAA | Moody's Seasoned Baa Corporate Bond Yield | Interest and exchange rates | FRED-MD |
| COMPAPFFx | 3-Month Commercial Paper Minus FEDFUNDS | Interest and exchange rates | FRED-MD |
| TB3SMFFM | 3-Month Treasury C Minus FEDFUNDS | Interest and exchange rates | FRED-MD |
| TB6SMFFM | 6-Month Treasury C Minus FEDFUNDS | Interest and exchange rates | FRED-MD |
| T1YFFM | 1-Year Treasury C Minus FEDFUNDS | Interest and exchange rates | FRED-MD |
| T5YFFM | 5-Year Treasury C Minus FEDFUNDS | Interest and exchange rates | FRED-MD |
| T10YFFM | 10-Year Treasury C Minus FEDFUNDS | Interest and exchange rates | FRED-MD |
| AAAFFM | Moody's Aaa Corporate Bond Minus FEDFUNDS | Interest and exchange rates | FRED-MD |
| BAAFFM | Moody's Baa Corporate Bond Minus FEDFUNDS | Interest and exchange rates | FRED-MD |
| EXSZUSx | Switzerland / U.S. Foreign Exchange Rate | Interest and exchange rates | FRED-MD |
| EXJPUSx | Japan / U.S. Foreign Exchange Rate | Interest and exchange rates | FRED-MD |
| EXUSUKx | U.S. / U.K. Foreign Exchange Rate | Interest and exchange rates | FRED-MD |
| EXCAUSx | Canada / U.S. Foreign Exchange Rate | Interest and exchange rates | FRED-MD |
| WPSFD49207 | PPI: Finished Goods | Prices | FRED-MD |
| WPSFD49502 | PPI: Finished Consumer Goods | Prices | FRED-MD |
| WPSID61 | PPI: Intermediate Materials | Prices | FRED-MD |
| WPSID62 | PPI: Crude Materials | Prices | FRED-MD |
| OILPRICEx | Crude Oil, spliced WTI and Cushing | Prices | FRED-MD |
| PPICMM | PPI: Metals and metal products: | Prices | FRED-MD |
| CPIAUCSL | CPI : All Items | Prices | FRED-MD |
| CPIAPPSL | CPI : Apparel | Prices | FRED-MD |
| CPITRNSL | CPI : Transportation | Prices | FRED-MD |
| CPIMEDSL | CPI : Medical Care | Prices | FRED-MD |
| CUSR0000SAC | CPI : Commodities | Prices | FRED-MD |
| CUSR0000SAD | CPI : Durables | Prices | FRED-MD |
| CUSR0000SAS | CPI : Services | Prices | FRED-MD |
| CPIULFSL | CPI : All Items Less Food | Prices | FRED-MD |
| CUSR0000SA0L2 | CPI : All items less shelter | Prices | FRED-MD |
| CUSR0000SA0L5 | CPI : All items less medical care | Prices | FRED-MD |
| PCEPI | Personal Cons. Expend.: Chain Index | Prices | FRED-MD |
| DDURRG3M086SBEA | Personal Cons. Exp: Durable goods | Prices | FRED-MD |
| DNDGRG3M086SBEA | Personal Cons. Exp: Nondurable goods | Prices | FRED-MD |
| DSERRG3M086SBEA | Personal Cons. Exp: Services | Prices | FRED-MD |
| CES0600000008 | Avg Hourly Earnings : Goods-Producing | Labor Market | FRED-MD |
| CES2000000008 | Avg Hourly Earnings : Construction | Labor Market | FRED-MD |
| CES3000000008 | Avg Hourly Earnings : Manufacturing | Labor Market | FRED-MD |
| DTCOLNVHFNM | Consumer Motor Vehicle Loans Outstanding | Money and credit | FRED-MD |
| DTCTHFNM | Total Consumer Loans and Leases Outstanding | Money and credit | FRED-MD |
| INVEST | Securities in Bank Credit at All Commercial Banks | Money and credit | FRED-MD |
| VIXCLSx | VIX | Stock market | FRED-MD |

# C Transaction Costs for Managed Portfolios

For each month $t$ and characteristic $j$, we implement the following procedure. First, we transform each characteristic into cross-sectional percentile ranks within each month, then linearly map these ranks to the interval $[-0.5, +0.5]$. This transformation ensures comparability across characteristics with different scales and distributions while preserving ordinal relationships.

Next, we sort all stocks into two size groups based on the NYSE median market capitalization, then within each size group, we further sort stocks into three groups using the 30th and 70th percentiles as breakpoints for a given characteristic. This yields six portfolios: {SL, SM, SH, BL, BM, BH}, where S/B denotes small/big size groups and L/M/H denotes low/medium/high characteristic groups.

Thus, the factor return for characteristic $j$ at time $t+1$ is computed as:

$$r_{j,t+1}^{LS} = \frac{1}{2}(r_{SH,t+1} + r_{BH,t+1}) - \frac{1}{2}(r_{SL,t+1} + r_{BL,t+1}) \tag{C.1}$$

where $r_{p,t+1}$ represents the value-weighted return of portfolio $p$. Within each portfolio, individual stock weights are proportional to their market capitalization for the value-weighted case.

The transaction cost for the long-short portfolio is computed by combining the costs from both legs:

$$TC_{j,t+1}^{LS} = \frac{1}{2}(TC_{SH,t+1} + TC_{BH,t+1}) + \frac{1}{2}(TC_{SL,t+1} + TC_{BL,t+1}) \tag{C.2}$$

Note that transaction costs are additive for both long and short positions since investors incur costs regardless of the direction of their trades. The factor $\frac{1}{2}$ reflects that each leg (long and short) contributes half of the total portfolio weight.

Following Barroso and Detzel (2021), we implement a conservative approach to transaction cost calculation that accounts for the asymmetric nature of long-short portfolios. The key aspect is normalizing the long and short legs separately after accounting for return-driven weight changes. For each portfolio $p \in \{SL, SM, SH, BL, BM, BH\}$ and month $t$, we calculate turnover as:

$$TO_{p,t} = \frac{1}{2}\sum_{i=1}^{N_t} |w_{i,p,t} - \widehat{w}_{i,p,t-1}| \tag{C.3}$$

where $w_{i,p,t}$ is the weight of stock $i$ in portfolio $p$ at time $t$ after rebalancing, and $\widehat{w}_{i,p,t-1} = \left(\sum_{j=1}^{N_t} w_{j,p,t-1}(1+r_{j,t})\right)^{-1} w_{i,p,t-1}(1+r_{i,t})$ is the weight of stock $i$ before rebalancing, adjusted for the return over period $t$ and normalized to ensure weights sum to one within each leg. The transaction cost for portfolio $p$ is then calculated as:

$$TC_{p,t} = \sum_{i=1}^{N_t} |w_{i,p,t} - \widehat{w}_{i,p,t-1}| \cdot c_{i,t} \tag{C.4}$$

where $c_{i,t}$ is the one-way transaction cost (effective half bid-ask spread) of stock $i$ at time $t$, and $\widehat{w}_{i,p,t-1}$ represents the normalized weight after accounting for differential returns across stocks in the portfolio.

The long leg consists of positions in SH and BH portfolios (each with weight 0.5), while the short leg consists of positions in SL and BL portfolios (each with weight -0.5). Following this separate normalization, the transaction cost for the long-short factor is:

$$TC_{j,t}^{LS} = \frac{1}{2}(TC_{SH,t} + TC_{BH,t}) + \frac{1}{2}(TC_{SL,t} + TC_{BL,t}) \tag{C.5}$$

The net-of-costs return for the long-short factor $j$ is:

$$r_{j,t}^{net} = r_{j,t}^{gross} - TC_{j,t}^{LS} \tag{C.6}$$

where $r_{j,t}^{gross} = \frac{1}{2}(r_{SH,t} + r_{BH,t}) - \frac{1}{2}(r_{SL,t} + r_{BL,t})$ is the gross factor return. The net $r_{j,t}^{net}$ and gross $r_{j,t}^{gross}$ form the basis for our maximum squared Sharpe ratio analysis in Section 4.3.

## C.1 Bootstrap Test for Sharpe Ratio Differences

To formally test whether the transaction cost-aware SDFs achieve statistically significant improvements in Sharpe ratios, we employ the robust bootstrap test of Ledoit and Wolf (2008). Consider two investment strategies with excess returns $r_{base,t}$ and $r_{TC,t}$. The difference in Sharpe ratios is:

$$\Delta = SR_{base} - SR_{TC} = \frac{\mu_{base}}{\sigma_{base}} - \frac{\mu_{TC}}{\sigma_{TC}} \tag{C.7}$$

The test constructs a studentized bootstrap confidence interval for $\Delta$ using circular block bootstrap. We implement the one-sided version to test:

$$H_0 : SR_{TC} \leq SR_{base} \qquad \text{vs} \qquad H_1 : SR_{TC} > SR_{base} \tag{C.8}$$

The procedure computes the observed difference $\widehat{\Delta} = \widehat{SR}_{base} - \widehat{SR}_{TC}$ and its HAC standard error $s(\widehat{\Delta})$ using the Parzen kernel. The one-sided test statistic is:

$$d = \frac{\widehat{\Delta}}{s(\widehat{\Delta})} \tag{C.9}$$

For each of 5,000 bootstrap samples using circular block resampling, we compute the centered statistic:

$$\tilde{d}^{*,m} = \frac{\widehat{\Delta}^{*,m} - \widehat{\Delta}}{s(\widehat{\Delta}^{*,m})} \tag{C.10}$$

The one-sided p-value is:

$$\text{PV} = \frac{\#\{\tilde{d}^{*,m} \leq d\} + 1}{M + 1} \tag{C.11}$$

This approach remains valid under heavy-tailed return distributions, accounts for serial correlation and volatility clustering, and provides accurate finite-sample inference without restrictive normality assumptions. The method is computationally intensive and relies on stationarity assumptions for bootstrap validity.

# D Additional Empirical Results

In this Section, we report a series of additional empirical results. We first report the SDF portfolio composition and characteristics for alternative leverage constraints. Next, we report the equivalent of Figure 6 when replacing the absolute rank difference with the Spearman rank correlation (SRC) between consecutive months as an alternative measure of characteristic stability. Finally, we report additional results for the linear SDF specification, the alternative samples, and the model estimation based on the Brandt et al. (2009) transaction cost estimates.

## D.1 SDF Under Alternative Leverage Constraints

Figure D.1 shows that transaction costs affect portfolio composition differently depending on constraint tightness. The top panel reports the results under mild constraints (±10%), whereas the bottom panels report the results for the tight constraints (±2.5%). These corresponds to $\kappa \in \{0.20, 0.05\}$ in Eq.(4), respectively.

Under tight constraints (±2.5%), diversification effects are strongest, with HHI ratios of 1.37 (GAN) and 1.69 (FFN) indicating that frictionless portfolios are substantially more concentrated. Under loose constraints (±10%), turnover reduction intensifies—GAN turnover ratios increase from 1.25 to 1.44 as constraints relax, showing that frictionless portfolios trade even more excessively when given greater flexibility.

Figure D.2 report the difference in portfolio characteristics under the ±10% and ±2.5% position constraints, respectively. Under loose constraints (±10%), characteristic exposures are substantially larger as frictionless models pursue aggressive high-turnover strategies that transaction costs discipline. Under tight constraints (±2.5%), differences become smaller since external constraints naturally limit high-turnover behavior.

**Characteristics turnover vs SDF characteristics.** Figure 6 in Section 4.1 reports the relationship between characteristics turnover and the extent to which transaction costs affect the SDF characteristics. Figure D.3 extends this analysis to alternative leverage constraints.

The results confirm the baseline findings across different constraint specifications. High-turnover characteristics such as short-term reversal (`STreversal`) and return skewness (`ReturnSkew`) consistently cluster in the upper-right quadrant, while stable fundamental characteristics like return-on-equity (`ROE`) and book leverage (`BookLeverage`) appear in the lower-left quadrant.

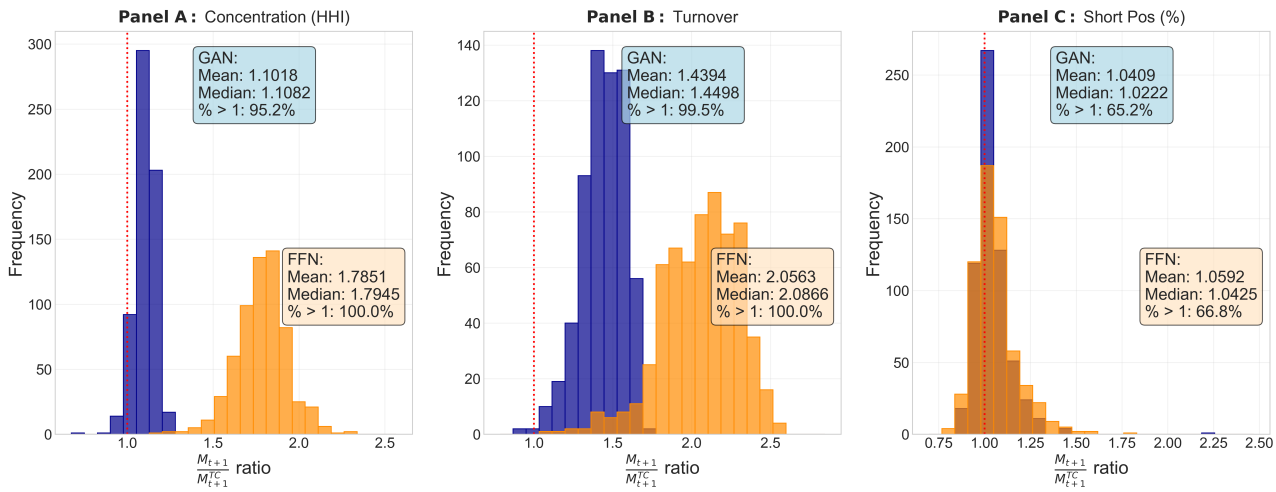
As an alternative stability measure, we also compute the Spearman rank correlation between consecutive months. For the overlapping set $S_t^{(j)} \cap S_{t-1}^{(j)}$, let

$$\rho_t^{(j)} = \text{SpearmanCorr}\Big(\{k_{i,t}^{(j)}\}_{i \in S_t^{(j)} \cap S_{t-1}^{(j)}},\ \{k_{i,t-1}^{(j)}\}_{i \in S_t^{(j)} \cap S_{t-1}^{(j)}}\Big)$$

where $k_{i,t}^{(j)} \in [-0.5, 0.5]$ is the rank-standardized $j$th characteristic for firm $i$ at time $t$. We define the Spearman Rank Correlation ($\text{SRC}^{(j)}$) for characteristic $j$ as

$$\text{SRC}^{(j)} = 1 - \frac{1}{T-1} \sum_{t=2}^{T} \rho_t^{(j)}. \tag{D.1}$$

**Panel A**: Milder leverage constraints $\omega_{it} \in (-10\%, +10\%)$.




**Panel B**: Tighter leverage constraints $\omega_{it} \in (-2.5\%, +2.5\%)$.

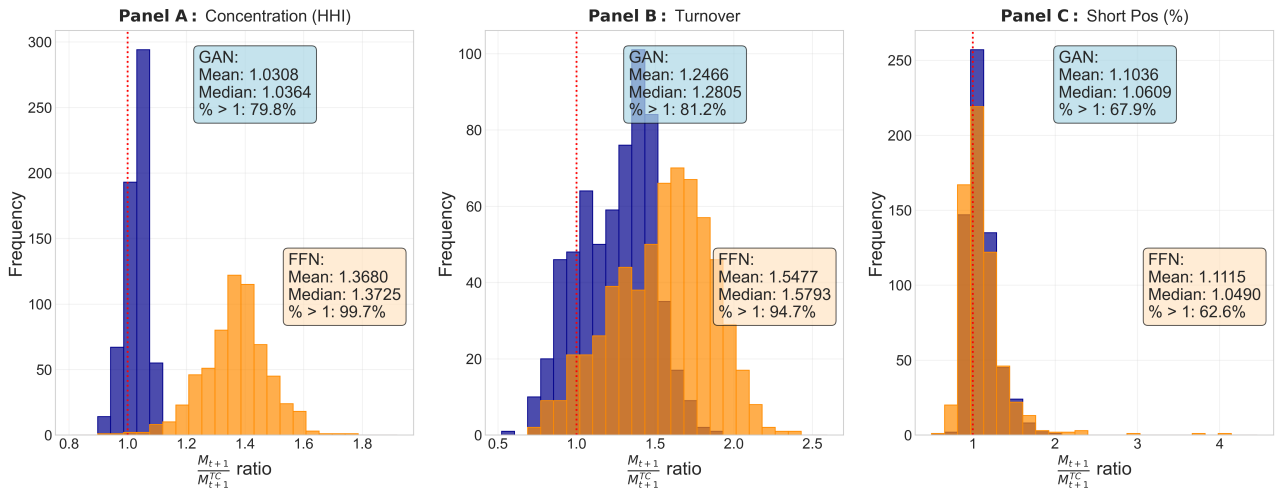
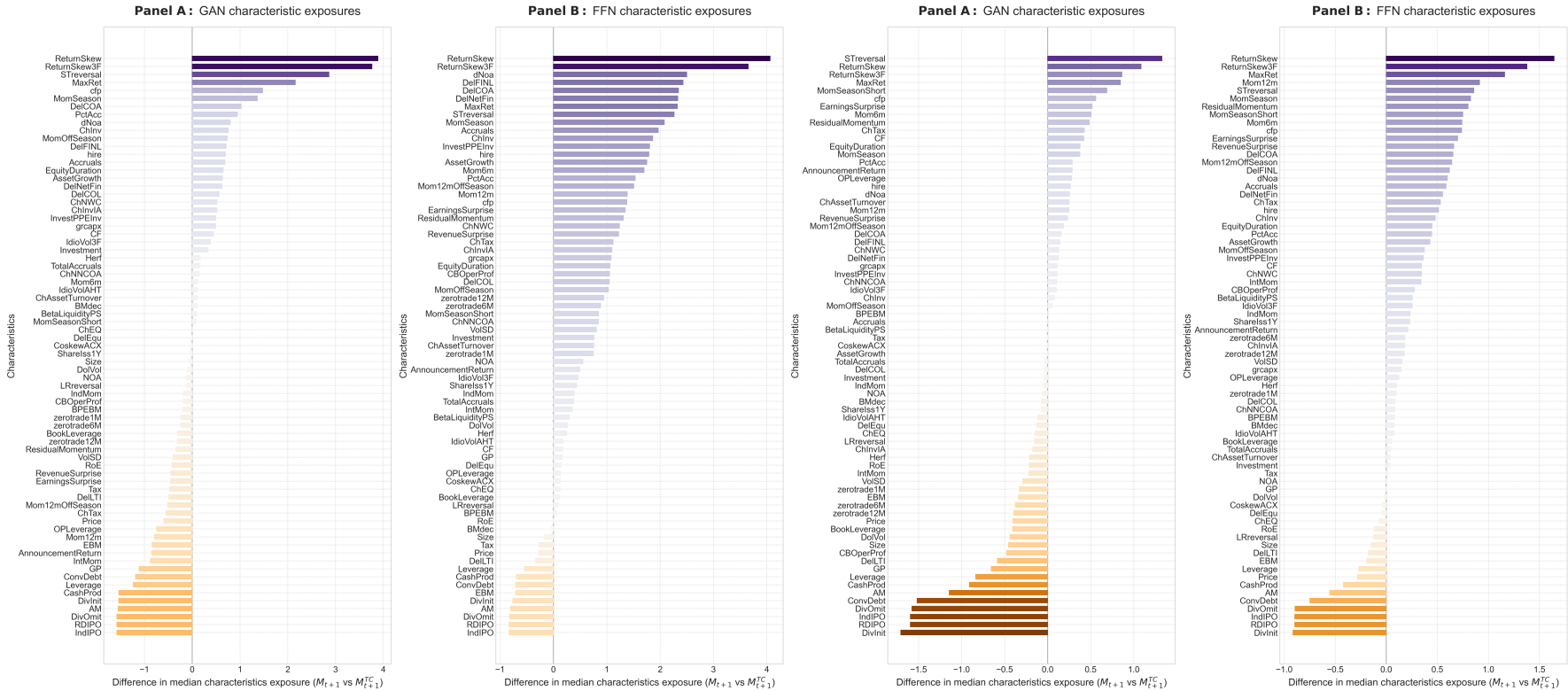



Figure D.1: **SDF portfolio composition for different leverage constraints**. This figure reports the ratio of three main descriptive statistics between the frictionless SDF ($M_{t+1}$) and the transaction-cost-aware SDF ($M^{TC}_{t+1}$) for different leverage constraints. The left panel reports the ratio of the HHI indexes, the middle panel the ratio of the aggregate turnover, and the right panel the ratio of the fraction of short sales in the tangency portfolio. We report the results for both the FFN (orange bars) and the GAN (blue bars). The results are for the out-of-sample period from January 1997 to December 2023.

Higher $\text{SRC}^{(j)}$ values indicate greater ranking instability and thus higher rebalancing costs.

Figure D.4 confirms the robustness of our main findings using this alternative measure. Finally, Figure D.5 expands on the definition of transaction costs in Section C and plots the average transaction costs required to construct characteristic-managed portfolios against their SDF reallocation. This helps to establish a more direct link between implementation costs and reduced SDF influence.

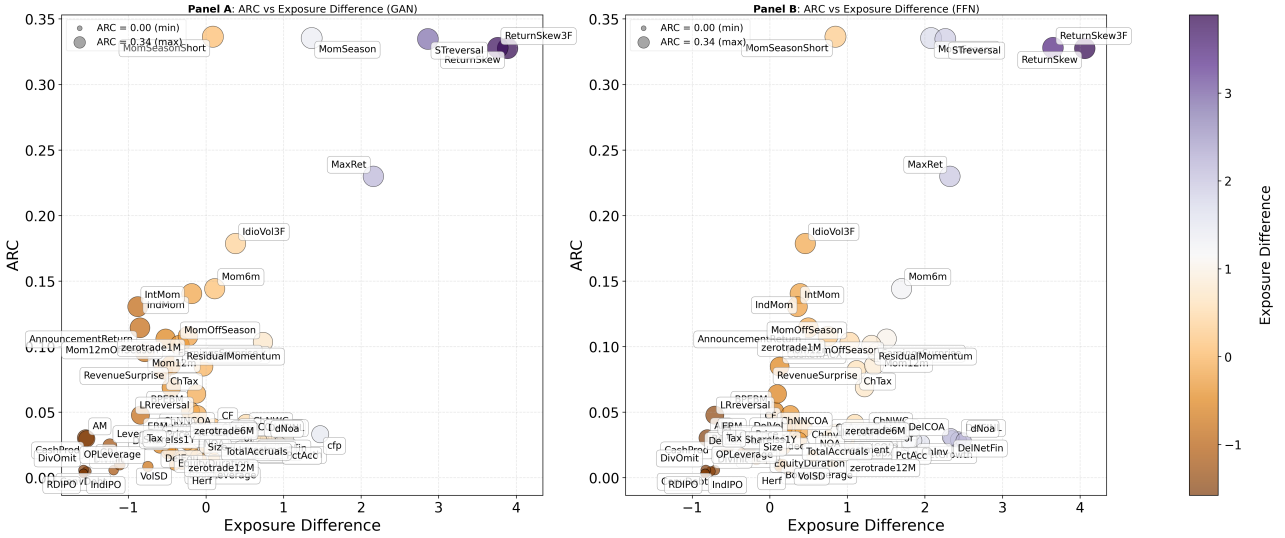



(a) Milder leverage constraints $\omega_{it} \in (-10\%, +10\%)$

(b) Tighter leverage constraints $\omega_{it} \in (-2.5\%, +2.5\%)$

Figure D.2: **SDF characteristics exposure difference under alternative leverage constraints.** This figure reports the difference between frictionless and TC-aware SDFs in the time-series median SDF portfolio characteristics. The left panel reports the results for the GAN estimates, whereas the right panel reports the results for the FFN estimates. The out-of-sample period is from January 1997 to November 2023.

**Panel A**: Milder leverage constraints $\omega_{it} \in (-10\%, +10\%)$.




**Panel B**: Tighter leverage constraints $\omega_{it} \in (-2.5\%, +2.5\%)$.

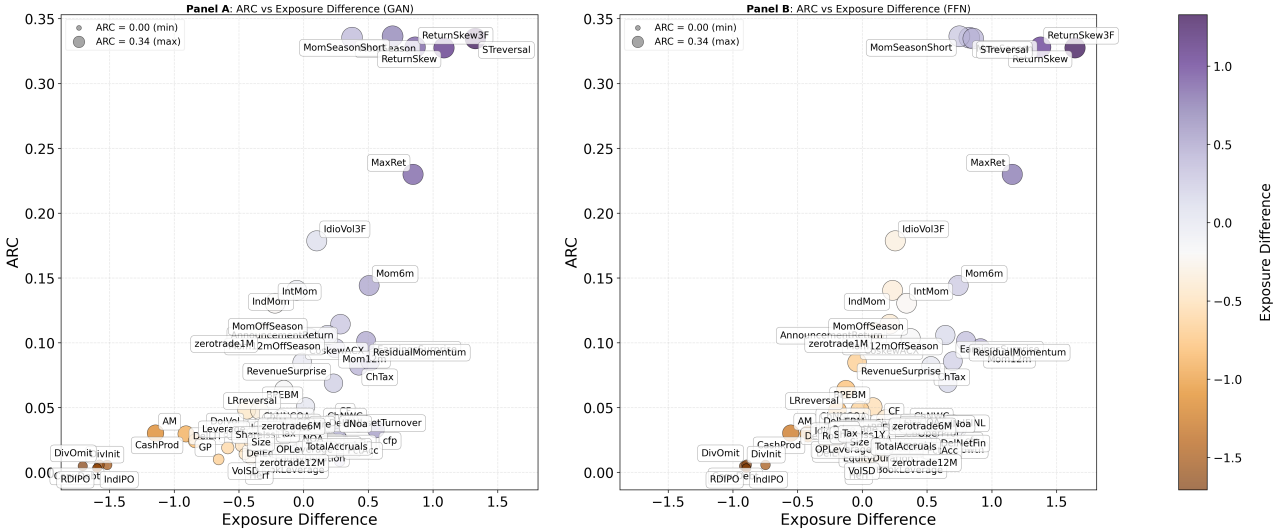



Figure D.3: **Characteristics turnover vs SDF reallocation with alternative leverage constraints.** This figure illustrates the relationship between characteristic turnover (ARC, y-axis) and the reduction in SDF exposure resulting from the inclusion of transaction costs (x-axis). The left panels present the results for the FFN estimates, while the right panels display the results for the GAN estimates. We report the alternative leverage constraints of $\omega_{it} \in (-10\%, +10\%)$ (Panel A) and $\omega_{it} \in (-2.5\%, +2.5\%)$ (Panel B). The results are for the out-of-sample period from January 1997 to November 2023.

## D.2 SDF Portfolio Performance

Figure D.6 reports the cumulative return spread between the TC-aware and the frictionless SDFs under the $\omega_{i,t} \in (-10\%, 10\%)$ (Panel A) and $\omega_{i,t} \in (-2.5\%, 2.5\%)$ (Panel B) leverage constraints. TC-aware SDFs underperform on a gross basis (black solid lines), but this relationship reverses when transaction costs are considered (dashed gray lines). For the GAN

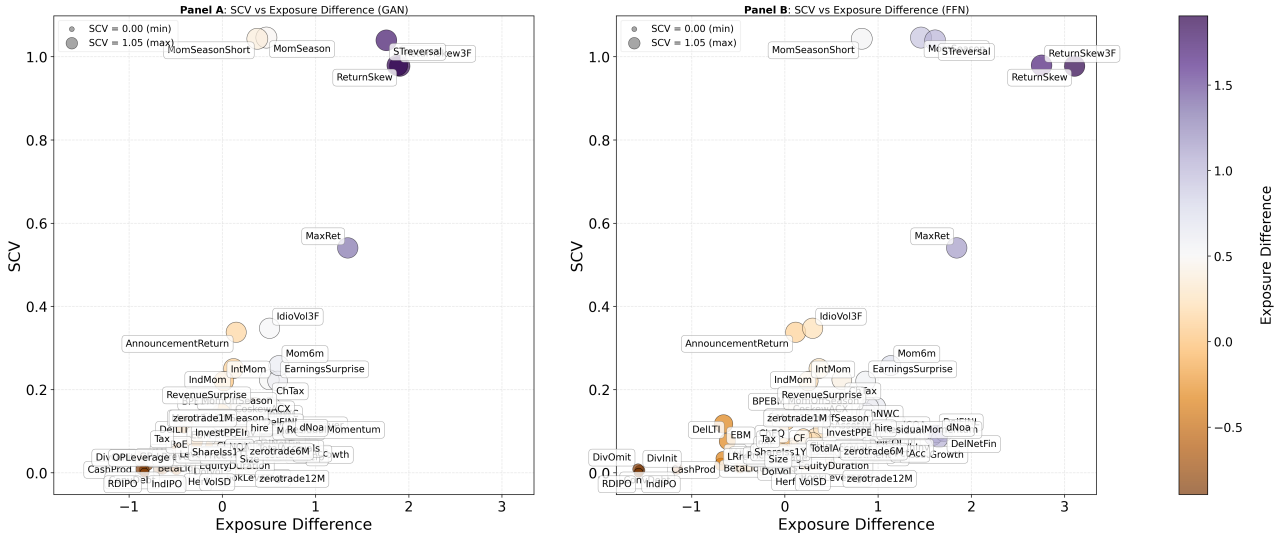



Figure D.4: **Rank correlation vs SDF reallocation.** This figure plots the Spearman rank correlation (SRC) as in Eq.(D.1) against the difference in SDF characteristic exposure when including transaction costs ($M_{t+1}$ vs $M_{t+1}^{TC}$). The left panel reports the results for the GAN estimates whereas the right panel reports the results for the FFN estimates. We focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The results are for the out-of-sample period from January 1997 to December 2023.

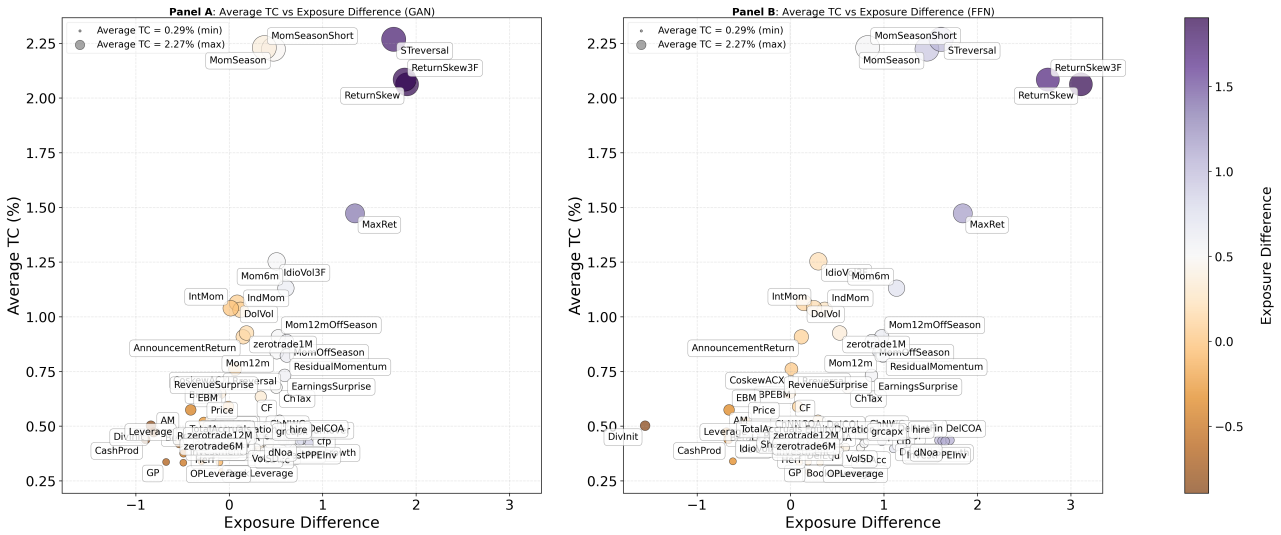
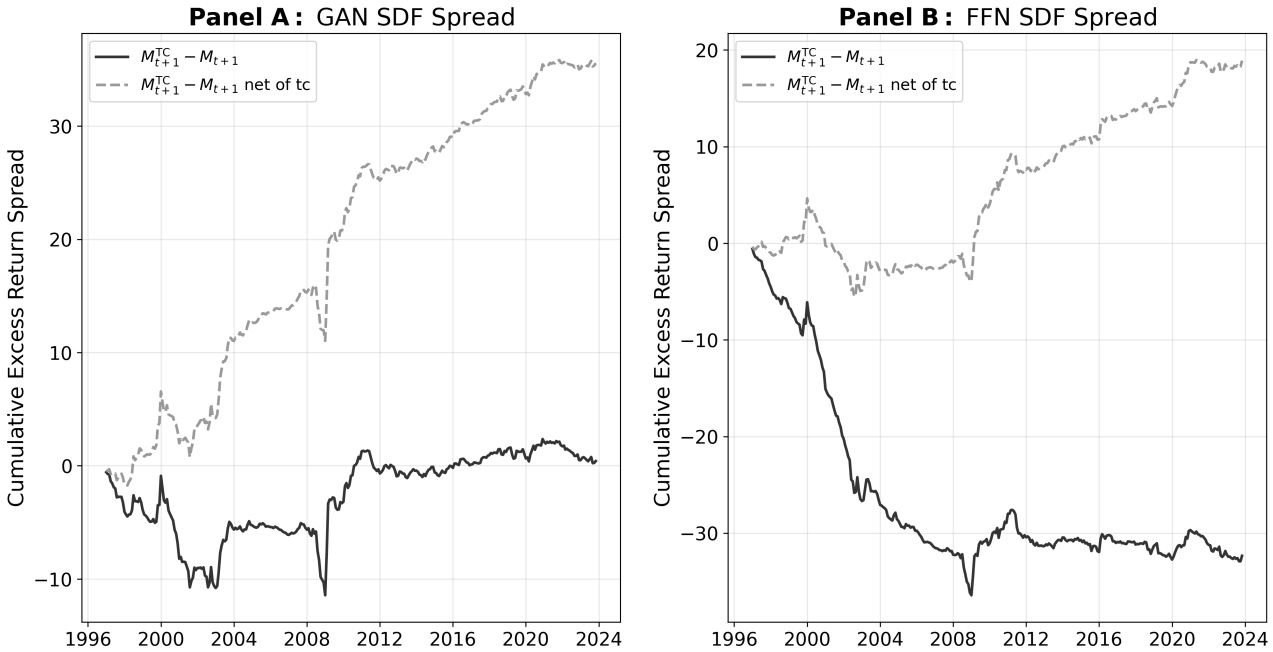



Figure D.5: **Transaction costs vs SDF reallocation.** This figure plots the average transaction costs from Section C against the difference in SDF characteristic exposure when including transaction costs ($M_{t+1}$ vs $M_{t+1}^{TC}$). The left panel presents the results for the GAN estimates, while the right panel displays the results for the FFN estimates. We focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The results are for the out-of-sample period from January 1997 to December 2023.

architecture (left panel), the net performance of $M_{t+1}^{TC}$ significantly exceeds that of $M_{t+1}$ over the out-of-sample period. This performance gap is even more pronounced for a conventional

neural network (right panel).

**Panel A**: Leverage constraints $\omega_{i,t} \in (-10\%, 10\%)$.




**Panel B**: Leverage constraints $\omega_{i,t} \in (-2.5\%, 2.5\%)$.

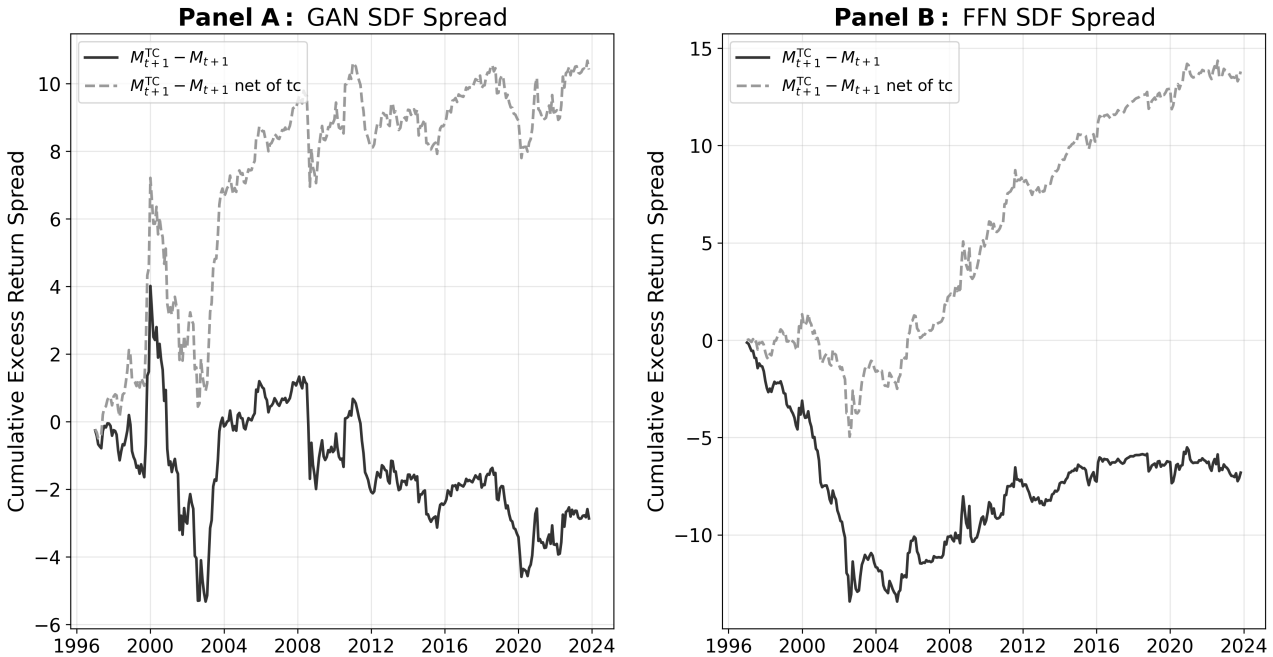



Figure D.6: **Cumulative SDF returns spread**. This figure shows cumulative log return spread between $M^{TC}_{t+1}$ and $M_{t+1}$ over the out-of-sample period (January 1997 to November 2023). The left panel reports the results from the GAN, and the right panel reports the results for the FFN architectures. The results pertain to the ±10% (Panel A) and ±2.5% (Panel B) leverage constraints.

**Beta-sorted portfolios.** Figure D.7 reports the cumulative returns of the beta-sorted portfolios. We note that the strength of this cross-sectional predictability varies meaningfully across leverage constraints, with clearer decile separation emerging under looser constraints where models have greater flexibility to exploit risk-return relationships.

The beneficial effect of transaction costs on low-beta portfolios (the improvement in dashed vs solid lines for Decile 1) is most pronounced under looser leverage constraints, where there is more scope for the frictionless model to make costly trading that the TC-aware model avoids. This pattern reinforces our earlier finding that transaction costs provide the greatest economic value precisely when investors have maximum flexibility.

## D.3 Linear SDF Specification

Figure D.8 compares portfolio composition for linear SDFs estimated with and without transaction costs. The results reveal that while transaction costs affect both linear and nonlinear models, the mechanisms differ fundamentally. Linear models exhibit more narrow responses to transaction costs. Both GAN and FFN architectures show more modest changes in concentration and turnover compared to their nonlinear counterparts. This explains why nonlinear models show larger benefits from transaction cost awareness. Nevertheless, transaction cost awareness does reduce the SDF portfolio turnover even in linear specifications.

Figure D.9 shows that linear TC-aware SDFs exhibit diffuse reallocation across many characteristics compared to frictionless SDFs, while nonlinear SDFs (Figure 5) concentrate their adjustments on fewer characteristics.

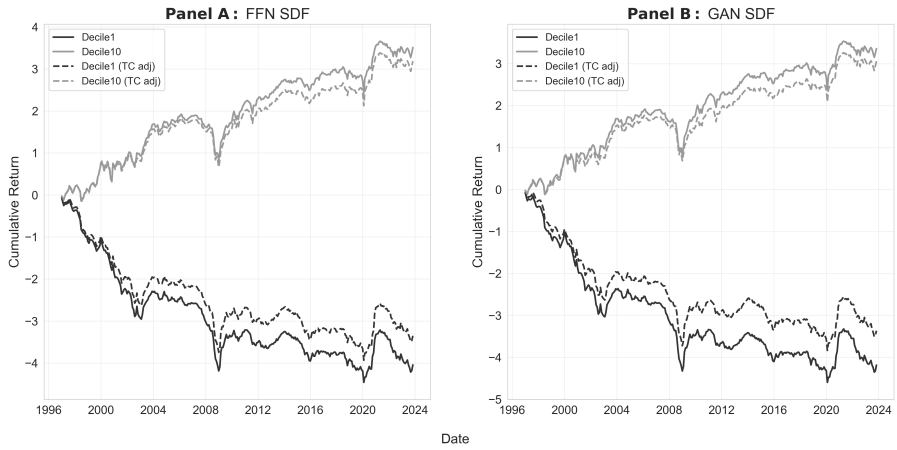
**Limits to arbitrage and economic restrictions.** To ensure that our findings are not merely artifacts of neural network flexibility, we examine the performance of transaction cost-aware SDFs under a strictly linear specification. We leverage the neural networks' architectural flexibility to impose linear constraints on the pricing kernel while maintaining the same input characteristics and estimation framework as our baseline non-linear approach.

Table D.4 reveals several key patterns. First, the linear specification performs substantially worse than the non-linear baseline across all metrics. Explained variation decreases by approximately 50% (from 0.098 to 0.051 for TC-aware GAN in the full sample), and cross-sectional $R^2$ drops from 0.223 to 0.109. This substantial performance degradation confirms that non-linear interactions among firm characteristics and macroeconomic variables are crucial for effective cross-sectional pricing.

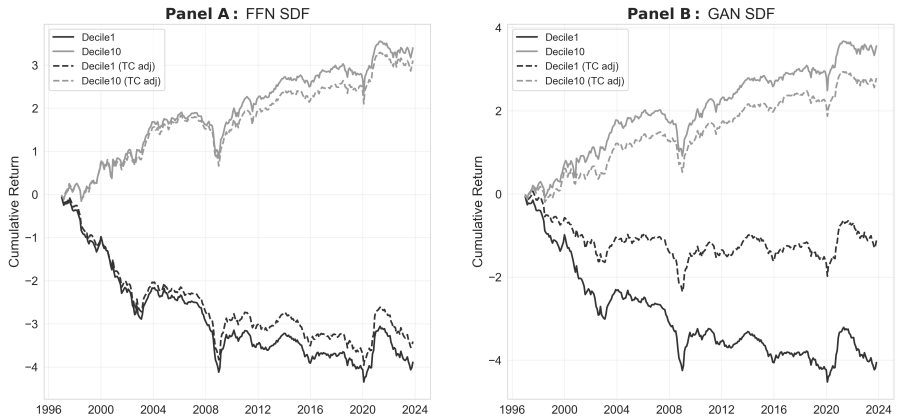
Second, despite this overall performance decline, transaction cost-aware SDFs consistently outperform their frictionless counterparts even under linear constraints. The improvement patterns mirror those observed in the non-linear specification: net Sharpe ratios improve meaningfully (from 0.028 to 0.084), and benefits are particularly pronounced during market stress periods.

Panel B of Table D.4 examines performance across different market conditions under the linear specification. The asymmetric benefits of transaction cost awareness persist. During tight financial conditions, TC-aware SDFs substantially improve net Sharpe ratios (from $-0.157$ to $-0.026$ for GAN), while the improvements during loose conditions are more modest (from 0.054 to 0.104). This pattern reinforces our main finding that transaction cost awareness is most valuable precisely when trading frictions are highest and arbitrage is most limited.

**Panel A**: Leverage constraints $\omega_{i,t} \in (-5\%, 5\%)$.




**Panel B**: Leverage constraints $\omega_{i,t} \in (-2.5\%, 2.5\%)$.




**Panel C**: Leverage constraints $\omega_{i,t} \in (-10\%, 10\%)$.

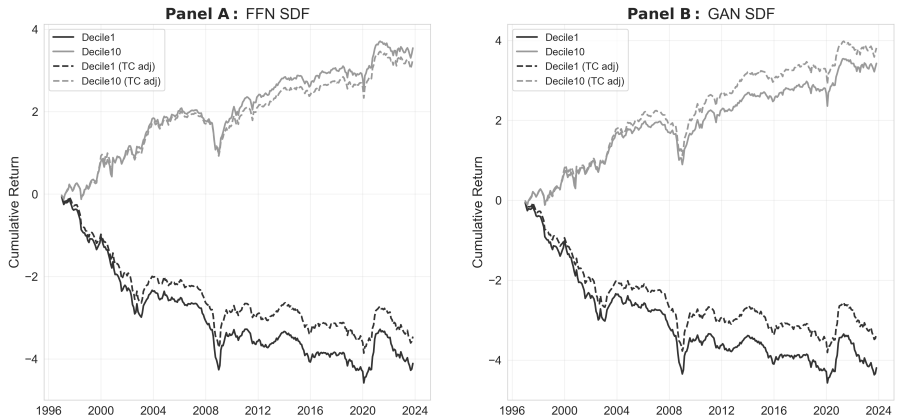



Figure D.7: **Beta sorted portfolios for different leverage constrains**. This figure shows cumulative returns for extreme decile portfolios (Decile 1: lowest beta, Decile 10: highest beta) sorted on model-implied $\beta_{i,t}$ over January 1997 to November 2023. The top panel reports the results for the tighter constraints ($\pm 5\%$), whereas the middle and bottom panel reports the results for the tighter ($\pm 2.5\%$) and looser constraints ($\pm 10\%$), respectively. Solid lines show frictionless SDF results, dashed lines show TC-aware SDF results.

**Panel A**: Market Capitalization Effects

| | Explained Variation | | | | XS-$R^2$ | | | | Sharpe Ratio (Raw) | | | | Sharpe Ratio (Net) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ |
| Full Sample | 0.047 | 0.051 | 0.056 | 0.050 | 0.116 | 0.109 | 0.127 | 0.121 | 0.256 | 0.268 | 0.281 | 0.301 | 0.028 | **0.084** | 0.051 | **0.141** |
| Excl. Microcap | 0.096 | 0.113 | 0.103 | 0.121 | 0.173 | 0.183 | 0.177 | 0.190 | 0.178 | **0.207** | 0.196 | 0.243 | 0.010 | **0.070** | 0.028 | **0.125** |

**Panel B**: Financial Conditions and Market Uncertainty

| | Tight Fin. Cond. | | | | Loose Fin. Cond. | | | | High VIX | | | | Low VIX | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | | GAN | | FFN | |
| | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ | $M_{t+1}$ | $M_{t+1}^{TC}$ |
| Sharpe Ratio (Raw) | 0.096 | 0.166 | 0.126 | 0.206 | 0.279 | 0.289 | 0.305 | 0.327 | 0.302 | 0.313 | 0.332 | 0.369 | 0.183 | 0.199 | 0.202 | 0.185 |
| Sharpe Ratio (Net) | -0.157 | -0.026 | -0.120 | **0.053** | 0.054 | **0.104** | 0.077 | **0.163** | 0.059 | **0.113** | 0.087 | **0.191** | -0.049 | 0.022 | -0.032 | **0.041** |
| Explained Variation | 0.067 | 0.072 | 0.070 | 0.077 | 0.043 | 0.047 | 0.047 | 0.052 | 0.052 | 0.057 | 0.056 | 0.062 | 0.032 | 0.034 | 0.035 | 0.038 |
| XS-$R^2$ | 0.040 | 0.044 | 0.040 | 0.049 | 0.113 | 0.103 | 0.117 | 0.123 | 0.083 | 0.085 | 0.090 | 0.098 | 0.086 | 0.071 | 0.086 | 0.086 |

Table D.4: **Transaction cost-aware SDF performance under linear specification**. This table reports performance metrics for stochastic discount factors under the linear specification with portfolio weight constraints of $\omega_{it} \in (-5\%, 5\%)$. Panel A examines market capitalization effects by comparing the full cross-section against a sample excluding the bottom 25% of stocks by market cap. Panel B analyzes performance across different market conditions: financial conditions (tight vs. loose based on NFCI) and market uncertainty (high vs. low VIX). Results are shown for both GAN and FFN architectures. We highlight in **red** the Sharpe ratio differentials (TC-aware minus frictionless SDF), which are statistically significant at a 5% confidence level. P-values are calculated based on the robust bootstrap test of Ledoit and Wolf (2008). $M_{t+1}$ denotes the standard frictionless SDF, while $M_{t+1}^{TC}$ represents the transaction cost-aware SDF. Sample periods: January 1972 to December 2023 (full sample and NFCI), January 1990 to December 2023 (VIX). All results are out-of-sample.

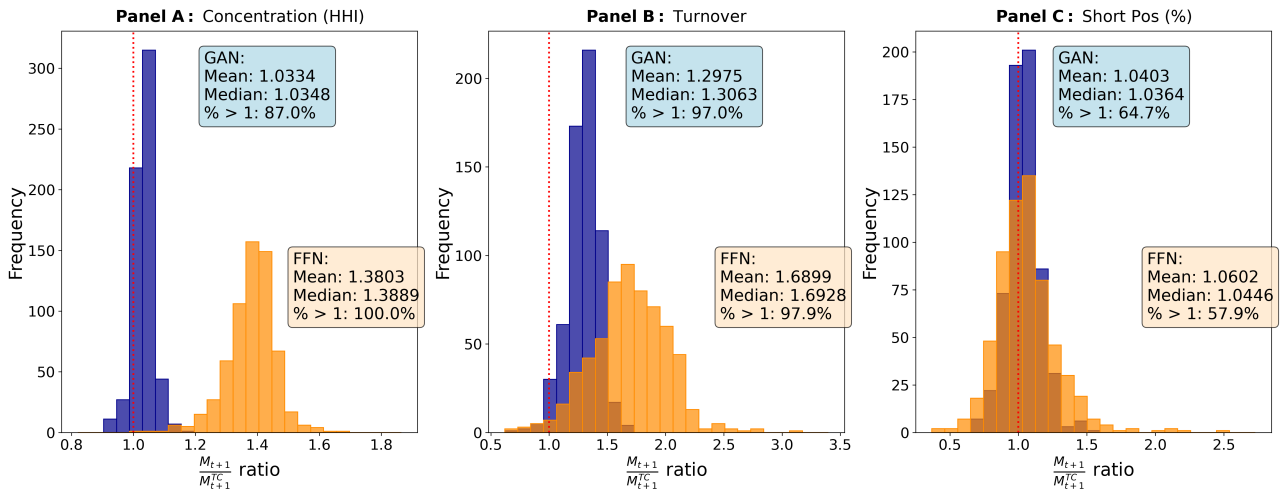



Figure D.8: **SDF portfolio composition for the linear specification**. This figure reports the ratio of three main descriptive statistics between the frictionless SDF ($M_{t+1}$) and the TC-aware SDF ($M_{t+1}^{TC}$) under the linear specification. The left panel reports the ratio of the HHI indexes, the middle panel the ratio of the aggregate turnover, and the right panel the ratio of the fraction of short sales in the tangency portfolio. We report the results for both the FFN (orange bars) and the GAN (blue bars). We focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The results are for the out-of-sample period from January 1997 to December 2023.

The linear specification also confirms that transaction cost benefits extend beyond the smallest, most illiquid securities. When excluding microcaps, performance improvements persist across all metrics, with net Sharpe ratios improving from 0.010 to 0.070 for GAN. This demonstrates that the benefits of transaction cost awareness are not concentrated solely in hard-to-trade securities but extend throughout the cross-section.

## D.4 Alternative Samples

In this Section, we show the SDF portfolio composition and characteristic exposure based on different sample configurations. We examine two alternative sample configurations that vary this trade-off: a longer sample starting in January 1967 (56 years, but only 62 characteristics available) and a shorter sample starting in January 1986 (37 years, but 88 characteristics available). Both samples maintain our standard 20-year training period and 5-year validation period, with testing periods of January 1991-2023 and January 2011-2023, respectively. Figure D.10 shows that the portfolio composition effects of transaction costs vary significantly across sample periods. Similarly, Figure D.11 in Appendix D.4 show the differential impact of transaction costs on the SDF characteristics. Models trained on the longer sample exhibit greater sensitivity to transaction costs, with pronounced aversion to high-turnover strategies like short-term reversal. In contrast, models trained on the post-1986 sample display more nuanced adjustments: moderate reductions in high-turnover characteristics while increasing weights on lower-frequency signals like dividend policy and IPO activity. These findings demonstrate that transaction cost effects are not static but depend critically on the historical cost environment during model training. Models learn from past market conditions, with those exposed to high-friction periods developing a stronger aversion to costly trading strategies.

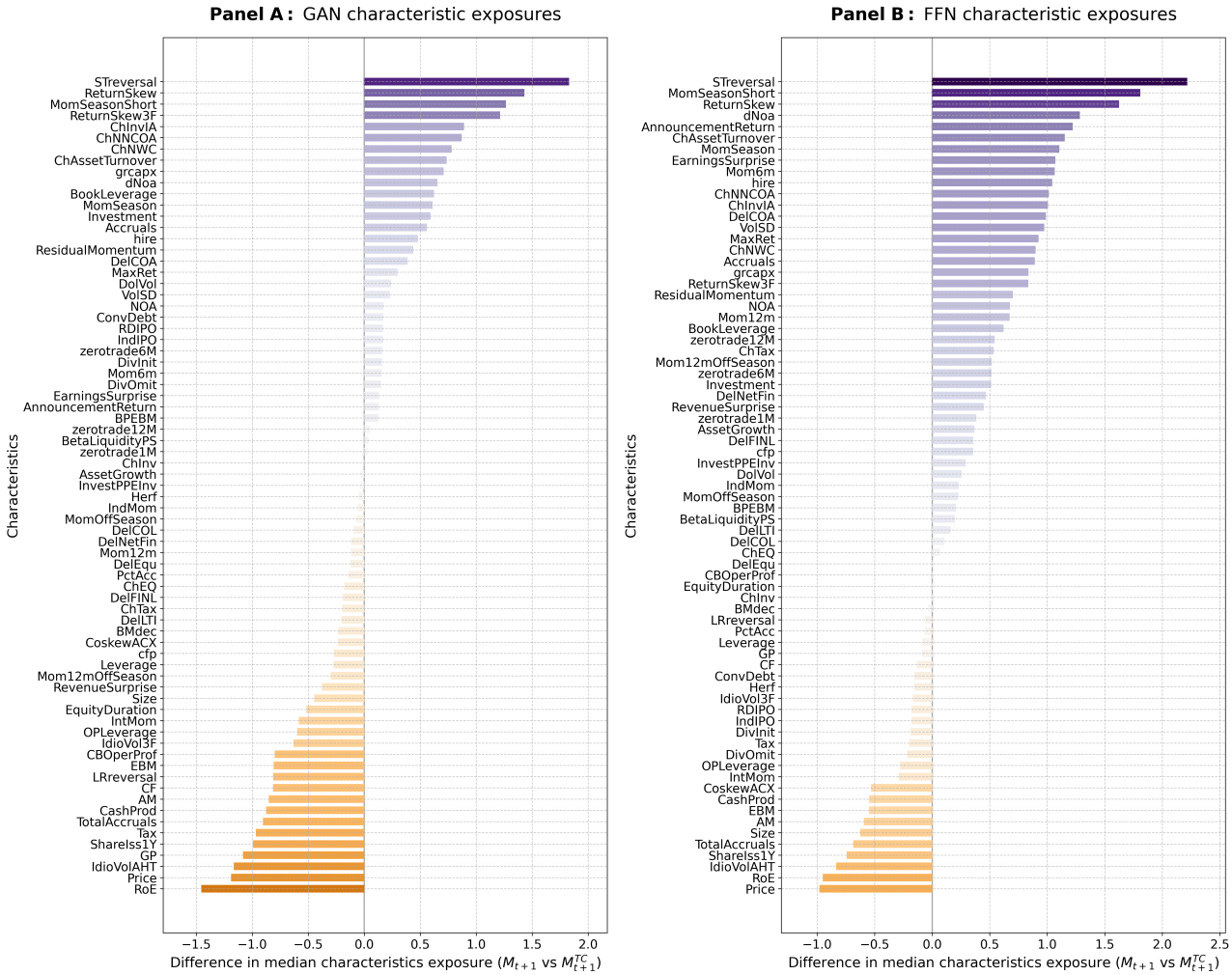
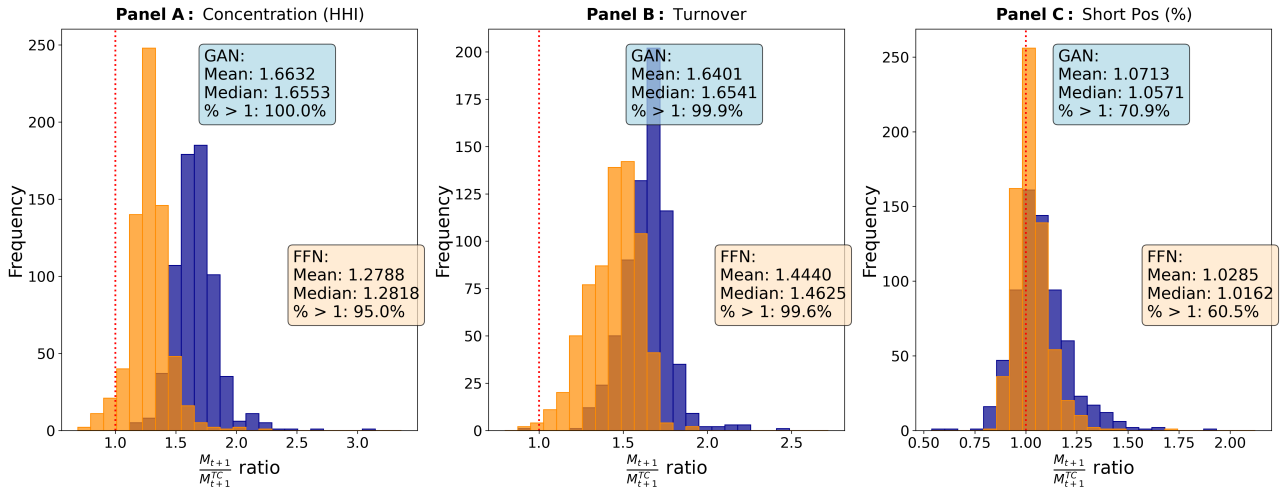



Figure D.9: **SDF characteristic exposure for the linear specification**. This figure reports the difference in the median characteristic exposure as in Eq.(16) between $M_{t+1}$ and $M^{TC}_{t+1}$ portfolios when using a linear specification. The left panel presents the results for the GAN estimates, while the right panel displays the results for the FFN estimates. We focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The results are for the out-of-sample period from January 1997 to December 2023.

**Panel A**: Longer sample (1967-2023)




**Panel B**: Shorter sample (1986-2023)

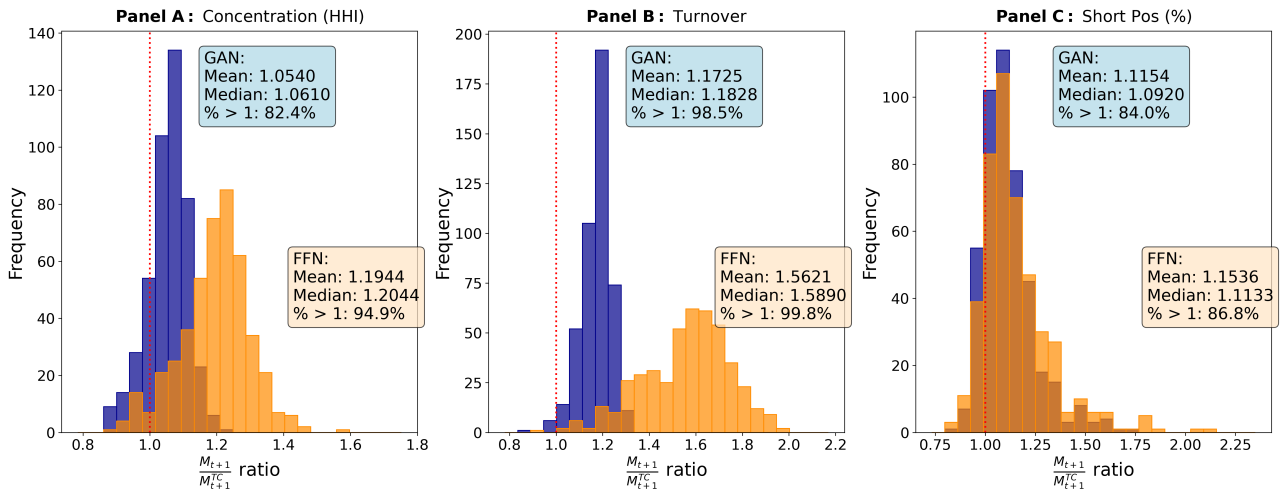


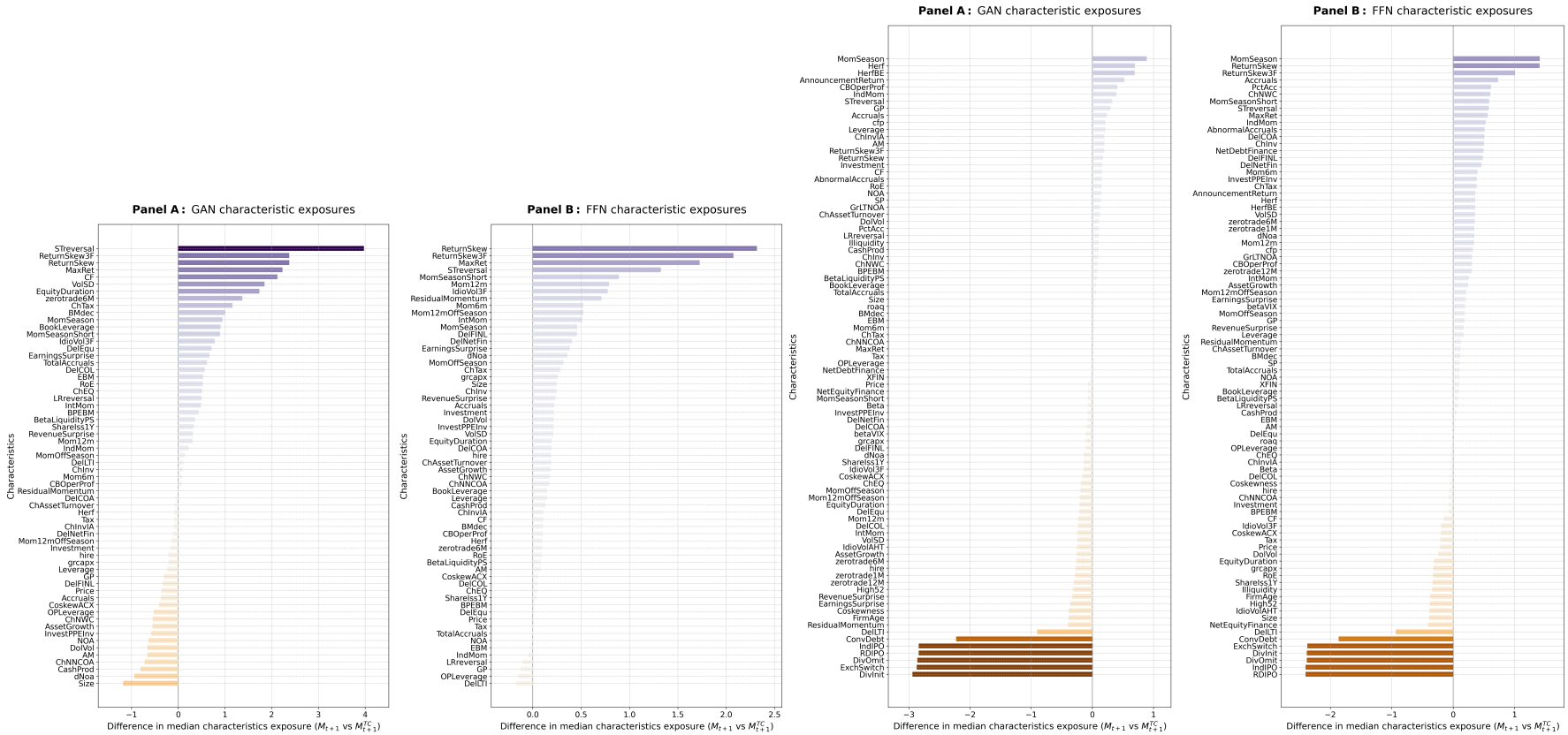

Figure D.10: **SDF portfolio composition for alternative samples**. This figure reports the ratio of three main descriptive statistics between the frictionless SDF ($M_{t+1}$) and the TC-aware SDF ($M^{TC}_{t+1}$) for the different subsamples. The left panels report the ratio of the HHI indexes, the middle panels the ratio of the aggregate turnover, and the right panels the ratio of the fraction of short sales in the tangency portfolio. We report the results for both the FFN (orange bars) and the GAN (blue bars). The results are for the out-of-sample period from January 1991 to December 2023 for the longer sample and from January 2011 to December 2023 for the shorter sample.

## D.5 Alternative Transaction Cost Measures for the SDF

We follow Brandt et al. (2009) and approximate one-way trading costs as $c_{i,t} = 0.006 - 0.0025 \times \text{me}_{i,t}$, where $\text{me}_{i,t}$ represents the market capitalization percentile of firm $i$ (normalized between 0 and 1). Figure D.12 reports the cross-sectional distribution of the transaction costs (in %) calculated over the out-of-sample period. This specification generates transaction costs ranging from 0.6% for the smallest firms to 0.35% for the largest firms, substantially lower than our baseline bid-ask spread estimates.

Figure D.13 shows that transaction costs still promote portfolio concentration (25% HHI

increase for GAN, 12% for FFN) and reduce turnover (22%-37% higher turnover ratios for frictionless models), confirming our main findings even under conservative cost assumptions (see Figure 4 in the main text). Figure D.14 demonstrates that both architectures continue to reduce exposure to high-turnover characteristics while increasing weights on stable characteristics.




(a) Longer sample, less characteristics

(b) Shorter sample, more characteristics

Figure D.11: **SDF characteristic exposure for the longer sample**. This figure reports the difference in the median characteristic exposure as in Eq.(16) between $M_{t+1}$ and $M^{TC}_{t+1}$ portfolios for the longer sampler. We focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is from January 1991 to December 2023.

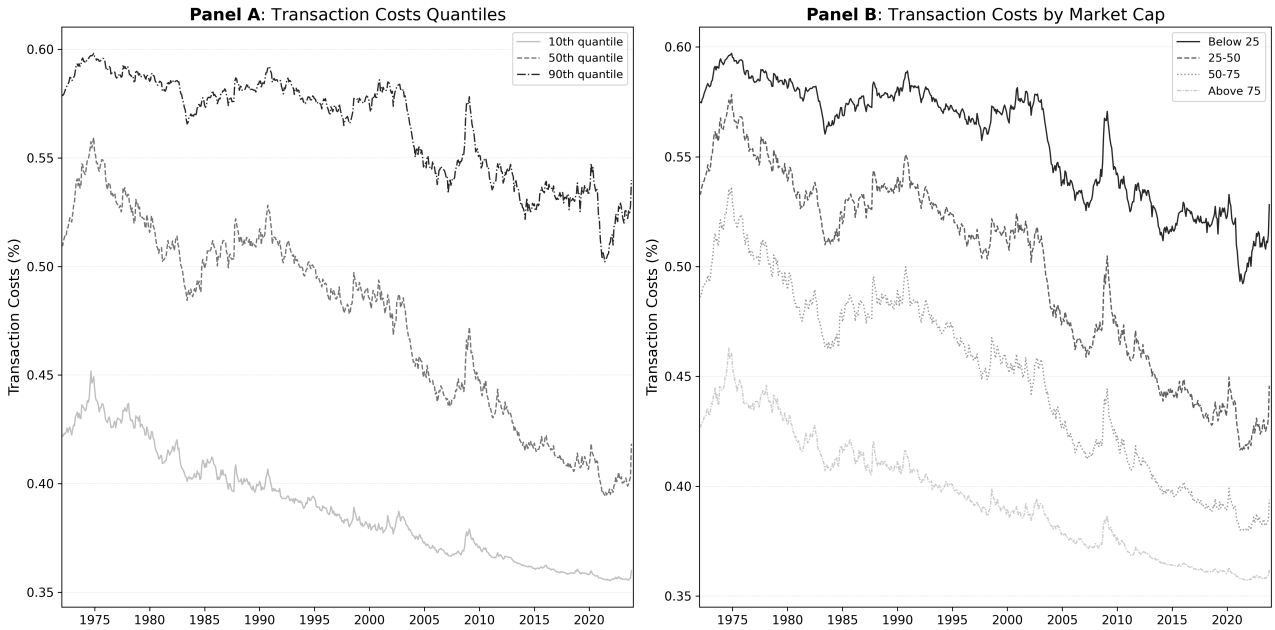



Figure D.12: **Sample variation of transaction costs as in Brandt et al. (2009)** . The left panel reports the 10th, 50th, and 90th percentiles of the cross-sectional distribution of the transaction costs (in %) calculated as in Brandt et al. (2009) over the out-of-sample period. The right panel reports their average (in %) for stocks sorted by market capitalization. The full sample period spans from January 1980 to December 2023.

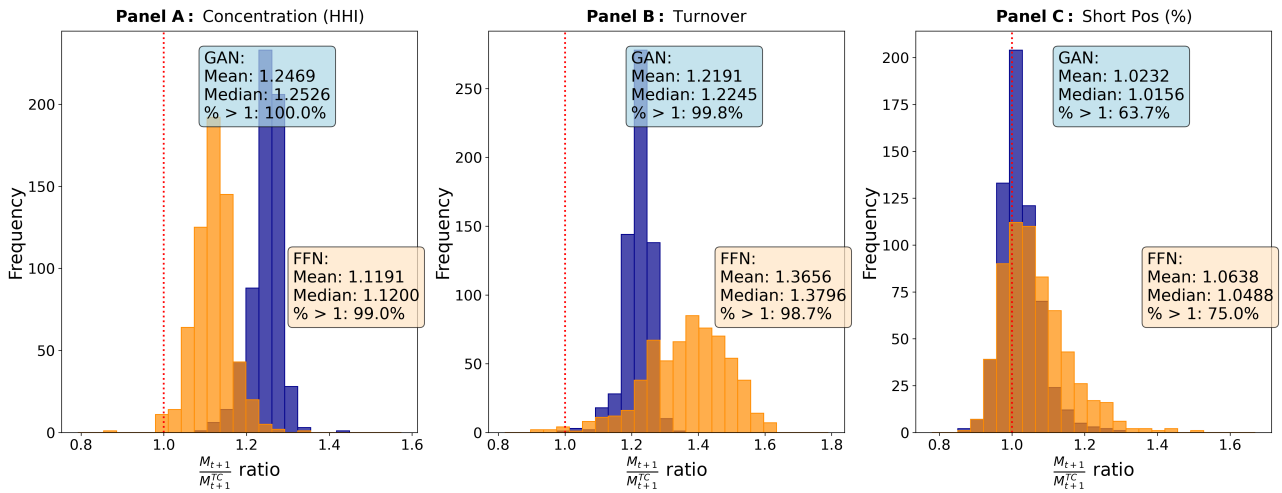



Figure D.13: **SDF portfolio composition with the Brandt et al. (2009) transaction costs**. This figure reports the ratio of three main descriptive statistics between the frictionless SDF ($M_{t+1}$) and the TC-aware SDF ($M^{TC}_{t+1}$) when transaction costs are approximated as in Brandt et al. (2009). The left panel reports the ratio of the HHI indexes, the middle panel the ratio of the aggregate turnover, and the right panel the ratio of the fraction of short sales in the tangency portfolio. We report the results for both the FFN (orange bars) and the GAN (blue bars). We focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The results are for the out-of-sample period from January 1997 to December 2023.

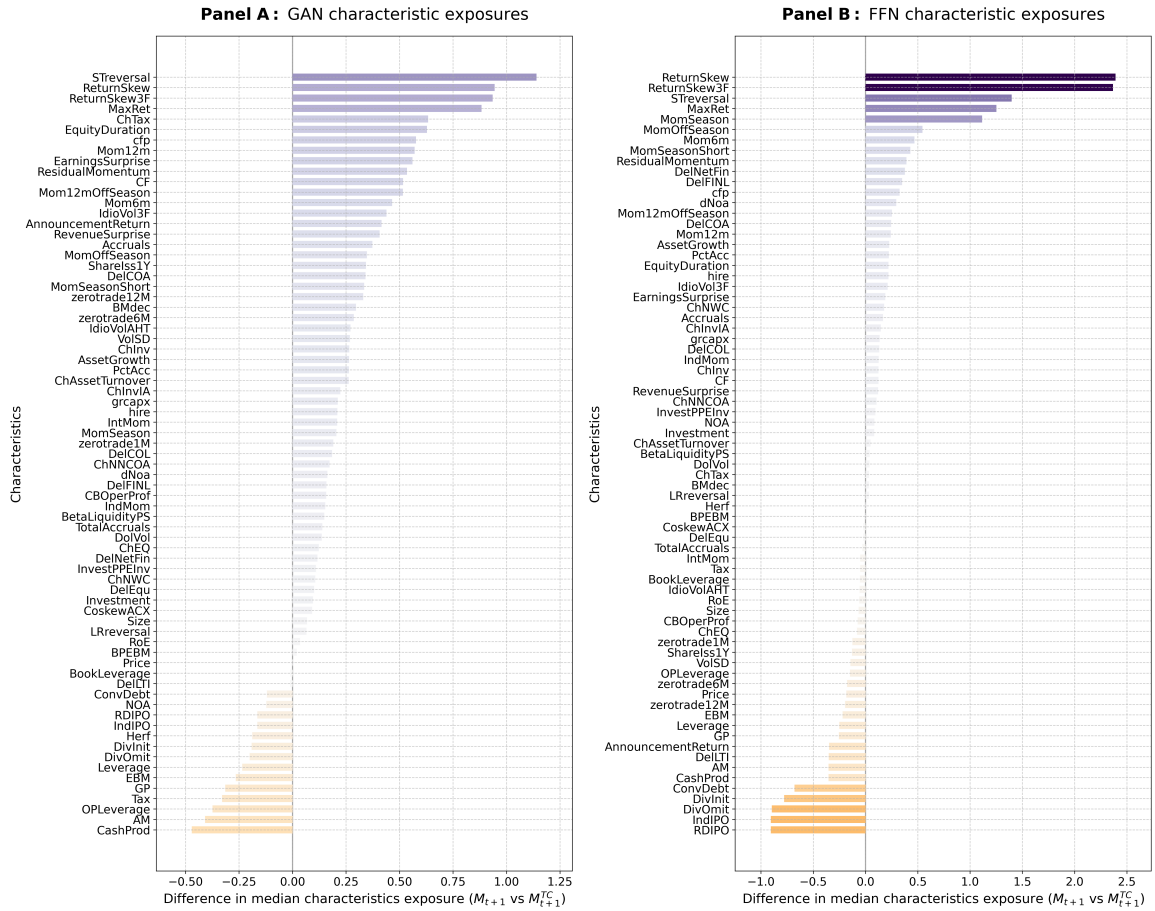



Figure D.14: **SDF characteristic exposure with the Brandt et al. (2009) transaction costs**. This figure reports the difference in the median characteristic exposure as in Eq.(16) between $M_{t+1}$ and $M^{TC}_{t+1}$ when transaction costs are approximated as in Brandt et al. (2009). We focus on the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is from January 2011 to December 2023.