Transaction Costs and the Stochastic Discount Factor*

Daniele Bianchi
Queen Mary,
University of London
d.bianchi@qmul.ac.uk

Teng Jiao
Queen Mary,
University of London
t.jiao@qmul.ac.uk

Hao Ma
Queen Mary,
University of London
hao.ma@qmul.ac.uk

Abstract

Machine-learning estimates of the stochastic discount factor (SDF) often ignore the cost of trading the tangency portfolio that spans them, and assign a premium to high-turnover characteristics that an investor who pays to trade cannot collect. We impose the no-arbitrage condition on the return of the tangency portfolio net of stock-specific trading costs, and estimate the resulting SDF as a nonlinear function of 73 stock characteristics and macroeconomic conditions with adversarial neural networks. Although charging for trading costs restricts the set of tangency portfolios the estimator can reach, the cost-aware SDF prices the cross-section of gross stock returns better out of sample, and the improvement grows with the severity of the costs. The cost-aware tangency portfolio holds less exposure to high-turnover characteristics, and the pricing gain is concentrated in the anomalies whose premia do not survive their own trading costs. Under price impact, the pricing gain grows with the size of the investor, and for a large investor the frictionless tangency portfolio loses money after costs.

Keywords: Stochastic discount factor, Transaction costs, Machine learning, Limits to arbitrage, Price impact.

JEL codes: G12, G14, C45, C58

*This version: October 7, 2026. We are thankful to Zhanhui Chen, Guanhao (Gavin) Feng, Junxiong Gao, Jiangyuan Li (discussant), Albert Menkveld, Mathis Moerke (discussant), Elisa Ossola, Markus Pelger, Seth Pruitt, Landon J. Ross (discussant), Tom Schmitz, Fabio Trojani, and Yu Zheng for their helpful comments and suggestions. We are also grateful to the participants in the 2026 Midwest Finance Association Annual Meeting, the 5th Frontiers of Factor Investing Conference at Lancaster University Management School, the 1st Liechtenstein Workshop on AI in Finance, and the 3rd FinEML at Erasmus University Rotterdam, as well as to seminars at Queen Mary, University of London, the Universitá degli Studi di Milano-Bicocca, and the Free University of Bozen-Bolzano, for their feedback.

1 Introduction

Asset pricing models are judged by how well the stochastic discount factor (SDF) explains the cross-section of expected stock returns. A growing literature estimates the SDF with machine learning methods, either by combining many stock characteristics into a tangency portfolio whose return prices the cross-section (Chen et al., 2024) or by estimating the portfolio weights directly by maximizing its Sharpe ratio (Feng et al., 2026). These methods are often agnostic about trading costs. The tangency portfolio can thus take large positions in high-turnover characteristics whose gross premium disappears, or turns negative, once the cost of that trading is considered (Novy-Marx and Velikov, 2016).

Existing work investigates how frictions reshape the implementable tangency portfolio by letting an investor trade off return against cost (Jensen et al., 2026), and how they change the ranking of factor models whose factors are built without costs (Detzel et al., 2023). Neither estimates the SDF by including cost considerations in a no-arbitrage condition that identifies it. This matters because, when costs bind, the pricing kernel differs from its frictionless counterpart (Korsaye et al., 2025), and thus the characteristics that matter for pricing may also differ. We ask whether a cost-aware pricing kernel prices the cross-section better than one built from the frictionless tangency portfolio, and why.

To answer this question, we embed stock-specific trading costs directly in the no-arbitrage condition that pins down the SDF. The resulting cost-aware kernel is affine in the net-of-cost tangency portfolio, so that the latter is exposed to an expensive signal only to the extent that it contributes meaningfully to asset pricing. Rather than imposing linearity, we estimate the cost-aware SDF as a nonlinear function of 73 stock characteristics and a large panel of macroeconomic series, using the adversarial neural network framework of Chen et al. (2024) on nearly 10,000 U.S. stocks from 1972 to 2023.

We consider both proportional transaction costs, such as effective bid-ask spreads, and price impact costs. Since price impact grows with the size of the investor, we estimate the kernel separately for investors with up to $500 million under management. We then compare the

frictionless and cost-aware SDFs on how well they price the cross-section of gross returns, how their tangency portfolios differ gross and net of costs, and how they explain the variation of realized returns across a wide range of characteristic-managed portfolios.

Although charging for trading costs restricts the set of tangency portfolios the estimator can reach, the cost-aware SDF prices the cross-section better than its frictionless counterpart out of sample. Under bid-ask costs, explained variation rises from 6.3% to 9.8%, the cross-sectional R^2 rises from 14.5% to 22.3%, and the gain grows with trading cost, reaching 14.3% and 25.8% at the \$500 million capacity. That a more restricted tangency portfolio prices the cross-section better suggests that part of the covariance the frictionless SDF relies on carries no information about expected returns.

To see where this covariance comes from, we turn to characteristic-managed portfolios. Under bid-ask costs, the cost-aware SDF prices 47 of 68 more accurately than the frictionless one, and the average absolute pricing error falls from 31.0 to 24.5 basis points per month. The largest improvements are in idiosyncratic volatility, maximum daily return, the zero-trade measures, and other liquidity characteristics such as dollar volume. The improvement depends on whether an anomaly survives its own trading costs. Only eleven of the 68, mostly value, profitability, and accrual signals, earn a positive return net of costs, and for these, the pricing error does not fall under bid-ask or \$50m price-impact costs. For the remaining 57, it falls by 24% under bid-ask costs and by 45% at \$50m.

The anomalies the cost-aware SDF explains best, idiosyncratic volatility, maximum daily return, and the zero-trade measures, also earn higher returns when the tangency portfolio is cheap to rebalance. Across anomalies, the correlation between the loading on the aggregate rebalancing cost and the pricing improvement is -0.47. Because the rebalancing cost is known when the position is taken, it carries no conditional price of risk, and we interpret these returns as premia that an investor who pays to trade cannot collect.

Comparing the two tangency portfolios shows how the cost-aware SDF moves away from the characteristics these anomalies are built on. Under bid-ask costs, the frictionless portfolio

is 29% more concentrated and trades 36% more than the cost-aware one, and it loads more heavily on characteristics whose cross-sectional rankings change most from month to month, such as short-term reversal, return skewness, and maximum daily return. The cost-aware SDF shifts exposure away from these signals. As a result, its tangency portfolio earns a lower monthly gross Sharpe ratio (0.317 vs. 0.348) and a higher one net of costs (0.113 vs. 0.064).

In a simplified linear setting with proportional costs, we show analytically that the SDF loading on each characteristic falls by an amount governed by the ratio of its trading cost to its contribution to pricing. Unlike ridge-type estimators, which shrink each characteristic's coefficient according to estimation noise and prior variance (Kozak et al., 2020), the cost penalty acts on the trades of the combined portfolio, which partly offset across characteristics (DeMiguel et al., 2020), so in our setting, how strongly a characteristic is affected depends on the cost of maintaining exposure to it alongside the others.

Under price impact, how much an investor pays to trade depends on the investor's size. For an investor with \$5 million under management, the frictionless and cost-aware tangency portfolios earn nearly the same net Sharpe ratio. For an investor with \$500 million, the frictionless portfolio earns a monthly net Sharpe ratio of -1.36, against 0.14 for the cost-aware one, and the pricing gain is the largest. Under a square-root impact function, the same ordering holds at \$5 billion, with net Sharpe ratios of -0.46 and 0.15. An investor of that size would lose money holding the frictionless tangency portfolio, so the SDF built from it is not the relevant pricing kernel for that investor.

We next ask whether the gains come from the estimator rather than from cost awareness. Holding everything else fixed, replacing the adversarial objective with a feedforward network that minimizes the average pricing error, or restricting the weights to be linear in characteristics, or replacing stock-level costs with their cross-sectional average, leaves the pricing gain intact (Section 5). Finally, the gain is not confined to small stocks, since excluding the smallest quartile of firms leaves it intact.

The pricing gain could also arise if the cost-aware SDF simply recovered the portfolio an

investor would hold under costs (Jensen et al., 2026). In that case, it should move together with the gain in net Sharpe ratio, which is not what we find. Under the mild size-based costs of Brandt et al. (2009), the frictionless portfolio already trades cheaply and the two net Sharpe ratios are nearly identical (0.167 against 0.165, monthly), yet explained variation still improves by 44%. Under tight position limits, which restrict portfolio turnover outside the estimation, the gap in net Sharpe ratios narrows while the pricing gain grows to 80%. In these cases, the no-arbitrage tangency portfolio chosen to price the cross-section net of costs is not the one an investor would choose to maximize net returns (Section 6).

This paper contributes to two strands of the literature. The first studies pricing kernels under frictions. Luttmer (1996) shows that trading costs turn the no-arbitrage condition into an inequality, widening the set of valid SDFs into a band. Korsaye et al. (2025) characterize the SDFs in that band for proportional costs. We provide an empirical counterpart by estimating a nonlinear, cost-aware SDF in which costs fall within the pricing condition. The second studies how trading costs affect the cross-section of returns, including which anomalies survive their own costs (Novy-Marx and Velikov, 2016, 2019), which characteristics a cost-aware investor should use (DeMiguel et al., 2020), and which factor models best describe returns net of costs (Detzel et al., 2023). Jensen et al. (2026), whose price-impact specification we adopt, characterize the frontier an investor of a given size can implement net of costs, and we characterize the SDF that prices the cross-section for the same investor.

Our anomaly-level results also relate to two further literatures. The first studies the stock-level liquidity premium (Amihud and Mendelson, 1986; Pástor and Stambaugh, 2003; Acharya and Pedersen, 2005) and asks whether illiquid stocks earn higher returns. We ask the analogous question for the pricing kernel: whether the anomalies the cost-aware SDF absorbs reflect compensation for risk or implementation costs. The second attributes the persistence of some anomalies to limits to arbitrage (Shleifer and Vishny, 1997; Pontiff, 2006; Stambaugh et al., 2015). Idiosyncratic volatility, the leading example in that literature, is among the anomalies the cost-aware SDF absorbs; consistent with this view, our pricing improvements are largest when financial conditions are tight and arbitrage capital is constrained.

2 A Cost-Aware Stochastic Discount Factor

We assume the existence of a strictly positive stochastic discount factor (or pricing kernel) 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. Throughout, a return dated $t+1$ is realized over $(t, t+1]$ and a quantity dated t is known at t. The minimum-variance SDF admits a representation as an affine function of the return on a portfolio of tradable assets (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 $\omega_{i,t}$ is the position in asset i at time t as a fraction of wealth. The positions are in excess returns and are not constrained to sum to one. Under this representation, the expected excess return on any asset is proportional to the covariance of its return with the portfolio return $\sum_i \omega_{i,t} r^e_{i,t+1}$, which is mean-variance efficient, and thus normally referred to as the tangency portfolio return.

Holding the tangency portfolio requires rebalancing, and we denote by $\mathcal{TC}_t(\Delta\boldsymbol{\omega}_t; \boldsymbol{\psi}_t) \geq 0$ the cost of moving from the inherited position to the new weights. The vector $\boldsymbol{\psi}_t$ collects the cost parameters known at time t, such as each stock's bid-ask spread or price-impact coefficient and the investor's wealth (Section 2.1). The trade in stock i is $\Delta\omega_{i,t} = \omega_{i,t} - \omega^+_{i,t-1}$, where $\omega^+_{i,t-1} = \omega_{i,t-1}(1 + r_{i,t})$ is the position inherited before any trade (DeMiguel et al., 2009, 2020).[1] The return the investor earns on the tangency portfolio is the gross return less this

[1] An alternative convention normalizes inherited weights by portfolio return, $\omega^{nt}_{i,t} = \omega_{i,t-1}(1+r_{i,t})/(1+r^p_t)$, so that they sum to one (Gârleanu and Pedersen, 2013; Jensen et al., 2026). For proportional costs, the difference is absorbed into the cost parameter $c_{i,t}$ and is quantitatively negligible at a monthly frequency. We adopt this normalization when implementing price-impact costs, where the distinction matters because costs are quadratic in trade size.

cost, $r^{\pi,\text{net}}_{t+1} = \sum_{i=1}^{N} \omega_{i,t}\, r^{e}_{i,t+1} - \mathcal{TC}_t$, and the kernel spanned by that return is

$$M^{TC}_{t+1} \;=\; 1 - r^{\pi,\text{net}}_{t+1} \;=\; \underbrace{1 - \sum_{i=1}^{N} \omega_{i,t}\, r^{e}_{i,t+1}}_{M_{t+1}} + \underbrace{\mathcal{TC}_t(\Delta\boldsymbol{\omega}_t;\, \boldsymbol{\psi}_t)}_{\text{TC component}} \tag{2}$$

We call M^{TC}_{t+1} the transaction-cost-aware, or cost-aware, SDF. It differs from M_{t+1} by the cost of maintaining the tangency portfolio and reduces to it when trading is free. Economically, M^{TC}_{t+1} prices assets for an investor whose wealth grows with the return on the tangency portfolio net of the cost of rebalancing it, so an asset's expected return reflects its covariance with returns that can be earned after paying for the trades needed to hold that portfolio.

Imposing $\mathbb{E}_t[M^{TC}_{t+1}\, r^{e}_{i,t+1}] = 0$ and rearranging gives

$$\mathbb{E}_t\big[r^{e}_{i,t+1}\big] \;=\; \frac{\text{Cov}_t\Big(\sum_j \omega_{j,t}\, r^{e}_{j,t+1},\; r^{e}_{i,t+1}\Big)}{1 - \mathbb{E}_t\Big[\sum_j \omega_{j,t}\, r^{e}_{j,t+1}\Big] + \mathcal{TC}_t}, \tag{3}$$

where the covariance is with the gross return because $\mathcal{TC}_t$ is known at t. Expected excess returns remain thus proportional to conditional covariances with the tangency portfolio return, as in the frictionless case. The cost enters the denominator—it shifts the price of covariance risk by the same amount for every asset, so it carries no price of risk of its own—and the weights—it changes which tangency portfolio the pricing condition selects, because the weights must now price the cross-section after paying the cost of maintaining them. Everything we document about how the cost-aware kernel prices the cross-section runs through the weights. Accordingly, Section 4.1 evaluates both kernels with betas on the gross factor $\sum_i \widehat{\omega}_{i,t} r^{e}_{i,t+1}$, whose conditional covariances with returns are those in Equation (3).

The cost term restricts the tangency portfolios the estimator can reach, so an SDF built from a restricted portfolio could, in principle, price worse. Better asset pricing performance implies that part of the covariance that M_{t+1} attributes to priced risk is covariance with characteristics whose exposure cannot be maintained once costs are paid, and the premia on those characteristics compensate for what an investor who pays the costs cannot collect. Because the

weights are functions of firm characteristics (Section 2.2), the cost of maintaining the tangency portfolio is highest for characteristics whose cross-sectional rankings change often. This means that the pricing condition net of costs keeps such a characteristic only if its contribution to pricing covers that cost.

Equation (2) is not the only SDF consistent with no arbitrage once trading is costly. Luttmer (1996) shows that with proportional costs the no-arbitrage condition becomes an inequality, with $|\mathbb{E}_t[M_{t+1}\, r^e_{i,t+1}]|$ bounded by the cost of a round-trip trade in asset i, so the set of valid SDFs widens from a single minimum-variance element to a band whose width grows with the cost of trading. Korsaye et al. (2025) characterize the minimum-dispersion SDFs in that band for sublinear cost functions and show that each is dual to a penalized portfolio problem, so every kernel in the band, including the frictionless one, is defined by a portfolio weight vector, and the cost constraint determines which one.[2]

Within this set, we select the net-of-cost analog of the minimum-variance SDF. We keep the pricing condition as an equality and charge the kernel for the realized cost of rebalancing the whole portfolio, which is what the investor pays, rather than imposing the band asset by asset. Proposition 1 works out the proportional-cost case, where the band is polyhedral, and we leave a band-constrained estimator for future work. Our objective also differs from that of Jensen et al. (2026), who select weights to maximize mean-variance utility net of trading costs. We select weights that satisfy no-arbitrage pricing restrictions net of costs. Section 6 examines whether the two lead to different portfolios.

2.1 Trading Cost Specifications

Our baseline specification uses proportional transaction costs,

$$\mathcal{TC}_t = \sum_{i=1}^{N} c_{i,t}\, |\Delta\omega_{i,t}|, \qquad \boldsymbol{\psi}_t = \{c_{i,t}\}_{i=1}^{N}, \tag{4}$$

[2]Proposition 2 in Korsaye et al. (2025) requires the cost function to be positively homogeneous of degree one and subadditive. Proportional costs and the polyhedral Luttmer (1996) bands satisfy this. Quadratic price impact is convex but homogeneous of degree two, so the duality does not extend to it, although existence does (their Remark 2).

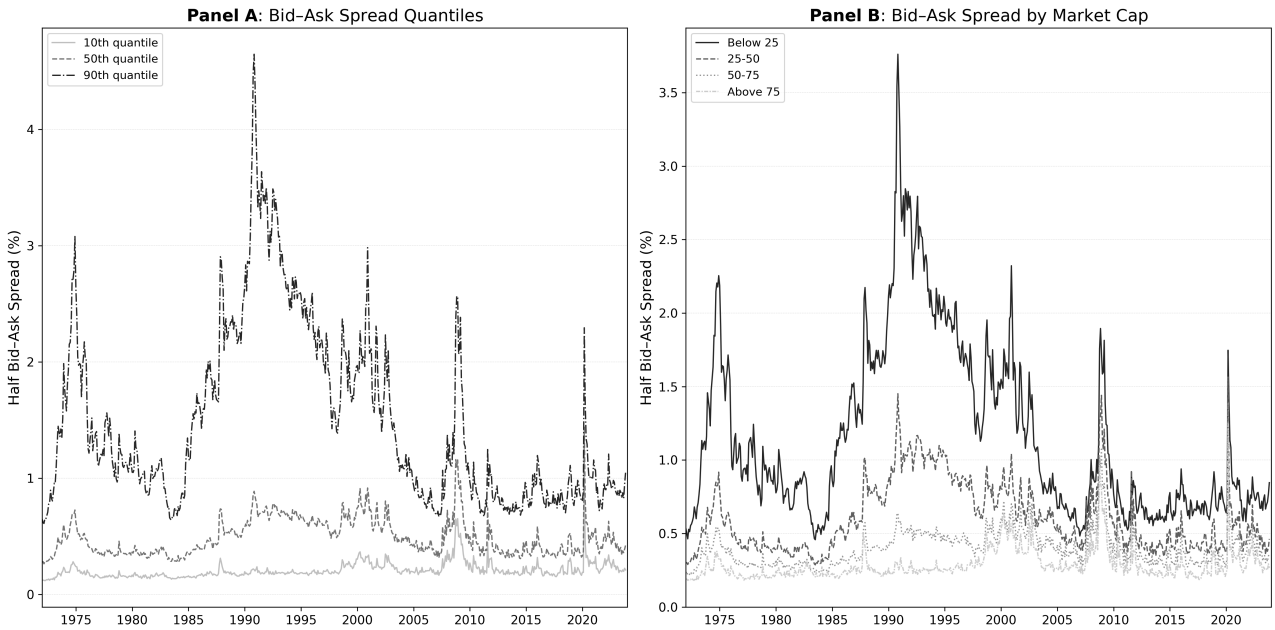


Figure 1: **Sample variation of bid-ask spreads**. The left panel reports the 10th, 50th, and 90th percentiles of the cross-sectional distribution of effective bid-ask spreads (in %). The right panel reports the average bid-ask spread (in %) for stocks sorted by market capitalization. The full sample spans from January 1972 to November 2023.

where $c_{i,t} \in \mathbb{R}^+$ is the effective half-spread for asset i at time t, taken as given and set by microstructure factors outside the marginal investor's control (Gârleanu and Pedersen, 2013; Korsaye et al., 2025).[3]

We measure $c_{i,t}$ as one half of the Corwin and Schultz (2012) spread estimate, scaled by the stock price (Bessembinder and Venkataraman, 2010), so that it is the one-way cost of a trade. The left panel of Figure 1 shows substantial time-series variation of half bid-ask spreads (in %) over the sample period, with pronounced spikes during periods of market stress. The wide gap between the 90th percentile and the median indicates considerable right skewness in the cross-sectional distribution, with the least liquid stocks facing much higher trading costs. Average spreads decline sharply after the early 2000s, so the training window (1972–1991) corresponds to a higher-cost regime than most of the out-of-sample evaluation (1997–2023). The right panel documents the inverse relationship between market capitalization and transaction costs

[3]Bid–ask spreads do adjust with market-wide order flow and volatility (Glosten and Milgrom, 1985; Kyle, 1985) but are taken as given by the representative investor. Our specification is symmetric in buys and sells. Shorting fees and margin requirements are not explicitly modeled. Incorporating asymmetric short costs would add a long-short compression channel on top of the within-leg mechanism. We do not model it, and its direction is untested here.

(Hasbrouck, 2009). Stocks in the lowest market capitalization quartile exhibit transaction costs approximately 3–4 times larger than those in the highest quartile, with the size-based differential widest in the early sample.

The second cost specification we consider follows the quadratic price-impact costs of Jensen et al. (2026). Unlike proportional costs, price impact charges different costs depending on the investor's wealth,

$$\mathcal{TC}_t = \frac{\mathrm{w}_t}{2} \sum_{i=1}^{N} \Lambda_{i,t} \left(\Delta\omega_{i,t}\right)^2, \qquad \boldsymbol{\psi}_t = \{\Lambda_{i,t},\, \mathrm{w}_t\}_{i=1}^{N}, \tag{5}$$

where $\Lambda_{i,t}$ is the stock-level price-impact coefficient and w_t is investor size, or the institution's assets under management (AUM).[4] The impact coefficient is calibrated following Jensen et al. (2026)'s assumption that trading 1% of a stock's daily dollar volume produces an average price impact of 10 bps. Let $V_{i,t}$ denote expected daily dollar volume. With a trade size $\tau_{i,t} = 0.01 V_{i,t}$, the impact is $\frac{1}{2}\Lambda_{i,t}\tau_{i,t}$, so the calibration implies

$$\Lambda_{i,t} = \frac{0.2}{V_{i,t}}. \tag{6}$$

Expected daily dollar volume $V_{i,t}$ is estimated from CRSP monthly data as a trailing six-month average of monthly dollar volume divided by 21 trading days.[5] It is distinct from the DolVol characteristic, the log of two-month lagged volume, which enters both the frictionless and the cost-aware SDF as a predictor, while $V_{i,t}$ enters only the cost penalty.

Under price impact, the trade in weight units is

$$\Delta\omega_{i,t} = \omega_{i,t} - \frac{1 + r_{i,t}}{1 + r_t^p}\, \omega_{i,t-1}, \tag{7}$$

where the normalization by portfolio return r_t^p keeps inherited weights summing to one, as

[4] In Section 6.1 we also assess sensitivity to the cost specification by replacing the baseline proportional costs with the size-based linear approximation of Brandt et al. (2009) and by replacing the quadratic cost with a square-root specification (Almgren et al., 2005; Frazzini et al., 2018), which captures the concave dependence of impact on trade size documented in the empirical microstructure literature.

[5] Appendix B provides the detailed construction of $V_{i,t}$ from the CRSP monthly stock file. The six-month lookback ensures $\Lambda_{i,t}$ varies through time with market-wide liquidity conditions but does not respond to the representative investor's own trading.

required when costs are quadratic in trade size.[6] The portfolio return net of transaction costs is then

$$r_{t+1}^{\pi,\text{net}} = \sum_{i=1}^{N} \omega_{i,t}\, r_{i,t+1}^{e} - \frac{\mathrm{w}_t}{2} \sum_{i=1}^{N} \Lambda_{i,t}\, (\Delta\omega_{i,t})^2. \tag{8}$$

Costs per unit of wealth scale with w_t, and for given weights, larger investors earn a lower net return. Following Jensen et al. (2026), we assume wealth grows with the value-weighted market return. We evaluate price impact costs at three investor wealth levels at the end of the testing sample: \$5m, \$50m, and \$500m in assets under management.

Because the penalty in Equation (5) depends on w_t, so does the SDF in Equation (2). A \$5m investor and a \$500m investor face different costs of maintaining the same portfolio, and the weights that satisfy the pricing condition net of those costs differ. The cost-aware SDF under price impact is therefore the kernel that prices the cross-section for an investor of a given size. A single kernel that prices assets for all investors at once would require a heterogeneous-agent framework, which is beyond this paper's scope.

Two assumptions warrant note. First, we treat $\Lambda_{i,t}$ as exogenous to the investor's own trading history. It responds to market-wide microstructure conditions but not to feedback from the marginal investor's policy. Second, the cost friction is myopic. It penalizes one-period rebalancing without amortizing costs across future holding periods (Gârleanu and Pedersen, 2013; Jensen et al., 2026). We expect relaxing either assumption to affect the magnitude of the reallocation more than its direction, since for a common holding horizon, the ranking of characteristics by implementation cost is unchanged. We do not estimate either extension.

2.2 Estimation

The weights in Equation (2) are unknown functions of the information available at t. We estimate them with the min-max objective

$$\min_{\omega} \max_{g} \frac{1}{N} \sum_{j=1}^{N} \left\| \mathbb{E}\left[M_{t+1}^{TC}(\vartheta_\omega) \cdot r_{j,t+1}^{e} \cdot g_{j,t} \right] \right\|^2 \tag{9}$$

[6] Aggregate wealth growth, $1+r_t^p$, is the value-weighted market portfolio return constructed from the universe of assets in our dataset.

where $M_{t+1}^{TC}(\vartheta_\omega)$ is the cost-aware SDF of Equation (2) and $g_{j,t} \in [-1,1]^d$ is a vector of d instruments for stock j built from information known at t. Each element of $r_{j,t+1}^e\, g_{j,t}$ is the excess return on a position in stock j scaled by one instrument, and its expectation against M_{t+1}^{TC} is the pricing error of that position. Equation (9) averages the $N \times d$ squared pricing errors with equal weights. It is a GMM criterion with an identity weighting matrix, where the inner maximization selects the instruments from the data.

The min-max structure follows Hansen and Jagannathan (1997), who show that the SDF minimizing the largest pricing error over normalized payoffs is the one closest, in a least-squares sense, to the set of SDFs that price all assets. The inner maximization searches for the managed positions that the current weights price worst, and the outer minimization adjusts the weights to price them. Our objective shares this structure but is not the Hansen-Jagannathan distance itself. That distance weights pricing errors by the inverse second-moment matrix of the test payoffs, whereas Equation (9) weights them equally and bounds the instruments. Choosing $g_{j,t}$ plays the role of instrument selection in GMM (Nagel and Singleton, 2011), with instruments selected for the size of the pricing error they expose (Chen et al., 2024). In the terminology of Kaji et al. (2023), the weights $\omega(\cdot\,;\vartheta_\omega)$ carry the parameters of interest, and $g_{j,t}$ is a nuisance parameter estimated in the inner step.[7]

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 state indicators I_t,

$$\omega_{i,t} = \omega(I_t, I_{i,t}; \vartheta_\omega) = \kappa \cdot (\sigma\left(f(I_t, I_{i,t}; \vartheta_\omega)\right) - 0.5) \tag{10}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function. Re-centering the output by subtracting 0.5 lets the portfolio take both long and short positions, whereas scaling by factor κ limits weights to $[-\kappa/2, \kappa/2]$. In the main empirical application, we focus on $\kappa = 0.10$, corresponding to maximum position sizes of $\pm 5\%$. In Section 6.4, we discuss robustness to tighter or looser

[7] Kaji et al. (2023) estimate structural parameters by simulation, with a discriminator that classifies observed against simulated observations under a cross-entropy loss. Our inner maximization selects moment conditions, and the loss is the squared pricing error. We adopt the adversarial framework, not their asymptotic results, which are derived for their objective.

constraints of $\pm 2.5\%$ and $\pm 10\%$ ($\kappa = \{0.05, 0.20\}$).

The function $f(I_t, I_{i,t}; \vartheta_\omega)$ is a feedforward neural network with two hidden layers of 64 units each and parameter vector ϑ_ω, and we parametrize $g_{j,t} = g(I_t, I_{j,t}; \vartheta_g)$ as a tanh transformation of a linear combination of its inputs, with no hidden layer, whose output dimension equals $d = 8$, the number of unconditional moment conditions, and which keeps each instrument in $[-1, 1]$. In both networks, the macroeconomic series enter through a recurrent (LSTM) layer that summarizes the history $(I_1, \ldots, I_t)$ in a low-dimensional state vector, following Chen et al. (2024). Substituting the parametric weights into Equation (2) yields the parametric cost-aware SDF:

$$M_{t+1}^{TC}(\vartheta_\omega) = 1 - \sum_{i=1}^{N} \omega(I_t, I_{i,t}; \vartheta_\omega)\, r_{i,t+1}^{e} + \mathcal{TC}_t(\Delta\boldsymbol{\omega}_t(\vartheta_\omega);\, \boldsymbol{\psi}_t) \tag{11}$$

where $\Delta\omega_{i,t}(\vartheta_\omega)$ is the rebalancing described above with $\omega_{i,t}$ as in Equation (10).

To estimate the parameter vector $(\vartheta_\omega, \vartheta_g)$, we follow Chen et al. (2024) and adopt a Generative Adversarial Network (GAN) approach in which two neural networks compete under the min–max objective in Eq. (9). The adversarial estimation alternates between two steps. First, holding the SDF parameters ϑ_ω fixed, we update the moment parameters ϑ_g by gradient ascent on the pricing loss, identifying the test portfolios that are hardest for the current SDF to price. Then, holding ϑ_g fixed, we update ϑ_ω via gradient descent, adjusting portfolio weights to minimize pricing errors on these challenging test assets. Algorithm 1 in Appendix A.1 summarizes the procedure, which evaluates the loss on the full training sample. Every pricing gain we report is therefore measured against the hardest available test assets. Section 5.1 re-estimates the model with $g_{j,t} \equiv 1$, removing the adversarial step and recovering a comparable improvement moving from a frictionless to a cost-aware SDF.

We hold the network architecture, hyperparameters, and initialization fixed across the estimates and use a common pre-training procedure to isolate the role of the cost awareness. Appendix A.1 provides the layer specifications for both networks and the LSTM state vectors, the hyperparameter values, the pre-training routine, and a schematic of the estimation.

3 Data

We obtain 161 stock characteristics from the Open Asset Pricing database, all classified as "clear predictors" by Chen and Zimmermann (2021).[8] Starting the sample in January 1972 provides the largest set of continuously available characteristics ($p = 129$) over a sufficiently long time series ($T = 624$ months). We further exclude characteristics available for fewer than 800 stocks in any month or with maximum cross-sectional coverage below 2,000 stocks, reducing the set to 73 characteristics.[9] Appendix B describes the sample selection procedure (see Figure B.1) and provides a description of each characteristic (see Table B.1).

We merge the 73 characteristics with monthly stock return data from CRSP and include the return of stock i in month t only if all 73 characteristics are available for that stock in that month (Chen et al., 2024). Therefore, a stock does not need a complete time series to enter the sample. 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 (Shumway and Warther, 1999), and compute each stock's excess return by subtracting the one-month Treasury bill yield from the realized return. 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).[10] Each month, firm-level characteristics are transformed into cross-sectional percentile ranks and linearly mapped into the range $[-0.5, 0.5]$ (Kelly et al., 2019). To reduce look-ahead bias, we lag all firm-level characteristics and macroeconomic variables by one month.

Excluding stock-months with missing characteristics shrinks the universe in a non-random way. Missingness is concentrated within small-cap stocks (Chen and McCoy, 2024; Freyberger et al., 2024; Bryzgalova et al., 2025), which are also the least liquid and the most costly to trade (Avramov et al., 2023). Our sample thus omits the stocks where trading costs bind most.

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

[9] 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).

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

Section 6.3 indicates the direction of this omission. Removing the smallest quartile of firms from the test sample partly shrinks the pricing gain from cost awareness, consistent with the exclusion working against our results.

Macroeconomic variables. We incorporate 122 macroeconomic variables from the FRED-MD database (McCracken and Ng, 2016). These series capture a broad spectrum of economic conditions, including real activity, labor markets, inflation, interest rates, and financial indicators. We transform each series using the recommended codes (typically log levels, log first differences, or growth rates) to ensure stationarity. Table B.2 in Appendix B summarizes each macroeconomic indicator.

Sample and constraints. Our main sample period is from January 1972 to November 2023, totaling 52 years. 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 (January 1997 – November 2023), following Chen et al. (2024). Throughout Section 4, we present results from the GAN architecture under the baseline leverage constraint of $\omega_{i,t} \in (-5\%, +5\%)$. Section 5 re-estimates the model without the adversarial step and without nonlinearity, and Section 6 varies the cost specification, sample, market regime, and position limits.

4 Empirical Results

In this Section, we compare the frictionless and cost-aware kernels out of sample, from January 1997 to November 2023. Section 4.1 asks whether the cost-aware kernel prices the cross-section of individual stock returns better. Section 4.2 examines which anomalies account for the difference and how the answer depends on whether an anomaly survives its own trading costs. Section 4.3 traces the difference to the weights of the two tangency portfolios and to the characteristics they load on. Section 4.4 reports what the two tangency portfolios earn, gross and net of costs, for investors of different sizes, and how they relate to the frontiers spanned by the anomalies and by observable factors.

		M_{t+1}^{TC}			
		Bid-Ask	Price Impact		
	M_{t+1}		\$5m	\$50m	\$500m
Panel A: Explained Variation					
Train	0.124	0.190	0.190	0.208	0.243
Validation	0.036	0.052	0.052	0.058	0.066
Test	0.063	0.098	0.100	0.114	0.143
Panel B: XS-R^2					
Train	0.134	0.240	0.234	0.246	0.274
Validation	0.147	0.195	0.196	0.196	0.186
Test	0.145	0.223	0.222	0.234	0.258

Table 1: **Cross-sectional asset pricing performance**. This table reports the explained variation (EV) and cross-sectional R^2 for the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) across training, validation, and testing periods. The cost-aware SDF is estimated under four transaction cost specifications: effective bid-ask spreads (baseline) and price-impact costs calibrated at three capacity levels corresponding to investor wealth of \$5m, \$50m, and \$500m. All results are based on the GAN architecture under the baseline leverage constraints of ±5%.

4.1 Cross-Sectional Asset Pricing

We measure how well each kernel explains the cross-section of individual stock returns with the explained variation (EV) and the cross-sectional R^2 (XS-R^2) of Chen et al. (2024). Both metrics use each stock's conditional exposure $\widehat{\beta}_{i,t}$ to the tangency portfolio return $F_{t+1} = \sum_{i=1}^{N_t} \widehat{\omega}_{i,t} r_{i,t+1}^e$, estimated by an auxiliary network that predicts the conditional covariance between each stock and F_{t+1}, and the residual $\widehat{\varepsilon}_{i,t+1}$ from a cross-sectional regression, without intercept, of $r_{i,t+1}^e$ on $\widehat{\beta}_{i,t}$ in each month. Appendix A.3 provides more details. EV compares the average squared residual with the average squared excess return and measures how much of the variation in returns the kernel captures. XS-R^2 compares the squared time-series means of the same two quantities and measures how much of the variation in average returns it captures.[11] The auxiliary network is used only for evaluation, and $\mathcal{TC}_t$ enters F_{t+1} only through the tangency portfolio weights.

[11] Following Chen et al. (2024), EV $= 1 - \frac{\frac{1}{T}\sum_t \frac{1}{N_t}\sum_{i=1}^{N_t} \widehat{\varepsilon}_{i,t+1}^2}{\frac{1}{T}\sum_t \frac{1}{N_t}\sum_{i=1}^{N_t} (r_{i,t+1}^e)^2}$ and XS-$R^2 = 1 - \frac{\frac{1}{N}\sum_i \frac{T_i}{T}\left(\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} \widehat{\varepsilon}_{i,t+1}\right)^2}{\frac{1}{N}\sum_i \frac{T_i}{T}\left(\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} r_{i,t+1}^e\right)^2}$, where N_t is the number of stocks in month t, $\mathcal{T}_i$ is the set of months in which stock i is in the sample, and T_i is its size.

Table 1 shows that the restricted kernel prices the cross-section better out of sample under every cost specification. Under bid-ask costs, EV rises from 6.3% to 9.8% and XS-R^2 from 14.5% to 22.3%, improvements of 56% and 54%. Under price impact, the cost of a trade of a given size rises with the investor's capacity, and so does the pricing gain. EV rises from 0.100 at \$5m to 0.114 at \$50m and 0.143 at \$500m, improvements of 59%, 81%, and 127%, and XS-R^2 rises from 0.222 to 0.234 and 0.258. The gain from charging the kernel for its trading grows with the cost of that trading.

A higher EV or XS-R^2 could come from a different ranking of systematic risk across stocks or from a closer fit within the same ranking. Appendix D.1 shows that when stocks are sorted into deciles on their estimated SDF betas, the rank ordering of decile returns is nearly the same under the frictionless kernel, the bid-ask specification, and the \$50m price-impact specification (correlation ≈ 0.98), all producing monotonic spreads with significant CAPM alphas. The two kernels therefore agree on which stocks carry more systematic risk, although the cost-aware betas fit realized and average returns more closely within that ordering. The largest change is in the lowest decile, whose average return rises from 0.4% a year under the frictionless kernel to 2.1–2.9% under cost-awareness.

Because both metrics depend on an estimated beta, Appendix D.2 repeats the evaluation with a linear specification to estimate $\widehat{\beta}_{i,t}$. Cost-awareness improves both metrics under every specification with linear betas. For EV under every specification, and for XS-R^2 under price impact, the improvement exceeds the largest change induced by switching the beta network itself. Under bid-ask costs, the XS-R^2 gain with linear betas, from 0.164 to 0.182, is smaller than that change.

4.2 Anomaly Pricing and the Cost of Arbitrage

To better understand where the pricing gain comes from, we turn our focus from individual stocks to portfolios built on the characteristics. For each characteristic—signed as in Chen and Zimmermann (2021) so that the high-minus-low portfolio reflects the documented premium—, we form a value-weighted long-short portfolio from 2×3 sorts on size and the characteristic,

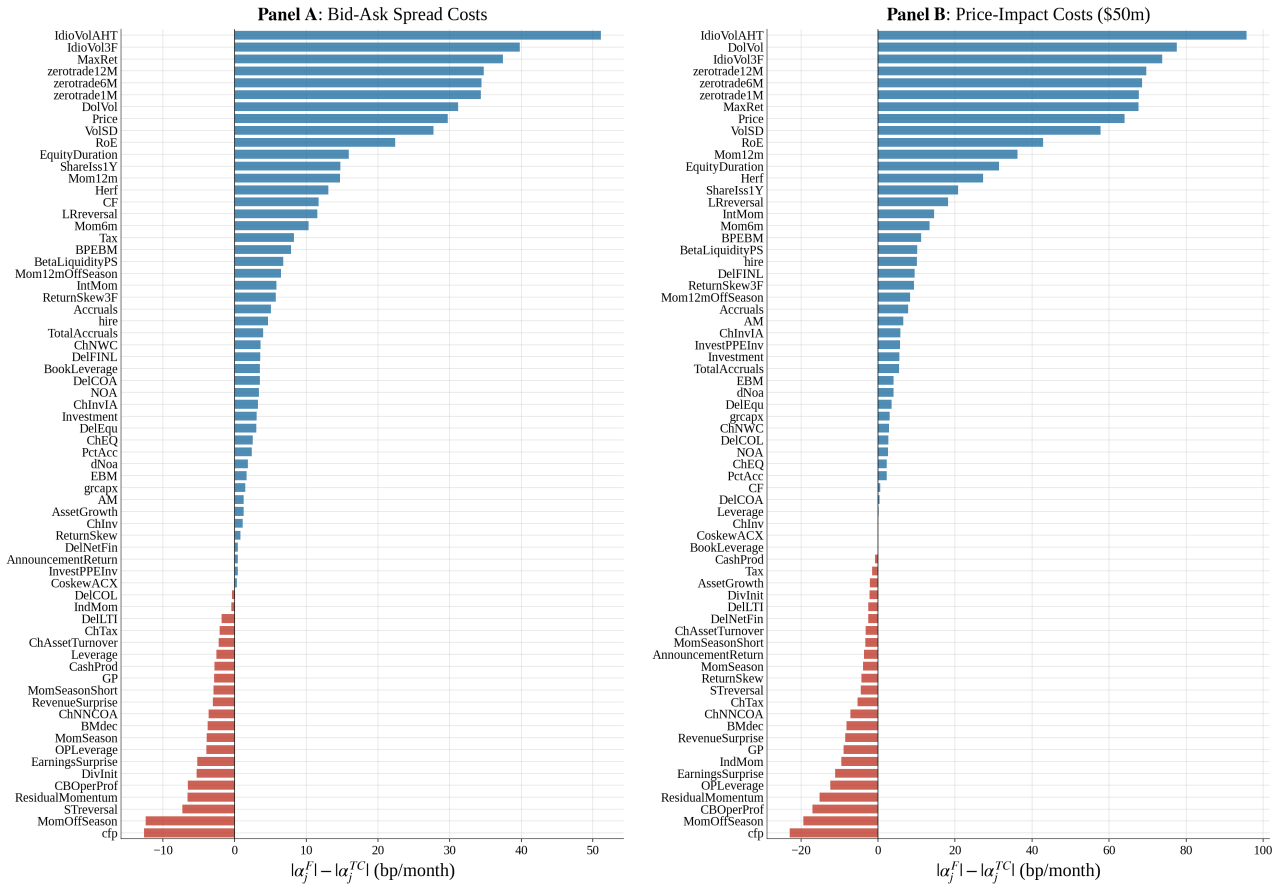


Figure 2: **Anomaly pricing implications.** Each bar reports $\Delta|\alpha_j| = |\alpha_j^F| - |\alpha_j^{TC}|$ (bp/month) for one of the 68 characteristic-managed portfolios, sorted from largest improvement to largest deterioration. Blue bars indicate anomalies that the cost-aware SDF prices better. Red bars indicate anomalies priced worse. Panel A: bid-ask spread costs. Panel B: price-impact costs ($w = \$50$m). The out-of-sample period is January 1997 to November 2023.

and charge each leg its own trading cost when net returns are needed. For this analysis, we consider 68 anomaly portfolios since five of the 73 characteristics cannot be formed as 2×3 factors.[12] Appendix C provides more details on the portfolio construction. We regress the gross return of each portfolio on the tangency portfolio return of each kernel, $R_{j,t} = \alpha_j + \beta_j X_t + \varepsilon_{j,t}$ with $X_t \in \{F_t, F_t^{TC}\}$, and measure the change in pricing accuracy as $\Delta|\alpha_j| = |\alpha_j^F| - |\alpha_j^{TC}|$. The spread $\Delta|\alpha_j|$ is positive when the cost-aware kernel prices anomaly j more accurately. The test assets are the same gross returns under both kernels, so all variation in the pricing errors comes from the SDF only.

Across all 68 portfolios, Figure 2 shows that the cost-aware SDF prices 47 more accurately

[12] `Size` defines the size split, and `ConvDebt`, `DivOmit`, `IndIPO` and `RDIPO` fall below the coverage threshold within size groups.

under bid-ask costs and 41 under both bid-ask and $50m price-impact costs. The average absolute pricing error falls from 31.0 bp per month to 24.5 bp and 19.4 bp, reductions of 21% and 38%, and the ranking of improvements is stable across the two specifications (correlation of 0.92). Twenty-one anomalies are priced worse under bid-ask costs, two of them by more than 10 bp, cash-flow-to-price (`cfp`, −13 bp) and off-season momentum (`MomOffSeason`, −12 bp), both of which carry small pricing errors under the frictionless kernel (9 and 3 bp). Appendix D.3 reports the $5m and $500m price impact, with mean improvements of 4.3 and 4.7 bp against 6.5 bp under bid-ask costs and 11.6 bp at $50m.

The largest improvements are in portfolios built on idiosyncratic volatility (`IdioVolAHT`, `IdioVol3F`), maximum daily return (`MaxRet`), the three zero-trade measures, dollar volume (`DolVol`), and the volatility of trading volume (`VolSD`). Under price impact at $50m, their pricing errors shrink by 58 to 96 bp per month. The zero-trade portfolios carry frictionless alphas of −111 to −121 bp ($|t|$ between 2.89 and 3.27), which fall to between −41 and −53 bp ($|t| < 1.6$), and `IdioVolAHT` moves from −163 bp ($t = -2.96$) to −67 bp ($t = -1.26$). Idiosyncratic volatility is the leading example in the literature of a mispricing that persists because it is costly to arbitrage (Stambaugh et al., 2015).

A common property of absorbed anomalies is that their expected returns do not cover trading costs. Following Detzel et al. (2023), we classify an anomaly as surviving its own trading costs when its mean return net of those costs is positive. Eleven of the 68 survive, mostly value, profitability, and accrual signals.[13] Among the 57 anomalies that do not survive, the average absolute pricing error falls from 32.0 bp per month under the frictionless kernel to 24.2 bp under bid-ask cost-awareness and 17.6 bp under price impact at $50m, reductions of 24% and 45%. Among the eleven surviving anomalies, it moves from 25.9 to 26.3 and 28.8 bp.

The improvement also rises with the cost of trading the anomaly itself. Sorting the 68 portfolios into terciles by the average monthly cost of their long-short position, the average improvement is 4.1 bp in the cheapest tercile and 8.0 bp in the most expensive under bid-ask

[13]The eleven are `AM`, `BMdec`, `cfp`, `GP`, `CBOperProf`, `OPLeverage`, `Accruals`, `PctAcc`, `NOA`, `ChInv` and `LRreversal`.

costs, and 6.5 and 15.8 bp under price impact. The cost-aware kernel therefore prices better the anomalies that are expensive to trade and whose premia do not survive that expense, while leaving the pricing of the survivors unchanged.

Exposure to the cost of rebalancing. To further investigate the relationship between anomaly returns and cost-awareness in the pricing kernel, we estimate for each portfolio the loading β_j^{TC} of gross returns on the aggregate rebalancing cost, $\mathcal{TC}_t$. Table D.4 in Appendix D.4 shows the results for the bid-ask and \$50m cost-aware SDFs. Across the 68 portfolios, β_j^{TC} correlates -0.47 with $\Delta|\alpha_j|$. The most negative loadings belong to momentum (`Mom12m`, `Mom12mOffSeason`, -0.06), the volatility of trading volume (-0.05), `IdioVolAHT` (-0.05) and the zero-trade measures (between -0.03 and -0.05). Their returns are lower when rebalancing the tangency portfolio is more costly. At the positive end sit short-term reversal, long-term reversal, and cash-flow-to-price (all $+0.07$), whose returns are higher when rebalancing is more expensive. Short-term reversal pays most when liquidity provision is expensive (Nagel, 2012). The cost-aware kernel prices short-term reversal and cash-flow-to-price worse under bid-ask costs (-7 and -13 bp) and long-term reversal better ($+12$ bp). In the limits-to-arbitrage view, a premium persists when the arbitrageurs who would eliminate it cannot trade cheaply (Shleifer and Vishny, 1997; Pontiff, 2006). The kernel reaches this conclusion through the tangency portfolio weights that the moment condition selects (Section 2).

Because $\mathcal{TC}_t$ is known at t, it carries no conditional price of risk (Section 2). Therefore, in this Section we use the loadings only to order the anomalies. However, unconditional loadings can still relate to average returns. Appendix D.5 shows that in cross-sectional regressions of average returns on β_j^{TC}, the slope is significant under bid-ask costs and at the \$500m capacity, with Shanken-corrected t-statistics of 2.76 and 2.70. It is not significant in monthly Fama-MacBeth regressions under any specification, where the t-statistics range from 0.26 to 1.30.

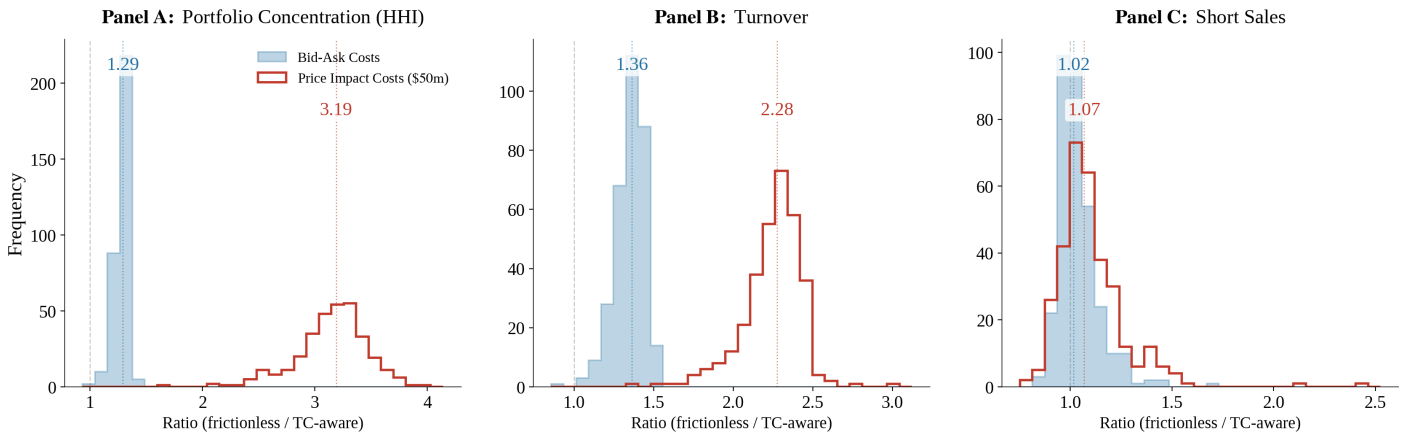


Figure 3: **Tangency portfolio composition.** This figure reports the distribution of the monthly ratio between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) for three portfolio statistics: the HHI index (Panel A), aggregate turnover (Panel B), and fraction of short sales (Panel C). The distributions are shown for bid-ask spread costs (filled blue) and price-impact costs with capacity \$50m (red outline). Dotted vertical lines mark median ratios. The dashed line at 1.0 indicates parity. The out-of-sample period is January 1997 to November 2023.

4.3 How the Two Tangency Portfolios Differ

The anomalies the cost-aware kernel absorbs are built on characteristics, so the reallocation that absorbs them should be visible in the two tangency portfolios' weights. Figure 3 plots the time-series distribution of the ratio between the frictionless and cost-aware portfolios for three weight statistics: the Herfindahl-Hirschman Index (HHI), aggregate turnover, and the fraction of short positions.[14] Under bid-ask spreads, the frictionless portfolio is 29% more concentrated (median HHI ratio of 1.29) and trades 36% more (turnover ratio of 1.36). Under price impact at \$50m, the two ratios rise to 3.19 and 2.28, and the HHI distributions of the two specifications do not overlap.

The share of short positions barely changes, with median ratios of 1.02 and 1.07. Short sales carry no separate cost in our specifications, so the loss gives the kernel no reason to compress the short leg relative to the long one; reallocation instead takes the form of diversification and lower turnover within each leg.

To trace the change in composition to individual characteristics, we measure for each characteristic j the exposure of the tangency portfolio as $\widehat{x}_t^j = \left| N_t^{-1} \sum_{i=1}^{N_t} \widehat{\omega}_{i,t} I_{i,t}^j \right|$ and take the differ-

[14] $\text{HHI}_t = \sum_{i=1}^{N_t} \widehat{\omega}_{i,t}^2$ and $TO_t = \sum_{i=1}^{N_t} |\Delta\widehat{\omega}_{i,t}|$, with $\Delta\widehat{\omega}_{i,t} = \widehat{\omega}_{i,t} - \widehat{\omega}_{i,t-1}(1 + r_{i,t})$.

ence in medians $\Delta\widehat{x}^j = \text{med}(\widehat{x}_t^j) - \text{med}(\widehat{x}_t^{TC,j})$, which is positive when the frictionless portfolio is more exposed. Exposure measures how much the tangency portfolio loads on a characteristic, not how much the weight network relies on it for pricing. To measure this reliance, we decompose the weight network $\omega(I_t, I_{i,t}; \vartheta_\omega)$ for each stock into contributions from the 73 stock characteristics in $I_{i,t}$ using DeepSHAP (Lundberg and Lee, 2017). DeepSHAP writes the difference between $\omega(I_t, I_{i,t}; \vartheta_\omega)$ and the average weight over a reference sample of 300 stock-months as a sum of contributions of the network inputs, where $\phi_{i,t}^j$ is the contribution of stock i's value of characteristic j. The mean absolute contribution $\bar{\phi}_{j,t}$ over the stocks in month t measures how much the weight network relies on characteristic j, and $\Delta\text{SHAP}^j = \text{med}(\bar{\phi}_{j,t}) - \text{med}(\bar{\phi}_{j,t}^{TC})$ is the difference between the frictionless and cost-aware kernels, positive when the frictionless network relies on j more. Appendix A.4 provides the computational details.

Figure 4 plots $\Delta\widehat{x}^j$ against ΔSHAP^j and codes each characteristic by the quartile of its average rank change,

$$\text{ARC}_j = \frac{1}{T-1}\sum_{t=2}^{T}\left[\frac{1}{|S_t^{(j)} \cap S_{t-1}^{(j)}|}\sum_{i \in S_t^{(j)} \cap S_{t-1}^{(j)}}\left|k_{i,t}^{(j)} - k_{i,t-1}^{(j)}\right|\right] \tag{12}$$

where $S_t^{(j)}$ is the set of firms with a value of characteristic j in month t and $k_{i,t}^{(j)} \in [-0.5, 0.5]$ is the rank-standardized value of characteristic j for firm i. ARC measures how much a characteristic's cross-sectional ranking changes from one month to the next, and thus how much trading it takes to keep a constant exposure to it. Across the 68 characteristic-managed portfolios of Section 4.2, ARC and realized trading cost correlate at 0.98.[15]

Under bid-ask costs, the changes in exposure and attribution correlate at 0.63, so the characteristics that lose exposure also lose weight in the network's attribution. Such a correlation is 0.47 under price impact. Nearly every characteristic in the top ARC quartile lies in the upper-right quadrant of both panels, led by short-term reversal, return skewness, and momentum. Short-term reversal loses exposure and, as Section 4.2 shows, the cost-aware kernel prices

[15] An alternative instability measure based on the rank correlation between consecutive months correlates with ARC at 0.97 (Pearson) and 0.90 (Spearman), and replacing ARC with its quartiles produces visually identical scatter patterns.

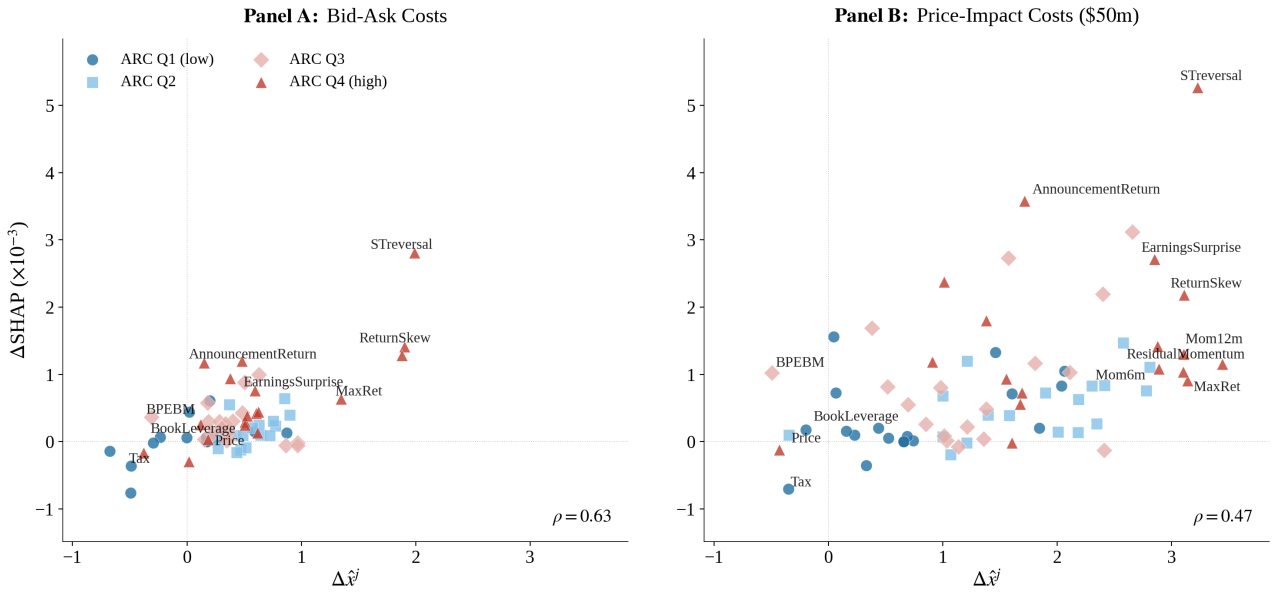


Figure 4: **Cost-driven characteristic reallocation.** Each point represents one of the 73 firm characteristics. The x-axis plots the change in tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs (positive values indicate higher frictionless exposure). The y-axis plots the corresponding change in DeepSHAP attribution ΔSHAP^j ($\times 10^{-3}$). Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). Panel A: bid-ask spread costs. Panel B: price-impact costs ($50m). The out-of-sample period is January 1997 to November 2023.

it worse: the kernel gives up the position that would capture its premium, which is what a premium masked by trading costs looks like from inside the kernel. Under price impact, the same characteristics move further. Appendix D.6 reports a compressed version of the pattern at $5m and a stronger one at $500m, where momentum and short-term reversal show the largest shifts. High turnover is not the only way a characteristic becomes expensive to hold. `IdioVolAHT` and the six- and twelve-month zero-trade measures sit in the two lowest ARC quartiles, and their long-short portfolios still fail to cover their own trading costs (Section 4.2).

Understanding the cost-pricing trade-off. A simplified version of the estimation problem gives the reallocation trade-off in closed form. Suppose the SDF weights are linear in characteristics, $\omega_{i,t} = \sum_{j=1}^{k} \theta_j I_{i,t}^j$, transaction costs are proportional, $\mathcal{TC}_t = \sum_{i=1}^{N} c_i |\Delta\omega_{i,t}|$, and there is no adversarial component, $g_{j,t} = 1$. Assume that the cross-sectional rank changes $\Delta I_{i,t}^j$ are independent across characteristics and have zero mean, that the pricing loss is approximately separable across characteristics, and that stock-level trading costs are uncorrelated

with stock-level turnover. Let $\bar{c} = N^{-1} \sum_i c_i$ denote the average trading cost.

Proposition 1 (Characteristic reallocation)**.** *Under the conditions above, the expected portfolio turnover attributable to characteristic j is proportional to $|\theta_j| \, \mathrm{ARC}_j$. The loadings θ_j^{TC}, $j = 1, \ldots, k$ that minimize the sum of squared pricing errors of the cost-aware SDF satisfy, to first order in $\bar{c}$:*

$$|\theta_j^{TC}| = \left(|\theta_j^F| - \frac{\lambda \, \mathrm{ARC}_j}{\pi_j} \right)^+, \quad \forall j \tag{13}$$

where θ_j^F is the frictionless loading, $\pi_j > 0$ is the curvature of the pricing loss in the direction of characteristic j (i.e., the sensitivity of the cross-sectional fit to perturbations in $|\theta_j|$), and $\lambda > 0$, proportional to $\bar{c}$, indexes the severity of transaction costs. The reallocation $|\theta_j^F| - |\theta_j^{TC}|$ is thus governed by the ratio ARC_j / π_j.

The proof is in Appendix A.5. In Equation (13), two characteristics with the same frictionless exposure receive different cost-aware loadings when their ARC differs, and a high-ARC characteristic keeps its loading when its contribution to pricing, π_j, is large enough. The proposition is consistent with the patterns in Figure 4. High-ARC characteristics with modest pricing contributions lose the most exposure, whereas low-ARC fundamentals stay near the origin because their cost term is small at any severity. In the proposition, a more severe cost corresponds to a larger λ, which extends the reallocation to more characteristics. Section 5.3 replaces each $c_{i,t}$ with its cross-sectional mean $\bar{c}_t$, so that ARC_j is the only source of variation in the penalty across characteristics.

Equation (13) has the form of a soft-thresholding rule in which the loading on characteristic j is shrunk by $\lambda \, \mathrm{ARC}_j / \pi_j$. The shrinkage penalties of Kozak et al. (2020) also shrink loadings, but their penalty on a characteristic depends on its estimation noise and prior variance, whereas $\lambda \, \mathrm{ARC}_j$ is the cost of keeping exposure to the characteristic. The proposition also simplifies the cost term of the estimated kernel in one respect. In Equation (2), $\mathcal{TC}_t$ charges the trades of the whole portfolio, and trades generated by different characteristics partly offset (DeMiguel et al., 2020). Therefore, the cost of keeping exposure to a characteristic depends on the other characteristics held alongside it, and the separable penalty $\lambda \, \mathrm{ARC}_j$ overstates it. With

		Gross		Net	
	Market	M_{t+1}	M^{TC}_{t+1}	M_{t+1}	M^{TC}_{t+1}
Panel A: Bid-Ask Spreads (Baseline)					
Train	0.088	1.921	1.441	1.297	1.169
Validation	0.345	1.178	1.043	0.201	0.353
Test	0.143	0.348	0.317	0.064	0.113
Panel B: All Cost Specifications (Test Period)					
Bid-Ask	0.143	0.348	0.317	0.064	0.113
Price Impact, \$5m		0.348	0.326	0.313	0.311
Price Impact, \$50m		0.348	0.244	0.002	0.225
Price Impact, \$500m		0.348	0.157	−1.355	0.139

Table 2: **SDF portfolio Sharpe ratios**. This table reports monthly Sharpe ratios for the market portfolio, the frictionless SDF (M_{t+1}), and the cost-aware SDF (M^{TC}_{t+1}). Panel A reports results under the baseline bid-ask spread specification across training, validation, and testing periods. Panel B compares all four transaction cost specifications in the out-of-sample testing period: effective bid-ask spreads and price-impact costs at three capacity levels corresponding to investor wealth of \$5M, \$50M, and \$500M. All results are based on the GAN architecture under ±5% leverage constraints.

correlated characteristics, the reallocation of each one also depends on its covariance with the others in the pricing loss.

4.4 Mean-Variance Efficiency

Table 2 reports the gross and net monthly Sharpe ratios of the two tangency portfolios. Under bid-ask costs, the cost-aware tangency portfolio earns a gross Sharpe ratio of 0.317 against 0.348 for the frictionless portfolio, and a net Sharpe ratio of 0.113 against 0.064, monthly. Trading costs remove 82% of the frictionless portfolio's gross Sharpe ratio and 64% of the cost-aware portfolio's, so the cost-aware portfolio gives up 9% of its gross Sharpe ratio and earns 77% more net of costs.[16] In Panel A of Figure 5, the gross wealth gap between the two portfolios is negative throughout the test period, and the net gap is positive and widens.

Under price impact, trade costs grow with an investor's wealth. For an investor with \$5 million under management, the frictionless and cost-aware tangency portfolios earn net monthly Sharpe ratios of 0.313 and 0.311 (Table 2, Panel B), yet EV rises by 59% and XS-R^2

[16]Both portfolios earn much higher Sharpe ratios in the training sample (Panel A), 1.921 and 1.441 gross, as is common for flexible estimators.

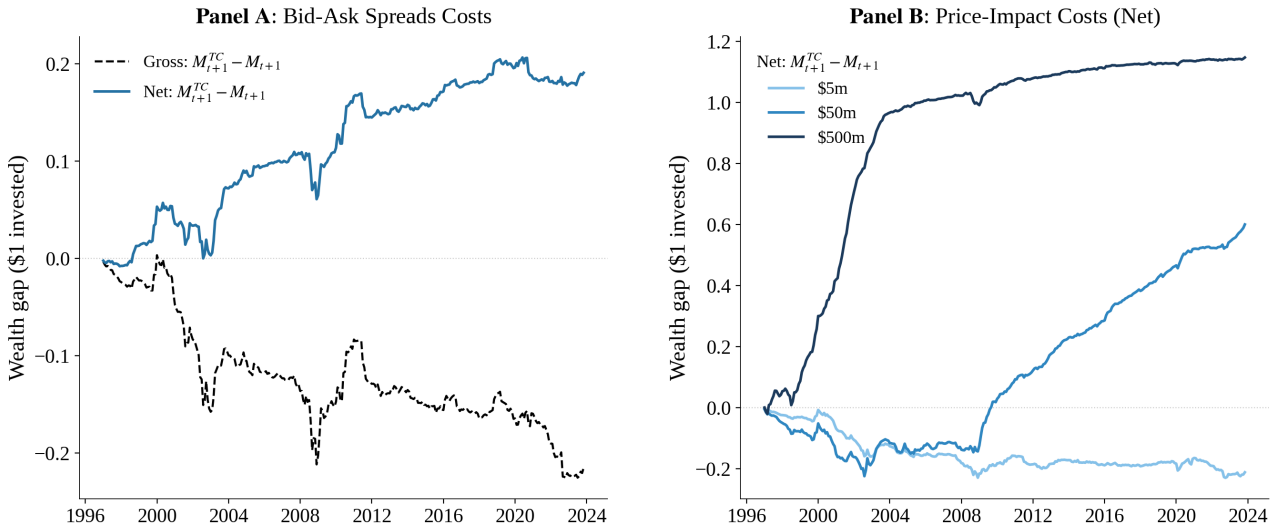


Figure 5: **Wealth gap between cost-aware and frictionless tangency portfolios.** Both panels plot the wealth gap $\prod_{\tau=1}^{t}(1+F_\tau^{TC}) - \prod_{\tau=1}^{t}(1+F_\tau)$, where F_τ^{TC} and F_τ denote monthly returns on the M_{t+1}^{TC} and M_{t+1} tangency portfolios. Panel A reports gross and net wealth gaps under bid-ask costs. Panel B reports net wealth gaps under price-impact costs at three capacity levels ($5m, $50m, $500m). The test period is January 1997 to November 2023.

by 53% (Table 1). At this investor size, the pricing gain comes with negligible net risk-adjusted returns. The weights still move, as the cost-aware portfolio shows a lower gross Sharpe ratio (0.326 vs. 0.348), while the improvement on the anomaly portfolios is the smallest of the four specifications (Section 4.2).

At $50m wealth, the frictionless portfolio's net Sharpe ratio falls to 0.002 while the cost-aware portfolio earns 0.225; at $500m, the frictionless portfolio earns −1.355 against 0.139, so an investor of that size could not hold the frictionless tangency portfolio. The cost-aware portfolio gives up gross performance to achieve this, with gross Sharpe ratios of 0.244 and 0.157 against 0.348; over the same range, EV rises by 81% and 127%. Panel B of Figure 5 plots the net wealth gap between the two portfolios: slightly negative at $5m, positive and widening at $50m, and reaching $1.15 per dollar invested at $500m by the end of the sample. Section 6.1 shows that under the square-root impact function of Almgren et al. (2005), the same ordering holds at $5 billion.

Frontier expansion. The pricing errors of Section 4.2 measure how far an anomaly's average return lies from what a kernel implies. We now ask what an investor who holds a kernel's

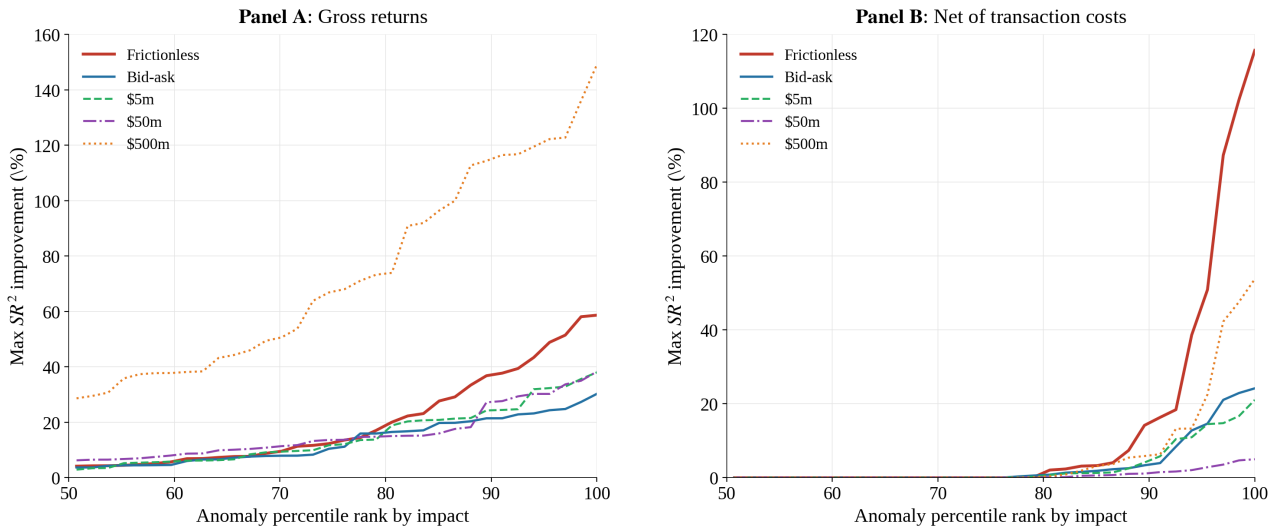


Figure 6: **Frontier expansion from adding anomalies to the SDF.** For each SDF specification M and each of the 68 characteristic-managed portfolios A, the figure plots $\%\Delta SR^2(M,A) = SR^2(M,A)/SR^2(M) - 1$, where SR^2 is the ex-post maximum squared Sharpe ratio attainable from non-negative positions in the long and short versions of each asset. We rank anomalies by frontier expansion, and the plot begins at the median. Panel A uses gross returns. Panel B deducts transaction costs from both the long and the short leg of every position. In Panel B, we evaluate the frictionless tangency portfolio net of bid-ask costs and each cost-aware portfolio net of its own cost specification. The out-of-sample period is January 1997 to November 2023.

tangency portfolio gains by adding the anomaly. For each SDF M and anomaly A, we form the opportunity set from long and short positions in both, charge each position the one-way cost of the leg it holds, and solve for the maximum squared Sharpe ratio subject to non-negative weights summing to one. The frontier expansion is $\%\Delta SR^2(M,A) = SR^2(M,A)/SR^2(M) - 1$.[17] Gross of costs, the long and short positions are mirror images, the constraint does not bind, and the increase in the squared Sharpe ratio equals the squared pricing error scaled by the residual variance, $\alpha_j^2/\mathrm{Var}(\varepsilon_j)$ (see Appendix C.1). An anomaly with a large pricing error thus expands the frontier little when that error is hard to separate from noise. Net of costs, each leg pays its own cost; symmetry breaks, and the expansion must be computed.

Gross of costs, almost every anomaly expands every frontier, and the kernels separate only in the upper quartile of Figure 6, Panel A. The most impactful anomaly expands the frictionless frontier by 59% and the cost-aware frontier by 30% under bid-ask costs and by 38% at $5m and

[17] Each asset enters twice, as a long and a short position, and we charge costs on both, since an investor pays to trade in either direction. Non-negativity prevents the optimizer from holding an asset and its mirror image at once, a position that earns nothing gross while paying to trade twice.

$50m. The $500m SDF shows larger percentage expansions, 28% at the median, because its own gross SR^2 is 0.025, so a given absolute gain registers as a large percentage (Appendix D.7).

Net of costs, only 16 to 17 of the 68 anomalies expand any frontier, in line with Detzel et al. (2023), which finds that most anomalies stop being anomalous once their trading costs are paid. Under bid-ask costs, where both kernels are evaluated net of the same costs, the remaining expansions are much larger for the frictionless kernel. At the 90th percentile, the frictionless frontier expands by 14.8%, and the cost-aware frontier by 3.5%, and the largest expansion falls from 116% to 24%. Several of the anomalies that still expand a frontier lose money net of costs on both legs and enter as hedges against the tangency portfolio. `IdioVolAHT` returns −58 bp per month net on its long leg and −2 bp on its short, and the six- and twelve-month zero-trade measures lose 14 to 44 bp either way. The exposure of a frictionless kernel to these anomalies is invisible in gross returns and large in net returns.

The same question can be asked of an investor who already holds the Fama–French factors. Appendix D.8 shows that, gross of costs, both tangency portfolios expand the frontier of the three-, five-, and six-factor models; net of bid-ask costs, neither adds much to it, and net of price impact, the cost-aware portfolio adds to it at $50m but not at $500m.

5 Attributing the Pricing Gain

Section 4 attributes the pricing gain to the moment condition stated net of costs. Part of the gains may not come from cost awareness per se, but from the mechanics of the estimation approach, such as the adversarial choice of test assets, the nonlinearity of the mapping between characteristics and tangency portfolio weights, and the stock-level variation in trading costs. This section removes each ingredient in turn, holding the others fixed, and re-estimates both SDFs.[18] If the pricing gain depended uniquely on one of these features, it should become negligible when that ingredient is removed. All results are out of sample, from January 1997 to November 2023, unless otherwise noted.

[18] We do not estimate cases in which two features are removed at once. Each exercise holds the others fixed, which is what isolates its contribution.

	Adversarial objective (baseline)				Feedforward network, $g_{j,t} \equiv 1$			
	EV	XS-R^2	SR		EV	XS-R^2	SR	
			Gross	Net			Gross	Net
M_{t+1} (frictionless)	0.063	0.145	0.348	0.064	0.081	0.172	0.315	0.051
M^{TC}_{t+1} (bid-ask)	0.098	0.223	0.317	0.113	0.109	0.230	0.253	0.118
M^{TC}_{t+1} ($5m)	0.100	0.222	0.326	0.311	0.118	0.241	0.270	0.262
M^{TC}_{t+1} ($50m)	0.114	0.234	0.244	0.225	0.112	0.237	0.205	0.192
M^{TC}_{t+1} ($500m)	0.143	0.258	0.157	0.139	0.124	0.242	0.115	0.104

Table 3: **Adversarial objective versus feedforward network.** This table reports out-of-sample performance for the baseline SDF, estimated under the min–max objective of Equation (9), and for a feedforward network that sets $g(h^g_t, I_{j,t}; \vartheta_g) \equiv 1$ and so minimizes the average pricing error across a fixed cross-section. Both use the same tangency portfolio architecture, inputs, and sigmoid output, so differences reflect the training objective. EV is explained variation and XS-R^2 the cross-sectional R^2. Sharpe ratios are monthly. Net Sharpe ratios deduct the cost specification of the corresponding row, so the frictionless row reports its value under bid-ask costs. Charged the remaining specifications, the frictionless SDF earns net Sharpe ratios of 0.313, 0.002 and -1.355 under the adversarial objective and 0.289, 0.052 and -1.241 under the feedforward network, at $5m, $50m and $500m capacity. The out-of-sample period is January 1997 to November 2023.

5.1 Fixing the Test Assets

Setting $g(h^g_t, I_{j,t}; \vartheta_g) \equiv 1$ in Equation (9) replaces the largest pricing error attainable from a managed portfolio with the average squared pricing error across a fixed cross-section of stocks. We call this setting the feedforward network (FFN) SDF. The weight network keeps the baseline architecture, inputs, and sigmoid output, so the two estimators differ only in their objective. Appendix A.1 provides details on the estimation algorithm.

Table 3 reports the results for the baseline adversarial objective and the feedforward network. Under the latter, the explained variation rises from 0.081 for the frictionless kernel to between 0.109 and 0.124 for the cost-aware kernels, and cross-sectional R^2 rises from 0.172 to between 0.230 and 0.242. Under the baseline adversarial network, explained variation moves from 0.063 to between 0.098 and 0.143, and cross-sectional R^2 from 0.145 to between 0.222 and 0.258. Therefore, cost-awareness improves cross-sectional pricing regardless of whether the test assets are chosen adversarially or fixed in advance as in standard practice.

Comparing the two estimators row by row, the feedforward network attains the higher

explained variation for the frictionless kernel and under bid-ask and \$5m costs (0.081, 0.109, and 0.118 against 0.063, 0.098, and 0.100), and the baseline attains the higher explained variation at \$50m and \$500m (0.114 and 0.143 against 0.112 and 0.124). Cross-sectional R^2 orders the two estimators in the same way except at \$50m, where the feedforward network remains marginally higher (0.237 against 0.234). The feedforward network minimizes the average squared pricing error across a fixed cross-section—which is close to what cross-sectional R^2 measures—, whereas the adversarial objective targets the largest pricing error attainable from a managed portfolio.

Appendix D.9 reports the time-series distribution of the ratio between the frictionless and cost-aware portfolios for HHI, aggregate turnover, and the fraction of short positions, as in Section 4.3. Under bid-ask costs, the median HHI of the frictionless feedforward portfolio is 1.54 times that of its cost-aware counterpart, its median turnover is 1.81 times as large, and its median short-position ratio is 1.07. Under price impact, the concentration ratio rises to 1.84, the turnover ratio falls to 1.70, and the short-position ratio is 1.10, so the shift toward more diversified, lower-turnover portfolios that Section 4.3 documents for the baseline also occurs without the adversarial objective.

At the characteristic level, Figure D.5 in Appendix D.9 shows that changes in exposure and DeepSHAP attribution correlate at 0.50 under bid-ask costs and 0.45 under price impact, compared with 0.63 and 0.47 for the baseline. Short-term reversal, return skewness, and maximum daily return lose both exposure and attribution under cost-awareness, and stable fundamentals such as book leverage stay near the origin of the axes.

5.2 Linear SDF Specification

Removing the activation functions from both networks makes the characteristic-to-weight mapping and the instrument linear in the characteristics and the macroeconomic state. To see this, assume for simplicity a simple network with two hidden layers; with linear activation functions the mapping collapses to $f^{\text{linear}}(h_t^\omega, I_{i,t}; \vartheta_\omega) = W_\omega^{\text{combined}}[h_t^\omega, I_{i,t}] + b_\omega^{\text{combined}}$, with $W_\omega^{\text{combined}} = W_\omega^{(2)} W_\omega^{(1)}$ and $b_\omega^{\text{combined}} = W_\omega^{(2)} b_\omega^{(1)} + b_\omega^{(2)}$, so that portfolio weights are a linear function of stock characteristics.

	Nonlinear SDF (baseline)				Linear SDF			
	EV	XS-R^2	SR		EV	XS-R^2	SR	
			Gross	Net			Gross	Net
M_{t+1} (frictionless)	0.063	0.145	0.348	0.064	0.053	0.160	0.256	0.028
M^{TC}_{t+1} (bid-ask)	0.098	0.223	0.317	0.113	0.078	0.198	0.268	0.084
M^{TC}_{t+1} ($5m)	0.100	0.222	0.326	0.311	0.076	0.193	0.310	0.294
M^{TC}_{t+1} ($50m)	0.114	0.234	0.244	0.225	0.127	0.243	0.248	0.212
M^{TC}_{t+1} ($500m)	0.143	0.258	0.157	0.139	0.143	0.257	0.123	0.088

Table 4: **Linear vs. nonlinear SDF.** This table reports out-of-sample performance metrics for the GAN architecture under each transaction cost specification and the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. EV is explained variation; XS-R^2 is cross-sectional R^2; SR (Gross) and SR (Net) are monthly Sharpe ratios before and after transaction costs. Net Sharpe ratios for the frictionless SDF are computed under bid-ask costs. Both panels use the nonlinear beta network of Appendix A.3; Appendix D.2 reports the full decomposition across SDF and beta network specifications. The test period is January 1997 to November 2023.

Table 4 compares linear vs non-linear SDF specifications, fixing the adversarial objective, the inputs, the hyperparameters, and the estimation procedure. Removing the nonlinearity changes the pricing metrics of the frictionless SDF little: explained variation falls from 0.063 to 0.053, and cross-sectional R^2 rises from 0.145 to 0.160. Within the linear model, cost-awareness raises explained variation by 0.023 to 0.090 and cross-sectional R^2 by 0.033 to 0.097 across the four specifications. These amounts are comparable to those within the nonlinear model. The two architectures converge as the cost binds. At $50m, the linear cost-aware kernel attains higher explained variation than the nonlinear one (0.127 vs. 0.114), and at $500m the two are indistinguishable, with explained variation of 0.143 for both and cross-sectional R^2 of 0.257 and 0.258. A binding cost penalty determines which characteristics the SDF can use, and once it does, flexibility in combining them adds little pricing accuracy. Appendix D.2 repeats the exercise for the auxiliary network that estimates the conditional betas and reaches the same conclusion.

Figure D.6 in Appendix D.10 shows that the composition of the linear portfolios changes along one margin only. Under bid-ask costs, the frictionless linear portfolio trades 29% more than its cost-aware counterpart (turnover ratio of 1.29, against 1.36 for the nonlinear baseline), while the two are equally concentrated (HHI ratio of 1.03, against 1.29).

5.3 Uniform Costs Across Stocks

The third main ingredient of the cost-aware SDF specification is stock-level variation in trading frictions. Proposition 1 attributes the reallocation to how much a characteristic trades. Stock-specific costs confound turnover with stock expense. Replacing each $c_{i,t}$ with its monthly cross-sectional mean removes the second source and leaves ARC_j in Proposition 1 as the only variation in the penalty across characteristics. The second row of Table 5 in Section 6.1 reports the results.

Pricing accuracy is essentially unchanged, with explained variation of 0.099 against 0.098 and cross-sectional R^2 of 0.225 against 0.223. The two SDFs have nearly the same tangency portfolio: characteristic exposures rank-correlate at 0.997 with the baseline and DeepSHAP attributions at 0.997, portfolio concentration correlates at 0.9999, the share of short positions at 0.998, and turnover at 0.995. Uniform costs raise turnover by 1.8%, and the exposure differences they produce rank-correlate at -0.31 with ARC, so the stock-level cost data change which stocks carry a given exposure and leave the pricing of the cross-section unchanged. This result holds only for proportional costs. Under price impact, the coefficient $\Lambda_{i,t}$ is stock-specific by construction and scales with wealth, so the cost cross-section cannot be flattened without completely changing the cost function.

Overall, the results show that the pricing gain survives removing the adversarial objective, the nonlinearity in the pricing kernel, and stock-month cost heterogeneity in the cross-section.

What the estimator adds to the portfolio. Under bid-ask costs, the frictionless feedforward portfolio earns a monthly net Sharpe ratio of 0.051 versus 0.118 for its cost-aware counterpart, and 0.028 versus 0.084 for the frictionless linear portfolio. At \$500m, the frictionless feedforward portfolio earns -1.241 against 0.104, so the ordering of net Sharpe ratios reported for the baseline in Section 4.3 holds in both alternative estimators.

Holding the cost specification fixed, the baseline earns a higher gross Sharpe ratio than the feedforward network in every row of Table 3, from 0.317 against 0.253, monthly, under bid-ask

costs to 0.157 against 0.115 at \$500m, and a higher net Sharpe ratio in every row except under bid-ask costs (0.113 against 0.118). Removing the nonlinearity lowers the monthly net Sharpe ratio under every specification in Table 4, from 0.064 to 0.028 for the frictionless kernel and from 0.139 to 0.088 at \$500m. The adversarial objective and the nonlinearity therefore raise the Sharpe ratio of the tangency portfolio, the quantity for which Hansen and Jagannathan (1991) states the bound, while the pricing gain comes from the moment condition stated net of costs.

6 Robustness

In this Section, we study the robustness of the results by stress-testing the baseline adversarial network along several dimensions: the cost function, the sample, the market regime, and the position limits. We report cross-sectional pricing accuracy, measured by explained variation and cross-sectional R^2, and the tangency portfolio's net Sharpe ratio under each specification. All results are out of sample, from January 1997 to November 2023, unless otherwise noted.

6.1 Cost Specifications

We replace the baseline costs in two ways. The first replaces bid-ask spreads with the size-based linear approximation of Brandt et al. (2009),

$$c_{i,t} = 0.006 - 0.0025 \times \text{me}_{i,t}, \tag{14}$$

where $\text{me}_{i,t}$ is the market-capitalization percentile of firm i normalized between 0 and 1. This specification charges one-way costs between 0.60% for the smallest firms and 0.35% for the largest, with little variation over time, so it is milder in the cross-section and less variable than the baseline half-spreads (median 0.5%, 90th percentile above 2%, spikes above 3% in stressed markets, Figure 1), although 0.35% for the largest firms is high by the standards of the later

	EV		XS-R^2		SR_{net}		Improvement	
Cost specification	M_{t+1}	M^{TC}_{t+1}	M_{t+1}	M^{TC}_{t+1}	M_{t+1}	M^{TC}_{t+1}	ΔEV	$\Delta\text{SR}_{\text{net}}$
Bid-ask costs	0.063	0.098	0.145	0.223	0.064	0.113	+56%	+0.049
Bid-ask, uniform $\bar{c}_t$	0.063	0.099	0.145	0.225	0.064	0.107	+56%	+0.043
Brandt size-based	0.063	0.091	0.145	0.220	0.167	0.165	+44%	−0.002
$\sqrt{}$-impact, \$500m	0.063	0.102	0.145	0.221	0.103	0.233	+62%	+0.130
$\sqrt{}$-impact, \$5bn	0.063	0.131	0.145	0.253	−0.462	0.146	+108%	+0.608

Table 5: **Pricing performance and SRs under alternative cost specifications.** The table reports out-of-sample pricing performance (EV and XS-R^2) and net Sharpe ratios (SR_{net}) across four cost specifications: bid-ask spreads (baseline), the size-based linear approximation of Brandt et al. (2009) (Equation 14), and the square-root price-impact specification (Equation 15) at two capacity levels. "M" denotes the frictionless SDF; "M^{TC}" denotes the cost-aware SDF. Within each row, performance is evaluated under the cost specification of that row, so SR_{net} for M varies across rows because the realized cost differs across cost technologies. "%" indicates the percentage improvement of M^{TC} relative to M. "Δ" denotes the difference in SR_{net} between M^{TC} and M.

part of our sample.[19] Appendix D.11 reports the Brandt cost distribution.

The second specification replaces the quadratic price impact with the square-root specification of Almgren et al. (2005),

$$\mathcal{TC}_t = \kappa \sum_i \sigma_{i,t} \left|\Delta\omega_{i,t}\right|^{3/2} \sqrt{w_t / V_{i,t}}, \qquad \kappa\,\bar{\sigma} = 0.01, \tag{15}$$

calibrated to the same anchor as the quadratic baseline, that is, a trade of 1% of average daily volume produces 10 basis points of impact.

Impact per dollar traded grows in proportion to the size of the trade under quadratic costs and in proportion to its square root under Equation (15), so large trades cost less under the square-root law, and an investor of a given size pays less to hold the same portfolio. At \$500m, the frictionless tangency portfolio earns a net Sharpe ratio of −1.355 under quadratic costs and 0.103 under square-root costs (Tables 2 and 5). We therefore evaluate the square-root specification at \$500m, the largest size under quadratic costs, and at \$5bn, where the frictionless monthly net Sharpe ratio falls to −0.462 as costs are most binding.

[19] The calibration of Brandt et al. (2009) is estimated on data that end in 2002, before most of the decline in spreads visible in Figure 1.

Table 5 reports pricing and net Sharpe ratios under the baseline and the alternative costs. The frictionless SDF incurs the same costs as the cost-aware one in each row. Under Brandt costs (third row in the table), the explained variation rises from 0.063 to 0.091 and cross-sectional R^2 from 0.145 to 0.220; these are improvements of 44% and 52%, close to the bid-ask improvements of 56% and 54%, although the costs charged are milder on average. Under square-root impact, the improvement grows with capacity, with explained variation rising by 62% at \$500 million and 108% at \$5 billion, and cross-sectional R^2 by 52% and 74%. Across specifications, the pricing gain increases with the severity of the cost the SDF charges.

The monthly net Sharpe ratio moves differently. Under Brandt costs, the frictionless portfolio earns 0.167 and the cost-aware portfolio 0.165, while pricing improves by 44%. Under square-root impact, the cost-aware portfolio earns 0.233 against 0.103 at \$500 million, and 0.146 against -0.462 at \$5 billion, where the frictionless tangency portfolio is one that no investor of that size could hold.

Appendix D.11 shows how the reallocation in Section 4.3 appears under each alternative cost and scales with its severity. The concentration and turnover ratios between the frictionless and cost-aware portfolios are 1.26 and 1.23 under Brandt costs, 1.76 and 1.94 under square-root impact at \$500 million, and 3.42 and 3.09 at \$5 billion. The shifts in characteristic exposure under Brandt costs rank-correlate at 0.77 with those under bid-ask costs, and the same turnover-intensive characteristics lead the reallocation under square-root impact.

6.2 Sample Construction

Figure 1 shows that bid-ask spreads were higher before the mid-1990s and lowest after 2010, so an earlier sample trains the adversarial network under heavier frictions than a later one. To gauge the effect of trading frictions in the training period, we re-estimate the frictionless and cost-aware SDF under different cost specifications over an early sample that begins in January 1967 with a reduced set of 62 characteristics. The testing period thus starts from January 1991. Table 6 reports the results.

	EV	XS-R^2	SR (Gross)	SR (Net)
Baseline (1972–2023), test: 1997–2023				
M_{t+1} (frictionless)	0.063	0.145	0.348	0.064
M_{t+1}^{TC} (bid-ask)	0.098	0.223	0.317	0.113
Early sample (1967–2023), test: 1991–2023				
M_{t+1} (frictionless)	0.086	0.224	0.232	−0.073
M_{t+1}^{TC} (bid-ask)	0.105	0.250	0.221	−0.005
M_{t+1}^{TC} ($5m)	0.093	0.232	0.244	0.228
M_{t+1}^{TC} ($50m)	0.025	0.121	0.179	0.165
M_{t+1}^{TC} ($500m)	0.067	0.195	0.140	0.109

Table 6: **Early sample period.** This table reports out-of-sample performance metrics for the GAN architecture under each transaction cost specification. EV is explained variation; XS-R^2 is cross-sectional R^2; SR (Gross) and SR (Net) are monthly Sharpe ratios before and after transaction costs. The leverage constraint is $\omega_{i,t} \in (-5\%, +5\%)$.

Under bid-ask costs, the cost-aware kernel raises explained variation from 0.086 to 0.105 and cross-sectional R^2 from 0.224 to 0.250, gains of 22% and 12%. The test period opens in the high-spread early 1990s; both tangency portfolios earn negative net Sharpe ratios, and cost-awareness narrows the shortfall from −0.073 to −0.005. At $5m price impact, explained variation rises to 0.093. At $50m and $500m, the cost-aware kernel earns positive net Sharpe ratios of 0.165 and 0.109, and its pricing metrics fall below the frictionless benchmark, with explained variation of 0.025 and 0.067.

This is the one configuration in the paper in which charging for costs worsens pricing. The capacities are fixed in nominal dollars, so a kernel trained over 1967–1986, when $50m or $500m was a far larger share of daily dollar volume than in the baseline training window, learns an aversion to turnover calibrated to that era and carries it into a test period where the frictionless benchmark faces much less impact.[20]

We also consider training under a later sample that begins in January 1986 with 88 characteristics. The testing period is from January 2011, so its test window excludes the high-cost 1990s. Each keeps the 20-year training and 5-year validation windows. Table D.7 in Ap-

[20]The baseline is not affected in the same way, because its training window (1972–1991) and test window (1997–2023) are closer in dollar volume, and because price impact at $500m still binds in the baseline test window, where the frictionless net Sharpe ratio is −1.36 (Table 2).

pendix D.12 shows the results. The explained variation and cross-sectional R^2 fall under bid-ask, \$5m, and \$50m costs, and rise only at \$500m, where explained variation reaches 0.137 against 0.073 and cross-sectional R^2 reaches 0.216 against 0.160. Under bid-ask costs, the net Sharpe ratio rises from 0.167 to 0.236. The test window is the cheapest in the sample,[21] and a moment condition stated net of a friction that barely binds carries little additional pricing information until the friction is large enough to bind, as it is at \$500m. The pricing gain in the baseline is therefore a property of a period in which costs bind, and not of the sample period as such.

Appendix D.12 shows that the reallocation scales with the cost environment across the three samples. Under bid-ask costs, the median ratio of frictionless to cost-aware portfolio concentration falls from 1.67 in the early sample to 1.29 in the baseline and 1.02 in the later sample, and the median turnover ratio falls from 1.65 to 1.36 and 1.19 (Figure D.15). Compared with the baseline, the absolute change in median exposure between the two kernels is, on average, 48% larger in the early sample and 22% smaller in the later sample (Figure D.16). Because the reallocation is largest for the SDF estimated under the heaviest frictions and smallest for the SDF evaluated in the cheapest window, the pricing gain in the baseline is a property of a period in which costs bind, which the 1997–2023 window is, and not of the sample period as such.

6.3 Market Conditions

Avramov et al. (2023) show that machine learning methods perform best in the illiquid corners of the cross-section and across periods of tight limit-to-arbitrage—which casts doubt on whether they can be implemented at all for pricing—, so an SDF that charges for trading should be judged in those states. We evaluate the estimated kernels across slices of the test period rather than re-estimating them, which bounds their adaptation from below.[22] We consider two time-series partitions that use the National Financial Conditions Index (NFCI), where

[21] Panel A of Figure 1 puts the 90th percentile of the half-spread distribution near 1% throughout, against peaks above 4% in the early 1990s and near 3% in 1975 and 2001.

[22] Re-estimating over subsamples would let the networks adjust to each environment. Our results therefore understate the cost-aware kernel's ability to adapt.

	M_{t+1} (frictionless)				M^{TC}_{t+1} (bid-ask)			
	EV	XS-R^2	SR (Gross)	SR (Net)	EV	XS-R^2	SR (Gross)	SR (Net)
Full cross-section	0.063	0.145	0.348	0.064	0.098	0.223	0.317	0.113
Excl. microcaps	0.117	0.196	0.223	0.022	0.147	0.226	0.218	0.067
Tight fin. conditions	0.090	0.077	−0.104	−0.423	0.139	0.135	−0.129	−0.350
Loose fin. conditions	0.059	0.141	0.409	0.131	0.090	0.216	0.381	0.180
High VIX	0.073	0.132	0.364	0.063	0.110	0.212	0.330	0.115
Low VIX	0.040	0.115	0.364	0.074	0.061	0.152	0.336	0.126

Table 7: **Performance across market conditions.** This table reports out-of-sample performance metrics for the GAN architecture under the baseline bid-ask cost specification and leverage constraint of $\omega_{i,t} \in (-5\%, +5\%)$. M_{t+1} denotes the frictionless SDF; M^{TC}_{t+1} denotes the TC-aware SDF. EV is explained variation; XS-R^2 is cross-sectional R^2; SR (G) and SR (N) are monthly Sharpe ratios before and after transaction costs. Financial conditions are classified using the National Financial Conditions Index (NFCI) from the Chicago Fed; market uncertainty is classified using the VIX median. The test period is January 1997 to November 2023.

positive values mark tighter-than-average conditions, and the VIX split at its median.[23] A third partition drops the bottom quartile of firms by market capitalization, which tests whether cost-awareness exploits microcap frictions alone.

Table 7 shows that pricing improves in every subsample and that the improvement grows with the severity of the friction. Cross-sectional R^2 rises by 75% during tight financial conditions, from 0.077 to 0.135, against 53% when conditions are loose, and by 61% in high-uncertainty months against 32% in low-uncertainty months. Net Sharpe ratios improve alongside the pricing metrics. Both SDFs earn negative net Sharpe ratios under tight financial conditions (−0.423 frictionless and −0.350 cost-aware), and under loose conditions, cost-awareness raises the net Sharpe ratio from 0.131 to 0.180.

Excluding the smallest quartile of firms directly addresses the objection of Avramov et al. (2023). Within the more liquid segment of the cross-section, the frictionless kernel prices better than in the full cross-section, with explained variation of 0.117 against 0.063 and cross-sectional R^2 of 0.196 against 0.145. The net Sharpe ratio of the tangency portfolio remains close to zero at 0.022, against 0.067 for the cost-aware kernel, and pricing improves further, with explained variation rising by 26% and cross-sectional R^2 by 15%. The gain from cost-awareness thus

[23] For each month t: conditions are Tight if $\text{NFCI}_t > 0$ and Loose otherwise. Uncertainty is High if VIX_t exceeds its median and Low otherwise.

	EV	XS-R^2	SR (Gross)	SR (Net)	ΔEV	ΔSR$_{\text{net}}$
Loose: $\pm 10\%$						
M_{t+1} (frictionless)	0.067	0.147	0.352	0.047		
M_{t+1}^{TC} (cost-aware)	0.096	0.209	0.334	0.143	+43%	+0.096
Baseline: $\pm 5\%$						
M_{t+1} (frictionless)	0.063	0.145	0.348	0.064		
M_{t+1}^{TC} (cost-aware)	0.098	0.223	0.317	0.113	+56%	+0.049
Tight: $\pm 2.5\%$						
M_{t+1} (frictionless)	0.065	0.143	0.286	0.063		
M_{t+1}^{TC} (cost-aware)	0.117	0.241	0.231	0.084	+80%	+0.021

Table 8: **Alternative leverage constraints.** This table reports out-of-sample performance metrics for the GAN architecture under the baseline bid-ask cost specification. EV is explained variation; XS-R^2 is cross-sectional R^2; SR (Gross) and SR (Net) are monthly Sharpe ratios before and after transaction costs. Constraints are ordered from loose to tight. The test period is January 1997 to November 2023.

extends beyond the most illiquid stocks. Appendix D.13 shows that under the three price-impact specifications, every regime-cost combination improves, and at \$500m the cost-aware kernel turns negative frictionless net Sharpe ratios positive in every regime.

6.4 Leverage Constraints

Intuitively, position limits restrict turnover outside the estimation, so a pricing gain from restricted trading alone should weaken as the limits tighten. In this Section, we re-estimate both kernels under loose ($\pm 10\%$) and tight ($\pm 2.5\%$) position limits alongside the baseline of $\pm 5\%$, under bid-ask costs.[24]

Table 8 reports the two kernels under the three limits. Under $\pm 10\%$ limits, the frictionless kernel trades the most, and its net Sharpe ratio falls to 0.047, below the 0.064 it earns at the baseline, while the cost-aware kernel earns 0.143. As limits tighten, they restrict turnover, and the gap in net Sharpe ratios narrows from 0.096 to 0.049 and 0.021, monthly. The pricing gain moves the other way. Explained variation improves by 43% under loose limits, 56% at the

[24] A strict no-short-sale constraint would remove the kernel's ability to take negative characteristic exposures, degrading pricing under both specifications, and would interact with estimation error in the manner described by Jagannathan and Ma (2003). Short-selling intensity is, in any case, nearly unchanged under cost-awareness, because shorts are costless in our specifications, so the reallocation operates within each leg.

baseline, and 80% under tight limits, where it rises from 0.065 to 0.117, and cross-sectional R^2 by 42%, 54%, and 69%. The pricing gain is largest when position limits come closest to doing the cost penalty's work on turnover, although tighter limits also change the feasible set in other ways. One reading is that a cap on the weight of any single stock forces a given characteristic exposure to be assembled from more stocks, including stocks the SDF would otherwise avoid because they are too costly.

Figure D.17 in Appendix D.14 shows that the turnover ratios between the frictionless and cost-aware portfolios rise from 1.27 under tight limits to 1.36 at the baseline and 1.44 under loose limits, and the correlation between exposure shifts and DeepSHAP attribution shifts rises from 0.43 to 0.75 (Figure D.18). The frictionless kernel's exposure to short-term reversal, return skewness, and maximum daily return grows with the room the position limits give it.

7 Conclusion

We estimate the stochastic discount factor from a no-arbitrage condition stated net of the cost of rebalancing its tangency portfolio. Out of sample, the cost-aware kernel prices individual stocks better under every cost specification, with explained variation rising from 0.063 to 0.098 under bid-ask costs and to 0.143 under price impact at \$500m. Its tangency portfolio is less concentrated, trades less, and holds less exposure to high-turnover characteristics. The pricing errors that fall are those of anomalies whose premia do not survive their own trading costs, led by idiosyncratic volatility, maximum daily return, and the zero-trade measures, while the pricing of the eleven anomalies that survive does not improve.

The gain persists without the adversarial objective, without nonlinearity, and with stock-level costs replaced by their cross-sectional average, and across sample periods, the reallocation behind it scales with how much costs bind. Under the milder costs of Brandt et al. (2009), the two tangency portfolios earn the same net Sharpe ratio while explained variation rises by 44%, whereas the two gains would move together if the cost-aware kernel merely recovered the portfolio an investor would hold under costs (Jensen et al., 2026).

Under price impact, the kernel prices the cross-section for an investor of a given size; at \$5 billion, the frictionless tangency portfolio earns a net Sharpe ratio of -0.462 versus 0.146 for the cost-aware one. A single kernel for investors of different sizes, and a direct comparison with an estimator that maximizes net-of-cost utility, remain open questions.

References

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

Almgren, R., Thum, C., Hauptmann, E., and Li, H. (2005). Direct estimation of equity market impact. *Risk*, 18(7):58–62.

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

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.

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.

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.

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

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.

Feng, G., Jiang, L., Li, J., Song, Y., and Wang, Y. (2026). Deep Tangency Portfolio. *Management Science*. Forthcoming.

Frazzini, A., Israel, R., and Moskowitz, T. J. (2018). Trading costs. *Available at SSRN 3229719*.

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.

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.

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.

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

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

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., Malamud, S., and Pedersen, L. H. (2026). Machine learning and the implementable efficient frontier. *The Review of Financial Studies*, page hhag022.

Kaji, T., Manresa, E., and Pouliot, G. (2023). An adversarial approach to structural estimation. *Econometrica*, 91(6):2041–2063.

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.

Korajczyk, R. A. and Sadka, R. (2004). Are momentum profits robust to trading costs? *The Journal of Finance*, 59(3):1039–1082.

Korsaye, S. A., Quaini, A., and Trojani, F. (2025). Smart stochastic discount factors. *Management Science*.

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.

Lesmond, D. A., Schill, M. J., and Zhou, C. (2004). The illusory nature of momentum profits. *Journal of financial economics*, 71(2):349–380.

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

Luttmer, E. G. (1996). Asset pricing in economies with frictions. *Econometrica: Journal of the Econometric Society*, pages 1439–1467.

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. (2012). Evaporating liquidity. *The Review of Financial Studies*, 25(7):2005–2039.

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.

Novy-Marx, R. and Velikov, M. (2019). Comparing cost-mitigation techniques. *Financial Analysts Journal*, 75(1):85–102.

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

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics*, 42(1-2):35–52.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *The Journal of finance*, 52(1):35–55.

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.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948.

Supplementary Appendix for:

Transaction Costs and the Stochastic Discount Factor

This appendix is organized as follows. Appendix A provides implementation details for the neural network estimation and the proof of Proposition 1. Appendix B describes the data, including the sample selection procedure, the construction of firm characteristics and macroeconomic variables, and the computation of expected daily dollar volume used for the price-impact calibration. Appendix C details the construction of characteristic-managed portfolios and their transaction costs, as well as the frontier-expansion measure used in the pairwise spanning tests. Appendix D reports additional empirical results referenced in the main text, covering beta-sorted portfolios and alternative beta estimates, anomaly pricing and characteristic reallocation at alternative investor sizes, frontier expansion and spanning against observable factors, the feedforward network and the linear SDF, alternative cost specifications, alternative samples, market conditions, and alternative leverage constraints.

A Estimation and Implementation Details

This section describes the network architecture and hyperparameter choices, the upstream pre-training procedure, the `DeepSHAP` attribution methodology, and the auxiliary beta estimation network.

A.1 Network Architecture and Loss Functions

Figure A.1 visualizes the GAN 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 that generates hidden states h_t^ω, which are combined with firm-level characteristics $I_{i,t}$ to produce portfolio weights $\widehat{\omega}(h_t^\omega, I_{i,t})$. The transaction cost penalty enters directly through the rebal-

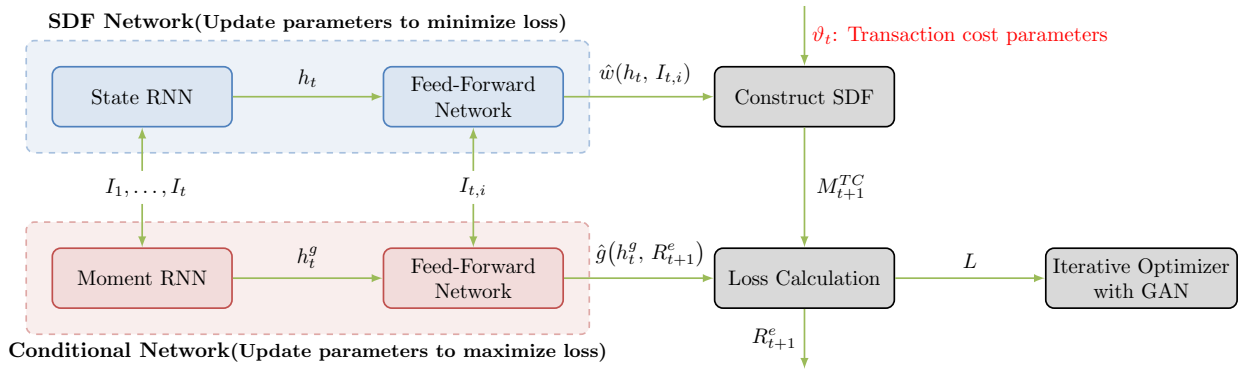


Figure A.1: **GAN architecture for cost-aware SDF estimation**. This figure illustrates the adversarial training framework used to estimate transaction-cost-aware stochastic discount factors.

ancing measure $\Delta\omega_{i,t}$ and the cost coefficient $\Lambda_{i,t}$ (or bid-ask spread $c_{i,t}$, depending on the cost specification). The instrument network (bottom, red) conditions on a parallel LSTM to generate hidden states h_t^g, which are combined with stock characteristics to produce the moment conditions. We discuss each element in turn.

LSTM specifications. As described in Section 2.2, we do not feed the macroeconomic indicators I_s directly into the feedforward networks. Instead, two parallel LSTM networks first process them, compressing the time-series history $(I_1, \ldots, I_t)$ into low-dimensional state vectors h_t^ω and h_t^g. Both LSTM networks share the same functional form. For $x \in \{\omega, g\}$, the LSTM update equations are:

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

$$o_s^x = \sigma\big(W_o^x I_s + U_o^x h_{s-1}^x + b_o^x\big), \qquad \tilde{c}_s^x = \tanh\big(W_c^x I_s + U_c^x h_{s-1}^x + b_c^x\big), \tag{A.2}$$

$$c_s^x = f_s^x \odot c_{s-1}^x \; + \; i_s^x \odot \tilde{c}_s^x, \qquad h_s^x = o_s^x \odot \tanh(c_s^x), \qquad s = 1, \ldots, t, \tag{A.3}$$

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; 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 the macroeconomic encoding. Following Chen et al. (2024), the SDF network uses LSTM$^\omega$ with hidden dimension 4, while the moment network uses LSTM^g with hidden dimension 32.

Feedforward network layers. The LSTM state vectors h_t^ω and h_t^g are concatenated with asset-level characteristics $I_{i,t}$ and passed to feedforward networks that produce portfolio weights and test assets, respectively. For a neural network with L hidden layers, the SDF weight function $f(h_t^\omega, I_{i,t}; \vartheta_\omega)$ in Equation (10) is computed as:

$$f(h_t^\omega, I_{i,t}; \vartheta_\omega) = W_\omega^{(L+1)} h_\omega^{(L)} + b_\omega^{(L+1)} \tag{A.4}$$

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

$$h_\omega^{(1)} = \text{ReLU}(W_\omega^{(1)}[h_t^\omega, I_{i,t}] + b_\omega^{(1)}) \tag{A.5}$$

$$h_\omega^{(l)} = \text{ReLU}(W_\omega^{(l)} h_\omega^{(l-1)} + b_\omega^{(l)}) \quad \text{for} \quad l \in \{2, 3, ..., L\} \tag{A.6}$$

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. The resulting portfolio weights are $\omega_{i,t} = \kappa\big(\sigma(f(h_t^\omega, I_{i,t}; \vartheta_\omega)) - 0.5\big)$. The instrument function $g_{j,t} = g(h_t^g, I_{j,t}; \vartheta_g)$ has no hidden layer. It applies a tanh transformation to a linear combination of $[h_t^g, I_{j,t}]$, with parameters ϑ_g, and returns the $d = 8$ instruments, one for each unconditional moment condition.

Empirical loss functions. Having defined the network architectures, we now describe the loss functions used to estimate their parameters. Following Chen et al. (2024), we replace the expectation in Equation (9) by the time-series average of each stock's pricing error over the training sample and evaluate the loss on the full training panel at every gradient step, with no mini-batching. The empirical loss for the GAN estimation in Section 2.2 is

$$\mathcal{L}^{\text{GAN-TC}}(\vartheta_\omega, \vartheta_g) = \frac{1}{N} \sum_{j=1}^{N} \frac{T_j}{T} \left\| \frac{1}{T_j} \sum_{t \in \mathcal{T}_j} M_{t+1}^{TC}(\vartheta_\omega)\, r_{j,t+1}^e\, g_{j,t} \right\|^2 \tag{A.7}$$

where $M_{t+1}^{TC}(\vartheta_\omega)$ is defined in Equation (11), $\mathcal{T}_j$ is the set of training months in which stock j is observed, T_j is its size, and T is the number of months in the training sample. We average the pricing error of each instrumented position over time before squaring it, and the weight

T_j/T gives more weight to moments estimated over longer histories in the unbalanced panel. All parameters—including the LSTM weights—are updated jointly through backpropagation. The moment network parameters ϑ_g are updated by gradient ascent on $\mathcal{L}^{\text{GAN-TC}}$, while the SDF parameters ϑ_ω are updated by gradient descent, as described in Algorithm 1.

GAN training algorithm. Algorithm 1 below summarizes the GAN procedure described in Section 2.2. The moment network parameters ϑ_g are updated by gradient ascent on the pricing loss. The SDF parameters ϑ_ω are updated by gradient descent.

Algorithm 1: Cost-Aware SDF Estimation with GAN

Input : Pretrained parameters $\vartheta_\omega^{\text{init}}$, $\vartheta_g^{\text{init}}$ from Algorithm 3 (Appendix A.2); hyperparameters: total cycles K, moment network epochs n_{mom}, SDF network epochs n_{sdf}, gradient steps per epoch S, learning rates η_g, η_ω

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**

      for *step* $s = 1$ **to** S **do**

        Fix ϑ_ω at current value;

        Compute loss $\mathcal{L}^{\text{GAN-TC}}(\vartheta_\omega, \vartheta_g)$ on the full training sample;

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

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

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

    for *step* $s = 1$ **to** S **do**

      Fix ϑ_g at current value;

      Compute loss $\mathcal{L}^{\text{GAN-TC}}(\vartheta_\omega, \vartheta_g)$ on the full training sample;

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

  Evaluate model on validation set;

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

Output: Parameters $\widehat{\vartheta}_\omega$ with the highest validation Sharpe ratio and corresponding $M_{t+1}^{\text{TC}}(\widehat{\vartheta}_\omega)$

In our baseline implementation, we set $n_{\text{mom}} = 64$ and $n_{\text{sdf}} = 256$ epochs for the instrument and SDF networks, respectively, and each epoch consists of $S = 4$ gradient steps on the full training sample. All gradient-based updates use Adam with learning rates $\eta_g = \eta_\omega = 0.001$. We run a single cycle ($K = 1$), as further cycles do not improve the fit (Chen et al., 2024). 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 cost-aware SDFs. Appendix A.2 provides a more detailed description of this pre-training routine.

FFN training algorithm. For the Feed-Forward Neural Network specification in Section 5.1, the empirical loss sets $g_{j,t} \equiv 1$ and is again evaluated on the full training sample:

$$\mathcal{L}^{\text{FFN-TC}}(\vartheta_\omega) = \frac{1}{N}\sum_{j=1}^{N}\frac{T_j}{T}\left(\frac{1}{T_j}\sum_{t\in\mathcal{T}_j} M_{t+1}^{TC}(\vartheta_\omega)\, r_{j,t+1}^{e}\right)^2 \tag{A.8}$$

where the cost-aware SDF $M_{t+1}^{TC}(\vartheta_\omega)$ maintains the same form as in Equation (11). Algorithm 2 summarizes the training procedure.

Algorithm 2: Cost-Aware SDF Estimation with FFN

Input : Pretrained parameters $\vartheta_\omega^{\text{init}}$; hyperparameters: total epochs E, gradient steps per epoch S, learning rate η_ω

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

for *epoch* $e = 1$ **to** E **do**

  for *step* $s = 1$ **to** S **do**

    Compute loss $\mathcal{L}^{\text{FFN-TC}}(\vartheta_\omega)$ according to Eq. (A.8) on the full training sample;

    Update $\vartheta_\omega \leftarrow \vartheta_\omega - \eta_\omega \nabla_{\vartheta_\omega}\mathcal{L}^{\text{FFN-TC}}$;

  Evaluate model on validation set;

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

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

Hyperparameter configurations. The SDF network has 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 and asset-specific information as inputs to generate portfolio weights. The feedforward component includes two hidden layers, each with 64 hidden units.

The instrument network uses an LSTM cell to encode macroeconomic dynamics into 32 hidden states and has no hidden layer, so the eight instruments are a tanh transformation of a linear combination of these hidden states and the asset characteristics, following the

generalized linear specification of Chen et al. (2024). The instruments capture pricing errors of portfolios constructed from asset characteristics or state-dependent payoffs under particular macroeconomic scenarios. The two state vectors differ in size because the hyperparameter search of Chen et al. (2024) selects four states for the SDF network and 32 for the instrument network, and the moment conditions that identify the SDF can depend on different states than the SDF weights (Chen et al., 2024, footnote 18).

We adopt the optimal network configuration and hyperparameter choices from Chen et al. (2024), who search across 384 distinct network configurations, fit the GAN model for each combination, and select the top configurations based on validation-set performance. Holding the network architecture and hyperparameters fixed across the frictionless and cost-aware SDFs ensures a level playing field for the comparison. Table A.1 summarizes the resulting choices.

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 0.0001	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. With no hidden layer in the instrument network, its 8 hidden units are the $d = 8$ instruments.

A.2 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): The empirical loss is

$$\mathcal{L}(\vartheta_\omega) = \frac{1}{N}\sum_{j=1}^{N}\frac{T_j}{T}\left(\frac{1}{T_j}\sum_{t\in\mathcal{T}_j} M_{t+1}(\vartheta_\omega)\, r^e_{j,t+1}\right)^2, \tag{A.9}$$

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 ϑ_ω. The moment network remains switched off $\big(g_{j,t} \equiv 1\big)$ and transaction costs are ignored. Gradient descent on $\mathcal{L}(\vartheta_\omega)$ is performed for n_{sdf} epochs using Adam with learning rate 0.001.

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

$$\mathcal{L}(\vartheta_g;\vartheta^\star_\omega) = \frac{1}{N}\sum_{j=1}^{N}\frac{T_j}{T}\left\|\frac{1}{T_j}\sum_{t\in\mathcal{T}_j} M_{t+1}(\vartheta^\star_\omega)\, r^e_{j,t+1}\, g(I_t, I_{j,t};\vartheta_g)\right\|^2, \tag{A.10}$$

with

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

We perform gradient ascent on $\mathcal{L}(\vartheta_g;\vartheta^\star_\omega)$ for n_{mom} epochs using Adam, while keeping all other parameters in the SDF network fixed. Algorithm 3 summarizes the pretraining steps.

After this pretraining procedure, we obtain $\vartheta^\star_\omega$ and $\vartheta^\star_g$, which serve as the initialization for the conditional SDF training described in Algorithm 1 in Section 2.2. These pretrained parameters ensure that all models start conditional training under identical initial conditions. In our empirical analysis, we set $n_{\text{sdf}} = 256$ 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.

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_{sdf} **do**
  for *step* $s = 1$ **to** S **do**
    `/* Update` ω `only;` ϑ_g `is frozen`
    (i) Compute the loss $\mathcal{L}(\vartheta_\omega)$ from (A.9) on the full training sample.;
    (ii) $\vartheta_\omega \leftarrow \vartheta_\omega - \eta_\omega \nabla_{\vartheta_\omega} \mathcal{L}(\vartheta_\omega)$;
  Evaluate $\vartheta_\omega^\star$ based on the loss;

Step 2 (Moments pretraining);
for *epoch = 1 to* n_{mom} **do**
  for *step* $s = 1$ **to** S **do**
    `/* Update` g `only;` $\vartheta_\omega = \vartheta_\omega^\star$ `is frozen`
    (i) Compute the loss $-\mathcal{L}(\vartheta_g; \vartheta_\omega^\star)$ from (A.10) on the full training sample.;
    (ii) $\vartheta_g \leftarrow \vartheta_g + \eta_g \nabla_{\vartheta_g} \mathcal{L}$;
  Evaluate $\vartheta_g^\star$ based on the loss;

A.3 SDF Risk Loadings (Betas)

Without loss 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:

$$\widehat{\beta}_{i,t} = \mathbb{E}_t[r^e_{i,t+1} \widehat{F}_{t+1}] \approx f(I_{i,t}; \widehat{\vartheta}_\beta) \tag{A.11}$$

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

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

$$y_{i,t+1} = r^e_{i,t+1} \cdot \widehat{F}_{t+1} \tag{A.12}$$

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

and 5% dropout. The network parameters ϑ_β are estimated by minimizing:

$$\mathcal{L}(\vartheta_\beta) = \frac{1}{|\mathcal{I}|} \sum_{(i,t)\in\mathcal{I}} \left(y_{i,t+1} - \widehat{\beta}_{i,t} \right)^2 \tag{A.13}$$

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:

$$\widehat{\beta}_{i,t} = f(I_{i,t}; \vartheta^*_\beta) \tag{A.14}$$

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.

A.4 DeepLIFT Attribution and SHAP Values

Consider our tangency portfolio weight network $\omega(h^\omega_t, I_{i,t}; \vartheta_\omega)$ that maps the input $\mathbf{x} = [h^\omega_t, I_{i,t}] \in \mathbb{R}^{77}$, the four LSTM states and the 73 characteristics, to portfolio weight $\omega_i \in \mathbb{R}$. Let $\mathbf{x}^0 = (x^0_1, \ldots, x^0_{77})$ denote the reference baseline input for each input feature and $\Delta x_j = x_j - x^0_j$ the difference-from-reference for feature j. DeepLIFT assigns attribution scores $C_{\Delta x_j, \Delta\omega}$ such that:

$$\Delta\omega = \sum_{j=1}^{77} C_{\Delta x_j, \Delta\omega} \cdot \Delta x_j \tag{A.15}$$

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

DeepLIFT computes these attribution scores by propagating contribution ratios backward through the network, layer by layer, rather than using gradients, which can suffer from saturation in deep networks with ReLU activations (Dubey et al., 2022).

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":[25]

$$C_{\Delta y_i^{(l)}, \Delta z_i^{(l)}} = \begin{cases} 1 & \text{if } y_i^{(l)} > 0 \text{ and } y_i^{(l),0} > 0 \\ 0 & \text{if } y_i^{(l)} \leq 0 \text{ and } y_i^{(l),0} \leq 0 \\ \frac{\Delta z_i^{(l)}}{\Delta y_i^{(l)}} & \text{if they cross the activation boundary} \end{cases} \tag{A.17}$$

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:

$$C_{\Delta x_j, \Delta \omega} = \sum_{\text{paths } P} \prod_{(k,i) \in P} C_{\Delta y_k, \Delta y_i} \tag{A.18}$$

[25] 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:

$$C_{\Delta y_j^{(l)}, \Delta y_i^{(l+1)}} = W_{ij} \frac{\Delta y_j^{(l)}}{\sum_k W_{ik} \Delta y_k^{(l)}} \tag{A.16}$$

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.

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 reference input $\mathbf{x}^0 = (0, 0)$. Suppose $\Delta x_1 = 0.8$ (high profitability) and $\Delta x_2 = -0.4$ (small size). The hidden layer activation is $y_1^{(1)} = \text{ReLU}(W_{11} \cdot 0.8 + W_{12} \cdot (-0.4))$, 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 reference input, DeepLIFT traces this difference back through the network. For instance, it might find $C_{\Delta x_1, \Delta\omega} = +0.030$ and $C_{\Delta x_2, \Delta\omega} = +0.010$, such that $\Delta\omega = 0.030 \times 0.8 + 0.010 \times (-0.4) = 0.024 - 0.004 = 0.02$.

From DeepLIFT to DeepSHAP. DeepSHAP approximates Shapley values by averaging DeepLIFT attributions over multiple reference points. Given a background distribution $\mathcal{D}_{\text{background}}$, the DeepSHAP attribution is:

$$\phi_j^{\text{DeepSHAP}} = \mathbb{E}_{\mathbf{x}^0 \sim \mathcal{D}_{\text{background}}}[C_{\Delta x_j, \Delta\omega} | \mathbf{x}^0] \tag{A.19}$$

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. We use the DeepExplainer implementation of Lundberg and Lee (2017). The background distribution $\mathcal{D}_{\text{background}}$ consists of 3000 stock-months drawn once at random from the test sample, and each serves as a reference input $\mathbf{x}^0$. Because the references come from different months, they differ from the input for stock i in month t in both characteristics and LSTM states. For each stock i and month t of the test sample, the attributions $\phi_{j,i,t}$ sum to the difference between $\omega_{i,t}$ and the average weight over the background distribution, and we keep the attributions of the 73 characteristics. We average their absolute values over the stocks in each month,

$$\bar{\phi}_{j,t} = \frac{1}{N_t} \sum_{i=1}^{N_t} |\phi_{j,i,t}|, \tag{A.20}$$

and ΔSHAP^j in Figure 4 is the difference between the medians over months of $\bar{\phi}_{j,t}$ for the frictionless and cost-aware SDFs.

The resulting SHAP values measure the contribution of each input characteristic to the portfolio weight, 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 capture only local sensitivity and can be zero even when inputs are crucial to the final decision due to ReLU saturation. Algorithm 4 summarizes the computational procedure for calculating SHAP values using the DeepLIFT approach.

A.5 Proof of Proposition 1

From $\omega_{i,t} = \sum_j \theta_j \, I^j_{i,t}$, the change in portfolio weight is $\Delta\omega_{i,t} = \sum_j \theta_j \, \Delta I^j_{i,t}$, where the return-induced drift in inherited positions is negligible at the monthly rebalancing frequency. The trade generated by characteristic j alone has expected size $\mathbb{E}|\theta_j \, \Delta I^j_{i,t}| = |\theta_j| \, \text{ARC}_j$, which is the first statement of the proposition. By the triangle inequality, $\mathbb{E}|\Delta\omega_{i,t}| \leq \sum_j |\theta_j| \, \text{ARC}_j$, and we take the right-hand side as the trading of the portfolio, which ignores the netting of trades across characteristics. Since stock-level costs are uncorrelated with stock-level turnover, aggregating across stocks gives $\mathbb{E}[\mathcal{TC}_t] = N\bar{c} \sum_j |\theta_j| \, \text{ARC}_j$.

Let $F^j_{t+1} = \sum_i I^j_{i,t} \, r^e_{i,t+1}$ be the return on the portfolio managed on characteristic j, so that $\sum_i \omega_{i,t} r^e_{i,t+1} = \boldsymbol{\theta}' \mathbf{F}_{t+1}$. The pricing error of the cost-aware SDF for test asset n is

$$e^{TC}_n(\boldsymbol{\theta}) = \mathbb{E}[M^{TC}_{t+1} \, r^e_{n,t+1}] = \mu_n - \boldsymbol{\theta}' \boldsymbol{\Sigma}_n + \mathbb{E}[\mathcal{TC}_t \, r^e_{n,t+1}],$$

where $\mu_n = \mathbb{E}[r^e_{n,t+1}]$ and $\boldsymbol{\Sigma}_n = \mathbb{E}[\mathbf{F}_{t+1} \, r^e_{n,t+1}]$. The last term decomposes as $\mathbb{E}[\mathcal{TC}_t \, r^e_{n,t+1}] = \mathbb{E}[\mathcal{TC}_t] \, \mu_n + \text{Cov}(\mathcal{TC}_t, \, r^e_{n,t+1})$. Measurability of $\mathcal{TC}_t$ at time t does not imply zero covariance, since predetermined characteristics forecast returns. We assume that the covariance is pro-

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 set $\mathcal{S}$ of $B = 300$ stock-months drawn once from the test sample

Initialize: $\boldsymbol{\phi}_{i,t} \leftarrow \mathbf{0} \in \mathbb{R}^{77}$;

`/* DeepSHAP computation over background distribution */`

for *each background stock-month* $k \in \mathcal{S}$ **do**

  Set reference point $\mathbf{x}^0 \leftarrow \mathbf{x}_k$;

  `/* 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**

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

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

  for *input feature* $j = 1$ **to** 77 **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 $\boldsymbol{\phi}_{i,t}^{\text{firm}} \leftarrow (\phi_{5,i,t},\ldots,\phi_{77,i,t})$;

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

portional to the level term, $\mathbb{E}[\mathcal{TC}_t\, r^e_{n,t+1}] = \zeta\,\mathbb{E}[\mathcal{TC}_t]\,\mu_n$ with $\zeta > 0$. In the data, a regression of $\mathbb{E}[\mathcal{TC}_t\, r^e_{n,t+1}]$ on $\mathbb{E}[\mathcal{TC}_t]\,\mu_n$ across the 73 characteristic-managed portfolios that enter the estimation moment conditions, which are the assets the proposition concerns and a wider set than the 68 sorted portfolios of Section 4.2, gives $R^2 = 0.88$ and a slope of 1.42 under the baseline proportional-cost specification, and the two rank-correlate at 0.90. The orthogonal residual has a cross-sectional dispersion equal to 54% of that of the level term.

Write $e_n = \mu_n - \boldsymbol{\theta}'\boldsymbol{\Sigma}_n$ for the frictionless pricing error, $\mathbf{S} = [\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_N]$, and $\kappa = \zeta N \bar{c}$, so that $e_n^{TC} = e_n + \kappa\, \mu_n \sum_j |\theta_j|\, \mathrm{ARC}_j$. The sum of squared pricing errors is

$$\sum_{n=1}^{N} \big(e_n^{TC}\big)^2 = \|\mathbf{e}\|^2 + 2\kappa\, \mathbf{e}'\boldsymbol{\mu} \sum_j \mathrm{ARC}_j\, |\theta_j| + \kappa^2\, \|\boldsymbol{\mu}\|^2 \Big(\sum_j \mathrm{ARC}_j\, |\theta_j|\Big)^2,$$

with $\mathbf{e} = \boldsymbol{\mu} - \mathbf{S}'\boldsymbol{\theta}$. The last term is of second order in $\bar{c}$ and is dropped. The frictionless solution is $\boldsymbol{\theta}^F = (\mathbf{S}\mathbf{S}')^{-1}\mathbf{S}\boldsymbol{\mu}$, at which $\mathbf{S}\mathbf{e}^F = \mathbf{0}$ and therefore $\mathbf{e}^{F\prime}\boldsymbol{\mu} = \|\mathbf{e}^F\|^2$. Since $\mathbf{e}'\boldsymbol{\mu} - \|\mathbf{e}^F\|^2 = (\boldsymbol{\theta}^F - \boldsymbol{\theta})'\mathbf{S}\boldsymbol{\mu}$ is itself of order $\bar{c}$ at the cost-aware solution, replacing $\mathbf{e}'\boldsymbol{\mu}$ with $\|\mathbf{e}^F\|^2$ changes the objective only at second order. To first order in $\bar{c}$ the estimation problem is therefore

$$\min_{\boldsymbol{\theta}} \; \|\boldsymbol{\mu} - \mathbf{S}'\boldsymbol{\theta}\|^2 \;+\; 2\lambda \sum_{j=1}^{k} \mathrm{ARC}_j\, |\theta_j|, \qquad \lambda = \zeta N \bar{c}\, \|\mathbf{e}^F\|^2 > 0.$$

The pricing loss has Hessian $\mathbf{H} = 2\mathbf{S}\mathbf{S}'$, positive definite for $N \gg k$, and $\|\boldsymbol{\mu} - \mathbf{S}'\boldsymbol{\theta}\|^2 = \|\mathbf{e}^F\|^2 + (\boldsymbol{\theta} - \boldsymbol{\theta}^F)'\mathbf{S}\mathbf{S}'(\boldsymbol{\theta} - \boldsymbol{\theta}^F)$. Under separability $\mathbf{S}\mathbf{S}'$ is diagonal, and with $\pi_j \equiv H_{jj}/2$ the problem decomposes into

$$\min_{\theta_j} \; \pi_j\, (\theta_j - \theta_j^F)^2 + 2\lambda\, \mathrm{ARC}_j\, |\theta_j|, \qquad j = 1, \ldots, k,$$

whose solution is $\theta_j^{TC} = \mathrm{sign}(\theta_j^F)\big(|\theta_j^F| - \lambda\, \mathrm{ARC}_j/\pi_j\big)^+$, which is Equation (13). Characteristic j is down-weighted whenever $\lambda\, \mathrm{ARC}_j/\pi_j > 0$, and set to zero when this ratio exceeds $|\theta_j^F|$. Since π_j depends on the characteristic's role in the pricing loss while ARC_j depends on its implementation cost, two characteristics with identical θ_j^F receive different cost-aware loadings whenever their ARC values differ. □

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 out-

standing, 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). We compute excess returns by subtracting the one-month Treasury bill yield, obtained from Kenneth French's data library, from each stock's return.

Sample selection procedure. 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 B.1 reports the number of characteristic-month (p, T) pair observations for different starting dates, 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.

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 ($p = 129$) for a moderately long time series ($T = 624$ months). The right panel reports the cross-section of stocks available over the resulting sample period.

We incorporate a wide range of asset characteristics from Chen and Zimmermann (2021). We retain a predictor 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.1 summarizes the resulting characteristics.

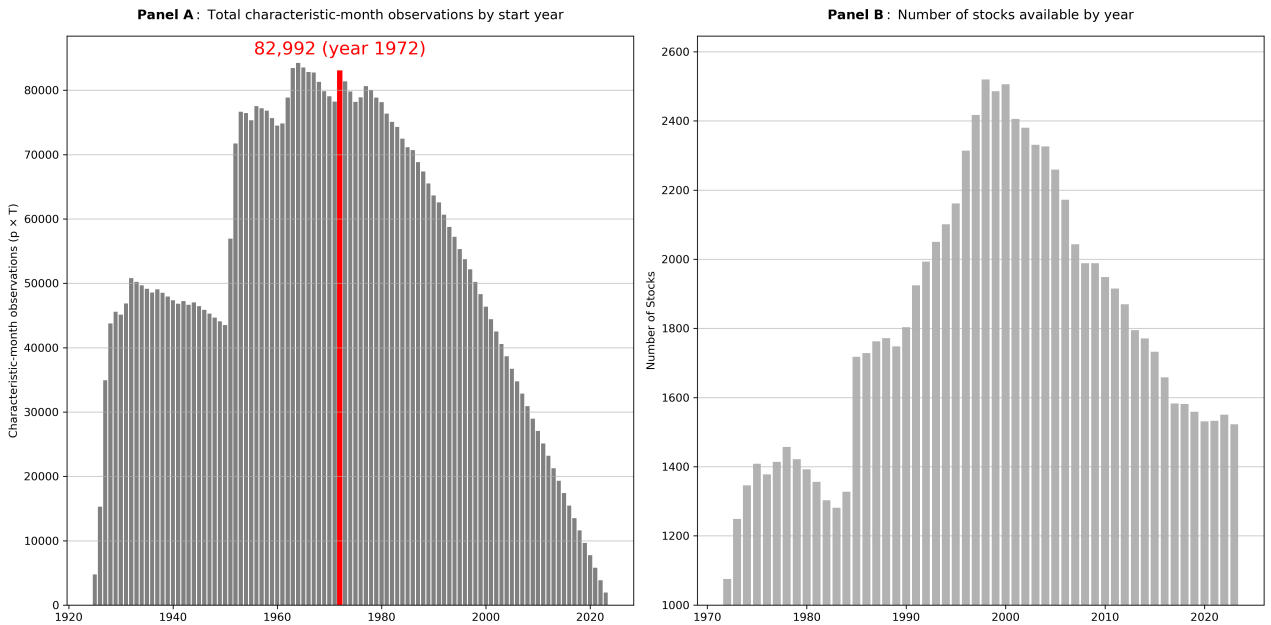


Figure B.1: **Sample selection**. The left panel reports the number of characteristic-month (p, T) pair observations for different starting dates, holding the end date fixed at November 2023. The right panel reports the cross-section of stocks available over the sample from January 1972 to November 2023.

Table B.1: **Summary of firm characteristics**. This table summarizes the firm-level characteristics $I_{i,t}$ used in the main empirical analysis as inputs to 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 $\times$ 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)

(continued)

(continued)

Acronym	Name	Definition	Reference
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)
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. We transform all series using the recommended procedures to ensure stationarity. We use these macroeconomic indicators in the main empirical analysis as input to the LSTM cells to extract latent factors, which we then include as input to the SDF and instrument networks (see Section 2.2 for a full description).

Table B.2: **Summary of macroeconomic indicators**: This table summarizes 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

(Continued on next page)

(Table continued from previous page)

Variable Name	Description	Group	Source
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
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

(Continued on next page)

(Table continued from previous page)

Variable Name	Description	Group	Source
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

Construction of expected daily dollar volume. The stock-level price-impact coefficient $\Lambda_{i,t}$ in Equation (6) requires an estimate of expected daily dollar volume $V_{i,t}$. We construct this measure from the CRSP monthly stock file (MSF) as follows. For each stock i in month m, we compute monthly dollar volume as

$$\mathrm{DVOL}^{\mathrm{m}}_{i,m} = P_{i,m} \cdot \big(100\,\mathrm{VOL}_{i,m}\big), \tag{B.1}$$

where $P_{i,m}$ is the month-end price and $\mathrm{VOL}_{i,m}$ is CRSP monthly trading volume reported in hundreds of shares. We then compute a trailing six-month rolling mean of monthly dollar volume, $\overline{\mathrm{DVOL}}^{\mathrm{m},6}_{i,m}$ (using all available history when fewer than six months exist). To map this monthly measure into expected daily dollar volume, we assume a constant number of trading days per month, $D = 21$, and set

$$V_{i,t} \approx \frac{\overline{\mathrm{DVOL}}^{\mathrm{m},6}_{i,m}}{D}, \tag{B.2}$$

where m is the month t in which the portfolio is rebalanced, so that $V_{i,t}$ uses the six months up to and including month t. Substituting into Equation (6) yields

$$\Lambda_{i,t} = \frac{0.2}{V_{i,t}} \approx \frac{0.2 \times 21}{\overline{\mathrm{DVOL}}^{\mathrm{m},6}_{i,m}}. \tag{B.3}$$

C Transaction Costs for Managed Portfolios

This appendix describes how we construct the 68 characteristic-managed portfolios used in the anomaly-level analysis of Section 4.2 and in the factor-model comparison of Appendix D.8, and how we compute their transaction costs following Barroso and Detzel (2021). We build one portfolio for each characteristic that supports a 2×3 double sort, excluding Size, the sorting variable itself, and four characteristics whose coverage is too thin to fill the six cells.

Portfolio construction. For each month t and characteristic j, we first transform the raw characteristic into cross-sectional percentile ranks, then linearly map these ranks to the interval $[-0.5, +0.5]$ to ensure comparability across characteristics with different scales and distribu-

tions. We then sort all stocks into two size groups based on NYSE median market capitalization, and within each size group further sort into three groups using the 30th and 70th percentiles of the characteristic. This yields six value-weighted portfolios: {SL, SM, SH, BL, BM, BH}, where S/B denotes small/big, and L/M/H denotes low/medium/high. The gross long-short factor return for characteristic j is:

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

where $r_{p,t+1}$ is the value-weighted return of portfolio p.

Turnover and transaction costs. To compute transaction costs, we account for the asymmetric nature of long-short portfolios by normalizing the long and short legs separately after return-driven weight changes. For each portfolio $p \in \{SL, SM, SH, BL, BM, BH\}$ and month t, we first compute the drift-adjusted weight of stock i before rebalancing:

$$\widehat{w}_{i,p,t-1} = \frac{w_{i,p,t-1}(1 + r_{i,t})}{\sum_{k=1}^{N_t} w_{k,p,t-1}(1 + r_{k,t})} \tag{C.2}$$

where $w_{i,p,t-1}$ is the weight after the previous rebalancing and $r_{i,t}$ is the intervening return. Portfolio turnover is then:

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

and the transaction cost for portfolio p is:

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

where $c_{i,t}$ is the one-way transaction cost (effective half bid-ask spread) of stock i at time t. Transaction costs are additive across the long and short legs, since investors incur costs regardless of trade direction. The long-short factor transaction cost is:

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

and the net-of-costs return is $r^{net}_{j,t} = r^{gross}_{j,t} - TC^{LS}_{j,t}$. Net and gross returns form the basis for the maximum squared Sharpe ratio analysis in Table D.6.

C.1 Frontier Expansion and Pricing Errors

Section 4.2 measures both the pricing error the SDF leaves on each anomaly and how much the anomaly expands the frontier attainable from the SDF. We set both out here, along with their relationship.

For each characteristic j, the investment opportunity set holds the SDF factor F_t and the long-short managed portfolio $C_{j,t}$, each entering as a long and a short position. Gross of costs the four payoffs are F_t, $-F_t$, $C_{j,t}$ and $-C_{j,t}$. Net of costs, every position pays the one-way cost of the leg it holds, so the long payoffs become $F_t - TC^F_t$ and $C_{j,t} - TC^C_{j,t}$ and the short payoffs $-F_t - TC^F_t$ and $-C_{j,t} - TC^C_{j,t}$. Following Novy-Marx and Velikov (2016, footnote 4) and Detzel et al. (2023), we compute $SR^2(F)$ and $SR^2(F, C_j)$ as the maximum squared Sharpe ratio over non-negative weights summing to one, and report the frontier expansion $\%\Delta SR^2 = SR^2(F, C_j)/SR^2(F) - 1$.

Gross of costs, this expansion rescales the pricing error reported in Section 4.2:

$$SR^2(F_t, C_{j,t}) - SR^2(F_t) = \frac{\alpha_j^2}{\mathrm{Var}(\varepsilon_{j,t})} \tag{C.6}$$

where α_j and $\varepsilon_{j,t}$ come from the time-series regression of $C_{j,t}$ on F_t. Non-negativity does not bind gross, because the long and short payoffs are exact mirror images and the constrained optimum coincides with the unconstrained one. Both sections therefore report the same information, just with different denominators. An anomaly with a large pricing error expands the frontier little when that error is hard to separate from residual noise. Net of costs, the relation fails. Charging each leg its own cost breaks the mirror symmetry, the constraint binds, and the two measures come apart, which is why both are reported.

All estimates use test-period data (January 1997 – November 2023). Pricing errors carry Newey-West standard errors with four lags. Test asset returns are gross throughout, so varia-

tion in α_j comes from switching the pricing kernel rather than from adjusting the test portfolios.

D Additional Empirical Results

This section reports the additional results referred to in the main text, in the order in which the main text refers to them. Appendix D.1 reports the beta-sorted portfolios, and Appendix D.2 the pricing performance under alternative estimates of the conditional SDF betas. Appendix D.3 extends the anomaly-level pricing analysis to the \$5m and \$500m price-impact constraints, Appendix D.4 tabulates the anomaly-level results of Section 4.2, and Appendix D.5 reports the cross-sectional regressions on exposure to the aggregate rebalancing cost whose ordering Section 4.2 uses. Appendix D.6 extends the characteristic reallocation to the \$5m and \$500m constraints. Appendix D.7 reports frontier expansion in levels and under equal weighting, and Appendix D.8 the spanning tests against the Fama–French factor models.

The remaining appendices support Sections 5 and 6. Appendix D.9 reports portfolio composition and characteristic attribution for the feedforward network, and Appendix D.10 for the linear SDF. Appendix D.11 documents the size-based linear cost specification of Brandt et al. (2009) and the square-root impact cost of Almgren et al. (2005). Appendix D.12 reports results for the later sample starting in 1986, and Appendix D.13 extends the market-regime analysis to the price-impact cost specifications. Finally, Appendix D.14 reports the results under alternative leverage constraints.

D.1 Beta-Sorted Portfolios

Table D.1 reports average excess returns and pricing errors for each beta-sorted decile under three specifications: the frictionless SDF, cost-aware bid-ask, and cost-aware price impact with \$50m capacity. All three produce monotonic return spreads with similar long-short returns: 27.8% annually for the frictionless SDF, 22.0% for cost-aware bid-ask, and 22.6% for cost-aware price impact. The CAPM alphas are uniformly large and significant (t-statistics between 7.40 and 8.42), and the rank ordering of decile returns across specifications is nearly identical

	Frictionless SDF					TC bid-ask SDF					TC price-impact SDF ($50m)				
		Market$-R_f$		FF5			Market$-R_f$		FF5			Market$-R_f$		FF5	
Decile	Avg	α	t-stat	α	t-stat	Avg	α	t-stat	α	t-stat	Avg	α	t-stat	α	t-stat
1	0.004	−0.028	−0.47	−0.006	−0.10	0.021	−0.006	−0.10	0.012	0.18	0.029	0.002	0.03	−0.059	−1.60
2	0.072	0.052	1.03	0.068	1.28	0.089	0.071	1.52	0.088	1.75	0.075	0.057	1.29	0.024	0.80
3	0.091	0.077	1.80	0.092	2.01	0.091	0.076	1.83	0.089	2.03	0.092	0.079	2.05	0.065	2.34
4	0.105	0.093	2.30	0.107	2.47	0.117	0.105	2.57	0.120	2.77	0.109	0.096	2.40	0.083	2.99
5	0.122	0.112	2.95	0.128	3.17	0.119	0.107	2.69	0.121	2.79	0.120	0.110	2.83	0.108	3.91
6	0.132	0.124	3.23	0.138	3.33	0.134	0.124	3.12	0.138	3.23	0.132	0.121	2.97	0.127	4.42
7	0.141	0.132	3.30	0.145	3.39	0.149	0.139	3.45	0.151	3.49	0.148	0.137	3.28	0.143	4.92
8	0.163	0.152	3.62	0.168	3.67	0.142	0.130	3.02	0.145	3.21	0.159	0.147	3.33	0.167	5.39
9	0.188	0.177	3.85	0.189	3.87	0.197	0.184	4.05	0.203	4.26	0.181	0.168	3.46	0.209	6.22
10	0.281	0.264	4.78	0.290	5.02	0.240	0.226	4.25	0.252	4.49	0.255	0.240	4.30	0.291	7.56
10−1	0.278	0.291	8.42	0.297	8.32	0.220	0.232	7.69	0.241	7.39	0.226	0.238	7.40	0.350	14.09
		GRS	p-value	GRS	p-value		GRS	p-value	GRS	p-value		GRS	p-value	GRS	p-value
		10.26	0.000	10.63	0.000		8.92	0.000	9.15	0.000		7.61	0.000	31.15	0.000

Table D.1: **Pricing errors of beta-sorted portfolios.** This table reports annualized average excess returns and pricing errors (alphas) for decile portfolios sorted on estimated SDF betas $\widehat{\beta}_{i,t}$. Decile 1 (10) contains stocks with the lowest (highest) estimated betas. For each specification, we report alphas and t-statistics from time-series regressions on the CAPM (Market$-R_f$) and the Fama and French (2015) five-factor model (FF5). The row "10−1" reports the long-short portfolio. The GRS statistic (Gibbons et al., 1989) tests the joint null that all decile alphas are zero. The out-of-sample period is January 1997 to November 2023.

(correlation ≈ 0.98).

This similarity is informative. Switching from the frictionless to a cost-aware SDF changes the tangency portfolio and, therefore, the conditional betas $\widehat{\beta}_{i,t}$ for every stock: the decile compositions differ across specifications, yet the resulting risk-return gradient is preserved. Large EV and XS-R^2 improvements documented in Table 1 therefore do not arise from a fundamentally different ranking of which stocks are risky. Instead, the pricing gains operate within, rather than across, risk tiers. Transaction costs constrain the pricing kernel without sacrificing the information needed for coarse risk differentiation.

Within this broadly similar ordering, one economically meaningful difference emerges at the extremes. The long-short spread narrows from 27.8% (frictionless) to 22.0–22.6% (cost-aware), with the adjustment concentrated at the bottom. Decile 1 returns increase by 1.7–2.5 percentage points, from a near-zero 0.4% annually under the frictionless SDF to 2.1–2.9% under cost-awareness. This pattern is consistent with the frictionless SDF, which relies on noisy covariance estimates for small, illiquid firms (Avramov et al., 2023). Constraining the tangency portfolio stabilizes these estimates and yields a more plausible risk-return gradient at the low end of the distribution.

The GRS test statistics (Gibbons et al., 1989) reject the null of zero joint pricing errors for all three specifications under both the CAPM (10.26, 8.92, and 7.61) and the Fama and French (2015) five-factor model (10.63, 9.15, and 31.15), confirming that every SDF, frictionless or cost-aware, captures a risk-return trade-off that standard linear factors do not span.[26] Frictionless and cost-aware bid-ask SDFs produce comparable FF5 alpha gradients (10−1 FF5 alpha of 29.7% and 24.1%), while the price-impact SDF generates a steeper gradient (35.0%), indicating that the risk dimension it captures is less spanned by standard factors.

Table D.2 extends the beta-sorted portfolio analysis of Table D.1 to the \$5m and \$500m price-impact capacity constraints. The frictionless SDF columns are identical across all specifications since the frictionless model does not depend on capacity.

[26]The GRS statistic for the price-impact specification under FF5 is notably large. This reflects the combination of a steep FF5 alpha gradient (10−1 of 35.0%, $t = 14.09$) and tight within-decile residual variance, both of which inflate the quadratic form $\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}$.

Under the $5m constraint, where price-impact costs barely lower the net Sharpe ratio of the frictionless portfolio, the cost-aware SDF produces a risk-return gradient that closely resembles the frictionless baseline. Here the long-short spread narrows modestly from 27.8% to 23.1%, and the CAPM GRS statistic remains highly significant (8.21 vs. 11.10). Both the rank ordering of decile returns and the magnitudes of the FF5 alphas are largely preserved, although explained variation is 59% higher than under the frictionless SDF at this size (Table 1).

Under the $500m constraint, the pattern changes qualitatively. The long-short spread compresses to 9.3%, less than one-third of the frictionless benchmark, as the severe price-impact costs restrict the tangency portfolio and narrow the cross-sectional range of estimated betas. Decile 1 returns rise from 0.8% to 7.8% annually, while Decile 10 returns fall from 28.6% to 17.1%. The resulting risk-return relationship remains monotonic but substantially flatter. The CAPM GRS declines to 1.94 ($p = 0.040$) and the FF5 GRS to 1.76 ($p = 0.068$). Because both statistics are quadratic forms in the estimated alphas, compressing the long-short spread to less than a third of its frictionless value reduces the dispersion the test can detect. This decline therefore follows from the narrower spread rather than from the cost-aware SDF becoming spanned by the standard factors.

D.2 Sensitivity to the Beta Estimation Approach

Table D.3 reports out-of-sample EV and XS-R^2 under four combinations of SDF and beta network specifications: nonlinear SDF with nonlinear betas (the baseline), nonlinear SDF with linear betas, linear SDF with nonlinear betas, and linear SDF with linear betas. We describe the nonlinear beta network in Appendix A.3. Its linear variant removes all hidden layers and nonlinear activations, mapping characteristics directly to conditional betas via a single linear transformation.

Two results emerge from this comparison. Cost-awareness improves cross-sectional pricing across all combinations, and the improvement from the frictionless to the cost-aware SDF holds whether the SDF is linear or nonlinear and whether betas are estimated with a flexible or a restricted network. Monotonic scaling with cost severity also holds in all four cases.

	Frictionless SDF					TC price-impact SDF ($5m)					TC price-impact SDF ($500m)				
		Market$-R_f$		FF5			Market$-R_f$		FF5			Market$-R_f$		FF5	
Decile	Avg	α	t-stat	α	t-stat	Avg	α	t-stat	α	t-stat	Avg	α	t-stat	α	t-stat
1	0.008	−0.024	−0.41	−0.001	−0.02	0.019	−0.007	−0.12	0.012	0.20	0.078	0.063	1.80	0.072	1.90
2	0.068	0.047	0.94	0.063	1.19	0.085	0.067	1.46	0.082	1.66	0.101	0.090	2.52	0.099	2.55
3	0.089	0.076	1.80	0.090	2.04	0.094	0.082	1.99	0.094	2.17	0.107	0.093	2.45	0.103	2.52
4	0.111	0.101	2.51	0.116	2.65	0.106	0.093	2.37	0.106	2.51	0.116	0.100	2.50	0.114	2.59
5	0.114	0.105	2.74	0.119	2.91	0.120	0.109	2.72	0.124	2.87	0.133	0.119	2.76	0.133	2.88
6	0.134	0.124	3.18	0.139	3.32	0.135	0.123	3.03	0.138	3.15	0.133	0.119	2.63	0.136	2.86
7	0.148	0.138	3.55	0.150	3.56	0.149	0.139	3.36	0.150	3.48	0.155	0.140	3.00	0.152	3.11
8	0.156	0.146	3.48	0.162	3.58	0.156	0.144	3.35	0.160	3.44	0.146	0.128	2.53	0.151	2.90
9	0.186	0.174	3.71	0.190	3.75	0.185	0.171	3.71	0.193	3.97	0.160	0.144	2.60	0.167	2.91
10	0.286	0.269	4.95	0.292	5.10	0.249	0.235	4.40	0.260	4.62	0.171	0.158	2.49	0.191	2.93
10−1	0.278	0.293	8.36	0.293	8.12	0.231	0.242	7.63	0.247	7.36	0.093	0.095	2.09	0.119	2.75
		GRS	p-value	GRS	p-value		GRS	p-value	GRS	p-value		GRS	p-value	GRS	p-value
		11.10	0.000	10.73	0.000		8.21	0.000	8.22	0.000		1.94	0.040	1.76	0.068

Table D.2: **Pricing errors of beta-sorted portfolios: alternative capacity constraints.** This table reports annualized average excess returns and pricing errors (alphas) for decile portfolios sorted on estimated SDF betas $\widehat{\beta}_{i,t}$. The frictionless SDF column repeats the baseline from Table D.1 since the frictionless model does not depend on capacity. For each specification, we report alphas and t-statistics from time-series regressions on the CAPM (Market$-R_f$) and the Fama and French (2015) five-factor model (FF5). The row "10−1" reports the long-short portfolio. The GRS statistic (Gibbons et al., 1989) tests the joint null that all decile alphas are zero. The out-of-sample period is January 1997 to November 2023.

	Frictionless		Bid-Ask		$5m		$50m		$500m	
	EV	XS-R^2	EV	XS-R^2	EV	XS-R^2	EV	XS-R^2	EV	XS-R^2
Panel A: Nonlinear Beta Network										
Nonlinear SDF (Baseline)	0.063	0.145	0.098	0.223	0.100	0.222	0.114	0.234	0.143	0.258
Linear SDF	0.053	0.160	0.078	0.198	0.076	0.193	0.127	0.243	0.143	0.257
Panel B: Linear Beta Network										
Nonlinear SDF	0.070	0.164	0.088	0.182	0.095	0.209	0.111	0.230	0.144	0.254
Linear SDF	0.061	0.167	0.077	0.182	0.078	0.189	0.120	0.235	0.142	0.250

Table D.3: **Sensitivity to the beta estimation network.** This table reports out-of-sample EV and XS-R^2 under four combinations of SDF and beta network specifications. Panel A uses the nonlinear beta network described in Appendix A.3; Panel B replaces it with a linear variant that removes all hidden layers and nonlinear activations. The first row of Panel A is the baseline from Table 1. All results use the GAN architecture under the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

Second, the choice of beta network affects absolute levels but not the qualitative pattern. Comparing Cases 1 and 4 (nonlinear SDF with nonlinear vs. linear betas), the frictionless EV shifts from 0.063 to 0.070 and XS-R^2 from 0.145 to 0.164, while the cost-aware bid-ask EV shifts from 0.098 to 0.088. This gap narrows modestly but remains economically meaningful across all specifications. Comparing Cases 2 and 3 (linear SDF with nonlinear vs. linear betas) yields a similar conclusion: absolute metrics change, but the cost-awareness improvement is robust to the beta estimation method.

Net Sharpe ratios are computed from portfolio returns and require no beta estimate, so they are identical across the two panels. Nonlinear and linear SDFs deliver 0.064 and 0.028 in the frictionless models. Under bid-ask cost-awareness, the figures are 0.113 and 0.084, and at $500m capacity, 0.139 and 0.088. Only the pricing metrics respond to the beta network.

Across the ten model-by-cost cells, replacing the nonlinear beta network with the linear one changes explained variation by at most 0.010 and cross-sectional R^2 by at most 0.041. Cost-awareness changes them by 0.015 to 0.113.

One ordering does depend on the beta network. Under the nonlinear specification, the linear SDF attains a higher frictionless cross-sectional R^2 than the nonlinear SDF (0.160 against 0.145), and the gap narrows to 0.003 once betas are linear (0.167 against 0.164). It reflects the interaction between the two networks rather than a property of the SDF.

D.3 Anomaly Pricing Under Alternative Capacity

Figure D.1 repeats the anomaly-level comparison of Figure 2 at \$5m and \$500m. At \$5m, the cost-aware SDF prices 42 of the 68 portfolios more accurately, with a mean improvement of 4.3 bp per month. The largest improvements are again in idiosyncratic volatility (`IdioVolAHT`, 35 bp, and `IdioVol3F`, 26 bp), maximum daily return (25 bp), and the zero-trade measures, and the largest deteriorations are in cash-flow-to-price (-13 bp) and short-term reversal (-7 bp).

At \$500m, the cost-aware SDF prices 29 of the 68 portfolios more accurately, and both the improvements and the deteriorations are larger. `IdioVolAHT` improves by 117 bp per month, the one-month zero-trade measure by 92 bp and `Price` by 87 bp, while the momentum portfolios (`MomOffSeason`, `Mom6m`, `Mom12mOffSeason` and `ResidualMomentum`) are priced worse by 39 to 43 bp and cash-flow-to-price by 37 bp. The mean improvement is 4.7 bp, close to the 4.3 bp at \$5m and below the 11.6 bp at \$50m (Section 4.2).

D.4 Anomaly-Level Results

Table D.4 reports, for each of the 68 characteristic-managed portfolios of Section 4.2, the mean gross return and trading cost of the long-short position, whether the anomaly survives its own trading costs, the pricing errors under the frictionless kernel and under the cost-aware kernel for bid-ask costs and price impact at \$50m, the pricing improvements, and the loading β_j^{TC} on the aggregate rebalancing cost. Every anomaly-level number in Section 4.2 is drawn from this table.

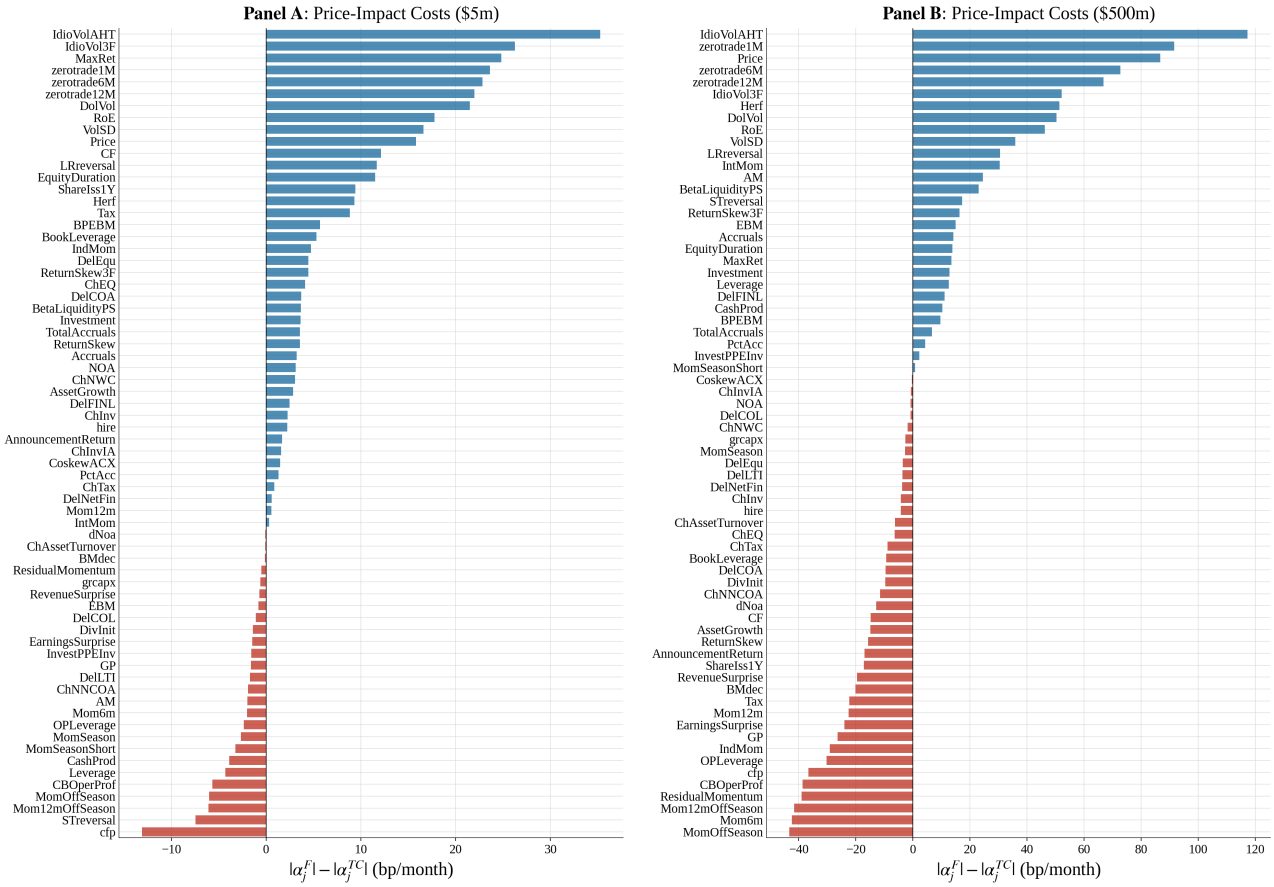


Figure D.1: **Anomaly pricing implications: alternative capacity constraints.** Each bar reports $\Delta|\alpha_j| = |\alpha_j^F| - |\alpha_j^{TC}|$ (bp/month) for one of the 68 characteristic-managed portfolios, sorted from largest improvement to largest deterioration. Blue bars indicate anomalies that the cost-aware SDF prices better. Red bars indicate anomalies priced worse. Panel A: price-impact costs ($5m). Panel B: price-impact costs ($500m). The out-of-sample period is January 1997 to November 2023.

Table D.4: **Anomaly-level results.** One row per characteristic-managed long-short portfolio (2×3 sorts on size and the characteristic, value-weighted, 68 portfolios), test period January 1997 to November 2023. Gross is the mean monthly gross return and Cost the mean monthly trading cost of the long-short position, both in basis points; an anomaly survives its own trading costs when Gross exceeds Cost (Detzel et al., 2023). α is the intercept, in basis points per month, from a time-series regression of the gross long-short return on the tangency portfolio return of the frictionless kernel (M_{t+1}) and of the cost-aware kernel under bid-ask costs and under price impact at \$50m, with t-statistics in parentheses. $\Delta|\alpha| = |\alpha^F| - |\alpha^{TC}|$ is positive when the cost-aware kernel prices the anomaly more accurately. β^{TC} is the loading of the gross long-short return on the aggregate rebalancing cost $\mathcal{TC}_t$ of the bid-ask cost-aware kernel. Rows are sorted by $\Delta|\alpha|$ under bid-ask costs. Across the 68 portfolios, β^{TC} correlates with $\Delta|\alpha|$ at -0.47 under bid-ask costs and at -0.38, -0.49 and -0.20 at the \$5m, \$50m and \$500m capacities, each loading computed with the $\mathcal{TC}_t$ series of its own specification. 11 of the 68 anomalies survive their own trading costs.

				M_{t+1}		M^{TC}_{t+1}, bid-ask		M^{TC}_{t+1}, \$50m		$\Delta\|\alpha\|$		
Anomaly	Gross	Cost	Survives	α	(t)	α	(t)	α	(t)	bid-ask	\$50m	β^{TC}
IdioVolAHT	-28	30	No	-163	(-2.96)	-112	(-1.98)	-67	(-1.26)	51.1	95.7	-0.049
IdioVol3F	-0	112	No	-107	(-2.61)	-67	(-1.60)	-33	(-0.83)	39.8	73.8	-0.032
MaxRet	14	121	No	-79	(-2.34)	-42	(-1.19)	-12	(-0.35)	37.4	67.6	0.010
zerotrade12M	-11	24	No	-111	(-2.89)	-76	(-1.87)	-41	(-1.05)	34.8	69.7	-0.043
zerotrade6M	-15	29	No	-111	(-3.10)	-77	(-2.03)	-43	(-1.17)	34.5	68.6	-0.034
zerotrade1M	-24	64	No	-121	(-3.27)	-87	(-2.33)	-53	(-1.53)	34.4	67.7	-0.045
DolVol	19	87	No	-110	(-1.48)	-79	(-1.10)	-33	(-0.50)	31.2	77.6	-0.013
Price	13	35	No	122	(2.46)	92	(1.81)	58	(1.15)	29.7	64.0	0.086
VolSD	11	22	No	-83	(-2.52)	-55	(-1.61)	-25	(-0.77)	27.7	57.8	-0.052
RoE	-9	22	No	-66	(-2.43)	-43	(-1.58)	-23	(-0.88)	22.4	42.9	-0.011
EquityDuration	10	22	No	-39	(-1.84)	-23	(-1.16)	-7	(-0.37)	16.0	31.4	0.018
ShareIss1Y	17	29	No	-25	(-1.25)	-10	(-0.51)	4	(0.21)	14.8	20.8	0.022
Mom12m	37	60	No	-38	(-0.90)	-23	(-0.61)	-2	(-0.05)	14.7	36.2	-0.055
Herf	14	15	No	54	(2.48)	41	(2.17)	27	(1.52)	13.1	27.3	0.023
CF	13	30	No	-14	(-0.49)	3	(0.11)	14	(0.56)	11.7	0.6	0.052
LRreversal	50	45	Yes	70	(2.22)	59	(1.97)	52	(1.88)	11.6	18.2	0.068
Mom6m	43	80	No	-20	(-0.53)	-10	(-0.30)	7	(0.23)	10.3	13.4	-0.033
Tax	15	21	No	-10	(-0.59)	2	(0.11)	12	(0.74)	8.3	-1.6	0.009
BPEBM	4	32	No	15	(0.61)	7	(0.34)	4	(0.20)	7.9	11.2	-0.019
BetaLiquidityPS	12	31	No	24	(1.44)	17	(1.01)	13	(0.83)	6.8	10.2	0.016
Mom12mOffSeason	45	64	No	-16	(-0.37)	-9	(-0.26)	7	(0.22)	6.5	8.4	-0.059
IntMom	-15	74	No	-40	(-1.33)	-34	(-1.17)	-25	(-0.88)	5.8	14.6	-0.043
ReturnSkew3F	-8	142	No	-21	(-2.17)	-15	(-1.61)	-12	(-1.30)	5.7	9.4	0.000
Accruals	25	24	Yes	33	(2.23)	28	(2.12)	25	(2.05)	5.1	7.8	0.034
hire	5	24	No	-10	(-0.53)	-5	(-0.31)	0	(0.00)	4.7	10.1	0.004

Table D.4, continued

Anomaly	Gross	Cost	Survives	α	(t)	α	(t)	α	(t)	bid-ask	\$50m	β^{TC}
TotalAccruals	11	25	No	15	(1.18)	11	(0.96)	10	(0.88)	4.0	5.5	0.009
ChNWC	-2	27	No	4	(0.33)	-0	(-0.03)	-1	(-0.11)	3.6	2.9	0.007
DelFINL	1	24	No	-14	(-1.22)	-11	(-0.93)	-5	(-0.42)	3.6	9.5	0.010
BookLeverage	-5	17	No	6	(0.24)	-3	(-0.11)	-6	(-0.30)	3.5	-0.1	-0.002
DelCOA	19	25	No	14	(0.76)	11	(0.60)	14	(0.86)	3.5	0.4	0.032
NOA	31	20	Yes	32	(2.34)	28	(2.33)	29	(2.58)	3.4	2.6	0.012
ChInvIA	3	29	No	-6	(-0.61)	-3	(-0.28)	0	(0.02)	3.3	5.8	0.016
Investment	12	25	No	20	(1.43)	17	(1.38)	14	(1.23)	3.1	5.6	0.019
DelEqu	21	24	No	19	(0.89)	16	(0.78)	16	(0.83)	3.0	3.5	0.028
ChEQ	22	24	No	18	(0.88)	15	(0.78)	15	(0.88)	2.5	2.3	0.027
PctAcc	30	22	Yes	29	(2.28)	27	(2.25)	27	(2.37)	2.4	2.3	0.033
dNoa	16	25	No	-7	(-0.50)	-5	(-0.38)	3	(0.21)	1.9	4.0	0.029
EBM	16	34	No	25	(1.38)	23	(1.33)	21	(1.20)	1.7	4.0	0.040
grcapx	5	24	No	-4	(-0.22)	-2	(-0.14)	1	(0.05)	1.5	3.0	0.023
AM	29	28	Yes	45	(1.07)	43	(1.14)	38	(1.04)	1.3	6.6	0.063
AssetGrowth	21	24	No	10	(0.51)	9	(0.47)	12	(0.71)	1.3	-2.1	0.030
ChInv	27	24	Yes	26	(1.69)	25	(1.72)	26	(1.99)	1.1	0.1	0.024
ReturnSkew	11	143	No	-4	(-0.44)	3	(0.36)	8	(0.96)	0.8	-4.3	0.010
DelNetFin	6	25	No	-1	(-0.06)	0	(0.02)	3	(0.29)	0.5	-2.5	0.004
AnnouncementReturn	36	64	No	20	(1.51)	20	(1.62)	24	(2.05)	0.4	-3.7	0.004
InvestPPEInv	5	23	No	-11	(-0.75)	-10	(-0.77)	-5	(-0.39)	0.4	5.7	0.013
CoskewACX	17	53	No	11	(0.50)	10	(0.52)	10	(0.57)	0.3	0.0	0.023
DelCOL	2	26	No	-7	(-0.40)	-7	(-0.44)	-4	(-0.28)	-0.3	2.7	0.012
IndMom	49	72	No	21	(0.67)	21	(0.90)	30	(1.44)	-0.5	-9.5	0.027
DelLTI	5	18	No	5	(0.60)	7	(0.86)	8	(1.00)	-1.8	-2.5	0.005
ChTax	34	45	No	25	(1.70)	27	(2.07)	30	(2.49)	-2.1	-5.3	0.002
ChAssetTurnover	9	24	No	3	(0.25)	5	(0.49)	6	(0.62)	-2.2	-3.2	0.014
Leverage	23	24	No	32	(0.83)	35	(0.98)	32	(0.94)	-2.5	0.2	0.054
CashProd	8	27	No	15	(0.50)	18	(0.64)	16	(0.58)	-2.8	-0.8	0.043
GP	38	17	Yes	19	(1.01)	22	(1.13)	28	(1.45)	-2.8	-9.0	0.002
MomSeasonShort	9	155	No	9	(0.43)	11	(0.60)	12	(0.63)	-2.9	-3.3	0.004
RevenueSurprise	24	39	No	10	(0.72)	13	(0.93)	18	(1.38)	-3.0	-8.5	-0.011
ChNNCOA	13	26	No	3	(0.31)	7	(0.65)	11	(1.04)	-3.6	-7.2	0.002
BMdec	28	21	Yes	9	(0.36)	13	(0.52)	18	(0.74)	-3.8	-8.2	0.033
MomSeason	25	156	No	19	(1.05)	23	(1.25)	23	(1.30)	-3.9	-3.9	0.051
OPLeverage	36	17	Yes	8	(0.55)	12	(0.91)	21	(1.59)	-4.0	-12.4	0.015

Table D.4, continued

Anomaly	Gross	Cost	Survives	α	(t)	α	(t)	α	(t)	bid-ask	\$50m	β^{TC}
EarningsSurprise	25	46	No	7	(0.49)	13	(0.87)	19	(1.35)	-5.2	-11.1	0.012
DivInit	4	35	No	-1	(-0.03)	6	(0.23)	3	(0.12)	-5.3	-2.2	-0.023
CBOperProf	32	21	Yes	5	(0.24)	12	(0.52)	22	(1.02)	-6.5	-17.0	0.014
ResidualMomentum	42	55	No	10	(0.47)	17	(0.85)	25	(1.36)	-6.6	-15.1	0.007
STreversal	43	157	No	48	(1.98)	56	(2.22)	53	(2.15)	-7.3	-4.5	0.068
MomOffSeason	30	61	No	3	(0.13)	16	(0.65)	23	(0.98)	-12.4	-19.4	0.059
cfp	34	30	Yes	9	(0.28)	21	(0.82)	32	(1.22)	-12.7	-22.9	0.074

	λ_{TC}	t	t (Shanken)	t (FM)	XS-R^2
Bid-ask spreads	0.022	3.58	2.76	0.85	0.163
Price impact \$5m	0.002	2.31	1.27	0.38	0.075
Price impact \$50m	0.001	2.22	0.92	0.26	0.070
Price impact \$500m	0.001	3.38	2.70	1.30	0.148

Table D.5: **Pricing of transaction-cost exposure.** This table reports the price of risk on exposure to the aggregate rebalancing cost $\mathcal{TC}_t$ across the 68 characteristic-managed long-short portfolios, constructed as in Section 4.2. For each portfolio, β_j^{TC} is the loading of the gross long-short return on $\mathcal{TC}_t$. λ_{TC} is the slope from a cross-sectional regression of average excess returns on β_j^{TC}, expressed in monthly units of the corresponding cost series and therefore not comparable across rows. The t-statistic is from that univariate cross-sectional regression. t (Shanken) is the errors-in-variables corrected statistic from the two-factor specification that also includes the frictionless SDF beta, and t (FM) is the Fama-MacBeth statistic from monthly cross-sections. XS-R^2 is from the univariate cross-sectional regression. Each row uses the $\mathcal{TC}_t$ series of the corresponding cost specification. The test period is January 1997 to November 2023.

D.5 Pricing of Transaction-Cost Exposure

Table D.5 reports cross-sectional regressions of average excess returns on the loadings β_j^{TC} used in Section 4.2, for each cost specification. Across all four, the full-sample slope is positive, with univariate t-statistics between 2.22 and 3.58, and the loadings alone account for between 7% and 16% of the cross-sectional variation in average returns. Adding the frictionless SDF beta as a control and applying the Shanken correction leaves the statistic above two under bid-ask costs and at \$500m (2.76 and 2.70) and below two at the two intermediate capacities (1.27 and 0.92). Monthly Fama-MacBeth statistics never exceed 1.30. The premium is therefore one that full-sample average returns identify but that monthly cross-sections estimate too imprecisely to detect, and none of these estimates is a conditional risk price, since $\mathcal{TC}_t$ enters no conditional beta. AR(1) innovations to $\mathcal{TC}_t$ behave the same way, with full-sample cross-sectional t-statistics between 0.17 and 2.98 and Fama-MacBeth statistics between -0.23 and 1.02, so they do not identify a separate priced dimension.

D.6 Characteristic Reallocation Under Alternative Capacity

Figure D.2 reports the portfolio-level consequences of cost-awareness under the \$5m and \$500m price-impact capacity constraints, extending the baseline results in Figure 3. Reallocation

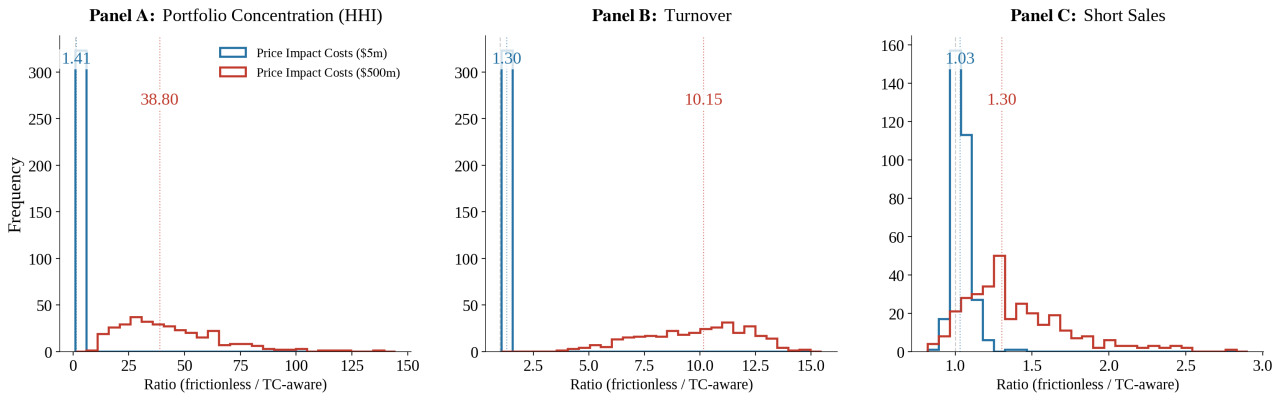


Figure D.2: **Tangency portfolio composition under alternative capacity constraints.** This figure reports the monthly ratio of three portfolio statistics between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) for the GAN architecture under price-impact costs. Panel A: Herfindahl-Hirschman Index (HHI) of portfolio concentration. Panel B: aggregate portfolio turnover. Panel C: fraction of short positions. Ratios above one indicate that the frictionless portfolio exhibits higher values. Blue histograms: \$5m capacity. Red outlines: \$500m capacity. Dashed vertical lines indicate median ratios. The baseline leverage constraint is $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

scales monotonically with cost severity. HHI ratios increase from 1.41 (\$5m) to 3.19 (\$50m, Figure 3) to 38.80 (\$500m); turnover ratios follow the same pattern: 1.30, 2.28, and 10.15. Under the \$500m constraint, the frictionless portfolio is nearly 40 times more concentrated and trades more than 10 times as much as its cost-aware counterpart, consistent with the deeply negative net Sharpe ratio (-1.355) documented in Table 2. Short-selling intensity increases modestly under \$500m (median ratio of 1.30), indicating that at extreme capacity levels cost-awareness also adjusts the long-short composition.

Figure D.3 extends the characteristic reallocation analysis of Figure 4 to the \$5m and \$500m price-impact capacity constraints. Under the \$5m constraint (Panel A), the positive correlation between exposure shifts ($\Delta\widehat{x}^j$) and SHAP attribution shifts (ΔSHAP^j) is $\rho = 0.63$, comparable to the baseline bid-ask result. The reallocation magnitudes are compressed relative to the \$50m baseline, consistent with the negligible economic impact of costs at this capacity level: most characteristics cluster near the origin, and only short-term reversal (`STreversal`) shows a substantial shift. The same high-turnover characteristics identified in the main text—announcement return, return skewness, maximum daily return—appear in the upper-right region, while stable fundamentals such as book leverage and tax remain near the origin.

Under the \$500m constraint (Panel B), the shifts in both exposure and attribution are

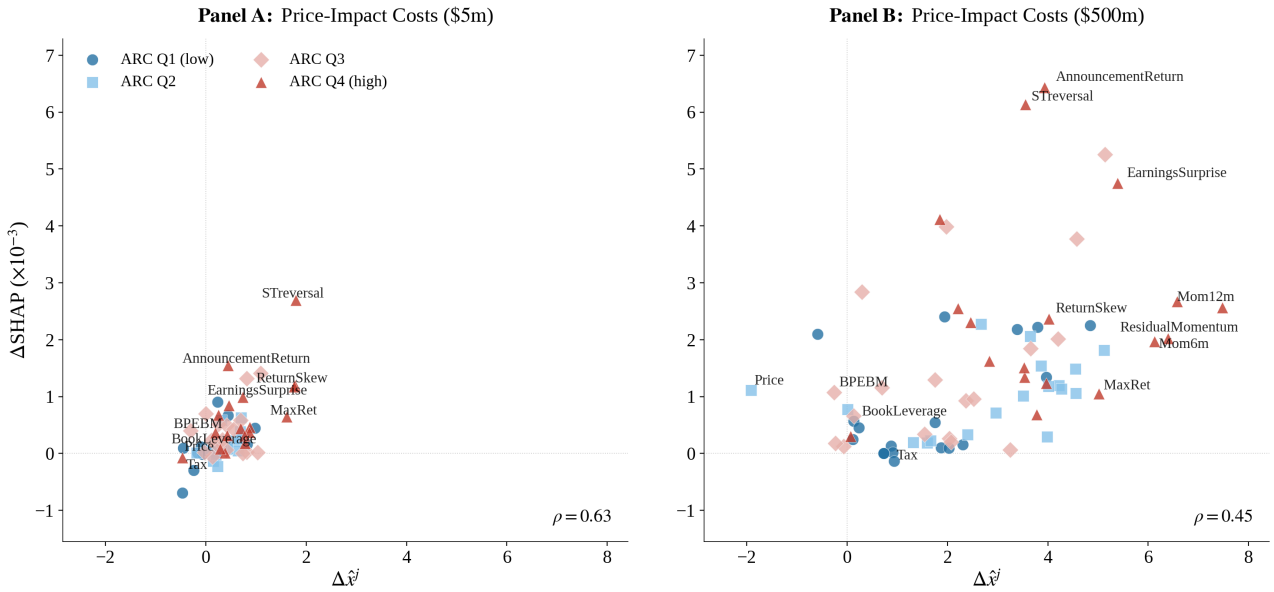


Figure D.3: **Characteristic reallocation under alternative capacity constraints.** Each point represents one of the 73 firm characteristics. The x-axis plots the change in tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs (positive values indicate higher frictionless exposure). The y-axis plots the corresponding change in DeepSHAP attribution ΔSHAP^j ($\times 10^{-3}$). Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). Panel A: price-impact costs ($5m). Panel B: price-impact costs ($500m). The baseline leverage constraint is $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

larger than at $50m. The largest exposure shifts are in the momentum signals (`Mom12m`, `Mom12mOffSeason`, `ResidualMomentum` and `Mom6m`), followed by earnings surprise, and the correlation between exposure and attribution shifts is $\rho = 0.45$.

D.7 Frontier Expansion: Levels and Equal Weighting

Figure 6 reports frontier expansion in percentage terms. Those levels are the maximum squared Sharpe ratio attainable from each SDF alone. Gross of costs, these run from 0.121 for the frictionless kernel to 0.101 under bid-ask costs, 0.106 at $5m, 0.059 at $50m, and 0.025 at $500m. Net of costs, the ordering reverses from 0.004 frictionless to 0.013, 0.097, 0.051, and 0.019, the aggregate result of Table D.6 seen one anomaly at a time.

Repeating the exercise on equal-weighted portfolios leaves the net panel qualitatively unchanged. Anomalies expand the frictionless frontier by 7.6 percent on average, compared with 0.6 to 2.9 percent for the cost-aware specifications, close to the 6.9 and 0.4 to 3.2 percent obtained under value weighting. Gross of costs, which is not the comparison the paper makes,

the average expansion under equal weighting reaches 23.8 percent at \$50m and 86.0 percent at \$500m against 19.2 percent for the frictionless kernel. Both figures follow from the small gross SR^2 of the constrained kernels, which inflates percentage expansion measured against a low base, rather than from larger expansions in levels.

D.8 Spanning Against Observable Factors

A tangency portfolio with a positive net Sharpe ratio may still add little to what an investor can obtain from observable factors. Following Detzel et al. (2023), we form opportunity sets from long and short positions in the factors of the Fama–French three-factor model (FF3; Fama and French, 1993), the five-factor model (FF5; Fama and French, 2015) and the five-factor model augmented with momentum (FF6; Jegadeesh and Titman, 1993; Carhart, 1997). We add each tangency portfolio to each factor set and compute the maximum squared Sharpe ratio subject to non-negative weights summing to one, with each leg charged its own cost.[27] Table D.6 reports the ex-ante SR^2, with weights estimated in the validation sample (January 1992–December 1996) and applied in the test period, and the ex-post SR^2, re-optimized over the test period.[28]

With FF5, the ex-ante gross SR^2 rises from 0.089 to 0.226 when the frictionless tangency portfolio is added and to 0.161 and 0.175 with the cost-aware portfolios under bid-ask costs and at \$5m. Net of costs, the frictionless portfolio raises it from 0.027 to 0.035. The table evaluates the frictionless portfolio net of bid-ask costs, so its like-for-like comparison is with the bid-ask cost-aware portfolio, which reaches 0.037; under bid-ask costs, the two kernels add about the same to the net frontier of observable factors.

Under price impact, the cost-aware portfolios raise the net SR^2 to 0.128 at \$5m and 0.057 at \$50m and add nothing at \$500m, where it equals that of FF5 (0.027). The comparison with

[27] 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: $f^{\text{net}}_{kt} = f_{kt} - TC^f_{kt}$ for long positions and $f^{S,\text{net}}_{kt} = -f_{kt} - TC^f_{kt}$ for short positions, where TC^f_{kt} represents the one-way transaction cost (Barroso and Detzel, 2021). Each model is then evaluated via $SR^2(f^{\text{net}}) = \max_{w\in\mathbb{R}^{2K}} \left[w'\mu/\sqrt{w'\Sigma w}\right]^2$ subject to $w_k \geq 0$ and $\sum_k w_k = 1$, where $\mu = E(f^{\text{net}})$ and $\Sigma = \text{cov}(f^{\text{net}})$.

[28] In-sample SR^2 (training period, January 1972–December 1991) is substantially higher for all SDF-augmented models, ranging from 0.44 to 3.19, for both frictionless and cost-aware specifications.

	Value-weighted portfolios				Equal-weighted portfolios			
	Gross		Net		Gross		Net	
Model	Ex-ante	Ex-post	Ex-ante	Ex-post	Ex-ante	Ex-post	Ex-ante	Ex-post
Panel A: Fama–French three-factor model								
FF3	0.031	0.032	0.022	0.025	0.046	0.049	0.021	0.025
FF3 + M_{t+1}	0.193	0.212	0.033	0.037	0.212	0.246	0.033	0.038
FF3 + M^{TC}_{t+1} (bid-ask)	0.129	0.152	0.038	0.043	0.148	0.099	0.037	0.045
FF3 + M^{TC}_{t+1} ($5m)	0.143	0.166	0.135	0.147	0.163	0.200	0.136	0.156
FF3 + M^{TC}_{t+1} ($50m)	0.073	0.089	0.061	0.074	0.092	0.115	0.059	0.077
FF3 + M^{TC}_{t+1} ($500m)	0.028	0.035	0.022	0.026	0.049	0.054	0.021	0.026
Panel B: Fama–French five-factor model								
FF5	0.089	0.127	0.027	0.047	0.191	0.216	0.030	0.041
FF5 + M_{t+1}	0.226	0.255	0.035	0.054	0.312	0.344	0.040	0.050
FF5 + M^{TC}_{t+1} (bid-ask)	0.161	0.199	0.037	0.057	0.253	0.297	0.043	0.055
FF5 + M^{TC}_{t+1} ($5m)	0.175	0.212	0.128	0.153	0.262	0.306	0.140	0.159
FF5 + M^{TC}_{t+1} ($50m)	0.109	0.154	0.057	0.084	0.166	0.248	0.066	0.085
FF5 + M^{TC}_{t+1} ($500m)	0.068	0.127	0.027	0.048	0.057	0.221	0.030	0.042
Panel C: Fama–French five-factor model + momentum								
FF6	0.094	0.132	0.027	0.047	0.187	0.220	0.030	0.041
FF6 + M_{t+1}	0.174	0.256	0.035	0.054	0.298	0.346	0.040	0.050
FF6 + M^{TC}_{t+1} (bid-ask)	0.134	0.199	0.037	0.057	0.119	0.297	0.043	0.055
FF6 + M^{TC}_{t+1} ($5m)	0.145	0.213	0.128	0.153	0.251	0.307	0.140	0.159
FF6 + M^{TC}_{t+1} ($50m)	0.095	0.155	0.058	0.084	0.188	0.248	0.066	0.085
FF6 + M^{TC}_{t+1} ($500m)	0.071	0.132	0.027	0.048	0.154	0.224	0.030	0.042

Table D.6: **Mean-variance frontier expansion: SDF portfolios vs. observable risk factors.** This table reports the maximum out-of-sample squared Sharpe ratio (SR^2) attainable from different investment opportunity sets, each consisting of long and short positions in the specified factors. "Gross" evaluates raw returns; "Net" deducts transaction costs for each factor's long and short legs (Appendix C). Ex-ante SR^2 estimates weights in the validation sample (January 1992–December 1996) and evaluates on the test period (January 1997–November 2023). Ex-post SR^2 re-optimizes over the test period. FF3 denotes Fama and French (1993); FF5 denotes Fama and French (2015); FF6 adds cross-sectional momentum. M_{t+1} is the frictionless specification, evaluated net of bid-ask costs, and M^{TC}_{t+1} denotes cost-aware SDFs under bid-ask spread costs or price-impact costs at the indicated capacity, each evaluated net of its own costs.

the frictionless portfolio at these capacities is the one in Panel B of Table 2, which evaluates both portfolios net of the same costs. The results are similar with FF3 and FF6 and with equal-weighted factors.[29]

Net of the baseline bid-ask costs, therefore, neither tangency portfolio adds much to the frontier an investor can reach with the Fama–French factors alone, and under price impact the

[29]Adding momentum to FF5 contributes nothing on a net basis: the FF6 net ex-ante SR^2 (0.027) is identical to FF5, indicating that momentum's trading costs fully consume its gross contribution (Korajczyk and Sadka, 2004; Lesmond et al., 2004).

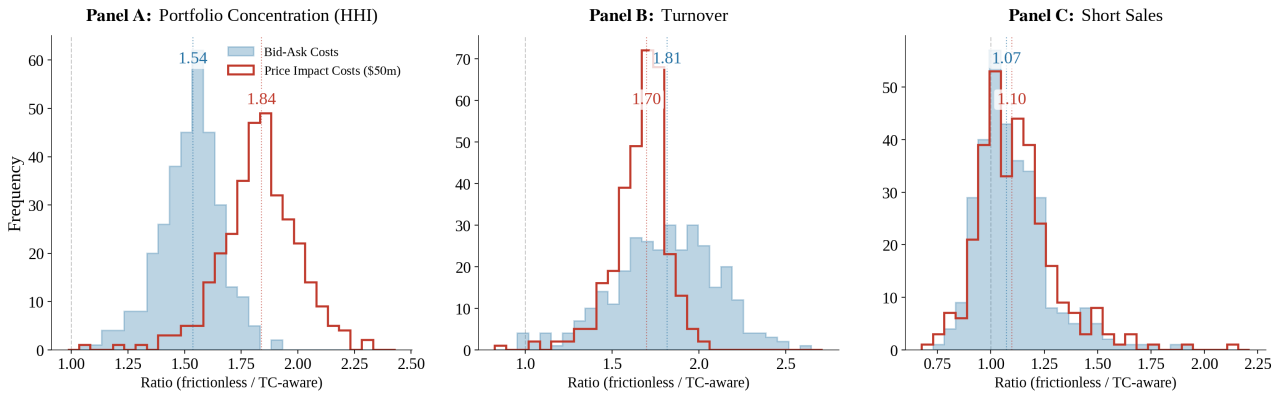


Figure D.4: **FFN portfolio composition: frictionless vs. cost-aware.** This figure reports the monthly ratio of three portfolio statistics between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) estimated with the feedforward neural network (FFN). Panel A: Herfindahl-Hirschman Index (HHI) of portfolio concentration. Panel B: aggregate portfolio turnover. Panel C: fraction of short positions. Ratios above one indicate that the frictionless portfolio exhibits higher values. Blue histograms correspond to bid-ask spread costs. Red outlines correspond to price-impact costs ($50m). Dashed vertical lines indicate median ratios. The baseline leverage constraint is $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

addition from the cost-aware portfolio is positive at $5m and $50m and zero at $500m. What a tangency portfolio is worth to such an investor depends on the costs that investor faces, which is the point of Section 4.4.

D.9 FFN Portfolio Composition

Figure D.4 reports the monthly ratio of portfolio statistics between the frictionless and cost-aware FFN tangency portfolios. These results mirror the GAN evidence in Figure 3. Panel A shows that frictionless FFN portfolios are substantially more concentrated than their cost-aware counterparts, with median HHI ratios of 1.54 (bid-ask) and 1.84 (price-impact). Panel B confirms that cost-awareness reduces turnover: the frictionless portfolio trades 81% more under bid-ask costs and 70% more under price-impact costs. Panel C shows that short-selling intensity is largely unaffected, with median ratios near unity (1.07 and 1.10), indicating that cost-awareness operates primarily through diversification and reduced turnover rather than through long-short composition. Broader dispersion under price-impact costs in Panels A and B is consistent with the quadratic cost structure penalizing large trades more severely, inducing greater portfolio adjustments.

Figure D.5 complements the portfolio-level evidence with characteristic-level attribution.

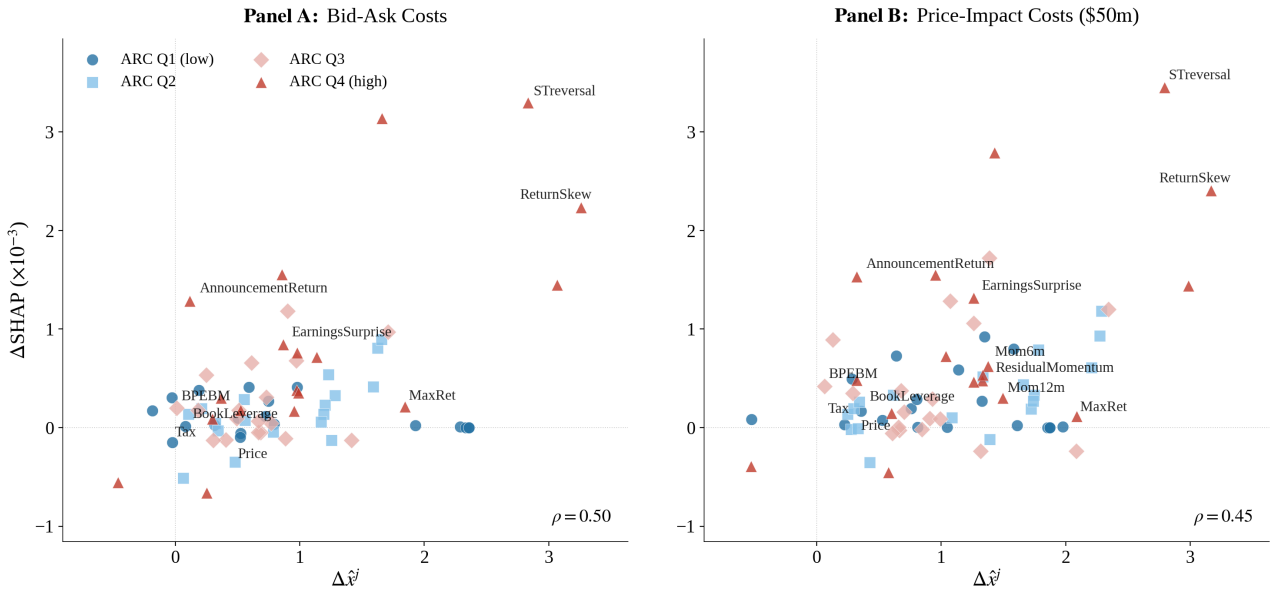


Figure D.5: **FFN characteristic reallocation under cost-awareness.** Each point represents one of the 73 firm characteristics. The x-axis plots the change in tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs (positive values indicate higher frictionless exposure). The y-axis plots the corresponding change in DeepSHAP attribution ΔSHAP^j ($\times 10^{-3}$). Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). Panel A: bid-ask spread costs. Panel B: price-impact costs ($50m). The baseline leverage constraint is $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

The positive correlation between $\Delta\widehat{x}^j$ and ΔSHAP^j is preserved under the FFN ($\rho = 0.50$ under bid-ask costs, $\rho = 0.45$ under price-impact), confirming that portfolio exposure shifts reflect active repricing rather than mechanical weight adjustments. Under the FFN, we find in the upper-right quadrant the same high-turnover characteristics that dominate it in the GAN, namely short-term reversal, return skewness, maximum daily return, and announcement return, while stable fundamentals such as book leverage, tax, and price cluster near the origin. Correlations are somewhat lower than in the GAN ($\rho = 0.63$ and 0.47), consistent with the FFN's unconditional objective of providing less discipline in isolating the most important test-asset portfolios, although the qualitative pattern is unchanged.

D.10 Linear SDF Specification

Figure D.6 compares portfolio composition under the linear specification for the frictionless and cost-aware SDFs. Here the reallocation operates primarily through turnover reduction: the frictionless linear portfolio trades 29% more than its cost-aware counterpart (median turnover

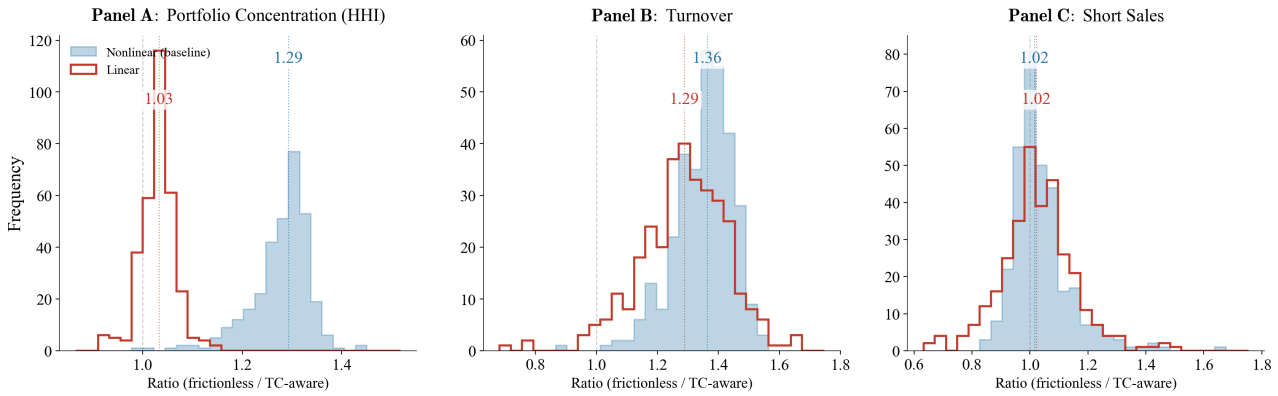


Figure D.6: **Tangency portfolio composition: nonlinear vs. linear specification.** This figure reports the monthly ratio of three portfolio statistics between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) for the GAN architecture under bid-ask costs. Panel A: Herfindahl-Hirschman Index (HHI) of portfolio concentration. Panel B: aggregate portfolio turnover. Panel C: fraction of short positions. Ratios above one indicate that the frictionless portfolio exhibits higher values. Blue histograms correspond to the nonlinear baseline specification. Red outlines correspond to the linear specification. Dashed vertical lines indicate median ratios. The baseline leverage constraint is $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

ratio of 1.29), comparable to the nonlinear baseline (1.36). By contrast, the concentration channel is largely absent, with an HHI ratio of just 1.03 (vs. 1.29 for the nonlinear model). This asymmetry reflects the linear model's inability to exploit characteristic interactions. Without nonlinear activations, the network cannot concentrate weights on specific subsets of the cross-section, limiting cost-awareness to disciplining aggregate turnover rather than reshaping the portfolio's characteristic composition.

At the characteristic level, Figure D.7 plots the exposure shift $\Delta\widehat{x}^j$ under the linear specification against the same quantity under the nonlinear baseline. This correlation is only $\rho = 0.27$, far below the cross-cost-specification correlation of 0.77 documented for the Brandt costs (Appendix D.11). The two models agree on the most prominent high-turnover characteristics—short-term reversal and return skewness remain in the upper-right quadrant—but diverge across most of the cross-section, with most points lying below the 45-degree line. This indicates that the linear model produces uniformly smaller reallocation magnitudes and targets a narrower set of characteristics, consistent with its restricted functional form.

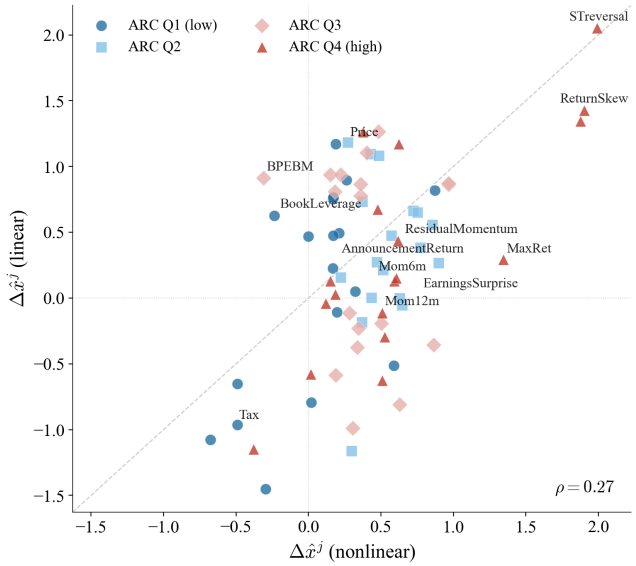


Figure D.7: **Characteristic reallocation: nonlinear vs. linear specification.** Each point represents one of the 73 firm characteristics. The x-axis plots the change in median tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs under the nonlinear (baseline) specification. The y-axis plots the same quantity under the linear specification. Positive values indicate higher frictionless exposure. Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). The dashed line is the 45-degree reference. All results use the GAN architecture under bid-ask costs and the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

D.11 Alternative Cost Specifications

Size-based transaction costs. Figure D.8 reports the cross-sectional distribution of the Brandt et al. (2009) size-based linear costs over the full sample. This specification generates one-way costs ranging from 0.60% for the smallest firms to 0.35% for the largest, with minimal time variation — substantially lower than the baseline bid-ask spread estimates documented in Figure 1.

Figure D.9 confirms that the portfolio reallocation mechanism operates under these milder costs. The frictionless portfolio is 26% more concentrated (median HHI ratio of 1.26) and trades 23% more (turnover ratio of 1.23) than the cost-aware counterpart, while short-selling intensity is essentially unchanged (ratio of 1.01). Magnitudes are smaller than under the baseline bid-ask specification (Figure 3), consistent with the lower cost levels, but the direction is preserved across all three channels.

At the characteristic level, Figure D.10 plots the exposure shift $\Delta\widehat{x}^j$ under size-based linear

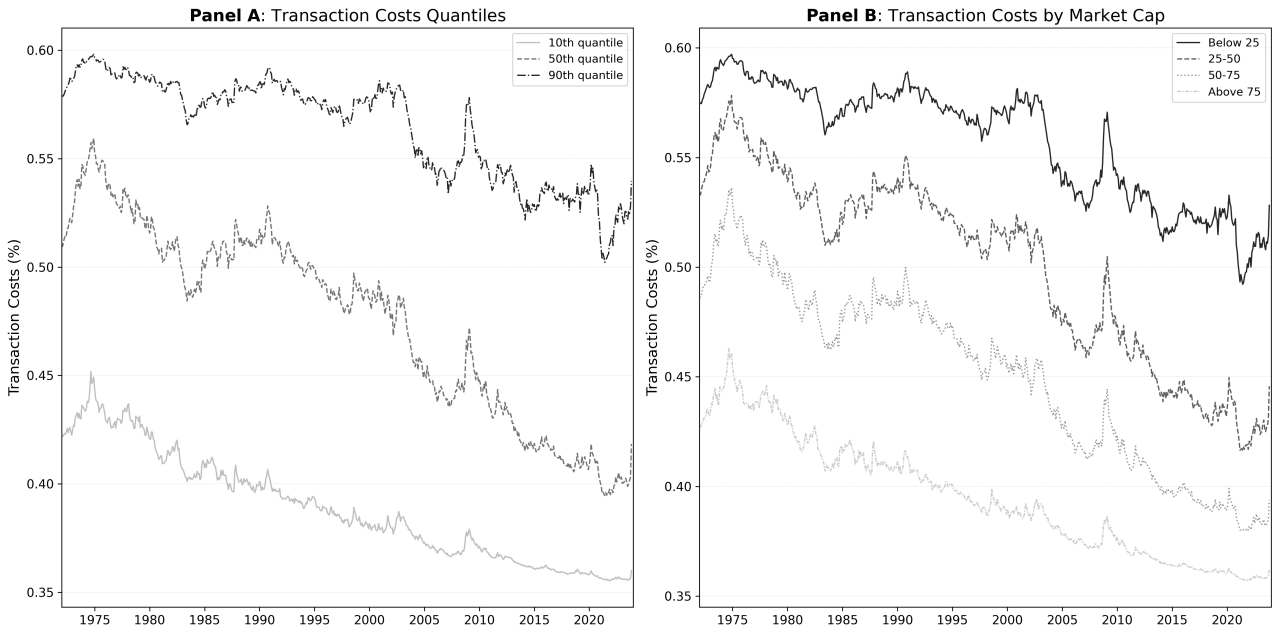


Figure D.8: **Sample variation of transaction costs as in Brandt et al. (2009).** The left panel reports the 10th, 50th, and 90th percentiles of the cross-sectional distribution of the transaction costs (in %). The right panel reports their average (in %) for stocks sorted by market capitalization. The sample runs from January 1972 to November 2023.

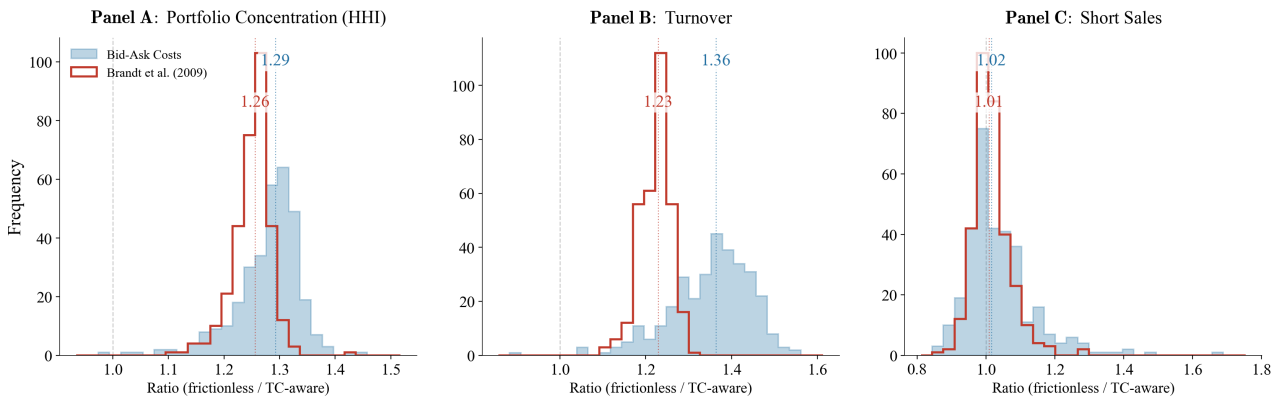


Figure D.9: **Tangency portfolio composition: bid-ask vs. size-based linear costs.** This figure reports the monthly ratio of three portfolio statistics between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) for the GAN architecture. Panel A: Herfindahl-Hirschman Index (HHI) of portfolio concentration. Panel B: aggregate portfolio turnover. Panel C: fraction of short positions. Ratios above one indicate that the frictionless portfolio exhibits higher values. Blue histograms correspond to bid-ask spread costs. Red outlines correspond to the Brandt et al. (2009) size-based linear costs. Dashed vertical lines indicate median ratios. The baseline leverage constraint is $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

costs against the same quantity under bid-ask costs. A strong positive correlation ($\rho = 0.77$) confirms that cost-awareness reallocates away from the same high-turnover characteristics under both cost specifications. Points sit below the 45-degree line, indicating that the reallocation is uniformly smaller under the milder Brandt costs but preserves the same structure.

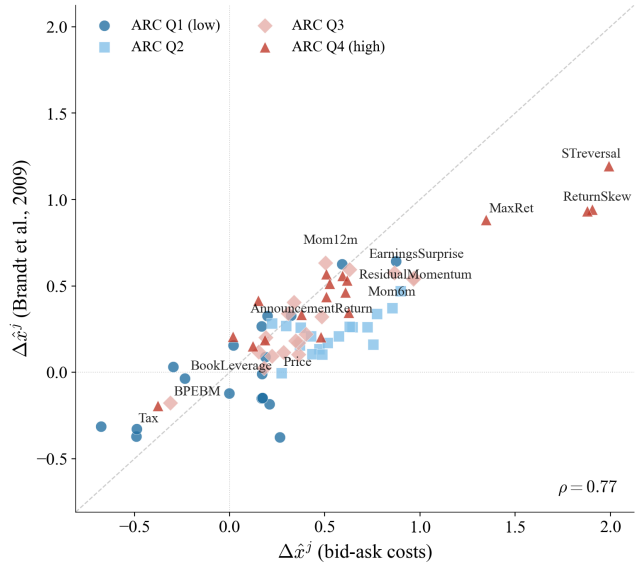


Figure D.10: **Characteristic reallocation: bid-ask vs. Brandt et al. (2009) costs.** Each point represents one of the 73 firm characteristics. The x-axis plots the change in median tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs under bid-ask costs; the y-axis plots the same quantity under the Brandt et al. (2009) cost specification. Positive values indicate higher frictionless exposure. Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). The dashed line is the 45-degree reference. All results use the GAN architecture under the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

Square-root price impact. The square-root specification in Equation (15) is calibrated to share the same anchor as the quadratic baseline of Jensen et al. (2026): a trade of 1% of average daily dollar volume produces 10 basis points of per-share impact. Under the square-root law, per-share impact at trade size $Q = 0.01\, V_{i,t}$ equals $\kappa\, \sigma_{i,t} \sqrt{Q/V_{i,t}} = 0.1\, \kappa\, \sigma_{i,t}$. Setting this equal to 0.001 at the cross-sectional mean volatility $\bar{\sigma}$ yields the anchor $\kappa\, \bar{\sigma} = 0.01$, with $\bar{\sigma}$ computed over the training period. We then model heterogeneity in per-stock implementation cost as a function of daily dollar volume ($V_{i,t}$) and return volatility ($\sigma_{i,t}$), measured as the 12-month rolling standard deviation of the firm's excess return.

Section 6.1 explains why we evaluate the square-root specification at larger sizes than the quadratic one.

Figures D.11 and D.12 report the portfolio composition ratios under the square-root specification at \$500m and \$5bn capacity. Reallocation intensifies sharply with capacity: at \$500m, the frictionless portfolio is 76% more concentrated and trades 94% more than the cost-aware counterpart; at \$5bn, the gaps widen to factors of 3.4 and 3.1, respectively. Short-sale ratios

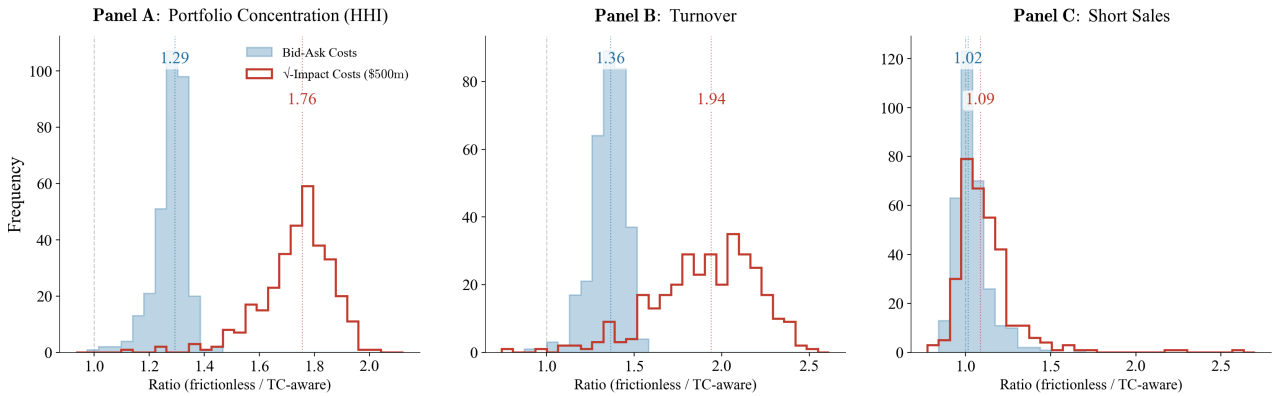


Figure D.11: **Tangency portfolio composition: bid-ask vs. square-root price impact at $500m.** This figure reports the monthly ratio of three portfolio statistics between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) for the GAN architecture. Panel A: Herfindahl-Hirschman Index (HHI). Panel B: aggregate portfolio turnover. Panel C: fraction of short positions. Ratios above one indicate that the frictionless portfolio exhibits higher values. Blue histograms correspond to bid-ask spread costs; red outlines correspond to the square-root price-impact specification at $500m capacity. Dashed vertical lines indicate median ratios. The baseline leverage constraint is $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

remain near unity (1.09 and 1.11) at both capacity levels, because shorts are costless in the cost specification. Figures D.13 and D.14 report the characteristic-level reallocation alongside the bid-ask baseline. The same high-turnover characteristics dominate the upper-right quadrant in all panels — short-term reversal, return skewness, maximum daily return, and (at $5bn) the momentum signals — while stable fundamentals such as book leverage, tax, and price remain near the origin. Exposure shifts and SHAP attribution shifts remain positively correlated at both capacities, though more weakly than under the bid-ask baseline ($\rho = 0.63$), because the binding cost constraint produces a more diffuse reallocation. The upper-right cluster of high-turnover signals is preserved across all specifications.

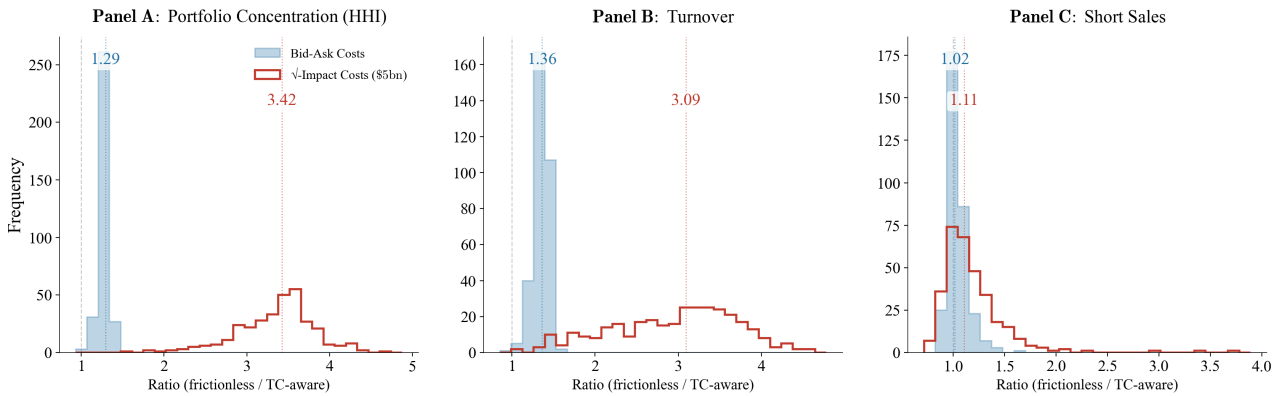


Figure D.12: **Tangency portfolio composition: bid-ask vs. square-root price impact at $5bn.** This figure reports the monthly ratio of three portfolio statistics between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) for the GAN architecture. Panel A: Herfindahl-Hirschman Index (HHI). Panel B: aggregate portfolio turnover. Panel C: fraction of short positions. Ratios above one indicate that the frictionless portfolio exhibits higher values. Blue histograms show bid-ask spread costs; red outlines show the square-root price-impact specification at $5bn capacity. Dashed vertical lines indicate median ratios. The baseline leverage constraint is $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

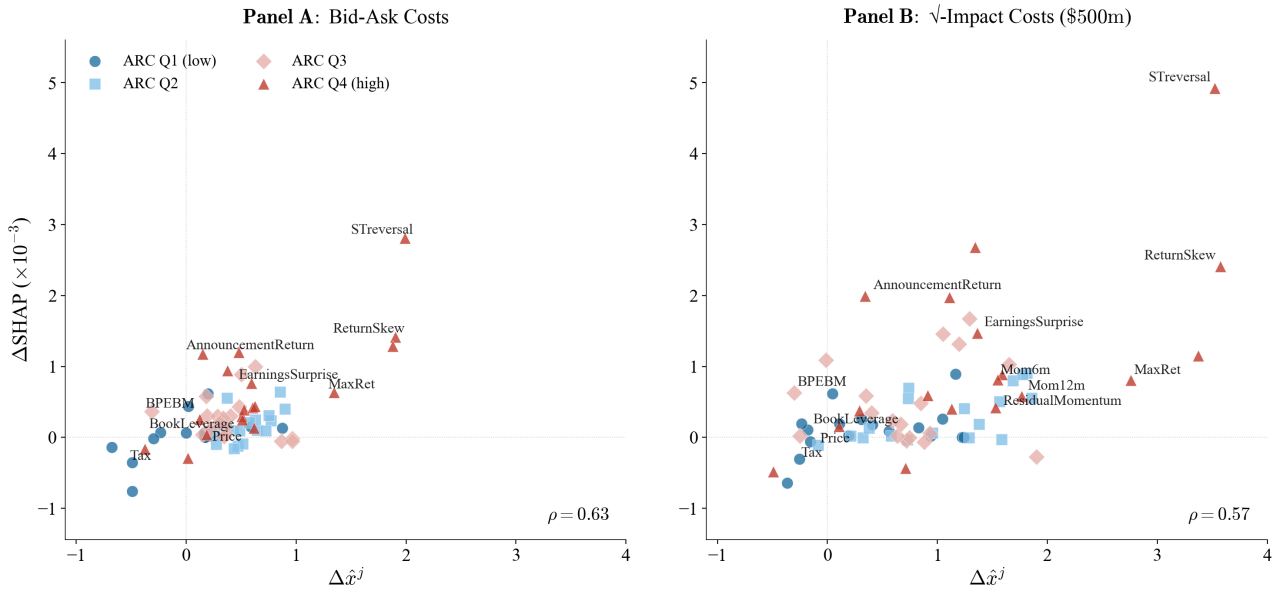


Figure D.13: **Characteristic-level reallocation: bid-ask vs. square-root price impact at $500m.** Each point represents one of the 73 firm characteristics. The x-axis plots the change in median tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs; the y-axis plots the change in SHAP attribution ΔSHAP^j ($\times 10^{-3}$). Panel A reports the scatter under bid-ask spread costs; Panel B reports the scatter under the square-root price-impact specification at $500m capacity. The Pearson correlation ρ is reported in the lower-right corner of each panel. Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). All results use the GAN architecture under the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

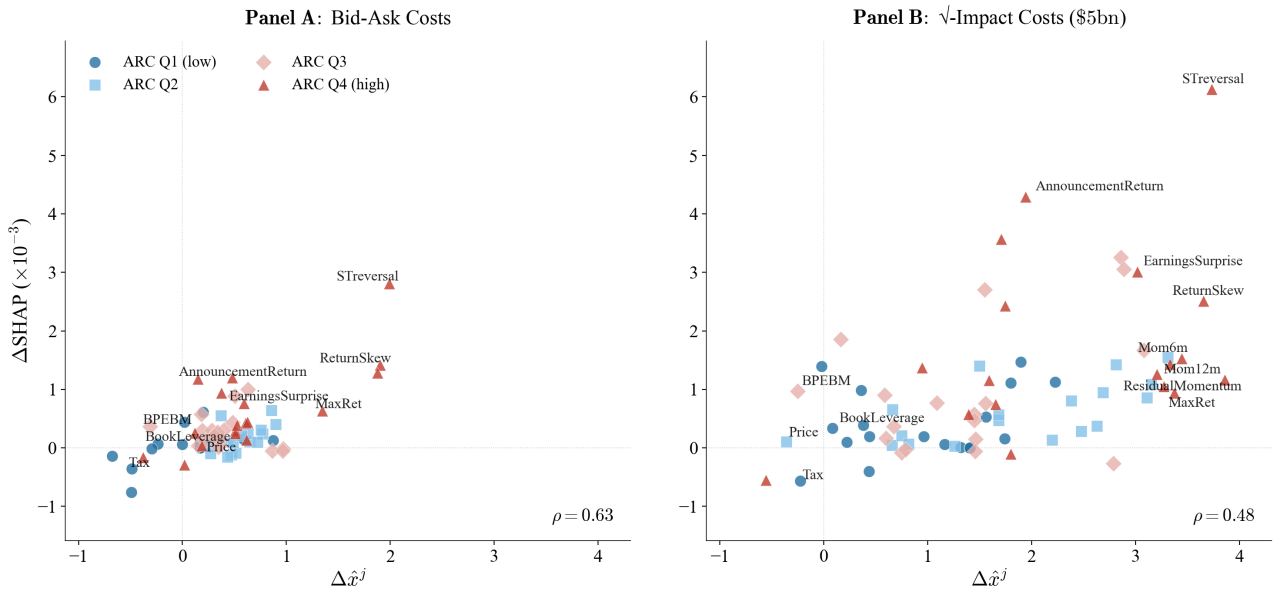


Figure D.14: **Characteristic-level reallocation: bid-ask vs. square-root price impact at $5bn.** Each point represents one of the 73 firm characteristics. The x-axis plots the change in median tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs; the y-axis plots the change in SHAP attribution ΔSHAP^j ($\times 10^{-3}$). Panel A reports the scatter under bid-ask spread costs; Panel B reports the scatter under the square-root price-impact specification at $5bn capacity. The Pearson correlation ρ is reported in the lower-right corner of each panel. Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). All results use the GAN architecture under the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. The out-of-sample period is January 1997 to November 2023.

	EV	XS-R^2	SR (Gross)	SR (Net)
Baseline (1972–2023), test: 1997–2023				
M_{t+1} (frictionless)	0.063	0.145	0.348	0.064
M_{t+1}^{TC} (bid-ask)	0.098	0.223	0.317	0.113
Later sample (1986–2023), test: 2011–2023				
M_{t+1} (frictionless)	0.073	0.160	0.407	0.167
M_{t+1}^{TC} (bid-ask)	0.059	0.117	0.386	0.236
M_{t+1}^{TC} ($5m)	0.068	0.124	0.387	0.350
M_{t+1}^{TC} ($50m)	0.061	0.141	0.334	0.251
M_{t+1}^{TC} ($500m)	0.137	0.216	0.245	0.182

Table D.7: **Later sample period.** This table reports out-of-sample performance metrics for the GAN architecture under each transaction cost specification. EV is explained variation. XS-R^2 is cross-sectional R^2. SR (Gross) and SR (Net) are monthly Sharpe ratios before and after transaction costs. The leverage constraint is $\omega_{it} \in (-5\%, +5\%)$.

D.12 Alternative Samples

Table D.7 reports out-of-sample performance for the later sample starting in January 1986 (38 years, 88 characteristics, test period January 2011 – November 2023). Figure D.15 reports portfolio composition ratios under both alternative samples alongside the baseline, and Figure D.16 compares the characteristic-level reallocation. Section 6.2 discusses the results. To increase readability, the portfolio composition ratios and the characteristic-level reallocation pertain only to bid-ask transaction costs relative to the frictionless SDF.

D.13 Market Conditions Under Price-Impact Costs

Table D.8 extends the market-regime analysis of Table 7 to the three price-impact specifications. Across all panels the frictionless model is the same. Columns differ only in the cost specification used to compute net Sharpe ratios.

Both main findings from the bid-ask analysis carry over to every price-impact specification. First, cost-awareness improves EV and XS-R^2 uniformly across all regime–cost combinations. Second, the magnitude of the pricing improvement scales with cost severity: the $500m constraint produces the largest EV and XS-R^2 gains, followed by $50m and then $5m. This ordering is consistent with the mechanism in Section 4.3: more binding capacity constraints drive

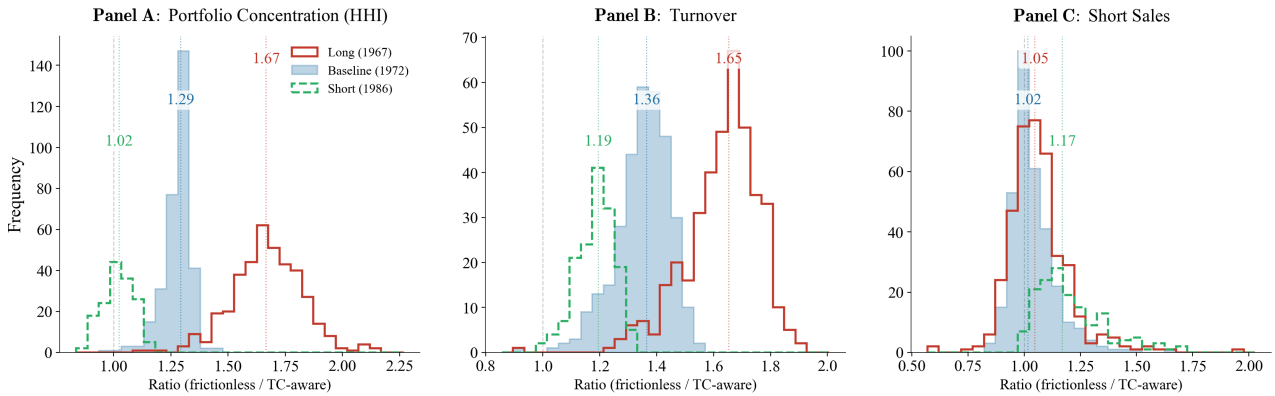


Figure D.15: **Tangency portfolio composition for alternative samples.** This figure reports the monthly ratio of three portfolio statistics between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) for the GAN architecture under bid-ask costs. Panel A: Herfindahl-Hirschman Index (HHI) of portfolio concentration. Panel B: aggregate portfolio turnover. Panel C: fraction of short positions. Ratios above one indicate that the frictionless portfolio exhibits higher values. Red outlines: long sample (1967–2023, test: 1991–2023); blue histograms: baseline (1972–2023, test: 1997–2023); green dashed: short sample (1986–2023, test: 2011–2023). Dashed vertical lines indicate median ratios. All results use the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$.

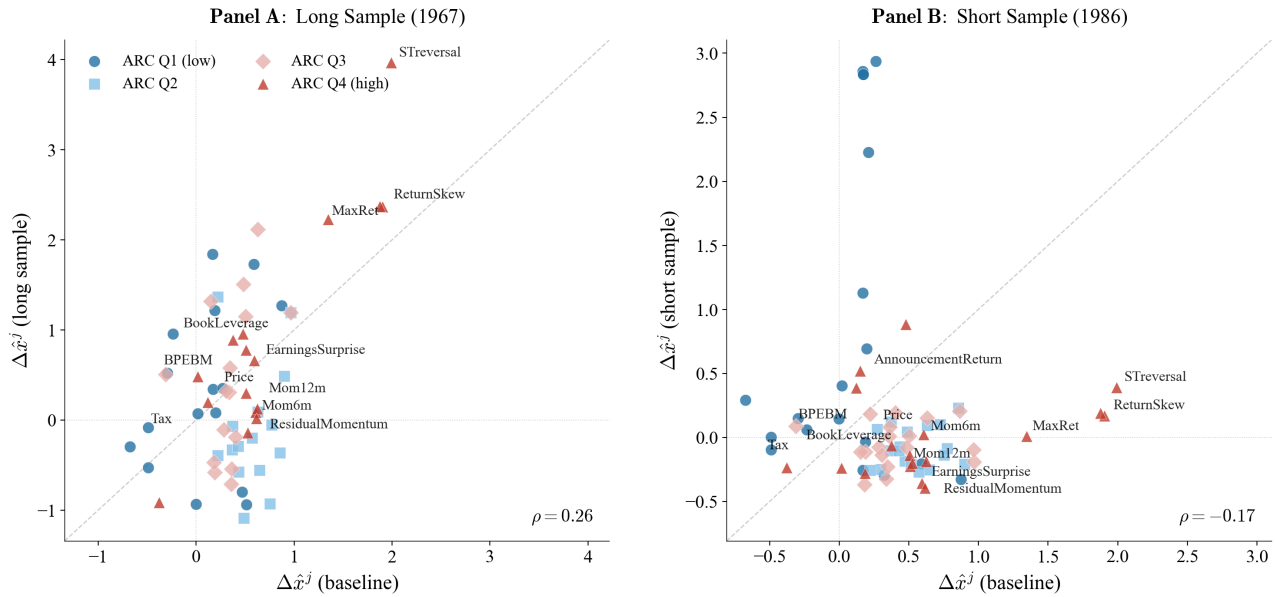


Figure D.16: **Characteristic reallocation: alternative samples vs. baseline.** Each point represents one firm characteristic. The x-axis plots the change in median tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs under the baseline specification (1972–2023). The y-axis plots the same quantity under each alternative sample. Positive values indicate higher frictionless exposure. Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). The dashed line is the 45-degree reference. Panel A: long sample (1967–2023, 62 characteristics). Panel B: short sample (1986–2023, 88 characteristics). Both use the GAN architecture under bid-ask costs and the baseline leverage constraint of $\omega_{it} \in (-5\%, +5\%)$.

greater reallocation away from high-turnover characteristics, yielding larger cross-sectional pricing improvements.

	M_{t+1} (frictionless)				M_{t+1}^{TC} (cost-aware)			
	EV	XS-R^2	SR (Gross)	SR (Net)	EV	XS-R^2	SR (Gross)	SR (Net)
Panel A: Price-impact costs ($5m)								
Full cross-section	0.060	0.137	0.347	0.313	0.100	0.222	0.326	0.311
Excl. microcaps	0.118	0.192	0.215	0.213	0.147	0.224	0.204	0.203
Tight fin. conditions	0.084	0.058	−0.119	−0.153	0.144	0.136	−0.183	−0.197
Loose fin. conditions	0.056	0.134	0.409	0.375	0.092	0.215	0.399	0.384
High VIX	0.066	0.105	0.362	0.331	0.110	0.191	0.328	0.315
Low VIX	0.042	0.099	0.366	0.317	0.070	0.138	0.377	0.355
Panel B: Price-impact costs ($50m)								
Full cross-section	0.060	0.137	0.347	0.002	0.114	0.234	0.244	0.225
Excl. microcaps	0.118	0.192	0.215	0.195	0.166	0.236	0.194	0.191
Tight fin. conditions	0.084	0.058	−0.119	−0.441	0.163	0.160	−0.107	−0.120
Loose fin. conditions	0.056	0.134	0.409	0.060	0.105	0.225	0.314	0.293
High VIX	0.066	0.105	0.362	0.046	0.125	0.208	0.232	0.216
Low VIX	0.042	0.099	0.366	−0.105	0.078	0.138	0.325	0.295
Panel C: Price-impact costs ($500m)								
Full cross-section	0.060	0.137	0.347	−1.355	0.143	0.258	0.157	0.139
Excl. microcaps	0.118	0.192	0.215	0.010	0.196	0.257	0.147	0.143
Tight fin. conditions	0.084	0.058	−0.119	−1.346	0.211	0.218	0.020	0.010
Loose fin. conditions	0.056	0.134	0.409	−1.384	0.132	0.243	0.201	0.180
High VIX	0.066	0.105	0.362	−1.293	0.158	0.254	0.146	0.130
Low VIX	0.042	0.099	0.366	−1.567	0.098	0.122	0.278	0.242

Table D.8: **Robustness to market conditions: price-impact costs.** This table reports out-of-sample performance metrics for the GAN architecture under price-impact cost specifications and leverage constraint of $\omega_{it} \in (-5\%, +5\%)$. M_{t+1} denotes the frictionless SDF; M_{t+1}^{TC} denotes the cost-aware SDF. EV is explained variation; XS-R^2 is cross-sectional R^2; SR (Gross) and SR (Net) are monthly Sharpe ratios before and after transaction costs. The frictionless model is the same across all three panels; columns differ only in the cost specification used to compute net Sharpe ratios. Financial conditions are classified using the National Financial Conditions Index (NFCI) from the Chicago Fed; market uncertainty is classified using the VIX median. The test period is January 1997 to November 2023.

Net Sharpe ratios highlight the implementability benefits of cost-awareness in the face of price impact. Under the $5m constraint, costs are small enough that even the frictionless portfolio retains most of its gross performance, so the cost-aware and frictionless net Sharpe ratios are similar. Under $50m, the frictionless portfolio begins to suffer meaningful implementation shortfalls, particularly during tight financial conditions (net SR of −0.441 vs. −0.120 for the cost-aware SDF). Under $500m, the frictionless portfolio is catastrophically eroded by price impact (net SR of −1.355 for the full cross-section), while the cost-aware SDF delivers positive

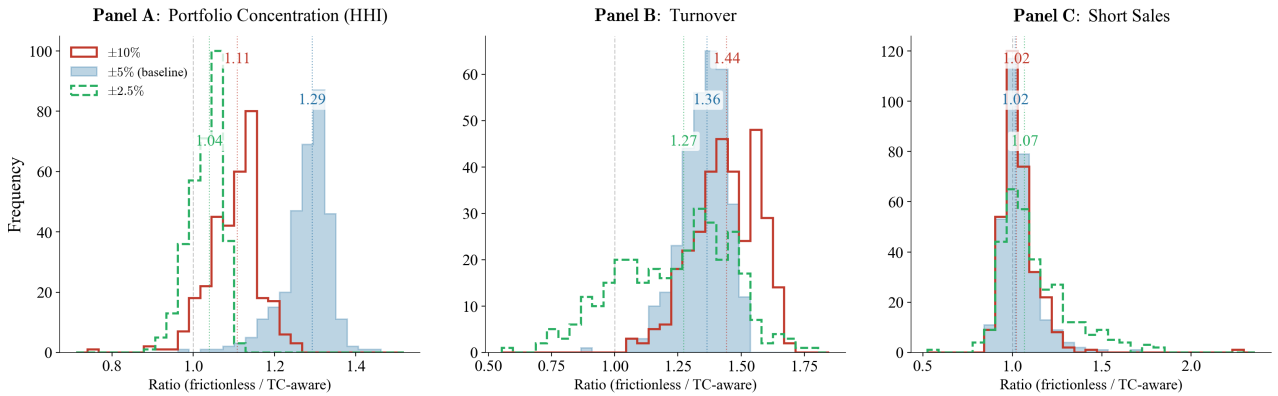


Figure D.17: **Tangency portfolio composition under alternative leverage constraints.** This figure reports the monthly ratio of three portfolio statistics between the frictionless SDF (M_{t+1}) and the cost-aware SDF (M_{t+1}^{TC}) for the GAN architecture under bid-ask costs. Panel A: Herfindahl-Hirschman Index (HHI) of portfolio concentration. Panel B: aggregate portfolio turnover. Panel C: fraction of short positions. Ratios above one indicate that the frictionless portfolio exhibits higher values. Red outlines: loose constraints $\omega_{it} \in (-10\%, +10\%)$; blue histograms: baseline constraints $\omega_{it} \in (-5\%, +5\%)$; green dashed: tight constraints $\omega_{it} \in (-2.5\%, +2.5\%)$. Dashed vertical lines indicate median ratios. The out-of-sample period is January 1997 to November 2023.

net Sharpe ratios across all regimes. Section 6.3 discusses the results.

D.14 SDF Under Alternative Leverage Constraints

Figure D.17 reports portfolio composition ratios under tight (±2.5%) and loose (±10%) leverage constraints alongside the baseline (±5%). These correspond to $\kappa \in \{0.05, 0.10, 0.20\}$ in Equation (10), respectively.

Turnover ratios increase monotonically as constraints loosen: 1.27 (tight), 1.36 (baseline), and 1.44 (loose), so the frictionless kernel trades more aggressively with more room. Concentration ratios are 1.04 under tight constraints, 1.29 at the baseline, and 1.11 under loose constraints, so under tight limits the two portfolios are about equally concentrated, and the reallocation runs through turnover.

At the characteristic level, Figure D.18 shows that the reallocation pattern amplifies as constraints loosen. Under tight constraints (Panel A), the correlation between exposure shifts and SHAP attribution shifts is $\rho = 0.43$, with most characteristics clustered near the origin. Under loose constraints (Panel B), the correlation rises to $\rho = 0.75$ as the frictionless model exploits its greater flexibility to load heavily on high-turnover characteristics — particularly

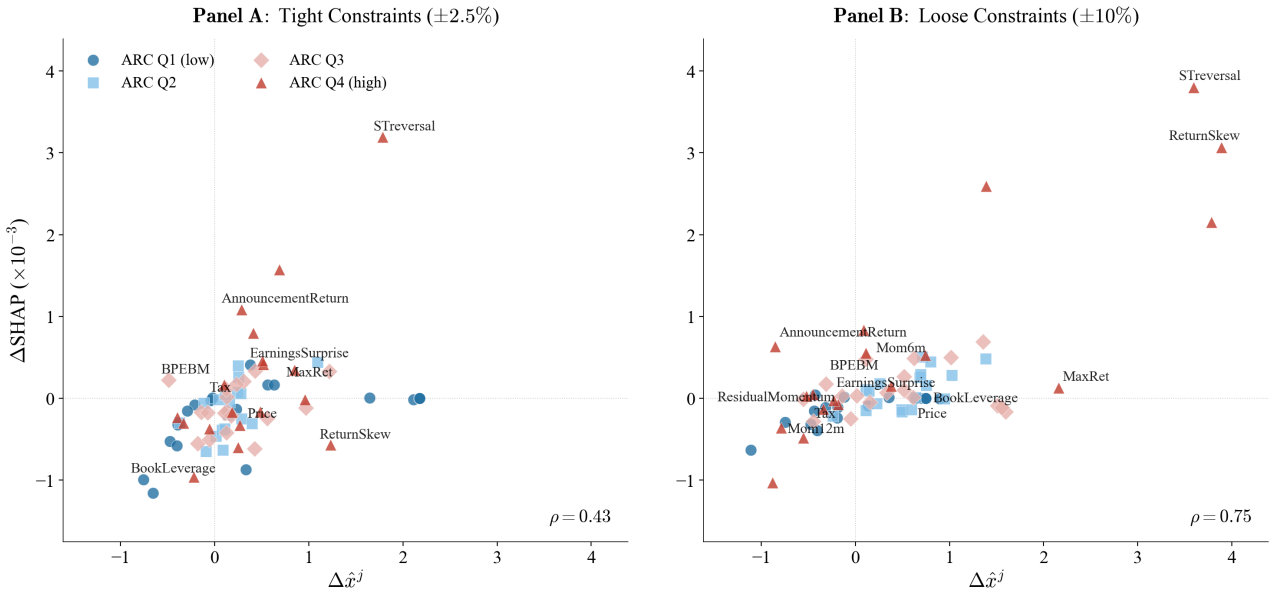


Figure D.18: **Characteristic reallocation under alternative leverage constraints.** Each point represents one of the 73 firm characteristics. The x-axis plots the change in median tangency portfolio exposure $\Delta\widehat{x}^j$ between frictionless and cost-aware SDFs. The y-axis plots the corresponding change in DeepSHAP attribution ΔSHAP^j ($\times 10^{-3}$). Positive values indicate higher frictionless exposure or attribution. Markers are color-coded by ARC quartile: Q1/low (dark blue circles), Q2 (light blue squares), Q3 (salmon diamonds), Q4/high (red triangles). Panel A: tight constraints $\omega_{it} \in (-2.5\%, +2.5\%)$. Panel B: loose constraints $\omega_{it} \in (-10\%, +10\%)$. All results use the GAN architecture under bid-ask costs. The out-of-sample period is January 1997 to November 2023.

short-term reversal, return skewness, and maximum daily return — which the cost-aware model then disciplines.