Focus or Spread? Attention Allocation under High-Dimensional Uncertainty

Hayong Yun*
November 23, 2026

Abstract

Many economic decisions involve more variables than anyone can track. We introduce rotate-then-quantize (RTQ) to economics for such problems. A random rotation turns a high-dimensional problem into many one-dimensional ones, with a worst-case loss guarantee that holds for any prior. We apply it to rational inattention and ask when attention should be focused on a few variables or spread across many. Focusing pays when a few risks dominate and spreading pays otherwise. Evidence from US equities is broadly consistent with the theory. RTQ is a promising tool wherever the curse of dimensionality plagues economics.

JEL Classification: D83, D81, G11
Keywords: Rational inattention, attention allocation, high-dimensional uncertainty, portfolio choice, vector quantization, curse of dimensionality

*Department of Finance, Eli Broad College of Business, Michigan State University, 667 North Shaw Lane, Room 339, East Lansing, MI 48824. Email: yunhayon@msu.edu.

Focus or Spread? Attention Allocation under High-Dimensional Uncertainty

Abstract

Many economic decisions involve more variables than anyone can track. We introduce rotate-then-quantize (RTQ) to economics for such problems. A random rotation turns a high-dimensional problem into many one-dimensional ones, with a worst-case loss guarantee that holds for any prior. We apply it to rational inattention and ask when attention should be focused on a few variables or spread across many. Focusing pays when a few risks dominate and spreading pays otherwise. Evidence from US equities is broadly consistent with the theory. RTQ is a promising tool wherever the curse of dimensionality plagues economics.

JEL Classification: D83, D81, G11
Keywords: Rational inattention, attention allocation, high-dimensional uncertainty, portfolio choice, vector quantization, curse of dimensionality

1 Introduction

Many economic decisions involve more variables than anyone can track. An investor chooses among thousands of stocks, a central bank watches hundreds of economic indicators, and a retailer sets prices for thousands of products. As the number of variables grows, the work needed to use them well grows much faster, a problem Bellman (1957) named the curse of dimensionality.[1] The problem is getting worse in the digital age, as decision makers have access to far more data than they can process.[2] In finance alone, researchers have documented hundreds of variables that appear to predict stock returns (Cochrane, 2011; Harvey et al., 2016), and central banks build models that draw on hundreds of data series (Bernanke et al., 2005). Forecasters who try to use all of these predictors run into exactly this problem (Giannone et al., 2021), and for investors with limited attention, more information can make decisions worse rather than better (Bernales et al., 2023).

This paper studies how to address the curse of dimensionality in rational inattention. Rational inattention, the economic theory of how people cope with more information than they can process, is the natural place to study it. In the theory of Sims (2003), a decision maker has a limited capacity to process information and must decide what to pay attention to. With many variables the theory is hard to apply unless uncertainty is normally distributed. In portfolio choice, Van Nieuwerburgh and Veldkamp (2010) show that an investor who chooses jointly what to learn and what to hold may specialize in a few assets, because learning about an asset makes it more attractive to hold and holding more of it makes learning more valuable, which helps explain why many portfolios are concentrated. Outside the normal case, the problem has been hard to solve in many dimensions, and a signal designed for one distribution can perform badly under another. Sims (2006) conjectured that the best signals would often be discrete rather than normal, but how to find them when there are many variables has remained open. So economists have had little guidance on how a decision maker should spread limited attention across many variables, or on when focusing on a few is best.

We address this problem with a recently developed method, rotate-then-quantize (RTQ), introduced by Zandieh et al. (2026) under the name TurboQuant. Its practical importance became clear in March 2026, when Google described TurboQuant as a way to cut the memory that large language models need, and memory-chip stocks such as Micron and SanDisk fell sharply on the news.[3] A random rotation mixes all the variables together, so that every rotated variable has the same known

[1] For a nontechnical account of why the curse of dimensionality limits what economic models can describe, see Ben Klemens, "The Heisenberg Uncertainty Principle of Social Science Modeling," *Scientific American*, July 7, 2020.

[2] Benoît Cœuré, then a member of the Executive Board of the European Central Bank, describes how data generated by technology in daily life give policymakers new information and new challenges in "Policy analysis with big data," speech at the Banque de France conference "Economic and Financial Regulation in the Era of Big Data," Paris, November 24, 2017.

[3] See, for example, Charlie Garcia, "Google just sucker-punched these highflying tech stocks - don't let the relief rally fool you," *MarketWatch*, April 7, 2026.

distribution, whatever the original data look like. A problem with N interacting variables then becomes N separate one-variable problems, each solved with the same simple rounding rule. This is how RTQ gets around the curse of dimensionality: its worst-case loss stays within a fixed multiple of the best that any signal with the same information capacity can achieve, however the variables are distributed, and the guarantee does not weaken as the number of variables grows. Section 2.2 shows in a picture why rotation works and relates it to the Johnson-Lindenstrauss lemma.

To see what this means for attention, consider an investor choosing among 500 stocks under a fixed attention budget. A focused investor watches five stocks closely and ignores the other 495. An RTQ investor instead builds 500 composite signals, each a weighted basket of all 500 stocks, and reads every composite at coarse resolution. Because every composite mixes all stocks, coarse readings add up to a rough estimate of every stock. Both investors use the same attention; one concentrates it and the other spreads it.

We use RTQ to build a framework for rational inattention in high dimensions. The framework does not require normal uncertainty and works the same way with ten variables or ten thousand. We do not claim that RTQ beats strategies built for a known prior: when uncertainty is normal, the classical solution, which allocates capacity across the main directions of risk, is optimal. RTQ's advantage is robustness. A signal designed for one prior can fail badly under another with the same covariance, as we show by simulation, while RTQ performs the same under every prior.

The framework delivers three results. First, RTQ's worst-case loss is within a known constant of the lowest worst-case loss that any signal with the same information capacity can achieve, for any prior and any quadratic loss. With many variables the constant is $C_T \cdot \chi(W) \approx 2.72\,\chi(W)$, where $C_T = \pi\sqrt{3}/2$ and $\chi(W)$ is the condition number of the loss weights W, and it does not grow with the number of variables. Zandieh et al. (2026) study the case $W = I$; we extend it to the general case that economic decisions require. Second, under RTQ the decision maker should pay most attention to the composite signals where mistakes are most costly, so stakes, not volatility, drive attention. Third, focusing on a few variables becomes arbitrarily worse than RTQ as the attention budget grows, because the variables left unwatched leave a loss that extra capacity cannot remove. Together these results answer the question of when to focus: focusing costs little when a few sources of risk dominate, and a great deal otherwise. Focusing bets that a few known directions carry most of the risk, and spreading with RTQ stays safe when that bet is wrong, so RTQ's robustness and the focus-or-spread question are two sides of the same result. Neither strategy is always better, and when conditions favor spreading, RTQ is an efficient way to do it.[4]

We take the framework to the US stock market, a setting with thousands of variables where the

[4]The Internet Appendix extends the framework to an agent facing many separate decisions under one attention budget, such as forecasting many assets. It gives a simple rule for when spreading beats concentrating and shows how to infer the number of characteristics that matter from how outcomes rise with the number of targets (Sections IA.2 and IA.3).

curse of dimensionality bites hard. Using the full CRSP universe from 1971 to 2026, we compare a broad allocation of attention with a concentrated benchmark that spends the same attention budget on the single most important source of risk. The broad allocation divides the budget across all directions of the estimated return covariance by the allocation rule of our method, which we call the RTQ allocation. It uses the covariance and no random rotation, so this test measures the value of spreading attention rather than concentrating it; the separate value of the random rotation is shown by the simulation in Section 4.6 and examined in the return data in Table 2. The RTQ allocation has the lower welfare cost in every quarter, including out of sample, where the allocation is chosen on the past 60 months and judged on the next 24. With two bits of attention per stock, the median out-of-sample gain is about one third of the cost of having no information, or 206 basis points per year in certainty-equivalent return on a scale where having no information costs 600. As the theory predicts, the gain is larger when risk is spread across many sources: 338 basis points when the top source of risk matters least and 166 when it matters most, although out of sample this relation is only marginally significant. The gain also shows up in returns an investor could have earned. Minimum-variance portfolios built with the RTQ allocation have annual volatility of 10.7 percent among the 500 largest stocks, against 12.7 percent with concentrated attention, and match portfolios built with full use of the estimated covariance. We also run RTQ itself, with a random rotation and without the covariance. Once the budget reaches two bits per stock, it beats the concentrated benchmark in every quarter among the 500 largest stocks and in 99 percent of quarters among all stocks, by a median of 41 and 96 basis points, and its loss varies little across quarters. In the typical quarter, however, it does worse than strategies that spread attention using the estimated covariance, which is informative for US stocks. A version that uses the covariance only to sort risks into a few groups, and applies the random rotation inside each group, matches the best of those strategies among the 500 largest stocks and has a lower loss among all stocks. As an illustration, not a practical application, of a setting where the covariance is hard to learn, we also track monthly venture funding rounds across 454 Crunchbase categories, with two years of history for each estimate. Measured as monthly changes, where risk is spread across many directions, RTQ has a lower loss than every other strategy in every quarter, including those that use the estimated covariance. Where a few broad directions dominate, as across 49 broader category groups and in the slow common trend of category levels, the covariance-based strategies do better, because such directions can be learned even from a short history. As between focusing and spreading, what decides is how much of the risk a few directions carry. Most of the reduction in portfolio volatility comes from paying some attention to every source of risk; how attention is divided among them adds little.[5]

[5] Among the 500 largest stocks, where the top source of risk matters more, the median out-of-sample welfare gain is 42 basis points per year. A calibration on an S&P 500 proxy is in Internet Appendix IA.1.

We also look at how institutional investors actually behave, using LSEG 13F holdings merged with CRSP from 1980 to 2024 (352,755 manager-quarters across 10,693 managers with at least 20 holdings). We do not argue that managers use RTQ. The RTQ allocation and the concentrated benchmark stand for a broad and a narrow strategy, and we ask when each kind of manager does better. Holdings do not reveal how managers gather information, so we treat this evidence as descriptive. Neither broad nor concentrated portfolios earn more on average, but the winner changes often: estimated year by year, concentrated portfolios earned significantly more in 18 of 45 years and broad portfolios in 16. This is consistent with the theory, under which neither strategy dominates in general, and with other evidence that broad strategies often do well: index funds tend to outperform actively managed funds after fees (Carhart, 1997; Fama and French, 2010), and multi-product firms adjust prices across many products rather than a few (Bhattarai and Schoenle, 2014; Bonomo et al., 2023). In the 13F data, however, the measure of risk concentration does not predict which kind of portfolio wins, so this evidence does not test the mechanism of the model; it is reported in Internet Appendix IA.7.4.[6]

Our main contribution is to introduce RTQ to economics and finance as a general tool for problems where the curse of dimensionality binds. To our knowledge, this is the first paper to apply TurboQuant in economics or finance. Random projections in the spirit of the Johnson-Lindenstrauss lemma are already used successfully in econometrics to shrink large forecasting models (Boot and Nibbering, 2019; Koop et al., 2019). RTQ is built for another purpose. Instead of keeping fewer variables at full precision, it keeps every variable at coarse precision. Because the rotation gives every coordinate the same known distribution and makes the coordinates nearly independent, one fixed rounding rule works for any data, without the covariance or a distributional assumption, and its worst-case loss is within a known factor of the best achievable. Economists already use tools built for high-dimensional data, such as shrinkage estimators of large covariance matrices (Ledoit and Wolf, 2004; Fan et al., 2013). RTQ adds a tool for a different task: deciding what information to acquire, with a guarantee that does not depend on the number of variables or on their distribution. TurboQuant's guarantee is stated for errors that count equally across variables, which suits its purpose of storing data. In many economic problems some variables matter more than others, so we extend the guarantee to losses that weight variables differently (Proposition 1) and derive how to divide attention across them (Proposition 3). Beyond portfolio choice, RTQ fits multi-product pricing, monetary policy with many state variables, and forecasting many variables at once, each of which deserves its own study. The framework also relates to the growing use of AI models in investment management such as Cong et al. (2026); Internet Appendix IA.4 shows that

[6]We also test this extension outside finance, in the Fisman et al. (2006) speed-dating experiment (551 participants, 8,378 four-minute conversations, 21 sessions with 5 to 22 partners per side). Regressing $\log(1 + \text{matches})$ on $\log(\text{partners})$ gives a slope of 0.472 (session-clustered standard error 0.063), which implies that about 1.89 traits matter; principal components of the trait ratings give 2.68. See Internet Appendix IA.3.

RTQ comes within a constant factor of the loss floor that an optimal linear cross-asset attention layer attains at a given bit budget.

We also contribute to the rational inattention literature. Sims (2003, 2006) develop the theory and the conjecture about discrete signals, Matějka and McKay (2015) characterize discrete choice under rational inattention, Maćkowiak and Wiederholt (2009, 2015) study price setting and business cycles, and Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk et al. (2016) study portfolio choice and fund managers' attention; Maćkowiak et al. (2023) survey the field. We add a way to approach the problem in many dimensions for any prior, within a known factor of the best worst-case loss, with discrete signals of the form Sims (2006) conjectured, and a criterion for when focusing or spreading is better. The rotation is a mathematical device rather than a literal description of behavior, in the same way that competitive equilibrium describes attainable allocations without requiring anyone to solve the auctioneer's problem.

Finally, we contribute to the empirical study of asset managers (Carhart, 1997; Fama and French, 2010; Kacperczyk et al., 2014). We show that broad and concentrated institutional portfolios each outperform in many years, with no stable winner, and that minimum-variance portfolios built with spread-out attention have lower realized risk than those built with concentrated attention. We also show that the share of portfolios that look closer to a broad benchmark than to a single-direction benchmark is largely mechanical, since random portfolios score as high, so such statistics need a random-portfolio comparison before they can be read.

Section 2 reviews rational inattention and explains why rotation helps. Section 3 sets up the high-dimensional attention problem. Section 4 presents RTQ and the theorems. Section 5 takes the framework to US stocks and institutional holdings. Section 6 concludes.

2 Background

This section reviews rational inattention, what it can solve and where a large number of variables makes it hard. It then explains, without formulas, what rotate-then-quantize is, why it works, and where it is useful. Section 4.1 states the method formally, and Internet Appendix IA.6 derives it.

2.1 Rational Inattention

Rational inattention starts with Sims (2003). An agent can process only a limited amount of information, measured in bits as the Shannon mutual information between the state and the signal she observes. In the baseline, the state is a single normally distributed variable and the agent minimizes quadratic loss. The best signal is then normal, and the remaining uncertainty has a simple formula that depends only on the prior variance and the capacity. Sims (2006) pointed out that outside this

normal-quadratic case the best signals are often discrete, which Jung et al. (2019) later confirmed for problems with a discrete set of actions. Saint-Paul (2017) builds a "quantized" version of rational inattention in which the action is a deterministic, step-like function of the state, so that limited information shows up as lumpy behavior rather than as noise. Our signals are also discrete; they come from rounding random composites of many variables, and we ask how to spread a bit budget across them.

Matějka and McKay (2015) extend the framework to discrete choice with general payoffs and show that rational inattention produces multinomial-logit choice probabilities, which connects it to demand estimation. Their result is for a finite set of actions. Caplin and Dean (2015) show how choice data can reveal whether behavior is consistent with some cost of acquiring information, and Caplin et al. (2019) use these ideas to study which options a decision maker considers at all. Denti (2022) characterizes the information costs consistent with standard axioms of rational behavior.

A large applied literature builds rational inattention into economic models. Maćkowiak and Wiederholt (2009) study price setting by firms with limited attention to cost and demand conditions, which produces sticky prices, and Woodford (2009) develops a related model of slow, state-dependent price adjustment. Maćkowiak and Wiederholt (2015) extend the analysis to a business-cycle model in which firms split attention between firm-specific and aggregate shocks. Steiner et al. (2017) study rational inattention over time and show that it produces inertia and delayed responses. In finance, Van Nieuwerburgh and Veldkamp (2010) show that limited attention makes investors hold concentrated portfolios of familiar assets. Maćkowiak et al. (2023) survey the field.[7]

To keep the problem tractable, these applied papers limit the number of state variables or the kind of signal, usually by assuming normal signals or letting the agent observe only a few linear combinations of the state. We offer a signal structure that works for any prior and any number of variables, with worst-case loss within a known constant of the best achievable.[8]

Recent work in machine learning builds portfolio models with the attention mechanism of Transformer models (Vaswani et al., 2017). Cong et al. (2026) introduce AlphaPortfolio, a Transformer-based model that attends across assets, is trained directly on the portfolio objective, and performs

[7] Other models of limited attention use a different cost. Kőszegi and Szeidl (2013) propose a model in which people overweight attributes that vary more across options, and Gabaix (2014) develops a sparsity-based model with closed-form solutions that is widely used in macroeconomics and finance. Like us, these papers ask how attention is spread across many dimensions; we keep the Shannon cost of Sims (2003). A separate literature studies how cognitive limits affect the use, rather than the acquisition, of information: Charles et al. (2024) show in experiments that investors pull valuations toward a default and act on stated beliefs only weakly, which can explain inelastic demand. Our question comes earlier: which signals an investor with limited capacity should acquire across many variables.

[8] Kamenica and Gentzkow (2011) introduce Bayesian persuasion, and Bergemann and Morris (2016) characterize the outcomes achievable across information structures in games. These papers ask what a designer should reveal; we ask what an agent should pay attention to. The rotation approach could also help an information designer facing many dimensions, since its guarantee holds for any prior with finite variance.

well out of sample. AlphaPortfolio and RTQ both spread attention across many assets. AlphaPortfolio learns its attention from data; RTQ is a closed-form rule with a worst-case guarantee and composite signals that are easy to interpret. Internet Appendix IA.4 shows that, at a given bit budget, RTQ comes within a constant factor of the loss of the best linear attention layer, for any prior with finite variance. The result covers the linear part of the architecture; we discuss the other parts of a Transformer there, where the connection is informal.

2.2 Rotate-Then-Quantize: The Idea

Rotate-then-quantize (RTQ) was developed to solve a storage problem in artificial intelligence. A large language model keeps a long list of numbers for every word in a conversation, and storing all of these lists at full precision uses a great deal of memory. Zandieh et al. (2026), who call the method TurboQuant, show that each number can be stored with only two to four bits, instead of the usual sixteen or thirty-two, with little loss of accuracy, provided the list is first scrambled by a random rotation. Earlier work on compressing data sent between computers used random rotations in a similar way (Vargaftik et al., 2021). Two features made TurboQuant attractive. It needs no training data, so the same rule works for every list and each list can be compressed the moment it arrives. And it comes with a proof that its worst-case error is close to the smallest worst-case error any method could achieve with the same number of bits. Finance has the same problem: a day of returns on 5,000 stocks is a list of 5,000 numbers, and an investor who can absorb only a little about each stock faces the same trade-off as the computer.

The simplest case shows the idea. Suppose an investor follows two stocks and can record each day's return on each stock only roughly, as "up a lot," "up a little," "down a little," or "down a lot." On most days this works well. But on a day when the first stock jumps 10 percent on earnings news and the second does not move, the four categories must stretch to cover 10 percent, and they become too coarse for the many days of small moves. Now replace the two stocks with two portfolios: an equal-weighted portfolio long both stocks, and a long-short portfolio long the first and short the second. This is a rotation of the two returns by 45 degrees, and it loses nothing, since the two stock returns can be recovered exactly from the two portfolio returns. On the earnings day, both portfolios move by about 7 percent, so the single large move is shared between them. A rotation does not change the size of a day's move; it spreads the move so no one number carries it.

With many stocks, the shape of a day's returns can vary a great deal, and this is what makes rounding hard. Suppose each return can be recorded with two bits, which means it must be rounded to one of four levels. Any fixed set of four levels suits some days and fails on others. On an earnings day, two stocks carry almost the entire move and the rest are close to zero. On a market day, all stocks move together by similar amounts. Levels set for a market day are far too small for the two

large moves on an earnings day, and levels wide enough for the earnings day round every return on a market day to the same value. Tailoring the levels to each day requires knowing the shape of the day in advance, which is exactly what an investor with limited attention does not know. RTQ avoids the problem by replacing the stocks with composite portfolios, as in the two-stock case. Each composite holds every stock, with weights that are random but fixed in advance and used every day, and the stock returns can be recovered exactly from the composite returns.

Figure 1 shows RTQ at work on 16 stocks on three kinds of day. Each column is one day: an earnings day on which two stocks move, a market day on which all stocks move together, and a rotation day on which winners and losers alternate. Row (a) shows the stock returns, row (b) the returns on the 16 composites with their rounded values (diamonds), and row (c) the stock returns recovered from the rounded composites (dots) against the true returns (bars). Look first at row (b). Although the three days look completely different in row (a), their composites in row (b) look alike: no composite dominates, and all are of similar size, so the same four rounding levels fit all three days. Then look at the error above each panel in row (c). Averaged over random choices of weights, RTQ's error is the same on all three days, about 11 percent of the day's squared size. Rounding each stock directly with the same four levels does as well or better on the market and rotation days, which happen to suit those levels, but it loses 40 percent on the earnings day. RTQ gives up a little on easy days in exchange for protection on hard ones, and it does not need to know in advance which kind of day it faces.

The composites look alike because of diversification. A portfolio of many stocks with spread-out weights is more predictable than any single stock, because the swings of individual stocks largely cancel. Each composite is such a portfolio, so whatever happened to individual stocks that day, its return falls in a predictable range. Because that range is known in advance, the four rounding levels can be chosen once to fit it, using the rounding rule that minimizes squared error for a known distribution (Lloyd, 1982; Max, 1960). One rule then serves every composite every day.

Figure 2 shows why this works, using 64 stocks. Panel (a) shows how the individual stock returns are spread out on the three kinds of day, and panel (b) does the same for the composite returns. In panel (a), the three days look nothing alike. In panel (b), all three fall on the same bell-shaped curve, and that curve depends only on the number of stocks, not on the day, the stocks, or any belief about their returns. The diamonds at the bottom of panel (b) are the four rounding levels chosen for that curve, closer together where values are common and farther apart in the tails. Panel (c) shows how the error changes as the number of stocks grows from 8 to 512. The solid line, RTQ, stays flat at about 11 percent, while the error from rounding each stock directly keeps rising, because days on which a few stocks dominate become harder to round as the list gets longer. Panel (d) illustrates the Johnson-Lindenstrauss lemma discussed below. Rotation turns every list

into one known shape, which keeps the error from growing with the number of variables.

Zandieh et al. (2026) prove that the pattern in panel (c) holds in general. For any list of numbers, RTQ's expected error has a known ceiling that does not depend on the list. For long lists, that ceiling is within a factor of about 2.7 of the smallest error that any method, however complex, could guarantee with the same number of bits. This is how RTQ escapes the curse of dimensionality: it needs one rotation and one rounding rule, and the guarantee does not weaken with more variables.

The same diversification logic underlies an older and better-known result. Johnson and Lindenstrauss (1984) show that a random projection can squeeze a large set of points into far fewer coordinates while keeping the distances between them almost unchanged. The number of coordinates needed depends on how many points there are, not on how many coordinates they started with. Each new coordinate is a random blend of all the original ones, and the squared distance after projection is an average over many such blends, which settles close to its true value as more coordinates are kept. Panel (d) of Figure 2 shows this for 200 points with 1,000 coordinates each. It plots, across all pairs of points, the distance after projection divided by the true distance. Keeping 20 coordinates gives a wide spread around one; keeping 320 gives a tight spike at one, so the distances are almost exactly preserved with a third of the coordinates. Random projections of this kind are already used in econometrics to shrink large forecasting models: Boot and Nibbering (2019) average forecasts over many random projections of the predictors and find that this improves forecasts of key macroeconomic series, and Koop et al. (2019) use random compression to estimate large vector autoregressions. The two methods compress in different ways. Johnson-Lindenstrauss keeps fewer numbers, each at full precision. RTQ keeps all the numbers, each at coarse precision. Zandieh et al. (2025) combine the two by keeping a single sign bit for each coordinate of a random projection. The fast form of the rotation used in practice, random signs followed by a Hadamard matrix, is the first step of the fast Johnson-Lindenstrauss transform of Ailon and Chazelle (2009).

RTQ is useful whenever a long list of numbers must be summarized with limited capacity and the shape of the list is not known in advance. In finance this includes tracking returns, risk exposures, or forecasts for thousands of securities; storing and comparing the long numerical summaries of filings and news that AI tools now produce, which is the use TurboQuant was built for; and monitoring many signals at once, as a risk manager, a regulator, or a central bank must. RTQ is least useful when a few variables carry almost all that matters; then focusing on those few is better, and Section 4.5 gives a criterion for telling the two cases apart.

In the attention problem, each composite is a signal that loads on every variable, much as a market index loads on every stock. RTQ describes an investor who pays coarse attention to many such composites rather than close attention to a few individual stocks. We take the rotation, the shape of the composites, the rounding rule, and the error guarantee for equally weighted errors from Zandieh et al. (2026). This paper adds the extension to losses that weight errors differently across

variables (Proposition 1), the allocation of attention across variables by stakes (Proposition 3), the loss floor of focusing on a few variables (Proposition 2), the criterion for when to focus and when to spread (Section 4.5), and the application to portfolio choice. Table IA.3 in the Internet Appendix lists each piece with its source.

3 Model

We consider an agent who must make many decisions at once under uncertainty, such as a portfolio manager choosing weights for hundreds of stocks or a retailer setting prices for thousands of products. The agent cannot fully observe the state and must decide in advance what to attend to.

The model nests the univariate case of Sims (2003) when $d = 1$. In high dimensions, two new forces appear: decisions interact (getting one stock weight wrong changes how costly it is to get another wrong), and the agent must spread a fixed attention budget across many variables. The number of ways to design signals about d interacting variables grows explosively with d; this is how the curse of dimensionality enters the problem.

3.1 Environment

State. The state is a vector $\mathbf{x} \in \mathbb{R}^d$ drawn from a prior F with mean μ and finite second moments. Each x_i is one piece of payoff-relevant information. We impose no other restriction on F: it may be normal, fat-tailed, skewed, or discrete. Three examples:

- A fund manager choosing weights for $d = 500$ stocks, where $\mathbf{x}$ is the vector of next-period excess returns. Some move together (tech stocks), others are nearly independent.
- A grocery chain setting prices for $d = 2{,}000$ products, where $\mathbf{x}$ is the vector of marginal costs. Some products are substitutes (Coke and Pepsi), so their costs matter jointly.
- A central bank tracking $d = 50$ macro indicators across countries, where $\mathbf{x}$ captures current conditions, some lagging, some leading.

Action. After observing a signal about $\mathbf{x}$, the agent chooses $\mathbf{a} \in \mathbb{R}^d$, one decision per state variable: a weight for each stock, a price for each product, a forecast for each indicator.

Loss function. The agent's loss from choosing $\mathbf{a}$ when the state is $\mathbf{x}$ is

$$L(\mathbf{a}, \mathbf{x}) = (\mathbf{a} - \mathbf{x})^\top W (\mathbf{a} - \mathbf{x}), \tag{1}$$

where W is a $d \times d$ symmetric positive definite matrix. The quadratic form is the second-order approximation to any smooth objective around the full-information optimum. If the agent has

objective $U(\mathbf{a},\mathbf{x})$ with full-information optimal action $\mathbf{a}^*(\mathbf{x}) = \mathbf{x}$, the loss from choosing $\mathbf{a}$ instead is $U(\mathbf{a}^*(\mathbf{x}),\mathbf{x}) - U(\mathbf{a},\mathbf{x}) \approx \frac{1}{2}(\mathbf{a}-\mathbf{x})^\top H(\mathbf{x})(\mathbf{a}-\mathbf{x})$, where $H(\mathbf{x}) = -\partial^2 U/\partial \mathbf{a}^2$, and W corresponds to a constant or expected Hessian. For CARA portfolio choice with risk aversion γ, $W = \gamma\Sigma$; for pricing, W is the Hessian of the profit function and reflects demand elasticities; for a central bank, W reflects how forecast errors feed into the welfare loss from suboptimal policy. The quadratic form therefore does not assume risk neutrality; as in Sims (2003), preferences enter through W.

The matrix W is the central object of the model. If $W = I$, equation (1) is the sum of squared errors, and the penalties for different variables do not interact. In practice, errors interact. If a fund manager overweights two correlated tech stocks, the combined mistake is worse than the sum of the individual mistakes, because the errors compound rather than diversify. If a retailer misprices Coke, the cost depends on whether she also mispriced Pepsi, because consumers substitute. The matrix W captures these *decision complementarities*: W_{ii} measures how costly it is to get variable i wrong in isolation, and W_{ij} measures how much the cost of an error in i depends on the error in j. The off-diagonal terms are what make the high-dimensional attention problem different from d copies of the one-dimensional problem. With $W = I$, the agent can treat each variable separately; with general W, the right attention for variable i depends on how much goes to variable j.

3.2 Information Constraint

The agent has limited capacity for processing information. Before observing $\mathbf{x}$, she chooses a *signal structure*: a conditional distribution $\pi(\cdot|\mathbf{x})$ that generates a signal $\mathbf{s}$ given $\mathbf{x}$, from which she forms a posterior about $\mathbf{x}$. Following Sims (2003), the amount of information in the signal is the *Shannon mutual information*

$$I(\mathbf{x};\mathbf{s}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{s}), \tag{2}$$

the reduction in uncertainty about $\mathbf{x}$ from observing $\mathbf{s}$, measured in bits (base-2 logarithms). One bit is a binary distinction, and each additional bit doubles the resolution of the signal. The capacity constraint is

$$I(\mathbf{x};\mathbf{s}) \leq \kappa, \tag{3}$$

with $\kappa > 0$ the agent's total capacity in bits. Higher κ means more information and lower expected loss. We write $\kappa = bd$, where b is the average number of bits per dimension. Small b (say, 2 or 3) means attention is severely constrained; large b approaches full information.

The key property of mutual information for our purpose is *additivity*: if $\mathbf{x} = (x_1,\ldots,x_d)$ has independent coordinates and the signal is designed independently for each,

$$I(\mathbf{x};\mathbf{s}) = \sum_{i=1}^{d} I(x_i;s_i). \tag{4}$$

Additivity is what lets the problem decompose after rotation: a budget of $\kappa = bd$ bits becomes b bits for each of d rotated coordinates. The *rate-distortion function* $R(D)$ gives the minimum number of bits needed to reach expected squared error at most D. It is a lower bound on how accurately any signal structure can represent the state at a given capacity, and our main result bounds the loss of RTQ by a constant multiple of this floor.

3.3 Timing

The sequence of events is:

1. The agent knows the prior F, the loss matrix W, and capacity κ.
2. She designs a signal structure $\pi(\cdot|\mathbf{x})$ subject to $I(\mathbf{x};\mathbf{s}) \leq \kappa$.
3. Nature draws $\mathbf{x} \sim F$.
4. The agent observes $\mathbf{s} \sim \pi(\cdot|\mathbf{x})$ and forms posterior beliefs.
5. She chooses $\mathbf{a}$ to minimize expected loss given $\mathbf{s}$.

Step 2 is critical: the agent designs her information filter *before* seeing the state, which is the sense in which attention is allocated, not just used.

3.4 Optimal Action and Expected Loss

For any signal realization $\mathbf{s}$, the best response in step 5 is the posterior mean,

$$\mathbf{a}^*(\mathbf{s}) = \mathbb{E}[\mathbf{x}|\mathbf{s}], \tag{5}$$

for any positive definite W.[9] Ex ante expected loss is then determined by the quality of the signal, measured by the posterior error $\varepsilon = \mathbb{E}[\mathbf{x}|\mathbf{s}] - \mathbf{x}$:

$$\mathscr{L}(\pi) = \mathbb{E}\left[\varepsilon^\top W \varepsilon\right] = \operatorname{tr}(W \cdot \Sigma_\varepsilon), \tag{6}$$

where $\Sigma_\varepsilon = \mathbb{E}[\varepsilon\varepsilon^\top]$ depends on π (we write $\Sigma_\varepsilon(\pi)$ when needed). With $W = I$, the loss is total mean squared error. With large off-diagonal entries, the loss also penalizes correlated errors: when two related variables are wrong in the same direction, the loss is amplified.

[9] $\mathbb{E}[(\mathbf{a}-\mathbf{x})^\top W(\mathbf{a}-\mathbf{x})|\mathbf{s}]$ is convex quadratic in $\mathbf{a}$ with minimum at $\mathbb{E}[\mathbf{x}|\mathbf{s}]$ for any positive definite W.

3.5 Assumptions

Assumption 1 (Shannon mutual information cost). As in Sims (2003), the cost of information is Shannon mutual information between state and signal. Denti (2022) shows that Shannon cost belongs to the class of posterior-separable costs, which depend only on the distribution of posterior beliefs. Alternatives include Fisher information and neighborhood-based costs (Maćkowiak et al., 2023), but Shannon remains standard. Our results rely on additivity across independent components (equation 4), which is specific to Shannon entropy.

Assumption 2 (Ex ante signal design). As in Sims (2003), the agent commits to a signal structure before observing the state, as a policy, not in response to each realization. A portfolio manager plans her research before the trading day, and a central bank its monitoring before data arrive.

Assumption 3 (Quadratic loss). The loss is the weighted quadratic (1), following Sims (2003). It is exact for CARA portfolio choice with normal returns and for linear-quadratic regulation, and approximate for problems with substantial nonlinearities (option payoffs, bankruptcy thresholds).

Assumption 4 (Arbitrary prior with finite second moments). Unlike Sims (2003), who relies on normal priors, we allow F to be any distribution on $\mathbb{R}^d$ with finite second moments. Outside the normal case, optimal signals can be discrete and depend on the distribution (Jung et al., 2019), and the normal restriction rules out the fat tails, skewness, and discrete mixtures common in economic data. Our results need only finite second moments because the random rotation works through a geometric property of high-dimensional spaces, not through distributional assumptions.

Assumption 5 (General positive definite loss matrix). W is symmetric positive definite but otherwise unrestricted. Most applied work assumes $W = I$ or diagonal, which rules out interactions across dimensions. The rotation-based strategy of Zandieh et al. (2026) was developed for $W = I$; extending it to general W is one of our contributions, and it is the source of the condition number $\chi(W) = \lambda_{\max}(W)/\lambda_{\min}(W)$ in the main bound. Positive definiteness means every direction matters at least slightly; in the rank-one limit (a single policy instrument), focusing is trivially optimal and our results are not needed.

Assumption 6 (Known and fixed W). W is known to the agent and does not depend on $\mathbf{x}$, as is implicit in the quadratic tracking problem of Sims (2003). This is realistic when W reflects stable features of the environment (long-run return covariance, cross-price elasticities) and less so when risk aversion or demand shifts unpredictably. Allowing uncertain W requires modeling joint uncertainty over the state and the decision weights, which we leave for future work.

Assumption 7 (Action dimension equals state dimension). The agent chooses one action per state variable. This fits portfolio choice, multi-product pricing, and forecasting many indicators. It rules out problems in which the agent aggregates many state variables into a lower-dimensional decision, such as a central bank that observes 50 indicators but sets a single interest rate. That case is the low-rank limit of our setup, where selective attention is appropriate.

Assumption 8 (Static, single-period). As in Sims (2003), the agent makes one decision.[10]

Assumption 9 (Single agent, no strategic interaction). As in Sims (2003), the agent's payoff depends only on her own action and the state, not on others' actions or attention.[11]

3.6 The Agent's Problem

The agent chooses the signal structure that minimizes expected loss, given a capacity of κ bits:

$$\min_{\pi} \quad \operatorname{tr}(W \cdot \Sigma_{\varepsilon}(\pi)) \qquad \text{subject to} \quad I(\mathbf{x};\mathbf{s}) \leq \kappa. \tag{7}$$

The problem is well defined but hard to solve in general: the agent chooses a whole conditional distribution π, and the constraint is nonlinear in π. When F is normal, the solution has a closed form: it allocates capacity across the eigenvectors of the loss-weighted prior covariance by water-filling. When F is not normal, the best signal can be discrete and depends on the details of F (Jung et al., 2019), and there is no closed form in the many dimensions that matter in practice. Section 4 gives a signal structure that comes within a constant factor of the lowest worst-case loss of any signal with the same capacity, for any prior.[12]

4 Main Result

This section formalizes RTQ and the results that let it handle many variables. Section 4.1 states the strategy and the properties of random rotation it relies on. Section 4.2 shows that its loss stays within a constant factor of the lowest worst-case loss of any signal with the same information capacity, for any prior and any number of variables, which is how it gets around the curse of dimensionality. Section 4.3 shows that focusing on a few variables leaves a loss floor that extra capacity cannot remove. Section 4.4 derives how RTQ allocates attention across directions, and Section 4.5 turns these results into a criterion for when to focus and when to spread. Section 4.6 illustrates the guarantee by simulation, and Section 4.7 extends it to market equilibrium.

[10] The static setting separates how attention is spread across variables at a point in time from how it is spread over time. Our main results concern the first. Multi-period models of rational inattention are surveyed in Maćkowiak et al. (2023), and Internet Appendix IA.5.3 gives a dynamic extension of our framework.

[11] This rules out settings where one agent's information acquisition affects another's payoff, as in Hellwig and Veldkamp (2009) and Myatt and Wallace (2012). Our main bound does not directly apply there. Still, our results could serve as a building block for games with rational inattention: conditional on opponents' strategies, each agent faces a single-agent attention problem to which RTQ provides an approximate best response.

[12] Internet Appendix IA.2 extends the setup to an agent who tracks many separate targets at once under a shared bit budget, the special case in which W is block diagonal, and compares spreading and concentrating attention there.

4.1 The Rotate-Then-Quantize Strategy

Section 2.2 explained the idea behind rotate-then-quantize (RTQ) with an example. This subsection states the strategy formally, gives the properties of random rotation that the rest of the paper uses, and summarizes what carries over to other economic problems. Steps 1-3 apply the vector quantization procedure of Zandieh et al. (2026), which combines the scalar quantizer of Han et al. (2026) with the random-rotation preconditioning of Zandieh et al. (2025). Step 4 and the allocation of bits across coordinates are our additions for the rational-inattention setting.

Step 1: Separate size and direction. Write $\mathbf{x}$ as its size $\rho = \|\mathbf{x}\|$ times its direction $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$. In the main case, ρ is public side information: the agent observes it under every strategy we compare, and it does not count against the budget, so all κ bits go to the direction and the capacity constraint reads $I(\mathbf{x};\mathbf{s} \mid \rho, R) \leq \kappa$. Alternatively, ρ can be rounded with b_0 bits, leaving $\kappa - b_0$ bits for the direction, at the cost of a looser bound (Appendix A.2).

Step 2: Rotate. Multiply the direction by a random orthogonal matrix, $\mathbf{r} = R\mathbf{u}$. The matrix R is drawn once, before any data arrive, and used for every state. Our results are for a Haar-random R. In practice a randomized Hadamard transform is used as a fast approximation; its rotated coordinates follow the distribution below only approximately.

Step 3: Round each coordinate. Give b_i bits to coordinate i of $\mathbf{r}$, with $b_0 + \sum_i b_i \leq \kappa$, and round each r_i separately with the scalar quantizer built for the $\text{Beta}(\frac{d-1}{2}, \frac{d-1}{2})$ distribution.

Step 4: Rotate back and act. Rotate the rounded vector $\hat{\mathbf{r}}$ back, $\tilde{\mathbf{u}} = R^\top \hat{\mathbf{r}}$, and rescale by the size to get the reconstruction $\widetilde{\mathbf{x}} = \rho\tilde{\mathbf{u}}$ (or $\hat{\rho}\tilde{\mathbf{u}}$ when the size is rounded). The agent then takes the action $\mathbf{a}(\mathbf{s}) = \mathbb{E}[\mathbf{x} \mid \mathbf{s}]$, the posterior mean given the signal $\mathbf{s} = (\hat{\rho}, \hat{\mathbf{r}})$. Because the posterior mean minimizes quadratic loss given the signal, its loss is no higher than that of the reconstruction $\widetilde{\mathbf{x}}$, so every upper bound we prove for $\widetilde{\mathbf{x}}$ also holds for the agent's action.[13]

Properties of the rotation. The next lemma collects the results from Zandieh et al. (2026) that the strategy relies on. Internet Appendix IA.6 derives each part and shows where it comes from.

Lemma 1 (Building blocks of RTQ)**.** *Let R be a Haar-random $d \times d$ orthogonal matrix, $d \geq 3$, and let $\mathbf{u} \in S^{d-1}$ be any fixed unit vector. Let $\mathbf{r} = R\mathbf{u}$.*

(i) Common distribution. $\mathbf{r}$ is uniformly distributed on S^{d-1}. Each coordinate r_i has density $f_d(r) \propto (1-r^2)^{(d-3)/2}$ on $[-1,1]$, so $(1+r_i)/2 \sim \text{Beta}(\frac{d-1}{2}, \frac{d-1}{2})$. The density depends only on d, not on $\mathbf{u}$, and is close to $N(0,1/d)$ for large d.

(ii) Weak dependence. The coordinates are uncorrelated, $\mathbb{E}[r_i r_j] = 0$ for $i \neq j$, and $\text{Corr}(r_i^2, r_j^2) = -1/(d-1)$.

[13] The reconstruction $\widetilde{\mathbf{x}}$ minimizes squared error under the Beta distribution of the rotated coordinates, not under the prior F. The two coincide only when F matches that distribution. For any other F, the posterior mean does at least as well as $\widetilde{\mathbf{x}}$.

(iii) Upper bound. If each r_i is rounded with the b-bit Lloyd-Max quantizer for f_d and the result is rotated back, $\tilde{\mathbf{u}} = R^\top q(\mathbf{r})$, then $\mathbb{E}_R\|\mathbf{u}-\tilde{\mathbf{u}}\|^2 \le C_T \cdot 4^{-b}$ with $C_T = \pi\sqrt{3}/2 \approx 2.72$, for every $\mathbf{u}$ and every $b \ge 1$.

(iv) Lower bound. Let $\mathbf{u}$ be uniformly distributed on S^{d-1}. For any signal $\mathbf{s}$ with mutual information $I(\mathbf{u};\mathbf{s}) \le bd$ bits, and any reconstruction $\hat{\mathbf{u}}(\mathbf{s})$, the expected squared error is at least $4^{-b}/\gamma_d(b)$, where

$$\gamma_d(b) \;=\; 4^{b(2d-1)/(d-1)^2} \cdot \Big(\frac{d+1}{2}\Big)^{2d/(d-1)^2} \;\ge\; 1, \qquad \gamma_d(b) \to 1 \text{ as } d \to \infty. \tag{8}$$

The bound covers every signal that satisfies the capacity constraint, including every code with at most 2^{bd} values.

Part (i) is the formal version of panel (b) of Figure 2: whatever the input, the rotated coordinates follow one known distribution, so one quantizer designed for it serves every coordinate of every state. Part (ii) says the remaining dependence across coordinates is weak and fades with d; it does not affect the expected squared error of coordinate-by-coordinate rounding, which is a sum over coordinates. Parts (iii) and (iv) together say that rounding coordinate by coordinate after rotation is within the factor $C_T\gamma_d(b)$ of the best that any signal with the same capacity can achieve when the direction is uniform, and therefore of the best possible worst-case error. The correction $\gamma_d(b)$ matters only in low dimensions: it is 1.03 at $d = 500$ and $b = 2$ and 1.007 at $d = 3{,}000$. Parts (i) to (iii) are from Zandieh et al. (2026), who state the lower bound as 4^{-b}. Because the unit sphere in $\mathbb{R}^d$ has dimension $d-1$, a signal can do slightly better than 4^{-b} in any finite dimension, so part (iv) uses a finite-d version, proved in Internet Appendix IA.6.5. The Johnson-Lindenstrauss lemma (Internet Appendix Lemma IA.3) is the dimension-reduction counterpart of part (i): a random projection to k of order $\varepsilon^{-2}\log N$ coordinates preserves all pairwise distances among N points within a factor $1 \pm \varepsilon$. The paper uses only Lemma 1; Appendix Lemma 2 restates part (iii) in the form used in the proofs.

What carries over to other problems. Four features are not specific to attention.

1. *No prior needed.* The strategy is set by R and the bit allocation $(b_0, b_1, \ldots, b_d)$, both chosen before any data arrive. The quantizer is precomputed from f_d, which depends only on d, so the agent does not need to know F. We call it *distribution-free* in this sense and because the upper bound in Proposition 1 holds for any F with finite second moments.

2. *Separability.* After rotation, expected squared error is a sum of d one-coordinate errors, and by additivity (4) the attention budget is a sum of bits. A problem in d interacting variables becomes d one-variable problems linked only through the budget.

3. *A guarantee that does not weaken with dimension.* The constant C_T in Lemma 1 does not depend on d, b, or F, and the correction $\gamma_d(b)$ shrinks to one as d grows. Proposition 1 shows that a general loss matrix W adds only its condition number $\chi(W)$.

4. *Low cost.* A randomized Hadamard transform (Ailon and Chazelle, 2009) applies the rotation in $O(d \log d)$ operations, and quantization is a table lookup.

Any problem in which an agent must summarize a high-dimensional state with limited capacity, such as a firm tracking costs across many products or a regulator monitoring many banks, can use the same three steps: rotate, allocate bits by stakes, and round with a precomputed rule.

Interpretation. The random rotation is a modeling device, not a literal action. Each rotated coordinate is a basket of all original variables with weights fixed in advance, much as decision makers read aggregates such as indexes, and the rotated coordinates behave like synthetic factors of similar variance. Just as competitive equilibrium describes attainable allocations without requiring agents to solve the Walrasian auctioneer's problem, RTQ describes attainable attention strategies without requiring agents to compute a Haar-random rotation. Modern AI portfolio models such as Cong et al. (2026) implement broadly similar cross-asset attention in practice, so the benchmark is implementable rather than purely abstract.

A technical note: the Beta distribution in Lemma 1 describes a fixed direction $\mathbf{u}$ over random draws of R. Once R is drawn and held fixed, the rotated coordinates of $R\mathbf{x}$ for $\mathbf{x} \sim F$ reflect the structure of F, and the strategy is a fixed rule that maps each state to a signal and each signal to an action. The quantizer is built for the Beta distribution because the guarantee is a worst case over directions, and every direction has Beta-distributed rotated coordinates over draws of R. Proposition 3 shows how to allocate bits across the rotated coordinates for general W.

4.2 RTQ Guarantee Against the Spherical Benchmark

We compare RTQ with the lowest worst-case loss that any signal with the same capacity can achieve, the minimax benchmark. This is the right benchmark when the prior F is not known. The result concerns the direction of the state. Its size $\rho = \|\mathbf{x}\|$ is public side information, available under every strategy, and the attention budget is spent on the direction. This split between size and direction is standard for TurboQuant (Zandieh et al., 2026). Observing the realized size exactly is a strong assumption, since it is more than knowing the average volatility of returns. We make it for every strategy alike, so that none gains from information outside its budget, and Appendix A.2 extends the result to the case in which ρ is also quantized.

The model limits mutual information, while RTQ uses a fixed number of bits. The two fit together. A signal with at most 2^κ values always satisfies the constraint, so RTQ is feasible. Our

lower bound is stated for the mutual-information constraint itself, so it covers every signal the agent could choose. Let $\mathscr{L}^{RTQ}(\kappa)$ denote expected loss under RTQ with optimal bit allocation.

Proposition 1 (RTQ guarantee against the minimax benchmark)**.** *Write* $\mathbf{x} = \rho\mathbf{u}$ *with unit direction* $\mathbf{u} \in S^{d-1}$ *and size* $\rho = \|\mathbf{x}\|$*, and suppose* ρ *is public side information, so the attention budget applies only to the direction. RTQ encodes the direction with* $b = \kappa/d$ *bits per coordinate and reconstructs* $\hat{\mathbf{x}} = \rho\tilde{\mathbf{u}}$*, where* $\tilde{\mathbf{u}}$ *is the reconstruction of Lemma 1(iii). For any prior* F *with finite second moments, any positive definite* W*, any* $d \geq 3$*, and any* $\kappa \geq 1$*,*

$$\mathscr{L}^{RTQ}(\kappa) \;\leq\; \lambda_{\max}(W) \cdot \mathbb{E}[\rho^2] \cdot C_T \cdot 4^{-b}, \tag{9}$$

with $C_T = \pi\sqrt{3}/2 \approx 2.72$*. Define the reference level*

$$\mathscr{L}^{\text{sphere}}_{\text{lb}}(\kappa) \;\equiv\; \lambda_{\min}(W) \cdot \mathbb{E}[\rho^2] \cdot 4^{-b}. \tag{10}$$

There is a prior with the same $\mathbb{E}[\rho^2]$*, in which the direction is uniform on the sphere and the size is constant, under which every signal with* $I(\mathbf{x};\mathbf{s} \mid \rho, R) \leq \kappa$ *has loss at least* $\mathscr{L}^{\text{sphere}}_{\text{lb}}(\kappa)/\gamma_d(b)$*, with* $\gamma_d(b)$ *as in* (8)*. Therefore RTQ's loss under any prior is at most*

$$C_T \cdot \chi(W) \cdot \gamma_d(b) \tag{11}$$

times this lower bound, where $\chi(W) = \lambda_{\max}(W)/\lambda_{\min}(W)$ *is the condition number of* W*, and* $\mathscr{L}^{RTQ}(\kappa) \leq C_T\,\chi(W)\,\mathscr{L}^{\text{sphere}}_{\text{lb}}(\kappa)$*. Since* $\gamma_d(b) \to 1$ *as* $d \to \infty$*, the factor approaches* $C_T\chi(W)$*.*

The proposition says that, with many variables, RTQ's loss is at most about $2.72\,\chi(W)$ times the lowest worst-case loss of any signal with the same capacity. The factor does not depend on the prior F, and its dependence on the number of variables works in RTQ's favor: the correction $\gamma_d(b)$ is 1.03 at $d = 500$ and $b = 2$ and falls toward one as d grows. When all errors are weighted equally ($W = I$), the factor is $C_T\gamma_d(b)$, which is the TurboQuant guarantee of Zandieh et al. (2026) with the finite-d lower bound. When some directions of W matter much more than others, $\chi(W)$ widens the gap; this is the cost of using one rule for every direction when the stakes differ across directions. In a portfolio of 500 stocks, W comes from the return covariance, and after shrinkage $\chi(W)$ is moderate, so RTQ is close to the best worst-case loss across all 500 assets.[14]

The benchmark is a worst case over directions, not the best possible signal for a particular prior. When the prior is known, a signal designed for it can do better; with a normal prior, for example, the classical solution allocates capacity across the principal directions by water-filling

[14]We state the result with the size ρ observed because the TurboQuant guarantee of Zandieh et al. (2026) is stated for unit vectors with the length stored separately.

and is optimal. The value of Proposition 1 is that RTQ achieves its guarantee without knowing F, and Section 4.6 shows how much a signal designed for the wrong prior can lose. The loss concentration ratio of Section 4.5 identifies settings in which the prior has enough structure that a tailored signal may beat RTQ.

Proof sketch. The full proof is in Appendix A.2. With ρ observed, the error is $\mathbf{x}-\hat{\mathbf{x}}=\rho(\mathbf{u}-\tilde{\mathbf{u}})$, and since $\mathbf{e}^\top W\mathbf{e} \le \lambda_{\max}(W)\|\mathbf{e}\|^2$, the loss is at most $\lambda_{\max}(W)\rho^2\|\mathbf{u}-\tilde{\mathbf{u}}\|^2$. For every direction, Lemma 1(iii) bounds the expected squared error by $C_T 4^{-b}$. Taking expectations over ρ and $\mathbf{u}$ gives (9). For the lower bound, Lemma 1(iv) gives an expected squared error of at least $4^{-b}/\gamma_d(b)$ when the direction is uniform, and $\mathbf{e}^\top W\mathbf{e} \ge \lambda_{\min}(W)\|\mathbf{e}\|^2$ converts it to the weighted loss. Dividing the two gives (11). □

The factor $\lambda_{\max}(W)$ in the proof is conservative for a typical draw of R. Lemma IA.1 in the Internet Appendix shows that the diagonal entries of $RWR^\top$ cluster around $\mathrm{tr}(W)/d$ when $\chi(W)$ is moderate, so a typical rotation spreads the stakes evenly across coordinates. This result helps interpret the allocation below but is not used in the worst-case bound.

4.3 Cost of Selective Attention: The Unmonitored Loss Floor

Proposition 1 shows what RTQ achieves. The next result shows the best that focusing on a few directions can achieve. Focusing puts all capacity on the watched directions and leaves the rest at their prior, which creates a floor on the loss.

Consider a selective-k strategy: the agent acquires signals about k chosen directions and uses the prior mean for all other directions, without inferring anything about them from the watched ones. This is the usual applied reading of "watch a few key indicators and ignore the rest," and we call it hard exclusion. We work in a basis in which W and the prior covariance Σ are both diagonal, the case used in nearly all applied work on rational inattention.

Proposition 2 (Selective attention with hard exclusion: unmonitored loss floor)**.** *Fix $k<d$ and consider the selective-k strategy with hard exclusion in a basis where $W=\mathrm{diag}(\omega_1,\dots,\omega_d)$ and $\mathrm{Cov}(x_j,x_\ell)=0$ for $j\neq\ell$, with $\omega_1\ge\cdots\ge\omega_d>0$ and prior variance σ_j^2 along direction j. Let $S\subset\{1,\dots,d\}$ be the chosen monitored set with $|S|=k$. The expected loss satisfies*

$$\mathscr{L}^S \;\ge\; \sum_{j\notin S}\omega_j\sigma_j^2. \tag{12}$$

The best set S consists of the k directions with the largest welfare exposure $\omega_j\sigma_j^2$, which are not necessarily those with the largest ω_j. The right-hand side does not depend on capacity: however much attention goes to the watched directions, the unwatched ones leave a loss floor.

Proof sketch. Under hard exclusion, the agent's action on $j \notin S$ is the prior mean, so posterior error on direction j has variance σ_j^2 and contributes $\omega_j \sigma_j^2$ to the weighted loss. Monitored directions contribute non-negative loss. Summing over the unmonitored set gives (12). To select S optimally, minimize $\sum_{j \notin S} \omega_j \sigma_j^2$ subject to $|S| = k$, which selects the k largest values of $\omega_j \sigma_j^2$. Full argument in Appendix A.3. □

The diagonal basis matters. If the prior covariance is not diagonal in the eigenbasis of W, signals about watched directions also carry information about unwatched ones, and a selective strategy that uses this information can achieve a lower loss than (12). The same can happen when the directions are uncorrelated but not independent, which is possible outside the normal case. The bound applies to hard exclusion, which is the form of selective attention used in applied work.

RTQ avoids the floor by spreading attention across all d directions. Its error in each direction falls at the rate $4^{-\kappa/d}$, and no direction is left at its prior. As capacity grows with k fixed, RTQ's loss goes to zero while the selective floor stays, so the ratio of the two losses grows without limit. With $k = 10$ and $d = 500$, RTQ spends one fiftieth as many bits on each direction as the selective strategy spends on each watched one, but it covers fifty times as many directions, while the selective strategy leaves 490 directions unwatched.

Internet Appendix IA.1.7 computes the bound for a diversified portfolio, multi-product pricing, and monetary policy with several instruments, using plausible condition numbers. Realized losses are far below this worst case.[15]

The worst-case RTQ bound (9) does not depend on d. Two forces make RTQ look better still as d rises. First, by Lemma IA.1, the stakes in the rotated coordinates become more even, so a typical rotation comes closer to the lower bound. Second, the loss floor of hard exclusion grows with the number of unwatched directions, while RTQ's loss does not. In complex environments, the common advice to concentrate attention can therefore be costly.[16]

4.4 Optimal Bit Allocation Across Dimensions

How should the agent split her capacity across the rotated coordinates? Let W have eigenvalues $\omega_1 \geq \cdots \geq \omega_d > 0$ with eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_d$, and define the *rotated loss weights* as the diagonal of $RWR^\top$:

$$w_i = (RWR^\top)_{ii} = \mathbf{e}_i^\top RWR^\top \mathbf{e}_i. \tag{13}$$

The weight w_i measures how costly an error in rotated coordinate i is.

[15] For example, in the S&P 500 calibration RTQ has a welfare cost of 0.5 basis points per year against 198 for VNV at the same budget (Table IA.1 in Internet Appendix IA.1).

[16] Internet Appendix IA.1.5 develops this intuition in more detail.

Proposition 3 (High-resolution water-filling approximation)**.** *Under the high-resolution Lloyd-Max MSE model* $\mathrm{MSE}_i(b_i) = \alpha_d \cdot 2^{-2b_i}$ *for the* $\mathrm{Beta}(\frac{d-1}{2}, \frac{d-1}{2})$ *distribution at* b_i *bits, the continuous bit allocation* $(b_1^*, \ldots, b_d^*)$ *that minimizes expected loss under RTQ, subject to* $\sum_{i=1}^d b_i \leq \kappa - b_0$*, satisfies*

$$b_i^* \;=\; \left[\frac{1}{2}\log_2\!\left(\frac{w_i}{\nu}\right)\right]^+, \tag{14}$$

where $[z]^+ = \max(z, 0)$ *and* $\nu > 0$ *is chosen so* $\sum_{i=1}^d b_i^* = \kappa - b_0$*. For finite and integer bit widths, the exact RTQ allocation solves the discrete separable problem*

$$\min_{b_i \in \mathbb{Z}_{\geq 0}} \sum_{i=1}^d w_i \cdot \mathrm{MSE}_i(b_i) \quad s.t. \quad \sum_{i=1}^d b_i \leq \kappa - b_0, \tag{15}$$

using a precomputed Lloyd-Max distortion table. The water-filling formula (14) *is the closed-form high-resolution limit of* (15) *as* $b_i \to \infty$*.*

The allocation is the water-filling rule of information theory. Bits go only to coordinates whose weight w_i exceeds a threshold ν, and among those, coordinates with higher stakes get more bits, in proportion to $\log_2 w_i$. The threshold adjusts so that the bits add up to the budget, and it can be read as the shadow price of capacity, the welfare value of one more bit.

The main implication is that, under the approximation in the proposition, the stakes of a decision, not the volatility of the data, determine how attention is allocated across the composite signals. In the one-variable setting of Sims (2003) there is nothing to allocate, and a common intuition from low-dimensional extensions is that an agent should attend to the most volatile variables. The proposition treats every rotated coordinate as having the same distortion function, which holds on average over rotations (Lemma 1) but only approximately for one fixed rotation and a particular prior. Under this approximation the allocation depends only on the weights w_i. For a portfolio manager, this means giving more attention to composites whose errors add most to portfolio risk, which depends on W, rather than to those that load on the most volatile stocks. Since each composite mixes all stocks, the result concerns composites, not individual stocks.

When $W = I$, all w_i are equal and the allocation is uniform, $b_i^* = (\kappa - b_0)/d$, which is the case covered by Zandieh et al. (2026). For a given d, the allocation becomes less even as $\chi(W)$ rises. In high dimensions, Lemma IA.1 pulls the w_i toward $\mathrm{tr}(W)/d$, so the allocation across rotated coordinates is close to uniform even when W is far from the identity.

Proof sketch. Under RTQ, expected loss decomposes in the high-resolution approximation as $\mathscr{L}^{RTQ} \approx \mathbb{E}[\rho^2] \sum_{i=1}^d w_i \cdot \mathrm{MSE}_i(b_i) + \mathscr{L}^{\mathrm{norm}}(b_0)$, where the off-diagonal cross-terms $\mathbb{E}[\delta_i \delta_j]$ for $i \neq j$ are assumed negligible (mean-zero scalar Lloyd-Max errors with cross-covariances of order $o(4^{-b})$ relative to diagonal MSE; see Appendix A.4). For the Beta distribution, $\mathrm{MSE}_i(b_i)$ is convex and

decreasing in b_i, so constrained minimization of this separable objective with a budget yields water-filling via KKT. The threshold ν is the Lagrange multiplier on the capacity constraint. The water-filling formula is the closed-form asymptotic result; the discrete integer-allocation formulation (15) is its finite-resolution counterpart and recovers it in the high-resolution limit. □

After rotation, total loss is approximately a sum of one term per coordinate and the budget is a sum of bits, so the problem is a standard allocation of a fixed budget across d uses with diminishing returns. The difference from textbook water-filling is that the w_i are the diagonal entries of $RWR^\top$ rather than the eigenvalues of W, because the rotation mixes the eigenvalues together.

4.5 When to Focus and When to Spread

Propositions 1 and 2 together give a criterion for when to focus and when to spread. It is based on the *loss concentration ratio*, the share of what is at stake that sits in the top k directions.

Definition 1 (Loss Concentration Ratio)**.** *Let W have eigenvalues $\omega_1, \ldots, \omega_d > 0$ with prior variances $\sigma_1^2, \ldots, \sigma_d^2$ along the corresponding eigenvectors. Reindex the eigenvectors so that the welfare exposures $\omega_j \sigma_j^2$ are sorted in decreasing order: $\omega_1\sigma_1^2 \geq \omega_2\sigma_2^2 \geq \cdots \geq \omega_d\sigma_d^2$. The loss concentration ratio at rank k is the share of total welfare-weighted prior loss captured by the top-k welfare exposures,*

$$\rho(k) \;=\; \frac{\sum_{j=1}^{k} \omega_j \sigma_j^2}{\sum_{j=1}^{d} \omega_j \sigma_j^2} \;=\; \max_{|S|=k} \frac{\sum_{j\in S} \omega_j \sigma_j^2}{\sum_{j=1}^{d} \omega_j \sigma_j^2}. \tag{16}$$

This ordering is the one Proposition 2 uses to choose the watched directions. When σ_j^2 is the same across directions, or proportional to ω_j as in mean-variance portfolio choice with $W = \gamma\Sigma$, it is the same as ordering by ω_j.

When $\rho(k)$ is close to one, the problem is effectively k-dimensional: almost everything at stake lies in k directions. When $\rho(k)$ is well below one, the other $d-k$ directions carry a large part of the loss, and ignoring them is costly.

Corollary 1 (Focus-vs-Spread Threshold)**.** *Under the exact-norm assumption of Proposition 1, RTQ achieves lower expected loss than the selective-k hard-exclusion strategy whenever*

$$\rho(k) \;<\; 1 - C_T \cdot \chi(W) \cdot \frac{\mathscr{L}_{\mathrm{lb}}^{\mathrm{sphere}}(\kappa)}{\sum_{j=1}^{d} \omega_j \sigma_j^2}, \tag{17}$$

where $C_T = \pi\sqrt{3}/2$. The condition can hold only if $C_T \cdot \chi(W) \cdot \mathscr{L}_{\mathrm{lb}}^{\mathrm{sphere}}(\kappa) < \sum_{j=1}^{d} \omega_j \sigma_j^2$, that is, when $\chi(W)$ is moderate and capacity is not too small. With norm quantization at b_0 bits, an

additional $L_\rho(b_0)$ term enters the right-hand side as in Proposition 1; under bounded support and high-rate conditions this term vanishes as $b_0 \to \infty$.

The condition is sufficient but not necessary, because it compares an upper bound on RTQ's loss with a lower bound on the selective strategy's loss; RTQ can do better even when it fails.[17] At the other extreme, when W has rank one, as with a single policy instrument, $\rho(1) = 1$ and focusing on that one direction is optimal.

Proof. Combining the lower bound from Proposition 2 with the exact-norm RTQ upper bound from Proposition 1: $\mathscr{L}^S \geq (1-\rho(k))\sum_{j=1}^{d} \omega_j \sigma_j^2$ and $\mathscr{L}^{RTQ} \leq C_T \cdot \chi(W) \cdot \mathscr{L}_{\text{lb}}^{\text{sphere}}(\kappa)$. RTQ wins whenever the upper bound on $\mathscr{L}^{RTQ}$ is below the lower bound on $\mathscr{L}^S$, which gives (17). With norm quantization at b_0 bits, the argument goes through with an extra $L_\rho(b_0)$ term on the right-hand side. □

The ratio $\rho(k)$ can be computed from the eigenvalues of W and the prior variances, so a researcher can plot it against k to see how many directions matter. If $\rho(k)$ is close to one for small k, the problem is effectively low-dimensional and focusing works well. If $\rho(k)$ stays well below one even for large k, the problem is truly high-dimensional and spreading attention is better.

4.6 What the Distribution-Free Guarantee Buys: A Simulation

Proposition 1 holds for any prior with finite second moments. A signal designed for a normal prior has no such protection. To see how much this matters, we hold the mean and covariance of the state fixed, at zero and the identity, and vary only the shape of the prior. A designer who assumes a normal prior builds the same signal in every case: b bits per coordinate with the Lloyd-Max quantizer for a standard normal variable. We compare it with RTQ at the same budget for four priors: normal; Student t with three degrees of freedom; a mixture of a calm regime (90 percent of draws) and a volatile regime (10 percent); and a sparse prior in which three randomly chosen coordinates receive large shocks and the rest are zero. Each run uses 4,000 draws, with $d \in \{64, 256\}$ and $b \in \{1, 2, 3\}$.

Figure 3 shows the result at $d = 256$ and $b = 2$. Each group of bars is one prior, and the height of a bar is the expected squared error as a share of $\mathbb{E}\|\mathbf{x}\|^2$, so zero is perfect recovery and one is the error of receiving no signal. The light gray bars are the design built for a normal prior, and the black bars are RTQ. RTQ's error is about 0.117 under all four priors, below its bound $C_T \cdot 4^{-2} = 0.170$

[17] For a portfolio with $d = 500$ and $\chi(W) = 3$, a typical value after shrinkage, suppose the reference level $\mathscr{L}_{\text{lb}}^{\text{sphere}}$ is 10 percent of the prior loss. The condition is then $\rho(k) < 1 - 2.72 \times 3 \times 0.1 \approx 0.18$: it guarantees that RTQ wins only when the top k directions carry less than 18 percent of the loss, which is a demanding test. In practice RTQ does much better than the condition suggests. In our S&P 500 sample, the loss of the selective strategy is 20 to 400 times that of RTQ even where the condition fails (Internet Appendix Figure IA.3).

(the dotted line). The normal design matches RTQ when the prior really is normal (0.118), but its error rises to 0.28 under the t prior, 0.31 under the two-regime prior, and 0.97 under the sparse prior, which is almost as bad as receiving nothing. A signal designed for the wrong prior can thus fail badly even when the mean and covariance are right, while RTQ's error is the same under every prior. The same pattern holds at $d = 64$ and at one and three bits per coordinate (Internet Appendix Table IA.4).

The simulation also separates the two parts of RTQ. The dark gray bars use the normal design but also store the length of the state, as RTQ does. This removes most of the loss under the fat-tailed and two-regime priors, which differ from the normal mainly in how the overall scale varies. It does not help under the sparse prior (0.95), where a few coordinates carry the whole state. Only the rotation fixes that, by spreading those few large values across all coordinates before rounding.

4.7 Extension to market equilibrium with heterogeneous attention

Internet Appendix IA.5 places the RTQ agent in a standard noisy rational-expectations equilibrium with normal shocks and CARA preferences. The market has three groups: well-informed investors (mass α) who observe a noisy signal of the state, investors who use RTQ with a budget of B bits (mass $1-\alpha$), and noise traders. To keep the equilibrium linear, this extension replaces the RTQ rounding error with normal noise of the same variance, so it describes an approximation to RTQ rather than RTQ itself. Internet Appendix Theorem IA.3 shows that the equilibrium price is unique and linear, with closed-form coefficients, and that the welfare loss of an RTQ investor has two parts: the noise from her own limited attention and the information she can learn from the price. Corollary IA.2 extends the result to many investor types with different budgets and risk aversion.

The equilibrium adds an asset-pricing prediction. As the share of well-informed investors grows, prices reveal more, and the cost of limited attention falls. This bears on the rise of AI-based investing: if more investors use high-capacity tools such as Transformer models, prices become more informative and the advantage of the best-informed investors over everyone else shrinks, consistent with the small alphas of active managers in Fama and French (2010).[18]

Together, the results describe rational inattention with many variables. RTQ's loss is within about $C_T\chi(W)$ of the lowest worst-case loss of any signal with the same capacity, whatever the prior; where attention goes is set by the stakes in the loss, not by the volatility of the data; and focusing on a few directions leaves a loss floor that extra capacity cannot remove. None of these results assumes a normal or other parametric prior. Section 5 takes them to the full CRSP universe

[18] Internet Appendix IA.5.3 also extends the framework to several periods, with an AR(1) state and an agent who can reallocate attention over time. The resulting dynamics show the three features that Steiner et al. (2017) find for dynamic rational inattention: inertia, slow response to structural breaks, and pooling of the attention budget across periods (Proposition IA.8, Lemma IA.2, and Corollary IA.3).

and the LSEG 13F holdings data.[19]

5 Empirical Tests

This section takes the framework to US equities, where thousands of stocks make the curse of dimensionality severe. We use the CRSP monthly panel to measure how much welfare a broad allocation of attention gains over a concentrated one, and how that gain depends on loss concentration $\rho(1)$. Descriptive evidence on institutional investors from LSEG 13F holdings is summarized at the end of the section and reported in Internet Appendix IA.7.4.[20]

The theory makes two predictions for the CRSP data. First, spreading attention across all sources of risk should do better than spending the same budget on the largest source of risk alone, and the gain should hold up out of sample, where errors in the estimated covariance work against any allocation tuned to it. Second, the gain should be large when $\rho(1)$ is low, so that no single direction carries most of the loss, and small when $\rho(1)$ is high, so that concentrating gives up little. For managers, the theory does not say that broad portfolios beat focused ones on average. It says that which kind does better depends on conditions, and names $\rho(1)$ as the one that should matter.

Our analysis sample uses common stocks on NYSE, AMEX, and NASDAQ in the merged CRSP and Compustat monthly file with at least 60 months of trailing return history. The rolling exercise uses 221 quarter-ends from 1971Q1 to 2026Q1. The eligible universe has 34 stocks at the first quarter-end, passes 500 stocks in 1974, and averages 2,422 stocks (maximum 3,381). The annual series of $\rho(1)$ and d_{eff} use 55 December endpoints from 1971 to 2025. We report welfare costs in basis points of certainty-equivalent return per year, $(L/P)\cdot\mu_p\cdot 10^4$, where L is the loss under an allocation rule, P is the loss with no signal, and $\mu_p = 0.06$ is the equity premium. Receiving no information therefore costs 600 basis points.

The broad strategy in most of this section is the *RTQ allocation*: the rule of Proposition 3 for dividing κ bits, applied to the eigendirections of the estimated covariance. It gives more bits to directions where a mistake costs more, as measured by λ_j^2, and none to directions that matter too little, and the loss in direction j is $\lambda_j^2 \min(C_T 4^{-b_j}, 1)$. The RTQ allocation uses the estimated covariance and no random rotation. We write RTQ without qualification only for the full method

[19] Internet Appendix IA.2 studies an agent who tracks many separate targets with a shared budget, and Internet Appendix IA.3 tests its predictions in the speed-dating experiment of Fisman et al. (2006). Internet Appendix IA.1 calibrates the results on a proxy for the S&P 500 and compares RTQ directly with the strategy of Van Nieuwerburgh and Veldkamp (2010).

[20] A test outside finance, on the Fisman et al. (2006) speed-dating experiment, is summarized at the end of this section and reported in full in Internet Appendix IA.3. A companion calibration on a 488-stock proxy of the top 500 CRSP commons by market cap is in Internet Appendix IA.1; Subsection IA.1.2 computes the loss concentration ratio for that covariance ($\rho(1) = 0.80$, $\rho(5) = 0.95$) and plots the spectrum diagnostics (Figures IA.1 and IA.2). Panel C of Table 1 repeats the rolling comparison on the 500 largest stocks at each date, in and out of sample.

with a random rotation, which does not use the covariance and is the one covered by the worst-case guarantee of Proposition 1. Dividing bits across eigendirections in this way is also what the classical solution prescribes under a normal prior, so the comparison with the RTQ allocation measures the value of spreading attention rather than concentrating it. It does not measure the separate value of the random rotation, which Section 4.6 isolates by simulation and Table 2 below examines in the return data. In the figures, the label RTQ stands for the RTQ allocation.

The concentrated benchmark is a single-direction benchmark: it spends the same κ bits on the top eigendirection and leaves the other directions at the prior. We label it fair VNV in the tables and figures, because it is motivated by the specialization result of Van Nieuwerburgh and Veldkamp (2010) and gets the same budget as the RTQ allocation. It is not their model. In their paper specialization comes from choosing jointly what to learn and what to hold, which our problem with a fixed loss matrix does not include, and their results depend on preferences and information costs. Comparing the RTQ allocation with receiving no information instead would overstate the case for spreading, since any information beats none.[21]

Loss concentration and effective dimension vary substantially over the sample. For each year 1971–2025, we estimate the $N \times N$ sample covariance on 60 trailing monthly returns of the eligible universe, apply Ledoit-Wolf (Ledoit and Wolf, 2004) shrinkage, and compute $\rho(k) = \sum_{j=1}^{k} \lambda_j^2 / \sum_{j=1}^{N} \lambda_j^2$ and $d_{\text{eff}} = (\text{tr}\Sigma)^2/\text{tr}(\Sigma^2)$. Here $\rho(1)$ is the share of total risk, weighted by its welfare cost, that sits in the largest direction, and d_{eff} counts roughly how many directions matter.

Figure 4 shows both measures by year. The top panel plots $\rho(1)$ (solid line) and $\rho(5)$ (dashed line), and the dotted line marks one half. Across the 55 years $\rho(1)$ ranges from 0.260 to 0.936, with a mean of 0.661. It is high, around 0.8 to 0.9, in the 1970s and 1980s, when one market-wide direction dominated, and falls below one half in the late 1990s and early 2000s and again from about 2019 to 2024. The bottom panel plots d_{eff} (solid line), which ranges from 9.6 to 587.6 (mean 146.7), against the number of stocks (dashed line). The US market is thus neither dominated by one source of risk nor spread evenly over thousands; it spends most of the sample in the middle range, where focusing and spreading call for different allocations, which makes the comparison informative.

We then run the rolling comparison. At each quarter-end, we estimate the covariance on the prior 60 months and choose the RTQ allocation and the fair VNV allocation from it. In sample, we evaluate the losses with the same covariance. Out of sample, we evaluate them in the same ex-ante eigendirections with the covariance of the next 24 months, so that each allocation is judged on risk it did not see. Each loss is in basis points of certainty-equivalent return per year, where no

[21] Earlier versions of this paper computed the loss of the RTQ allocation as $C_T^2 \lambda_j 4^{-b_j}$ against a prior of λ_j^2. The two terms have different units, so the welfare gain changed with the units of returns, and the negative gains in some early windows came from that mismatch. Table IA.5 in the Internet Appendix compares the two formulas.

information costs 600 and full information nothing. The scale is a normalization, so each number is a model-implied welfare gain: dividing by 600 gives it as a share of the no-information loss.

Table 1 reports the gain from spreading attention, measured as the welfare cost of fair VNV minus that of the RTQ allocation, so a positive number means that spreading does better. Each row is an attention budget of one, two, or three bits per stock ($\kappa = N$, $2N$, $3N$). The first column is the median gain in sample, the next four describe the gain out of sample, and the last column compares the RTQ allocation with receiving no information. The key number is the out-of-sample median in Panel A at two bits per stock: 206 basis points per year, about one third of the loss from having no information, with a mean of 247 and a Newey-West t-statistic of 4.84. This means that an investor who spreads a budget of two bits per stock across all directions is better off, by the equivalent of about 2 percentage points of return per year, than one who spends the same budget on the largest source of risk, even when the allocation is judged on risk it did not see. The gain is smaller with one bit per stock (median 86) and larger with three (median 239). It is smaller when the covariance is estimated without shrinkage (Panel B, median 39 at two bits per stock) and among the 500 largest stocks (Panel C, median 42), where $\rho(1)$ is higher. In every specification, the RTQ allocation has the lower loss in every quarter, in sample and out of sample. Spreading attention thus beats concentrating it in the US market throughout the sample, and by more when the investor covers more stocks. Internet Appendix Table IA.6 repeats the comparison with 36- and 120-month estimation windows and with three other covariance estimators. The RTQ allocation has the lower loss in every quarter in every case, and the out-of-sample median gain at two bits per stock is 39 to 206 basis points; the window matters little (159 to 206).

Figure 5 shows the gain quarter by quarter. The black line is the gain over fair VNV and the gray line, near the top, is the gain over receiving no information. Panel (a) is in sample and panel (b) is out of sample. The black line in panel (b) moves much more than in panel (a), because the next 24 months can differ a lot from the previous 60, and in a few quarters it comes close to zero, but it never falls below zero. So the advantage of spreading attention is not driven by a few periods: it holds in every quarter from 1971 to 2026, though its size varies a great deal.

How large the gain is should depend on $\rho(1)$. When one direction carries most of the risk, concentrating on it loses little; when risk is spread out, concentrating leaves much of it unwatched. In sample this relation is close to an identity. With two or more bits per stock, the RTQ allocation removes almost all of the loss, while fair VNV removes only the top direction's share, so the in-sample gain is approximately $600 \cdot (1 - \rho(1))$ basis points (correlation 0.999 across quarters).

Figure 6 shows the gain against ex-ante $\rho(1)$, with one point per quarter. The left panel is in sample, where the points lie on a line by construction. The right panel is out of sample, which is the informative test. The points are widely scattered, but the fitted line (dashed) still slopes down: the slope is -369 basis points per unit of $\rho(1)$ (Newey-West $t = -1.79$, $R^2 = 0.13$). Sorting the

213 out-of-sample quarters into terciles of ex-ante $\rho(1)$ gives the same message: the mean gain is 338 basis points in the low tercile, 237 in the middle, and 166 in the high tercile. Spreading attention thus pays most when risk is spread across many directions, as the theory predicts, but out of sample the relation is only marginally significant. It is slightly stronger at three bits per stock (-414, $t = -1.91$), weaker at one bit per stock ($t = -0.97$), and statistically insignificant among the 500 largest stocks, where $\rho(1)$ varies less. It is similar with 36- and 120-month windows ($t = -1.68$ and -1.77) but absent with the sample covariance, constant-correlation shrinkage, and POET, whose noisier estimates of the top eigenvalue weaken the link between ex-ante $\rho(1)$ and later losses. The same pattern holds across market-cap deciles at the end of 2024: the gain over fair VNV is 429 basis points in the decile with the lowest $\rho(1)$ (0.16) and 57 basis points in the decile with the highest (0.91) (Table IA.8 in the Internet Appendix).

The welfare cost of each rule also depends on how many stocks the investor covers. Figure 7 shows the cost of each rule as the investor covers more stocks, adding them in order of market capitalization, at the end of 1990 and the end of 2024, with two bits per stock. Both axes are on a log scale. The black circles are the RTQ allocation, the gray squares are fair VNV, and the dashed line at the top is the cost of receiving no information (600 basis points). The cost of the RTQ allocation stays below one basis point per year at every universe size, while the fair VNV cost ranges from 36 to 278 basis points and rises as smaller stocks enter and $\rho(1)$ falls. Covering more stocks thus makes concentrated attention more costly but leaves spread attention almost unaffected, which is the sense in which spreading attention escapes the curse of dimensionality. In the same calculations, the loss of the RTQ allocation is far below the distribution-free bound of Proposition 1: the median ratio of the two is below one percent (Table IA.7). The bound is a worst case over all priors, and an allocation that uses the actual return covariance does much better.

The comparison so far uses the RTQ allocation, which relies on the estimated covariance to decide where attention goes. Table 2 shows how RTQ itself performs, with a random rotation and without the covariance, and compares it with strategies that use the estimated covariance. Nothing here comes from a distortion formula: each strategy encodes every monthly return vector with a fixed number of bits and decodes it, and the loss is the realized error on the returns of the next 24 months, on the same scale as before. Panel A uses the 500 largest stocks and Panel B all eligible stocks. The first two rows are RTQ, with the same bits on every rotated coordinate and with bits allocated by Proposition 3. The next three rows spread attention across the principal components of the estimated covariance: two round each component with a quantizer built for a normal prior, without and with the length of the return vector, and the third is the classical normal-prior solution, which is a benchmark and not a code that can be run. The last four rows put all bits on the top 1, 5, 20, or 100 principal components. For each budget, the columns give the median and mean loss across quarters and the share of quarters in which RTQ with equal bits has the lower loss.

The table has three messages. First, RTQ's loss is nearly the same in every quarter. Among the 500 largest stocks at two bits per stock, the median is 10.2 basis points and the mean 11.0, and the loss falls by a factor of about eight with each added bit (83.8, 10.2, and 1.4). Second, against focused strategies RTQ needs enough capacity. The loss from watching only the top direction is 51.9 basis points at every budget, because extra bits cannot reduce the loss in directions that are not watched (Proposition 2). With one bit per stock RTQ does worse than this strategy among the 500 largest stocks, where it has the lower loss in only 24 percent of quarters. With two bits it has the lower loss in every quarter, and with three bits it also beats the top 100 components in 75 percent of quarters (94 percent among all stocks). Third, strategies that spread attention with the help of the estimated covariance have a lower median loss than RTQ at every budget: 1.0 against 10.2 basis points at two bits per stock. US stock returns have a strong factor structure, the estimated covariance captures much of it, and RTQ with equal bits does not use it. This is the price of robustness. The benefit appears in the quarters where the estimated covariance describes the next two years poorly. The principal-component design without the length has a mean loss far above its median (7.0 against 1.0), and among all stocks its mean loss is above RTQ's at two and three bits per stock (31.3 against 12.4, and 28.3 against 2.0). Giving that design the length of the return vector, which RTQ has, lowers its mean loss among all stocks to 8.4 at two bits, below RTQ's, and to 5.7 at three bits, still above RTQ's 2.0. Its twelve worst quarters are all windows formed between September 2006 and June 2009, whose evaluation periods cover the financial crisis; its largest loss is 128 basis points at two bits, against 25 for RTQ.

These numbers should be read with one fact in mind: RTQ with equal bits does not use the covariance, and every other strategy in Table 2 does. The focused strategies need it as well, because watching the top direction requires knowing which direction that is. Taking the covariance as known is a strong assumption when there are many variables. With about 2,500 stocks, the covariance has about 3.1 million distinct entries, and 60 months of returns give 150,000 observations, so the estimate rests heavily on shrinkage toward a simple target. The covariance also changes over time, as the crisis quarters show. For an agent with limited attention, learning the covariance of everything she follows is itself costly, and the classical solution takes that knowledge as free. In some settings, such as newly listed firms, private assets, or new products, there is no history to estimate it from. US equities, with decades of returns and a strong factor structure, are therefore a setting that favors methods built on the covariance. RTQ's loss of 10 basis points at two bits per stock is what an investor obtains without that knowledge, and neither the procedure nor its guarantee changes when the covariance is unknown. So the US evidence supports spreading attention over concentrating it, and that an investor who trusts her covariance estimate should use it to decide how to spread. RTQ is the choice for an investor who does not have a covariance she can trust, which is common when the number of variables is large: it beats focused strategies once

the budget reaches about two bits per stock, and its loss varies little across quarters.

The random rotation and the covariance can also be used together. Table 2 includes a version we call grouped RTQ. It sorts the principal components of the estimated covariance into a few groups with similar stakes, divides the bits across groups, and inside each group mixes the components with a random rotation and rounds every rotated coordinate alike. It uses the covariance to form the groups and to decide how much attention each gets, not to tell the directions inside a group apart. We form the groups in two ways, both set in advance: five fixed groups, and groups that start anew each time the eigenvalue has halved (six or seven groups in the median quarter). Among the 500 largest stocks, grouped RTQ is level with the principal-component design that has the length of the return vector: with groups by eigenvalue, the median loss at two bits per stock is 0.73 basis points against 0.74, and it is the lower of the two in 56 percent of quarters. Among all stocks it is ahead: the median loss is 1.8 basis points against 2.4, and it is the lower of the two in 92 percent of quarters, with similar results at one and three bits. Fixed groups give median losses up to 10 percent higher and a larger loss in the worst quarter among the 500 largest stocks. Grouped RTQ does not remove the bad quarters, however: its largest losses are about the same as those of the principal-component design (123 against 128 basis points at two bits among all stocks), because both rely on the same estimated covariance. So the random rotation costs nothing once the common factors are handled separately, and helps where individual directions are poorly estimated, as among all stocks. Only RTQ without the covariance avoids the bad quarters.

The theory also says where RTQ should be closest to the covariance-based design: where loss concentration is low. Internet Appendix Table IA.9 tests this by removing the top 1, 5, and 20 principal components from returns and repeating the comparison on what remains, which is closer to stock-specific risk. Removing them lowers median $\rho(1)$ among the 500 largest stocks from 0.88 to 0.32, 0.10, and 0.06. The pattern is as predicted. At two bits per stock, RTQ's median loss is about 15 times that of the principal-component design with the length before any factor is removed, 2.3 times after one, 1.5 times after five, and 1.2 times after twenty. At three bits per stock the two are about equal once five factors are removed, and among all stocks RTQ has the lower loss in most quarters once one factor is removed (54, 92, and 70 percent of quarters). At one and two bits the covariance-based design stays ahead in most quarters throughout. The estimated covariance is therefore the better guide for factor risk, and for stock-specific risk RTQ comes close to it or matches it without using the covariance.

The welfare comparison measures losses inside the model. To check that the gain shows up in returns an investor could have earned, we build portfolios with limited attention. At each quarter-end, an investor with κ bits learns the risk in eigendirection j of the trailing 60-month covariance up to distortion $D_j = \min(C_T 4^{-b_j}, 1)$, and treats what it has not learned as average risk, so its risk estimate in direction j is $(1-D_j)\lambda_j + D_j\bar{\lambda}$. The RTQ allocation divides the bits across directions

and VNV puts them all on the top direction; we call the two resulting portfolios the RTQ portfolio and the VNV portfolio for short. This rule can also be read as shrinking the estimated eigenvalues toward their mean, with more shrinkage where less attention is paid, so the comparison reflects in part the benefit of regularizing the covariance; "full information" below means full use of the estimated covariance, not knowledge of the true one.

The rule has a simple information interpretation. Suppose the investor knows the eigenvectors and the average eigenvalue $\bar{\lambda}$, and receives a signal about each λ_j whose error variance is the fraction D_j of her prior variance. With a linear signal, her posterior mean puts weight $1 - D_j$ on the signal and D_j on $\bar{\lambda}$. The rule is the expected value of this posterior mean given the true λ_j, so it describes the investor's estimate on average and leaves out the signal noise. Each investor holds the minimum-variance portfolio of its risk estimate for the next three months. Minimum variance is the natural target here because it depends only on the covariance, which is what attention is spent on, and needs no forecast of expected returns.

Table 3 reports how these portfolios performed from 1971 to 2026. Panel A uses the 500 largest stocks and Panel B all eligible stocks. Each row is one investor: 1/N, which uses no information; VNV; RTQ with one and two bits per stock; and an investor with full information on the estimated covariance. The columns are the mean return, the volatility, their ratio, and gross leverage. Since every investor targets minimum variance, the column to focus on is volatility. Among the 500 largest stocks, the RTQ portfolio has an annual volatility of 10.7 percent against 12.7 percent for VNV, and among all eligible stocks 10.9 percent against 14.4 percent. The row "RTQ minus VNV" shows that both differences are statistically significant ($t \approx 4$). The RTQ portfolio matches the full-information portfolio (10.9 and 11.0 percent) while using only one or two bits per stock, and it has a higher ratio of mean return to volatility than VNV (1.00 against 0.93 among the 500 largest stocks, 1.21 against 1.07 among all stocks). VNV earns a higher mean return, by 1.1 and 2.3 percentage points per year, but the difference is not statistically significant ($t = -1.14$ and -1.69), so the gain from spreading attention comes through lower risk rather than higher return. The portfolios differ in gross leverage (1.79 for VNV, 2.59 for RTQ, and 2.79 for full information among the 500 largest stocks), because a more detailed risk estimate supports larger offsetting positions; the comparison is before trading and financing costs, which would weigh more on the more levered portfolios. Panel C splits the quarters by ex-ante $\rho(1)$. The volatility reduction is largest when $\rho(1)$ is low, that is, when risk is spread across many directions: 3.6 percentage points among all stocks in the lowest tercile, against 1.7 and 2.9 in the middle and highest terciles. Panel D asks how much of this comes from where attention goes and how much from shrinkage as such. A portfolio that shrinks every eigenvalue by the same amount, the average shrinkage of RTQ, has a volatility of 10.82 percent among the 500 largest stocks, close to RTQ's 10.74 and far below VNV's 12.72. RTQ's remaining advantage over uniform shrinkage is statistically significant but small, 0.08 and 0.10

percentage points ($t = -2.79$ and -3.37). When the investor's risk estimate also carries the noise of her signal, the advantage is 0.31 and 0.20 percentage points ($t = -2.12$ and -1.84), and RTQ still has lower volatility than VNV by 1.9 and 3.5 percentage points. So an investor who spreads limited attention would have held a portfolio a sixth to a quarter less volatile than one who put all of it on the largest source of risk, and that most of this gain comes from paying some attention to every direction; how the attention is divided among them adds little.

Figure 8 shows the same comparison over time, as rolling 36-month volatility. The left panel is the 500 largest stocks and the right panel all eligible stocks. The solid black line (RTQ) lies below the dashed line (VNV) in most periods and close to the full-information line, and the gap is widest in periods of high market volatility. The lower risk of the RTQ portfolio is thus not confined to a few episodes. When breadth pays, the RTQ allocation turns limited attention into a better portfolio.

The US evidence leaves one question open. RTQ needs no covariance, but in US stocks a strong market factor dominates the covariance and is easy to estimate, and strategies that use it do better in the typical quarter. To illustrate RTQ where the covariance is hard to learn, we turn to venture capital. From the Crunchbase bulk export of April 2026, we count venture funding rounds each month from January 2008 to March 2026 in each of 454 Crunchbase categories with at least 500 rounds, and in Crunchbase's 49 broader category groups. Where venture money goes matters: inflows into venture funds raise the prices paid for new deals (Gompers and Lerner, 2000), and startups first funded in hot markets fail more often but are more innovative when they succeed (Nanda and Rhodes-Kropf, 2013). Funds and their investors therefore track deal flow sector by sector, with hundreds of variables and only a few years of history from which to learn how they move together. We do not offer this as a practical application, since the counts themselves are easy to look up. It is an illustration of data with many variables and a short history, where the covariance is hard to learn. The bits stand for how much of such data a decision maker can absorb and act on, not how much a computer can store, and deal counts stand in for what a VC would pay to learn sector by sector, such as deal quality or valuations. We repeat the encode-and-decode exercise of Table 2: at each quarter-end, strategies that use a covariance estimate it from the previous 24 months, and every strategy encodes each of the next 12 monthly vectors with 2 bits per variable. Every category counts equally in the loss, TurboQuant's own setting, so no strategy gains from a loss matrix built on the covariance. The loss measures how accurately each strategy tracks deal flow, not a payoff, so the exercise shows when the covariance helps; it is not a welfare comparison.

With 454 categories measured as monthly changes, RTQ has the lowest loss in Table 4, 11.7 percent of the loss with no information, against 13.3 percent for spreading across principal components and 81.6 percent for focusing on the top component, and it beats every other strategy in every quarter. Measured as levels, which share slow movements common to all categories, such as the rise in recorded rounds up to 2021 and their fall afterward, the covariance-based design given the

length of the vector keeps an edge (10.1 against 11.7 percent), because the common movement is a strong direction that the covariance picks up. Across the 49 groups the covariance-based design wins clearly, as in US stocks. This is not because the covariance of the groups is well estimated: with 49 groups and 24 months there are about as many covariance entries as observations. It is because the top component carries most of the variance (a median share of 0.64, against 0.16 for the categories), and so strong a direction can be learned from a short history. RTQ's loss is almost the same in all four columns, as the theory promises. Internet Appendix IA.7.3 shows that the pattern holds with a 60-month window and at one and three bits, and that with one bit per group, focusing on the top component beats RTQ in most quarters, as the theory predicts when attention is scarce and one direction dominates. When a few directions dominate, even a noisy covariance finds them, and using it pays. When risk is spread across many weak directions and the history is short, the covariance adds mostly noise, and RTQ does better without it.

We also look at institutional investors, using LSEG 13F holdings from 1980 to 2024 (352,755 manager-quarters and 10,693 managers with at least 20 holdings); Internet Appendix IA.7.4 reports the tests. We do not argue that managers use RTQ, and holdings do not reveal how a manager acquires information, so this evidence is descriptive. We rank each manager's portfolio from broad to concentrated and relate the ranking to the active return over the next quarter. Pooled over 45 years, neither kind of portfolio earns more. Estimated year by year, the winner switches often: concentrated portfolios earned significantly more in 18 years and broad portfolios in 16. The model says that concentrating should do relatively better when $\rho(1)$ is high. The 13F data do not show this: the interaction with ex-ante $\rho(1)$ is statistically insignificant and its point estimate has the opposite sign from the prediction (Table IA.15). The 13F results therefore do not validate the mechanism of the model, and the evidence for the $\rho(1)$ criterion comes from CRSP.

Internet Appendix IA.7 collects further results and robustness checks. Adjusting each manager's return for its own FF5+UMD loadings (Table IA.20), using all 13F filers (Tables IA.16, IA.17, and IA.18), and winsorizing returns (Table IA.19) leave the conclusions unchanged. Table IA.21 relates fit distance to portfolio concentration, active share, and size and momentum tilts.

To summarize, the CRSP evidence supports the theory's welfare predictions. Spreading attention with the RTQ allocation beats concentrating the same budget on one direction in every quarter from 1971 to 2026, out of sample as well as in sample. Minimum-variance portfolios built with the RTQ allocation have lower realized volatility than those built with VNV attention and match portfolios built with full information. RTQ with a random rotation, which uses no covariance, beats focused strategies from about two bits per stock but not the strategies that use the estimated covariance; grouped RTQ, which uses the covariance to form a few groups and rotates at random inside them, matches or beats those strategies. In venture deal flow across 454 categories measured as monthly changes, where risk is spread across many directions and the history is short,

RTQ beats the covariance-based strategies in every quarter; where a few directions dominate, the covariance-based strategies win. The gain from spreading is larger when loss is spread across many directions and smaller when one direction dominates, which is when concentrating gives up little, although out of sample this relation is only marginally significant. The 13F evidence is descriptive and fits the same message from the side of investors: neither broad nor concentrated portfolios win on average, and each wins in many years. Which approach pays depends on conditions, and when breadth pays, RTQ gives a disciplined way to spread attention.

The CRSP and 13F evidence tests the framework in finance. We also test it outside finance, on the Fisman et al. (2006) Columbia speed-dating experiment: 8,378 four-minute conversations among 551 participants in 21 sessions, where the number of partners per session varies from 5 to 22 for reasons unrelated to the participants. The parallel-target version of the model (Internet Appendix IA.2) predicts that a participant who spreads attention across more partners gets more matches, and that the slope β of $\log(1+\text{matches})$ on $\log(\text{partners})$ pins down the effective dimension of the traits participants care about, $d_{\text{eff}} = 1/(1-\beta)$ (Proposition IA.4). The estimated slope is 0.472 (session-clustered standard error 0.063), which implies $d_{\text{eff}} = 1.89$. Principal components of the six trait ratings give an independent estimate of 2.68. The slope is positive for both genders, across age groups, and in count-data models, and a placebo based on seating position is statistically insignificant. A model in which only a fixed number of partners receives attention would not produce a positive slope (Internet Appendix IA.3).

6 Conclusion

Many economic problems suffer from the curse of dimensionality, including rational inattention, the theory of how people cope with more information than they can process. This paper introduces rotate-then-quantize (RTQ) to economics to address it. A random rotation turns a problem with many interacting variables into many one-variable problems. RTQ's loss then stays within a known constant factor of the lowest worst-case loss of any signal with the same information capacity, for any prior with finite second moments, and the guarantee does not weaken with more variables.

The framework answers two questions. First, it gives a practical way to allocate attention across many variables, with this guarantee. Second, it gives a criterion for when to focus and when to spread: focusing costs little when a few sources of risk dominate and a great deal otherwise.[22] The evidence from US stocks is consistent with both. In the CRSP universe, spreading attention with the RTQ allocation, which uses the estimated covariance, beats spending the same budget on the largest source of risk in every quarter from 1971 to 2026, in and out of sample, and the

[22]Internet Appendix IA.2 gives the corresponding criterion for an agent who tracks many separate targets at once, and Internet Appendix IA.3 tests it in the speed-dating experiment of Fisman et al. (2006).

gain is larger when ex-ante loss concentration is lower (only marginally significant out of sample). Minimum-variance portfolios built with spread attention have lower realized volatility than those built with concentrated attention and match those built with full information, though equal attention to every direction does almost as well. RTQ with a random rotation, which does not use the covariance, beats focused strategies from about two bits per stock and has a stable loss across quarters; spreading with the estimated covariance does better in the typical quarter and worse when that estimate fails. A random rotation inside a few groups of risks formed from the covariance matches or beats both. In monthly venture funding across 454 Crunchbase categories, where histories are short and risk is spread across many weak directions, RTQ has a lower loss than every strategy that uses the estimated covariance; where a few directions dominate, those strategies do better. In the 13F data, neither broad nor concentrated portfolios win consistently, and loss concentration does not predict the winner; this evidence is descriptive, not a test of the model.

The main lesson is about robustness. A signal designed for a particular prior is best when that prior is right and can fail badly when it is wrong, even if the covariance is right. RTQ gives up a little in the first case in exchange for protection in the second, and its guarantee does not depend on the prior. Concentrated attention, as in Van Nieuwerburgh and Veldkamp (2010), where learning about an asset and holding it reinforce each other, remains the right answer when a few sources of risk dominate. Applied work that assumes concentrated attention for tractability could check the assumption with the focus-or-spread criterion.

The curse of dimensionality limits economic analysis well beyond portfolio choice, and RTQ is likely to be useful there too. Retailers and manufacturers set prices for thousands of products whose costs and demand move together (Bhattarai and Schoenle, 2014; Bonomo et al., 2023), and RTQ can spread limited attention across all of them. Central banks track hundreds of economic series (Bernanke et al., 2005), and the focus-or-spread criterion can say when watching a few headline indicators is enough. Forecasters and investors face hundreds of candidate predictors (Harvey et al., 2016; Giannone et al., 2021), and RTQ can use all of them coarsely rather than a few precisely. Bank supervisors monitor exposures across many institutions and markets, where an unwatched risk can be costly. Finally, economists and investors increasingly work with long numerical summaries of text, such as filings and news, produced by AI tools, which TurboQuant was built to compress. Each of these settings deserves its own study, and we hope other researchers will take RTQ to them.

RTQ also relates to the growing use of AI in investment management: Internet Appendix IA.4 shows that RTQ comes within a constant factor of the best linear cross-asset attention layer of models such as Cong et al. (2026), which gives these models a simple benchmark with an interpretation in terms of factor portfolios. Other directions for future work are threshold payoffs rather than quadratic losses, and general equilibrium with many inattentive agents.

References

Ailon, Nir and Bernard Chazelle, "The Fast Johnson–Lindenstrauss Transform and Approximate Nearest Neighbors," *SIAM Journal on Computing*, 2009, *39* (1), 302–322.

Bellman, Richard, *Dynamic Programming*, Princeton, NJ: Princeton University Press, 1957.

Bergemann, Dirk and Stephen Morris, "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 2016, *11* (2), 487–522.

Bernales, Alejandro, Marcela Valenzuela, and Ilknur Zer, "Effects of Information Overload on Financial Markets: How Much Is Too Much?," International Finance Discussion Papers 1372, Board of Governors of the Federal Reserve System 2023.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz, "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 2005, *120* (1), 387–422.

Bhatia, Rajendra, *Matrix Analysis*, Vol. 169 of *Graduate Texts in Mathematics*, Springer, 1997.

Bhattarai, Saroj and Raphael Schoenle, "Multiproduct Firms and Price-Setting: Theory and Evidence from U.S. Producer Prices," *Journal of Monetary Economics*, 2014, *66*, 178–192.

Bonomo, Marco, Carlos Carvalho, Oleksiy Kryvtsov, Sigal Ribon, and Rodolfo Rigato, "Multi-Product Pricing: Theory and Evidence from Large Retailers," *The Economic Journal*, 2023, *133* (651), 905–927.

Boot, Tom and Didier Nibbering, "Forecasting Using Random Subspace Methods," *Journal of Econometrics*, 2019, *209* (2), 391–406.

Caplin, Andrew and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

_ , _ , and John Leahy, "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies*, 2019, *86* (3), 1061–1094.

Carhart, Mark M., "On Persistence in Mutual Fund Performance," *Journal of Finance*, 1997, *52* (1), 57–82.

Cass, David, "Optimum Growth in an Aggregative Model of Capital Accumulation," *Review of Economic Studies*, 1965, *32* (3), 233–240.

Charles, Constantin, Cary Frydman, and Mete Kilic, "Insensitive Investors," *Journal of Finance*, 2024, *79* (4), 2473–2503.

Cochrane, John H., "Presidential Address: Discount Rates," *Journal of Finance*, 2011, *66* (4), 1047–1108.

Cong, Lin William, Ke Tang, and Jingyuan Wang, "AlphaPortfolio: Goal-Oriented Investment Management Through Deep Reinforcement Learning," NBER Working Paper No. 35195 2026. National Bureau of Economic Research.

Cover, Thomas M. and Joy A. Thomas, *Elements of Information Theory*, 2nd ed., Hoboken, NJ: Wiley-Interscience, 2006.

Denti, Tommaso, "Posterior Separable Cost of Information," *American Economic Review*, 2022, *112* (10), 3215–3259.

Fairhall, Adrienne L., Eric Shea-Brown, and Andrea Barreiro, "Information Theoretic Approaches to Understanding Circuit Function," *Current Opinion in Neurobiology*, 2012, *22* (4), 653–659.

Fama, Eugene F. and Kenneth R. French, "Luck versus Skill in the Cross-Section of Mutual Fund Returns," *Journal of Finance*, 2010, *65* (5), 1915–1947.

Fan, Jianqing, Yuan Liao, and Martina Mincheva, "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society, Series B*, 2013, *75* (4), 603–680.

Fisman, Raymond, Sheena S. Iyengar, Emir Kamenica, and Itamar Simonson, "Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment," *Quarterly Journal of Economics*, 2006, *121* (2), 673–697.

Gabaix, Xavier, "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri, "Economic Predictions With Big Data: The Illusion of Sparsity," *Econometrica*, 2021, *89* (5), 2409–2437.

Gompers, Paul and Josh Lerner, "Money Chasing Deals? The Impact of Fund Inflows on Private Equity Valuation," *Journal of Financial Economics*, 2000, *55* (2), 281–325.

Grossman, Sanford J. and Joseph E. Stiglitz, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 1980, *70* (3), 393–408.

Han, Insu, Praneeth Kacham, Amin Karbasi, Vahab Mirrokni, and Amir Zandieh, "PolarQuant: Vector Quantization with Polar Transformation," in "Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS)," Vol. 300 of *Proceedings of Machine Learning Research* PMLR 2026, pp. 5113–5121.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, "... and the Cross-Section of Expected Returns," *Review of Financial Studies*, 2016, *29* (1), 5–68.

Hellwig, Christian and Laura Veldkamp, "Knowing What Others Know: Coordination Motives in Information Acquisition," *Review of Economic Studies*, 2009, *76* (1), 223–251.

Johnson, William B. and Joram Lindenstrauss, "Extensions of Lipschitz Mappings into a Hilbert Space," in "Conference in Modern Analysis and Probability," Vol. 26 of *Contemporary Mathematics*, Providence, RI: American Mathematical Society, 1984, pp. 189–206.

Jung, Junehyuk, Jeong Ho Kim, Filip Matějka, and Christopher A. Sims, "Discrete Actions in Information-Constrained Decision Problems," *Review of Economic Studies*, 2019, *86* (6), 2643–2667.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, "Time-Varying Fund Manager Skill," *Journal of Finance*, 2014, *69* (4), 1455–1484.

_ , _ , and _ , "A Rational Theory of Mutual Funds' Attention Allocation," *Econometrica*, 2016, *84* (2), 571–626.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

Koop, Gary, Dimitris Korobilis, and Davide Pettenuzzo, "Bayesian Compressed Vector Autoregressions," *Journal of Econometrics*, 2019, *210* (1), 135–154.

Kőszegi, Botond and Adam Szeidl, "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics*, 2013, *128* (1), 53–104.

Ledoit, Olivier and Michael Wolf, "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices," *Journal of Multivariate Analysis*, 2004, *88* (2), 365–411.

Lloyd, Stuart P., "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, 1982, *28* (2), 129–137.

Maćkowiak, Bartosz and Mirko Wiederholt, "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 2009, *99* (3), 769–803.

_ and _ , "Business Cycle Dynamics under Rational Inattention," *Review of Economic Studies*, 2015, *82* (4), 1502–1532.

_ , Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Review," *Journal of Economic Literature*, 2023, *61* (1), 226–273.

Matějka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 2015, *105* (1), 272–298.

Max, Joel, "Quantizing for Minimum Distortion," *IRE Transactions on Information Theory*, 1960, *6* (1), 7–12.

McKenzie, Lionel W., "Turnpike Theory," *Econometrica*, 1976, *44* (5), 841–865.

Myatt, David P. and Chris Wallace, "Endogenous Information Acquisition in Coordination Games," *Review of Economic Studies*, 2012, *79* (1), 340–374.

Nanda, Ramana and Matthew Rhodes-Kropf, "Investment Cycles and Startup Innovation," *Journal of Financial Economics*, 2013, *110* (2), 403–418.

Nieuwerburgh, Stijn Van and Laura Veldkamp, "Information Acquisition and Under-Diversification," *Review of Economic Studies*, 2010, *77* (2), 779–805.

Panter, P. F. and W. Dite, "Quantization Distortion in Pulse-Count Modulation with Nonuniform Spacing of Levels," *Proceedings of the IRE*, 1951, *39* (1), 44–48.

Saint-Paul, Gilles, "A 'Quantized' Approach to Rational Inattention," *European Economic Review*, 2017, *100*, 50–71.

Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

_ , "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 2006, *96* (2), 158–163.

Steiner, Jakub, Colin Stewart, and Filip Matějka, "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 2017, *85* (2), 521–553.

Sun, Yeneng, "The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks," *Journal of Economic Theory*, 2006, *126* (1), 31–69.

Vargaftik, Shay, Ran Ben Basat, Amit Portnoy, Gal Mendelson, Yaniv Ben-Itzhak, and Michael Mitzenmacher, "DRIVE: One-bit Distributed Mean Estimation," in "Advances in Neural Information Processing Systems," Vol. 34 2021, pp. 362–377.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention Is All You Need," in "Advances in Neural Information Processing Systems (NeurIPS)," Vol. 30 2017, pp. 5998–6008.

Vershynin, Roman, *High-Dimensional Probability: An Introduction with Applications in Data Science* Cambridge Series in Statistical and Probabilistic Mathematics, 1st ed., Cambridge University Press, 2018.

Woodford, Michael, "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics*, 2009, *56*, S100–S124.

Zandieh, Amir, Majid Daliri, and Insu Han, "QJL: 1-Bit Quantized JL Transform for KV Cache Quantization with Zero Overhead," in "Proceedings of the AAAI Conference on Artificial Intelligence," Vol. 39 2025, pp. 25805–25813.

_ , _ , Majid Hadian, and Vahab Mirrokni, "TurboQuant: Online Vector Quantization with Near-optimal Distortion Rate," in "International Conference on Learning Representations (ICLR)" 2026.

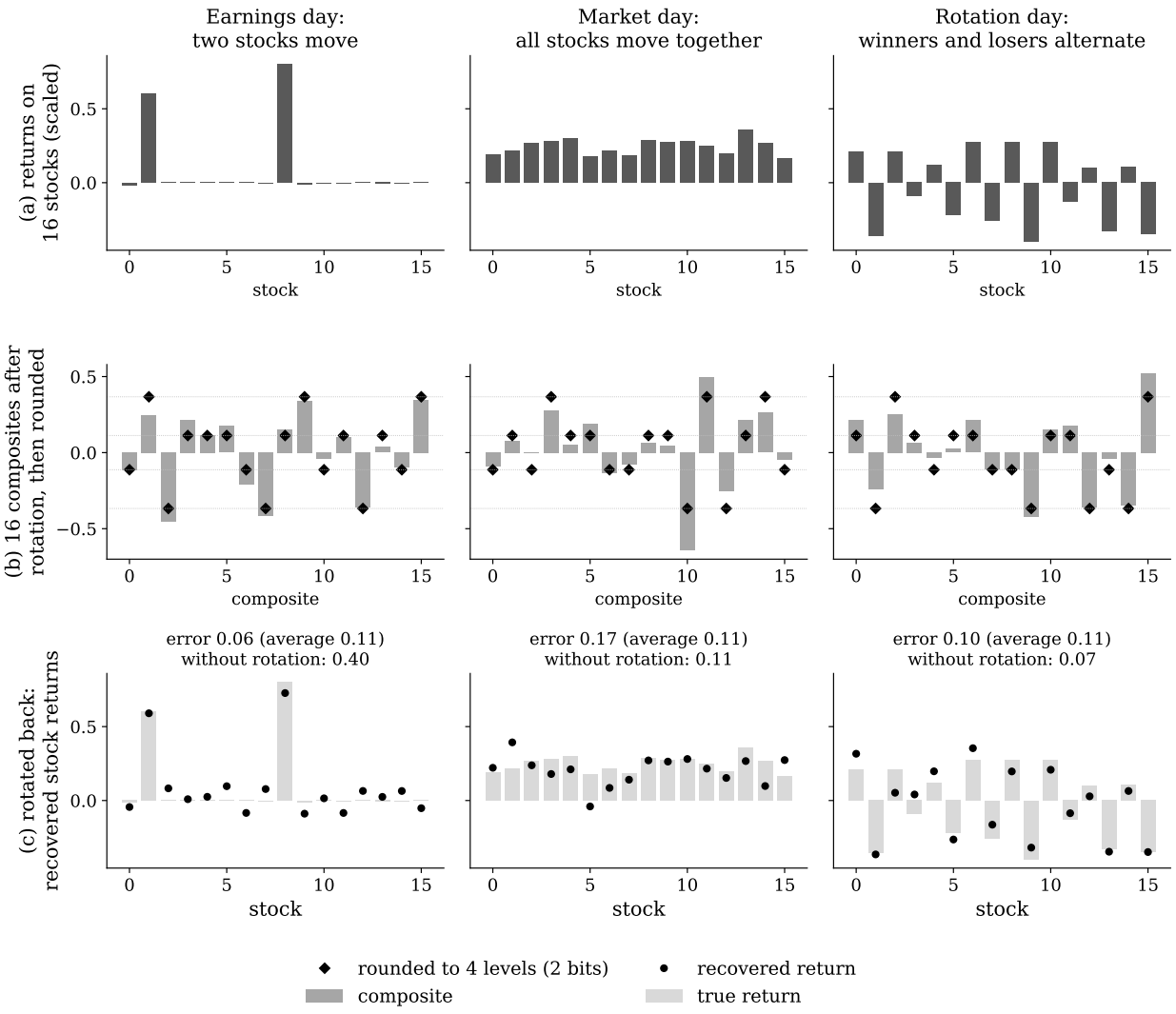

Figure 1: **Rotate-then-quantize in a small example.** Returns on 16 stocks on three stylized days, each scaled so that the squared returns sum to one. Row (a): the stock returns. Row (b): the returns on 16 composite portfolios after one random rotation (bars) and after rounding each to one of four levels, that is, two bits per composite (diamonds; dotted lines mark the four levels). Row (c): the rounded composites converted back into stock returns (dots) against the true returns (bars). Panel titles report the squared error as a share of the day's squared size, for this rotation, on average over 2,000 random rotations, and for rounding each stock's return directly with the same four levels.

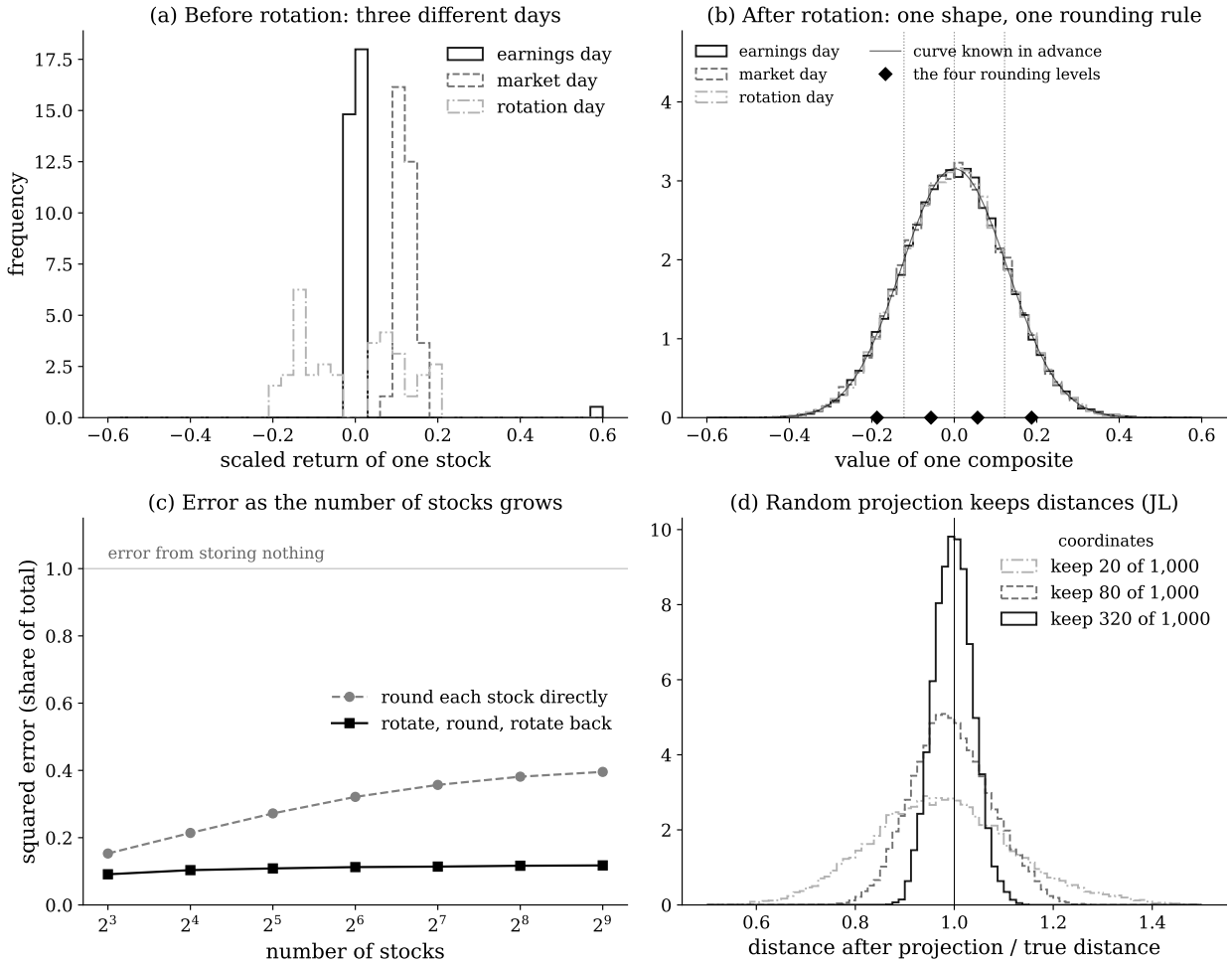

Figure 2: **Why rotation works.** (a) Distribution of scaled returns on 64 stocks on three stylized days (earnings, market and rotation days, as in Figure 1). (b) Distribution of the composite returns after random rotation, pooled over 400 rotations; the thin solid line is the distribution known in advance, which depends only on the number of stocks, and the diamonds are the four two-bit rounding levels chosen for it. (c) Average squared error (as a share of the day's squared size) at two bits per number, as the number of stocks grows, for the same rounding rule applied with and without rotation; averages over 30 rotations and the three kinds of day. (d) Johnson-Lindenstrauss: for 200 points with 1,000 coordinates, the distribution across all pairs of the distance after a random projection divided by the true distance, keeping 20, 80, or 320 coordinates.

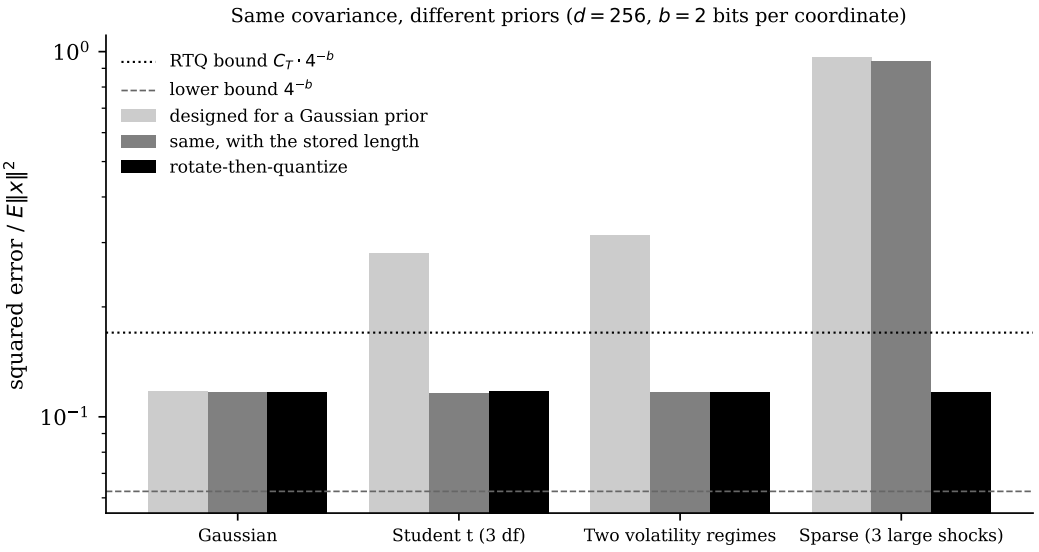

Figure 3: **Same covariance, different priors.** Expected squared error relative to $\mathbb{E}\|\mathbf{x}\|^2$ at $d = 256$ and $b = 2$ bits per coordinate, 4,000 draws per prior. Every prior has mean zero and identity covariance. Light gray: b-bit Lloyd-Max quantizer for $N(0,1)$ applied to each coordinate (the design for a Gaussian prior). Dark gray: the same design given the length of the state. Black: rotate-then-quantize. Dotted line: the RTQ bound $C_T \cdot 4^{-b}$. Dashed line: the large-d lower bound 4^{-b} on the worst-case error of any signal with the same capacity (the exact bound is $4^{-b}/\gamma_d(b)$, about six percent lower at $d = 256$).

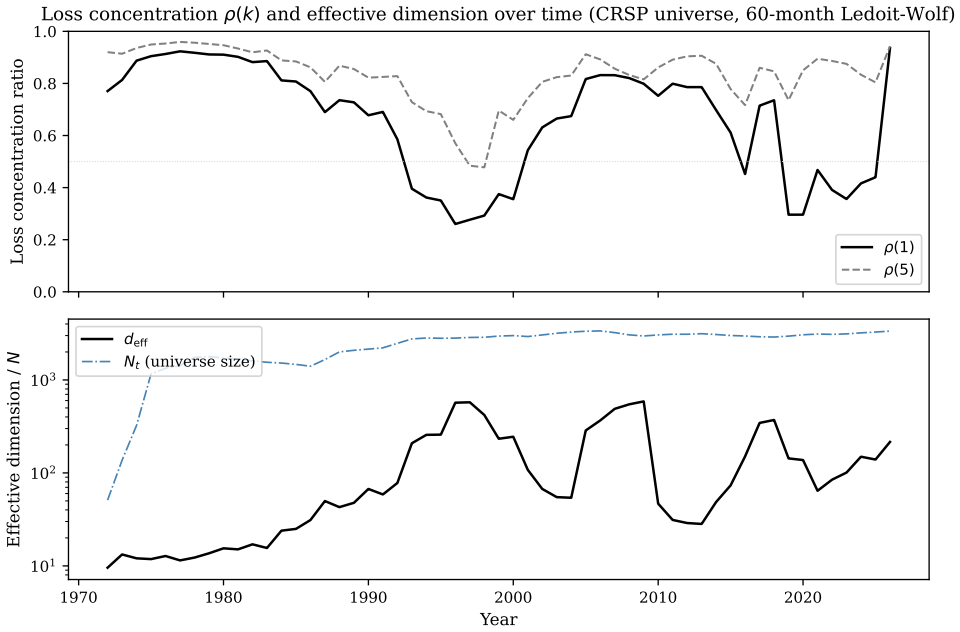

Figure 4: **Loss concentration $\rho(k)$ and effective dimension in the CRSP universe, 1971–2025.** At each year-end, Σ_t is the 60-month Ledoit and Wolf (2004) shrinkage estimator on monthly returns of the eligible CRSP universe (common stocks on NYSE/AMEX/NASDAQ with 60+ months). Top: $\rho(1)$ (solid) and $\rho(5)$ (dashed) with $\rho = 0.5$ dotted. Bottom: d_{eff} (solid, left) and universe size N_t (dot-dash, right). Mean $\rho(1) = 0.66$, range $[0.26, 0.94]$; mean $d_{\text{eff}} = 147$, range $[9.6, 588]$. Neither flat nor rank-one, which is the range where spreading and focusing differ.

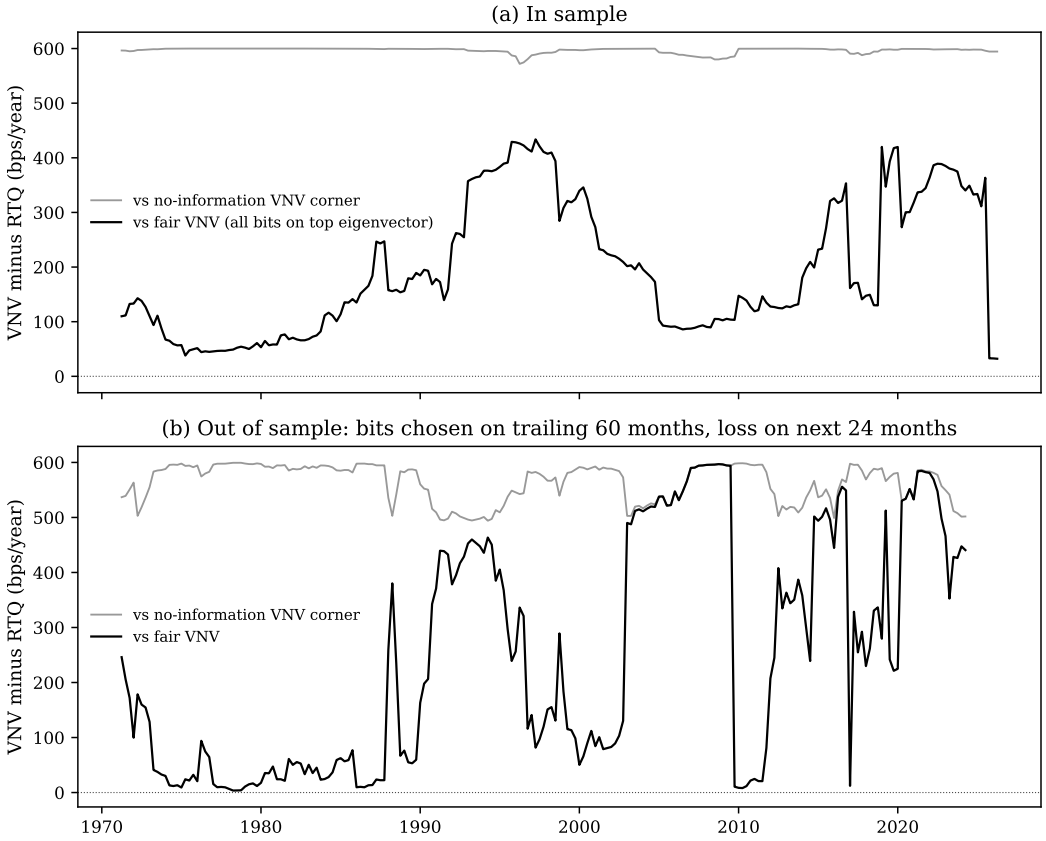

Figure 5: **Rolling welfare gain from spreading attention, 1971–2026.** At each quarter-end, the RTQ allocation and fair VNV (the same $\kappa = 2N$ bits on the top eigendirection) are chosen from the 60-month Ledoit-Wolf covariance. Black: welfare cost of fair VNV minus that of the RTQ allocation, in basis points per year. Gray: the gain over receiving no information. Panel (a): losses evaluated with the same covariance. Panel (b): losses evaluated in the ex-ante eigendirections with the covariance of the next 24 months. The gain is positive in every quarter in both panels.

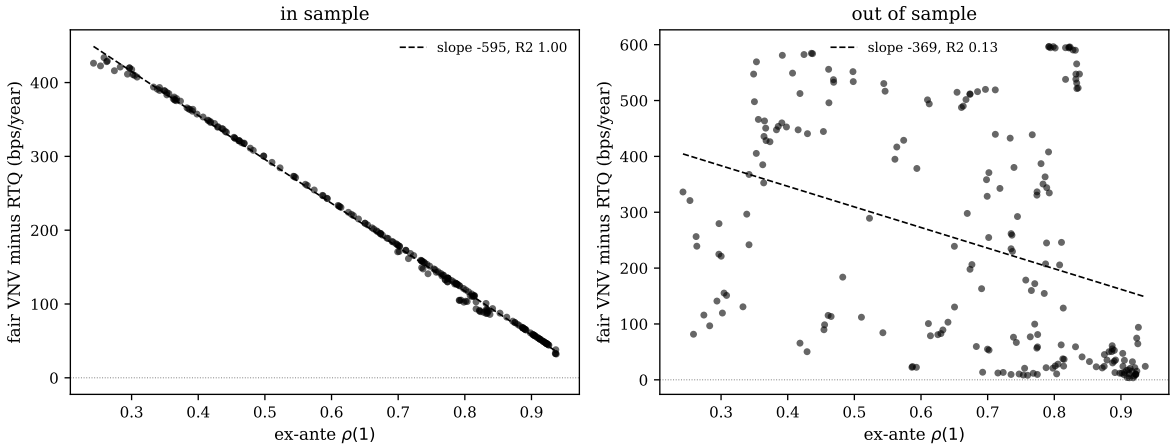

Figure 6: **Loss concentration and the gain from spreading attention.** Each point is a quarter-end, 1971–2026. Horizontal: ex-ante $\rho(1)$ from the trailing 60-month Ledoit-Wolf covariance. Vertical: welfare cost of fair VNV minus that of the RTQ allocation in basis points per year, $\kappa = 2N$. Left: in sample, where the gain is close to $600 \cdot (1 - \rho(1))$ by construction. Right: out of sample, with losses evaluated on the next 24 months. Dashed lines are OLS fits.

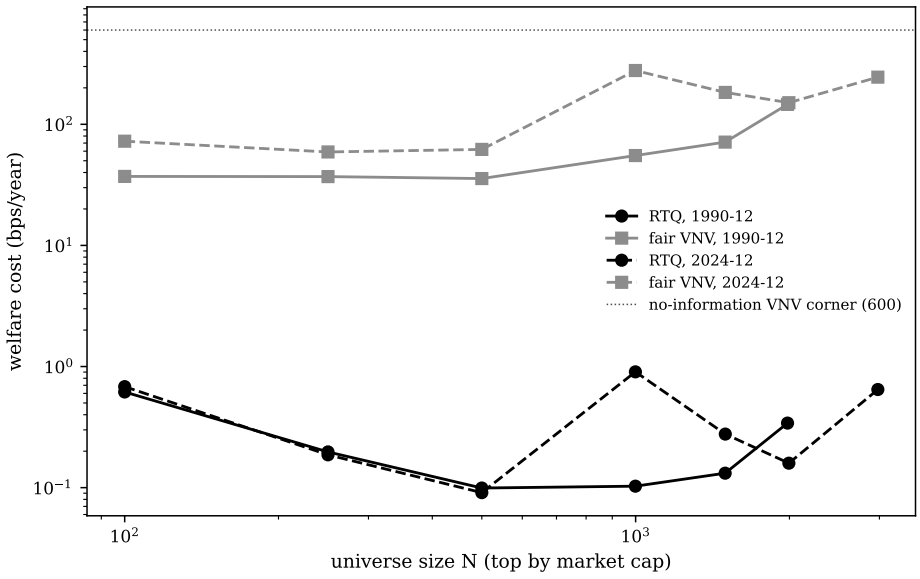

Figure 7: **Welfare cost across universe size.** At the end of 1990 and 2024, CRSP common stocks are ranked by market capitalization and the top N are taken. For each N, Σ is the 60-month Ledoit-Wolf covariance and $\kappa = 2N$. Both axes are log scale. Black circles: RTQ allocation. Gray squares: fair VNV. Dotted line: receiving no information (600 basis points).

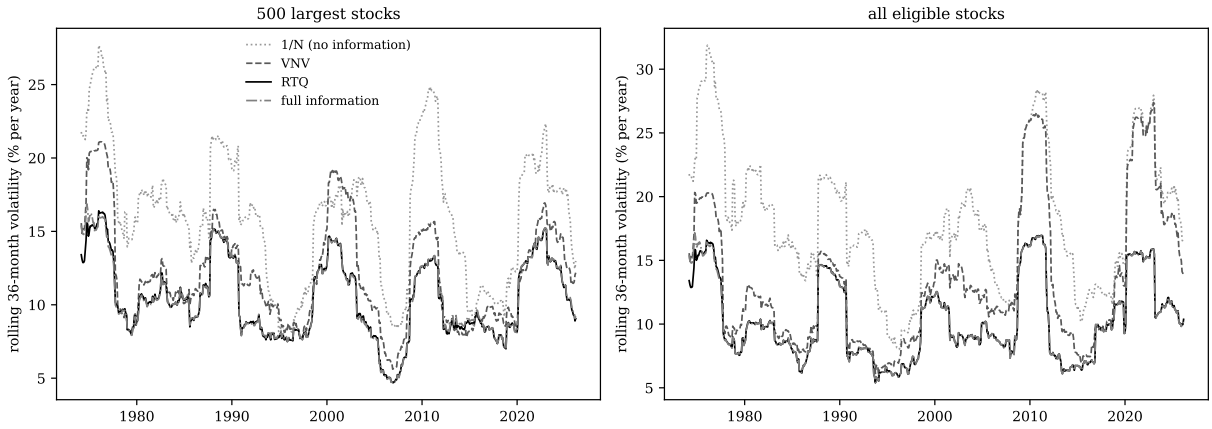

Figure 8: **Realized volatility of minimum-variance portfolios built with limited attention.** Rolling 36-month volatility of monthly returns, in percent per year, for the portfolios in Table 3 with $\kappa = 2N$. Solid black: RTQ allocation. Dashed: VNV. Dash-dot: full information on the covariance. Dotted: 1/N.

	Fair VNV minus RTQ allocation					No information minus
	In sample	Out of sample				(median, out of sample)
Budget κ	Median	Median	Mean	95% CI	NW t	RTQ allocation
A. All eligible stocks, Ledoit-Wolf covariance						
N	155	86	145	[68, 218]	3.63	530
$2N$	158	206	247	[145, 337]	4.84	582
$3N$	159	239	274	[163, 368]	5.02	596
B. All eligible stocks, sample covariance						
N	152	30	152	[53, 254]	2.90	586
$2N$	152	39	159	[58, 263]	2.96	588
$3N$	152	61	177	[74, 281]	3.24	595
C. 500 largest stocks, Ledoit-Wolf covariance						
N	75	29	45	[30, 53]	7.01	578
$2N$	76	42	79	[51, 102]	5.97	592
$3N$	76	50	93	[59, 123]	5.67	598

Table 1: **Welfare gain from spreading attention, rolling CRSP windows.** Each quarter-end from 1970 to 2025, the covariance of monthly returns is estimated on the trailing 60 months. The RTQ allocation divides κ bits across all eigendirections of the estimated covariance by the rule of Proposition 3; it uses no random rotation. Fair VNV spends the same κ bits on the top eigendirection only. Entries are differences in welfare cost, so a positive number is a gain from the RTQ allocation, in basis points of annual certainty-equivalent return, with the loss in each direction equal to $\lambda_j^2 \min(C_T 4^{-b_j}, 1)$ (Proposition 3). In sample, bits are chosen and losses evaluated with the same covariance (221 quarters). Out of sample, bits are chosen with the trailing covariance and losses are evaluated in the same ex-ante eigendirections with the covariance of the next 24 months (213 quarters). The 95% confidence interval is from a block bootstrap; NW t uses Newey-West standard errors with 20 lags. The RTQ allocation has the lower loss in every quarter in every specification. The last column compares it with receiving no information, the corner used in earlier drafts. N is the number of stocks in the window.

	1 bit per stock			2 bits per stock			3 bits per stock		
	Median	Mean	RTQ lower	Median	Mean	RTQ lower	Median	Mean	RTQ lower
A. 500 largest stocks									
Random rotation									
RTQ, equal bits	83.8	85.0	–	10.2	11.0	–	1.4	1.6	–
RTQ, allocated bits	48.4	54.2	0%	6.2	7.2	0%	1.0	1.3	1%
Random rotation inside groups									
Grouped RTQ, fixed groups	1.7	3.7	0%	0.76	1.8	0%	0.35	0.88	0%
Grouped RTQ, groups by eigenvalue	1.5	3.2	0%	0.73	1.4	0%	0.34	0.63	0%
Spread, using the estimated covariance									
Principal components	2.3	8.8	0%	1.0	7.0	15%	0.42	6.3	21%
Principal components, with length	1.6	3.2	0%	0.74	1.4	0%	0.34	0.64	0%
Normal-prior optimum	0.75	1.6	0%	0.14	0.33	0%	0.03	0.08	0%
Focused on the top principal components									
Top 1 (fair VNV)	51.9	62.8	24%	51.9	62.8	100%	51.9	62.8	100%
Top 5	19.0	26.5	3%	19.0	26.5	79%	19.0	26.5	100%
Top 20	7.8	14.8	0%	7.7	14.6	38%	7.7	14.6	99%
Top 100	2.5	8.8	0%	2.4	8.7	16%	2.2	8.6	75%
B. All eligible stocks									
Random rotation									
RTQ, equal bits	85.1	88.1	–	11.0	12.4	–	1.6	2.0	–
RTQ, allocated bits	82.0	77.4	25%	10.7	11.1	5%	1.6	1.9	20%
Random rotation inside groups									
Grouped RTQ, fixed groups	4.9	15.2	5%	2.0	8.4	7%	0.70	5.6	15%
Grouped RTQ, groups by eigenvalue	4.7	14.1	4%	1.8	7.8	8%	0.63	5.4	12%
Spread, using the estimated covariance									
Principal components	9.4	38.9	18%	4.3	31.3	31%	2.5	28.3	46%
Principal components, with length	5.7	14.9	4%	2.4	8.4	7%	0.94	5.7	23%
Normal-prior optimum	1.7	7.0	0%	0.34	1.8	0%	0.08	0.45	3%
Focused on the top principal components									
Top 1 (fair VNV)	111.0	167.8	57%	111.0	167.8	99%	111.0	167.8	100%
Top 5	46.1	79.6	33%	46.0	79.6	90%	46.0	79.6	100%
Top 20	24.8	56.9	24%	24.8	56.7	72%	24.8	56.7	99%
Top 100	13.2	44.8	21%	13.2	44.7	53%	13.1	44.6	94%

Table 2: **RTQ with a random rotation against strategies that use the estimated covariance.** Every strategy encodes each monthly return vector with bN bits and decodes it; no distortion formula is used. At each quarter-end the covariance is estimated on the trailing 60 months (Ledoit-Wolf) and also serves as the loss matrix. The loss is the realized weighted squared error on the returns of the next 24 months, in basis points per year on the scale where no information costs 600 (213 quarters, 1971–2024). RTQ rotates the return vector with a random rotation, takes its length as given, and rounds every rotated coordinate with the same b bits (equal bits, which uses no information about the covariance) or with bits allocated by the loss weights as in Proposition 3 (allocated bits). Grouped RTQ sorts the principal components into groups, divides bits across groups by their loss weights, and inside each group applies a random rotation and rounds every rotated coordinate alike, given the length. Fixed groups: components 1, 2–5, 6–20, 21–100, and the rest. Groups by eigenvalue: a new group each time the eigenvalue has halved. Principal components: each principal component of the estimated covariance is rounded with the quantizer for a normal prior, with bits allocated across components; "with length" gives this design the length of the return vector, as RTQ has. Normal-prior optimum: the classical solution under a normal prior, a benchmark that adds normal noise and is not a finite code. Top k: all bits go to the first k principal components and the others are left at the prior. Top 100 has 209 quarters. "RTQ lower" is the share of quarters in which RTQ with equal bits has the lower loss.

	Mean (% per year)	Volatility (% per year)	Mean / volatility	Gross leverage
A. 500 largest stocks				
1/N (no information)	13.1	16.5	0.79	1.00
VNV	11.8	12.7	0.93	1.79
RTQ, $\kappa = N$	10.8	10.8	1.00	2.34
RTQ, $\kappa = 2N$	10.7	10.7	1.00	2.59
Full information	10.8	10.9	0.99	2.79
RTQ ($\kappa = 2N$) minus VNV	-1.1	-2.0		
t-statistic	(-1.14)	(+3.94)		
B. All eligible stocks				
1/N (no information)	18.2	18.7	0.97	1.00
VNV	15.5	14.4	1.07	1.50
RTQ, $\kappa = N$	13.2	11.0	1.21	1.90
RTQ, $\kappa = 2N$	13.2	10.9	1.21	1.94
Full information	13.4	11.0	1.22	1.98
RTQ ($\kappa = 2N$) minus VNV	-2.3	-3.5		
t-statistic	(-1.69)	(+4.00)		

C. Volatility of VNV minus RTQ ($\kappa = 2N$), percentage points per year, by $\rho(1)$ tercile

	Low $\rho(1)$	Middle $\rho(1)$	High $\rho(1)$
500 largest stocks	2.3	0.6	1.6
	(4.05)	(1.23)	(2.61)
$\rho(1)$ range	0.45–0.82	0.82–0.90	0.90–0.95
All eligible stocks	3.6	1.7	2.9
	(4.01)	(5.18)	(3.78)
$\rho(1)$ range	0.24–0.59	0.61–0.79	0.79–0.94

D. Volatility (% per year) against uniform shrinkage, $\kappa = 2N$

	VNV	Uniform	RTQ	RTQ minus uniform
Expected risk estimate				
500 largest stocks	12.72	10.82	10.74	-0.08 (-2.79)
All eligible stocks	14.42	10.99	10.89	-0.10 (-3.37)
Risk estimate with signal noise				
500 largest stocks	12.77	11.22	10.91	-0.31 (-2.12)
All eligible stocks	14.43	11.12	10.92	-0.20 (-1.84)

Table 3: **Minimum-variance portfolios built with limited attention, 1971–2026.** At each quarter-end, the covariance of monthly returns is estimated on the trailing 60 months (Ledoit-Wolf). An investor with κ bits learns the risk in eigendirection j up to distortion $D_j = \min(C_T 4^{-b_j}, 1)$ and treats what is not learned as average risk, so the risk estimate in direction j is $(1-D_j)\lambda_j + D_j\bar{\lambda}$. RTQ denotes the RTQ allocation (Proposition 3, no random rotation); VNV puts all bits on the top direction (its results are the same at $\kappa = N$ and $2N$). Each investor holds the global minimum-variance portfolio of its risk estimate for the next three months; 660 monthly returns. Full information uses the estimated covariance itself; 1/N uses none. Mean / volatility uses raw returns. Panels A and B: the t-statistic under the mean is for the difference in mean returns and the one under volatility is for the difference in variances, both with Newey-West standard errors (12 lags). Panel C: within-quarter volatility of VNV minus RTQ, averaged over the quarters in each tercile of ex-ante $\rho(1)$, with Newey-West t-statistics in parentheses. Panel D: "Uniform" shrinks every eigenvalue toward the mean by the average D_j of RTQ, so it pays equal attention to every direction. "Expected risk estimate" is the rule of Panels A and B. The rows with signal noise use the posterior mean of an investor whose signal about each log eigenvalue has error share D_j, under a normal prior on log eigenvalues. The last column is the volatility difference, with the Newey-West t-statistic (6 lags) for the difference in variances in parentheses.

	454 categories		49 category groups	
	Monthly change	Log level	Monthly change	Log level
Top component's share of variance	0.16	0.14	0.64	0.57
Uses no covariance				
RTQ	11.7	11.7	11.5	11.1
Round each variable, no rotation	14.8 (100)	13.9 (65)	15.4 (100)	12.9 (77)
Uses the estimated covariance				
Focus: top component (fair VNV)	81.6 (100)	67.3 (100)	27.1 (98)	26.1 (90)
Spread across principal components	13.3 (100)	12.6 (74)	3.73 (2)	3.37 (0)
with length	14.7 (100)	10.1 (23)	5.61 (2)	3.79 (0)

Table 4: **Venture deal flow: RTQ with many variables and a short history.** The state each month is the vector of $\log(1+\text{rounds})$ of venture funding across Crunchbase categories or category groups, measured as its monthly change or as its level. At each quarter-end, strategies that use a covariance estimate it (Ledoit-Wolf) from the previous 24 months; every strategy then encodes each of the next 12 monthly vectors with 2 bits per variable, and the realized squared error is measured, with every variable counting equally. Entries are median losses as a percent of the loss with no information; in parentheses, the percent of quarter-ends in which RTQ has the lower loss. Top component's share of variance is the median share of the largest eigenvalue of the estimated covariance. Venture rounds from January 2008 to March 2026; categories with at least 500 rounds; 61 or 62 quarter-ends. Internet Appendix Table IA.10 reports 1 and 3 bits, a 60-month window, and the full definitions.

A Proofs

This appendix proves the propositions of Section 4, starting with the TurboQuant MSE lemma they rely on. The concentration lemma used only for intuition is Lemma IA.1 in the Internet Appendix.

A.1 TurboQuant MSE lemma

Lemma 2 (Worst-case scalar-quantization MSE on the unit sphere)**.** *Fix any unit vector* $\mathbf{u} \in S^{d-1}$. *Let R be a Haar-random orthogonal matrix, drawn independently of* $\mathbf{u}$. *Let* $\hat{\mathbf{r}}$ *be the vector obtained by scalar quantization of* $R\mathbf{u}$ *at b bits per coordinate using the Lloyd-Max quantizer for the* $\text{Beta}(\frac{d-1}{2}, \frac{d-1}{2})$ *distribution, and let* $\hat{\mathbf{u}} = R^\top \hat{\mathbf{r}}$. *Then the worst-case MSE satisfies*

$$\sup_{\mathbf{u} \in S^{d-1}} \mathbb{E}\left[\|\mathbf{u} - \hat{\mathbf{u}}\|^2\right] \;\le\; C_T \cdot 4^{-b}, \qquad C_T \;=\; \pi\sqrt{3}/2 \approx 2.72, \tag{18}$$

where the expectation is over R and the quantization, and C_T is independent of d, b, and $\mathbf{u}$.

Proof. This is Theorem 1 of Zandieh et al. (2026). (A randomized Hadamard transform is a fast substitute for R in practice; the exact statement requires the Haar rotation.) The Haar rotation makes any fixed $\mathbf{u}$ uniform on S^{d-1}, so each coordinate of $R\mathbf{u}$ is $\text{Beta}(\frac{d-1}{2}, \frac{d-1}{2})$ marginally with pairwise correlation $O(1/d)$. The Lloyd-Max quantizer designed for that density achieves per-coordinate MSE $C_T/d \cdot 4^{-b}$ (Panter-Dite high-rate constant evaluated against the Beta density), so summing over d coordinates gives the stated dimension-free bound. The precursor components are in Han et al. (2026) (polar/Beta scalar quantization) and Zandieh et al. (2025) (random-rotation preconditioning). □

A.2 Proof of Proposition 1 (RTQ guarantee, normalized-state case)

Proof. Step 1: Decompose with stored norm. Decompose the state as $\mathbf{x} = \rho\mathbf{u}$, where $\rho = \|\mathbf{x}\|$ and $\mathbf{u} \in S^{d-1}$. Because ρ is observed or stored exactly, the agent allocates the entire bit budget to encoding the direction. RTQ rotates $\mathbf{u}$ by Haar-random R to get $\mathbf{r} = R\mathbf{u}$, scalar-quantizes each coordinate at $b = \kappa/d$ bits using the Lloyd-Max quantizer for the $\text{Beta}(\frac{d-1}{2}, \frac{d-1}{2})$ distribution to get $\hat{\mathbf{r}}$, and unrotates: $\tilde{\mathbf{u}} = R^\top \hat{\mathbf{r}}$. The reconstruction is $\hat{\mathbf{x}} = \rho\tilde{\mathbf{u}}$, and the reconstruction error is

$$\|\mathbf{x} - \hat{\mathbf{x}}\|^2 \;=\; \|\rho\mathbf{u} - \rho\tilde{\mathbf{u}}\|^2 \;=\; \rho^2 \|\mathbf{u} - \tilde{\mathbf{u}}\|^2, \tag{19}$$

exactly.

Step 2: Apply the spectral and TurboQuant bounds. For any positive definite W, the spectral

inequality $W \preceq \lambda_{\max}(W) \cdot I$ gives

$$(\mathbf{x}-\hat{\mathbf{x}})^\top W(\mathbf{x}-\hat{\mathbf{x}}) \;\le\; \lambda_{\max}(W) \cdot \rho^2 \, \|\mathbf{u}-\tilde{\mathbf{u}}\|^2. \tag{20}$$

Since R is drawn independently of $(\rho, \mathbf{u})$, which may be correlated with each other under F, we condition on $(\rho, \mathbf{u})$. Lemma 2 holds uniformly over unit directions, so for every $(\rho, \mathbf{u})$,

$$\mathbb{E}\big(\|\mathbf{u}-\tilde{\mathbf{u}}\|^2 \,\big|\, \rho, \mathbf{u}\big) \;\le\; C_T \cdot 4^{-b}, \qquad C_T = \pi\sqrt{3}/2. \tag{21}$$

By the tower property,

$$\mathbb{E}[\rho^2 \|\mathbf{u}-\tilde{\mathbf{u}}\|^2] \;=\; \mathbb{E}\big[\rho^2 \mathbb{E}(\|\mathbf{u}-\tilde{\mathbf{u}}\|^2 \mid \rho, \mathbf{u})\big] \;\le\; \mathbb{E}[\rho^2] \cdot C_T \cdot 4^{-b}. \tag{22}$$

Combining with the spectral inequality,

$$\mathscr{L}^{RTQ} \;=\; \mathbb{E}[(\mathbf{x}-\hat{\mathbf{x}})^\top W(\mathbf{x}-\hat{\mathbf{x}})] \;\le\; \lambda_{\max}(W) \cdot \mathbb{E}[\rho^2] \cdot C_T \cdot 4^{-b}, \tag{23}$$

which is the upper bound (9) in the proposition.

Step 3: The lower bound. Consider the prior in which ρ is constant and $\mathbf{u}$ is uniform on S^{d-1}. By Lemma IA.7 in the Internet Appendix, every signal with $I(\mathbf{u};\mathbf{s} \mid R) \le bd$ bits and every reconstruction $\hat{\mathbf{u}}(\mathbf{s})$ satisfy

$$\mathbb{E}\|\mathbf{u}-\hat{\mathbf{u}}\|^2 \;\ge\; 4^{-b}/\gamma_d(b), \qquad \gamma_d(b) = 4^{\,b(2d-1)/(d-1)^2} \left(\tfrac{d+1}{2}\right)^{2d/(d-1)^2}. \tag{24}$$

Since $W \succeq \lambda_{\min}(W) \cdot I$, the weighted loss under this prior is at least $\lambda_{\min}(W)\,\rho^2 4^{-b}/\gamma_d(b) = \mathscr{L}^{\text{sphere}}_{\text{lb}}(\kappa)/\gamma_d(b)$, with

$$\mathscr{L}^{\text{sphere}}_{\text{lb}}(\kappa) \;\equiv\; \lambda_{\min}(W) \cdot \mathbb{E}[\rho^2] \cdot 4^{-b}. \tag{25}$$

This is a worst case over priors, not the prior-specific optimum $\mathscr{L}^*(F,\kappa)$, which can be smaller for structured priors.

Step 4: Combine. Dividing (23) by (25) gives $\mathscr{L}^{RTQ}(\kappa)/\mathscr{L}^{\text{sphere}}_{\text{lb}}(\kappa) \le C_T \chi(W)$, so RTQ's loss is at most $C_T\,\chi(W)\,\gamma_d(b)$ times the lower bound of Step 3, which is (11). □

The constant $C_T = \pi\sqrt{3}/2 \approx 2.72$ is the TurboQuant MSE constant of Zandieh et al. (2026). The derivation requires the norm to be stored exactly, so that $\hat{\mathbf{x}} = \rho\tilde{\mathbf{u}}$ and $\|\mathbf{x}-\hat{\mathbf{x}}\|^2 = \rho^2\|\mathbf{u}-\tilde{\mathbf{u}}\|^2$ pointwise. We adopt this assumption in the main statement because (i) the TurboQuant MSE theorem in Zandieh et al. (2026) is stated for unit-norm vectors with norms stored separately, and (ii) in finance applications the variance scale of returns is typically known and handled outside the attention problem.

The bound extends to the case where the norm itself is quantized. If the norm is also quantized with b_0 bits to produce $\hat{\rho}$, the reconstruction is $\hat{\mathbf{x}} = \hat{\rho}\tilde{\mathbf{u}}$ and the standard inequality $\|\rho\mathbf{u} - \hat{\rho}\tilde{\mathbf{u}}\|^2 \leq 2\rho^2\|\mathbf{u} - \tilde{\mathbf{u}}\|^2 + 2(\rho - \hat{\rho})^2\|\tilde{\mathbf{u}}\|^2$ gives the looser ratio bound

$$\frac{\mathscr{L}^{RTQ}}{\mathscr{L}^{\text{sphere}}_{\text{lb}}} \;\leq\; 2C_T\chi(W) \;+\; \frac{L_\rho(b_0)}{\lambda_{\min}(W)\,\mathbb{E}[\rho^2]\,4^{-b}}, \qquad L_\rho(b_0) \equiv 2\lambda_{\max}(W)\cdot\mathbb{E}[(\rho - \hat{\rho})^2\|\tilde{\mathbf{u}}\|^2],$$

where the factor of two reflects the loose triangle inequality. Under bounded support and high-rate conditions on ρ, standard scalar quantization gives $L_\rho(b_0) \to 0$ as $b_0 \to \infty$ at a distribution-dependent rate; finite second moments alone do not yield a universal 2^{-2b_0} rate. The main statement uses the exact-norm case, which matches Zandieh et al. (2026) directly.

The proof above uses only the trivial eigenvalue bound $\mathbf{v}^\top W\mathbf{v} \leq \lambda_{\max}(W)$ in Step 2. Lemma IA.1 (Internet Appendix) sharpens this in expectation over R: the diagonal entries of $RWR^\top$ concentrate around $\text{tr}(W)/d$, so the worst-case factor $\lambda_{\max}(W)$ is conservative for typical realizations of R. It is not used in the worst-case bound.

A.3 Proof of Proposition 2 (Selective attention with hard exclusion)

Proof. The proposition assumes a basis in which W and the prior covariance Σ are jointly diagonal: $W = \text{diag}(\omega_1, \ldots, \omega_d)$ with $\omega_1 \geq \cdots \geq \omega_d > 0$, and $\text{Cov}(x_j, x_\ell) = 0$ for $j \neq \ell$ with $\text{Var}(x_j) = \sigma_j^2$. Under hard exclusion, the agent acquires signals about a chosen monitored set $S \subset \{1, \ldots, d\}$ with $|S| = k$ and uses the prior mean as the action on the orthogonal complement, without inferring information across directions.

For each $j \notin S$, hard exclusion sets the action to the prior mean $\mathbb{E}[x_j]$. This is a restriction on the strategy. It coincides with the Bayes-optimal action when x_j is independent of the monitored coordinates, as with a normal prior, but zero correlation alone does not imply independence. The loss on direction $j \notin S$ is therefore

$$\mathbb{E}[\omega_j(a_j - x_j)^2] \;=\; \omega_j \cdot \text{Var}(x_j) \;=\; \omega_j\sigma_j^2. \tag{26}$$

The contribution from monitored directions is non-negative. Summing over the unmonitored set,

$$\mathscr{L}^S \;\geq\; \sum_{j \notin S} \omega_j\sigma_j^2, \tag{27}$$

which is the bound stated in the proposition.

To minimize the right-hand side over S with $|S| = k$, choose the k directions with the largest welfare exposure $\omega_j\sigma_j^2$, which are not necessarily those with the largest ω_j. The two rankings

coincide when σ_j^2 is constant across directions.

The bound (27) is tight: as $\kappa \to \infty$, the loss on the monitored directions vanishes and the strategy attains the floor. □

Combined with Proposition 1, the bound on $\mathscr{L}^S$ characterizes the ratio $\mathscr{L}^S/\mathscr{L}^{RTQ}$ in two limits. We use the exact-norm RTQ bound; norm quantization adds a residual $L_\rho(b_0)$ that vanishes under standard high-rate conditions.

The welfare ratio diverges with capacity. Combining the floor in (27) with the RTQ upper bound from Proposition 1,

$$\frac{\mathscr{L}^S}{\mathscr{L}^{RTQ}} \geq \frac{\sum_{j=k+1}^{d} \omega_j \sigma_j^2}{\lambda_{\max}(W) \cdot \mathbb{E}[\rho^2] \cdot C_T \cdot 4^{-b}}.$$

With d and k fixed, the numerator is constant and the denominator vanishes as $b = \kappa/d \to \infty$, so $\mathscr{L}^S/\mathscr{L}^{RTQ} \to \infty$: the selective loss is floored by the tail while RTQ's loss vanishes. This is the sense in which selective attention becomes arbitrarily worse than RTQ as capacity grows.

Consider next the large-d regime at fixed per-coordinate budget $b = \kappa/d$. The RTQ total loss grows at most as $\bar{\omega} \cdot d \cdot \mathrm{MSE}(b)$ with $\bar{\omega} = \mathrm{tr}(W)/d$, while selective loss on the ignored dimensions grows as $(d-k)\bar{\omega}\,\bar{\sigma}^2$. The ratio approaches $\bar{\sigma}^2/\mathrm{MSE}(b)$ as $d \to \infty$, a large constant when $b \geq 1$, though it does not diverge in d alone at fixed b. In insurance terms, selective attention leaves $d-k$ dimensions uninsured at the full prior loss σ_j^2, while RTQ insures every dimension at b bits and reduces the loss in each to $\mathrm{MSE}(b)$.

A.4 Proof of Proposition 3 (High-Resolution Water-Filling)

Proof. Write $\delta = \mathbf{r} - \hat{\mathbf{r}}$ for the per-coordinate quantization error in the rotated frame. By the loss decomposition in the proof of Proposition 1 and expanding the quadratic form,

$$\mathscr{L}^{RTQ}(b_1,\ldots,b_d) - \mathscr{L}^{\mathrm{norm}}(b_0) = \mathbb{E}[\rho^2] \cdot \mathbb{E}[\delta^\top R W R^\top \delta] = \mathbb{E}[\rho^2] \sum_{i=1}^{d} (RWR^\top)_{ii}\, \mathbb{E}[\delta_i^2] + \text{cross terms}.$$

Under the high-resolution approximation, the cross-terms $\mathbb{E}[\delta_i \delta_j]$ for $i \neq j$ are assumed negligible: scalar Lloyd-Max quantization errors on the rotated coordinates are approximately mean-zero and have cross-covariances that vanish at rate $o(4^{-b})$ relative to the diagonal MSE. The standard justification is that the rotated coordinates have approximately matched Beta marginals (Lemma IA.1) and each coordinate is quantized separately, so the cross-covariances of the errors are $O(4^{-b}/d)$, small relative to the diagonal terms. Subtractive dithering at the encoder makes them exactly zero if a fully rigorous statement is required; we use the standard high-resolution approximation, which is

asymptotically tight. Under the high-resolution Lloyd-Max model $\mathrm{MSE}_i(b_i) = \alpha_d \cdot 2^{-2b_i}$ (asymptotically tight as $b \to \infty$, Cover and Thomas, 2006, Ch. 10; α_d from the Panter-Dite formula applied to the Beta density), the agent's problem becomes

$$\min_{b_1,\ldots,b_d \geq 0} \sum_{i=1}^{d} w_i \alpha_d 2^{-2b_i} \quad \text{s.t.} \quad \sum_{i=1}^{d} b_i \leq \kappa - b_0, \tag{28}$$

with $w_i = (RWR^\top)_{ii}$. The objective is a sum of strictly convex functions and the constraint is linear, so KKT is necessary and sufficient. The interior first-order condition $2\ln 2\, w_i \alpha_d 2^{-2b_i^*} = \nu$ gives

$$b_i^* \;=\; \left[\tfrac{1}{2}\log_2(w_i/\nu)\right]^+,$$

after absorbing $2\ln 2\,\alpha_d$ into the redefined multiplier ν and enforcing non-negativity. The threshold ν is chosen so $\sum_i b_i^* = \kappa - b_0$; existence and uniqueness follow because $B(\nu) = \sum_i [\tfrac{1}{2}\log_2(w_i/\nu)]^+$ is continuous and strictly decreasing on $(0, \max_i w_i]$ with $B(\nu) \to \infty$ as $\nu \to 0^+$ and $B(\nu) = 0$ for $\nu \geq \max_i w_i$. □

Water-filling equalizes marginal returns. The reduction in loss from one more bit on coordinate i is proportional to $w_i \cdot 2^{-2b_i}$, which rises with the cost of errors in that coordinate and falls with the bits already allocated. At the optimum this marginal return is equal across all coordinates that receive attention. Coordinates with high w_i receive more bits, and coordinates with w_i below the threshold ν receive none, because their marginal return at $b_i = 0$ is already below the common level.

Internet Appendix

This internet appendix contains additional results for the paper titled "Focus or Spread? Attention Allocation under High-Dimensional Uncertainty". Throughout, $\mathscr{L}_{\text{lb}}^{\text{sphere}}$ and the phrase "minimax floor" refer to the reference level 4^{-b} per unit of squared length defined in Proposition 1. The worst-case lower bound for codes with the same number of bits equals this level divided by the factor $\gamma_d(b)$ of Lemma IA.7, which is within a few percent of one in the dimensions used here, so each statement that RTQ is within a constant of the floor holds with that constant multiplied by $\gamma_d(b)$.

Contents

IA.1 S&P 500 portfolio application **57**
IA.1.1 Setup . . . 57
IA.1.2 S&P 500 implementation . . . 58
IA.1.3 Welfare in basis points per year . . . 60
IA.1.4 Comparison with the classical rational-inattention solution . . . 60
IA.1.5 Why high dimensions matter . . . 63
IA.1.6 Auxiliary concentration lemma . . . 64
IA.1.7 Quantitative bounds for canonical settings . . . 64

IA.2 Parallel-target theorems and regime diagram **64**
IA.2.1 Setup and reference strategies . . . 65
IA.2.2 Spread coverage and concentration floor . . . 66
IA.2.3 Sufficient statistic, scaling, and regime diagram . . . 68

IA.3 Fisman speed-dating cross-domain test **71**
IA.3.1 Structural model and slope identity . . . 72
IA.3.2 Main regression and effective-dimension comparison . . . 74
IA.3.3 Per-coordinate bit budget and out-of-sample stability . . . 77

IA.4 Connection to Transformer-Based Portfolio Construction **79**
IA.4.1 Setup . . . 79
IA.4.2 Optimal linear-attention projection . . . 79
IA.4.3 RTQ within a constant factor of optimal linear attention . . . 80
IA.4.4 Explainability: RTQ composite signals as welfare-weighted factor portfolios . . . 81
IA.4.5 Discussion and limitations . . . 82

IA.5 Extensions **84**
IA.5.1 Equilibrium with heterogeneous attention . . . 84
IA.5.2 Heterogeneous bit budgets and risk aversion . . . 88
IA.5.3 Dynamic extension with attention revision over time . . . 90

IA.6 Technical background: from Johnson-Lindenstrauss to rotate-then-quantize **97**
IA.6.1 The Johnson-Lindenstrauss lemma . . . 97
IA.6.2 Coordinates of a randomly rotated vector . . . 98
IA.6.3 Dependence across rotated coordinates . . . 99
IA.6.4 The Lloyd-Max quantizer and the constant C_T . . . 100

IA.6.5 The lower bound 100
IA.6.6 QJL and the inner-product version of TurboQuant 102
IA.6.7 What the paper takes and what it adds 104
IA.6.8 Simulation with non-Gaussian priors: full results 104
IA.7 Additional empirical results **105**
IA.7.1 Welfare comparison 106
IA.7.2 RTQ after common factors are removed 108
IA.7.3 Venture deal flow: RTQ with many variables and a short history 110
IA.7.4 13F tests 114

IA.1 S&P 500 portfolio application

This appendix puts the theorems of Section 4 to work on one full application: portfolio choice across the S&P 500 universe. The goal is to move from theory to a welfare number an applied reader can compare with a classical rational-inattention solution. We fold in the eigenvalue diagnostic, the selective-versus-RTQ loss-ratio profile, and a head-to-head comparison with Van Nieuwerburgh and Veldkamp (2010). VNV is the optimum within the Gaussian-restricted signal class and remains the right answer in that class; we evaluate it here outside its design assumptions, alongside RTQ, as the canonical reference for high-dimensional portfolio attention. The exercise is not a horse race between a true and a false model. It is a comparison of two feasible signal structures on the same welfare metric. The full CRSP universe extension of this exercise is in Section 5 of the main text.

IA.1.1 Setup

An investor allocates wealth across d risky assets. Excess returns $\mathbf{x}$ have mean μ and covariance Σ. Under full information and mean-variance preferences with risk aversion $\gamma > 0$, the optimal portfolio is $\mathbf{a}^* = \Sigma^{-1}\mu/\gamma$. With a finite attention budget, the investor observes $\mathbf{x}$ through a noisy internal signal and chooses $\mathbf{a}$ on the posterior. A second-order expansion of the utility shortfall gives a quadratic loss with $W = \gamma\Sigma$. In the eigenbasis of Σ, expected welfare loss reduces to $\mathbb{E}[\mathscr{L}] \leq C\sum_j \lambda_j^2 2^{-2b_j}$ with $C = \pi\sqrt{3}/2$ and per-direction weights $c_j = \lambda_j^2$. We compare two strategies that use the same budget differently.

Under RTQ, the investor rotates into the eigenbasis of W and allocates bits across the d rotated coordinates by the high-resolution water-filling rule of Proposition 3. This is the RTQ allocation of Section 5: it uses the covariance, through the eigenbasis of W, and no random rotation, and the distribution-free guarantee applies when the allocation is combined with a random rotation. Expected loss is within $C_T \cdot \chi(W)$ of the spherical/minimax rate-distortion lower bound (Proposition 1), with $C_T = \pi\sqrt{3}/2 \approx 2.72$ matching the TurboQuant MSE constant and $\chi(W) = \lambda_{\max}(W)/\lambda_{\min}(W)$ the condition number of W. The bound holds for any distribution of $\mathbf{x}$ with finite second moments and any d.

The classical alternative is selective attention. Van Nieuwerburgh and Veldkamp (2010) solve the Gaussian rational-inattention problem for the same objective and find a corner solution: diagonalize Σ, rank rotated coordinates by Sharpe-weighted stakes, and pour all capacity into the top k directions. Under the Shannon constraint, the Gaussian posterior is sharp on monitored directions and equals the prior on the rest. The selective investor pays full prior variance on the $d-k$ ignored directions, the floor from Proposition 2.

For each strategy s, let $\mathscr{L}^s(\kappa) \equiv \mathbb{E}[(\hat{\mathbf{a}}^s - \mathbf{a}^*)^\top \Sigma (\hat{\mathbf{a}}^s - \mathbf{a}^*)]$ denote the expected squared tracking

error at capacity κ, where $\hat{\mathbf{a}}^s$ is the strategy's portfolio and $\mathbf{a}^* = \Sigma^{-1}\mu/\gamma$ is the full-information mean-variance optimum. We report the welfare cost in basis points per year via

$$\text{CE-bps/year} \;=\; \frac{\gamma \mathscr{L}^s}{2\mu_p} \cdot 10^4, \tag{IA.1}$$

where μ_p is the expected annual return on the full-information portfolio. The factor $\gamma\mathscr{L}^s/2$ is the certainty-equivalent welfare loss in return units, obtained from the second-order expansion of mean-variance utility around the full-information optimum; dividing by μ_p and multiplying by 10^4 expresses it as a fraction of the equity premium in basis points per year. The conversion does not require estimating $\mathbf{a}^*$ in sample and sidesteps the plug-in Markowitz pathology. Baseline: $\gamma = 5$ and $\mu_p = 10\%$.

IA.1.2 S&P 500 implementation

We use monthly returns for the top 500 CRSP common stocks by market capitalization, our S&P 500 proxy. The proxy is rebalanced each month, so a stock enters the universe when it sits in the top 500 by capitalization and exits otherwise. Building the universe from CRSP itself avoids the survivorship bias of pulling current S&P 500 constituents back in time and uses the same data vendor as the rest of the paper.[23] The sample covers 1970–2026 (675 months). For the static cross-section we keep stocks present in the top 500 for at least 240 of those months, leaving $d = 488$. We estimate Σ with Ledoit-Wolf shrinkage (Ledoit and Wolf, 2004); the rolling out-of-sample exercise below uses 60-month rolling windows with quarterly stepping.

The loss concentration ratio $\rho(k) = \sum_{j\le k}\lambda_j^2 / \sum_{j\le d}\lambda_j^2$ on the full-sample covariance gives $\rho(1) = 0.80$ at the market factor and $\rho(5) = 0.95$. Figure IA.1 plots the full profile.

Standard PCA tells a different story: the first principal component explains only 20 percent of return variance, and cumulative PCA does not reach 0.95 until k is well into the spectrum. The λ_j^2 weighting in $\rho(k)$ counts large eigenvalues doubly because each eigendirection enters portfolio loss through two channels (prior uncertainty proportional to λ_j and a cost of error also proportional to λ_j), which compresses the effective dimension far below what PCA indicates. Figure IA.2 plots the full eigenvalue spectrum side by side with the cumulative PCA and ρ curves.

[23] The proxy is built from the CRSP monthly file rather than a third-party feed so that the universe is reproducible after publication, in contrast to a Yahoo Finance pull that changes silently between extracts. Replication code will be made available on publication.

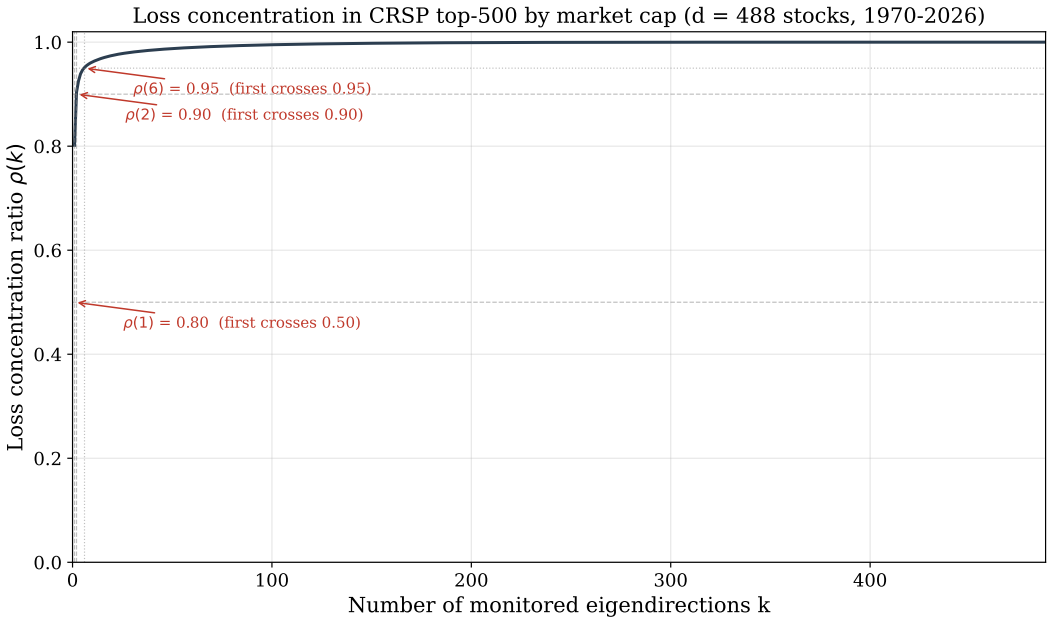

Figure IA.1: **Loss concentration ratio for the S&P 500 proxy.** $\rho(k) = \sum_{j \le k} \lambda_j^2 / \sum_{j \le d} \lambda_j^2$ measures the loss-weighted share of prior uncertainty in the top k eigendirections, where λ_j are eigenvalues of the monthly-return covariance. Inputs: top 500 CRSP commons by market cap, monthly returns from January 1970 to March 2026, restricted to stocks present in the top 500 for at least 240 of those months ($d = 488$). Horizontal axis: k. Vertical axis: $\rho(k)$. $\rho(1) = 0.80$ at the market factor; $\rho(5) = 0.95$. The first principal component explains only 20 percent of return variance, but the λ_j^2 weighting concentrates 80 percent of portfolio loss in the market factor.

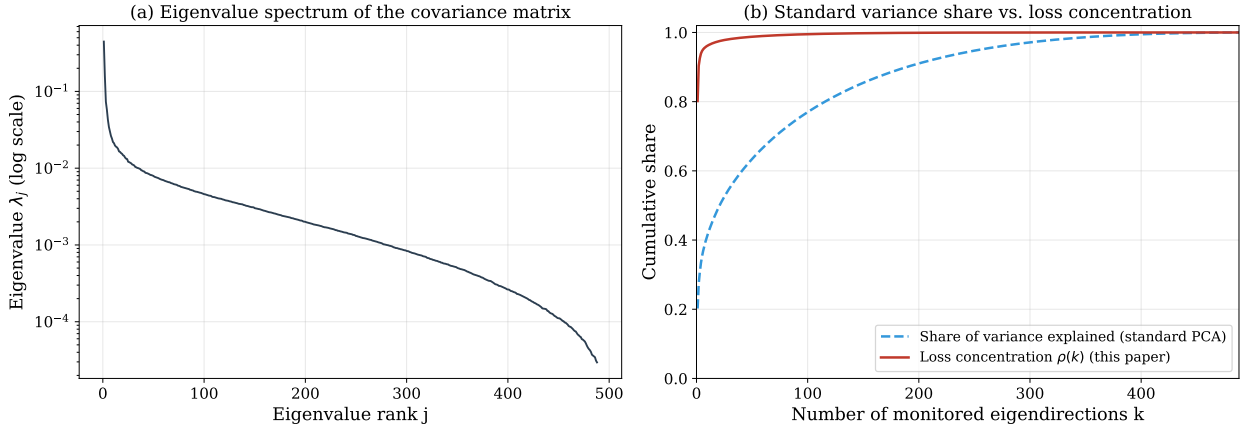

Figure IA.2: **Eigenvalue spectrum and cumulative concentration for the S&P 500 proxy covariance.** Inputs: top 500 CRSP commons by market cap, monthly returns 1970–2026 ($d = 488$, $T = 675$ months). Left: eigenvalues λ_j on log scale versus rank j. Right: PCA cumulative variance $\sum_{j \le k} \lambda_j / \sum_j \lambda_j$ (dashed) and loss concentration $\rho(k) = \sum_{j \le k} \lambda_j^2 / \sum_j \lambda_j^2$ (solid). PCA reaches 0.95 only deep into the spectrum; ρ crosses 0.95 at $k = 6$ (1.2 percent of d). The visual gap between the two curves shows how the λ_j^2 weighting compresses effective dimension below what PCA suggests.

IA.1.3 Welfare in basis points per year

Table IA.1 translates the loss numbers into basis points per year at $\gamma = 5$ and $\mu_p = 10\%$ using (IA.1). Selective-k losses are almost flat in b once the budget saturates the k monitored directions; the floor from ignored directions is all that remains. RTQ losses keep falling because water-filling shifts bits to tail directions at no marginal cost to the top. This decline is the high-rate regime where the Panter-Dite rate-distortion approximation holds; at very low per-coordinate rates the approximation breaks down, as the robustness discussion below makes explicit.

Strategy	$b = 0.5$	$b = 1$	$b = 2$	$b = 3$	$b = 4$
Selective, $k = 1$ (VNV corner)	198.20	198.20	198.20	198.20	198.20
Selective, $k = 5$	53.59	53.59	53.59	53.59	53.59
Selective, $k = 20$	25.48	25.48	25.48	25.48	25.48
RTQ (water-filling)	4.03	2.02	0.50	0.13	0.03
VNV / RTQ	49	98	393	1,572	6,293

Table IA.1: **Annual welfare cost in basis points of the equity premium, S&P 500 proxy.** Each cell reports the certainty-equivalent cost CE cost $= \gamma \cdot L/(2\mu_p) \cdot 10{,}000$ in bps per year, where L is the expected squared tracking error, $\gamma = 5$ is relative risk aversion, and $\mu_p = 10\%$ is the annual equity premium. Inputs: top 500 CRSP commons by market cap, monthly-return sample covariance, 1970–2026 ($d = 488$, $T = 675$ months). Columns: $b = \kappa/d$. Rows: selective at $k \in \{1, 5, 20\}$, RTQ, VNV/RTQ. Selective flattens once extra bits stop improving the k monitored directions. RTQ keeps falling within the high-rate range where the rate-distortion approximation is accurate. The last row shows VNV is two orders of magnitude worse than RTQ at $b = 1$ and four orders of magnitude worse at $b = 4$.

At $b = 2$, the VNV corner leaves 198 bps per year on the table relative to RTQ's 0.5 bps, a 198 bps gap.[24] The gap grows with the budget because Proposition 2 pins the VNV floor at the tail sum $\sum_{j>1} \lambda_j^2$ that extra capacity cannot remove, while RTQ spends the marginal bit on another tail direction. Figure IA.3 traces this divergence across budgets and across k.

IA.1.4 Comparison with the classical rational-inattention solution

Solving VNV and RTQ for subsets of the proxy at $N \in \{25, 50, 100, 250, 488\}$ and fixed budget $\kappa = 2N$ shows the gap is positive at every N, with VNV's cost flat as the corner captures the same fraction of risk regardless of N and RTQ's cost falling as more directions fit under the budget. The gap is 357 bps at $N = 25$, 260 at $N = 100$, and 198 at $N = 488$. Figure IA.4 plots the full scaling.

The corner solution does best in the large, market-dominated limit where $\rho(1)$ approaches 1; even there, the welfare cost difference is 198 bps per year. Rolling through 1975–2026 in 60-month

[24] Compounded over 30 years, 198 bps annually costs roughly 45 percent of lifetime consumption.

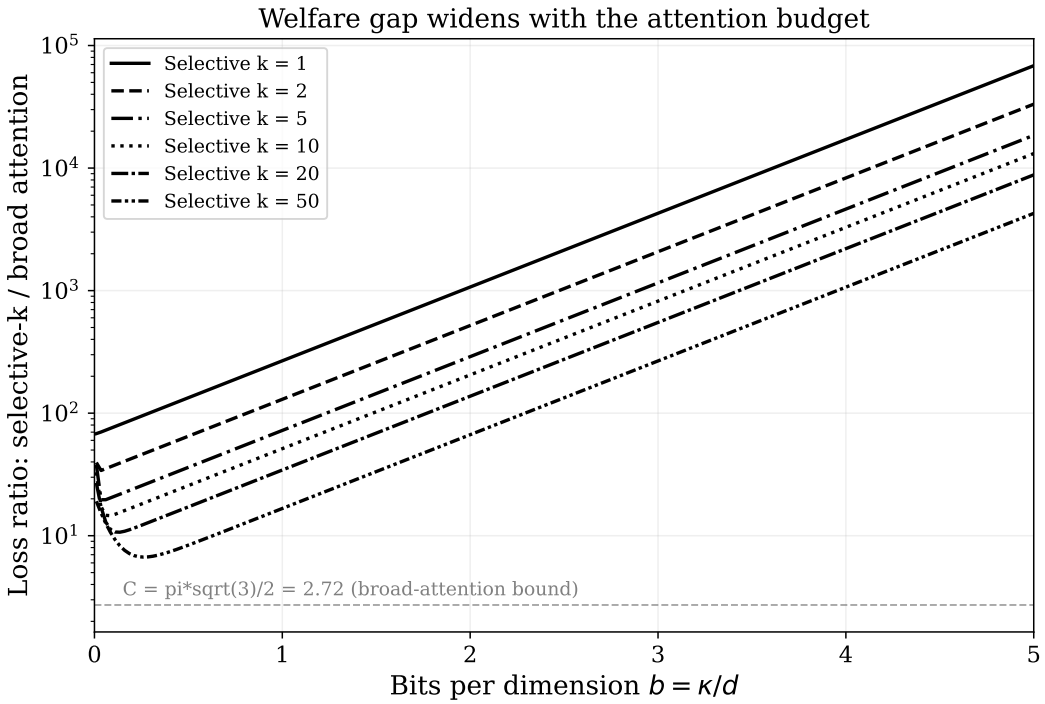

Figure IA.3: **Selective-k loss relative to RTQ loss, S&P 500 proxy.** Inputs: top 500 CRSP commons by market cap, monthly returns 1970–2026 ($d = 488$). Horizontal axis: $b = \kappa/d$ (log scale). Vertical axis: L^{S_k}/L^{RTQ} (log scale), from closed-form loss expressions with the sample covariance. Each curve fixes $k \in \{1, 5, 20, 50\}$ under the selective strategy. Every selective curve rises without bound as b grows; RTQ stays within $C \approx 2.7$ of the Shannon bound (gray dashed). At $k = 1$, $b = 2$ (VNV corner), selective is 393 times RTQ; at $b = 4$, over 6,000 times.

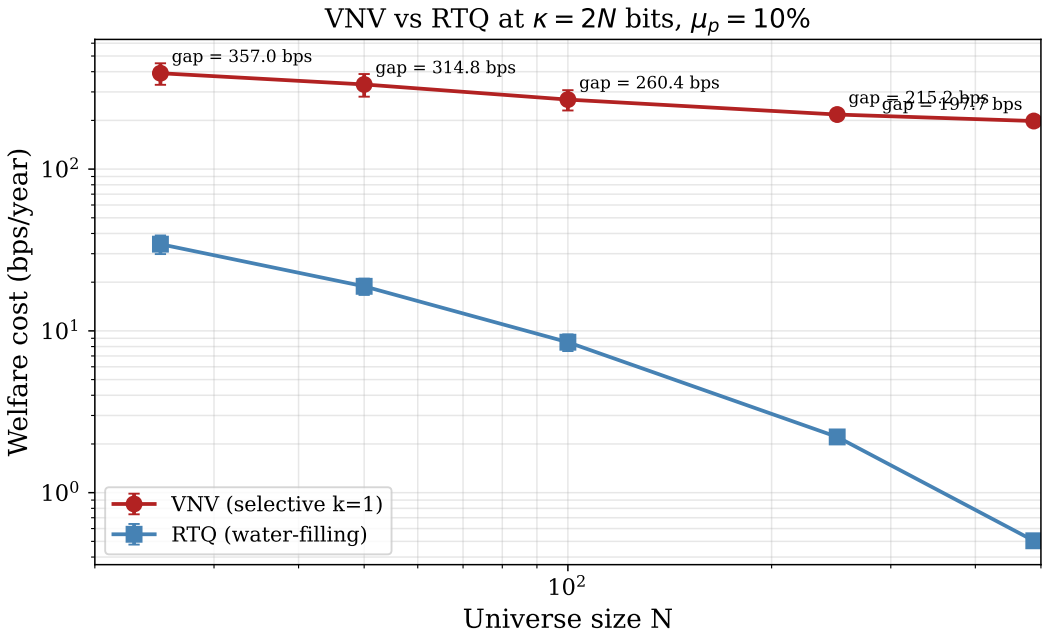

Figure IA.4: **VNV versus RTQ welfare cost across universe size, S&P 500 proxy subsets.** Horizontal axis: $N \in \{25, 50, 100, 250, 488\}$. Vertical axis: welfare cost in bps per year (CE metric from Table IA.1). Both strategies use budget $\kappa = 2N$. VNV: selective-$k = 1$ corner; RTQ: water-filling. Each point averages 30 random subsets drawn from the proxy covariance, 1970–2026 ($d = 488$). The gap falls from 357 bps at $N = 25$ to 198 bps at $N = 488$, and remains positive at every N. The gap is largest in small subsets where no single factor dominates and narrows as N grows and the market captures more loss-weighted uncertainty.

windows with Ledoit-Wolf shrinkage and quarterly re-estimation gives an average gap of 163 bps per year across 205 overlapping quarterly windows, with the gap positive in every window. These numbers are in sample: the allocation and the loss use the same covariance. Panel C of Table 1 reports an out-of-sample version on the 500 largest stocks at each date, with losses evaluated on the next 24 months; there the median gain over the concentrated benchmark is 42 bps per year at $\kappa = 2N$, and the gain is again positive in every quarter.[25][26]

IA.1.5 Why high dimensions matter

Section 4.3 of the main text states the headline implication of Proposition 2 in compressed form. This subsection develops the intuition at the length the argument deserves, using the S&P 500 calibration above as the running example.

Conventional rational inattention uses selective attention: monitor the top k state variables and ignore the rest. This is sound when d is small. When $d = 500$ and $k = 10$, the agent bears full prior uncertainty in 490 dimensions, and even small per-dimension loss sums to a large total. Proposition 2 makes this precise: "concentrate and ignore" is not just suboptimal but *increasingly* so as d grows.

Think of it as insurance. Selective insures k houses and leaves the rest uninsured; each ignored dimension contributes its full prior loss $\omega_j \sigma_j^2$. RTQ spreads a thin layer of coverage across every house. Per-house coverage is modest, but total expected loss stays controlled. As capacity grows with k fixed, RTQ loss falls while selective loss is floored by the tail. In high dimensions, broad thin coverage dominates narrow deep coverage.

Three statements summarize the role of dimension under RTQ. First, the worst-case bound of Proposition 1 is dimension-free: the constant $C_T \chi(W)$ does not depend on d. Second, the average-case rotation argument in Lemma IA.1 homogenizes the diagonal entries of W through concentration on the sphere and tightens as d grows. Third, the welfare comparison against hard-exclusion selective attention becomes more favorable as the ignored tail grows: with $k = 5$, moving from $d = 10$ to $d = 500$ raises ignored dimensions from 5 to 495, and the tail loss scales with d. Proposition 2 considers fixed k with total capacity κ growing, where the ratio of selective to RTQ loss diverges because selective is floored while RTQ loss decays.

[25] Median 124 bps, IQR 88 to 224, range 48 to 499. The gap is larger in high-volatility episodes (early 1980s, 2002, 2008–2009, 2020) because the eigenvalue tail flattens when idiosyncratic risk rises, widening the VNV floor. The window count of 205 is large because the 56-year sample covers many more rolling windows than a shorter post-2000 sample would.

[26] VNV is a formal solution to rational inattention under Shannon costs. That RTQ delivers lower loss does not mean VNV is wrong on its own terms; it means the Shannon cost, combined with corner-solution tractability of the Gaussian case, points applied work toward concentrated allocations that are a poor fit for the high-dimensional limit. Relaxing the tractability constraint, which RTQ does, reveals a different prescription for the same problem.

IA.1.6 Auxiliary concentration lemma

The high-dimensions intuition above invokes a concentration argument: a Haar-random rotation homogenizes the diagonal entries of W. The formal statement is below. The lemma is auxiliary and is not used in any worst-case bound; the main-appendix proofs cite it only for intuition about why the symmetric strategy is close to optimal on typical realizations of the rotation.

Lemma IA.1 (Concentration of rotated quadratic forms)**.** *Let W be a $d \times d$ symmetric positive definite matrix with eigenvalues $\lambda_{\max}(W) \geq \cdots \geq \lambda_{\min}(W) > 0$, and let R be Haar-uniform on $O(d)$. For any standard basis vector $\mathbf{e}_i$,*

$$\mathbb{E}\left[\mathbf{e}_i^\top RWR^\top \mathbf{e}_i\right] = \frac{\operatorname{tr}(W)}{d}, \quad \Pr\left[\left|\mathbf{e}_i^\top RWR^\top \mathbf{e}_i - \tfrac{\operatorname{tr}(W)}{d}\right| > \delta \cdot \tfrac{\operatorname{tr}(W)}{d}\right] \leq 2\exp\left(-\tfrac{c\,d\,\delta^2}{\chi(W)^2}\right),$$

where $\chi(W) = \lambda_{\max}(W)/\lambda_{\min}(W)$ and $c > 0$ is a universal constant.

Proof. $\mathbf{v} = R^\top \mathbf{e}_i$ is uniform on S^{d-1}, so $\mathbb{E}[\mathbf{v}\mathbf{v}^\top] = I_d/d$ and $\mathbb{E}[\mathbf{v}^\top W\mathbf{v}] = \operatorname{tr}(W)/d$. The function $f(\mathbf{v}) = \mathbf{v}^\top W\mathbf{v}$ is Lipschitz with constant $L = 2\lambda_{\max}(W)$ on the sphere, so Levy's inequality (Vershynin, 2018, Theorem 3.4.6) gives $\Pr[|f - \mathbb{E}f| > t] \leq 2\exp(-c\,d\,t^2/L^2)$. Setting $t = \delta \operatorname{tr}(W)/d$ and using $\operatorname{tr}(W) \geq d\,\lambda_{\min}(W)$ produces the stated bound. □

IA.1.7 Quantitative bounds for canonical settings

Section 4.2 of the main text states the bound $\mathscr{L}^{RTQ}/\mathscr{L}^{\text{sphere}}_{\text{lb}} \leq C_T \cdot \chi(W)$. This subsection computes the constant for three representative settings using back-of-the-envelope condition numbers drawn from typical empirical magnitudes rather than specific estimates. *Well-diversified portfolio* (500 stocks, $W = \gamma\Sigma$): $\chi(W) \approx 5$ to 20, so with $\chi(W) = 10$ the bound is $27 \cdot \mathscr{L}^{\text{sphere}}_{\text{lb}}$. *Multi-product pricing with substitutes* (200 products, W the Hessian of profit): cross-price elasticities of 2 to 3 give $\chi(W) \approx 8$ to 15, so with $\chi(W) = 12$ the bound is $32 \cdot \mathscr{L}^{\text{sphere}}_{\text{lb}}$. *Monetary policy with multiple instruments* ($m = 3$ rates monitoring $d = 50$ indicators): $\chi(W) \approx 5$ to 10, so with $\chi(W) = 7$ the bound is $19 \cdot \mathscr{L}^{\text{sphere}}_{\text{lb}}$. Across these settings, the worst-case envelope is 19 to 33 times the spherical floor; the bound depends only on $\chi(W)$, not on d. Realized performance routinely beats this envelope because real priors are not adversarial.

IA.2 Parallel-target theorems and regime diagram

This section develops the theory for the parallel-target setting: one agent estimating N multi-dimensional targets at once under a shared bit budget. The single-agent propositions of Section 4 of the main text have direct counterparts here, sharper because every target carries the same stakes,

so $\chi(W) = 1$ and drops out. The three parameters (N, d, B) collapse to a single scaling variable. The Fisman cross-domain test in Section IA.3 below uses the spread theorem and the effective-dimension reading that follow.

IA.2.1 Setup and reference strategies

The single-agent setup in the main text covers one agent facing one high-dimensional decision. Many economic settings have a different shape: one agent estimates many targets at once, each itself multi-dimensional, splitting a single shared attention budget across both layers. A participant at a speed-dating event forms an estimate of each of N partners, each described by a d-dimensional attribute vector (kindness, intelligence, humor, shared interests). A trader prices N securities, each with multiple risk characteristics. A central bank analyst forecasts N macro indicators, each with multiple underlying drivers. The action is one estimate per target, and loss sums over targets.

We use the same mathematical structure. Index targets by $i \in \{1, \ldots, N\}$. Each has attribute vector $x_i \in \mathbb{R}^d$ with unit Euclidean norm and prior variance $\sigma^2 = \mathbb{E}[\|x_i - \bar{x}\|_2^2]$. The agent produces estimates $\{\hat{x}_i\}$ with loss

$$L \;=\; \mathbb{E}\left[\sum_{i=1}^{N} \|x_i - \hat{x}_i\|_2^2\right] \tag{IA.2}$$

and total bit budget B to spread across the N targets. Two reference strategies span the space of reasonable allocations:

Concentration strategy. Choose a subset $S \subseteq \{1, \ldots, N\}$ of size k. Allocate B/k bits to each target in S, zero to targets outside S. For $i \in S$, estimate $\hat{x}_i$ with a bit-budget-constrained Lloyd-Max quantizer. For $i \notin S$, set $\hat{x}_i = \bar{x}$. This is the multi-target analog of selective attention.

Spread-coverage strategy (RTQ). Stack the N vectors into $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^{Nd}$. Draw a random orthogonal matrix $\Pi \in \mathbb{R}^{Nd \times Nd}$ (Haar-uniform). Compute $\mathbf{y} = \Pi\mathbf{x}$, quantize each coordinate with a Lloyd-Max scalar quantizer at $b = B/(Nd)$ bits, and decode via $\hat{\mathbf{x}} = \Pi^\top Q(\mathbf{y})$.

Both use the same total bit budget and differ only in how it is spread. Concentration places all attention on k targets and ignores the rest. Spread places the same total attention across all N at coarse resolution.

The parallel-target setup is the single-agent setup with a block-diagonal loss matrix. Taking $W = I_{Nd}$ and stacking the targets into a single Nd-dimensional state makes (IA.2) a special case of (6). Concentration corresponds to selective attention on kd of the Nd stacked coordinates; spread corresponds to RTQ on the stacked vector. Every single-agent theorem therefore has a parallel-target counterpart. The counterparts are sharper because the equal-weight block structure eliminates constants that appear in the general case. The condition number $\chi(W)$ drops out because

$W = I$, and the loss ratio reduces to a single summary number derived below.[27]

IA.2.2 Spread coverage and concentration floor

Theorem IA.1 (Spread coverage against the spherical/minimax benchmark)**.** *Stack the N attribute vectors $\{x_i\}_{i=1}^N$ with $\|x_i\|_2 = 1$ into $\mathbf{x} \in \mathbb{R}^{Nd}$. The spread-coverage strategy applies a Haar-random rotation in $\mathbb{R}^{Nd}$ and quantizes each rotated coordinate at $b = B/(Nd)$ bits per coordinate. For any bit budget B,*

$$L^{\text{spread}} \;\le\; \frac{\pi\sqrt{3}}{2} \cdot N \cdot 4^{-b} \;\approx\; 2.72 \cdot N \cdot 4^{-b}.$$

The worst-case lower bound for any quantizer of a unit vector in $\mathbb{R}^{Nd}$ with the same number of bits is $N \cdot 4^{-b}/\gamma^c_{Nd}(b)$ (Lemma IA.6, rescaled to squared length N). The spread loss is therefore at most $(\pi\sqrt{3}/2)\,\gamma^c_{Nd}(b)$ times this lower bound, for any N, d, and realized vectors $\{x_i\}$, and $\gamma^c_{Nd}(b) \to 1$ as Nd grows.

Proof. The proof scales TurboQuant's worst-case unit-sphere MSE bound directly, without going through a Gaussian / Poincare approximation.

Stack the attribute vectors into $\mathbf{X} = (x_1^\top, \ldots, x_N^\top)^\top \in \mathbb{R}^{Nd}$. Since each $\|x_i\| = 1$, $\|\mathbf{X}\|^2 = N$. Define the unit-norm vector

$$\mathbf{z} \;\equiv\; \mathbf{X}/\sqrt{N} \;\in\; S^{Nd-1}. \tag{IA.3}$$

The spread-coverage strategy applies TurboQuant in dimension Nd with $b = B/(Nd)$ bits per coordinate to $\mathbf{z}$. By Lemma 2 (Theorem 3.1 of Zandieh et al. (2026)), for the worst-case unit input $\mathbf{z} \in S^{Nd-1}$,

$$\mathbb{E}\|\mathbf{z} - \hat{\mathbf{z}}\|^2 \;\le\; C_T \cdot 4^{-b}, \qquad C_T \;=\; \pi\sqrt{3}/2. \tag{IA.4}$$

Rescale: $\hat{\mathbf{X}} = \sqrt{N}\,\hat{\mathbf{z}}$. Then

$$L^{\text{spread}} \;=\; \mathbb{E}\|\mathbf{X} - \hat{\mathbf{X}}\|^2 \;=\; N \cdot \mathbb{E}\|\mathbf{z} - \hat{\mathbf{z}}\|^2 \;\le\; C_T \cdot N \cdot 4^{-b}. \tag{IA.5}$$

This is the upper bound. The argument is exact: it does not invoke a high-rate Gaussian approximation, and it does not require Poincare convergence. The constant C_T is the universal TurboQuant constant from Zandieh et al. (2026), and the bound holds for any realized $\{x_i\}$ on the unit sphere.

[27] The attribute dimension d has a specific interpretation in parallel-target applications. In speed-dating, d is the number of distinct traits the agent cares about (core trait ratings plus activity categories, typically on the order of twenty). When traits are correlated, the effective dimension of the trait covariance can be far smaller than nominal d; an empirical recovery using the slope test of Section IA.3 bears this out. In compressed sensing, d is the signal dimension. In the population-coding application of Fairhall et al. (2012), d is the number of stimulus features a single neuron encodes. Higher d strengthens the spread advantage: the concentration floor scales linearly with target count, while spread's per-coordinate precision scales only with $B/(Nd)$, so the gap widens as either N or d grows.

For the lower bound, Lemma IA.6 applied to $\mathbf{z} \in S^{Nd-1}$ at b bits per coordinate says any quantizer has a hard input direction with MSE at least $4^{-b}/\gamma_{Nd}^{c}(b)$; in the large-dimension limit this is

$$\inf_{Q} \sup_{\mathbf{z} \in S^{Nd-1}} \mathbb{E}\|\mathbf{z} - \hat{\mathbf{z}}\|^2 \;\geq\; 4^{-b}.$$

Rescaling by N, the worst-case spread loss is bounded below by

$$L_{\text{lb}}^{\text{sphere}} \;\geq\; N \cdot 4^{-b}.$$

Combining,

$$\frac{L^{\text{spread}}}{L_{\text{lb}}^{\text{sphere}}} \;\leq\; C_T \;=\; \frac{\pi\sqrt{3}}{2},$$

uniformly in N, d, and the realized vectors $\{x_i\}$. The bound is relative to the worst-case (spherical) lower bound, not to the prior-specific rate-distortion function for any joint distribution of $\{x_i\}$ that has exploitable structure. □

The benchmark is the spherical/minimax bound for the stacked vector in $\mathbb{R}^{Nd}$. RTQ on the stacked vector comes within $\pi\sqrt{3}/2$ of the worst-case rate-distortion floor, uniformly in N, d, and the realized inputs. The guarantee is distribution-free: no normality, no thin tails, no specific factor structure. Spread always works, and the worst-case loss relative to the spherical floor is a fixed constant that does not grow with problem size.

Theorem IA.2 (Concentration-strategy floor)**.** *For the concentration strategy with subset size k and any budget B,*

$$L^{\text{conc}} \;\geq\; (N-k)\,\sigma^2,$$

where $\sigma^2 = \mathbb{E}[\|x_i - \bar{x}\|_2^2]$. This lower bound is independent of B.

Proof. For the concentration strategy with subset S of size k, the estimate on each target $i \notin S$ is the unconditional mean $\bar{x}$. Each such target contributes $\mathbb{E}[\|x_i - \bar{x}\|_2^2] = \sigma^2$ to the expected squared error. There are $N-k$ such targets, so their total contribution is $(N-k)\sigma^2$. Reallocating bits within S does not change the contribution from targets outside S, so the bound holds independent of B. □

Each ignored target contributes its full prior variance σ^2 because the estimate on it is the unconditional mean. Reallocating bits within S does not change this. You cannot learn about a partner you never speak to, a stock you never price, or a country you never follow. The floor is independent of B, which is the key economic feature: even a well-resourced analyst with a narrow watchlist still faces this floor on everything off the list. The result is the multi-target analog of Proposition 2.

Dividing Theorem IA.2 by Theorem IA.1 gives

$$\frac{L^{\text{conc}}}{L^{\text{spread}}} \geq \frac{(N-k)\sigma^2}{(\pi\sqrt{3}/2)\cdot N\cdot 4^{-b}} \to \frac{\sigma^2}{\pi\sqrt{3}/2}\cdot 4^b \qquad (N\to\infty,\ k \text{ fixed}),$$

which grows exponentially in b. Concentration is not merely worse than spread; the welfare cost grows exponentially in the per-coordinate budget. With more capacity, the penalty for ignoring targets rises faster than the gain from studying a narrow subset more carefully.

IA.2.3 Sufficient statistic, scaling, and regime diagram

Theorems IA.1 and IA.2 compress the three-parameter problem into one summary statistic and one sufficient-condition regime diagram. Define

$$\xi \equiv N\cdot 4^{-B/(Nd)}. \tag{IA.6}$$

N is the number of targets. $4^{-B/(Nd)}$ is the per-coordinate precision of spread: each bit cuts quantization variance by four, and each coordinate receives $B/(Nd)$ bits after the total budget is spread evenly. So ξ combines problem size with per-coordinate precision.

Proposition IA.1 (Sufficient statistic for parallel-target loss; unit-norm targets, $\sigma^2=1$)**.** *Under the assumptions of Theorem IA.1 with per-target prior variance normalized to $\sigma^2=1$,*

$$L^{\text{spread}} = \frac{\pi\sqrt{3}}{2}\cdot\xi\cdot(1+o(1)),$$

where $o(1)\to 0$ as $Nd,B\to\infty$ with $b=B/(Nd)$ bounded away from zero. The general $\sigma^2\neq 1$ case rescales the loss by σ^2. All (N,d,B) triples mapping to the same ξ produce the same normalized spread loss in the large-system limit.

Proof. By Theorem IA.1, $L^{\text{spread}}/\sigma^2 \leq (\pi\sqrt{3}/2)\cdot N\cdot 4^{-b}/\sigma^2$, which at unit prior variance $\sigma^2=1$ equals $(\pi\sqrt{3}/2)\xi$ with $\xi = N\cdot 4^{-b}$. Substituting $b=B/(Nd)$ gives (IA.6). In the high-rate regime (b bounded away from zero as $Nd,B\to\infty$), the TurboQuant MSE constant is asymptotically tight by the Panter-Dite formula applied to the Beta marginal of the Haar-rotated coordinates, so the upper bound and the achievable spread loss agree up to $o(1)$. □

The three-parameter problem reduces to one parameter. An analyst studying $N=500$ stocks with $d=20$ characteristics at budget B faces the same normalized loss as an analyst studying $N=20$ targets with $d=6$ attributes at the matching ξ. The plot of loss against ξ is common across applications, so a calibration done once transfers to any other setting with the same ξ.

Figure IA.5 confirms the relationship at five target counts, three total budgets, and two attribute dimensions.

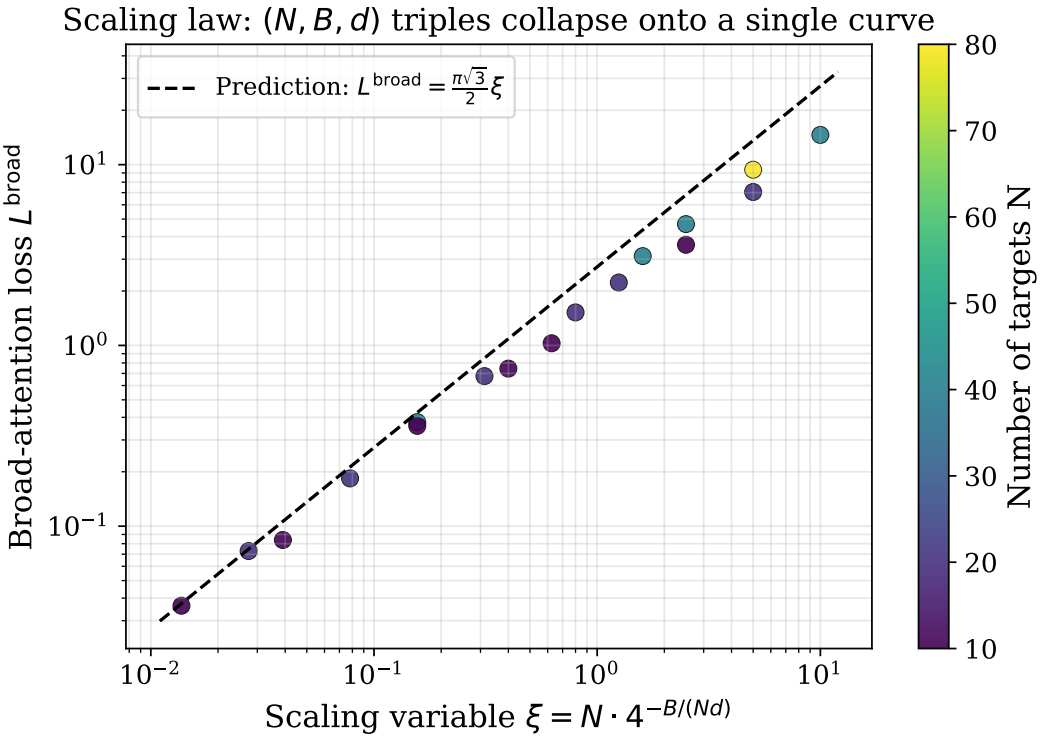

Figure IA.5: **One summary statistic governs the spread-strategy loss.** Numerical confirmation that all (N,d,B) triples producing the same ξ give the same loss. Monte Carlo simulation of the parallel-target setup. In each trial, N attribute vectors $x_i \in \mathbb{R}^d$ are drawn i.i.d. standard normal and normalized to the unit sphere; the spread (RTQ) strategy stacks them into $\mathbb{R}^{Nd}$, applies a Haar-random rotation, and scalar-quantizes each rotated coordinate with a Lloyd-Max quantizer at $K = \text{round}(2^{B/(Nd)})$ levels (capped at 256). Each circle is one (N,d,B) triple, loss averaged over 30 independent trials. Sweep: $N \in \{10,20,40,80,160\}$, $d \in \{3,5,8\}$, $B \in \{60,120,240,480\}$, restricted to triples with $b = B/(Nd) \in [1.5,5]$. Horizontal axis: $\xi = N \cdot 4^{-b}$ (log scale). Vertical axis: spread loss L^{spread} (log scale). Color: number of targets N. Dashed line: theoretical prediction $L^{\text{spread}} = (\pi\sqrt{3}/2)\,\xi$ from Proposition IA.1. Equation (IA.6) reduces the three-parameter problem to one.

The distribution-free guarantee in Theorem IA.1 is a property of the random rotation, not of the input distribution. Figure IA.6 verifies this on a simulation grid of four distributions (uniform on the sphere, heavy-tailed $t(3)$, three-mode Gaussian mixture, rank-one adversarial) crossed with three total dimensions $D = Nd \in \{64, 256, 1024\}$ and four per-coordinate bit rates $b \in \{2,3,4,5\}$, for $4 \times 3 \times 4 = 48$ simulated configurations. The empirical ratio $L^{\text{spread}}/L^{\text{sphere}}_{\text{lb}}$ stays below $\pi\sqrt{3}/2 \approx 2.72$ in every case. The rank-one adversarial input, which places all prior mass in a single direction, is the worst case for random rotation; even there, the ratio stays under 2.5 at $D = 1024$. The bound is scale-free in a directly useful sense: an upper bound on welfare loss under RTQ depends on the per-coordinate bit budget b alone, so applied researchers can size welfare loss ex ante without specifying or estimating the input distribution.

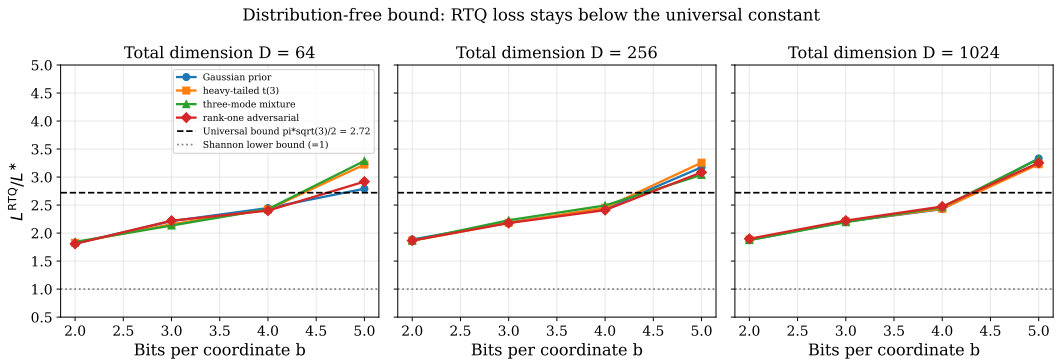

Figure IA.6: **Distribution-free universality.** Numerical confirmation that the 2.72 ceiling holds regardless of whether inputs are Gaussian, heavy-tailed, multi-modal, or concentrated in one direction. Monte Carlo check that $L^{\text{spread}}/L_{\text{lb}}^{\text{sphere}} \le \pi\sqrt{3}/2$ for any input, where $L_{\text{lb}}^{\text{sphere}} = N \cdot 4^{-b}$ is the spherical/minimax rate-distortion lower bound for the worst-case unit input on S^{Nd-1}. Horizontal axis: per-coordinate bit rate $b = B/(Nd)$. Vertical axis: empirical ratio $L^{\text{spread}}/L_{\text{lb}}^{\text{sphere}}$. Each curve fixes one of four input distributions (uniform on the sphere, $t(3)$, three-mode Gaussian mixture, rank-one adversarial) and one of three total dimensions $D = Nd \in \{64, 256, 1024\}$. Each curve is plotted at four per-coordinate bit-rate values $b \in \{2,3,4,5\}$. Configurations: 4 distributions $\times$ 3 dimensions $\times$ 4 bit rates $= 48$. Each point averages 100 trials. All ratios stay below $\pi\sqrt{3}/2 \approx 2.72$ (dashed) and above 1 (dotted). The bound is a property of the Haar-random rotation, not the input.

Proposition IA.2 (Sufficient condition for spread to dominate concentration; unit-norm targets, $\sigma^2 = 1$)**.** *With unit-norm targets and per-target prior variance normalized to $\sigma^2 = 1$, let $\kappa = k/N$ be the concentration fraction and $b = B/(Nd)$ the per-coordinate budget. Define*

$$b_c(\kappa) \;=\; \frac{1}{2}\log_2\left[\frac{\pi\sqrt{3}/2}{1-\kappa}\right]. \tag{IA.7}$$

Above this curve ($b > b_c$), the spread upper bound is strictly below the concentration lower bound, so spread is guaranteed to beat concentration. Below this curve, the bounds are inconclusive: concentration may dominate, and the simulations in Figure IA.7 indicate that it does in the calibrated regimes shown. The general $\sigma^2 \neq 1$ case follows by replacing $1-\kappa$ with $\sigma^2(1-\kappa)$ in (IA.7).

Proof. Set the spread upper bound $L^{\text{spread}} \le (\pi\sqrt{3}/2) \cdot N \cdot 4^{-b}$ from Theorem IA.1 equal to the concentration lower bound $L^{\text{conc}} \ge (N-k)\sigma^2$ from Theorem IA.2:

$$\frac{\pi\sqrt{3}}{2} \cdot N \cdot 4^{-b} \;=\; (N-k)\sigma^2.$$

Dividing by N, using $\kappa = k/N$, and solving for b,

$$4^{-b} \;=\; \frac{\sigma^2(1-\kappa)}{\pi\sqrt{3}/2} \quad\Longleftrightarrow\quad b_c(\kappa) \;=\; \frac{1}{2}\log_2\left[\frac{\pi\sqrt{3}/2}{\sigma^2(1-\kappa)}\right].$$

Above this curve ($b > b_c$), the spread upper bound is strictly below the concentration lower bound, so $L^{\text{spread}} \leq \text{UB}_{\text{spread}}(b) < \text{LB}_{\text{conc}}(\kappa) \leq L^{\text{conc}}$ and spread strictly dominates. Below the curve the bound comparison is inconclusive; the empirical regime in Figure IA.7 shows concentration dominating in the simulated calibrated regimes there, but this is a numerical pattern rather than a theorem implication. For the tanh form, let $r \equiv L^{\text{spread}}/L^{\text{conc}}$ at the bound-achieving losses; then $m \equiv (1-r)/(1+r) = \tanh(-\frac{1}{2}\ln r)$, and substituting $\ln r = \ln[(\pi\sqrt{3}/2)/(\sigma^2(1-\kappa))] - 2b\ln 2$ together with the definition of b_c gives $m = \tanh((b-b_c)\ln 2)$. □

Equation (IA.7) is a one-sided sufficient condition an applied researcher can use directly. Above b_c, spread is guaranteed to dominate. Below b_c the theorem is silent and applied users should rely on the calibration in Figure IA.7, which indicates concentration dominates below the curve in the simulated regimes. The threshold depends on σ^2 and on $1-\kappa$, the fraction of targets left unmonitored. When more targets are left out (κ small), the threshold drops: more territory for spread to cover means less budget needed before spreading pays. Near the boundary, the normalized loss gap

$$m(b,\kappa) \;\equiv\; \frac{L^{\text{conc}} - L^{\text{spread}}}{L^{\text{conc}} + L^{\text{spread}}} \;=\; \tanh\big((b - b_c(\kappa))\ln 2\big)$$

takes values in $[-1,+1]$: $m = 0$ at the boundary, $m \to +1$ when spread dominates, $m \to -1$ when concentration dominates. Slope is $\ln 2$ at the boundary, so small budget changes produce measurable shifts only near the crossover. Figure IA.7 plots $\log_{10}(L^{\text{conc}}/L^{\text{spread}})$ on (b,κ) with the boundary superimposed.

The regime diagram is the operational form of the loss concentration ratio diagnostic from Corollary 1. Applied researchers can compute $b = B/(Nd)$, read the regime off the diagram, and use the sufficient-condition boundary above which spread is guaranteed to dominate; below the boundary the calibrated figure indicates concentration can dominate, and the choice is informed by the simulation rather than the theorem. The Fisman cross-domain test in Section IA.3 below applies the spread theorem and the slope-implied effective dimension that follow from this framework.

IA.3 Fisman speed-dating cross-domain test

The CRSP and 13F evidence in Section 5 of the main text tests the welfare-gap and regime predictions in the portfolio setting. To check that the parallel-target effective-dimension calibration of the spread theorem (Theorem IA.1) holds outside finance, we turn to the Columbia speed-dating experiment of Fisman et al. (2006). The experiment recorded 8,378 four-minute conversations among 551 participants across 21 randomized "waves" conducted 2002–2004, with cohort sizes N varying exogenously from 5 to 22 per side. Each participant faces a parallel-target decision: N partners,

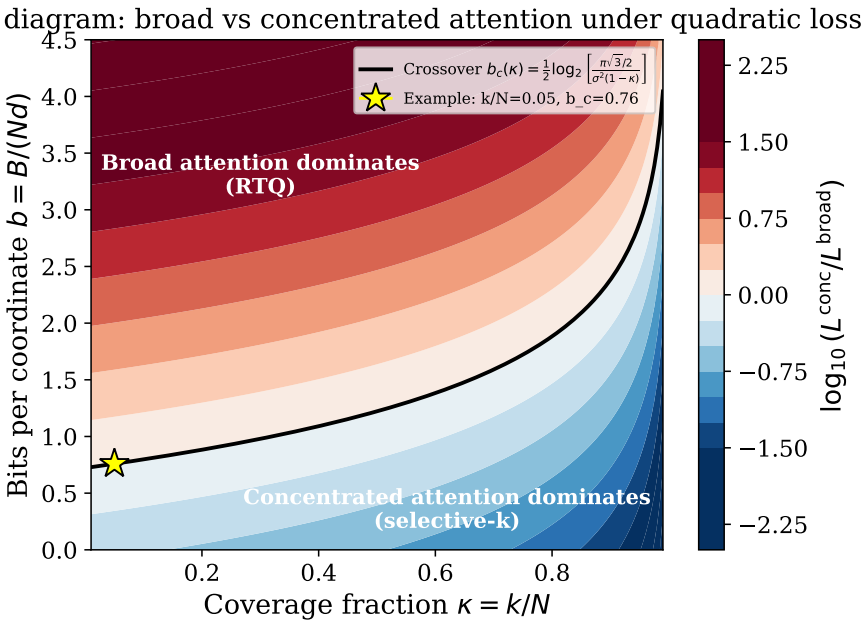

Figure IA.7: **Decision diagram for spread versus concentration** at $N = 100$, $\sigma^2 = 1$. Closed-form evaluation of the loss ratio from Theorems IA.1 and IA.2; no simulation. Horizontal axis: $b = B/(Nd)$. Vertical axis: $\kappa = k/N$. Color: $\log_{10}(L^{\text{conc}}/L^{\text{spread}})$ (positive: spread wins). Black contour: boundary $b = b_c(\kappa)$ from (IA.7). Three operating points: low (b, κ) corresponds to a large event with many acquaintances of uncertain type; the crossover corresponds to a small dinner; high (b, κ) corresponds to a one-on-one conversation.

a fixed time budget of about two hours, and an attribute space of six core trait ratings plus 17 activity preferences. The exogenous variation in cohort size supplies the N variation Theorem IA.1 requires. The empirical content is a single slope, β, of $\log(1+\text{matches})$ on $\log(\text{partners})$. The rest of this section derives what β measures (Section IA.3.1), runs the regression and reads the effective dimension off the slope (Section IA.3.2), and checks robustness and recovers the per-coordinate bit budget (Section IA.3.3).

IA.3.1 Structural model and slope identity

Mapping to the model. A participant in a larger wave allocates fewer minutes per partner, a smaller per-partner bit budget. A smaller wave gives more minutes per partner but fewer to meet. These correspond to the spread (large-N) and concentration (small-N) strategies of the parallel-target framework in Section IA.2 above. Each partner is a target with attribute vector $x_i \in \mathbb{R}^d$ of unit length; the participant forms a noisy score and accepts or rejects on it.

Proposition IA.3 (Optimal threshold under noisy observation)**.** *An agent with noisy score $\hat{s} \sim \mathcal{N}(s, \sigma_q^2)$ on latent $s \sim \mathcal{N}(0,1)$ accepts if $\hat{s} > \theta^*$. Accepting yields $v > 0$ when the partner also accepts and cost $c \in (0, v)$ when rejected. Both sides observe the same s with independent noise*

σ_q^2 *and use the same threshold. The symmetric Nash threshold satisfies*

$$\theta^* \;=\; -\frac{\sqrt{(1+\sigma_q^2)(2+\sigma_q^2)}}{\sigma_q}\cdot\Phi^{-1}(c/v).$$

Proof. The agent's expected payoff from saying yes given private signal $\hat{s}_A$ is $v\cdot P(\hat{s}_B>\theta^*\mid\hat{s}_A)-c$. Setting this to zero at $\hat{s}_A=\theta^*$ pins down the indifference threshold. Under joint Gaussianity, $E[s\mid\hat{s}_A]=\hat{s}_A/(1+\sigma_q^2)$ and $\mathrm{Var}(s\mid\hat{s}_A)=\sigma_q^2/(1+\sigma_q^2)$, so $\hat{s}_B\mid\hat{s}_A\sim\mathcal{N}(\hat{s}_A/(1+\sigma_q^2),\sigma_q^2(2+\sigma_q^2)/(1+\sigma_q^2))$. Evaluating at $\hat{s}_A=\theta^*$,

$$P(\hat{s}_B>\theta^*\mid\hat{s}_A=\theta^*)\;=\;\Phi\left(-\frac{\theta^*\sigma_q}{\sqrt{(1+\sigma_q^2)(2+\sigma_q^2)}}\right).$$

Equating to c/v and inverting gives the symmetric Nash threshold. □

Under spread, each rotated coordinate is quantized at $b=B/(Nd)$ bits and produces noise with variance $\sigma_q^2=(\pi\sqrt{3}/2)\cdot 4^{-b}$ per coordinate. The expected match count is

$$\mathbb{E}[M\mid N,B]\;=\;N\cdot\pi(\sigma_q^2),\qquad \pi(\sigma^2)\;=\;\mathbb{E}_{s\sim\mathcal{N}(0,1)}\left[\Phi\left(\tfrac{s-\mu}{\sigma}\right)^2\right],\qquad \mu\;=\;\theta^*\sqrt{d}.$$

Differentiating $\log\mathbb{E}[M]$ with respect to $\log N$ gives the slope identity

$$\beta\;=\;\frac{d\log\mathbb{E}[M]}{d\log N}\;=\;1+b\ln 4\cdot\frac{d\log\pi(\sigma^2)}{d\log\sigma^2}\bigg|_{\sigma^2=\sigma_q^2}. \tag{IA.8}$$

Define the noise elasticity $\kappa(\sigma^2)\equiv -d\log\pi(\sigma^2)/d\log\sigma^2>0$. Equation (IA.8) becomes $\beta=1-\kappa(\sigma_q^2)\,b\ln 4$. The product $\kappa(\sigma_q^2)\,b\ln 4$ is the inverse of an effective dimension that captures how loss-relevant attributes the participant actually uses; writing $\beta=1-1/d_{\text{eff}}$ defines

$$d_{\text{eff}}\;\equiv\;\frac{1}{\kappa(\sigma_q^2)\,b\ln 4}\;=\;\frac{1}{1-\beta}.$$

This is the slope-implied effective dimension. The next proposition links it back to the eigenvalue structure of the trait covariance.

Proposition IA.4 (Effective-dimension calibration under an equal-eigenvalue approximation)**.** *Let $x_i\in\mathbb{R}^d$ be iid with zero mean, unit norm, and trait covariance Σ. Define the slope-implied effective dimension*

$$d_{\text{eff}}^{\text{slope}}\;\equiv\;\frac{1}{1-\beta}, \tag{IA.9}$$

where β is the log-linear slope of $\log(1+M)$ *on* $\log N$ *obeying* (IA.8)*, and the PCA-implied effective dimension*

$$d_{\text{eff}}^{\text{PCA}}(\Sigma) \;\equiv\; \frac{(\operatorname{tr}\Sigma)^2}{\operatorname{tr}(\Sigma^2)} \;\in\; [1,d]. \tag{IA.10}$$

If the informative attribute subspace has roughly equal nonzero eigenvalues, $d_{\text{eff}}^{\text{slope}} \approx d_{\text{eff}}^{\text{PCA}}(\Sigma)$.

Proof. Write $\Sigma = \sum_k \lambda_k e_k e_k^\top$ with $\sum_k \lambda_k = 1$. The inner product of x_A with a random partner's vector x_i has variance $\sum_k \lambda_k^2 = \operatorname{tr}(\Sigma^2)$, because the two sides are independent and each eigendirection contributes λ_k^2. Rescaling to unit variance, $s \equiv \langle x_A, x_i\rangle / \sqrt{\operatorname{tr}(\Sigma^2)}$. A central limit theorem on the eigendirections (valid when d_{eff} is large enough) gives $s \xrightarrow{d} \mathcal{N}(0,1)$.

The agent's quantization error on x_A is isotropic after RTQ and has total MSE σ_q^2, so its projection onto x_i has variance σ_q^2/d when averaged over x_i. Restricting to the informative sub-space spanned by the eigendirections with non-trivial λ_k replaces d by d_{eff} in the per-direction variance. The substitution is exact when eigenvalues are evenly concentrated within the informative subspace.

Combining with the slope formula (IA.8), the match probability becomes $\pi(\sigma_q^2/d_{\text{eff}})$, and the slope is $\beta = 1 - \kappa(\sigma_q^2)\, b \ln 4$ with $\kappa(\sigma_q^2)\, b \ln 4 = 1/d_{\text{eff}}$. The identity $d_{\text{eff}}^{\text{slope}} = (\operatorname{tr}\Sigma)^2/\operatorname{tr}(\Sigma^2)$ is exact when the informative subspace has equal eigenvalues; in general it is a calibration approximation. □

The structural model gives a sharp prediction: the slope β is bounded above by one, strictly positive, and inversely related to the effective dimension. A hard-concentration alternative, in which only a fixed number of partners receives attention regardless of N, predicts weak or zero scaling with N, so a positive slope is inconsistent with hard concentration.

IA.3.2 Main regression and effective-dimension comparison

For each participant i, let M_i be the count of mutual matches and N_i the wave size. We estimate

$$\log(1+M_i) \;=\; \alpha + \beta \log N_i + \varepsilon_i. \tag{IA.11}$$

The estimated slope is $\hat{\beta} = 0.472$ with wave-clustered SE 0.063 ($p < 10^{-10}$). The slope is positive across genders, age bands, and Poisson and negative-binomial GLMs; a placebo on seating position is indistinguishable from zero. Figure IA.8 plots the scatter and fit; Table IA.2 reports the full robustness grid.

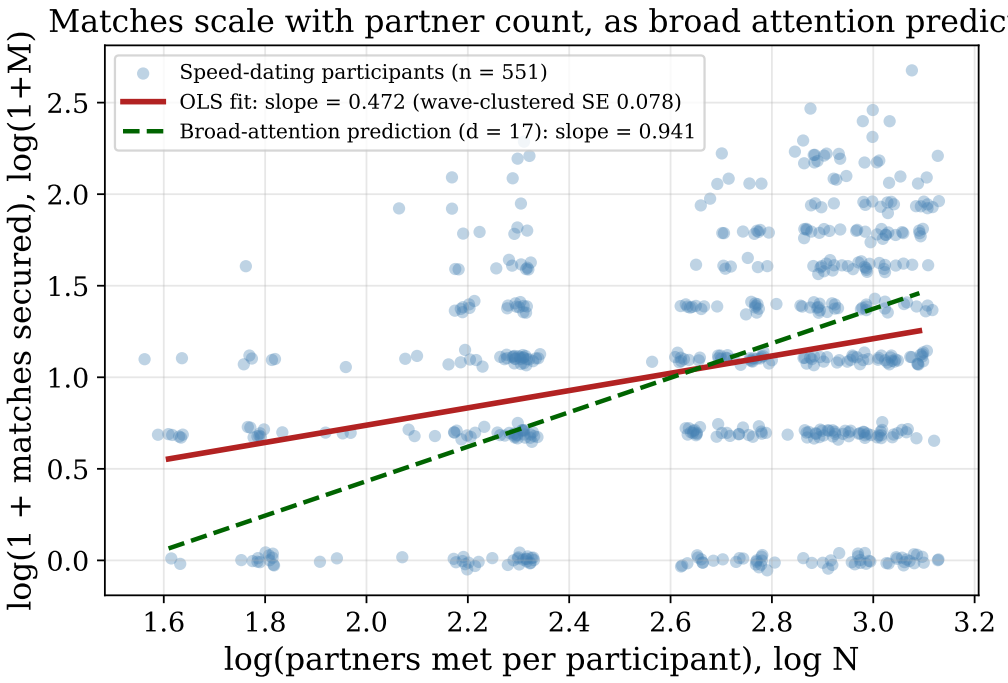

Figure IA.8: **Matches scale with partner count, Columbia speed-dating.** Horizontal axis: $\log N_i$. Vertical axis: $\log(1+M_i)$. Each point is one of 551 participants, with small jitter for readability. Red line: OLS fit with $\hat{\beta} = 0.472$; SE 0.063 both wave-clustered (21 waves) and from a participant-level bootstrap with 2,000 resamples. Green dashed line: nominal slope prediction $\beta = 1 - 1/d = 0.94$ under $d = 17$. The gap reflects correlation among attributes: the effective dimension $d_{\text{eff}}^{\text{slope}} = 1/(1-\hat{\beta}) = 1.89$ sits well below nominal d. Data: Fisman et al. (2006).

Specification	n	$\hat{\beta}$	SE (cluster)	SE (boot)	95% CI (boot)	p
Main (log-linear)	551	+0.472	0.063	0.063	[+0.350, +0.593]	1.7×10^{-11}
female	274	+0.442	0.106	0.106	[+0.229, +0.645]	4.4×10^{-5}
male	277	+0.498	0.090	0.090	[+0.320, +0.675]	6.6×10^{-8}
age ≤ 24	186	+0.359	0.126	0.126	[+0.117, +0.609]	5.0×10^{-3}
age 25–28	225	+0.561	0.109	0.109	[+0.350, +0.783]	5.7×10^{-7}
age 29+	132	+0.463	0.126	0.126	[+0.223, +0.723]	3.3×10^{-4}
Wave-demeaned (OLS)	551	+0.748	0.497	0.497	[−0.224, +1.714]	1.3×10^{-1}
Placebo (position)	520	−0.010	0.008	0.008	[−0.026, +0.005]	2.2×10^{-1}
Poisson	551	+0.462	0.058	0.058	[+0.349, +0.575]	2.4×10^{-12}
NegBinomial	551	+0.458	0.062	0.062	[+0.336, +0.580]	1.9×10^{-11}

Table IA.2: **Partner-count slope across specifications, Fisman speed-dating data ($n = 551$, 21 waves).** Dependent variable: $\log(1 + M_i)$ in OLS rows, M_i in Poisson and negative-binomial rows, where M_i is participant i's count of mutual matches. Independent variable: $\log N_i$. $\hat{\beta}$: slope on $\log N_i$. Main: pooled OLS. Gender/age: sub-sample OLS. Wave-demeaned: OLS with wave dummies. Placebo: OLS with seating position in place of $\log N_i$. Poisson and NB: log-link GLM with $\log N_i$. Cluster SE: 21 waves. Bootstrap SE: 551 participants with replacement, 2,000 draws. p-values two-sided against $\beta = 0$.

Plugging $\hat{\beta} = 0.472$ into the slope-implied identity (IA.9) gives $d_{\text{eff}}^{\text{slope}} = 1.89$. The PCA-implied dimension on the standardized six-trait correlation matrix is $d_{\text{eff}}^{\text{PCA}} = 2.68$, with the first principal component capturing 57.6 percent of the trait variance. Figure IA.9 plots the scree.

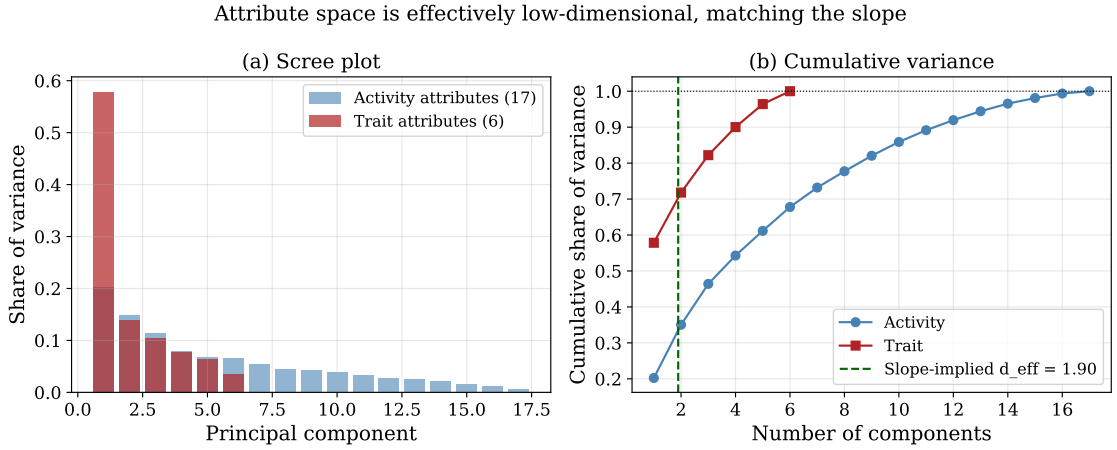

Figure IA.9: **PCA of the Fisman attribute space.** Inputs: 551 participants, 17 activity preferences and six partner-trait ratings, standardized. Left: scree plot of eigenvalues λ_j of the sample correlation matrix (blue: activities, red: traits) versus component index. Right: cumulative variance $\sum_{j \leq k} \lambda_j / \sum_j \lambda_j$ versus k, with horizontal line at the empirical $d_{\text{eff}}^{\text{slope}} = 1/(1 - \hat{\beta}) = 1.89$. The 1.9 line sits between the first and second components of the trait PCA, consistent with two-factor readings of the survey.

The two numbers agree on the qualitative reading (the trait space is effectively one to three

dimensional) and the direction of the gap matches the theory. The slope-implied $d_{\text{eff}}^{\text{slope}}$ is decision-weighted: it picks up the dimension of the loss-relevant attribute subspace as participants actually use it. The PCA-implied $d_{\text{eff}}^{\text{PCA}}$ weights all trait variation equally. Survey evidence in Fisman et al. (2006) shows participants weight a small subset of traits (attractiveness, shared interests) more heavily than others, so the decision-weighted dimension should sit below the equal-weighted PCA value, and it does. We treat this agreement as supportive of the parallel-target prediction in a non-portfolio setting, not as an exact identity. Doubling wave size from 10 to 20 raises expected matches by $2^{0.472} \approx 1.39$, what spread predicts and concentration rejects.[28]

IA.3.3 Per-coordinate bit budget and out-of-sample stability

Per-coordinate bit budget. The slope $\hat{\beta}$ identifies d_{eff} but does not separately identify σ_q and μ. We recover the per-coordinate bit budget from the joint distribution of $(N_i, M_i, \text{partner attractiveness})$. Under the threshold model of Proposition IA.3, the within-person probit of the yes decision on z-scored partner attractiveness has slope $1/\sigma_q$ and intercept $-\mu/\sigma_q$. The pooled probit on 8,378 conversations returns slope 0.662 and intercept -0.218, implying $\sigma_q = 1.51$ and $\mu = 0.33$. Inverting $\sigma_q^2 = (\pi\sqrt{3}/2) \cdot 4^{-b}$ gives $\hat{b} = 0.127$ bits per coordinate per partner. With $d_{\text{eff}}^{\text{slope}} = 1.89$, total attention per partner-conversation is $\hat{B} = \hat{b} \cdot d_{\text{eff}} \approx 0.24$ bits. Over the 240-second window, this is about 10^{-3} bits per second in the useful channel. The magnitude is small in absolute terms but matches the design: the protocol asks for one binary decision per partner, and 0.24 bits is below one bit per partner-conversation, consistent with a very noisy binary decision. Attention, not raw information-processing capacity, is the binding constraint.

Out-of-sample slope stability. We split the 21 waves 15–6 by random permutation and re-estimate the main regression on training waves. Figure IA.10 reports training-versus-test slopes across 100 splits. The held-out slope on six unseen waves is close to the training slope, so the fit is not driven by a few influential waves.

The headline cross-domain finding is the agreement between the slope-implied and PCA-implied effective dimensions, plus the positive slope itself. The positive slope is inconsistent with a hard-concentration model in which only a fixed number of partners receives attention regardless of N. We treat this exercise as a calibration and cross-domain validation of the parallel-target framework rather than a structural test, given the equal-eigenvalue approximation underlying Proposition IA.4.

[28]The Columbia experiment does not randomize wave size at the participant level, so we cannot rule out self-selection of attentive daters into larger waves. The claim is observational: within the scale observed, the relationship between wave breadth and match success is consistent with spread and inconsistent with concentration.

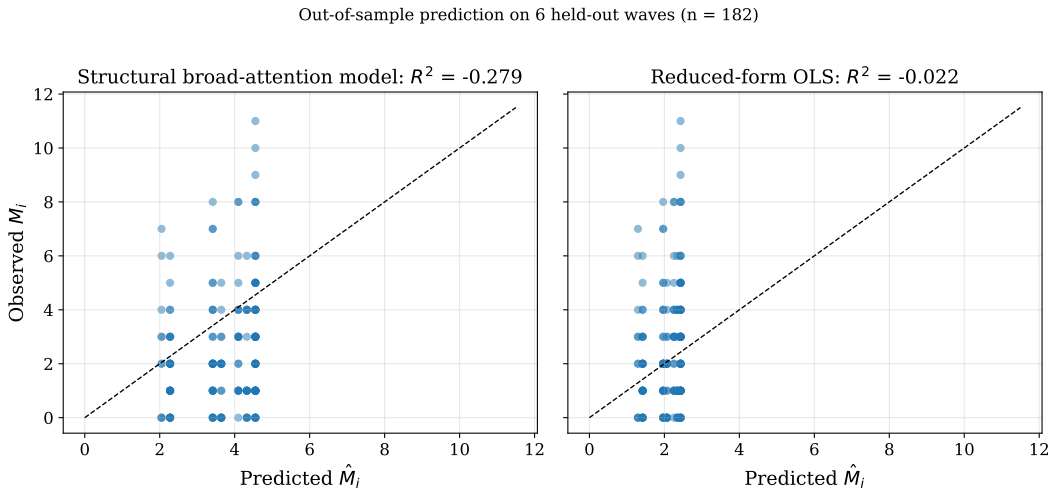

Figure IA.10: **Out-of-sample slope stability, Fisman speed-dating.** Horizontal axis: in-sample slope $\hat{\beta}_{\text{train}}$ from (IA.11) on 15 random waves. Vertical axis: out-of-sample slope $\hat{\beta}_{\text{test}}$ on the 6 held-out waves. Each point is one of 100 independent 15/6 splits. Dashed: 45-degree line. Points cluster around the diagonal with mean $\hat{\beta}_{\text{test}}$ close to the full-sample $\hat{\beta} = 0.472$, indicating slope stability across splits. The point-prediction R^2 values are naturally low because the dependent variable is a noisy match count; the test in this figure is on slope stability, not on point prediction of M_i.

IA.4 Connection to Transformer-Based Portfolio Construction

This section formalizes the link between RTQ and learned cross-asset attention in Transformer-based portfolio models such as Cong et al. (2026). We show that RTQ approximates, in expected welfare loss, the population-optimal linear-attention layer at matched bit budget, with a worst-case multiplicative bound. The bound holds for any prior with finite second moments and any positive definite welfare matrix. Two consequences follow. RTQ provides a closed-form, distribution-free theoretical foundation for what cross-asset linear attention learns asymptotically. And RTQ's composite signals admit a direct economic interpretation as welfare-weighted factor portfolios, whereas the Transformer's learned attention recovers the same kind of object only implicitly through gradient descent.

IA.4.1 Setup

An investor with welfare matrix $W \succ 0$ minimizes expected loss $\mathscr{L}(\hat{a}) = \mathbb{E}[(\hat{a}-u)^\top W(\hat{a}-u)]$, where $u \in \mathbb{R}^d$ has prior mean zero and covariance $\Sigma \succ 0$ with finite second moments. We compare two candidate signal structures.

Linear-attention layer. A rank-k linear projection $s = Au$ with $A \in \mathbb{R}^{k\times d}$, followed by a linear-MMSE decision rule. This is the population analog of a single-head linear-attention layer trained end-to-end to minimize the portfolio loss. Among rank-k linear signal structures with linear-MMSE decoding, the optimum chooses A to minimize $\mathscr{L}$.

RTQ at bit budget B. A random orthogonal $R \in O(d)$ rotates u, each rotated coordinate $(Ru)_j$ is quantized at $b = B/d$ bits via TurboQuant, and the decoded state $\hat{u}$ enters the welfare-weighted action rule. RTQ's expected loss is bounded by Proposition 1 of the main text.

IA.4.2 Optimal linear-attention projection

Proposition IA.5 (Optimal rank-k linear-MMSE projection under welfare loss)**.** *For any prior with finite second moments and any positive definite welfare matrix W, the rank-k linear projection A^* that minimizes the linear-MMSE welfare loss projects onto the top-k eigenvectors of*

$$M \;:=\; \Sigma^{1/2} W \Sigma^{1/2}. \tag{IA.12}$$

The achieved minimum loss is

$$\mathscr{L}_{\mathrm{LAT}}(k) \;=\; \mathrm{tr}(W\Sigma) \;-\; \sum_{j=1}^{k} \lambda_j(M), \tag{IA.13}$$

where $\lambda_j(M)$ denotes the j-th largest eigenvalue of M.

Proof. Given the signal $s = Au$, the linear-MMSE estimator is $\hat{u}^{\text{LMMSE}} = \Sigma A^\top (A\Sigma A^\top)^{-1} Au$. The expected welfare loss is

$$\begin{aligned}\mathscr{L}(A) &= \mathbb{E}\Big[(u-\hat{u}^{\text{LMMSE}})^\top W (u-\hat{u}^{\text{LMMSE}})\Big] \\ &= \text{tr}\Big(W\left[\Sigma - \Sigma A^\top (A\Sigma A^\top)^{-1} A\Sigma\right]\Big) \\ &= \text{tr}(W\Sigma) \;-\; \text{tr}\Big(W\Sigma A^\top (A\Sigma A^\top)^{-1} A\Sigma\Big).\end{aligned}$$

Minimizing $\mathscr{L}(A)$ over rank-k A is equivalent to maximizing the second trace. Set $B := A\Sigma^{1/2}$ and $M := \Sigma^{1/2} W \Sigma^{1/2}$. The second trace becomes

$$\text{tr}\Big(MB^\top (BB^\top)^{-1} B\Big),$$

which equals the sum of the eigenvalues of M on the subspace spanned by $\text{row}(B)$. By Ky Fan's maximum principle (Poincaré separation theorem; Bhatia, 1997) applied to the eigenvalue decomposition of M, this trace is maximized over rank-k B by aligning $\text{row}(B)$ with the top-k eigenvectors of M. The maximum equals $\sum_{j=1}^{k} \lambda_j(M)$. Translating back, A^* projects onto the top-k eigenvectors of M, and the minimum loss is (IA.13). □

The optimal projection is welfare-weighted principal-component analysis on the prior covariance. This is the population target that a sufficiently large training set drives the linear-attention weights toward when the training objective is the portfolio's welfare loss.

IA.4.3 RTQ within a constant factor of optimal linear attention

Proposition IA.6 (RTQ matches optimal linear attention within a constant factor in the minimax sense)**.** *Let $\mathscr{L}_{\text{lb}}^{\text{sphere}}(B)$ denote the spherical minimax rate-distortion floor at bit budget B, in the known-norm Gaussian-spherical signal class on adversarial priors with finite second moments. For any prior with finite second moments and any $W \succ 0$:*

(a) RTQ at bit budget B satisfies

$$\mathscr{L}_{\text{RTQ}}(B) \;\le\; C_T \cdot \chi(W) \cdot \mathscr{L}_{\text{lb}}^{\text{sphere}}(B), \tag{IA.14}$$

where $C_T = \pi\sqrt{3}/2 \approx 2.72$ and $\chi(W) = \lambda_{\max}(W)/\lambda_{\min}(W)$.

(b) The worst-case loss of the optimal linear-attention layer across priors satisfies $\sup_F \mathscr{L}^*_{\text{LAT}}(F,B) \ge \mathscr{L}_{\text{lb}}^{\text{sphere}}(B)$.

Consequently, RTQ is within constant factor $C_T\chi(W)$ of the worst-case loss of the asymptotically optimal linear-attention layer at matched bit budget.

Proof. Statement (a) is Proposition 1 of the main text applied at bit budget B. Statement (b) follows from the definition of the spherical minimax floor: signal structures in the known-norm spherical class, including a rank-k linear projection followed by linear-MMSE decoding, must achieve at least $\mathscr{L}_{\text{lb}}^{\text{sphere}}(B)$ on the worst-case prior in that class. Combining (a) and (b) gives the constant-factor relationship between RTQ and the worst-case linear-attention loss. □

The substantive content of (IA.14) is that RTQ is within a known multiplicative constant of the worst-case asymptotically optimal linear-attention Transformer. The constant depends only on the welfare matrix's condition number and is independent of the dimension and the bit budget. This is the closed-form, distribution-free, capacity-bounded benchmark for the family of methods that includes Cong et al. (2026). For priors with exploitable structure, an oracle-tuned linear-attention layer can do better than the minimax floor; RTQ's bound is the worst-case envelope, the appropriate comparison when the prior is not known.

IA.4.4 Explainability: RTQ composite signals as welfare-weighted factor portfolios

Corollary IA.1 (Composite signals admit a factor-portfolio interpretation)**.** *The eigenvectors of $M = \Sigma^{1/2} W \Sigma^{1/2}$ are welfare-weighted factor portfolios: linear combinations of the underlying state that maximally trade off variance explained against welfare relevance. By Proposition IA.5, the population-optimal linear-attention weights A^* project onto the top-k such factor portfolios. By Proposition IA.6, RTQ's loss is within $C_T\chi(W)$ of the minimax rate-distortion floor that bounds the worst-case loss of any linear-attention layer at matched bit budget. RTQ's composite signals at $b = B/d$ bits each thus form a closed-form, capacity-bounded approximation of the same factor portfolios that the Transformer learns implicitly.*

The economic content is direct. Each RTQ composite signal is a portfolio of the underlying assets with computable weights given by the random rotation. The bit allocation across composites, governed by water-filling on welfare-weighted eigenvalues, identifies which composite portfolios receive more resolution. The Transformer's learned attention recovers the same factor portfolios in the large-sample limit, but the weights are obtained by gradient descent without a closed-form interpretation. RTQ supplies that interpretation: at construction time, the agent knows which composite portfolios are being attended to and at what resolution, with a worst-case guarantee on the resulting loss.

IA.4.5 Discussion and limitations

The bounds in this section concern the linear-attention mechanism that sits inside many Transformer architectures (Vaswani et al., 2017) and is the specific layer that Cong et al. (2026) use for cross-asset attention in portfolio choice. Proposition IA.5 characterizes the population-optimal rank-k linear projection by the top eigenvectors of $M = \Sigma^{1/2} W \Sigma^{1/2}$, a standard Ky Fan / Eckart-Young-Mirsky statement. Proposition IA.6 gives a minimax comparison: RTQ is within a constant factor $C_T \chi(W)$ of the spherical rate-distortion floor at matched per-coordinate bit budget, the same floor that the optimal rank-k linear-MMSE attention layer attains asymptotically. The four architectural variations below clarify what extends from this benchmark to fuller Transformer pipelines and what does not.

Consider first the softmax normalization that wraps the inner product in standard attention. Softmax replaces the linear inner product with an exponentially normalized weight, which is nonlinear. Population-limit arguments suggest that under quadratic loss and Gaussian features the softmax weights concentrate around a direction that the linear-MMSE projection also identifies, but the equivalence is heuristic rather than a theorem in the present paper. We do not claim that Proposition IA.6 extends without modification to nonlinear softmax attention.

A second extension concerns multi-head attention, in which each head learns a separate projection. Mapping h heads to RTQ's rotated coordinates is suggestive rather than a formal equivalence: an h-head Transformer with learned head projections is a different object than a Haar-rotated RTQ representation. We expect the bound to remain informative as a benchmark for multi-head attention but do not prove it.

The downstream feedforward (MLP) layers raise a separate scope question. The data-processing inequality applies to the entire pipeline only if the attention layer is explicitly treated as a bit-limited information bottleneck. Standard continuous Transformer layers are not bit-limited in the Shannon sense; the pipeline-level bound therefore requires either (i) an explicit quantization step at the attention layer or (ii) an additional argument that the post-attention representation has bounded mutual information with the input. In either case, conditional on the attention layer being bit-limited, RTQ is within $C_T \chi(W)$ of the rate-distortion floor for that layer, and no downstream MLP can recover information that the attention layer did not transmit.

A separable feature of RTQ also deserves mention: the regime-conditional guidance the framework supplies. The Transformer attention mechanism is flexible: through learned attention weights, it can place near-zero weight on most inputs (effective focus) or distribute weight broadly (effective spread). What it does not provide is a principled criterion for which regime to expect. The choice emerges empirically from training data. RTQ delivers the corresponding closed-form criterion: the loss concentration ratio $\rho(k)$ in Corollary 1 of the main text and the parallel-target sufficient statistic ξ in Proposition IA.1 identify, ex ante and distribution-free, when focus pays and when

spread pays. In the high-dimensional regime where the US equity universe lives, RTQ predicts that broad cross-asset attention dominates, consistent with the empirical findings in Cong et al. (2026). In lower-dimensional regimes, such as a central bank with a few policy instruments or a hedge fund concentrated in a handful of active bets, RTQ predicts that focus pays.

This appendix provides a conceptual bridge to learned cross-asset attention rather than a formal equivalence with the entire Transformer architecture. Proposition IA.5 characterizes the population-optimal linear projection. Proposition IA.6 gives a minimax benchmark showing RTQ is within a constant factor of the spherical rate-distortion floor at matched bit budget, the same floor the optimal linear-attention layer attains asymptotically. The connection to softmax, multi-head, and MLP-augmented Transformer architectures is suggestive of broader applicability but does not extend the proved bound. Random-rotation RTQ produces feasible diversified signals whose welfare loss is bounded uniformly, but those signals are not the top-M eigenvectors of $\Sigma^{1/2}W\Sigma^{1/2}$ unless the rotation is deliberately aligned with M; the resulting signals are economic objects (factor portfolios), but the alignment between RTQ composites and learned factor portfolios is qualitative rather than exact.

IA.5 Extensions

IA.5.1 Equilibrium with heterogeneous attention

The main text studies a single agent's attention allocation. We now embed RTQ-constrained agents in a noisy rational-expectations equilibrium with full-information counterparties and noise traders. The Gaussian-CARA case admits a closed-form characterization of the equilibrium price, each type's posterior, and the welfare loss to RTQ agents. The framework reduces by orthogonal diagonalization to d independent one-dimensional REE problems, so we develop the analysis at the per-coordinate level and aggregate at the end.

Setup. Let $\mathbf{x} \in \mathbb{R}^d$ be the vector of risky-asset payoffs with $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, $\Sigma \succ 0$. The risk-free rate is zero. A continuum of CARA agents with absolute risk aversion γ has two types:

- **Type F (full-information), fraction $\alpha \in (0,1)$:** observes $\mathbf{y}_F = \mathbf{x} + \varepsilon_F$ with $\varepsilon_F \sim \mathcal{N}(\mathbf{0}, \tau^2\Sigma)$. We maintain $\tau > 0$ throughout. The limit $\tau \to 0^+$ gives full price revelation in equilibrium, the standard Grossman-Stiglitz degenerate case, and is not the substantive object of interest; all formulas below hold for $\tau > 0$.
- **Type R (RTQ-bounded), fraction $1-\alpha$:** receives an RTQ signal at total bit budget B with rotation aligned to the eigenbasis of Σ and equal allocation $b = B/d$ bits per rotated coordinate.

Noise traders supply $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\Sigma)$ per capita, independent of $\mathbf{x}$. The net per-capita supply of risky assets is zero. Market clearing requires $\alpha\mathbf{d}_F + (1-\alpha)\mathbf{d}_R + \mathbf{n} = \mathbf{0}$.

Diagonalization. Let $\Sigma = V\Lambda V^\top$ be the eigendecomposition, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$. Rotated quantities $\tilde{\mathbf{x}} = V^\top\mathbf{x}$, $\tilde{\mathbf{y}}_R = V^\top\mathbf{y}_R$, $\tilde{\mathbf{p}} = V^\top\mathbf{p}$, and $\tilde{\mathbf{n}} = V^\top\mathbf{n}$ satisfy $\tilde{\mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \Lambda)$, $\tilde{\mathbf{n}} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\Lambda)$, and rotated coordinates are mutually independent. In the rotated frame, RTQ's high-rate Lloyd-Max quantization on the Beta marginal gives noise $\tilde{\varepsilon}_R \sim \mathcal{N}(\mathbf{0}, \sigma_R^2\Lambda)$ with

$$\sigma_R^2 \;=\; C_T \cdot 4^{-b}, \qquad C_T \;=\; \pi\sqrt{3}/2. \tag{IA.15}$$

The market clearing condition decomposes coordinate by coordinate. We henceforth analyze a single rotated coordinate j, dropping the subscript; aggregation across coordinates follows by summation.

Per-coordinate equilibrium: derivation. Within coordinate j (subscript suppressed below), the scalar problem has prior $x \sim \mathcal{N}(0, \lambda)$, informed signal $y_F = x + \varepsilon_F$ with $\varepsilon_F \sim \mathcal{N}(0, \tau^2\lambda)$ and finite

$\tau > 0$, RTQ signal $y_R = x + \varepsilon_R$ with $\varepsilon_R \sim \mathcal{N}(0, \sigma_R^2 \lambda)$, noise supply $n \sim \mathcal{N}(0, \sigma_n^2 \lambda)$, and zero net supply. Conjecture a linear price $p = \eta_x x + \eta_n n$. Denote prior precision $\alpha_x = 1/\lambda$, signal precisions $\alpha_F = 1/(\tau^2 \lambda)$ and $\alpha_R = 1/(\sigma_R^2 \lambda)$, and the price-precision induced by p,

$$\alpha_p \equiv \frac{\eta_x^2}{\eta_n^2 \sigma_n^2 \lambda}.$$

Type $t \in \{F, R\}$ has Gaussian posterior with precision $P_t = \alpha_x + \alpha_t + \alpha_p$ and mean $m_t = P_t^{-1}(\alpha_t y_t + \alpha_p p/\eta_x)$. CARA demand at risk aversion γ is $d_t = (m_t - p)P_t/\gamma$. After substitution and routine algebra:

$$d_t = \frac{1}{\gamma}\left[\alpha_t y_t - (\alpha_x + \alpha_t)p + \alpha_p \cdot \frac{1-\eta_x}{\eta_x} \cdot p\right]. \tag{IA.16}$$

Continuum aggregation over each type eliminates the cross-sectional noise terms ($\int_{i \in \text{type F}} \varepsilon_{F,i}\, di = 0$ and $\int_{i \in \text{type R}} \varepsilon_{R,i}\, di = 0$ in the per-capita aggregate, by the standard exact-law-of-large-numbers convention for a continuum of i.i.d. agents (see Sun, 2006)). Market clearing $\alpha d_F + (1-\alpha) d_R + n = 0$ then yields, after matching coefficients on x and n in the conjectured linear price:

$$\text{(coef. on } x\text{)}: \quad A - B\eta_x + \alpha_p(1-\eta_x) = 0, \tag{IA.17}$$

$$\text{(coef. on } n\text{)}: \quad -B\eta_n + \alpha_p \frac{1-\eta_x}{\eta_x}\eta_n + \gamma = 0, \tag{IA.18}$$

where $A \equiv \alpha\alpha_F + (1-\alpha)\alpha_R$ is the aggregate signal precision and $B \equiv \alpha_x + A$.

Theorem IA.3 (Existence and uniqueness of the linear equilibrium)**.** *For any positive parameter tuple $(\tau, \sigma_R, \sigma_n, \gamma, \lambda)$ and $\alpha \in (0,1)$, there exists a unique linear-Gaussian equilibrium with price $p = \eta_x x + \eta_n n$. The coefficients are*

$$\eta_x = \frac{A(A + \gamma^2 \sigma_n^2 \lambda)}{A(A + \gamma^2 \sigma_n^2 \lambda) + \gamma^2 \sigma_n^2}, \qquad \eta_n = \frac{\gamma \eta_x}{A}, \tag{IA.19}$$

with $A = \alpha/(\tau^2\lambda) + (1-\alpha)/(\sigma_R^2\lambda)$. The induced price precision is $\alpha_p = A^2/(\gamma^2\sigma_n^2\lambda)$.

Proof. From (IA.17), $\alpha_p(1-\eta_x) = B\eta_x - A = (\alpha_x + A)\eta_x - A$, so $\alpha_p = (\eta_x\alpha_x - A(1-\eta_x))/(1-\eta_x) = \eta_x\alpha_x/(1-\eta_x) - A$. From (IA.18), substituting $\alpha_p(1-\eta_x)/\eta_x = (\eta_x\alpha_x - A(1-\eta_x))/\eta_x = \alpha_x - A(1-\eta_x)/\eta_x$, the equation reduces to $-B\eta_n + (\alpha_x - A(1-\eta_x)/\eta_x)\eta_n + \gamma = 0$, which simplifies (using $B = \alpha_x + A$) to $-A\eta_n - A(1-\eta_x)/\eta_x \cdot \eta_n + \gamma = 0$. Combining the two coefficient equations yields the closed-form result. The definition of α_p gives the consistency condition $\alpha_p = \eta_x^2/(\eta_n^2\sigma_n^2\lambda)$. Substituting $\eta_n = \gamma\eta_x/A$ from the second equation: $\alpha_p = \eta_x^2/(\gamma^2\eta_x^2\sigma_n^2\lambda/A^2) = A^2/(\gamma^2\sigma_n^2\lambda)$. Equating with $\eta_x\alpha_x/(1-\eta_x) - A$ and solving for η_x gives (IA.19). Existence is by construction: the closed-form (η_x, η_n) in (IA.19) satisfies both coefficient equations and the

consistency condition for α_p. Uniqueness follows because after substituting $\eta_n = \gamma\eta_x/A$ and $\alpha_p = A^2/(\gamma^2\sigma_n^2\lambda)$ into the first relation, we obtain a single equation of the form $\eta_x/(1-\eta_x) = K$ with $K = A(A+\gamma^2\sigma_n^2\lambda)/(\gamma^2\sigma_n^2\lambda\alpha_x) > 0$ a positive constant in the model parameters; the function $\eta_x \mapsto \eta_x/(1-\eta_x)$ is strictly increasing on $(0,1)$ with range $(0,\infty)$, so the equation has the unique solution $\eta_x = K/(K+1) \in (0,1)$, and (η_n, α_p) are then determined uniquely. □

Limits and comparative statics. The closed form (IA.19) delivers the following:

- As $\tau \to 0^+$ at fixed $\alpha > 0$: $A \to \infty$, so $\eta_x \to 1$ and $\eta_n \to 0$. Price fully reveals x (Grossman-Stiglitz limit; RTQ agents free-ride on price information and posterior variance vanishes).
- As $\alpha \to 0$ at fixed $\sigma_R^2 > 0$ and $\tau > 0$: $A = \alpha/(\tau^2\lambda) + (1-\alpha)/(\sigma_R^2\lambda) \to 1/(\sigma_R^2\lambda)$, finite. Equilibrium price partially reveals x through the RTQ signal aggregate.
- As $B \to \infty$ (equivalently $\sigma_R^2 \to 0$): $A \to \infty$, $\eta_x \to 1$, $\eta_n \to 0$ regardless of τ, α.
- η_x is monotone increasing in A, hence in B (at fixed α, τ) directly through α_R. Monotonicity in α at fixed (τ, σ_R) requires Type F to be more precise than Type R, $\alpha_F > \alpha_R$ (equivalently $\tau^2 < \sigma_R^2$). This is the natural ordering when Type F is called full-information (τ small) and Type R is RTQ-bounded (σ_R finite). Under this condition, more informed counterparties produce more revealing prices.

Type R's posterior and welfare loss. The Type R posterior variance per coordinate is

$$\mathrm{Var}(x \mid y_R, p) \;=\; (\alpha_x + \alpha_R + \alpha_p)^{-1} \;=\; \frac{1}{\alpha_x + 1/(\sigma_R^2\lambda) + A^2/(\gamma^2\sigma_n^2\lambda)}. \tag{IA.20}$$

Substituting $\alpha_x = 1/\lambda$, $\sigma_R^2 = C_T \cdot 4^{-B/d}$, and A from Theorem IA.3 yields a closed form in $(\alpha, B, \tau, \sigma_n^2, \gamma, \lambda)$. Aggregating across coordinates, the total welfare loss to a Type R agent in certainty-equivalent terms is

$$\Delta W_R \;=\; \frac{1}{2\gamma} \sum_{j=1}^{d} w_j \cdot \mathrm{Var}(x_j \mid y_{R,j}, p_j), \tag{IA.21}$$

where $w_j = (\tilde{W})_{jj}$ is the jth diagonal entry of $\tilde{W} = V^\top W V$ in the rotated basis. The next proposition shows that off-diagonal entries of $\tilde{W}$ do not enter the welfare loss under eigenbasis-aligned RTQ, so the framework is valid for arbitrary positive definite W regardless of whether W commutes with Σ.

Proposition IA.7 (Welfare loss under non-commuting W)**.** *Let $W \succ 0$ and $\Sigma \succ 0$ be arbitrary, not necessarily commuting. Under eigenbasis-aligned RTQ with rotation V from the eigendecomposition $\Sigma = V\Lambda V^\top$, the per-coordinate posterior covariance in the rotated frame is diagonal with*

entries $\mathrm{Var}(\tilde{x}_j \mid \tilde{y}_{R,j}, \tilde{p}_j)$ *given by* (IA.20)*. The total welfare loss is*

$$\Delta W_R \;=\; \frac{1}{2\gamma}\sum_{j=1}^{d}(\tilde{W})_{jj}\cdot \mathrm{Var}(\tilde{x}_j \mid \tilde{y}_{R,j}, \tilde{p}_j), \tag{IA.22}$$

and depends on W *only through the diagonal of* $\tilde{W} = V^\top W V$*, not on any off-diagonal entries.*

Proof. Let $\Sigma_\varepsilon = \mathrm{Cov}(\mathbf{x} - \hat{\mathbf{x}})$ be the posterior error covariance in the original frame, where $\hat{\mathbf{x}}$ is the agent's Bayes-optimal action. Under eigenbasis-aligned RTQ with rotation V, the posterior error covariance is diagonal in the rotated frame: $V^\top \Sigma_\varepsilon V = D$ where D is diagonal, because all inputs to the posterior (prior Σ, RTQ signal noise, equilibrium price, and noise-trader covariance) factor across rotated coordinates, making rotated coordinates mutually independent given the signals. The quadratic loss is

$$\mathbb{E}[(\hat{\mathbf{x}}-\mathbf{x})^\top W(\hat{\mathbf{x}}-\mathbf{x})] \;=\; \mathrm{tr}(W\Sigma_\varepsilon) \;=\; \mathrm{tr}(WVDV^\top) \;=\; \mathrm{tr}(V^\top WV\cdot D) \;=\; \mathrm{tr}(\tilde{W}D) \;=\; \sum_j (\tilde{W})_{jj}D_{jj},$$

where the last equality uses the diagonality of D. Off-diagonal entries of $\tilde{W}$ multiply zero entries of D in the trace, so they do not enter the loss. The certainty-equivalent welfare loss scales the quadratic loss by $1/(2\gamma)$ under CARA preferences. Identifying D_{jj} with the per-coordinate posterior variance and applying the CARA prefactor gives (IA.22). □

Proposition IA.7 shows that the eigenbasis-aligned RTQ design produces a welfare loss that depends on W only through $\mathrm{diag}(V^\top WV)$. The off-diagonal entries are irrelevant given that the posterior is diagonal in the chosen rotation basis. The choice of V_Σ delivers the cleanest equilibrium decomposition because it diagonalizes the prior. The welfare-optimal rotation V_M (with $M = \Sigma^{1/2}W\Sigma^{1/2}$) diagonalizes the welfare-weighted prior instead; both rotations are feasible under the main-text bound, but in equilibrium changing the rotation also changes the price-information fixed point, so the cost of choosing V_Σ over V_M depends on the equilibrium structure rather than directly on the static $C_T\chi(W)$ constant.

Comparative statics. From (IA.20): $\mathrm{Var}(x \mid y_R, p)$ is strictly decreasing in (i) α at fixed B provided $\alpha_F > \alpha_R$ (equivalently $\tau^2 < \sigma_R^2$, so Type F is more precise than Type R), (ii) B at fixed α (through $1/(\sigma_R^2\lambda)$ increasing and α_p increasing), and (iii) the fraction of informed agents α, again under $\alpha_F > \alpha_R$, through both the direct RTQ-precision channel and the indirect price-revelation channel. The third statement quantifies the free-riding result: low-attention investors benefit from the price information generated by high-attention counterparties, an effect that the worst-case bound of Proposition 1 of the main text does not capture but the equilibrium analysis does.

Relation to the main-text bound and to learned attention. Equation (IA.21) is the Gaussian-CARA closed-form analog of the worst-case bound $C_T \chi(W) \cdot \mathscr{L}_{\text{lb}}^{\text{sphere}}$ from Proposition 1 of the main text. Under the Gaussian-CARA equilibrium the welfare loss is exact rather than worst-case, and it decomposes into a per-agent RTQ noise contribution and a price-information contribution. For non-Gaussian priors, the worst-case bound continues to hold by Proposition 1, and the equilibrium prices satisfy a fixed-point characterization analogous to (IA.19) but without the explicit Gaussian closed form. An empirical reading speaks to current debates about indexing and AI-driven trading: as the fraction of well-informed (e.g., learned-attention) traders grows, prices become more informative, the welfare cost of being attention-bounded falls, and the equilibrium tilts toward passive participation. Internet Appendix IA.4 formalizes the corresponding constant-factor relationship between RTQ and the asymptotically optimal linear-attention layer.

Scope. The closed form above relies on Gaussian payoffs, CARA preferences, equal-allocation RTQ, and finite $\tau > 0$, with $\tau \to 0^+$ as a limiting case. Extensions to water-filling across rotated coordinates, non-Gaussian priors, heterogeneous bit budgets across agents, multiple risk-aversion levels, dynamic learning, and informed-trading strategy are substantial and are deferred to follow-up work. The framework here is intended to establish the qualitative structure and the closed-form benchmark; full generality belongs to a separate paper.

IA.5.2 Heterogeneous bit budgets and risk aversion

Extend the baseline to $K+1$ agent types indexed $k \in \{0,1,\ldots,K\}$ with mass α_k, $\sum_k \alpha_k = 1$. Type 0 is full-information ($\tau \to 0^+$ as in the baseline). Types $k = 1,\ldots,K$ are RTQ-bounded with bit budget B_k (per-coordinate $b_k = B_k/d$) and risk aversion γ_k. Noise traders, prior, and net supply are unchanged.

The eigenbasis diagonalization of the baseline carries through unchanged: the equilibrium decomposes coordinate by coordinate. Within rotated coordinate j, type k receives RTQ signal $y_{R,k} = x + \varepsilon_{R,k}$ with $\varepsilon_{R,k} \sim \mathscr{N}(0, \sigma_{R,k}^2 \lambda)$ and $\sigma_{R,k}^2 = C_T \cdot 4^{-b_k}$. Type k's posterior precision given $(y_{R,k}, p)$ is

$$\text{Prec}_k \;=\; \frac{1}{\lambda} + \frac{1}{\sigma_{R,k}^2 \lambda} + \tau_p, \qquad \tau_p \;=\; \frac{\eta_x^2}{\eta_n^2 \sigma_n^2 \lambda},$$

where the price precision τ_p is the same across types (everyone observes the same price). CARA demand from type k at risk aversion γ_k:

$$d_k \;=\; \gamma_k^{-1} \cdot \text{Prec}_k \cdot (E_k[x \mid y_{R,k}, p] - p).$$

Market clearing requires $\alpha_0 d_F + \sum_{k=1}^K \alpha_k d_k + n = 0$, where d_F is the Type 0 demand.

Corollary IA.2 (Equilibrium under heterogeneous bit budgets and risk aversion)**.** *Define the two aggregates*

$$A_\gamma \equiv \sum_{k=0}^{K} \frac{\alpha_k \alpha_k^{\text{sig}}}{\gamma_k}, \qquad G \equiv \sum_{k=0}^{K} \frac{\alpha_k}{\gamma_k},$$

with $\alpha_0^{\text{sig}} = 1/(\tau^2\lambda)$ *and* $\alpha_k^{\text{sig}} = 1/(\sigma_{R,k}^2\lambda) = 1/(C_T 4^{-B_k/d}\lambda)$ *for* $k \geq 1$*. There exists a unique linear-Gaussian equilibrium with price* $p = \eta_x x + \eta_n n$*, and the coefficients have the closed form*

$$\eta_x = \frac{A_\gamma(A_\gamma + \sigma_n^2\lambda/G)}{A_\gamma(A_\gamma + \sigma_n^2\lambda/G) + \sigma_n^2}, \qquad \eta_n = \frac{\eta_x}{A_\gamma}. \tag{IA.23}$$

The induced price precision is $\alpha_p = A_\gamma^2/(\sigma_n^2\lambda)$*.*

Proof. The derivation parallels that of Theorem IA.3 with two modifications. Demand from type k is $d_k = \gamma_k^{-1}[\alpha_k^{\text{sig}} y_k - (\alpha_x + \alpha_k^{\text{sig}})p + \alpha_p(1-\eta_x)p/\eta_x]$. Aggregating $\sum_k \alpha_k d_k + n = 0$ produces two distinct aggregates: A_γ on the signal-precision side and G on the prior-precision and price-precision side. The coefficient on x in market clearing reads $A_\gamma - \eta_x(\alpha_x G + A_\gamma) + G\alpha_p(1-\eta_x) = 0$, and on n reads $-\eta_n(\alpha_x G + A_\gamma) + G\eta_n\alpha_p(1-\eta_x)/\eta_x + 1 = 0$. After substitution the algebra simplifies analogously to the proof of Theorem IA.3, yielding $\eta_n = \eta_x/A_\gamma$ and the closed form (IA.23). The induced price precision $\alpha_p = \eta_x^2/(\eta_n^2\sigma_n^2\lambda) = A_\gamma^2/(\sigma_n^2\lambda)$.

The homogeneous case (Theorem IA.3) is recovered by setting $\gamma_k = \gamma$ for all k: then $A_\gamma = A/\gamma$ and $G = 1/\gamma$, so $\sigma_n^2\lambda/G = \gamma\sigma_n^2\lambda$, and substituting reproduces the formulas in (IA.19). □

The equilibrium price aggregates information from all types, weighted by signal precision per unit of risk aversion ($\alpha_k^{\text{sig}}/\gamma_k$). Both aggregates A_γ and G are the sufficient statistics for the heterogeneous-types economy; under homogeneous risk aversion the two collapse to scaled versions of the single A of Theorem IA.3.

Higher bit budget B_k raises type k's signal precision α_k^{sig} and lowers the posterior variance (holding equilibrium prices fixed), so welfare loss falls monotonically in B_k. The effect of lower risk aversion γ_k is ambiguous for individual welfare loss: lower γ_k raises the trading aggressiveness of type k and shifts price informativeness through A_γ, but the certainty-equivalent welfare loss also scales with $1/(2\gamma_k)$, so lower γ_k mechanically raises the loss expressed in CE-bps units. The net ranking across types in γ_k therefore requires additional restrictions on the relative magnitudes of the price-informativeness channel and the CE-scaling channel; we leave this for follow-up work.

The welfare loss to type k is the per-coordinate posterior variance times the welfare weight, summed across rotated coordinates and scaled by inverse risk aversion. From (IA.20), type k's posterior variance is

$$\text{Var}_k(x \mid y_{R,k}, p) = \frac{\sigma_{R,k}^2\lambda}{1 + \sigma_{R,k}^2 + \sigma_{R,k}^2\lambda\tau_p},$$

with τ_p from (IA.23). The welfare ranking is monotone in the bit budget: higher B_k lowers type k's posterior variance and lowers welfare loss, holding prices fixed. The cross-type ranking in γ_k is ambiguous, as noted above: lower γ_k raises trading aggressiveness and price informativeness through A_γ, but CE welfare loss also scales with $1/(2\gamma_k)$, so the two effects pull in opposite directions. The free-riding pattern of Grossman and Stiglitz (1980) carries through: low-budget agents benefit from price revelation generated by high-budget agents, with the size of the benefit increasing in the precision-weighted mass of well-informed types.

IA.5.3 Dynamic extension with attention revision over time

The baseline equilibrium is one-shot: the agent commits to a signal structure, observes the signal, takes an action, and incurs loss. We now develop the dynamic extension in which the agent allocates attention across multiple periods, observes signals sequentially, and revises beliefs as new information arrives. Under Gaussian AR(1) dynamics the model admits a closed-form characterization via Kalman filtering, the optimal allocation reduces to a tractable Bellman recursion, and the three predicted features (inertia, sluggish response, budget pooling) follow as derived consequences. We focus on the single-agent problem; embedding the dynamic agent in the heterogeneous-attention equilibrium of Subsection IA.5.2 follows the same logic.

Setup. Discrete time $t = 1, 2, \ldots, T$. The state $x_t \in \mathbb{R}^d$ follows a Gaussian AR(1) process

$$x_t \;=\; \rho\, x_{t-1} + v_t, \qquad v_t \sim \mathcal{N}(\mathbf{0}, Q), \qquad x_0 \sim \mathcal{N}(\mathbf{0}, \Sigma_0), \tag{IA.24}$$

with persistence $|\rho| < 1$ for stationarity and innovation covariance $Q \succ 0$ in the same eigenbasis as Σ_0. The agent has quadratic welfare loss with matrix $W \succ 0$, time discount $\beta \in (0, 1]$, and total bit budget $\bar{B} \geq 0$ to be allocated across periods, $\sum_{t=1}^{T} B_t \leq \bar{B}$ with $B_t \geq 0$. Within each period, the agent applies RTQ to the rotated state at bit budget B_t, observes the quantized signal, and updates beliefs.

High-rate regime. We maintain the high-rate Lloyd-Max approximation throughout: per-period bit allocations are assumed to satisfy $B_t/d \geq \underline{b}$ for a threshold $\underline{b}$ large enough that the Panter-Dite asymptotic MSE $\approx C_T \cdot 4^{-B_t/d} \cdot \mathrm{Var}$ is accurate. In numerical practice $\underline{b} \approx 1$ bit per coordinate suffices; the formulas below break down at very low per-coordinate rates and require a separate low-rate analysis with bounded posterior variance.

Per-period Kalman update. Let $P_{t|t-1} = \mathrm{Var}(x_t \mid \mathscr{F}_{t-1})$ denote the pre-update posterior covariance, where $\mathscr{F}_{t-1}$ is the agent's information set after observing signals through period $t-1$. Diag-

onalize in the time-invariant eigenbasis V of Σ_0 (and of Q, by assumption), so $P_{t|t-1} = V\Lambda_{t|t-1}V^\top$ with $\Lambda_{t|t-1}$ diagonal.

The RTQ signal at period t, in the rotated frame and under the high-rate Lloyd-Max model on the Beta marginal, is

$$\tilde{y}_t \;=\; \tilde{x}_t + \tilde{\varepsilon}_t, \qquad \tilde{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \sigma_t^2 \Lambda_{t|t-1}), \qquad \sigma_t^2 \;=\; C_T \cdot 4^{-B_t/d}, \quad C_T = \pi\sqrt{3}/2. \tag{IA.25}$$

The Kalman update gives the post-update covariance

$$\Lambda_{t|t} \;=\; \Lambda_{t|t-1} \cdot \phi(B_t), \qquad \phi(B) \;\equiv\; \frac{C_T \cdot 4^{-B/d}}{1 + C_T \cdot 4^{-B/d}}, \qquad B/d \geq \underline{b}, \tag{IA.26}$$

where $\phi(B)$ is the per-period RTQ shrinkage factor on the admissible high-rate range $B/d \geq \underline{b}$. On this range ϕ is strictly positive, strictly decreasing in B, and converges to 0 as $B \to \infty$ (full attention, posterior collapses to a point). The low-rate boundary $B \to 0$ is outside the high-rate regime; under the no-attention convention $\phi(0) = 1$ (posterior unchanged), but the formulas below are stated and used only for $B/d \geq \underline{b}$. The pre-update covariance at $t+1$ follows the standard prediction step:

$$\Lambda_{t+1|t} \;=\; \rho^2 \Lambda_{t|t} + Q \;=\; \rho^2 \phi(B_t)\, \Lambda_{t|t-1} + Q. \tag{IA.27}$$

Equations (IA.26) and (IA.27) together fully characterize the dynamics of the agent's posterior covariance as a function of the bit allocation $\{B_t\}$.

Stationary covariance: pre- and post-update fixed points. The recursion (IA.26)-(IA.27) has two distinct stationary objects, which we distinguish:

$$\Lambda_\infty^{\text{pre}}(B) \;=\; \frac{Q}{1 - \rho^2 \phi(B)}, \qquad \Lambda_\infty^{\text{post}}(B) \;=\; \phi(B) \cdot \Lambda_\infty^{\text{pre}}(B) \;=\; \frac{\phi(B)\, Q}{1 - \rho^2 \phi(B)}. \tag{IA.28}$$

Under a constant allocation $B_t = B$, the pre-update covariance $\Lambda_{t|t-1}$ converges to $\Lambda_\infty^{\text{pre}}(B)$ and the post-update covariance $\Lambda_{t|t}$ converges to $\Lambda_\infty^{\text{post}}(B)$. The Bellman state in (IA.30) is the pre-update covariance, so $\Lambda_\infty^{\text{pre}}$ is the relevant fixed point for the state recursion; the per-period welfare loss $\text{tr}(\tilde{W}\,\Lambda_{t|t})$ involves the post-update covariance, which equals $\phi(B)\text{tr}(\tilde{W}\,\Lambda_\infty^{\text{pre}})$ in stationarity. Both are continuous and monotone-decreasing in B under the high-rate regime ($B/d \geq \underline{b} > 0$), with $\Lambda_\infty^{\text{post}}(\infty) = 0$ as $B \to \infty$.

Bellman recursion for optimal allocation. The agent minimizes total discounted welfare loss

$$\mathscr{L}_{\text{total}}(\{B_t\}) \;=\; \sum_{t=1}^{T} \beta^{t-1} \cdot \operatorname{tr}\big(\tilde{W}\,\Lambda_{t|t}\big) \;=\; \sum_{t=1}^{T} \beta^{t-1} \cdot \phi(B_t) \cdot \operatorname{tr}\big(\tilde{W}\,\Lambda_{t|t-1}\big)\,, \tag{IA.29}$$

subject to $\sum_t B_t \le \bar{B}$, $B_t \ge 0$, and the dynamics (IA.26)-(IA.27); here $\tilde{W} = V^\top W V$ is the welfare matrix in the rotated frame. Let $V_t(\Lambda, b)$ denote the optimal cost-to-go starting at period t with pre-update covariance $\Lambda_{t|t-1} = \Lambda$ and remaining budget $b = \bar{B} - \sum_{s<t} B_s$. The Bellman equation is

$$V_t(\Lambda, b) \;=\; \min_{0 \le B_t \le b} \;\; \phi(B_t) \cdot \operatorname{tr}(\tilde{W}\Lambda) + \beta \cdot V_{t+1}\big(\rho^2 \phi(B_t)\Lambda + Q,\; b - B_t\big)\,, \tag{IA.30}$$

with terminal condition $V_{T+1} \equiv 0$. The first-order condition with respect to B_t gives the Euler equation linking marginal welfare savings at t to the resulting reduction in V_{t+1} through the predict step. Because ϕ is decreasing in B_t, the marginal benefit of an extra bit is largest where the pre-update covariance $\Lambda_{t|t-1}$ is largest in the welfare-weighted sense.

Optimal stationary policy in the infinite-horizon case.

Proposition IA.8 (Optimal stationary allocation in the infinite-horizon discounted case)**.** *Consider the infinite-horizon problem ($T = \infty$) with discount factor $\beta \in (0,1)$, stationary AR(1) dynamics (ρ, Q), and the high-rate regime maintained for the bit allocation. The agent's stationary problem is closed either by a per-period information cost μB with shadow price $\mu > 0$ on the bit budget, or equivalently by an exogenous per-period budget $\bar{B}_{\text{pp}} > 0$ that pins $B \equiv \bar{B}_{\text{pp}}$ with μ as the shadow value of relaxing that constraint; without one of these closures the welfare-loss minimization sends $B \to \infty$. Within the class of stationary (time-invariant) policies $B_t \equiv B \ge 0$, and given $\mu > 0$ (whether structural or as the shadow value of a fixed per-period budget), the optimal stationary allocation B^* solves the Euler condition*

$$\big|\phi'(B^*)\big| \cdot \frac{\operatorname{tr}\big(\tilde{W}\,\Lambda_\infty^{\text{pre}}(B^*)\big)}{1 - \beta\rho^2\phi(B^*)} \;=\; \mu, \tag{IA.31}$$

where $\Lambda_\infty^{\text{pre}}(B) = Q/(1-\rho^2\phi(B))$ is the pre-update stationary covariance from (IA.28)*. On the high-rate range where the per-period loss $g(B) = \phi(B)\operatorname{tr}(\tilde{W}\Lambda_\infty^{\text{pre}}(B))/(1-\beta\rho^2\phi(B))$ is strictly convex (Lemma IA.2 below), the Euler condition pins down a unique B^*. Under stationary initial conditions $\Lambda_{1|0} = \Lambda_\infty^{\text{pre}}(B^*)$, the constant policy $B_t \equiv B^*$ attains the steady-state per-period welfare loss $\phi(B^*)\operatorname{tr}(\tilde{W}\,\Lambda_\infty^{\text{pre}}(B^*))$ at every period.*

Proof. With stationary dynamics and stationary cost, the Bellman recursion admits a time-invariant value function $V_\infty(\Lambda)$ on the pre-update state Λ in the infinite-horizon case. Substituting the linear

ansatz $V_\infty(\Lambda) = c + \mathrm{tr}(P\Lambda)$ into the Bellman equation $V_\infty(\Lambda) = \phi(B)\mathrm{tr}(\tilde{W}\Lambda) + \beta V_\infty(\rho^2\phi(B)\Lambda + Q)$ and matching coefficients on Λ yields $P = \phi(B)\tilde{W}/(1-\beta\rho^2\phi(B))$. The first-order condition with respect to B, evaluated at the stationary pre-update covariance $\Lambda_\infty^{\mathrm{pre}}(B^*)$, gives

$$\phi'(B^*)\mathrm{tr}(\tilde{W}\Lambda_\infty^{\mathrm{pre}}(B^*)) + \beta\rho^2\phi'(B^*)\mathrm{tr}(P\Lambda_\infty^{\mathrm{pre}}(B^*)) + \mu \;=\; 0.$$

Substituting $P = \phi(B^*)\tilde{W}/(1-\beta\rho^2\phi(B^*))$ and collecting:

$$\phi'(B^*)\mathrm{tr}(\tilde{W}\Lambda_\infty^{\mathrm{pre}}(B^*))\left[1 + \frac{\beta\rho^2\phi(B^*)}{1-\beta\rho^2\phi(B^*)}\right] = -\mu,$$

and simplifying $1 + \beta\rho^2\phi/(1-\beta\rho^2\phi) = 1/(1-\beta\rho^2\phi)$ gives (IA.31) after taking absolute values (since $\phi'(B^*) < 0$). Uniqueness on the high-rate range follows from Lemma IA.2. □

Lemma IA.2 (Strict convexity of stationary loss on the high-rate range)**.** *Define* $g(B) \equiv \phi(B)\mathrm{tr}(\tilde{W}\Lambda_\infty^{\mathrm{pre}}(B))/(1-\beta\rho^2\phi(B))$ *on the high-rate range* $B/d \geq \underline{b}$*. There exists* $\underline{b}^* \geq \underline{b}$ *such that* g *is strictly convex on* $\{B : B/d \geq \underline{b}^*\}$*, hence the Euler condition* (IA.31) *has at most one stationary minimizer on that range.*

Proof. Write $u = \phi(B)$; on the high-rate range ϕ is a strictly decreasing, strictly convex, and analytic function of B with image $u \in (0, u_{\max}]$ where $u_{\max} = \phi(\underline{b}d) < 1$. The stationary loss is $g(B) = \mathrm{tr}(\tilde{W}Q)\cdot u/[(1-\rho^2 u)(1-\beta\rho^2 u)]$, a smooth function of u. Its second derivative with respect to u is positive on a neighborhood of $u = 0$ and the second derivative with respect to B is $g_{BB} = g_{uu}(\phi')^2 + g_u\phi''$, with both $\phi''(B) > 0$ on the high-rate range and $g_u = \mathrm{tr}(\tilde{W}Q)\cdot[(1-\rho^2 u)(1-\beta\rho^2 u) + u(\rho^2(1-\beta\rho^2 u) + \beta\rho^2(1-\rho^2 u))]/[(1-\rho^2 u)^2(1-\beta\rho^2 u)^2] > 0$. For u small enough (equivalently, B/d large enough), $g_{BB} > 0$ strictly; let $\underline{b}^*$ be any threshold above which both terms remain positive. Strict convexity on $\{B/d \geq \underline{b}^*\}$ follows, and uniqueness of the Euler-condition root on this range is immediate. □

The proposition restricts to stationary policies and the infinite-horizon case. For finite horizon $T < \infty$, the global optimum is given by the Bellman Euler condition in the next proposition and may deviate from any stationary policy. In short horizons, front-loading attention can be globally optimal: concentrating the bit budget in early periods reduces early posterior variance enough that the welfare savings dominate the cost of higher late-period variance. As T grows, the value of front-loading shrinks and the optimum approaches the stationary B^*. The following proposition characterizes the optimum for any finite T and any initial conditions.

Proposition IA.9 (Bellman characterization under maintained regularity)**.** *The Bellman recursion* (IA.30) *characterizes the finite-horizon optimum. At each* (t, Λ, b) *with* $\Lambda \succ 0$*, under the*

maintained high-rate convexity conditions of Lemma IA.2 together with the regularity hypothesis that $V_{t+1}(\cdot, b-B)$ is non-decreasing in the Loewner order on Λ and convex in b, the one-step Bellman summand is strictly convex in B, so the period-t minimizer $B_t^*(\Lambda, b)$ is unique. By Bellman's principle of optimality, the resulting trajectory $\{B_t^*\}$ defines the finite-horizon optimum.

Proof. Under the stated regularity hypothesis on V_{t+1}, the Bellman summand $\phi(B)\mathrm{tr}(\tilde{W}\Lambda)+\beta V_{t+1}(\rho^2\phi(B)\Lambda+Q, b-B)$ is a sum of a strictly convex term (by Lemma IA.2 applied to the first piece) and a convex term in B on the high-rate range. The minimum over the compact convex set $[0,b]$ is therefore attained uniquely. Bellman's principle of optimality then gives the finite-horizon optimum trajectory by per-step minimization. □

Proposition IA.9 gives a per-step uniqueness statement and a computational characterization of the optimum, conditional on the regularity hypothesis on V_{t+1}. We do not establish that hypothesis recursively from $V_{T+1} \equiv 0$, because the minimum over B of a family of convex-in-Λ functions is not in general convex in Λ; verifying that the value function inherits monotonicity and convexity from the per-step problem would require additional structure on the Bellman operator that goes beyond what we use in this paper. The per-step characterization remains sufficient for the substantive uses below: under stationary initial conditions, Proposition IA.8 gives the optimal stationary policy with the cleaner Lemma IA.2 uniqueness argument; under non-stationary initial conditions, the optimal allocation front-loads or back-loads attention depending on whether $\Lambda_{1|0}$ is above or below $\Lambda_\infty^{\mathrm{pre}}(B^*)$, with the trajectory characterized by the Bellman recursion subject to the stated regularity.

Turnpike intuition (informal implication). The finite-T optimum is expected to approach the stationary policy in the interior of long horizons. Under the high-rate regime, ϕ is Lipschitz and bounded, the Kalman recursion has a bounded invariant set for the pre-update covariance (any constant policy keeps $\Lambda_{t|t-1}$ in the compact range $[\Lambda_\infty^{\mathrm{pre}}(\overline{B}), Q/(1-\rho^2)]$), and the discounted Bellman operator on continuous bounded functions over the compact state set is a contraction with modulus $\beta < 1$. The Banach fixed-point theorem then gives $V_T(\cdot,t) \to V_\infty(\cdot)$ uniformly as $T-t \to \infty$. If the Lagrange multiplier sequence on the total budget constraint converges to its stationary value μ^* and the optimal policy is continuous in the value-function derivative (which holds under standard regularity by the implicit-function theorem applied to the strictly convex per-step problem of Proposition IA.9), then for any $\varepsilon > 0$ and $\delta \in (0, 1/2)$ there exists $T_0 = T_0(\varepsilon, \delta)$ such that

$$\left| B_t^{*,T} - B^* \right| < \varepsilon \qquad \text{for all } T \geq T_0 \text{ and all } t \in [\delta T, (1-\delta)T], \tag{IA.32}$$

at geometric rate in the horizon distance to the boundary. This is the standard turnpike pattern of Cass (1965) and McKenzie (1976) for budget-constrained dynamic programs. We state this

property as an informal implication rather than as a theorem: a fully rigorous proof requires additional regularity conditions on the value-function derivative and on the multiplier sequence that go beyond the scope of the present paper, and is left for follow-up work. In the empirical sections we use the Bellman recursion as the computational characterization (Proposition IA.9) and report stationary-policy results for stationary initial conditions (Proposition IA.8).

Three derived features. The Bellman recursion produces three qualitative features observed in the dynamic-RI literature.

Inertia. Equation (IA.26) shows that the post-update covariance is the pre-update covariance scaled by $\phi(B_t) \in (0,1]$, not replaced. The agent's belief at t is a precision-weighted average of the new RTQ signal and the carry-over from $t-1$, with weight $\rho^2\phi(B_{t-1})$ on the past and weight $1-\phi(B_t)$ on the new signal. As $\rho \to 1$ (high persistence), the past weight rises and the agent's belief moves slowly. This is the formal sense in which RTQ reuses past attention rather than re-rotating each period.

Sluggish response. A structural break at $t = t^*$ that shifts Q to $Q' \neq Q$ propagates through (IA.27) only after the agent observes signals reflecting the new innovation variance. The pre-update covariance $\Lambda_{t^*+k|t^*+k-1}$ converges to the new stationary pre-update value $\Lambda_\infty^{\text{pre},\prime}(B^*) = Q'/(1-\rho^2\phi(B^*))$ at geometric rate $(\rho^2\phi(B^*))^k$; for small B^* or high ρ, this convergence is slow. The agent's belief catches up to the new regime sluggishly, with half-life $-\log 2/\log(\rho^2\phi(B^*))$ periods.

Budget pooling. In the infinite-horizon discounted case with stationary dynamics, the optimal stationary policy is constant per-period allocation $B_t^* = B^*$ (Proposition IA.8); the agent does not front-load or back-load. In finite T, the optimum can deviate from equal allocation, with front-loading favored when initial uncertainty $\Lambda_{1|0}$ is above the stationary pre-update covariance and back-loading favored when it is below. The turnpike intuition above implies that for sufficiently long horizons the interior of the policy approaches the stationary B^*, with the front-/back-loading effects confined to the boundary periods.

Connection to the static framework. Under the stationary policy and stationary initial conditions, the dynamic problem reduces to a sequence of static RTQ problems with per-period budget $\bar{B}/T$ and per-period welfare-weighted covariance $\tilde{W}\,\Lambda_\infty^{\text{post}}(\bar{B}/T)$, equivalent to $\phi(\bar{B}/T)\tilde{W}\,\Lambda_\infty^{\text{pre}}(\bar{B}/T)$. The parallel-target sufficient statistic of Section IA.2 generalizes from the static $\xi = N \cdot 4^{-B/(Nd)}$ to the dynamic $\xi_{\text{dyn}} = T \cdot 4^{-\bar{B}/(Td)}$, with the persistence parameter entering through the stationary covariance scaling $1/(1-\rho^2\phi)$.

Corollary IA.3 (Dynamic loss bound under equal allocation, Gaussian AR(1))**.** *Under Gaussian AR(1) state dynamics with innovation covariance Q, equal allocation $B_t \equiv \bar{B}/T$, and stationary*

initial conditions $\Lambda_{1|0} = \Lambda_\infty^{\text{pre}}(\bar{B}/T)$*, the total welfare loss is bounded by*

$$\mathscr{L}_{\text{total}} \;\le\; C_T \cdot \chi(\tilde{W}) \cdot \frac{1-\beta^T}{1-\beta} \cdot \mathscr{L}_{\text{lb}}^{\text{sphere}}(\bar{B}/T), \tag{IA.33}$$

where $\mathscr{L}_{\text{lb}}^{\text{sphere}}(b)$ *is the static minimax rate-distortion floor at per-coordinate bit rate* b/d *from Proposition 1 of the main text.*

Proof. Under the equal-allocation policy $B_t \equiv \bar{B}/T$, applying the Kalman recursion (IA.26)-(IA.27) starting from $\Lambda_{1|0} = \Lambda_\infty^{\text{pre}}(\bar{B}/T)$ gives $\Lambda_{t|t-1} = \Lambda_\infty^{\text{pre}}(\bar{B}/T)$ for all t: the stationary pre-update covariance is preserved by the constant-allocation policy, by direct substitution into (IA.26)-(IA.27). Proposition 1 of the main text, applied to a single period at per-coordinate bit rate $\bar{B}/(Td)$, gives $\phi(\bar{B}/T) \cdot \text{tr}(\tilde{W}\,\Lambda_\infty^{\text{pre}}(\bar{B}/T)) \le C_T \chi(\tilde{W}) \cdot \mathscr{L}_{\text{lb}}^{\text{sphere}}(\bar{B}/T)$. Summing the discounted per-period losses $\sum_{t=1}^{T} \beta^{t-1} = (1-\beta^T)/(1-\beta)$ gives the bound. □

Corollary IA.3 bounds the dynamic welfare loss under equal allocation in the Gaussian AR(1) case, where the Kalman recursion gives the closed-form posterior and the stationary covariance has the closed form $\Lambda_\infty^{\text{pre}}(B) = Q/(1-\rho^2\phi(B))$. For non-Gaussian innovations with finite second moments, the static main-text bound continues to apply period by period to the corresponding posterior error covariance, but the closed-form Kalman recursion and the stationary covariance formula do not carry over; the dynamic worst-case bound in the non-Gaussian case requires the per-period covariance trajectory under whatever signal structure the agent uses, which is no longer closed-form. The Gaussian-AR(1) case in (IA.33) is the substantive content.

Comparison to Steiner-Stewart-Matejka and scope. Steiner et al. (2017) characterize dynamic rational inattention under Shannon capacity and find the same three features (inertia, sluggish response, an interior stationary allocation). The development above replaces Shannon capacity with the RTQ rate-distortion bit budget and inherits the static RTQ worst-case guarantee period by period; the closed-form covariance recursion holds for Gaussian AR(1). General non-Gaussian state dynamics inherit the static worst-case bound applied at each period to the corresponding posterior error covariance, but the closed-form Kalman recursion does not carry over. Embedding the dynamic agent in the equilibrium of Subsection IA.5.2 with full-information counterparties and noise traders is a substantial extension and a natural direction for follow-up work; the per-period equilibrium pricing follows the static analysis of Subsection IA.5.1 applied at each t with $\sigma_R^2 = C_T \cdot 4^{-B_t/d}$ and posterior covariance from (IA.26).

IA.6 Technical background: from Johnson-Lindenstrauss to rotate-then-quantize

This section derives the results that rotate-then-quantize relies on, in the notation of the paper. Section 2.2 gives the intuition with an example, and Lemma 1 in Section 4.1 states the results formally. The aim here is to show where each result comes from, using tools familiar to economists (the law of large numbers, a change of variables, and a counting argument), rather than to reproduce the original proofs, for which we cite the sources. Table IA.3 at the end separates what the paper takes from this literature from what it adds.

IA.6.1 The Johnson-Lindenstrauss lemma

Lemma IA.3 (Johnson-Lindenstrauss)**.** *Let $\varepsilon \in (0,1)$ and let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be points in $\mathbb{R}^n$. There is a universal constant C such that if $k \geq C\varepsilon^{-2}\log N$, the map $f(\mathbf{x}) = k^{-1/2}G\mathbf{x}$, where G is a $k \times n$ matrix with independent standard normal entries, satisfies*

$$(1-\varepsilon)\,\|\mathbf{x}_i - \mathbf{x}_j\|^2 \;\leq\; \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 \;\leq\; (1+\varepsilon)\,\|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad \textit{for all } i, j$$

with probability at least $1/2$.

The lemma is due to Johnson and Lindenstrauss (1984); Vershynin (2018, Section 5.3) gives a textbook proof. The reason the original dimension n does not enter the required k is short enough to state. Fix one vector $\mathbf{x}$ and write $\mathbf{g}_i$ for row i of G. Each $\mathbf{g}_i^\top \mathbf{x}$ is normal with mean zero and variance $\|\mathbf{x}\|^2$, whatever n is, and the k rows are independent. So

$$\|f(\mathbf{x})\|^2 \;=\; \frac{1}{k}\sum_{i=1}^{k}(\mathbf{g}_i^\top \mathbf{x})^2 \;=\; \|\mathbf{x}\|^2 \cdot \frac{\chi_k^2}{k},$$

an average of k independent draws whose mean is $\|\mathbf{x}\|^2$. The dimension n disappears at the first step and never returns. A standard chi-square tail bound gives $\Pr\big(|\chi_k^2/k - 1| > \varepsilon\big) \leq 2\exp(-ck\varepsilon^2)$ for a universal constant $c > 0$. The standard deviation of χ_k^2/k is $\sqrt{2/k}$, so quadrupling the number of kept coordinates halves the typical distortion. This is the diversification argument of Section 2.2: any single term $(\mathbf{g}_i^\top \mathbf{x})^2$ is very noisy, but the average of many independent terms settles near its mean, and panel (d) of Figure 2 shows the tightening. A union bound over the fewer than N^2 pairwise differences then requires only k of order $\varepsilon^{-2}\log N$, because the tail probability falls exponentially in k while the number of pairs grows only polynomially in N.

The paper does not use Lemma IA.3 in any proof. We include it for two reasons. First, the averaging argument above is the simplest example of concentration of measure, the property of

high-dimensional random matrices that also makes rotate-then-quantize work. Second, Zandieh et al. (2025) build their one-bit quantizer, QJL, directly on the Johnson-Lindenstrauss map (Section IA.6.6).

Lemma IA.3 and rotate-then-quantize compress in different ways. The Johnson-Lindenstrauss map reduces the number of coordinates from n to k and keeps each coordinate as a real number. Rotate-then-quantize keeps all d coordinates and reduces each to b bits. In the attention problem, the first corresponds to close attention to a few composite signals and the second to coarse attention to every direction.

IA.6.2 Coordinates of a randomly rotated vector

Let R be a random $d \times d$ orthogonal matrix drawn from the Haar (uniform) distribution on the orthogonal group, and let $\mathbf{u} \in S^{d-1}$ be a fixed unit vector.

Lemma IA.4 (Distribution of a rotated coordinate)**.** *The vector* $\mathbf{r} = R\mathbf{u}$ *is uniformly distributed on* S^{d-1}*. Each coordinate* r_i *has density*

$$f_d(r) \;=\; \frac{\Gamma(d/2)}{\sqrt{\pi}\,\Gamma\big((d-1)/2\big)}\,(1-r^2)^{(d-3)/2}, \qquad r \in [-1,1],$$

so that $(1+r_i)/2 \sim \mathrm{Beta}\big(\frac{d-1}{2}, \frac{d-1}{2}\big)$.

Proof. For any fixed orthogonal Q, the matrix QR has the same Haar distribution as R, so $Q\mathbf{r}$ has the same distribution as $\mathbf{r}$. The uniform distribution is the only distribution on S^{d-1} that is unchanged by every rotation, which gives the first claim.

For the density, take $i = d$ by symmetry. The set of points on S^{d-1} with $r_d = r$ is a sphere of dimension $d-2$ and radius $\sqrt{1-r^2}$, whose surface area is $A_{d-1}(1-r^2)^{(d-2)/2}$, where $A_k = 2\pi^{k/2}/\Gamma(k/2)$ is the surface area of the unit sphere in $\mathbb{R}^k$. A slab of height dr cuts the sphere along a band whose width along the surface is $dr/\sqrt{1-r^2}$. The probability that $r_d \in [r, r+dr]$ is the area of that band divided by A_d:

$$\frac{A_{d-1}}{A_d}\,(1-r^2)^{(d-2)/2}\,\frac{dr}{\sqrt{1-r^2}} \;=\; \frac{\Gamma(d/2)}{\sqrt{\pi}\,\Gamma\big((d-1)/2\big)}\,(1-r^2)^{(d-3)/2}\,dr.$$

□

A shortcut explains why the rotated coordinates are nearly normal. A uniform point on the sphere can be generated as $\mathbf{r} = \mathbf{g}/\|\mathbf{g}\|$ with $\mathbf{g}$ standard normal in $\mathbb{R}^d$, because the standard normal distribution looks the same in every direction. By the law of large numbers, $\|\mathbf{g}\|^2/d \to 1$, so $r_i \approx g_i/\sqrt{d}$, a normal variable with mean zero and variance $1/d$. Nothing about the original vector

$\mathbf{u}$ enters this argument. The same representation shows why the coordinates are only weakly dependent: they share the single common factor $\|\mathbf{g}\|$, whose relative variation vanishes as d grows.

The shape of f_d changes with d. At $d = 2$ it is the U-shaped arcsine density, at $d = 3$ it is uniform on $[-1, 1]$, and at $d = 4$ it is the semicircle density. For larger d it concentrates near zero: Stirling's formula gives $\Gamma(d/2)/\Gamma((d-1)/2) \approx \sqrt{d/2}$, and $(1-r^2)^{(d-3)/2} \approx e^{-dr^2/2}$ for small r, so f_d is close to the normal density with mean zero and variance $1/d$. This density depends only on d. It does not depend on the vector $\mathbf{u}$ or on the prior F, which is what allows one quantizer, computed once, to serve every coordinate of every state.

As Section 4.1 notes, the lemma describes the distribution over draws of R for a fixed $\mathbf{u}$. Once R is drawn and held fixed, the coordinates of $R\mathbf{x}$ for $\mathbf{x} \sim F$ follow whatever structure F has. The worst-case guarantee is a statement about expected loss over R for each fixed direction.

IA.6.3 Dependence across rotated coordinates

The rotated coordinates are not independent, since $\sum_i r_i^2 = 1$. The next lemma shows that the dependence is weak and fades with d.

Lemma IA.5 (Moments of rotated coordinates)**.** *For* $\mathbf{r}$ *uniform on* S^{d-1} *and* $i \neq j$*:* $\mathbb{E}[r_i] = 0$*,* $\mathbb{E}[r_i^2] = 1/d$*,* $\mathbb{E}[r_i r_j] = 0$*,* $\mathbb{E}[r_i^4] = 3/\big(d(d+2)\big)$*, and* $\mathbb{E}[r_i^2 r_j^2] = 1/\big(d(d+2)\big)$*. Hence* $\mathrm{Corr}(r_i^2, r_j^2) = -1/(d-1)$.

Proof. Write $\mathbf{r} = \mathbf{g}/\|\mathbf{g}\|$ with $\mathbf{g}$ standard normal in $\mathbb{R}^d$. The direction $\mathbf{r}$ is independent of the length $\|\mathbf{g}\|$. The first three moments follow from symmetry and $\sum_i r_i^2 = 1$. For the fourth moments, $g_i^2 g_j^2 = \|\mathbf{g}\|^4 r_i^2 r_j^2$ and independence give $\mathbb{E}[g_i^2 g_j^2] = \mathbb{E}\|\mathbf{g}\|^4 \cdot \mathbb{E}[r_i^2 r_j^2]$. Since $\|\mathbf{g}\|^2$ is chi-square with d degrees of freedom, $\mathbb{E}\|\mathbf{g}\|^4 = d(d+2)$. With $\mathbb{E}[g_i^4] = 3$ and $\mathbb{E}[g_i^2 g_j^2] = 1$, the two fourth moments follow. Then $\mathrm{Var}(r_i^2) = 2(d-1)/\big(d^2(d+2)\big)$ and $\mathrm{Cov}(r_i^2, r_j^2) = -2/\big(d^2(d+2)\big)$, and their ratio is $-1/(d-1)$. □

The weak dependence does not affect the expected error of coordinate-by-coordinate quantization. Let $q(\cdot)$ be a scalar quantizer applied to each rotated coordinate and $\tilde{\mathbf{u}} = R^\top q(\mathbf{r})$ the reconstruction. Since R preserves lengths,

$$\mathbb{E}\|\mathbf{u} - \tilde{\mathbf{u}}\|^2 \;=\; \mathbb{E}\|\mathbf{r} - q(\mathbf{r})\|^2 \;=\; \sum_{i=1}^{d} \mathbb{E}\big[(r_i - q(r_i))^2\big] \;=\; d \cdot D_b(f_d),$$

where $D_b(f_d)$ is the expected squared error of q on one coordinate with density f_d. The second equality is linearity of expectation and holds whatever the dependence across coordinates. The dependence matters only for whether a quantizer that codes all coordinates jointly could do better.

Section IA.6.5 shows that it can improve by at most the factor $C_T \gamma_d^c(b)$, which is close to C_T when d is large.

IA.6.4 The Lloyd-Max quantizer and the constant C_T

A b-bit scalar quantizer divides $[-1,1]$ into 2^b cells and assigns each cell a representative value. For a given density, the quantizer that minimizes expected squared error satisfies two conditions (Lloyd, 1982; Max, 1960). First, each boundary between cells is the midpoint between the two neighboring representative values. Second, each representative value is the mean of the density within its cell. For a log-concave density, which f_d is for $d \geq 3$, these conditions have a unique solution, and alternating between them converges to it. Because f_d depends only on d, the quantizer is computed once, before any data arrive. The rate 4^{-b} has a simple source. Each extra bit doubles the number of cells, which roughly halves the width of each cell, and the squared error within a cell is proportional to the square of its width. So each extra bit cuts the squared error by a factor of about four.

Zandieh et al. (2026, Theorem 1) show that for every unit vector $\mathbf{u}$ and every $b \geq 1$,

$$\mathbb{E}_R \|\mathbf{u} - \tilde{\mathbf{u}}\|^2 \;=\; d \cdot D_b(f_d) \;\leq\; C_T \cdot 4^{-b}, \qquad C_T = \frac{\pi\sqrt{3}}{2} \approx 2.72.$$

The constant comes from the standard high-resolution approximation for the error of an optimal scalar quantizer, $D_b(f) \approx \frac{1}{12}\left(\int f^{1/3}\right)^3 4^{-b}$ (Panter and Dite, 1951). For the normal density with variance $1/d$, $\left(\int f^{1/3}\right)^3 = 6\sqrt{3}\,\pi/d$, so $d \cdot D_b \approx (\pi\sqrt{3}/2)\,4^{-b}$; the exact density f_d gives the same leading term. At small bit budgets the exact Lloyd-Max errors are well below the bound. Zandieh et al. (2026) report $d \cdot D_b \approx 0.36$, 0.117, 0.030, and 0.009 for $b = 1,2,3,4$, against $C_T 4^{-b} = 0.68$, 0.17, 0.043, and 0.011. Lemma 2 in Appendix A restates this bound in the form used in the proof of Proposition 1.

IA.6.5 The lower bound

No quantizer can do much better. A counting argument shows why 4^{-b} is the right order. With bd bits the quantizer can name at most 2^{bd} distinct reconstructions. To place every point of the sphere within distance ε of one of them requires on the order of $\varepsilon^{-(d-1)}$ points, since the sphere has dimension $d-1$. Matching the two counts gives ε of order $2^{-bd/(d-1)}$, and hence squared error of order $4^{-bd/(d-1)}$, which is close to 4^{-b} when d is large.

Zandieh et al. (2026, Theorem 3) state the lower bound as exactly 4^{-b} for every d. In low dimensions a code can do slightly better, because the exponent is $bd/(d-1)$ rather than b. For example, with $d = 2$ and $b = 2$, sixteen equally spaced points on the unit circle give a worst-case

squared error of $4\sin^2(\pi/32) \approx 0.038$, below $4^{-2} = 0.0625$. The next two lemmas give lower bounds that hold in every dimension and approach 4^{-b} as d grows. The first covers codes with a fixed number of bits. The second covers every signal that satisfies a mutual-information constraint, which is the constraint in the model, and is the version used in Lemma 1(iv) and Proposition 1.

Lemma IA.6 (Lower bound for codes)**.** *Let $d \geq 2$ and let Q be any quantizer, deterministic or randomized, that maps unit vectors in S^{d-1} to at most $N = 2^{bd} \geq 2$ reconstruction points in $\mathbb{R}^d$. Then there is a unit vector $\mathbf{u}$ with*

$$\mathbb{E}\|\mathbf{u} - Q(\mathbf{u})\|^2 \;\geq\; \frac{d-1}{d+1}\left(\frac{2}{N}\right)^{2/(d-1)} \;=\; \frac{4^{-b}}{\gamma_d^c(b)}, \qquad \gamma_d^c(b) = \frac{d+1}{d-1}\cdot 2^{2(b-1)/(d-1)}.$$

Proof. The worst case over $\mathbf{u}$ is at least the average when $\mathbf{u}$ is uniform on the sphere, and for a uniform input a randomized quantizer is an average of deterministic ones, so it is enough to bound the average error of a deterministic quantizer with reconstruction points $\mathbf{c}_1, \ldots, \mathbf{c}_N$. Fix $r \in [0,1]$ and a point $\mathbf{c} \in \mathbb{R}^d$. The set of unit vectors within distance r of $\mathbf{c}$ is a spherical cap, and over all $\mathbf{c}$ its angular radius is largest, equal to $\theta = \arcsin r$, when $\|\mathbf{c}\| = \sqrt{1-r^2}$. The share of the sphere covered by a cap of angular radius $\theta \leq \pi/2$ is

$$\sigma_d(\theta) = \frac{\int_0^\theta \sin^{d-2}\varphi\, d\varphi}{\int_0^\pi \sin^{d-2}\varphi\, d\varphi} \;\leq\; \tfrac{1}{2}\sin^{d-1}\theta.$$

To see the inequality, let $g(\theta) = \frac{1}{2}I_d \sin^{d-1}\theta - \int_0^\theta \sin^{d-2}\varphi\, d\varphi$ with $I_d = \int_0^\pi \sin^{d-2}\varphi\, d\varphi$. Then $g(0) = g(\pi/2) = 0$ and $g'(\theta) = \sin^{d-2}\theta\left[\frac{1}{2}(d-1)I_d\cos\theta - 1\right]$, which is positive and then negative on $(0,\pi/2)$ because $\frac{1}{2}(d-1)I_d > 1$ for every $d \geq 2$. So $g \geq 0$ on $[0,\pi/2]$. By the union bound over the N points, the probability that a uniform $\mathbf{u}$ lies within distance r of its nearest reconstruction point is at most $\frac{N}{2}r^{d-1}$. Let $r_0 = (2/N)^{1/(d-1)} \leq 1$. Writing the expected squared distance as the integral of its tail probability,

$$\mathbb{E}\min_i \|\mathbf{u} - \mathbf{c}_i\|^2 \;\geq\; \int_0^{r_0^2}\left(1 - \tfrac{N}{2}t^{(d-1)/2}\right)dt \;=\; r_0^2\left(1 - \frac{2}{d+1}\right) \;=\; \frac{d-1}{d+1}\left(\frac{2}{N}\right)^{2/(d-1)}.$$

Substituting $N = 2^{bd}$ gives the second expression. □

Lemma IA.7 (Lower bound for any signal with limited mutual information)**.** *Let $d \geq 2$, let $\mathbf{u}$ be uniform on S^{d-1}, and let $\mathbf{s}$ be any signal with $I(\mathbf{u};\mathbf{s}) \leq bd$ bits. Then every reconstruction $\hat{\mathbf{u}} = \hat{\mathbf{u}}(\mathbf{s}) \in \mathbb{R}^d$ satisfies*

$$\mathbb{E}\|\mathbf{u} - \hat{\mathbf{u}}\|^2 \;\geq\; \frac{4^{-b}}{\gamma_d(b)}, \qquad \gamma_d(b) = 4^{b(2d-1)/(d-1)^2}\cdot\left(\frac{d+1}{2}\right)^{2d/(d-1)^2}.$$

Proof. Let σ denote the uniform distribution on the sphere and $P_{\hat{\mathbf{u}}}$ the distribution of $\mathbf{u}$ given the reconstruction. Since $\hat{\mathbf{u}}$ is a function of $\mathbf{s}$ and the marginal distribution of $\mathbf{u}$ is σ,

$$I(\mathbf{u};\mathbf{s}) \;\geq\; I(\mathbf{u};\hat{\mathbf{u}}) \;=\; \mathbb{E}_{\hat{\mathbf{u}}}\big[\mathrm{KL}(P_{\hat{\mathbf{u}}} \,\|\, \sigma)\big],$$

with information measured in nats. For any distribution P on the sphere and any function f, the Donsker-Varadhan inequality gives $\mathrm{KL}(P\|\sigma) \geq \mathbb{E}_P[f] - \log \mathbb{E}_\sigma[e^f]$. Fix a point $\mathbf{c} \in \mathbb{R}^d$ and $\alpha \in (0,(d-1)/2)$, and take $f(\mathbf{u}) = -\alpha \log \|\mathbf{u}-\mathbf{c}\|^2$. By the cap bound in the proof of Lemma IA.6, $\sigma(\|\mathbf{u}-\mathbf{c}\| \leq r) \leq \frac{1}{2} r^{d-1}$ for $r \leq 1$, so

$$\mathbb{E}_\sigma\big[\|\mathbf{u}-\mathbf{c}\|^{-2\alpha}\big] \;\leq\; 1 + \int_1^\infty \sigma\big(\|\mathbf{u}-\mathbf{c}\| < t^{-1/(2\alpha)}\big)\, dt \;\leq\; 1 + \frac{1}{2}\int_1^\infty t^{-(d-1)/(2\alpha)}\, dt \;=\; 1 + \frac{\alpha}{d-1-2\alpha} \;\equiv\; K_\alpha.$$

By Jensen's inequality, $\mathbb{E}_P[f] \geq -\alpha \log \mathbb{E}_P\|\mathbf{u}-\mathbf{c}\|^2$. Applying both bounds with $P = P_{\hat{\mathbf{u}}}$ and $\mathbf{c} = \hat{\mathbf{u}}$, and writing $D_{\hat{\mathbf{u}}} = \mathbb{E}[\|\mathbf{u}-\hat{\mathbf{u}}\|^2 \mid \hat{\mathbf{u}}]$,

$$\mathrm{KL}(P_{\hat{\mathbf{u}}} \,\|\, \sigma) \;\geq\; \alpha \log \frac{1}{D_{\hat{\mathbf{u}}}} - \log K_\alpha.$$

Taking expectations over $\hat{\mathbf{u}}$ and using Jensen's inequality once more, since $-\log$ is convex, $I(\mathbf{u};\mathbf{s}) \geq \alpha \log(1/D) - \log K_\alpha$ with $D = \mathbb{E}\|\mathbf{u}-\hat{\mathbf{u}}\|^2$. With $I(\mathbf{u};\mathbf{s}) \leq bd \log 2$ nats this gives

$$D \;\geq\; 2^{-bd/\alpha} K_\alpha^{-1/\alpha}.$$

Choosing $\alpha = (d-1)^2/(2d)$ gives $K_\alpha = (d+1)/2$, $2^{-bd/\alpha} = 4^{-bd^2/(d-1)^2}$, and $K_\alpha^{-1/\alpha} = ((d+1)/2)^{-2d/(d-1)^2}$, which is the stated bound. □

Both correction factors are close to one in the dimensions that matter here. With $b = 2$, the factor for codes, $\gamma_d^c(b)$, is 1.055 at $d = 64$, 1.013 at $d = 256$, 1.007 at $d = 500$, and 1.001 at $d = 3{,}000$. The factor for any signal with limited mutual information, $\gamma_d(b)$, is 1.22 at $d = 64$, 1.06 at $d = 256$, 1.03 at $d = 500$, and 1.007 at $d = 3{,}000$. A code with 2^{bd} values is a special case of a signal with bd bits of mutual information, so Lemma IA.7 also applies to codes; Lemma IA.6 is slightly sharper for them. Combined with Section IA.6.4, rotate-then-quantize is within the factor $C_T \gamma_d(b)$ of the best that any signal with the same capacity can achieve when the direction is uniform, which is about $C_T \approx 2.72$ when d is large. Proposition 1 extends both bounds to weighted quadratic loss, which adds the condition number $\chi(W)$.

IA.6.6 QJL and the inner-product version of TurboQuant

Zandieh et al. (2025) propose QJL, a one-bit version of the Johnson-Lindenstrauss map. For a $k \times d$

standard normal matrix S, a vector $\mathbf{y}$ is stored as $\operatorname{sign}(S\mathbf{y}) \in \{-1,+1\}^k$ together with its length $\|\mathbf{y}\|$. For any other vector $\mathbf{z}$, the estimator $\sqrt{\pi/2}\,k^{-1}\|\mathbf{y}\|\,(S\mathbf{z})^\top \operatorname{sign}(S\mathbf{y})$ is unbiased for the inner product $\mathbf{z}^\top\mathbf{y}$. Zandieh et al. (2026) use QJL as a second stage. The quantizer of Section IA.6.4 minimizes squared error but gives biased estimates of inner products, so their inner-product version spends one more bit per coordinate on a QJL sketch of the reconstruction error, which removes the bias.

The paper uses only the squared-error version. The agent's loss is quadratic in the gap between the action and the state, so expected loss depends on the reconstruction error through its second moments, and the action is the posterior mean, which already corrects for any bias in the mechanical reconstruction (Section 4.1). The inner-product stage is not needed.

IA.6.7 What the paper takes and what it adds

Ingredient	Source	Where used
Taken from the computer-science literature		
Random projection preserves distances (concentration of measure)	Johnson and Lindenstrauss (1984)	Intuition only (Lemma IA.3)
One-bit sketch for inner products (QJL)	Zandieh et al. (2025)	Not used; discussed in Section IA.6.6
Random rotation; Beta marginal of rotated coordinates	Zandieh et al. (2026), Han et al. (2026)	Section 4.1; Lemma IA.4
Optimal scalar quantizer for a known density	Lloyd (1982), Max (1960)	Step 3 of RTQ
Worst-case squared-error bound $C_T 4^{-b}$ for unweighted loss	Zandieh et al. (2026, Theorem 1)	Lemma 2
Minimax lower bound 4^{-b}	Zandieh et al. (2026, Theorem 3), Cover and Thomas (2006)	Proposition 1
Added in this paper		
Weighted quadratic loss and the factor $\chi(W)$	This paper	Proposition 1
Allocation of bits across coordinates by stakes (water-filling)	This paper	Proposition 3
Permanent loss floor of selective attention	This paper	Proposition 2
Loss concentration ratio $\rho(k)$ as a criterion for focus versus spread	This paper	Section 4.5
Parallel-target extension and regime diagram	This paper	Internet Appendix IA.2
Rational-inattention interpretation and portfolio application	This paper	Sections 3 and 5

Table IA.3: **What rotate-then-quantize takes from the computer-science literature and what this paper adds.** The upper part lists results the paper uses as given, with their sources. The lower part lists the paper's own results.

IA.6.8 Simulation with non-Gaussian priors: full results

Table IA.4 reports every case of the simulation in Section 4.6. All priors have mean zero and identity covariance. The Student t prior is scaled to unit variance. The two-regime prior draws

a standard deviation of 0.6 with probability 0.9 and 2.6 otherwise. The sparse prior sets three randomly chosen coordinates to normal draws scaled by $\sqrt{d/3}$ and the rest to zero. RTQ uses one random rotation per dimension and the Lloyd-Max quantizer for f_d; the Gaussian design uses the Lloyd-Max quantizer for $N(0,1)$. The code is in `TQ_RTQ_intuition_figure.ipynb`.

d	b	Prior	Gaussian design	Gaussian + length	RTQ	$C_T 4^{-b}$	4^{-b}
64	1	Gaussian	0.363	0.358	0.358	0.680	0.250
64	1	Student t (3 df)	0.598	0.359	0.360	0.680	0.250
64	1	Two volatility regimes	0.617	0.358	0.361	0.680	0.250
64	1	Sparse (3 large shocks)	1.362	1.337	0.359	0.680	0.250
64	2	Gaussian	0.117	0.114	0.114	0.170	0.062
64	2	Student t (3 df)	0.319	0.115	0.112	0.170	0.062
64	2	Two volatility regimes	0.316	0.115	0.115	0.170	0.062
64	2	Sparse (3 large shocks)	0.773	0.730	0.116	0.170	0.062
64	3	Gaussian	0.034	0.033	0.033	0.043	0.016
64	3	Student t (3 df)	0.179	0.034	0.034	0.043	0.016
64	3	Two volatility regimes	0.174	0.034	0.033	0.043	0.016
64	3	Sparse (3 large shocks)	0.503	0.452	0.034	0.043	0.016
256	1	Gaussian	0.363	0.362	0.362	0.680	0.250
256	1	Student t (3 df)	0.633	0.358	0.361	0.680	0.250
256	1	Two volatility regimes	0.619	0.362	0.363	0.680	0.250
256	1	Sparse (3 large shocks)	1.494	1.487	0.362	0.680	0.250
256	2	Gaussian	0.117	0.117	0.117	0.170	0.062
256	2	Student t (3 df)	0.281	0.116	0.118	0.170	0.062
256	2	Two volatility regimes	0.315	0.117	0.117	0.170	0.062
256	2	Sparse (3 large shocks)	0.967	0.945	0.117	0.170	0.062
256	3	Gaussian	0.035	0.034	0.034	0.043	0.016
256	3	Student t (3 df)	0.161	0.034	0.034	0.043	0.016
256	3	Two volatility regimes	0.172	0.034	0.034	0.043	0.016
256	3	Sparse (3 large shocks)	0.740	0.707	0.034	0.043	0.016

Table IA.4: **Squared error relative to $\mathbb{E}\|\mathbf{x}\|^2$ under four priors with the same covariance.** 4,000 draws per cell. $C_T 4^{-b}$ is the RTQ upper bound and 4^{-b} the large-d worst-case lower bound for any code with the same number of bits (the exact finite-d bound is $4^{-b}/\gamma_d(b)$, Lemma IA.6).

IA.7 Additional empirical results

This section collects the tables behind the robustness statements in Section 5. As in that section, RTQ in the welfare comparison and in the 13F tests refers to the RTQ allocation: the rule of

Proposition 3 applied to the eigendirections of the estimated covariance, without a random rotation. Subsections IA.7.2 and IA.7.3 run RTQ with a random rotation.

IA.7.1 Welfare comparison

Table IA.5 compares the welfare formula used in earlier versions of the paper with the corrected one. The earlier formula charged direction j a loss of $C_T^2 \lambda_j 4^{-b_j}$ against a prior of λ_j^2, so the comparison depended on the units of returns. Under the corrected formula, RTQ captures close to the full value of information in every window, and the informative comparison is with fair VNV. Table IA.6 repeats the comparison across estimation windows and covariance estimators. Table IA.7 shows how far the realized RTQ loss sits below the distribution-free bound of Proposition 1. Table IA.8 reports the gain over fair VNV across market-cap deciles and across universe sizes.

		Earlier formula		Corrected formula			
		vs. no information		vs. no information		vs. fair VNV	
Sample	κ	Median	% wins	Median	% wins	Median	% wins
All stocks, Ledoit-Wolf	N	-141	43	597	100	155	100
	$2N$	412	75	599	100	158	100
	$3N$	553	96	600	100	159	100
All stocks, sample	N	600	97	600	100	152	100
	$2N$	600	99	600	100	152	100
	$3N$	600	100	600	100	152	100
500 largest, Ledoit-Wolf	N	-548	10	599	100	75	100
	$2N$	310	76	600	100	76	100
	$3N$	527	94	600	100	76	100

Table IA.5: **Correction to the welfare loss formula.** In-sample gains in basis points, 221 quarters. The earlier draft computed the RTQ loss in direction j as $C_T^2 \lambda_j 4^{-b_j}$ and compared it with a prior loss of λ_j^2. The two terms have different units, so the gain changed when returns were measured in percent rather than decimals, and the sign of the gain in early windows came from this mismatch. The corrected loss is $\lambda_j^2 \min(C_T 4^{-b_j}, 1)$, as in Proposition 3, and does not depend on units. Against no information, the corrected gain is close to the full value of information in every window, so the informative comparison is with fair VNV, which spends the same bit budget on the top eigendirection.

	$\kappa = 2N$				Out-of-sample median		Slope on
	In sample	Out of sample					ex-ante $\rho(1)$
Specification	Median	Median	Mean	NW t	$\kappa = N$	$\kappa = 3N$	(t), $\kappa = 2N$
Ledoit-Wolf, 60 months (baseline)	158	206	247	4.84	86	239	−369 (-1.79)
Ledoit-Wolf, 36 months	192	159	233	5.07	84	186	−318 (-1.68)
Ledoit-Wolf, 120 months	145	174	233	5.19	53	203	−363 (-1.77)
Sample covariance, 60 m.	152	39	159	2.96	30	61	−29 (-0.11)
Constant correlation, 60 m.	104	76	186	3.57	69	80	−84 (-0.31)
POET, 60 months	149	56	173	3.20	53	57	−70 (-0.26)
500 largest, Ledoit-Wolf, 60 months	76	42	79	5.97	29	50	−39 (-0.50)

Table IA.6: **Welfare gain from spreading attention across covariance estimators and lookback windows.** Fair VNV minus RTQ welfare cost in basis points of annual certainty-equivalent return, computed as in Table 1. Each row changes one element of the baseline: the lookback window used to estimate the covariance, the estimator (sample covariance, constant-correlation shrinkage, or POET of Fan et al. (2013)), or the universe. Out of sample, losses are evaluated in the ex-ante eigendirections with the covariance of the next 24 months, estimated the same way. There are 213 out-of-sample quarters (197 with the 120-month lookback, which starts later). RTQ has the lower loss in every quarter, in and out of sample, at every budget, in every specification. The last column is the slope of the out-of-sample gain on ex-ante $\rho(1)$, with a Newey-West t-statistic (20 lags).

	Realized RTQ loss	Realized loss / bound	
Specification	(bps), median	Median	Maximum
Ledoit-Wolf, 60 months	0.845	0.008	0.275
Ledoit-Wolf, 36 months	1.086	0.011	0.285
Ledoit-Wolf, 120 months	1.119	0.011	0.462
Sample covariance, 60 months	0.000	0.000	0.014
Constant correlation, 60 months	0.019	0.000	0.034
POET, 60 months	0.000	0.000	0.013
500 largest stocks, Ledoit-Wolf, 60 months	0.226	0.002	0.195

Table IA.7: **How far below the Proposition 1 bound the realized RTQ loss sits.** Budget $\kappa = 2N$. The bound is the worst-case RTQ loss $C_T 4^{-\kappa/N}$ times the prior loss, 102 basis points at $\kappa = 2N$ (408 at $\kappa = N$ and 26 at $\kappa = 3N$). Realized: the corrected water-filling loss in each rolling window, in sample. The ratios are the same at $\kappa = N$ and $3N$ to three decimals. The bound is distribution-free and conservative; with actual return covariances the loss is far below it because water-filling moves bits toward the few large eigendirections.

A. Market-cap deciles, December 2024

Decile	N	$\rho(1)$	RTQ minus fair VNV	RTQ minus no information	Earlier formula
1	320	0.16	429	522	-714
2	320	0.30	365	547	-1268
3	325	0.47	309	592	-224
4	324	0.51	286	594	-153
5	328	0.81	116	599	283
6	324	0.74	152	599	248
7	328	0.87	75	599	345
8	327	0.86	83	599	354
9	328	0.68	167	577	-2092
10	328	0.91	57	600	390

B. Scaling with the number of stocks

Date	N	$\rho(1)$	Fair VNV loss	RTQ loss	Earlier RTQ loss
1990-12	100	0.94	37	0.62	1171
1990-12	250	0.94	37	0.20	374
1990-12	500	0.94	36	0.10	166
1990-12	999	0.91	55	0.10	111
1990-12	1,497	0.88	71	0.13	100
1990-12	1,983	0.76	146	0.34	149
2024-12	100	0.88	73	0.68	1166
2024-12	250	0.90	59	0.19	302
2024-12	500	0.90	62	0.09	136
2024-12	999	0.54	278	0.90	252
2024-12	1,498	0.69	183	0.28	96
2024-12	1,995	0.75	151	0.16	58
2024-12	2,979	0.59	245	0.65	98

Table IA.8: **Loss concentration and the gain from spreading attention.** Panel A: stocks in each market-cap decile at December 2024, 60-month Ledoit-Wolf covariance, $\kappa = 2N$; gains in basis points of annual certainty-equivalent return. Across deciles the gain over fair VNV falls with $\rho(1)$ (slope -510 basis points per unit of $\rho(1)$, $R^2 = 0.99$): when the top eigendirection carries most of the loss, concentrating attention gives up little; when loss is spread out, it gives up a lot. Panel B: losses in basis points for the largest N stocks at two dates, $\kappa = 2N$. The last column of each panel uses the earlier formula (see Table IA.5).

IA.7.2 RTQ after common factors are removed

Table IA.9 repeats the comparison of Table 2 after the top K principal components are removed from returns, for $K = 1$, 5, and 20. The theory predicts that RTQ should do relatively better as loss

concentration $\rho(1)$ falls, and removing common factors lowers $\rho(1)$ step by step, as the second column shows. Each row is one value of K, and the row $K = 0$ repeats Table 2. For each budget, the columns give the median loss of RTQ with equal bits, the median loss of the principal-component design that is given the length of the return vector, and the share of quarters in which RTQ has the lower loss. Losses are on a scale where no information about the remaining returns costs 600 basis points, so they are comparable across columns within a row but not across rows.

The thing to look at is how the two loss columns move together as K rises. Among the 500 largest stocks at two bits per stock, RTQ's loss is 10.2 against 0.70 with no factor removed, 23.5 against 10.3 after one, 40.5 against 27.7 after five, and 52.1 against 42.8 after twenty. At three bits per stock the losses are about equal from $K = 5$ (11.0 against 11.3, and 14.8 against 14.2), and among all stocks RTQ has the lower median loss for every $K \geq 1$. At one bit per stock the principal-component design stays ahead by a wide margin. Two further results are in the underlying file and not in the table. Without the length of the return vector, the principal-component design does worse: at three bits per stock RTQ has the lower loss in 67 and 80 percent of quarters among the 500 largest stocks for $K = 5$ and 20, and in 63, 87, and 97 percent among all stocks for $K = 1$, 5, and 20. And once one factor is removed, RTQ has the lower loss than the top direction in every quarter at every budget, including one bit per stock, because no single direction carries much of what remains. The takeaway is that the advantage of the covariance-based design comes mainly from the common factors. For the stock-specific part of returns, RTQ is close to it at two bits per stock and level with it at three, without using the covariance.

		1 bit per stock			2 bits per stock			3 bits per stock		
Factors removed	Median $\rho(1)$	RTQ	Principal components	RTQ lower	RTQ	Principal components	RTQ lower	RTQ	Principal components	RTQ lower
A. 500 largest stocks										
0	0.88	83.7	1.53	0%	10.2	0.70	0%	1.41	0.34	0%
1	0.32	113.0	23.9	0%	23.5	10.3	0%	5.47	4.33	3%
5	0.10	150.8	63.5	0%	40.5	27.7	0%	11.0	11.3	51%
20	0.06	176.5	120.3	0%	52.1	42.8	0%	14.8	14.2	31%
B. All eligible stocks										
0	0.72	85.0	5.71	4%	10.9	2.34	7%	1.56	0.92	24%
1	0.24	116.4	33.8	5%	24.8	15.1	8%	6.05	6.72	54%
5	0.08	145.6	72.5	1%	37.6	30.0	10%	10.2	12.3	92%
20	0.05	175.0	127.3	0%	51.2	46.5	18%	14.5	15.8	70%

Table IA.9: **RTQ and the principal-component design after common factors are removed.** The comparison of Table 2 is repeated after the top K principal components of the 60-month covariance are removed from the covariance and from the returns, in the estimation window and in the next 24 months, using the directions estimated in the window. $K = 0$ is the comparison of Table 2 on the same quarters. Median $\rho(1)$ is the loss concentration of what remains. RTQ: random rotation with the same b bits on every rotated coordinate, which uses no information about the covariance. Principal components: each principal component of what remains is rounded with the quantizer for a normal prior, with bits allocated across components, and the design is given the length of the return vector, as RTQ is. Entries are median out-of-sample losses in basis points per year, on the scale where no information about the remaining returns costs 600. "RTQ lower" is the share of quarters in which RTQ has the lower loss. 209 quarter-ends from 1972 to 2024 with at least 100 stocks.

IA.7.3 Venture deal flow: RTQ with many variables and a short history

In US stocks the covariance is estimated from many years of returns, and a strong market factor dominates it. Strategies that use it do better than RTQ in the typical quarter (Table 2). This subsection asks what happens when there are many more variables than periods, which is the setting RTQ is built for. We use the Crunchbase bulk export of April 2026 and count venture funding rounds each month in each Crunchbase category, from January 2008 to March 2026. Venture rounds are pre-seed through series J, rounds of unknown series, corporate rounds, convertible notes, equity crowdfunding, and undisclosed rounds; grants, debt, private equity, and post-IPO rounds are excluded. A round tagged with several categories counts once in each of them. We keep the 454 categories with at least 500 rounds, and we also aggregate to Crunchbase's 49 category groups. The state each month is the vector of $\log(1+\text{rounds})$ across categories, measured either as a level or as a monthly change. An investor who tracks where venture activity is moving faces exactly this problem: hundreds of variables, and only a few years of history from which to learn how they move together.

The design follows Table 2, with two differences. Every category counts equally in the loss,

which is TurboQuant's own setting, so no strategy gains from a loss matrix built on the estimated covariance. And losses are reported as a percent of the loss with no information. At each quarter-end, strategies that use a covariance estimate it (Ledoit-Wolf) from the previous 24 or 60 months, and every strategy encodes each of the next 12 monthly vectors with 1, 2, or 3 bits per variable. Besides RTQ, which uses no covariance, Table IA.10 reports rounding each variable directly without the rotation, focusing all bits on the top principal component (fair VNV), and spreading bits across principal components, with and without the length of the vector. The file behind the table also has three further strategies that use no covariance: focusing on aggregate deal flow, and focusing on the 5 or 20 most active categories. Among the 454 categories, RTQ has the lower loss than all three in every quarter at two and three bits.

With 454 categories and monthly changes (Panels A and B, left), RTQ has the lowest loss of any strategy. At two bits per category its median loss is 11.7 percent of the no-information loss, against 13.3 percent for spreading across principal components and 81.6 percent for focusing on the top component, with a 24-month window. RTQ has the lower loss than every other strategy in every quarter at two and three bits, with either window. At one bit it is level with spreading across principal components with a 24-month window (RTQ lower in 49 percent of quarters) and ahead with a 60-month window (82 percent). The longer window does not help the covariance-based design here: its median loss at two bits rises from 13.3 to 17.2 percent, which suggests that the way categories move together changes over time.

Measured as levels (Panels A and B, right), the covariance-based design keeps an edge at low budgets. With the length of the vector and a 24-month window, it has a median loss of 28.4 percent at one bit and 10.1 percent at two bits, against 36.3 and 11.7 percent for RTQ; at three bits the two are about level (3.2 against 3.4 percent, RTQ lower in 44 percent of quarters), and with a 60-month window RTQ is ahead at three bits (70 percent). Levels contain slow movements common to all categories, such as the rise in recorded rounds up to 2021 and their fall afterward, and the covariance picks these up; monthly changes remove them.

Across the 49 category groups (Panels C and D), where the top component carries 57 to 71 percent of the variance, the covariance-based design wins clearly, as in US stocks: at two bits its median loss is 1.2 to 3.7 percent against about 11 percent for RTQ. With 49 groups and a 24-month window there are about as many covariance entries as observations, so this advantage does not come from a well-estimated covariance. It comes from a few strong directions, which can be learned even from a short history. These panels also show the focus result of the theory. With one bit per group, focusing on the top component has a lower loss than RTQ in 68 to 82 percent of quarters, because attention is scarce and one direction dominates. From two bits RTQ has the lower loss than focusing in most quarters.

RTQ's loss is almost the same in all eight settings: 35 to 36 percent at one bit, 11 to 12 percent

at two, and 3.1 to 3.5 percent at three, whatever the data and however hard the covariance is to estimate. This is the property the theory promises. The takeaway is that the estimated covariance is the better guide when a few directions dominate, as for the 49 groups and for the slow common movements in levels, because such directions can be learned even from a short history. When risk is spread across many weak directions, as for the 454 categories measured as monthly changes, a short history cannot pin them down, and RTQ does better than the covariance-based design without using the covariance at all.

	Monthly change			Log level		
Bits per variable	1	2	3	1	2	3
A. Categories ($N = 454$), 24-month window						
RTQ (no covariance)	36.2	11.7	3.42	36.3	11.7	3.43
Round each variable, no rotation	40.8 (95)	14.8 (100)	4.72 (100)	35.3 (40)	13.9 (65)	4.49 (65)
Focus: top component (fair VNV)	81.6 (100)	81.6 (100)	81.6 (100)	67.3 (100)	67.3 (100)	67.3 (100)
Spread across principal components	36.0 (49)	13.3 (100)	4.62 (100)	30.8 (3)	12.6 (74)	4.89 (97)
with length	37.0 (61)	14.7 (100)	5.62 (100)	28.4 (2)	10.1 (23)	3.21 (44)
B. Categories ($N = 454$), 60-month window						
RTQ (no covariance)	36.3	11.8	3.46	36.3	11.8	3.45
Round each variable, no rotation	40.5 (98)	15.0 (100)	4.83 (100)	34.5 (24)	12.6 (62)	3.80 (64)
Focus: top component (fair VNV)	79.4 (100)	79.4 (100)	79.4 (100)	48.8 (70)	48.8 (100)	48.8 (100)
Spread across principal components	39.8 (82)	17.2 (100)	7.45 (100)	24.2 (4)	12.1 (56)	6.58 (100)
with length	41.4 (96)	19.1 (100)	9.37 (100)	23.1 (4)	10.4 (34)	4.82 (70)
C. Category groups ($N = 49$), 24-month window						
RTQ (no covariance)	35.6	11.5	3.26	35.5	11.1	3.15
Round each variable, no rotation	20.2 (3)	15.4 (100)	4.05 (84)	22.6 (0)	12.9 (77)	3.36 (48)
Focus: top component (fair VNV)	27.1 (30)	27.1 (98)	27.1 (100)	26.1 (18)	26.1 (90)	26.1 (100)
Spread across principal components	9.97 (0)	3.73 (2)	1.15 (2)	8.48 (0)	3.37 (0)	1.12 (8)
with length	12.1 (0)	5.61 (2)	2.05 (11)	9.76 (0)	3.79 (0)	1.14 (0)
D. Category groups ($N = 49$), 60-month window						
RTQ (no covariance)	35.7	11.1	3.16	35.6	11.2	3.23
Round each variable, no rotation	18.6 (2)	15.4 (100)	3.78 (90)	27.1 (2)	12.0 (76)	3.34 (60)
Focus: top component (fair VNV)	23.8 (18)	23.8 (100)	23.8 (100)	24.7 (32)	24.7 (70)	24.7 (100)
Spread across principal components	8.49 (0)	3.04 (0)	0.89 (4)	3.49 (0)	1.21 (0)	0.39 (0)
with length	9.20 (0)	4.26 (2)	1.73 (18)	3.69 (0)	1.56 (0)	0.49 (0)

Table IA.10: **Venture deal flow: RTQ with many variables and a short history.** The state each month is the vector of venture funding rounds across Crunchbase categories (Panels A and B) or category groups (Panels C and D), measured as the monthly change in $\log(1 + \text{rounds})$ or as its level. At each quarter-end, strategies that use a covariance estimate it (Ledoit-Wolf) from the previous 24 or 60 months; every strategy then encodes each of the next 12 monthly vectors with 1, 2, or 3 bits per variable, and the realized squared error is measured, with every variable counting equally. Entries are median losses as a percent of the loss with no information. In parentheses: the percent of quarter-ends in which RTQ has the lower loss. RTQ: random rotation with the same bits on every rotated coordinate; it uses no covariance. Round each variable: each variable rounded directly with the quantizer for a normal prior, given the length of the vector, as RTQ is. Focus: all bits on the top principal component of the estimated covariance. Spread across principal components: each component rounded with the quantizer for a normal prior, bits allocated across components; "with length is also given the length of the vector. Venture rounds (pre-seed to series J, unknown series, corporate rounds, convertible notes, equity crowdfunding, undisclosed), January 2008 to March 2026; categories with at least 500 rounds. 49 to 62 quarter-ends, depending on window and measure.

IA.7.4 13F tests

This subsection reports manager-level evidence from LSEG 13F institutional holdings. We do not argue that fund managers use RTQ, which was introduced only recently. RTQ and VNV serve here as benchmarks for a broad strategy that covers many directions and a narrow one that concentrates on a few. To place each manager between the two benchmarks, we translate each manager's share positions into active weights (in excess of market-cap weights), project them onto the top 50 eigendirections of the eligible-universe covariance, which capture essentially all loss-weighted concentration ($\rho(50) \approx 0.99$), and compute the cosine distances d_{RTQ} and d_{VNV} to the RTQ and VNV attention profiles built from the prior 60 months. The difference $d_{\text{RTQ}} - d_{\text{VNV}}$ is higher for more concentrated, VNV-like portfolios. The panel covers every quarter-end from 1980Q1 to 2024Q4 and every manager with at least \$1M in reported US equity: 451,061 manager-quarters and 13,569 managers. Our main sample keeps managers with at least 20 holdings (352,755 manager-quarters, 10,693 managers). Filers with only a few positions are mostly private equity and venture firms or corporations reporting a stake in a single company, which are not allocating attention across a portfolio; the tables at the end of this subsection repeat every test on all filers.

Holdings do not reveal how a manager acquires information. Observed portfolios reflect beliefs, preferences, mandates (investment-style rules, benchmark-tracking constraints, position-size limits), and attention together, and the fit distances cannot separate them. We therefore treat the 13F results as descriptive evidence on broad and focused portfolios, not as a measurement of attention. Our approach is similar in spirit to Kacperczyk et al. (2014, 2016), who relate mutual-fund portfolio composition to attention under auxiliary assumptions, and it carries the same limitation.

We use the fit distances only to rank portfolios from broad to concentrated, not as evidence that managers follow RTQ. Both distances are typically close to one, because actual portfolios resemble neither benchmark closely, so neither distance means much on its own; what we use is their difference, which orders portfolios from broad to concentrated. In 95.3 percent of manager-quarters the portfolio is closer to the RTQ profile than to the VNV profile, but this share is mechanical: the same managers' weights placed on randomly drawn stocks are closer to RTQ 97.9 percent of the time. Table IA.13 reports this and other benchmarks (random weights, a placebo spectrum, a VNV profile matched to each manager's breadth, and a test of the shape of the attention profile), and Table IA.12 reports the fit by investor type.

We next ask whether broad or concentrated portfolios earn more over the following quarter. For each manager-quarter we compute the three-month forward active return, $r^{\text{act}}_{i,t+1:t+3} = w'_{i,t} r_{t+1:t+3} - w'_{\text{mkt},t} r_{t+1:t+3}$, and regress it on the fit distances with quarter fixed effects,

$$r^{\text{act}}_{i,t+1:t+3} = \alpha_{q(t)} + \gamma_1 \, d_{\text{RTQ},i,t} + \gamma_2 \, d_{\text{VNV},i,t} + \varepsilon_{i,t}. \tag{IA.34}$$

Quarter fixed effects absorb market conditions common to all managers in a quarter, including any factor-model adjustment that applies the same factor returns to every manager.[29]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
d_{RTQ}	-4.54^{*}	-4.54^{*}	-4.54	-4.54	-6.92^{***}	-6.92	-38.57
	(-1.85)	(-1.80)	(-0.51)	(-0.51)	(-2.69)	(-0.91)	(-1.50)
d_{VNV}	-0.33	-0.33	-0.33	-0.33	-2.15^{***}	-2.15	$+85.77$
	(-0.61)	(-0.59)	(-0.09)	(-0.09)	(-3.30)	(-0.62)	$(+1.57)$
Separate regressions on the difference							
$d_{\text{RTQ}} - d_{\text{VNV}}$	$+0.27$	$+0.27$	$+0.27$	$+0.27$	$+1.91^{***}$	$+1.91$	-15.27
	$(+0.51)$	$(+0.48)$	$(+0.08)$	$(+0.08)$	$(+2.97)$	$(+0.56)$	(-0.99)
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager FE	No	No	No	No	Yes	Yes	No
Manager controls	No	No	No	No	No	No	Yes
Standard errors	HC1	Mgr	Qtr	Mgr, qtr	Mgr	Mgr, qtr	Mgr, qtr

Table IA.11: **Fit distance and next-quarter active return.** Managers with at least 20 holdings: 352,755 manager-quarters, 10,693 managers, 180 quarters, 1980Q1 to 2024Q4. Dependent variable: three-month forward active return of the manager's 13F portfolio, in percent. Columns (1) to (6) differ only in fixed effects and in how standard errors are clustered. Column (7) adds log number of holdings, log portfolio value, the largest position's share of active weight, and investor-type dummies. Table IA.16 repeats the table for all 13F filers. t-statistics in parentheses. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.

Table IA.11 reports the estimates. All seven columns use the same data and differ only in fixed effects, controls, and how standard errors are computed, as listed at the bottom of the table. The row to focus on is $d_{\text{RTQ}} - d_{\text{VNV}}$, from a separate regression on the difference alone. Since the difference is higher for more concentrated portfolios, a positive coefficient means that concentrated portfolios earned more over the next quarter, and a negative one that broad portfolios did. The coefficient is $+0.27$ percent per unit of the difference in columns (1) to (4), with t-statistics of 0.51 or less however the standard errors are computed. Adding manager fixed effects raises it to $+1.91$ in column (5), which is significant with standard errors clustered by manager ($t = 2.97$), but not once they are also clustered by quarter (column (6), $t = 0.56$). Controls for portfolio size, number of holdings, and investor type in column (7) turn it negative and insignificant. Clustering by quarter matters because managers in the same quarter hold overlapping stocks and face the same market, so their returns are correlated. The takeaway is that, pooled over 45 years, neither broad nor concentrated portfolios earn more on average.

[29]For the same reason, a pooled FF5+UMD adjustment leaves the estimates unchanged. Table IA.20 instead adjusts each manager's return with that manager's own factor loadings; the conclusions are the same.

The pooled estimate hides large differences across years. Figure IA.11 shows the coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ estimated separately for each year, with its 95 percent confidence interval. A point above zero (solid dot) is a year in which concentrated portfolios earned more, and a point below zero (open square) a year in which broad portfolios earned more. The vertical axis is stretched beyond ± 20 so that the largest years fit. The thing to look at is how often the points switch sides of zero. In 34 of the 45 years the coefficient is statistically significant at the 10 percent level: in 18 years concentrated portfolios earned more, and in 16 years broad portfolios earned more. The winner switches often, sometimes from one year to the next. Concentrated portfolios did well in the early 1990s, in 1999, and in 2020 and 2021; broad portfolios did well in 1996, 2002, 2004, 2010, 2013, and 2023. Treating each year as one observation, the mean coefficient is 8.4 with a t-statistic of 1.62, and without 1999, 2020, and 2021 it is 1.2 ($t = 0.48$). Winsorizing active returns at 1 and 99 percent within each quarter leaves these numbers almost unchanged (Table IA.19). The takeaway is that the average of zero in Table IA.11 is not the absence of differences but the average of large differences that change sign over time, which is consistent with the theory, under which neither strategy dominates in general.[30]

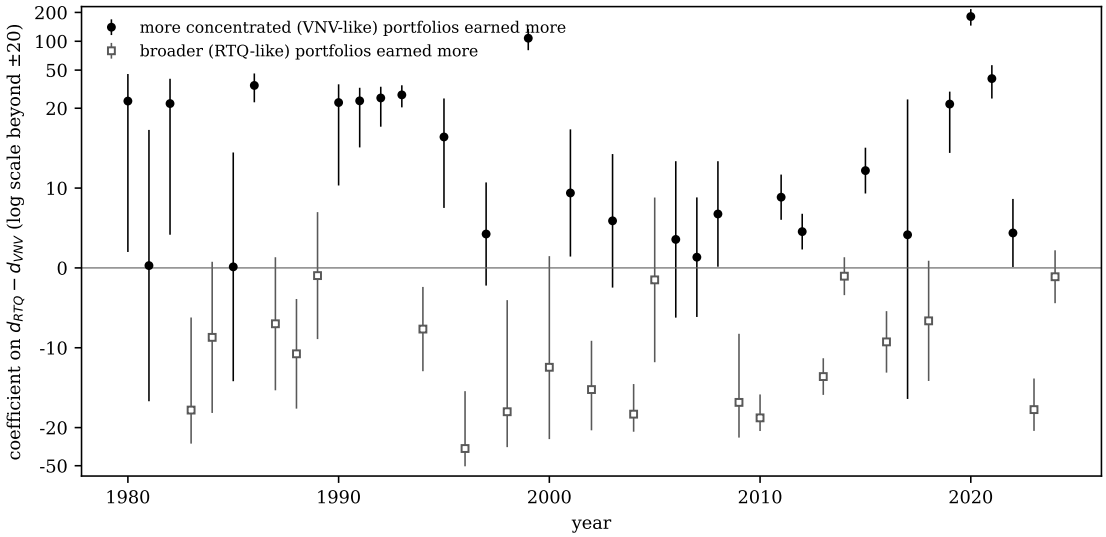

Figure IA.11: **Broad versus concentrated portfolios, year by year.** For each year 1980–2024, the coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ in a regression of three-month forward active return in percent, with quarter fixed effects and standard errors clustered by manager; 95 percent confidence intervals. Managers with at least 20 holdings. Solid dots: years in which more concentrated (VNV-like) portfolios earned more. Open squares: years in which broader (RTQ-like) portfolios earned more. The vertical axis is linear between -20 and 20 and logarithmic beyond.

Splitting the sample at regulatory events that changed the private information available to investors gives the same picture. Concentrated portfolios earned more before Reg FD (October

[30]Within-year t-statistics cluster by manager only, since each year has four quarters, and may overstate precision because managers hold the same stocks. The size of the swings, from -33 to $+181$, does not depend on the standard errors.

2000), which banned selective disclosure of material non-public information, and broad portfolios earned more after it and after Dodd-Frank (July 2010). Figure IA.11 shows, however, that the sign changes many times within each period, and with standard errors clustered by manager and quarter only the pre-Reg FD coefficient is marginally significant ($t = 1.93$; Table IA.14). The regulatory dates do not separate two stable periods.

The model points to loss concentration $\rho(1)$ as the condition that should decide the winner: concentrating should do relatively better when $\rho(1)$ is high. In the 13F data, ex-ante $\rho(1)$ does not predict which kind of portfolio wins: the interaction is statistically insignificant and its point estimate has the opposite sign from the prediction (Table IA.15). The 13F results therefore do not validate the mechanism of the model, and the alternation of winners across years has many possible explanations besides attention. The test has little power, since $\rho(1)$ varies only over time and the test rests on 180 quarters, and 13F holdings measure attention with considerable noise. The evidence for the $\rho(1)$ criterion thus comes from the CRSP welfare comparison in Section 5.

The remaining tables give supporting results. Table IA.12 and Figure IA.12 report the fit distances by investor type. Table IA.13 compares the actual fit with portfolios built from the same holdings with equal or random weights, with the managers' weights placed on randomly drawn stocks, with a VNV profile spread over as many directions as each portfolio uses, and with an RTQ profile computed from the covariance five years earlier. It also reports a shape test: the RTQ allocation gives more attention to directions with larger eigenvalues, while actual attention profiles fall with log eigenvalue on average. Table IA.14 splits the main sample at Reg FD and Dodd-Frank, and Table IA.15 tests whether ex-ante $\rho(1)$ predicts the coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$; within terciles of $\rho(1)$ the coefficient falls from $+6.9$ to -6.6, and no estimate is statistically significant.

Investor type	manager-quarters	unique managers	$\overline{d_{\text{RTQ}}}$	$\overline{d_{\text{VNV}}}$	frac closer to RTQ
Insurance	5,069	114	0.954	1.000	0.975
Investment cos.	7,374	595	0.963	0.999	0.970
Advisors	164,536	6,580	0.966	1.000	0.956
Other	250,502	5,686	0.966	0.999	0.952
Banks	23,640	564	0.955	0.997	0.933
All	451,121	13,571	0.966	0.999	0.953

Table IA.12: **Manager fit distances by LSEG TYPECODE.** For each manager-quarter with at least \$1M in reported US equity, share positions are translated to active weights (excess of market-cap benchmark) and projected onto the top 50 eigendirections of the prior-60-month Ledoit-Wolf shrinkage covariance. d_{RTQ}: cosine distance to the RTQ water-filling portfolio; d_{VNV}: same to the VNV selective $k = 1$. Last column: within-group fraction of manager-quarters with $d_{\text{RTQ}} < d_{\text{VNV}}$. Sample: 451,121 manager-quarters, 13,571 unique managers, LSEG S34 13F merged with CRSP, 1980Q1 to 2024Q4 (180 quarter-ends, all four quarters per year, no rank cap). Kruskal-Wallis $H = 2{,}704.89$ ($p \approx 0$) rejects equality. Mann-Whitney pairwise contrasts all reject equality at vanishing p-values given the panel size.

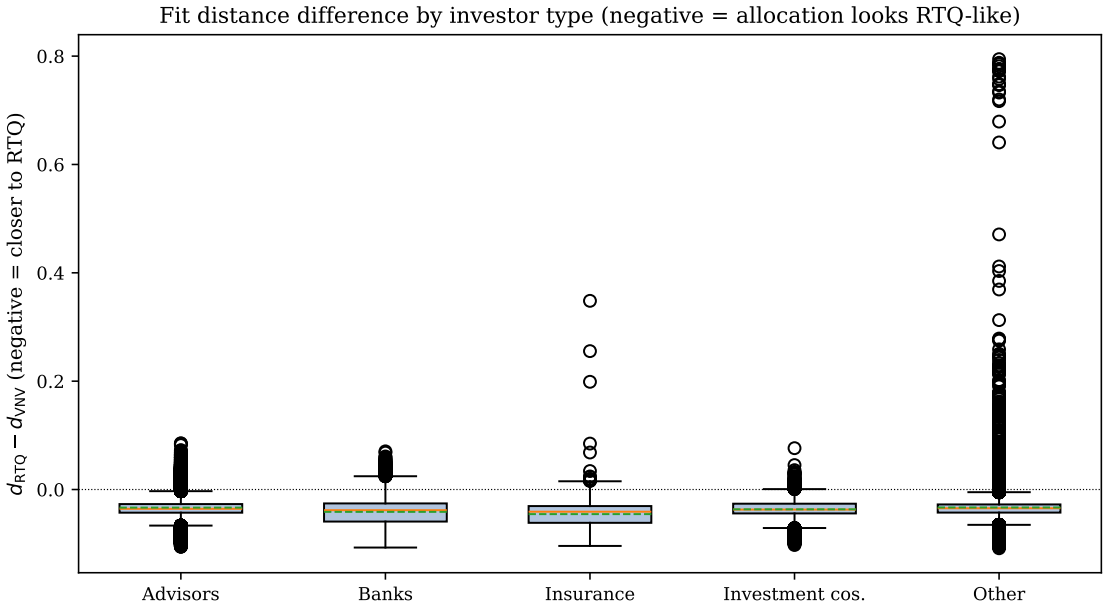

Figure IA.12: **Fit distance difference by investor type.** Box plot of $d_{\text{RTQ}} - d_{\text{VNV}}$ across 451,121 manager-quarters by LSEG TYPECODE. Sample: LSEG S34 13F merged with CRSP, 1980Q1 to 2024Q4 (180 quarter-ends, no rank cap, \$1M minimum reported US equity). Negative means closer to RTQ. Insurance ($n = 5{,}069$) is the most RTQ-like, followed by Investment companies ($n = 7{,}374$), Advisors ($n = 164{,}536$), Other ($n = 250{,}502$), and Banks ($n = 23{,}640$). Dotted line at zero marks equidistance. Orange: medians. Green dashed: means. Whiskers: 1.5-IQR Tukey limits.

Regime	n	(a)	(b)	(c)	(d)
Pre-Reg FD (1980–2000)	73,564	$+11.15^{***}$	$+11.15^{*}$	$+14.34$	$+6.75$
		$(+13.31)$	$(+1.93)$	$(+1.59)$	$(+1.31)$
Reg FD to Dodd-Frank (2000–2010)	75,849	-5.11^{***}	-5.11	-7.63	-4.56
		(-3.85)	(-0.80)	(-1.16)	(-0.77)
Post-Dodd-Frank (2010–2024)	203,342	-2.72^{***}	-2.72	-1.01	$+1.32$
		(-5.78)	(-0.61)	(-0.22)	$(+0.26)$
Quarter FE		No	No	Yes	Yes
Manager FE		No	No	No	Yes
Standard errors		HC1	Mgr, qtr	Mgr, qtr	Mgr, qtr

Table IA.14: **Regulatory regimes.** Managers with at least 20 holdings: 352,755 manager-quarters, 10,693 managers, 180 quarters, 1980Q1 to 2024Q4. Coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ (positive when more concentrated, VNV-like portfolios earned more) in regressions of three-month forward active return in percent, estimated separately in each regime. The regimes contain 83, 39, and 58 quarters. Column (a) is the specification of earlier drafts. Table IA.17 repeats the table for all 13F filers. t-statistics in parentheses. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.

A. Share of portfolios closer to RTQ than to VNV (%)	
Actual 13F portfolios	95.4
Equal weights on the manager's own holdings	95.4
Random weights on the manager's own holdings	97.1
The manager's weights on randomly drawn stocks	97.9
Actual closer to RTQ than to VNV with matched k	17.3
B. Mean fit distance to RTQ	
Current spectrum	0.965
Spectrum from five years earlier (placebo)	0.963
C. Slope of attention on log eigenvalue, top 50 directions	
RTQ water-filling (prediction)	$+0.0047$
Actual portfolios, mean	-0.0018
Actual portfolios, share with positive slope (%)	28.1
Equal weights on own holdings, mean	-0.0009
Random weights on own holdings, mean	-0.0017
Own weights on random stocks, mean	-0.0008
Actual percentile within random-stock draws, mean	0.42

Table IA.13: **Benchmarks for the fit of 13F portfolios.** 117,735 manager-quarters, fourth quarter of each year 1980 to 2024, 20 random draws per manager-quarter. Panel A: the share of actual portfolios closer to the RTQ attention profile than to the VNV profile, compared with portfolios built from the same manager's holdings with equal or random weights, and with the manager's weights placed on randomly drawn stocks. Matched k: VNV spread over the top k directions, with k equal to the manager's effective number of directions. Panel B: the RTQ profile computed from the covariance five years earlier. Panel C: RTQ water-filling gives more attention to directions with larger eigenvalues, so the slope of the attention profile on log eigenvalue is positive; the table reports this slope for RTQ, for actual portfolios, and for the null portfolios.

	(1)	(2)
A. Interaction with standardized $\rho(1)$		
$d_{\text{RTQ}} - d_{\text{VNV}}$	$+0.84$	$+2.60$
	$(+0.23)$	$(+0.76)$
$(d_{\text{RTQ}} - d_{\text{VNV}}) \times \rho(1)$	-3.56	-3.22
	(-0.94)	(-0.93)
B. Coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ *within* $\rho(1)$ *terciles*		
Low $\rho(1)$ (0.24–0.51; n = 159,701)	$+6.89$	$+8.02$
	$(+1.05)$	$(+1.46)$
Mid $\rho(1)$ (0.52–0.76; n = 99,096)	-0.63	$+2.90$
	(-0.11)	$(+0.49)$
High $\rho(1)$ (0.76–0.91; n = 93,958)	-6.60	-4.66
	(-1.24)	(-0.69)
Quarter FE	Yes	Yes
Manager FE	No	Yes
Standard errors	Mgr, qtr	Mgr, qtr

Table IA.15: **Does loss concentration predict when concentrated portfolios do better?** The model predicts that concentrating attention costs less when loss is concentrated in the top eigendirection, so the coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ (positive when more VNV-like portfolios earn more) should rise with $\rho(1)$. $\rho(1)$ is measured ex ante from the trailing 60-month Ledoit-Wolf covariance of all eligible stocks at the same quarter-end. Managers with at least 20 holdings: 352,755 manager-quarters, 10,693 managers, 180 quarters, 1980Q1 to 2024Q4. Dependent variable: three-month forward active return in percent. Terciles split the 180 quarters by $\rho(1)$, 60 quarters each. t-statistics in parentheses. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.

Tables IA.16, IA.17, and IA.18 repeat Tables IA.11, IA.14, and IA.15 on all 13F filers, including those with fewer than 20 holdings. The conclusions are the same. Table IA.19 lists every yearly coefficient behind Figure IA.11, for the main sample and for all filers, with and without winsorizing. Table IA.20 replaces active return with active return net of each manager's own factor exposures. Table IA.21 relates the fit distances to portfolio style and adds style characteristics as controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
d_{RTQ}	-8.08^{**}	-8.08^{**}	-8.08	-8.08	-13.37^{***}	-13.37	-53.99
	(-2.40)	(-2.40)	(-0.73)	(-0.73)	(-3.82)	(-1.34)	(-1.59)
d_{VNV}	$+2.36^{***}$	$+2.36^{***}$	$+2.36$	$+2.36$	$+0.50$	$+0.50$	$+109.01$
	$(+3.30)$	$(+3.60)$	$(+0.54)$	$(+0.54)$	$(+0.45)$	$(+0.12)$	$(+1.53)$
Separate regressions on the difference							
$d_{\text{RTQ}} - d_{\text{VNV}}$	-2.42^{***}	-2.42^{***}	-2.42	-2.42	-0.81	-0.81	-26.22
	(-3.39)	(-3.70)	(-0.56)	(-0.56)	(-0.75)	(-0.20)	(-1.34)
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager FE	No	No	No	No	Yes	Yes	No
Manager controls	No	No	No	No	No	No	Yes
Standard errors	HC1	Mgr	Qtr	Mgr, qtr	Mgr	Mgr, qtr	Mgr, qtr

Table IA.16: **Fit distance and next-quarter active return, all 13F filers.** All 13F filers: 451,061 manager-quarters, 13,569 managers, 180 quarters, 1980Q1 to 2024Q4. Dependent variable: three-month forward active return of the manager's 13F portfolio, in percent. Columns (1) to (6) differ only in fixed effects and in how standard errors are clustered. Column (7) adds log number of holdings, log portfolio value, the largest position's share of active weight, and investor-type dummies. Column (1) reproduces the regression reported in earlier drafts. Those drafts labelled the dependent variable an FF5+UMD alpha; that adjustment subtracts the same factor-implied return from every manager in a quarter, so quarter fixed effects absorb it and the coefficients equal those for raw active return. t-statistics in parentheses. *,**,***: $p < 0.10, 0.05, 0.01$.

Regime	n	(a)	(b)	(c)	(d)
Pre-Reg FD (1980–2000)	87,082	$+10.60^{***}$	$+10.60^{*}$	$+12.85$	$+4.91$
		$(+11.51)$	$(+1.90)$	$(+1.24)$	$(+0.93)$
Reg FD to Dodd-Frank (2000–2010)	95,574	-9.61^{***}	-9.61	-10.80	-9.89^{*}
		(-5.84)	(-1.45)	(-1.54)	(-1.67)
Post-Dodd-Frank (2010–2024)	268,405	-4.58^{***}	-4.58	-1.80	$+0.30$
		(-7.04)	(-0.78)	(-0.28)	$(+0.05)$
Quarter FE		No	No	Yes	Yes
Manager FE		No	No	No	Yes
Standard errors		HC1	Mgr, qtr	Mgr, qtr	Mgr, qtr

Table IA.17: **Regulatory regimes, all 13F filers.** All 13F filers: 451,061 manager-quarters, 13,569 managers, 180 quarters, 1980Q1 to 2024Q4. Coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ (positive when more concentrated, VNV-like portfolios earned more) in regressions of three-month forward active return in percent, estimated separately in each regime. The regimes contain 83, 39, and 58 quarters. Column (a) reproduces the regime table of earlier drafts. t-statistics in parentheses. *,**,***: $p < 0.10, 0.05, 0.01$.

	(1)	(2)
A. Interaction with standardized $\rho(1)$		
$d_{\text{RTQ}} - d_{\text{VNV}}$	-1.49	$+0.26$
	(-0.29)	$(+0.06)$
$(d_{\text{RTQ}} - d_{\text{VNV}}) \times \rho(1)$	-3.27	-3.36
	(-0.69)	(-0.82)
B. Coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ *within* $\rho(1)$ *terciles*		
Low $\rho(1)$ (0.24–0.51; n = 206,061)	$+5.03$	$+3.44$
	$(+0.52)$	$(+0.44)$
Mid $\rho(1)$ (0.52–0.76; n = 124,872)	-4.72	$+0.98$
	(-0.56)	$(+0.13)$
High $\rho(1)$ (0.76–0.91; n = 120,128)	-6.74	-5.46
	(-1.57)	(-0.82)
Quarter FE	Yes	Yes
Manager FE	No	Yes
Standard errors	Mgr, qtr	Mgr, qtr

Table IA.18: **Loss concentration and the return to concentration, all 13F filers.** The model predicts that concentrating attention costs less when loss is concentrated in the top eigendirection, so the coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ (positive when more VNV-like portfolios earn more) should rise with $\rho(1)$. $\rho(1)$ is measured ex ante from the trailing 60-month Ledoit-Wolf covariance of all eligible stocks at the same quarter-end. All 13F filers: 451,061 manager-quarters, 13,569 managers, 180 quarters, 1980Q1 to 2024Q4. Dependent variable: three-month forward active return in percent. Terciles split the 180 quarters by $\rho(1)$, 60 quarters each. t-statistics in parentheses. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.

	Main sample			All		Main sample			All
Year	Coef.	t	Winsorized	filers	Year	Coef.	t	Winsorized	filers
1980	$+23.8^{**}$	2.14	$+25.4^{**}$	+23.9	2003	+5.9	1.38	+5.2	+6.8
1981	+0.3	0.03	+1.0	-5.3	2004	-18.3^{***}	-9.52	-18.5^{***}	-25.9
1982	$+22.4^{**}$	2.41	$+22.8^{**}$	+27.5	2005	−1.5	-0.28	−1.3	+2.2
1983	-17.8^{***}	-3.01	-17.7^{***}	-26.5	2006	+3.6	0.71	+3.7	-5.2
1984	-8.7^{*}	-1.80	-8.6^{*}	-10.7	2007	+1.3	0.35	+1.3	-6.1
1985	+0.1	0.02	−3.1	+10.1	2008	$+6.8^{**}$	2.00	$+6.8^{**}$	-9.3
1986	$+34.6^{***}$	5.88	$+34.0^{***}$	+40.3	2009	-16.8^{***}	-3.84	-16.8^{***}	-7.7
1987	-7.0^{*}	-1.65	-7.0^{*}	-15.6	2010	-18.8^{***}	-12.55	-18.8^{***}	-24.0
1988	-10.8^{***}	-3.07	-10.4^{***}	-17.5	2011	$+8.9^{***}$	6.13	$+9.4^{***}$	+10.5
1989	−1.0	-0.24	−0.6	+4.6	2012	$+4.5^{***}$	3.99	$+4.5^{***}$	+6.2
1990	$+22.9^{***}$	3.56	$+23.2^{***}$	+39.3	2013	-13.6^{***}	-11.63	-13.6^{***}	-19.2
1991	$+23.9^{***}$	5.32	$+24.0^{***}$	+26.7	2014	−1.0	-0.86	−1.0	-3.2
1992	$+25.6^{***}$	6.34	$+25.1^{***}$	+23.4	2015	$+12.2^{***}$	8.33	$+12.3^{***}$	+12.5
1993	$+27.6^{***}$	7.56	$+27.8^{***}$	+27.7	2016	-9.3^{***}	-4.72	-9.1^{***}	-20.3
1994	-7.7^{***}	-2.84	-8.1^{***}	-8.8	2017	+4.2	0.40	+3.0	+22.1
1995	$+16.4^{***}$	3.62	$+16.3^{***}$	+16.7	2018	-6.6^{*}	-1.73	-6.6^{*}	-10.8
1996	-33.1^{***}	-3.68	-32.7^{***}	-27.6	2019	$+22.1^{***}$	5.62	$+22.1^{***}$	+14.3
1997	+4.2	1.29	+4.6	+10.9	2020	$+181.2^{***}$	10.01	$+174.4^{***}$	+275.6
1998	-18.0^{**}	-2.53	-19.5^{***}	-15.0	2021	$+40.8^{***}$	5.15	$+38.5^{***}$	+120.2
1999	$+107.7^{***}$	7.83	$+105.9^{***}$	+119.5	2022	$+4.4^{**}$	2.01	$+4.2^{**}$	+2.7
2000	-12.4^{*}	-1.75	-12.7^{*}	-42.8	2023	-17.7^{***}	-8.93	-17.6^{***}	-20.0
2001	$+9.4^{**}$	2.31	$+8.6^{**}$	+14.2	2024	−1.1	-0.65	−0.6	-6.6
2002	-15.2^{***}	-4.89	-15.4^{***}	-31.4					

	All 45 years		Without 1999, 2020, 2021		Years with $\lvert t\rvert > 1.645$	
	Mean	t	Mean	t	Positive	Negative
Main sample, raw	+8.4	1.62	+1.2	0.48	18	16
Main sample, winsorized	+8.1	1.61	+1.1	0.45	18	16
All filers, raw	+11.1	1.45	-0.4	-0.13	17	19
All filers, winsorized	+10.3	1.50	+0.0	0.01	18	18

Table IA.19: **Year-by-year coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$.** Top panel: for each year, a regression of three-month forward active return in percent on $d_{\text{RTQ}} - d_{\text{VNV}}$ with quarter fixed effects and standard errors clustered by manager. A positive coefficient means more concentrated, VNV-like portfolios earned more that year. Main sample: managers with at least 20 holdings. Winsorized: active return winsorized at 1% and 99% within each quarter. Bottom panel: the mean of the yearly coefficients, with a t-statistic from their spread across years (one observation per year), and the number of years with a significant positive or negative coefficient. Within-year t-statistics cluster by manager only, since each year has four quarters, and may overstate precision when managers hold the same stocks. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
A. Pooled regressions						
d_{RTQ}	$+7.45^{***}$	$+7.45^{***}$	$+7.45^{*}$	$+7.45^{*}$	$+2.82$	$+2.82$
	$(+3.74)$	$(+3.42)$	$(+1.86)$	$(+1.82)$	$(+1.33)$	$(+0.76)$
d_{VNV}	-1.36^{***}	-1.36^{***}	-1.36	-1.36	-3.08^{***}	-3.08
	(-3.39)	(-2.67)	(-0.73)	(-0.72)	(-6.01)	(-1.53)
$d_{\text{RTQ}} - d_{\text{VNV}}$ (separate)	$+1.43^{***}$	$+1.43^{***}$	$+1.43$	$+1.43$	$+3.07^{***}$	$+3.07$
	$(+3.59)$	$(+2.83)$	$(+0.78)$	$(+0.77)$	$(+6.05)$	$(+1.54)$
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Manager FE	No	No	No	No	Yes	Yes
Standard errors	HC1	Mgr	Qtr	Mgr, qtr	Mgr	Mgr, qtr

B. Coefficient on $d_{\text{RTQ}} - d_{\text{VNV}}$ by regime	(a)	(b)	(c)	(d)
Pre-Reg FD (n = 72,181)	$+9.89^{***}$	$+9.89^{**}$	$+12.88^{**}$	$+4.96$
	$(+16.09)$	$(+2.46)$	$(+2.48)$	$(+1.17)$
Reg FD to Dodd-Frank (n = 74,467)	-3.82^{***}	-3.82	-4.06	-3.35
	(-4.41)	(-1.21)	(-1.28)	(-0.89)
Post-Dodd-Frank (n = 193,346)	$+0.67^{*}$	$+0.67$	-0.04	$+2.87$
	$(+1.79)$	$(+0.30)$	(-0.02)	$(+1.17)$
Quarter FE	No	No	Yes	Yes
Manager FE	No	No	No	Yes
Standard errors	HC1	Mgr, qtr	Mgr, qtr	Mgr, qtr

C. $\rho(1)$ test (quarter FE, clustered by manager and quarter)	$d_{\text{RTQ}} - d_{\text{VNV}}$	$\times \rho(1)$	Low $\rho(1)$	Mid $\rho(1)$	High $\rho(1)$
	$+1.85$	-2.36	$+5.79^{*}$	-0.35	-1.36
	$(+0.98)$	(-1.22)	$(+1.68)$	(-0.14)	(-0.38)

Table IA.20: **Results with manager-specific factor adjustment.** Dependent variable: three-month forward active return in percent minus the return implied by the manager's own FF5+UMD loadings, estimated on the manager's full history of quarterly active returns (managers with at least 12 quarters and at least 20 holdings: 339,994 manager-quarters, 7,983 managers). Panel A follows Table IA.11, Panel B Table IA.14, and Panel C Table IA.15. t-statistics in parentheses. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.

A. Share closer to RTQ by portfolio style tercile

Characteristic	Tercile	Actual (%)	Random-stock null (%)	Percentile vs. null
Concentration (HHI)	low	92.6	96.3	0.45
	mid	95.0	98.1	0.47
	high	98.5	99.3	0.50
Active share	low	92.7	96.6	0.44
	mid	96.1	98.1	0.46
	high	97.4	98.9	0.52
Size tilt	low	96.9	98.3	0.50
	mid	96.6	97.7	0.43
	high	92.6	97.7	0.49
Momentum tilt	low	93.4	98.4	0.51
	mid	96.2	97.4	0.45
	high	96.5	97.9	0.46

B. Next-quarter active return on fit distance, with style controls

	(1)	(2)
d_{RTQ}	-36.81^{*}	-37.99^{**}
	(-1.93)	(-2.15)
d_{VNV}	$+10.91$	$+9.41$
	$(+1.16)$	$(+1.05)$
$d_{\text{RTQ}} - d_{\text{VNV}}$ (separate regression)	-11.16	-9.77
	(-1.21)	(-1.11)
Concentration (HHI)		-0.17
		(-0.32)
Active share		$+0.93$
		$(+0.79)$
Size tilt		-0.21
		(-0.08)
Momentum tilt		-1.17
		(-0.30)
Investor-type dummies	No	Yes

Table IA.21: **Fit to RTQ and portfolio style.** Fourth-quarter sample, 117,710 manager-quarters, 1980 to 2024. Size and momentum tilts are the portfolio-weighted percentile ranks of market capitalization and 12-month return within the eligible universe. Panel B: dependent variable is three-month forward active return in percent; quarter fixed effects; standard errors clustered by manager and quarter. t-statistics in parentheses. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.