Expertise and Specialization in Bank Trading*

Chase P. Ross† Sharon Y. Ross‡ Kaushik Vasudevan§

March 26, 2026

Abstract

Firms must choose which activities to pursue and which to leave to competitors, but empirical examination of the determinants of organizational scope remains limited by measurement challenges. Using novel supervisory data on individual trading desks at U.S. banks, we find that, distinct from prevailing theories of diversification driving scope, banks concentrate information-processing capacity in clusters of related activities. We document that banks specialize in activity clusters where they earn higher profits and have more precise assessment models. Examining reallocations, we find that specialized labor flows to expert banks, particularly in response to changes in risk limits tied to forecasted demand. Finally, banks contract non-core activities as they approach constraints. Our results highlight the importance of labor and organizational frictions for trading.

JEL Codes: G1, G12, G14, G2, G24, L2, L22

Keywords: firm organization, firm boundaries, specialization, segmentation, trading, human capital, risk factors

*The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors of the Federal Reserve System. We thank Aditya Chaudhry, Arvind Krishnamurthy, Taisiya Sikorskaya, Amir Sufi, and Quentin Vandeweyer for helpful comments and conversations. The initial draft of this paper was written while Vasudevan was visiting the University of Chicago, whose hospitality he gratefully acknowledges.

†Board of Governors of the Federal Reserve System. Email: chase.p.ross@frb.gov

‡Board of Governors of the Federal Reserve System. Email: sharon.y.ross@frb.gov

§Daniels School of Business, Purdue University. Email: kvasude@purdue.edu

1 Introduction

Firms choose which activities to perform and which to leave to competitors. What determines this scope of activities is an important question across industries, with implications for market structure, pricing, and resource allocation. However, analyzing the determinants of scope is empirically challenging because doing so requires fine-grained data on firms' activities, and measures of their marginal and total productivity in those activities.

We study scope and its determinants in the trading activities of the largest U.S. banks using unique supervisory data. Our main dataset, which comes from regulatory form FR VV-1, covers the trading activities of banks with at least $20 billion in average gross trading assets and liabilities, which accounted for roughly 93% of total U.S.-based trading activity as of 2019. These data provide detailed daily information on banks' allocations across different activities and asset classes, and also their profits, risk limits, measures of reasonably expected near-term customer demand, and profit attribution models. These data allow us to study, at a detailed level, how banks allocate capital across various risk factors and asset classes, and how these allocations respond to customer demand and bank constraints. We link these supervisory data to resume data on the traders employed by banks. Our combined dataset further allows us to explore the link between specialized labor and firm scope.

We document four main findings. First, banks specialize in their trading activities. Second, banks earn higher returns in their areas of specialization, and have more precise profit attribution models in those areas. Third, in response to binding constraints, banks contract peripheral activities and focus on core areas of expertise. Fourth, traders tend to reallocate toward banks where their expertise is more core, particularly following shocks to expected customer demand in those areas.

Our results are a departure from the predominant diversification-based perspective on trading activities, and indicate the importance of specialized expertise and labor in the scope of trading activities banks pursue. They highlight the role of increasing returns to scale in related activities that generate knowledge spillovers to one another (or decreasing returns to scope at the bank level). When faced with binding constraints on how large they can grow and with increasing returns to scale in related areas, banks specialize in different clusters of activities. Our evidence underscores the importance of taking an organizational- and labor-based perspective on trading, and particularly highlights the role of specialized labor and its higher marginal value in specialist banks.

We begin the paper by presenting a theoretical model to organize our empirical analysis and to provide guidance on the economic mechanisms at play. In the model, a continuum of ex-ante identical banks hire labor, where a unit of labor provides a bank with more precise information about a specific cluster of factors; for example, hiring a crude oil trader improves a bank's expertise in energy futures trading. The two key features in the model are that banks face a constraint on their overall headcount, and that their labor allocations have increasing returns to scale in producing information at the cluster-level. We solve the model in two stages: first, prices are determined as

a standard noisy rational expectations equilibrium taking banks' hiring as given; anticipating the corresponding pricing equilibrium, banks make their hiring decisions.

The first prediction of the model is that in equilibrium, banks specialize in different clusters of factors, due to a combination of increasing returns to scale and constraints on headcount. The second and third predictions of the model are direct consequences of cluster-level increasing returns to scale: that specializing banks should earn higher returns in their areas of specialization, and have more accurate models in those areas. The next set of model predictions concerns reallocations of activity. In response to binding constraints, specializing banks contract their peripheral activities, which is less costly to them than contracting core activities due to scale effects. And finally, banks scale up activity more in response to shocks to expected customer demand in their areas of specialization, because their already-developed expertise makes additional scale more valuable to them.

With our theoretical model in hand, we turn to the data. Before testing the model's predictions, we present summary statistics on banks' trading. The sample spans 814 trading desks across banks, with average desk profitability of $630,000 per day and a standard deviation of $5.5 million; summed across desks in our sample, this comes to $334 million daily. The typical desk has 9 different types of risk limits, including VaR limits, position limits, sensitivity limits, and RENTD limits (limits based on reasonably expected near-term demand).

Our first main result is that banks specialize in their trading activities. By regulation, banks are required to attribute their profit and loss to a set of risk factors; illustrative examples for risk factors for an equity desk include "U.S. large cap delta," "short-term interest rate delta," and "U.S. equity index basis." Most asset classes have hundreds of risk factors, which we sort into different "clusters" of about 130 related risk factors per asset class. Analyzing these clusters, we find that they tend to be highly concentrated, with an average cluster HHI of 8,800. This concentration provides evidence of banks choosing to focus on different trading strategies. Moreover, within banks, the top 14 clusters account for 50% of a bank's absolute P&L on average, indicating specialization by banks. Directly analyzing overlap, we find that it is low in top clusters across banks, providing further evidence that banks choose to focus on different trading activities. Finally, we document that core clusters are stable within a bank over time, with AR(1) coefficients of around 0.7 at horizons of 1-50 days and 0.5 at a 750-day horizon, providing evidence for the persistence of specialization.

Our second main result is that specialists outperform non-specialists. In our preferred specification, we find that a one standard deviation increase in specialization corresponds to a 6.1% higher ratio of daily returns per unit of VaR, an economically meaningful difference. Additionally, we analyze banks' profit attribution models, which attribute the share of desk-level P&L to several components, including movements in risk factors, components not mapped to a named risk factor, and mechanical components related to carry, cash flows, and changes in reserves. We compute the *unattributable* share of P&L as the ratio of unexplained P&L to total P&L for each desk at each bank. We show that this unattributable share is strictly decreasing in the share of activity that the desk contributes to the overall bank and the tenure of the desk at the bank. These results

underscore that banks have *better* information in their areas of specialization, which translates into stronger performance.

Our third result shows that banks retrench into their core areas of specialization when constraints bind. To test this prediction, we use our unique data on risk limits at the desk level. For each limit faced by a desk, we compute the desk's proximity to that limit, and compute the maximum proximity across different limits for each desk. We then compute a bank measure of constraints as the sum of desk-level proximities, weighted by the desks' VaR. We find that as the proximity to constraints tightens, banks immediately respond by retracting from non-core areas in the near term. This result emphasizes the relatively higher marginal value of additional resources in core activities versus peripheral activities for the bank.

Our final result focuses on trader reallocations. We obtain resume data from Revelio Labs and link it to our bank-level trading data. We show that, on average, traders tend to move to banks where their expertise is closer to the banks' specialization. Moreover, we show that following positive shocks to expected customer demand in a particular area (measured from RENTD limits), specialist banks are more responsive in hiring expert traders than non-specialist banks. This result indicates that the expertise and information-processing capacity of specializing banks are embodied in human capital, and is consistent with expert human capital having more marginal value in banks that operate heavily in matching areas.

Taken together, our results push against the traditional, diversification-based perspective on trading activities, and indicate the importance of specialized expertise in banks' trading activities. Moreover, they highlight the important role of human capital for expertise and information-processing capacities in trading. Finally, we show theoretically that our results are consistent with two important frictions: increasing returns to scale in clustered activities, and constraints at the bank level in information processing capacity. This link highlights the importance of labor- and organizational frictions in financial institutions' trading.

Related literature. Our paper relates most closely to work on specialization in financial markets. Prior work empirically documents industry specialization in banks' loan-making activities (Paravisini et al., 2023; Blickle et al., 2025). In lending, specialization is naturally associated with expertise because banks' core tasks are screening borrowers ex-ante and monitoring them ex-post, which are inherently considered to be information-intensive, and subject to potentially increasing returns to knowledge. Our paper documents similar specialization in banks' trading behavior, which cuts against the traditional intuition that investors typically choose portfolios to capture the gains from diversification. Against that backdrop, our results on specialization in trading provide direct empirical evidence for the importance of informational and organizational expertise for trading, consistent with a prior body of theoretical work (Van Nieuwerburgh and Veldkamp, 2010; Eisfeldt et al., 2023).[1]

[1]In a related vein, Kacperczyk et al. (2005) document stronger performance in concentrated industries among mutual funds and Wittwer and Uthemann (2025) document specialization by Canadian dealers in stocks, bonds, derivatives

Our results also relate to prior empirical work showing that markets are segmented, with limited participation by intermediaries in arbitrage trading, options markets, and currency markets (Siriwardane et al., 2025; Bryzgalova et al., 2025; Moskowitz et al., 2026). Specialization of the sort that we document naturally gives rise to segmentation. We empirically tie such segmentation to expertise and specialization in information production. Moreover, our theoretical results highlight that specialization can arise when related activities exhibit locally increasing returns to scale, but banks face constraints on the total amount of activity they can pursue. In this regard, our paper also relates to a literature in organizational economics emphasizing that firms' scope reflects organizational frictions, including returns to specialization, coordination costs, and span-of-control constraints (Lucas, 1978; Becker and Murphy, 1992; Garicano, 2000; Stein, 1997, 2002; Garicano and Hubbard, 2007), and emphasizes that such frictions are important for understanding financial institutions' trading.

Finally, our paper relates to work emphasizing the importance of labor in financial intermediation and asset pricing. Work on intermediary asset pricing and delegated asset management emphasizes the importance of managerial incentives (He and Krishnamurthy, 2013; Cen et al., 2023), and a broader literature studies managerial skill, matching, and mobility (Berk and van Binsbergen, 2015; Berk et al., 2017; Bonelli, 2025). Distinct from the literature's focus on managers, we focus on the importance of traders. Our results emphasize the importance of specialized traders within banks, where different traders have different strategy-specific productivities, and where putting related traders together generates organizational scale economies. Our analysis more broadly suggests the potential importance of specialized labor for understanding financial market and asset pricing outcomes.

The rest of our paper proceeds as follows. Section 2 presents our model. Section 3 describes the supervisory data we use. Section 4 presents evidence for specialization and the returns to specialization. Section 5 relates bank activity to customer demand. Section 6 tests the model's predictions on the reallocation of capital and labor in response to demand shocks and binding constraints. Section 7 concludes.

2 Model

In this section, we present a stylized model. The model highlights the roles of increasing returns to scale in related activities and bank constraints in giving rise to market specialization, and derives empirical predictions that we take to the data.

corresponding to stronger trading performance. In addition to documenting a similar empirical result for bank trading, we also provide direct evidence linking such performance improvements to information and its marginal value by analyzing banks' profit attribution models, risk limits, and hiring.

2.1 Model Setup

Environment. There is a continuum of ex-ante identical banks indexed by $b \in [0,1]$. There are $C > 2$ clusters (asset classes), each consisting of L risk factors; a risk factor is a portfolio of traded assets.

Each risk factor f trades for an endogenously determined price of p_f, and has a fundamental value of $\theta_f \sim N(0, \sigma_\theta^2)$, which is independent across factors. The terminal payoff from investing in factor f is equal to $\theta_f + \eta_f$, where $\eta_f \sim N(0, \sigma_\eta^2)$, reflects unlearnable noise (for example, execution risk).

Customers exogenously demand $D_f \sim N(0, \sigma_D^2)$ of risk factor f, which is independent across factors and from fundamentals and payoffs. Bank b takes a position of $z_{b,f}$ in f, and in aggregate, banks absorb customer demand, i.e., for every f, $\int_{b=0}^{1} z_{b,f} = D_f$.

Banks' Hiring and Portfolio Choice Problems. Bank b acquires information about risk factors by choosing labor supply in each cluster c, $n_{b,c}$, subject to a bank-level constraint $\sum_{c=1}^{C} n_{b,c} \leq N$. This labor supply choice gives rise to *expertise* in cluster c,

$$x_{b,c} = \alpha n_{b,c} + \frac{\beta n_{b,c}^2}{1 + \lambda n_{b,c}}, \quad \alpha, \beta, \lambda > 0.$$

Given their hiring choice, for each factor f in cluster c, bank b produces a private signal

$$q_{b,f} = \theta_f + u_{b,f}, \quad u_{b,f} \sim N(0, x_{b,c}^{-1}),$$

where $u_{b,f}$ is independent across factors and banks, and independent from customer demand. The form of expertise indicates increasing returns to scale at the cluster level. [2]

Each bank b has identical exponential utility with risk-aversion γ. After its hiring choices and after signals are observed, denoting the vector of bank b's expectations of payoffs as $m_b = \mathbb{E}_b(\theta_1 - p_1, \ldots, \theta_F - p_F)'$, and b's posterior variance-covariance matrix of risks as Σ_b, we can write b's portfolio choice problem as:

$$\max_{z_b} \; z_b' m_b - \frac{\gamma}{2} z_b' \Sigma_b z_b, \tag{1}$$

where $\mathbf{z}_b = (z_{b,1}, \ldots, z_{b,f})$ is a vector of its positions. Anticipating this portfolio choice problem, we can write bank b's hiring problem as

$$\max_{\{n_{b,c}\}} \mathbb{E}\left[\{\max_{z_b} \; z_b' m_b - \frac{\gamma}{2} z_b' \Sigma_b z_b\}\right] \quad s.t. \quad \sum_{c=1}^{C} n_{b,c} \leq N, \tag{2}$$

[2] We interpret $x_{b,c}$ as capturing the idea that putting more related workers together improves expertise, especially at first, though the gains taper as the desk grows. This is in the spirit of Becker and Murphy (1992), where specialization raises productivity but is limited by coordination costs.

where expectations ($\mathbb{E}$) are formed over future realizations of signals and customer demand.

Equilibrium. *An equilibrium is a collection*

$$\Big(\{n_{b,c}\}_{b\in[0,1],\, c=1,\ldots,C}, \{p_f\}_f, \{z_{b,f}\}_{b\in[0,1],\, f}\Big)$$

such that:

1. *given the price system $\{p_f\}_f$, each bank b chooses $\{n_{b,c}\}_{c=1}^{C}$ to solve* (2)*;*
2. *conditional on its hiring choices, realized signals, and realized prices, each bank forms posterior beliefs by Bayes' rule and chooses its portfolio to solve* (1)*;*
3. *for every factor f, factor markets clear:*

$$\int_0^1 z_{b,f}\, db = D_f;$$

4. *the price system $\{p_f\}_f$ is consistent with the beliefs and portfolios induced by the equilibrium hiring profile and clears all factor markets.*

2.2 Model Solution

We solve the model first by solving the asset market equilibrium conditional on banks' hiring decisions. Then we solve for equilibrium in hiring choices, and discuss conditions under which banks specialize in different clusters. Given that the primary goal of the model is to organize our empirical analysis, we provide a brief description of the model solution here and relegate the mathematical details to the appendix.

Proposition 1 (Linear Equilibrium Pricing Function)**.** *Given banks' hiring choices, factors in any cluster where a positive measure of banks hire admits a unique linear equilibrium pricing function of the form*

$$p_f = A_c\theta_f - B_c D_f, \qquad A_c, B_c > 0.$$

Proof. All proofs in Appendix IA.C. □

A key aspect of the pricing equilibrium in Proposition 1, arising from banks' first order conditions, is that banks with higher expertise in a factor take larger positions in that factor, because they face less risk. Given the linear equilibrium pricing function, which treats banks' hiring choices as exogenous, we then compute conditions under which banks specialize in different clusters.

Proposition 2 (Specialization under intermediate returns to scale)**.** *There is a $\bar{N} > 0$ such that, for $N < \bar{N}$, there are finite and positive values $\underline{\beta}(N) < \overline{\beta}(N)$ such that for every $\beta \in (\underline{\beta}(N), \overline{\beta}(N))$,*

1. *there is no equilibrium in which all banks choose the same hiring strategy across asset classes;*

2. there is an equilibrium in which banks specialize by hiring differently across clusters.

The proof of Proposition 2 proceeds in two steps. First, we rule out equilibria in which all banks choose the same hiring strategy across clusters. Second, we construct an equilibrium in which each bank concentrates its hiring and banks are distributed across clusters in equilibrium. The relevant equilibrium object is a symmetric mixing distribution over ex-ante identical agents, which captures how different banks specialize in different clusters in the cross-section.

The logic of Proposition 2 is as follows. The β parameter controls the returns to scale in cluster expertise. The lower bound ensures that hiring additional workers in the same cluster raises expertise quickly enough that banks prefer to concentrate hiring; the upper bound eliminates an equilibrium where all banks crowd into the same cluster because the gains from concentration are sufficiently strong. The restriction that N be sufficiently small means that hiring capacity is scarce enough for this tradeoff to matter: banks must decide where limited expertise is most valuable.[3]

2.3 Model Predictions

With our model in hand, we next turn to its predictions. We present formal prediction statements and proofs in the appendix.

Our first set of predictions concerns banks' specialization and the average performance of the banks in the areas they specialize in.

Prediction 1 (Specialization)**.** *Banks specialize across different clusters.*

Prediction 2 (Specialist Outperformance)**.** *Banks earn higher risk-adjusted returns in the clusters they specialize in.*

Prediction 3 (Specialists and Model Precision)**.** *Banks have more precise models in the clusters they specialize in.*

In the model, banks have increasing returns to scale at the cluster level, resulting in higher risk-adjusted returns in the cluster they specialize in (Prediction 2). The specific mechanism in the model is that banks' expertise takes the form of more precise information (Prediction 3), which we directly observe a proxy for given our unique supervisory data.

Our next set of predictions has to do with how banks allocate resources and reallocate them in response to demand and binding constraints.

Prediction 4 (Expected Demand and Positions)**.** *Clusters with higher expected customer demand attract more resources.*

[3]The modeling and proof approach here relates to Van Nieuwerburgh and Veldkamp (2010), who show that specialization arises in a model of rational inattention where investors face entropy-based costs to acquire information. There, the entropy-based cost function makes investors prefer to acquire information in a concentrated area. We model information as arising from hiring, with the benefits of concentrating expertise stemming from increasing-returns-to-scale within a cluster.

Prediction 5 (Tightening Risk Limits and Peripheral Contraction)**.** *When banks' risk limits tighten, they contract positions in areas outside of their core expertise.*

Prediction 6 (Expert Labor Flows)**.** *Expert traders move to firms that specialize in their areas of expertise, particularly following positive shocks to expected customer demand.*

Prediction 4 is straightforward: we should observe more activity by banks in areas where customer demand is forecast to be higher. This prediction is particularly useful to empirically test our measures of expected demand, which we use in Prediction 6.

Predictions 5 and 6 provide further evidence of one of the core mechanisms of our model: that on the margin, capital and expert workers carry higher value to firms within their areas of specialization than outside of them. Prediction 6 also highlights the important role of expert labor as the vehicle of banks' expertise.

3 Data

We use data covering the largest banks' trading desks collected by the Federal Reserve to implement the Volcker Rule's prohibition on proprietary trading, from form FR VV-1. The data is available daily from 2014 to 2025, with the most recent data focused on large U.S. and foreign banks that have at least $20 billion of average gross trading assets and liabilities, excluding U.S. government securities. Although the data covers trading activities at only the largest banks, large banks dominate trading so the data spans roughly 93 percent of total U.S.-based trading activity as of 2019. We focus on the set of banks that consistently provide data through the full sample. We describe how we clean the data in the Internet Appendix.[4]

We use the data to form three panels that underlie the bulk of our empirical work:

1. Date × desk: Each desk is required to provide data on its profitability, value-at-risk (VaR), and trading volume, along with the desk's names and a description of its trading activities. For example, a bank may have a desk named "U.S. equity market making" with a description "The desk makes markets in U.S. equities, indices, and delta-one linked derivatives and conducting related hedging activities" (this is not an actual desk, but an illustrative example).

2. Date × desk × risk-factor: Banks are required to attribute their profit and loss to a set of risk factors; the bank provides both the name of the risk factor and a brief description. For example, the risk factors for the equities desk example above might be "U.S. large cap delta," "short-term interest rate delta," and "U.S. equity index basis" (again, these are illustrative examples).

[4]Other academic papers that have used these data include Falato et al. (2025), who show bank curtailed equity-market trading risk after the Volcker rule; Lu and Wallen (2025), who show that bank trading desks behave like toll-taking intermediaries and bear little market risk; Barbiero et al. (2024), who show that dealer risk limit shocks affect currency market prices; and Bräuning and Stein (2024) and Li et al. (2025), who study dealer risk limits and Treasury markets.

3. Date × desk × risk-limit: Banks are required to enumerate the risk-limits that constrain their risk, giving both the limit and the current utilization of that limit. Limits vary substantially, including VaR limits or limits related to underlying risks (which can be, but do not have to be, the same as the risk factors mentioned in the previous panel). In the second half of the sample, banks were required to report additional details on the limits, including the type of limit (VaR, position limit, sensitivity, inventory age, etc.) and the source of the limit (risk appetite, regulatory capital, reasonably expected near term demand (RENTD), risk reduction/mitigation, etc.). For example, the equities desk may report both a VaR limit in dollar terms of $100, a position limit in dollar terms based on its expectation of near-term demand of |$500|, and a sensitivity limit based on an equity index basis in basis points. To make risk limits more comparable across samples and limit types, we construct a risk limit *proximity* variable, which ranges from 0 (far from limit) to 100 (at limit), and values above 100 indicate a breach; we describe its construction in the Internet Appendix.

It is not possible to directly calculate a risk factor's return because no data are reported on the size of the desk's portfolio or the share of the portfolio committed to a specific risk factor. That is, if we know a risk factor earns $100 in profits, we do not know if that is on a portfolio worth $1,000 or $1 billion. For this reason, we define returns motivated in spirit by Falato et al. (2025):

$$Return_t = PnL/\sigma(PnL)$$

where $\sigma(PnL)$ is the volatility of the risk factor's PnL over the previous three months. Henceforth, any reference to the return of a risk factor or desk refers to this return proxy.

Banks do not provide uniform classifications of their desks and risk factors, so comparison across banks is non-trivial. We use a standard TF-IDF approach to classify each desk into one of ten asset classes based on the desk's name and description: credit, equities, rates, foreign exchange, securitized and structured products, collateral and financing, commodities, overlay and counterparty risk, municipal, and legacy. We describe the asset class classification procedure in the Internet Appendix.

3.1 Summary Statistics

Table 1 reports trading desk summary statistics. Panel A summarizes profitability and VaR by asset class, spanning 814 trading desks and 1.3 million desk–trading day observations. There are 176 credit trading desks over the full sample, followed by equities (151), rates (134), and securitized/structured products (128). The remaining desk types constitute smaller shares—ranging from 26 to 82 desks—but all constitute economically meaningful shares in trading activity. On any given day, there are 100 equities desks, 83 credit desks, and 74 rates desks, among others.

Desk profitability varies substantially, averaging $630,000 (equal-weighted across asset classes) with an average standard deviation of $5.5 million. The average desk is profitable, with 90th-

percentile profits of $3.2 million compared to 10th-percentile profits of -$1 million. Across all desks within an asset class, equity desks have the largest average daily profits at $146 million (100 desks $\times$ $1.46 million), followed by foreign exchange at $50 million. All asset classes generate positive average profits. On average, trading desks in our sample generate about $334 million in profits each day (summing the dot products of average desk per day and mean profitability across all asset classes), amounting to $56 million per bank per day or $84 billion on an annual basis across all banks in the sample. The large dollar values in the table show that the largest trading desks contribute a material share of banks' overall revenues.

The table also shows that more profitable desks tend to operate with higher VaR. Forex desks average $7.6 million in VaR with a large standard deviation of $15.8 million, followed by rates desks with an average $7.4 million in VaR and standard deviation of $15.5 million. The 10th percentile for most desks is 0, implying that desks do not always operate with material risk exposures, although most asset classes have 90th percentile VaR near $13 million. Across all banks, total VaR is no more than $2.6 billion (the dot product of desks per day and the mean VaR), but this overstates risk because VaR is not strictly additive as exposures across desks within a bank can offset and hedge one another.

Table 2 summarizes the risk limit panel, restricted to the post-2020 sample where banks are required to report the type and source of each limit. The typical desk carries 9 limits on a given day, though the distribution is wide: desks at the 90th percentile face 16 simultaneous constraints. The volume of limits reflects the layered structure of risk governance at large dealers, where a single trading desk may be subject to a VaR limit, one or more sensitivity limits, a position limit, and a RENTD limit, each potentially originating from a different source within the bank's risk framework.

Panel B reports the distribution of our risk limit proximity measure, which captures how close each limit's current utilization is to the allowable bound (Appendix IA.A.2). The mean proximity across all limit-date observations is 20 percent, indicating that the typical limit has substantial headroom. The distribution varies meaningfully across limit types. Other than "other" limit types, position limits have the largest average usage (39%), followed by VaR (28%). Sensitivity and scenario limits tend to be less binding. The breakdown by limit source reveals that RENTD—which reflects the bank's ex-ante forecast of expected customer demand— is by far the most common form of analysis used to set risk limits. Figure IA.4 plots weekly averages of proximity and the share of limits near capacity over the sample period.

Panel C sharpens the picture by reporting the share of observations exceeding specific proximity thresholds. Across all limit types, only 14% of limit-days have proximity above 50 and 0.2% are above 100%. The rates differ substantially across types: position limits are above 50 on 37% of days, compared to 15% for sensitivity limits and 19% for VaR limits. By source, RENTD limits are above 50 on 13% of days, compared to 22% for risk appetite limits and 5% for risk reducing limits, although the latter two are much less common. The relatively low utilization of RENTD limits is consistent with banks calibrating expected demand for wide buffers: banks set these limits based on

expected customer flow, and actual flow rarely approaches or exceeds the forecast.

Panel D shows that binding constraints are typically transient. The median near-capacity spell—defined as consecutive days with proximity above 80—lasts 2 days, with the 75th percentile at 5 days. Such short spells further indicate that limits are binding.

Panel E documents how frequently limits are adjusted, broken out by both source and type. On any given day, 0.5% of limits experience a change in size, with increases and decreases occurring at roughly similar rates. The median spell between consecutive changes is 183 days across all limit sources, and 162 days for RENTD limits, indicating that limit changes are relatively infrequent. By type, aging limits change most frequently at 1.4% of days, while VaR limits change on 0.5% of days. The variation in change frequency across sources is central to our later demand shock tests, which rely on changes in RENTD limit size as a signal that the bank's expectations about customer demand have shifted.

3.2 Sorting Factors into Clusters

Our analysis in the paper focuses on different clusters of related activities. To identify such clusters, we sort risk factors into clusters to understand which desks and banks specialize in which risks, and to what extent those risks are correlated. We cluster each desk's reported risk factors within each asset class using the correlations of the risk factors' three-month return proxies.

We adopt a two-stage taxonomy and assignment procedure to handle the disparate lifespans of trading strategies. In the first step, we identify the stable structure of the market. We restrict our sample to "stable" risk factors with at least a year of data and compute pairwise correlations only when two factors have at least a year of overlapping observations. We apply hierarchical agglomerative clustering using average linkage and a distance metric of one minus the absolute correlation motivated by Jensen et al. (2023). We define clusters by cutting the resulting dendrogram at a distance threshold equivalent to a minimum within-cluster absolute correlation of 0.10. This choice guarantees that stable risk factors that comove are grouped into the same cluster. We will shortly discuss the robustness of the 0.10 threshold choice.

In the second stage, we populate these clusters with short-lived risk factors (those with at least 10 daily observations). Because short-lived factors may not overlap temporally with a cluster's median or centroid, we employ a best overlap member approach: a short-lived factor is assigned to a cluster if it exhibits statistically significant comovement with any stable member of that cluster with which it overlaps. For each assignable factor, we attempt to match it to an existing cluster based on realized comovement: within each candidate cluster, we identify the cluster member with which the short-lived factor has the largest temporal overlap and compute the correlation over that overlap. We assess significance using a Fisher z-test against the null of zero correlation, while also requiring the correlation magnitude to meet the cluster's minimum threshold (0.10). Factors that fail to match any existing cluster are labeled as unclassified, representing idiosyncratic or transient risks that do not statistically significantly load on underlying factors. This two-stage approach has

the advantage of allowing us to include a wide set of risk factors, but relying on the longer-lived factors to create the correlation-based clustering, where those correlations are better estimated relative to the correlations estimated for the shorter-lived factors.

For exposition, we then label each cluster using TF-IDF on the risk factors' names and descriptions, which provides interpretable labels such as "sovereign/credit/european" that capture the shared characteristics within each cluster. We find that the text-based cluster labels are relatively poor descriptors because many risk factor names and descriptions provided by the firms are too broad or general. Therefore, we find that the best information about the factor stems from its realized profits. The process yields risk factor classifications that allow us to compare trading strategies across banks, which we plot in Figure 1 for the 10 largest risk factors in each asset class.

Table 3 reports factor summary statistics. Panel A shows the characteristics; most asset classes have hundreds of risk factors, and the average desk is exposed to 8 factors on a given day. Given that each desk is exposed to several risk factors, the average profits attributable to a factor are small, ranging from -$21,000 (securitized and structured products) to $34,000 (equities). The risk factors for equities and rates tend to have larger tails, with 90th-percentile profit of $400,000 in equities and $363,000 in rates.

Table 3 Panel B shows the summary statistics for the clusters to which we sort risk factors. The average asset class has about 130 unique clusters. The number of clusters aligns with our priors about the relative complexity of asset classes: municipal bond trading has the fewest clusters, about 30, whereas credit and equities both have the most, given the wide range of tradable products in those asset classes. Again, the stability in the number of clusters across the two subsamples shows that the estimation is consistent and stable. The last two columns show that even though there are many clusters, most have more than a single risk factor. On average, 6 percent of clusters are singletons.

The large number of distinct clusters suggests that many of the factors are esoteric. Figure 2 confirms this by calculating the average correlation within an asset class and the average correlation of risk factors across asset classes: virtually none of the buckets have a meaningful average correlation coefficient, even when looking at the average correlation of risk factors in the same asset class. This likely reflects the wide range of risk factors, many of which are both esoteric and small. The figure also shows the average correlations if we limit to the largest cluster (or clusters) in each asset class as measured by the sum of profit volatility across the factors in the cluster, and the correlation coefficients now become much larger, partly because we cluster using correlations. Within each asset class, a small set of clusters is widely traded, and these widely traded factors tend to be the largest in terms of profitability volatility. However, the long tail of smaller factors tends to be relatively more esoteric.

Table 4 studies how well common aggregate measures explain the trading desks' returns with Fama and French (1993) style regressions against aggregate indices. We calculate an aggregate return index from the risk factors using either equal-weighted returns or value-weighted returns,

weighted by the standard deviation of the previous month's PnL. We calculate analogous indices for each asset class. The first two columns show the mean and standard deviation of these indices; recall, because we do not observe the portfolio sizes, we infer returns by comparing the daily PnL relative to the volatility of PnL over the previous three months, hence, while they are return proxies, they are unitless and not directly comparable to percentages. The aggregate value-weighted index has a daily return proxy of 0.006, meaning that the average PnL is 0.006 times its volatility over the past three months, with a standard deviation of 0.066.

The remaining columns in Table 4 show that aggregate measures do a poor job of explaining variation across the risk factors. Columns (5)–(7) show the median beta, t-stat, and R^2 across risk factors in an asset class when regressing that risk factor's return proxy on the value-weighted index. Securitized products have the largest beta to the index, 0.38, but R^2 is only 0.31% (not 31%), low relative to Fama and French (1993)'s R^2s for equity portfolios in the range of 70 to 90%; and the t-stat shows that the average risk factor has no significant relationship with the index. The result does not change if we use value-weighted indices constructed using only the factors within an asset class. Even with these more targeted asset-class-specific indices, betas are largely less than 0.2, and R^2 is still well below 1%. The last three columns show that the standard Fama–French market index does no better at explaining risk factors' variation, with all betas near zero—even for risk factors at equity desks.

Sorting Robustness Two choices are particularly important in our risk factor sorting. First, we focus on the absolute value of correlations because we are interested in whether a desk trades a specific risk factor, not whether they are long or short the factor. Trading desks can typically switch between long and short expressions of a risk factor depending on their preference to hedge, make markets, manage their inventory, or satisfy changing client demand. A signed-correlation sorting procedure would erroneously separate desks active in the same risk but trading on opposite sides.

Second, we choose a correlation threshold of $\rho = 0.10$. At face value, this seems like a relatively low value. We choose the cutoff by balancing the statistical power of identifying true non-zero relationships against economic intuition about the number of clusters that are reasonable within an asset class. Figure 3 and Table 5 show the tradeoff: equities desks have about 1,500 risk factors, sorting into 273 clusters with the 0.10 threshold. But if we increase the threshold to 0.15, there are 420 clusters; to 0.25 there are 707, and if we impose strong comovement of $\rho = 0.75$ there are 988 clusters—a large share of which are singleton clusters containing a single risk factor. We choose 0.10 because it balances the number of clusters with correlations that are statistically different from zero, since a correlation of 0.10 is roughly equal to 1.6 standard deviations: under the null of zero correlation and $n = 252$, then $\mathbb{E}[\rho] = \sqrt{2/\pi}\sqrt{n-1} \approx 0.05$, so $|\rho| = 0.10$ is roughly twice the noise floor we would expect from random variation alone.[5]

[5]Figures IA.2 and IA.3 plot additional summary statistics of the correlations of risk factors across banks, asset classes, and clusters when clustering based on either the absolute value or signed value of the 90-day return proxy.

The immediate concern is that correlations of 0.10, even if yielding a reasonable number of clusters, are insufficiently large to correctly cluster similar risks and might errantly cluster together risk factors that are spuriously correlated. Figure 4 plots the distribution of correlations across pairs of risk factors within an asset class, comparing risk factors within and between clusters. The mass of the distribution for the between clusters spikes just above 0, showing that the correlation of risk factors in the same asset class but different clusters is not different from zero (since we take absolute values, uncorrelated pairs will necessarily exhibit a small positive average correlation). The mass of within cluster correlations is shifted well to the right, as expected, but importantly shows that even though 0.10 is the threshold there is a nontrivial share of correlations that are well above it. The mean correlation within a cluster is 0.17, with a 75th percentile of 0.30. By contrast, the average correlation of risk factors in the same asset class but different clusters averages 0.04, with a 75th percentile of 0.09.

We more formally test whether the correlation-based clustering correctly captures persistent comovement relative to noise around the threshold in Table 6. Since the clusters were formed based on correlation, it is partly mechanical that the average correlation within cluster is higher than across clusters. Instead, we show that the separation in correlations within versus across clusters is quantitatively strong and replicates across different subsamples. We summarize this separation using the area-under-the-curve (AUC) defined as $Pr(|\rho|_{within} > |\rho|_{between})$. An AUC of 0.5 implies no separation, while higher values show that pairs with higher correlations are consistently concentrated within a cluster. In the full sample, the average AUC across asset classes is 0.88, confirming that any given within-cluster pair has higher comovement than a randomly chosen between-cluster pair 88 percent of the time.

To ensure this pattern reflects meaningful separation rather than noise around the threshold, we compute the AUC using only odd or only even trading days by recomputing the (absolute) correlation from the filtered subset (e.g., ρ calculated only using odd days). If clustering were primarily driven by sampling noise around the threshold, then we would expect the within-cluster advantage to attenuate when randomly splitting the sample. The data reject this: the AUC for both odd and even days remains well above 0.5 and similar to the full-sample values. Moreover, the table shows the hit rate—defined as the percent of pairs with correlations exceeding the 0.10 threshold—to check the stability of the clustering around the threshold. If many within-cluster pairs are only slightly above 0.10 due to sampling error, then we would expect them to frequently fail to reach the threshold in the subsamples. Within-cluster pairs clear the threshold in both subsamples 42 percent of the time versus only 3 percent for between-cluster pairs, implying that the within-cluster pairs are about 14 times more likely to exhibit stable, above-threshold relationships across the subsamples compared to between cluster pairs. Together, these results show that the clustering captures persistent return comovement rather than spurious correlations near the threshold.

4 Specialization and the Returns to Specialization

With our clusters of risk factors in hand, we next turn to evaluating our models' empirical predictions. Section 4.1 documents specialization at the cluster level. Section 4.2 then asks whether banks earn higher risk-adjusted returns and have better risk-model performance in asset classes where they are more specialized. Together, these results suggest that the clustering of intermediation reflects real informational advantages rather than institutional inertia.

4.1 Prediction 1: Banks Specialize

Prediction 1 states that trading desks specialize in different risk factors. This specialization is widely understood by those in the industry: among the largest U.S. investment banks, some have developed a reputation as specialists in certain segments. For example, some banks' indices serve as blue-chip indices for certain markets. Goldman Sachs is often viewed as a commodities specialist given its acquisition of the large commodities trading brokerage J. Aron, and that historical expertise is evident in the leading commodities index being the GSCI, originally named the Goldman Sachs Commodities Index. In emerging-market sovereign debt, the J.P. Morgan EMBI family is widely treated as the canonical hard-currency benchmark. Morgan Stanley has a reputation for equities intermediation, spanning its equity and equity derivative intermediation, equity financing, and equity capital market underwriting businesses.[6]

To test Prediction 1, we say a bank is specialized if its risk-taking is concentrated in a bank-specific core set of clusters and that allocation is stable over time. We define specialization using three allocation-based components:

1. Concentration: bank b's activity is disproportionately concentrated in a small set of clusters (its "core").

2. Persistence: core clusters are stable over time.

3. Heterogeneity: core clusters differ across banks.

First, we test whether risk factors are more correlated when they are traded by the same bank, in the same asset class, or in the same cluster. The model suggests that banks specialize in specific risks and therefore we would expect bank i's risk factors to be more correlated than risk factors across banks. We test this by calculating the correlation ρ_{ij} for all risk factors i and j that are

[6]Morgan Stanley's CEO recently said: "We have talked for many years about what it takes to run a global fully laid out equities business. The old nine boxes where you have Prime Brokerage, derivatives, and importantly cash, and then the three regions [Americas, Asia, EMEA]. And the cost to run that business, just dollars in the ground to turn on the lights every year, those costs, as you know, have only gone up. And so not surprisingly, the top of the competitive heap is able to endure operating leverage and above-market returns, and the rest have to slug it out just to sort of make the nut."

sorted into a cluster in the process described above. We then run the regression

$$\begin{aligned}|\rho_{ij}| \times 100 = \beta_0 &+ \beta_1 \mathbb{I}_{\text{SameBank}_{ij}} + \beta_2 \mathbb{I}_{\text{SameCluster}_{ij}} + \beta_3 \mathbb{I}_{\text{SameAC}_{ij}} \\ &+ \beta_4 (\mathbb{I}_{\text{SameBank}} \times \mathbb{I}_{\text{SameCluster}})_{ij} \\ &+ \beta_5 (\mathbb{I}_{\text{SameBank}} \times \mathbb{I}_{\text{SameAC}})_{ij} + \varepsilon_{ij} \end{aligned} \tag{3}$$

Column (1) of Table 7 shows that risk factors within a bank are more highly correlated; same-bank risk factors have 11 percent higher absolute correlation than cross-bank pairs (1.00/8.78). Column (6) estimates the full specification: same-bank series in the same cluster and asset class have 3.6 percentage points higher absolute correlation than non-same-bank series in the same cluster and asset class. The correlation patterns above are consistent with specialization but could also reflect bank-specific shocks that affect all risk factors within a bank, like common funding shocks or shared risk models within banks. We distinguish these explanations by examining the cross-sectional heterogeneity and persistence of bank allocations next.[7]

We would see similar patterns if banks allocated similarly across each cluster but faced common shocks or model adjustments. But we find that most clusters are dominated by a small set of banks, typically only one or two. We compute the Herfindahl-Hirschman Index (HHI) of bank allocations, where each bank's share is measured by its proportion of total volatility (standard deviation of PnL) in that cluster over the past three months. Figure 5 shows the average when sorting clusters based on the size of the cluster. The clusters at the far left—the largest clusters in terms of PnL volatility summed across banks—tend to have more banks, as indicated by the lighter shades. Table 8 shows the average cluster HHI is 8,800. This extreme concentration confirms substantial heterogeneity in which clusters each bank specializes.

The heterogeneity we observe reflects natural specialization boundaries. Table 8 also shows a strong negative correlation between the HHI and the size of a cluster. Larger, more liquid clusters attract more participants: a one-standard-deviation increase in log cluster size is associated with 580 point decline in cluster HHI. Banks specialize in distinct niches, with only the largest clusters supporting multiple specialist participants.

Thus far, we have shown that clusters are dominated by a small set of banks, but banks themselves are also dominated by a small set of clusters. Figure 6 shows the cumulative share of activity accounted for by a bank's top-ranked clusters. Across all three measures, concentration is steep: the top 14 clusters account for 50% of a bank's absolute P&L, and 68 clusters account for 90% (vs. 366 clusters total). The pattern is similar for profits and losses considered separately, with 8 and 35 clusters reaching the 50% and 90% thresholds for profits, and 8 and 34 for losses. Moreover, Figure 7 shows that this concentration is bank-specific rather than system-wide. The median pairwise Jaccard similarity of banks' top-3 cluster sets is just 0.2, indicating that banks

[7]Table IA.1 shows the analogous table, but when clustering on signed returns (rather than the absolute value of returns). The within-bank correlation in Table IA.1 is consistent with banks trading both sides of that risk rather than taking directional bets, so risk factors for the specializing bank tend to hedge one another.

largely specialize in distinct clusters.

Finally, the concentration is also stable over time. We regress each bank-cluster's share of total risk on its lagged values:

$$\text{Share}_{b,c,t} = \alpha + \beta \, \text{Share}_{b,c,t-\tau} + \varepsilon_{b,c,t} \tag{4}$$

where $\text{Share}_{b,c,t}$ is bank b's percentage of total risk in cluster c on date t. Figure 8 plots the coefficients β_τ. Cluster shares exhibit strong persistence: the coefficient is 0.7 at short horizons (1-50 days) and remains 0.5 at 750 days. Banks maintain stable relative allocations to their specialized clusters rather than churning across different risks. Because the return and attribution outcomes in the next subsection are observed at the desk level, the tests in Section 6.2 use asset-class specialization as a coarser proxy for the cluster-level specialization documented here.

4.2 Specialist Outperformance

Section 4.1 defined specialization as persistent, bank-specific concentration in a core set of clusters. In this subsection, because the relevant outcome variables are observed at the desk level, we use an asset-class-level specialization measure as a coarser proxy for that cluster-level specialization. We test whether banks perform better in asset classes where they are more specialized in two ways. First, we ask whether banks earn higher risk-adjusted returns in specialist asset classes. Second, we ask whether banks' risk models better explain P&L in those specialist asset classes.

Prediction 2: Specialists have higher risk-adjusted returns Table 9 shows that banks earn higher profits in asset classes where they are more specialized. The table shows the results of the regression of

$$r_{b,i,t} = \alpha + \beta \, \text{Specialist}_{b,i,t} + \Gamma' X + \varepsilon_{b,i,t}$$

where $r_{b,i,t}$ is the risk-adjusted return of bank b in group i on date t, where i indexes asset classes. Returns are defined as the desk's P&L divided by its lagged VaR. $\text{Specialist}_{b,i,t}$ is the standardized average share of absolute P&L in asset class i over the previous three months (excluding date t). Our preferred specification is column (3), which includes asset-class-by-date fixed effects and bank-by-date fixed effects. The group-by-date fixed effects ensure that we compare returns across banks within the same asset class on the same date, absorbing common returns to a particular group on a particular day. The bank-by-date fixed effects absorb bank-specific daily profitability, ensuring that the coefficient does not reflect specialist banks happening to have good days for unrelated reasons. Together, these fixed effects isolate a narrow but clean test: within a given bank on a given day, asset classes in which the bank specializes earn higher risk-adjusted returns than those in which it does not. In all specifications, however, greater specialization is associated with higher risk-adjusted profits.

In our preferred specifications, a one-standard-deviation increase in specialization is associated with a 6.1% higher ratio at the desk level. To put the magnitude in context, the daily coefficient of 0.061 corresponds to an annualized risk-adjusted return difference of approximately 97 percent ($0.061 \times \sqrt{252}$), suggesting specialists earn an economically meaningful premium.

Prediction 3: Specialization and Model Accuracy Here we use an asset-class-level specialization measure. We exploit the structure of the trading desk data to test whether banks have better profitability attribution in asset classes where they are more specialized. We note that attribution data are observed at the desk rather than the cluster level.

The trading desk data requires banks to attribute their profitability to several components, including movements in named risk factors, and other components that are not mapped to any named risk factor, along with other mechanical components related to carry, cash flows, and changes in reserves. We construct a daily desk-specific measure to capture the share of PnL that is not explained by the bank's risk models insofar as there are profits or losses that are not attributable to named risk factors. We construct an *unattributable* share for each desk-day as the ratio of unexplained PnL to total valuation-related PnL from existing positions:

$$\text{Unattributable Share}_{i,d,t} = \frac{|\text{Other Attributable Profit and Loss}_{i,d,t}| + |\text{Other Unattributed Profit and Loss}_{i,d,t}|}{\sum_k |\text{Risk Factor P\&L}_{i,d,k,t}| + |\text{Other Attributable Profit and Loss}_{i,d,t}| + |\text{Other Unattributed Profit and Loss}_{i,d,t}|} \tag{5}$$

where the denominator sums across the absolute value of risk factor PnL moves.[8] The ratio is bounded between 0 and 1, with lower values indicating that the desk's risk factor model more fully describes its profitability. We do not use the netted total of risk factor PnL because risk factors can offset one another.

An example makes the ratio clear: suppose a desk attributes its PnL to two risk factors: risk factor A earns +\$100, B earns −\$101, and there is unattributed PnL of \$1. Of the \$202 (in absolute value) movement in the desk's positions in any direction, the desk's risk factors captured \$201/\$202=99.5%. Hence, the unattributable share in this case is 1/202=0.5%. Empirically, the desks' risk factors describe total PnL well: the median unattributable share is 1.6%.

Figure 9 provides a first look at the relationship between specialization and model attribution.

[8] "Other Attributable Profit and Loss" (Item 5 in the FR VV-1 data) is "... any portion of the daily profit and loss produced by revaluing the positions held at the end of the previous day using the market data at the end of the current day that is not specifically attributed to changes in risk factors monitored and managed as part of the trading desk's overall risk management policies and procedures." "Risk Factor P&L" (Item 4) is "... the profit and loss from existing positions due to changes in the specific risk factors monitored and managed as part of the trading desk's overall risk management policies and procedures ..." Item 4 and 5 combined equal the daily PnL stemming from revaluing positions held at the end of the previous day using market data at the end of the current day. "Other Unattributed Profit and Loss" (Item 10) is "any other profit and loss on existing positions that is not reported [in the other items]."

Specialization is defined as in Table 10: the trailing 90-day share of a bank's gross risk factor activity in the desk's asset class. The chart shows that desks in more specialized asset classes have lower unattributable shares.

The figure also includes another proxy for specialization: the desk's tenure. We use this variable as a robustness for specialization that is not directly tied to profitability, although we note the natural selection bias: banks will close unprofitable desks. However, the variable still allows us to test if long-lasting desks have lower unattributable shares. We construct the variable by counting the days since the desk first appeared in the data, although we exclude desks reported on the bank's first reporting date, as those desks existed before the data collection began. Hence, the tenure measure captures the tenure of desks that we observe entering the sample. The right panel of Figure 9 shows that longer-lasting desks have lower error, as expected. Notably, though, it appears that banks learn over time, as there's a steady decline in the unattributable share over time.

Table 10 presents formal regression estimates. The dependent variable is the unattributable share and the independent variables capture specialization and tenure (both standardized to have mean 0 and unit variance). The first three columns use increasingly demanding fixed effect specifications to test the relationship between unattributable shares and specialization. Our preferred specification is column (3), which includes asset-class-by-date and bank-by-date fixed effects. The asset-class-by-date removes the possibility that some asset classes experience hard-to-explain returns across all banks, and the bank-by-date fixed effects absorb variation in the banks' overall model quality. The two fixed effects isolate a within-bank-day comparison: on a given day, does a bank's risk model explain a greater share of PnL in its specialist asset classes compared to its peripheral ones? The coefficient on specialist indicates that a 1 standard deviation increase in specialization is associated with a 3.2pp reduction in the unattributable share.

Column (4) shows that tenure also independently describes attribution quality, a sort of learning-by-doing channel. The bank-by-date fixed effects absorb any bank-specific time trends that could stem from the bank's models generally improving over time. Hence, the coefficient is identified from cross-desk variation within the same bank on a given day: a bank's older desks have better attribution than its newer desks, holding fixed general institutional improvement in risk modelling. The remaining columns include horse races of specialization and tenure. The stable coefficients across specifications show that the specialization effect is not simply a proxy of experience. Banks that currently concentrate in a particular segment have better risk factor coverage, beyond any gains that they could expect from having been active in the segment for longer.

5 Demand and Allocation of Resources

The previous sections characterize the cross-sectional organization of intermediation and show that it reflects genuine expertise. We now ask how this structure responds to changes in the economic environment. The model predicts that higher demand for a risk factor attracts more intermediation

capacity and that, conditional on a demand shock, specialists respond more aggressively than non-specialists. This section tests both predictions. We first show that factors with greater demand are served by more banks and desks, consistent with the model's entry margin. We then examine how existing participants respond to demand shocks and find that specialists increase their positions by more, consistent with the intensive margin channel in which lower posterior variance makes each unit of spread more valuable to a better-informed bank.

5.1 Prediction 4: Demand and Resource Allocation

If specialization reflects banks' equilibrium allocation of limited trading capacity, then high-demand risks should attract more activity. We test this prediction by studying if higher demand corresponds to future entry of new desks and banks.

We estimate demand using three proxies. The first two are derived from activity-based measures of realized trading activity, either the trailing 3-month average of total gross risk factor PnL across all banks (termed "demand level") or its standard deviation ("demand volatility").

The last measure exploits the richness of the Volcker data's risk limit reporting; banks often report limits on specific risk factors. We isolate limits denominated in USD that are based on the banks' expectation of RENTD, the bank's internal forecast of customer trading activity. A single risk factor can be subject to many limits: for example, a risk factor might have a limit on its delta exposure, its vega exposure, and a stress scenario loss, each denominated in USD but measuring a distinct dimension of the bank's expected customer-facing behavior. We measure limit-derived demand capacity as the maximum absolute value of the upper and lower bounds. We then aggregate to the cluster-date level in two stages. In the first stage, we aggregate across limits within each bank-cluster-date, yielding the average limit size across all limits (capturing the average scale of expected demand along each risk dimension). In the second stage, we average the measure across banks within a cluster-date, weighting banks equally, so that a cluster's demand reflects the typical bank's assessment rather than being dominated by banks with more granular limit structures.

Table 11 shows that higher demand attracts more desks and banks. The dependent variable is the number of distinct banks (Panel A) or desks (Panel B) over the subsequent month or year, and the independent variables are the (log of) the demand measures in dollars (standardized to have a mean of zero and unit variance). We include both cluster and asset-class-by-date fixed effects. The fixed effect specification is important to correctly understand the dynamics between risk demand and banks' allocation. Without cluster fixed effects, the regression would ask if permanently larger risk clusters—perhaps major Treasury market-making desks—have both high demand and many active banks, reflecting an almost tautological relationship between the cross-sectional distribution of cluster sizes rather than dynamic responses. The cluster fixed effects absorb these permanent differences and ensure that identification comes from within-cluster variation over time: when a given cluster experiences higher-than-usual demand relative to its own history, do more banks subsequently participate? This is the variation that speaks to whether demand attracts entry, rather

than simply reflecting that large clusters are large. The asset-class-by-date fixed effects further absorb common time-varying conditions within each asset class.

The first six columns show that a one-standard-deviation increase in demand increases the number of banks in a cluster by 0.12–0.20 (0.03–0.12) at the 1 month (1 year) horizon, and coefficients across the different demand proxies occupy a similar order of magnitude. The last six columns repeat the analysis but look instead at the number of distinct desks and find a similarly positive relationship. The RENTD results are particularly informative because the demand measure is derived from the bank's own risk management process and set before trading occurs, ruling out the possibility that the relationship is driven by reverse causality from entry to measured demand.

Coefficients are consistently larger at the short horizon compared to the long horizon. A larger long-term coefficient would indicate that entry is gradual. That is not what we find: the larger short-term coefficients likely indicate that our demand proxies comingle transitory and persistent demand shocks, with banks and desks entering in response to a temporary shock but then subsequently exiting as the shock dissipates. We will examine the persistence of demand shocks shortly. The consistency of results across all three proxies, spanning both realized activity and ex-ante demand forecasts, provides robust evidence that high-demand clusters attract additional intermediation resources.

6 Reallocations within and Across Banks

The final set of predictions concerns what happens at the margins of the firm: how banks adjust their scope under stress, and where displaced expertise ends up. The model predicts that when balance sheet capacity tightens, banks shed peripheral activities before core ones, retreating toward their cluster of specialization. It further predicts that traders working in a bank's peripheral factors generate higher surplus at firms where those factors are core, creating a force for efficient sorting across firms. This section tests both predictions. We first show that banks facing binding constraints disproportionately reduce activity outside their primary clusters. We then use trader transition data to show that traders in peripheral roles are more likely to move to firms where their expertise is central, and that this pattern is stronger following positive demand shocks in the relevant factors.

6.1 Prediction 5: Binding Constraints and Core Retrenchment

If specialization reflects a deliberate allocation of limited intermediation capacity, then banks facing binding constraints should contract activity in areas outside of their core expertise. The logic is straightforward: when a bank's risk budget tightens, it must reduce exposure somewhere. If the bank has invested in building expertise in certain clusters—and if that expertise is costly to rebuild once lost—it should preserve activity in those core clusters and cut from the periphery. We now show that when risk budgets tighten, banks preserve their core expertise while shedding peripheral activity, consistent with Prediction 5.

Within each forward horizon—1 day, 3 days, and 5 days—we report specifications using VaR-weighted utilization and VaR-weighted near-capacity share as the constraint measure. In all six specifications, the interaction coefficient is positive and statistically significant. The magnitude in column (1) implies that on days when the VaR-weighted share of the bank's desks near capacity increases by one standard deviation, the gap in next-day activity growth between a cluster one standard deviation above the mean in specialization and a cluster at the mean widens by 12 percentage points. The more specialized the bank is in a given cluster, the more it preserves activity there when constrained.

The horizon comparison reveals the dynamics of the reallocation. The interaction coefficient falls from 12.2pp at the 1-day horizon to 2.7pp at the 5-day horizon, suggesting that short-term changes happen outside the bank's core clusters, but longer-horizon changes are more equally distributed. The results are similar whether we measure constraints using the continuous utilization score or the near-capacity share, confirming that the finding is not sensitive to how we define "constrained."

6.2 Prediction 6: Trader Sorting Across Firms

If specialization reflects genuine expertise that is costly to build, then the labor market should respect these investments (prediction 6). A trader working in a peripheral asset class at one bank—e.g., a commodities trader at a bank that primarily intermediates rates—holds expertise that is possibly underinvested in at the current employer but potentially valuable elsewhere. The prediction is that such traders are more likely to move, and conditional on moving, more likely to land at a bank where their asset class is core. Two channels could amplify this sorting: demand shocks that raise the value of the trader's expertise at receiving banks (a pull channel), and binding constraints at the origin bank that force it to shed peripheral activity first (a push channel).

We test this using a trader-level panel constructed from Revelio Labs data, matched to the banks in our sample. The panel tracks individual traders at monthly frequency, recording the trader's employer, position and job description, and seniority level. We use these values to map traders to the most likely asset class, which we describe in the Internet Appendix. In total, the data provides information on 25,000 individuals with trading roles.

We identify a move as a change in employer between consecutive months. We observe two types of departures: there are 3,374 hires by banks in our sample, and 1,098 hires where the employee is moving between two covered banks. About 400 of the moves overlap with periods in which we observe trading data for the origin and destination bank. We classify a trader's asset class as *peripheral* at the origin bank if the bank's three-month trailing share of gross risk factor activity in that asset class falls below the bank's within-month median across all asset classes. We define *core* at the destination symmetrically.

Table 13 presents the transition matrix for all moves between covered banks. Of the 99 traders who move from a peripheral position, 50% land at a bank where their asset class is core. By contrast, among traders moving from a core position, 85% end up at a bank where they are also core. The

asymmetry is stark: peripheral traders disproportionately sort toward banks in which their asset class specialization is core at the destination bank.

To test this more formally, we regress an indicator for whether the trader moved on the trader's asset class share at its origin bank:

$$\mathbf{I}[\text{Moved}]_{i,t} = \alpha + \beta \, \text{Origin Asset Class Share}_{i,t} + \mathbf{X}'_{i,t}\gamma + \varepsilon_{i,t}$$

where i indexes traders, t indexes months, and Origin Bank Activity Share is the bank's three-month trailing share of gross risk factor activity in the trader's asset class. Higher values indicate a more core position. $\mathbf{X}_{it}$ includes seniority controls and fixed effects.

Columns (1)—(4) of Table 14 show that periphery traders are more likely to leave their bank. A 1pp increase in the origin bank's activity share in the trader's asset class reduces the trader's move probability by 0.2–0.5pp. The estimate is stable across specifications that add month × asset class fixed effects (column 3) to absorb common shocks (e.g., regulatory changes, profitability shocks) that affect all traders in a given asset class-month. Column (2) shows that senior traders are less likely to move overall. Column (4) restricts to senior traders only and finds the same.

One consideration is that peripheral traders are more likely to leave when their original bank grows more constrained, insofar as it gets closer to its risk limits. This "push" mechanism does not appear strong in the data: column (5) includes an interaction of the origin asset class share and the bank's constraint (measured as the three-month trailing VaR-weighted share of the origin bank's desks with risk limit proximity above 80). There is no significant relationship between the two, although the uninteracted bank constraint variable is negative, perhaps because periods when a single bank's constraints are binding coincide with periods when other banks' constraints are binding and therefore they are less likely to expand trader headcount.

A "push" interpretation is that peripheral traders leave when their origin bank becomes constrained and sheds peripheral activity. A "pull" interpretation is that traders sort to banks where demand for their expertise rises. To test the push mechanism, column (5) interacts the origin asset class share with a constraint measure: the bank's three-month trailing VaR-weighted share of desks with risk limit proximity above 80%. The interaction is insignificant. The uninteracted constraint variable is negative, consistent with constrained banks experiencing higher turnover, but this does not disproportionately affect peripheral traders.

Column (6)—(7) of Table 14 tests the "pull" interpretation with the regression:

$$\Delta\text{Asset Class Share}_{i,t} = \alpha + \beta \, \mathbb{I}\mathit{PeripheralatOriginBank}_{i,t} + \mathbf{X}'_{i,t}\gamma + \varepsilon_{i,t}$$

where ΔAsset Class Share$i$ = Dest. Bank Activity Share$i$ − Origin Bank Activity Share$i$ measures the change in the bank's activity share in trader i's asset class from origin to destination. The sample is restricted to movers between covered banks. A positive β indicates that peripheral traders disproportionately move to banks where their asset class is more central. $\mathbf{X}_i$ includes asset class

and month fixed effects and seniority controls.

Column (6) shows that traders moving from peripheral positions experience a 7.9pp larger increase in asset class share at the destination relative to traders moving from core positions. The effect is large considering that the asset class share of the average periphery trader at their bank is 6%, hence moving periphery traders more than double their asset class share. The estimate is similar when controlling for seniority in column (7).

Columns (8)–(9) test whether sorting intensifies during demand shocks. We interact the peripheral indicator with a standardized measure of market-level capacity expansion: the smoothed change in aggregate RENTD limit size for risk factors in the trader's asset class relative to its trailing average. Increased RENTD limits mechanically reflect increased demand expectations, raising demand for specialized traders. The interaction coefficient is positive and significant, indicating that peripheral traders disproportionately move to better-matched banks when market-level capacity in their asset class expands.

Taken together, these results paint a picture of pull-driven reallocation. Peripheral traders leave not because they are forced out by constrained employers, but because the labor market offers them better matches elsewhere. The sorting is efficient: traders flow from banks that undervalue their asset class to banks that value it more, and this flow intensifies when demand for their expertise is high. The implication is that specialization is not merely an organizational feature of banks but is partly embodied in the human capital of individual traders, and the labor market acts as a mechanism that reinforces and sharpens the pattern of specialization over time.

7 Conclusion

We study the scope of bank intermediation, which risk factors a desk chooses to trade, and why. We show that banks organize their trading around clusters of related risks, that this clustering reflects genuine knowledge complementarities, and that complementarities shape how the financial system prices risk and reallocates resources. Specialists earn higher returns, produce more accurate forecasts, and respond more aggressively to demand. When capacity tightens, banks shed peripheral activities first. Displaced traders sort toward firms where their expertise is central. Some of these patterns are individually consistent with alternatives: clustering with risk diversification, specialist returns with selection, contraction ordering with capital allocation. But two findings are not. Denser coverage of related factors improves forecasts for existing activities, which diversification and incentive theories do not predict. And expertise is non-portable across firms, which selection on individual ability does not predict. A simple model with clustered signal precision and a balance sheet constraint generates each of these patterns without entry costs, learning lags, or search frictions.

The results suggest that the knowledge structure of intermediaries is a first-order determinant of how risk is priced and allocated. If the organization of expertise determines who intermediates what, then policies that reshape bank scope—capital requirements or activity restrictions—reshape

the knowledge structure of the market itself. Understanding those consequences requires taking seriously that intermediation capacity is not fungible. Our evidence shows it is not.

References

Omar Barbiero, Falk Bräuning, Gustavo Joaquim, and Hillary Stein. Dealer risk limits and currency returns. Technical Report 24-11, Federal Reserve Bank of Boston, 2024.

Gary S. Becker and Kevin M. Murphy. The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, 107(4):1137–1160, 1992.

Jonathan B. Berk and Jules H. van Binsbergen. Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1):1–20, 2015.

Jonathan B. Berk, Jules H. van Binsbergen, and Binying Liu. Matching capital and labor. *The Journal of Finance*, 72(6):2467–2504, 2017.

Kristian Blickle, Cecilia Parlatore, and Anthony Saunders. Specialization in banking. Technical report, 2025.

Maxime Bonelli. Up the capital ladder: The role of labor mobility in the mutual fund industry, 2025.

Falk Bräuning and Hillary Stein. The effect of primary dealer constraints on intermediation in the treasury market. Technical Report 24-7, Federal Reserve Bank of Boston, 2024.

Svetlana Bryzgalova, Anna Pavlova, and Taisiya Sikorskaya. Strategic arbitrage in segmented markets. *Journal of Financial Economics*, 166:104008, 2025.

Xiao Cen, Winston Wei Dou, Leonid Kogan, and Wei Wu. Fund flows and income risk of fund managers, 2023. NBER Working Paper 31986.

Andrea L Eisfeldt, Hanno Lustig, and Lei Zhang. Complex asset markets. *Journal of Finance*, 78 (5):2519–2562, 2023.

Antonio Falato, Diana Iercosan, and Filip Zikes. Banks as regulated traders. *Journal of Financial Economics*, 170:104080, 2025. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2025.104080.

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

Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108(5):874–904, 2000.

Luis Garicano and Thomas N. Hubbard. Managerial leverage is limited by the extent of the market: Hierarchies, specialization, and the utilization of lawyers' human capital. *Journal of Law and Economics*, 50(1):1–43, 2007.

Zhiguo He and Arvind Krishnamurthy. Intermediary asset pricing. *American Economic Review*, 103 (2):732–770, 2013.

Theis Ingerslev Jensen, Bryan Kelly, and Lasse Heje Pedersen. Is there a replication crisis in finance? *The Journal of Finance*, 78(5):2465–2518, 2023.

Marcin Kacperczyk, Clemens Sialm, and Lu Zheng. On the industry concentration of actively managed equity mutual funds. *Journal of Finance*, 60(4):1983–2011, 2005.

Dan Li, Lubomir Petrasek, and Mary Tian. Risk-averse dealers in a risk-free market: The role of trading desk risk limits. Technical Report 2025-034, Board of Governors of the Federal Reserve System, 2025.

Lina Lu and Jonathan Wallen. What do bank trading desks do? Technical report, Harvard Working Paper, 2025.

Robert E. Jr. Lucas. On the size distribution of business firms. *The Bell Journal of Economics*, 9 (2):508–523, 1978.

Tobias J Moskowitz, Chase P Ross, Sharon Y Ross, and Kaushik Vasudevan. Risk, specialization, and covered-interest parity. *Journal of Finance, Forthcoming*, 2026.

Daniel Paravisini, Veronica Rappoport, and Philipp Schnabl. Specialization in bank lending: Evidence from exporting firms. *Journal of Finance*, 78(4):2049–2085, 2023.

Emil N Siriwardane, Adi Sunderam, and Jonathan Wallen. Segmented arbitrage. *The Journal of Finance*, 80(5):2543–2590, 2025.

Jeremy C Stein. Internal capital markets and the competition for corporate resources. *Journal of Finance*, 52(1):111–133, 1997.

Jeremy C. Stein. Information production and capital allocation: Decentralized versus hierarchical firms. *The Journal of Finance*, 57(5):1891–1921, 2002.

Stijn Van Nieuwerburgh and Laura Veldkamp. Information acquisition and under-diversification. *Review of Economic Studies*, 77(2):779–805, 2010.

Milena Wittwer and Andreas Uthemann. Specialization in financial markets. 2025.

8 Figures

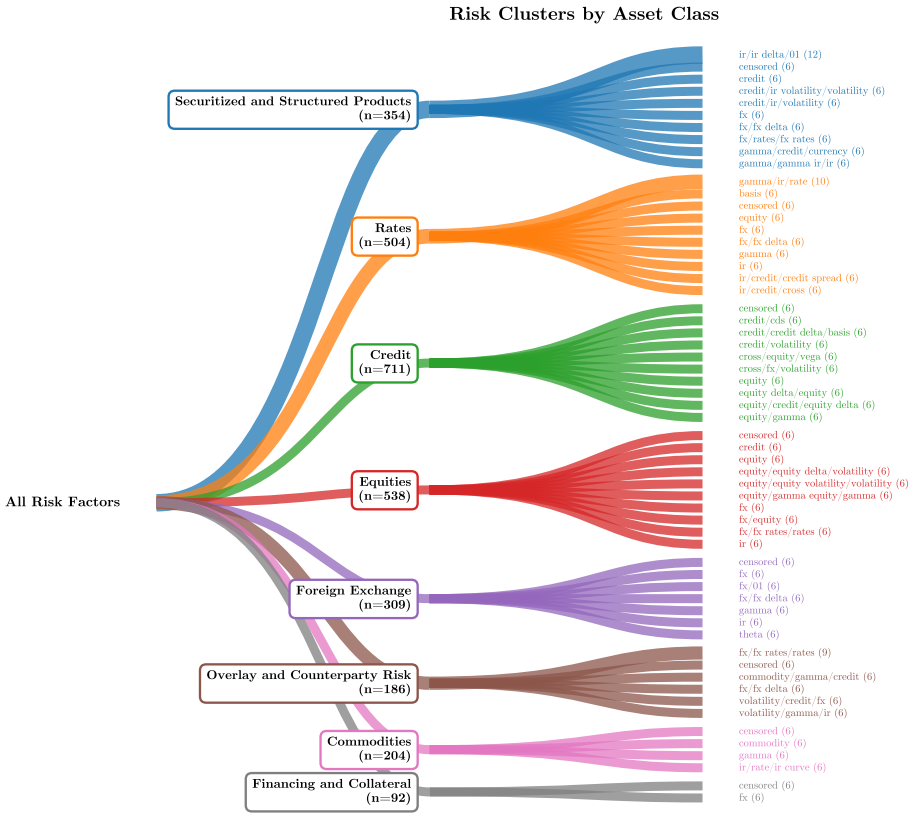


Figure 1: Desk and Risk Classification. Figure shows the classification of desks into asset classes and the 10 largest risk factor clusters within each asset class that include all banks in our sample. Figure provides text labels for risk factors only when the text label is derived from risk factors provided by the full set of banks; if a risk factor has too few banks or the text label is used by too few banks, we label it "censored" given anonymity requirements for confidential data.

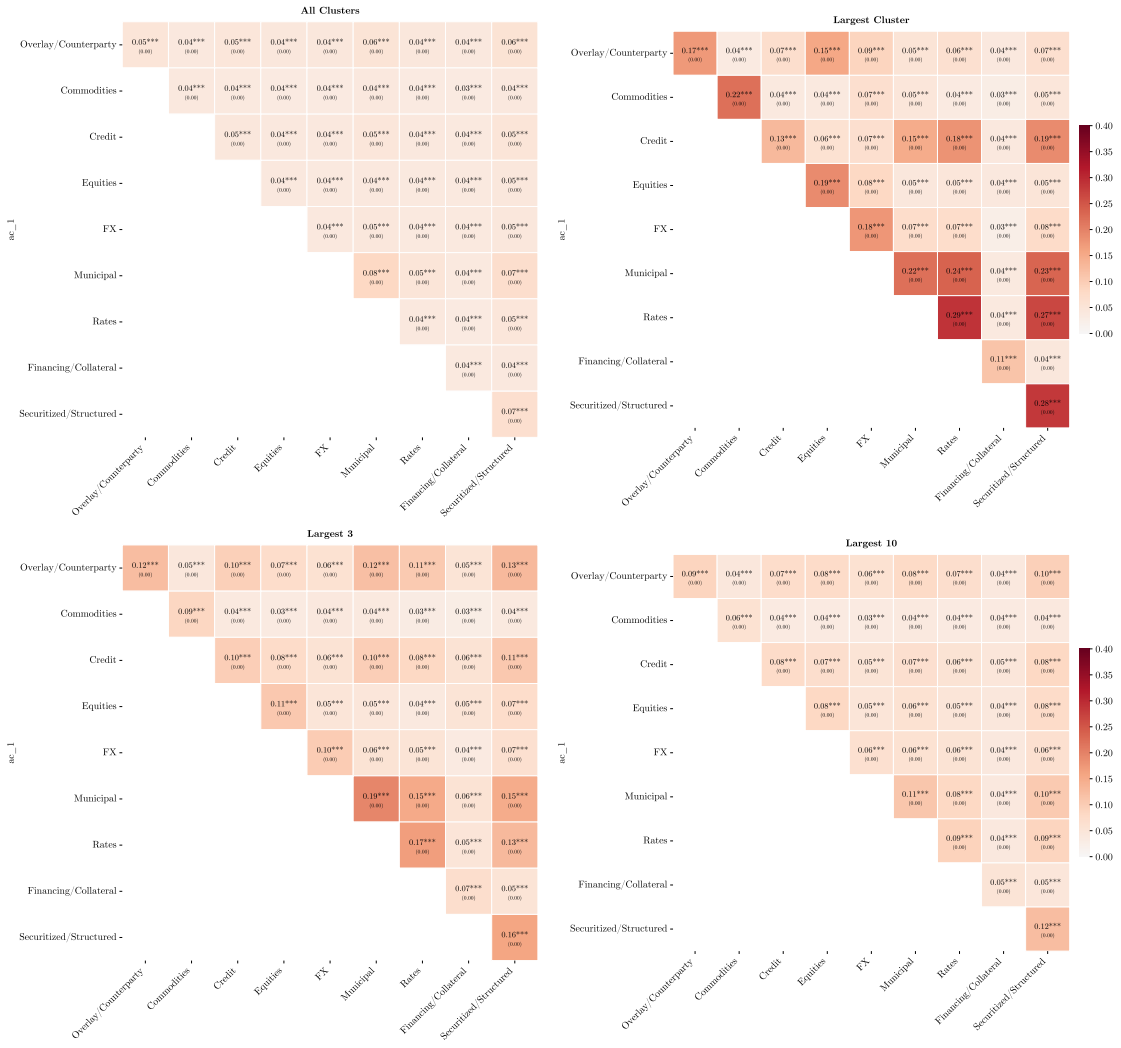


Figure 2: Correlation of Risk Factors Across Asset Classes. Each cell reports the average absolute correlation of returns (defined above) between risk factor pairs across asset classes. Correlations are aggregated using Fisher's z-transformation, weighting by both the inverse variance of each pairwise estimate and by risk factor volatility. Cluster size is defined as the sum of constituent risk factors' P&L standard deviations. p-values in parentheses; *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively. Sample restricted to post-2021 observations.



Figure 3: Clusters and Singletons across Correlation Thresholds. Figure plots the number of resulting clusters, along with number of singleton clusters that contain only a single risk factor, when categorizing risk factors based on the absolute value correlation threshold of our 90 day return proxy. Figure based on post-2020 sample clustering.

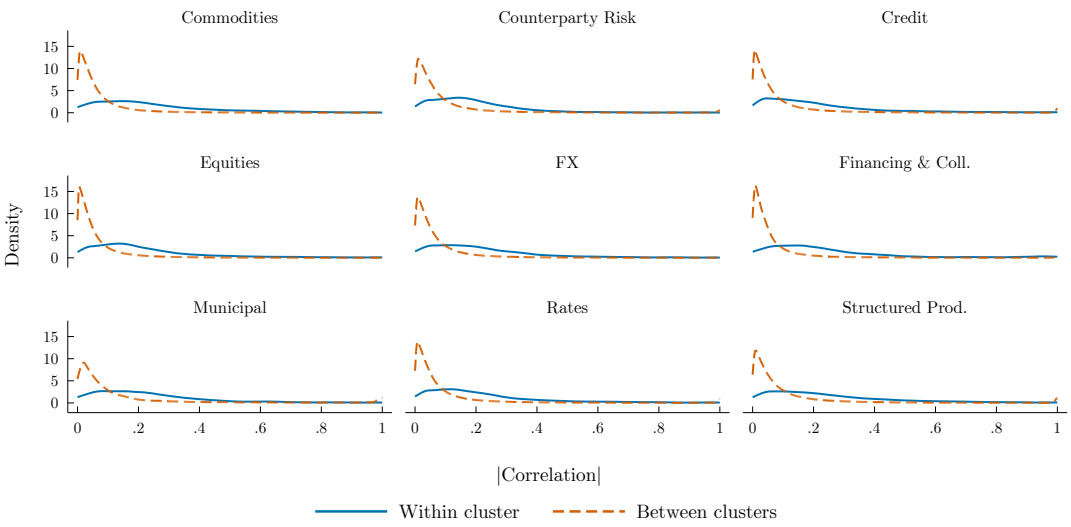


Figure 4: Distribution of $|\rho|$ across risk factor pairs within an asset class. Figure shows the distribution of $|\rho|$ for pairs of risk factors in the same asset class that are either in the same cluster or in separate clusters. We require the risk factor pair have at least one year of overlapping data.

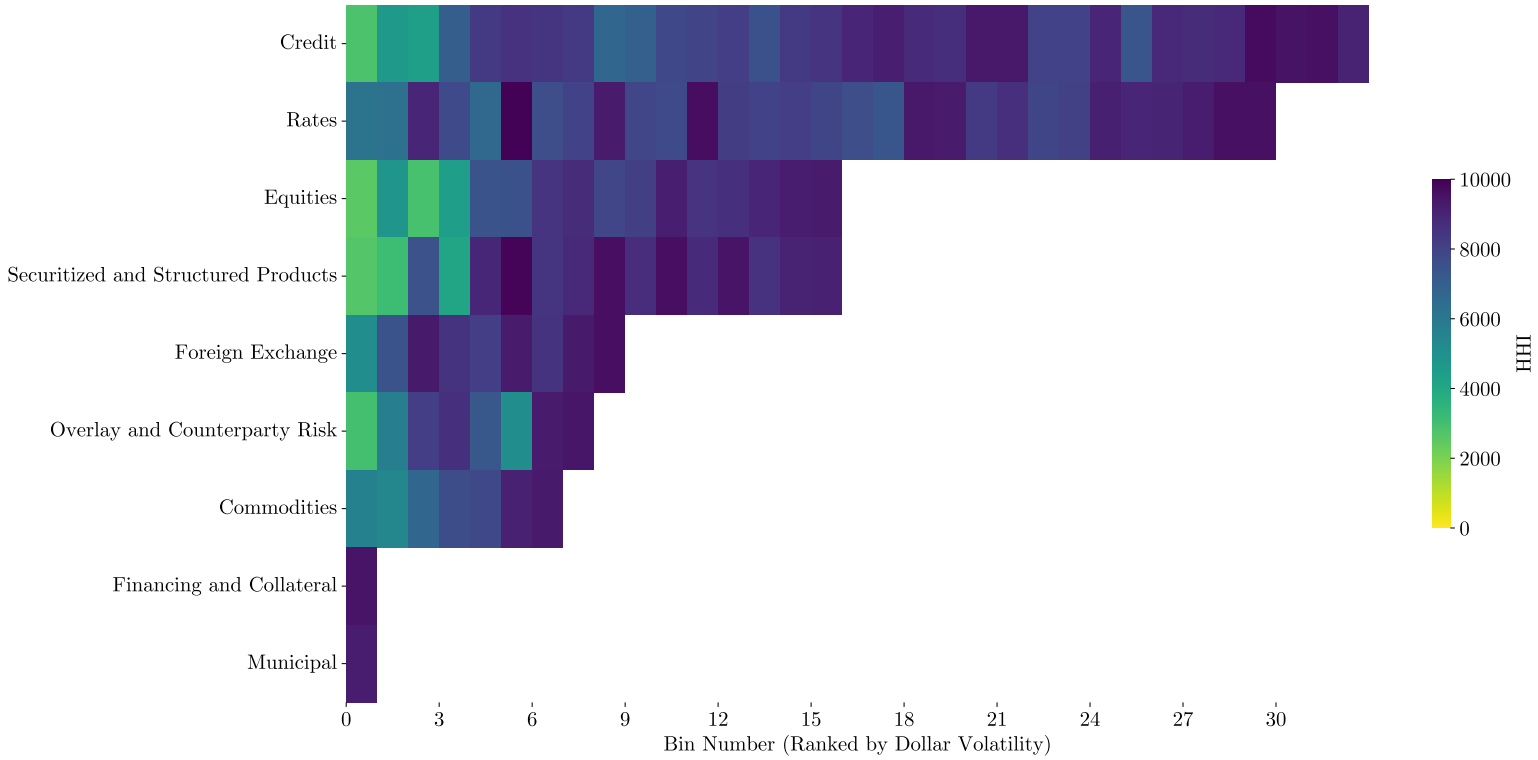


Figure 5: Bank Concentration within Clusters and Cluster Size. This figure displays the degree of bank specialization across risk factor clusters, organized by asset class. Each row represents an asset class. Within each asset class, clusters are ranked from left to right by average monthly dollar PnL volatility (highest to lowest). Adjacent clusters are grouped into bins such that each bin contains observations from all banks in our sample. The color intensity represents the Herfindahl-Hirschman Index (HHI) across banks within each bin, averaged over the constituent clusters. Darker colors indicate higher concentration (specialized risk-taking by fewer banks), while lighter colors indicate lower concentration (more diversified risk-taking across banks). Market shares are calculated as each bank's standard deviation of PnL within a cluster over the past 3 months divided by the sum of all banks' standard deviations in that cluster over the past 3 months. The plotted HHI is the time-series average of daily values. Based on risk clusters estimated in the post-2020 sample.

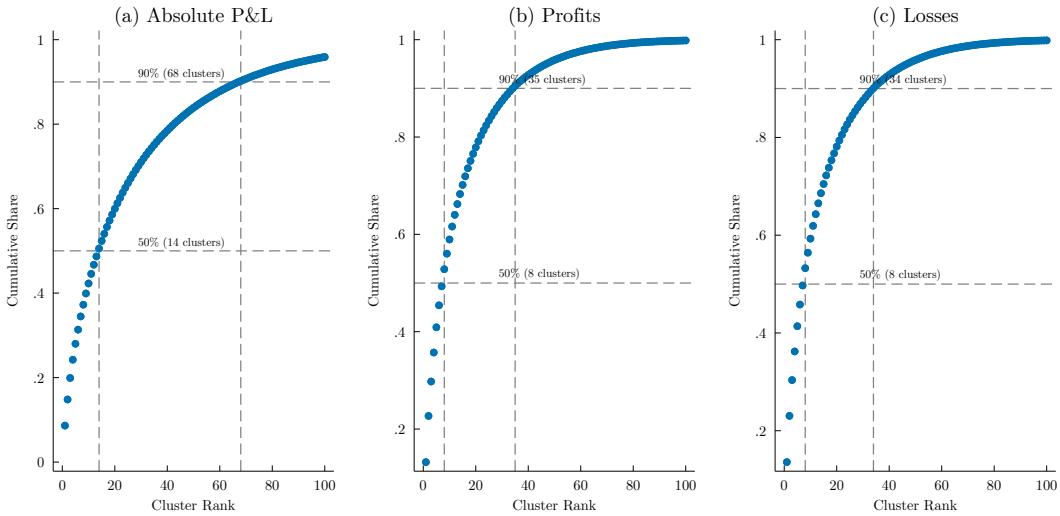


Figure 6: Concentration of Activity Across Risk Factor Clusters. Figure shows the cumulative share of bank activity accounted for by the top 100 risk factor clusters. For each bank-day, we rank clusters by their share of total activity and compute the cumulative share at each rank. We then take the median across days within each bank, followed by the median across banks. Panel (a) uses the absolute value of P&L as the activity measure, Panel (b) uses profits (positive P&L only), and Panel (c) uses losses (absolute value of negative P&L). Dashed lines indicate the number of clusters required to account for 50% and 90% of activity.

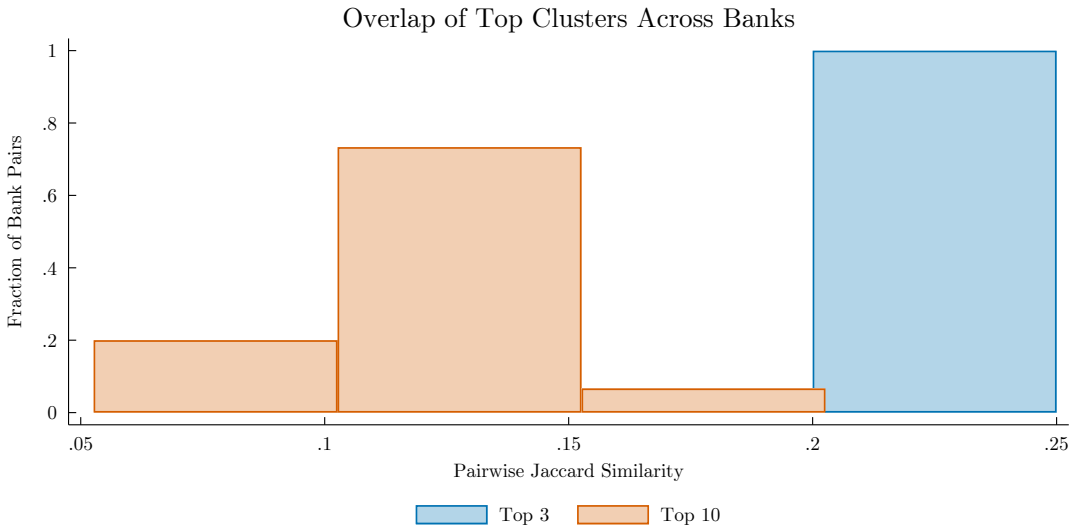


Figure 7: Overlap of Top Clusters Across Banks. Figure shows the distribution of pairwise Jaccard similarity scores across all bank pairs for the top 3 and top 10 risk factor clusters by activity share. For each bank-month, we rank clusters by their share of total absolute P&L and select the top K. For each bank pair within a month, we compute the Jaccard similarity (size of intersection divided by size of union) of their top-K sets, then take the median across months for each pair. The histogram displays the resulting cross-pair distribution. Values near zero indicate that banks concentrate their activity in largely distinct sets of clusters; a value of one would indicate identical top-K sets.

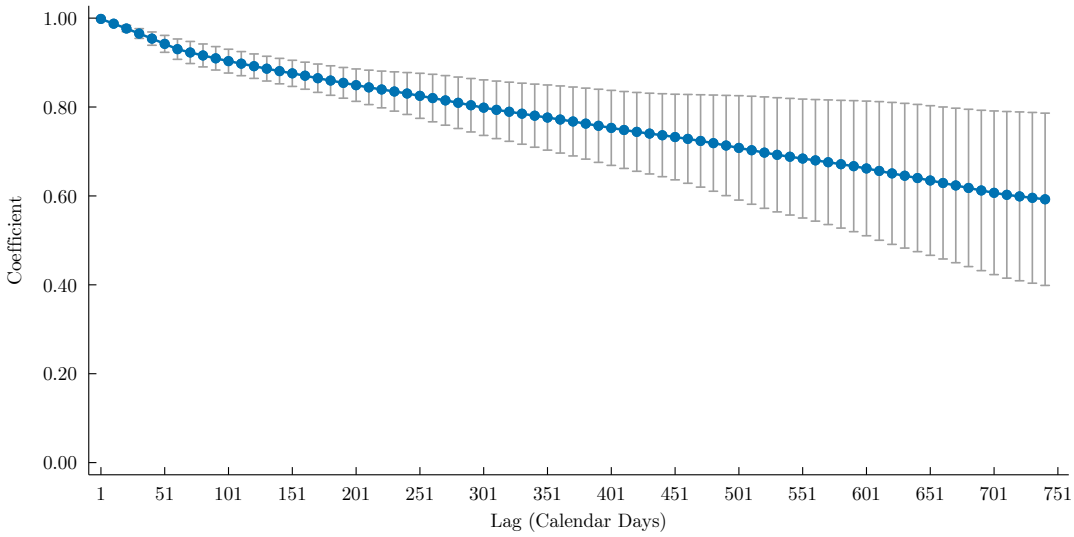


Figure 8: Cluster Share Persistence. Figure plots the coefficient from the regression $\text{Share}_{b,c,t} = \alpha + \beta\,\text{Share}_{b,c,t-\tau} + \varepsilon_{b,c,t}$ where $\text{Share}_{b,c,t}$ is bank b's percentage of total risk (standard deviation of PnL over previous three months) in cluster c on date t. Based on risk clusters estimated in the post-2020 sample.

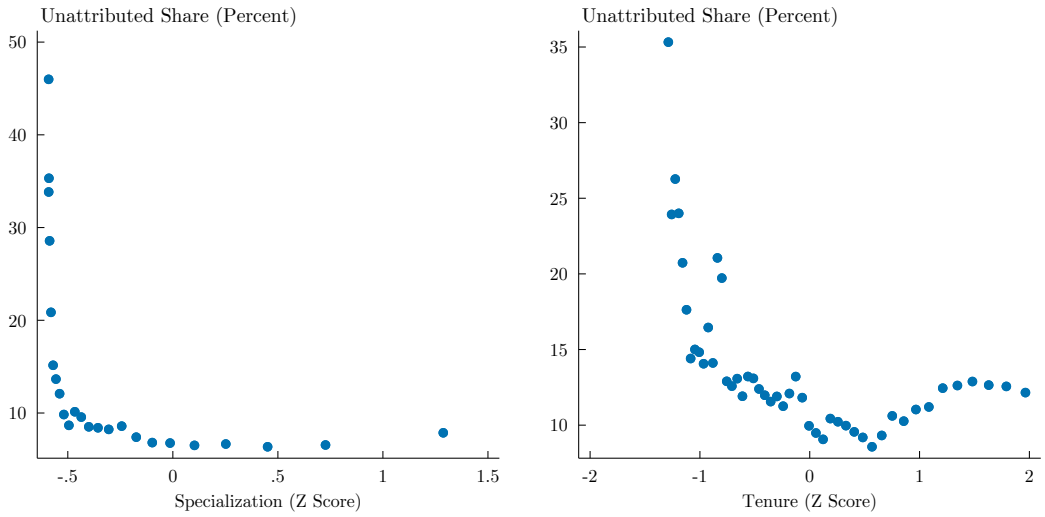


Figure 9: Specialization and Model Accuracy. Figure plots the relationship between the unattributable share of existing-position P&L and two measures of bank expertise. The unattributable share is the ratio of unexplained P&L (Items 5 and 10 under the FR VV-1 reporting requirements) to total valuation-related P&L from existing positions, with all components entering in absolute value. Panel (a) plots bin means of the unattributable share against the standardized specialization measure, defined as the trailing 90-day average share of a bank's gross risk factor activity in each asset class. Panel (b) plots bin means against the standardized cumulative number of active trading days. Bins are constructed to ensure each contains observations from all banks in our sample, and figures show the range from ± 2 standard deviations of the x-axis variable. The sample in Panel (b) excludes desks that were already active on the bank's first reporting date to avoid left-censored tenure.

9 Tables

				Profitability ($ mn)					Value-at-Risk ($ mn)				
Asset Class	Count	Desk–Day N	Avg. Desks/Day	Mean	StDev	10th	90th	Skew	Mean	StDev	10th	90th	Skew
Commodities	45	79,780	28	0.7	-1.5	3.5	4.4	-1.6	6.8	0.4	15.2	8.5	4.2
Credit	176	232,251	83	0.6	-0.5	2.3	4.8	-12.2	4.5	0.0	9.7	10.7	11.5
Equities	151	280,618	100	1.5	-0.6	4.9	4.4	-1.1	5.2	0.0	12.8	7.3	3.9
Financing and Collateral	37	62,613	22	0.7	-0.8	3.0	3.8	-7.5	3.6	0.1	8.3	4.1	2.7
Foreign Exchange	82	108,756	39	1.3	-1.0	5.2	4.4	-3.1	7.6	0.0	15.5	15.8	5.8
Municipal	26	42,224	15	0.2	-0.2	1.0	1.5	-3.3	2.9	0.0	8.8	5.1	3.1
Overlay and Counterparty Risk	35	61,998	22	0.0	-2.3	2.5	5.9	-18.9	4.7	0.0	14.1	8.4	3.9
Rates	134	208,373	74	0.3	-1.2	3.9	16.4	-17.3	7.4	0.0	16.0	15.5	7.1
Securitized and Structured Products	128	189,124	67	0.4	-0.6	2.0	4.4	-10.2	6.8	0.0	17.3	13.1	5.8

Table 1: Trading Desk Summary Statistics. Table reports summary statistics at the trading-desk level. We report the number of desks by asset class and bank, along with the distribution of desk-level P&L. Statistics are computed over the full sample period; see data section for details.

Panel A: Structure	*N*	Mean	SD	p10	p25	p50	p75	p90
Limits per desk-date	519,109	8.6	8.4	2	3	6	12	16
Panel B: Risk Limit Usage (%)	*N*	Mean	SD	p10	p25	p50	p75	p90
All Limits Types	4,451,870	19.6	25.0	0.0	8.2	32.9	58.0	58.0
Aging	423,099	24.3	33.8	0.0	0.2	48.1	80.3	80.3
Other	2,060	44.3	31.7	5.6	47.4	72.8	84.8	84.8
Position	151,924	38.8	29.4	12.0	36.8	62.5	80.3	80.3
Scenario	1,528,811	8.0	18.1	0.0	0.0	3.8	33.5	33.5
Sensitivity	1,566,027	23.6	23.4	4.3	16.5	36.9	59.1	59.1
VaR	779,949	27.8	24.4	6.6	23.7	43.1	61.9	61.9
All Limit Sources	4,451,870	19.6	25.0	0.0	8.2	32.9	58.0	58.0
RENTD	3,011,972	17.6	23.8	0.0	5.1	29.6	55.1	63.7
Risk Appetite	903	28.5	26.5	6.5	20.6	46.0	72.5	55.1
Risk Reducing	410	20.5	17.0	6.3	23.0	30.8	35.8	72.5
Other	1,438,585	23.8	26.9	0.0	15.2	38.4	63.7	35.8
Panel C: Proximity Thresholds (%)	Share >50	>80	>90	>95	>100			
All Limits Types	13.9	3.0	1.4	0.9	0.2			
Aging	24.1	10.1	6.4	4.8	0.6			
Other	48.2	15.0	5.8	1.4	0.0			
Position	37.0	10.2	4.2	2.2	0.2			
Scenario	5.1	0.9	0.4	0.2	0.1			
Sensitivity	15.1	2.6	0.9	0.4	0.2			
VaR	18.5	2.8	1.1	0.6	0.3			
All Limit Sources								
RENTD	12.5	2.3	0.8	0.5	0.2			
Risk Appetite	22.1	3.9	2.7	2.2	0.1			
Risk Reducing	5.4	0.5	0.2	0.2	0.2			
Other	16.9	4.6	2.5	1.7	0.3			
Panel D: Persistence	*N*	Mean	SD	p10	p25	p50	p75	p90
Near-cap spell (>80%, days)	19,046	7	24	1	1	2	5	14
Panel E: Limit Changes								
By Limit Source	All Limit Sources	RENTD	Risk Appetite	Risk Reducing	Other			
% days with any change	0.53	0.45	0.22	0.24	0.73			
% days with increase	0.33	0.30	0.00	0.24	0.39			
% days with decrease	0.33	0.30	0.22	0.00	0.39			
Med. $\|\Delta\|$ (%, cond.)	20	25	19.5	100	12.5			
Med. days between changes	183	162	19	61	216			
By Limit Type	All Limits Types	Aging	Other	Position	Scenario	Sensitivity	VaR	
% days with any change	0.53	1.44	0.53	0.43	0.41	0.43	0.53	
% days with increase	0.33	0.77	0.49	0.30	0.20	0.37	0.27	
% days with decrease	0.33	0.69	0.44	0.27	0.21	0.36	0.30	
Med. $\|\Delta\|$ (%, cond.)	20	5	27	20	30	22	20	
Med. days between changes	183	230	66	145	159	187	159	

Table 2: Risk Limit Summary Statistics. Each observation is a specific risk limit on a specific trading desk on a given day from the FR VV-1 reporting form. Panel A reports the distribution of limits per desk-date. Panel B reports the distribution of proximity scores overall and by limit type (VaR, sensitivity, position, scenario, aging) and by limit source (risk appetite, RENTD, risk reducing). Proximity measures how close current usage is to the allowable bound, on a scale where 0 indicates maximum slack, 100 indicates usage at the limit, and values above 100 indicate a breach (see Appendix IA.A.2). Panel C reports the share of limit-date observations exceeding various proximity thresholds, by limit type and source. Panel D reports the duration of near-capacity spells, defined as consecutive days with proximity above 80. Panel E reports the frequency and magnitude of limit changes, broken out by limit source and by limit type; the conditional median absolute change and median days between changes are computed on the subset of limit-days with a nonzero change. Limited to the Post-2020 sample.

Panel A: Characteristics	Avg. RF/Desk/Day	Factors (Count)		Profitability ($ mn)			
		Pre-2021	Post-2020	Mean	StDev	10th	90th
Commodities	9.1	580	353	0.02	1.42	-0.25	0.31
Credit	8.8	1,823	1,142	-0.01	1.02	-0.15	0.12
Equities	8.6	1,561	1,243	0.03	1.66	-0.30	0.40
Financing and Collateral	7.7	317	265	0.01	0.99	-0.10	0.11
Foreign Exchange	8.3	813	427	0.00	1.78	-0.34	0.33
Municipal	5.2	175	127	-0.01	0.86	-0.13	0.13
Overlay and Counterparty Risk	12.2	494	420	-0.01	1.79	-0.31	0.30
Rates	7.8	1,187	947	-0.01	2.14	-0.39	0.36
Securitized and Structured Products	6.7	1,060	728	-0.02	1.69	-0.31	0.26

Panel B: Clusters	Classified Factors (Count)		Clusters		Singleton Clusters (%)	
	Pre-2021	Post-2020	Pre-2021	Post-2020	Pre-2021	Post-2020
Commodities	580	353	103	93	9	12
Credit	1823	1142	230	185	0	2
Equities	1561	1243	273	260	4	6
Financing and Collateral	317	265	92	77	11	9
Foreign Exchange	813	427	117	105	2	4
Municipal	175	127	34	28	9	7
Overlay and Counterparty Risk	494	420	76	95	4	9
Rates	1187	947	175	192	3	5
Securitized and Structured Products	1060	728	133	99	5	1

Table 3: Risk-Factor Cluster Summary Statistics. Table summarizes the risk factors and their clusters; data calculated from daily factor returns. Count is the number of the set of unique risk factors-desk observations reported across all desks and asset classes before excluding factors that last for less than a year. Profitability moments are calculated from the date × desk × risk factor panel. Classified factors are the count of factors that are classified; the difference between total factors and classified factors are those that do not have data for at least a year. Clusters is the number of risk factor clusters we estimated following the methodology described in section 3.2. Singleton clusters columns give the number of clusters that contain a single factor only.

	Value-Weighted Index		Equal-Weighted Index		vs. Aggregate VW Index			vs. AC VW Index			vs. Fama-French Mkt-RF		
	Mean	Stdev	Mean	Stdev	b	t-stat	R^2 (%)	b	t-stat	R^2 (%)	b	t-stat	R^2 (%)
Commodities	0.039	0.203	0.023	0.113	0.35	0.32	0.22	0.15	0.40	0.20	0.00	-0.03	0.00
Credit	-0.043	0.139	-0.012	0.081	0.31	0.26	0.26	0.14	0.22	0.31	-0.01	-0.13	0.00
Equities	0.058	0.108	0.021	0.078	0.31	0.32	0.19	0.25	0.39	0.20	0.00	0.02	0.00
Financing and Collateral	0.043	0.285	0.014	0.164	0.20	0.19	0.15	0.06	0.27	0.19	0.00	0.04	0.00
Foreign Exchange	0.005	0.165	0.004	0.112	0.37	0.35	0.24	0.15	0.34	0.28	-0.01	-0.14	0.00
Municipal	-0.032	0.290	-0.013	0.215	0.04	0.04	0.47	0.01	-0.03	0.69	0.01	0.20	0.01
Overlay and Counterparty Risk	-0.014	0.228	-0.014	0.115	0.29	0.32	0.23	0.09	0.27	0.56	-0.01	-0.12	0.00
Rates	-0.011	0.141	-0.014	0.083	0.37	0.36	0.25	0.17	0.32	0.26	0.00	-0.09	0.00
Securitized and Structured Products	-0.034	0.144	-0.049	0.117	0.38	0.30	0.31	0.18	0.29	0.45	-0.01	-0.16	0.00
Aggregate	0.006	0.066	-0.004	0.049									

Table 4: Risk-Factor Variation Regressions. Table summarizes risk factor indices and their ability to explain variation across risk factors. The first four columns present the moments from the return indices calculated either equal weighted (across all risk factors, or across all risk factors within an asset class) and an analogous value-weighted index which uses the standard deviation of a risk factor's profits in the previous month as weights, for risk factors with at least 15 trading days of data in the previous month. Remaining columns show the median beta and R^2 across factors within an asset class from regression of Risk Factor$_t^i = \alpha + \beta \text{Index}_t + \varepsilon_t$, where Index$_t$ is either the value-weighted (VW) aggregate risk factor index, the value-weighted risk factor index calculated across risk factors within the same asset class (AC), or the Fama–French equity market excess return. R^2 columns are in percent (e.g., 0.20 = 0.20%).

Asset Class	Period	Correlation Threshold						
		0.05	0.1	0.15	0.2	0.25	0.5	0.75
Commodities	Pre-2021	37	103	155	186	224	300	316
	Post-2020	30	93	139	190	218	291	301
Credit	Pre-2021	89	230	350	470	591	920	984
	Post-2020	68	185	308	416	518	756	823
Equities	Pre-2021	103	273	420	576	707	960	988
	Post-2020	105	260	411	550	640	857	883
Foreign Exchange	Pre-2021	46	117	177	231	281	387	401
	Post-2020	43	105	174	220	266	327	337
Financing and Collateral	Pre-2021	39	92	135	162	183	238	241
	Post-2020	34	77	114	134	150	181	183
Municipal	Pre-2021	14	34	45	54	64	89	95
	Post-2020	10	28	38	51	57	84	91
Overlay and Counterparty Risk	Pre-2021	27	76	121	166	206	321	344
	Post-2020	34	95	146	193	223	303	317
Rates	Pre-2021	70	175	281	375	464	672	709
	Post-2020	70	192	292	390	477	648	677
Securitized and Structured Products	Pre-2021	56	133	192	240	301	469	533
	Post-2020	37	99	153	203	251	392	450

Table 5: Cluster Count by Correlation Cutoff Threshold. Table shows the number of clusters resulting from the clustering process described above; all columns are based on cluster using the absolute value of correlations of the trailing 3-month return proxy within an asset class. The columns show the resulting clusters when adjusting the required (absolute value) correlation threshold when sorting risk factors into clusters.

	Pairs		AUC			Within Hit Rate			Between Hit Rate		
Asset Class	Within	Between	Full	Odd	Even	Odd	Even	Both	Odd	Even	Both
Commodities	2,549	148,937	0.84	0.75	0.75	0.70	0.68	0.48	0.29	0.29	0.14
Credit	15,579	1,382,665	0.78	0.71	0.69	0.66	0.62	0.44	0.32	0.32	0.16
Equities	9,676	1,461,704	0.83	0.72	0.75	0.63	0.68	0.44	0.27	0.27	0.13
Financing and Collateral	960	66,519	0.84	0.76	0.73	0.67	0.61	0.43	0.25	0.25	0.12
Foreign Exchange	3,804	265,172	0.81	0.73	0.71	0.67	0.63	0.44	0.31	0.30	0.15
Municipal	962	17,395	0.77	0.69	0.69	0.71	0.73	0.56	0.39	0.40	0.23
Overlay and Counterparty Risk	5,526	162,109	0.79	0.73	0.68	0.71	0.63	0.49	0.33	0.33	0.17
Rates	9,657	835,731	0.80	0.73	0.69	0.69	0.61	0.45	0.31	0.32	0.16
Securitized and Structured Products	11,918	511,980	0.79	0.72	0.72	0.71	0.72	0.57	0.36	0.37	0.20

Table 6: Within-Cluster Comovement. Table reports diagnostics that validate the correlation-based clustering of risk factors within each asset class. For each asset class, the first two columns report the number of factor pairs classified as within the same cluster ("Within") versus in different clusters ("Between"), restricting to pairs within the same asset class. Columns *AUC* report the area-under-the-curve statistic $\Pr(|\rho|_{\text{within}} > |\rho|_{\text{between}})$ computed using absolute correlations estimated on the full sample, odd trading days only, and even trading days only. Columns *Within Hit Rate* report the share of within-cluster pairs with $|\rho| \geq 0.10$ when correlations are estimated on odd days, on even days, and on both subsamples simultaneously; *Between Hit Rate* reports the same quantities for between-cluster pairs. AUC values substantially above 0.5 and large gaps in hit rates between within- and between-cluster pairs indicate that clusters capture persistent return comovement rather than correlations arising from sampling noise around the $|\rho| = 0.10$ cutoff.

	$\lvert\rho_{ij}\rvert$					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Bank	1.00***			0.99***	0.97***	0.99***
	(11.81)			(11.29)	(11.50)	(11.29)
Same Asset Class		0.10 **		0.07		−0.06
		(2.05)		(1.48)		(−1.34)
Same Cluster			13.77***		12.85***	12.90***
			(54.91)		(46.58)	(46.43)
Same Bank × Same AC				0.01		−0.13
				(0.08)		(−1.60)
Same Bank × Same Cluster					2.58***	2.69***
					(7.00)	(7.26)
Constant	8.78***	8.96***	8.95***	8.77***	8.76***	8.77***
	(109.46)	(108.96)	(109.71)	(108.80)	(109.17)	(108.80)
N	34,582,464	34,582,464	34,582,464	34,582,464	34,582,464	34,582,464
R^2	0.001	0.000	0.002	0.001	0.002	0.002

Table 7: Bank Specialization in Risk Factors. Table reports regressions of risk factor pair (absolute value of) correlations on dummies for whether the risk factors are in the same bank, same asset class, or same cluster. Sample includes all pairs that have at least one year of overlapping daily observations. Standard errors are two way clustered on the two factors in the pair. Coefficients are multiplied by 100 for readability.

	N	Mean	Std.Dev	Min	Max
Cluster HHI	2,901,082	8805.4	1961.5	1731.3	10000.0
$\ln(1 + \text{SD of PnL})$	2,901,082	12.2	2.8	0.1	19.2
Correlation	-0.29***				

Table 8: Bank Specialization Summary Statistics. Table reports summary statistics for two cluster-level measures computed at the cluster-by-day level. Cluster HHI is the Herfindahl-Hirschman Index of bank market shares within each risk factor cluster, where market shares are defined as a bank's sum of 90-day rolling standard deviations of PnL in the cluster divided by the total across all banks in that cluster on a given day. $\ln(1 + \text{SD of PnL})$ is the log of one plus the total 90-day rolling standard deviation of PnL across all banks in a cluster on a given day, capturing cluster size in dollar risk terms. The bottom panel reports the pairwise correlation between the two measures where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Based on risk clusters estimated in the post-2020 sample.

	(1)	(2)	(3)
Specialist	0.0888**	0.0624*	0.0606*
	(2.37)	(1.87)	(1.78)
Date FE	Yes		
Cluster × Date FE			
Bank FE		Yes	
Asset Class FE			
Asset Class × Date FE		Yes	Yes
Bank × Date FE			Yes
N	1,086,785	1,086,677	1,086,677
R^2	0.006	0.003	0.003

Table 9: Specialist Returns. Table reports estimates from regressions of risk-adjusted returns on a measure of bank specialization. The dependent variable is daily P&L divided by lagged Value-at-Risk at the desk level. *Specialist* is the bank's average share of absolute P&L allocated to each asset class over the previous 3 months (excluding date t), standardized to have zero mean and unit variance. Returns are winsorized at the 5th and 95th percentiles. Within R^2 reported. t-statistics shown using robust standard errors clustered by date and desk where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Unattributable Share					
	(1)	(2)	(3)	(4)	(5)	(6)
Specialist	−1.859**	−3.093***	−3.227***		−2.910***	−3.021***
	(−2.31)	(−3.32)	(−3.74)		(−3.23)	(−3.64)
Tenure				−3.363***	−2.804***	−2.978***
				(−2.97)	(−2.59)	(−2.72)
Asset Class × Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE		Yes			Yes	
Bank × Date FE			Yes	Yes		Yes
N	551052	551052	550955	550955	551052	550955

Table 10: Specialization and Model Accuracy. Table reports regressions of the unattributable share on measures of bank expertise. The unattributable share is the ratio of unexplained P&L (Items 5 and 10 under the FR VV-1 reporting requirements) to total valuation-related P&L from existing positions, with all components entering in absolute value. Specialist is the standardized trailing 90-day average share of a bank's gross risk factor activity allocated to each asset class. Tenure is the standardized cumulative number of active trading days in the asset class, restricted to desks whose first appearance postdates the bank's first reporting date. Both Specialist and Tenure are standardized to have zero mean and unit variance within each panel.Within R^2 reported. t-statistics shown using robust standard errors clustered by date and desk where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	# Distinct Banks over Horizon						# Distinct Desks over Horizon					
	Level		Volatility		RENTD		Level		Volatility		RENTD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Activity Level	0.208***	0.130***					0.519***	0.375***				
	(23.53)	(14.50)					(20.11)	(16.15)				
Activity Volatility			0.182***	0.115***					0.436***	0.317***		
			(23.55)	(14.41)					(20.54)	(16.01)		
RENTD Limit					0.050	0.002					1.013**	0.829**
					(0.73)	(0.04)					(2.58)	(1.97)
Horizon	1m	1y	1m	1y	1m	1y	1m	1y	1m	1y	1m	1y
Cluster FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Asset Class × Date FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	2,916,966	2,666,612	2,916,902	2,666,548	261,438	208,387	2,916,966	2,666,612	2,916,902	2,666,548	261,438	208,387
R^2	0.058	0.026	0.050	0.022	0.001	0.000	0.070	0.046	0.055	0.037	0.030	0.029

Table 11: Demand and Resource Allocation. Table reports regressions of future cluster participation on measures of demand. The dependent variable in Panel A is the maximum number of distinct banks active in the cluster over the subsequent 30 or 365 days; Panel B replaces banks with desks. Three demand proxies are considered: *Activity Level* is the log trailing 90-day average of total gross risk factor P&L across all banks in the cluster. *Activity Volatility* is the log trailing 90-day standard deviation of the same measure. *RENTD Limit* is derived from banks' internal position limits justified by Reasonably Expected Near-Term Demand analysis. For each USD-denominated RENTD li Xmit, we take the maximum absolute value of the upper and lower bounds as a measure of demand capacity. We aggregate to the cluster-date level by averaging across limits within each bank-cluster and then across banks. All independent variables are log-transformed and standardized to have zero mean and unit variance. Independent variables are standardized to have zero mean and unit variance within each panel. Within R^2 reported. t-statistics shown using robust standard errors clustered by date and cluster where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Dep. Var: $\Delta\text{Activity}_{bct}$					
	1-Day		3-Day		5-Day	
	Util	Near Cap	Util	Near Cap	Util	Near Cap
	(1)	(2)	(3)	(4)	(5)	(6)
Specialist_{bct}	−1.089	−5.028 **	−1.272 *	−3.514 **	−0.386	−0.729
	(−1.18)	(−2.26)	(−1.82)	(−2.31)	(−1.07)	(−0.98)
$\text{Near Capacity}_{bt} \times \text{Specialist}_{bct}$	11.690***		7.248***		2.592**	
	(4.15)		(3.49)		(2.12)	
$\text{Utilization}_{bt} \times \text{Specialist}_{bct}$		0.108***		0.064***		0.016
		(3.30)		(2.82)		(1.39)
N	5,598,020	5,598,020	6,956,742	6,956,742	7,150,618	7,150,618
R^2	0.000	0.000	0.000	0.000	0.000	0.000

Table 12: Binding Constraints and Reallocation Toward Core Expertise. Table reports regressions of changes in bank-cluster activity on the interaction between bank-level constraint tightness and cluster-level specialization. The dependent variable is $100 \times [\ln(\overline{\text{Activity}}^{\text{fwd}} + 1) - \ln(\overline{\text{Activity}}^{\text{trail}} + 1)]$, where forward activity is averaged over the subsequent 1, 3, or 5 days and trailing activity over the previous 10 days. Specialist is the standardized trailing 90-day average share of a bank's gross risk factor activity allocated to each cluster. Utilization is the VaR-weighted average of each desk's maximum limit proximity score across the bank, where proximity measures how close usage is to the allowable bound (0 = maximum slack, 100 = at limit). Near Capacity is the VaR-weighted share of desks with proximity above 80. All specifications include bank-by-date and cluster-by-date fixed effects. Within R^2 reported. t-statistics shown using robust standard errors clustered by date and cluster where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Destination			
Origin	Peripheral	Core	Total	% Core
Core	42	240	282	85.1%
Peripheral	49	50	99	50.5%
Total	91	290	381	

Table 13: Transition Matrix: Origin Peripherality and Destination Specialization. Table reports the number of trader moves between covered banks, classified by whether the trader's asset class was peripheral at the origin bank and core at the destination bank. A trader's asset class is classified as peripheral (core) if the origin bank's trailing share of gross risk factor activity in that asset class falls below (above) the bank's within-month median. Destination core is defined symmetrically. Sample includes all moves between covered banks where both origin and destination specialization data are available. The final column reports the share of movers landing at a core destination.

	Dep. Var: $\mathbf{1}[\text{Moved}_{it}]$					Dep. Var: ΔAC Share$_i$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Origin AC Share$_{it}$	-0.195***	-0.347***	-0.356***	-0.511*	-0.423**				
	(-2.70)	(-2.99)	(-3.06)	(-1.81)	(-2.28)				
Peripheral$_{it}$						7.860***	7.869***	9.349***	10.383***
						(6.40)	(6.34)	(4.29)	(4.37)
Origin AC Share$_{it}$ × Bank Constraint$_{bt}$					0.397				
					(1.03)				
Peripheral$_{it}$ × Demand Shock$_{at}$								4.485**	4.402*
								(2.12)	(2.03)
Bank Constraint$_{bt}$					-0.213**				
					(-2.14)				
Demand Shock$_{at}$								6.059***	5.146**
								(3.13)	(2.22)
Seniority$_i$			-0.043***		-0.043***		-0.482		-1.016
			(-7.04)		(-7.14)		(-1.04)		(-1.14)
Sample	All	All	All	Senior	All	Movers	Movers	Movers	Movers
AC FE					Yes	Yes	Yes	Yes	Yes
Month FE					Yes	Yes	Yes	Yes	Yes
AC × Month FE		Yes	Yes	Yes					
N	428,310	428,310	428,310	52,931	428,310	357	357	121	121
R^2	0.000	0.000	0.000	0.000	0.000	0.081	0.087	0.159	0.177

Table 14: Trader Sorting: Peripheral Traders Leave and Improve Match Quality. Columns (1)–(4) regress an indicator for departure to any firm on the origin bank's share of activity in the trader's asset class. The sample is all trader-months at covered banks, excluding final observations; column (4) restricts to senior traders (seniority ≥ 6). Column (5) tests the push channel: whether the gap in departure rates between core and peripheral traders widens when the origin bank faces binding risk limit constraints. Bank Constraint is the smoothed VaR-weighted share of the origin bank's desks with risk limit proximity above 80; a positive interaction with Origin AC Share indicates that core traders are relatively more retained when the bank is constrained, while peripheral traders are shed. Columns (6)–(9) regress the change in bank-AC specialization from origin to destination (dest AC share − origin AC share, multiplied by 100) on an indicator for whether the trader was peripheral at the origin bank. The sample is moves between covered banks where both origin and destination specialization are observed. Demand Shock is the smoothed standardized change in market-level RENTD limit size in the trader's asset class relative to its trailing average. A trader's asset class is classified as peripheral if the origin bank's share of activity in that asset class falls below the bank's within-month median. Dependent variables multiplied by 100 for readability. t-statistics in parentheses using standard errors clustered by month where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Internet Appendix to

Expertise and Specialization in Bank Trading

Chase P. Ross, Sharon Y. Ross, Kaushik Vasudevan

The Internet Appendix consists of three sections. Section IA.A provides additional details about the data. Section IA.B presents additional tables and figures. Section IA.C presents proofs for the model.

IA.A Data Cleaning and Construction

We clean the data in several ways.

We construct our analysis sample from regulatory filings submitted by banks under the Federal Reserve's Volcker Rule trading desk reporting requirements. The data come in two distinct reporting formats: the pre-2020 format and the post-2020 format, which we merge to create a consistent time series. The reporting instructions were updated in December 2020 to provide additional granularity on certain fields, particularly limit characteristics and desk classifications. We clean the data so observations before and after 2020 are directly comparable.

We focus on the banks that consistently report throughout the sample. All monetary values are converted to U.S. dollars using the bank-reported currency conversion rates. We exclude observations missing any of the key identifiers (bank, desk, or risk factor). We drop trading desks that never report meaningful activity throughout their listing period. Specifically, we exclude desks where both comprehensive P&L and VaR are either missing or zero for all dates. Similarly, we drop individual risk factors within a desk if their risk-move P&L is missing or zero for all dates.

We winsorize extreme observations to limit the influence of outliers. For risk factor P&L and desk-level variables (comprehensive P&L, VaR, and P&L components), we winsorize at the 0.1st and 99.9th percentiles. We balance the panel by identifying each desk's first and last active dates—defined as dates when the desk reports non-zero and non-missing P&L or VaR.

IA.A.1 Desk Classifications

We sort desks into ten groups: credit, equities, rates, foreign exchange, securitized and structured products, collateral and financing, commodities, overlay and counterparty risk, municipal, and legacy.

1. Credit: desks that trade corporate debt, credit derivatives, or support the firm's debt capital markets and leveraged loan businesses.

2. Equities: desks that trade equities, equity derivatives, or convertible products.

3. Rates: desks that trade government bonds, agency debt, interest rate swaps, swaptions, or macro products.

4. Foreign exchange: desks that trade currencies or currency derivatives.

5. Commodities: desks that trade commodities, including energy, metals, and agriculture.

6. Securitized and structured products: desks that trade asset-backed securities, securitized products (e.g., CLOs, CMOs, CDOs), non-agency mortgage products, or other structured products.

7. Municipal: desks that trade municipal, tax-exempt, or local government obligations.

8. Collateral and financing: desks that trade repos, collateral swaps, provide financing for internal counterparties or prime brokerage clients, or engage in securities financing.

9. Overlay and counterparty risk: desks that conduct valuation adjustment (VA) trading; including valuation adjustments to hedge counterparty risk (CVA), hedges to funding conditions (FVA), or desks that focus on centralized hedging of other desks' net positions or focus on regulatory capital or risk weighted asset optimization.

10. Legacy: desks that manage portfolios or lines of business that the firm is exiting or otherwise unwinding.

The classification uses a hybrid approach combining keyword matching with supervised machine learning, trained on a subset of desks where we have high confidence in the asset class assignment. For each asset class, we manually construct keyword lists based on domain knowledge of trading desk naming conventions. These lists include both standard terminology (e.g., "equity," "stock," "ipo" for the equity class) and specialized jargon (e.g., "sonia," "sofr," "eurodollar" for the rates class). We designate certain keywords as "dominant" when they unambiguously signal a particular asset class (e.g., "xva," "cva," "counterparty credit risk" for XVA desks). When a desk name or description matches a dominant keyword, we immediately assign that asset class without further disambiguation. For the remaining desks, we apply keyword matching to the combined desk name and description text. If text matches keywords from only one asset class, we assign that class. If text matches keywords from multiple asset classes we disambiguate using TF-IDF (term frequency-inverse document frequency) similarity. We exclude desks in the rare case they have no desk name or desk description from the classification, rather than setting them to unclassified.

IA.A.2 Measuring Risk Limit Utilization

We construct a unified measure of how close each trading desk is to its regulatory risk limits using data from the FR VV-1 reporting form. The measure captures the proximity of a desk's current risk exposure to its allowable bounds, expressed on a scale where 0 indicates maximum slack, 100

indicates usage exactly at the limit, and values above 100 indicate a breach. Our construction addresses reporting format changes across our sample period.

In the second subsample, banks report explicit upper and lower bounds along with the usage value for each limit, so the allowable range is directly observed. In the first subsample, banks report only a single limit size, a usage value, and a limit percent used field. To harmonize the two formats, we must determine whether each first-subsample limit represents a symmetric constraint (where exposure must stay within $\pm|\text{limit size}|$) or a directional constraint (where the limit size is a one-sided bound). For example, if a delta limit of +\$100M is symmetric, then usage of −\$80M represents 80% utilization; if it is directional, the desk is far from the constraint.

We exploit the bank's own reported limit percent used to classify each limit. When usage and the limit size have opposite signs, the signed ratio usage/limit size is negative, but a bank using an absolute value convention will report a positive limit percent used equal to |usage|/|limit size|. We flag a limit as symmetric whenever we observe a day on which the signed ratio is negative yet the reported limit percent used is approximately equal to the absolute value of that ratio (within a tolerance of one percentage point). If any such day exists for a given limit, we classify the entire limit as symmetric. Symmetric limits receive both bounds: an upper bound of $+|\text{limit size}|$ and a lower bound of $-|\text{limit size}|$. Remaining limits are treated as directional, with positive limit sizes assigned as upper bounds and negative limit sizes as lower bounds.

Given the harmonized bounds, we compute a proximity score for each limit-date observation based on how close the current usage is to the nearest binding constraint. For each limit, we compute the headroom as the distance from current usage to the nearest bound: for two-sided limits with upper bound u and lower bound l, this is $\min(u - x,\, x - l)$; for one-sided limits, it is the distance to the single reported bound. We normalize headroom by the half-width of the allowable range—$(u - l)/2$ for two-sided limits, or the absolute value of the bound for one-sided limits—so that the resulting proximity score equals 0 at the midpoint of the allowable range, 100 at either boundary, and exceeds 100 in the event of a breach. For one-sided limits, usage on the opposite side of the constraint can produce negative proximity values, indicating that the desk is moving away from the relevant bound. We clamp values at zero from below and winsorize the right tail of proximity at the 99.9th percentile, so that all observations far from their relevant constraint register as unconstrained.

We aggregate per-limit proximity scores to the desk-date level by taking the maximum across all limits on the desk, capturing the most binding constraint. As a robustness check, we also compute the mean of the top three proximity scores, which is less sensitive to outliers from data errors or discrete limit resets. To construct a bank-date constraint measure, we aggregate across desks using Value-at-Risk weights, so that a small desk near capacity contributes less than a large desk under pressure. We also construct binary indicators for whether any desk on the bank experienced a breach (proximity above 100) or was near capacity (proximity above 80) on a given day.

IA.B Figures and Tables

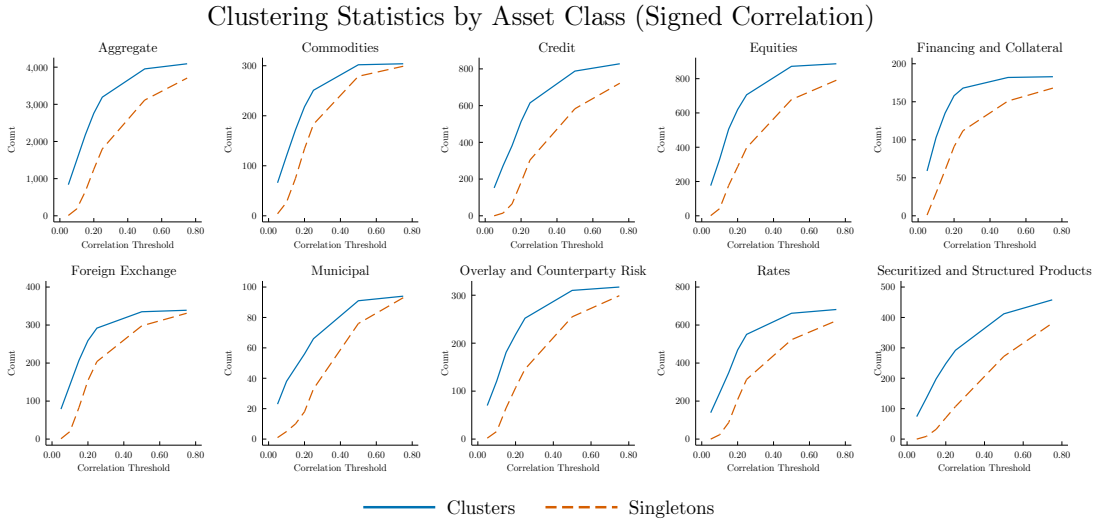


Figure IA.1: Clusters and Singletons across Correlation Thresholds. Figure plots the number of resulting clusters, along with number of singleton clusters that contain only a single risk factor, when categorizing risk factors based on the signed correlation threshold of our 90 day return proxy (as opposed to the benchmark clustering based on the absolute value of correlations). Figure based on post-2020 sample clustering.

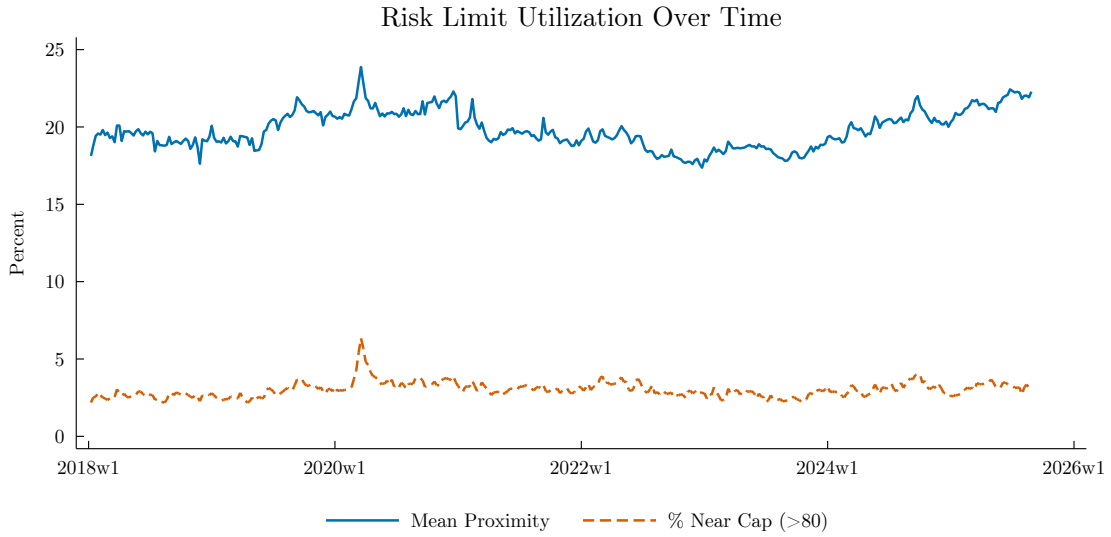


Figure IA.4: Risk Limit Utilization Over Time. Weekly averages of risk limit proximity (solid) and the share of limit-date observations near capacity (proximity above 80, dashed). Proximity measures how close current usage is to the allowable bound, on a scale where 0 indicates maximum slack and 100 indicates usage at the limit. The sample includes all limit types and sources from the FR VV-1 reporting form.

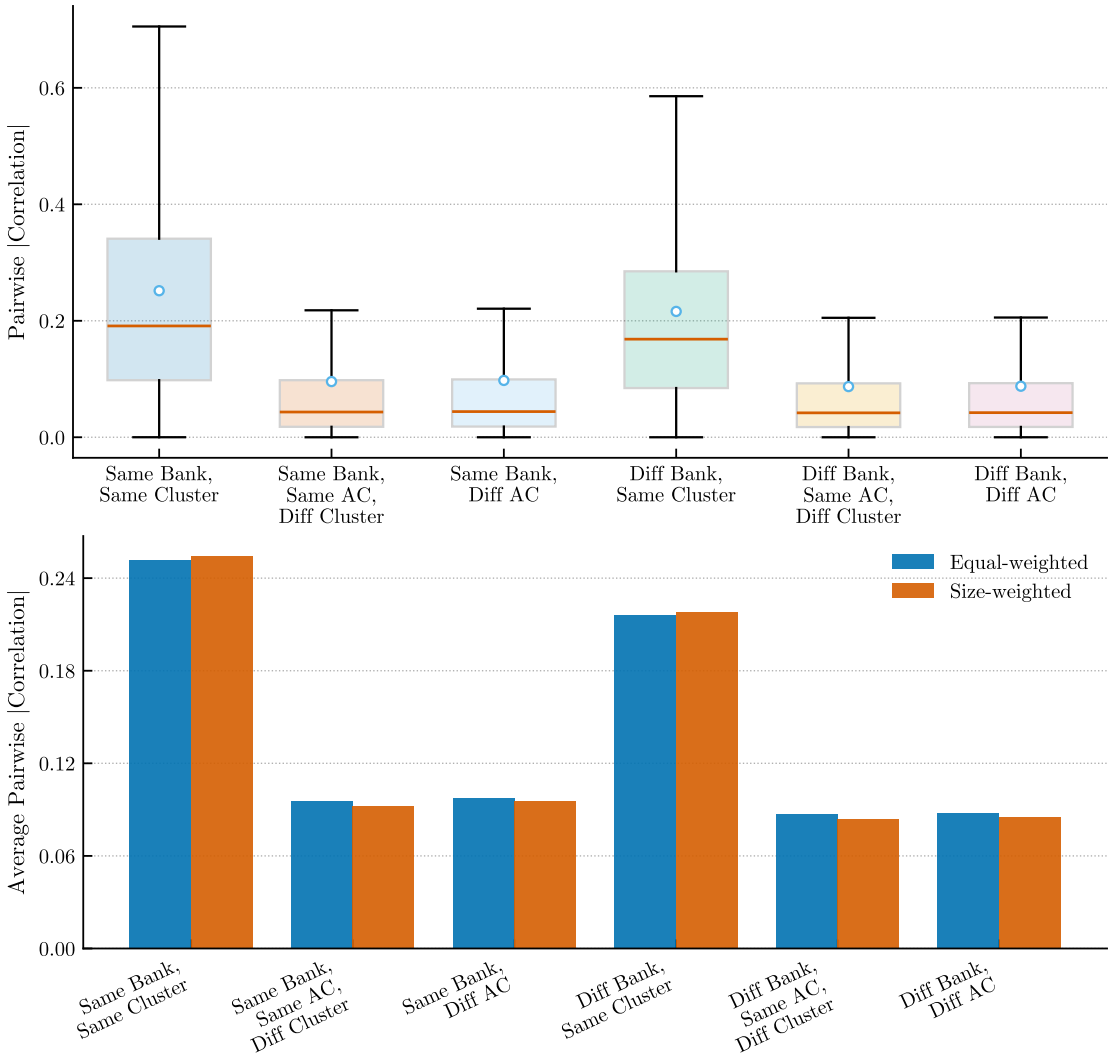


Figure IA.2: Correlations of Risk Factors. Figure plots the (absolute value) of correlations of risk factors across clusters, banks, and asset classes (AC), where the clusters are estimated using the benchmark approach using the absolute value of our 90 day return proxy.

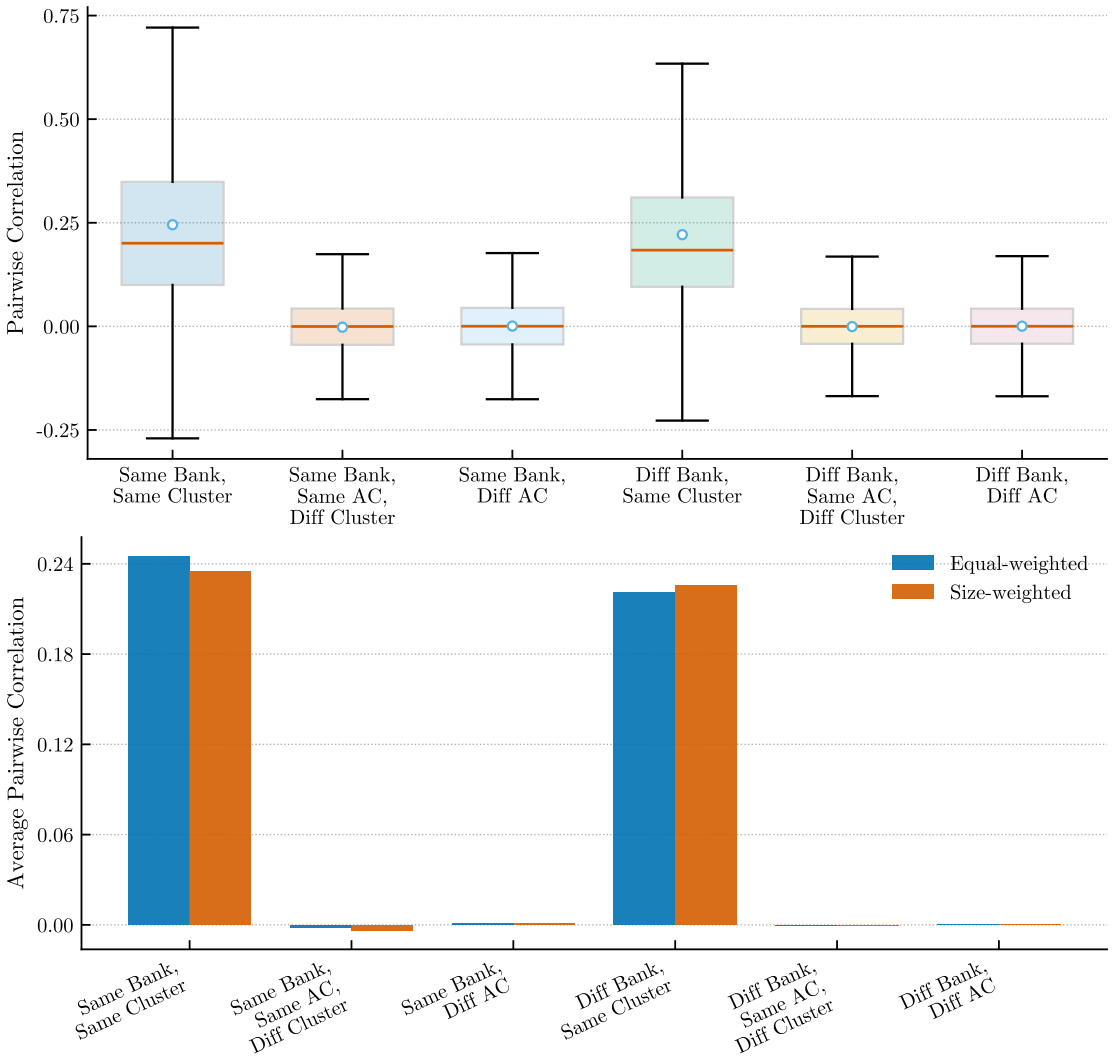


Figure IA.3: Correlations of Risk Factors when clustering with signed return proxy. Figure plots the (absolute value) of correlations of risk factors across clusters, banks, and asset classes, where the clusters are estimated using the benchmark approach using the correlations of signed 90 day return proxy.

	ρ_{ij}					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Bank	0.02 *			0.03 ***	−0.00	0.03 ***
	(1.74)			(2.72)	(−0.50)	(2.72)
Same Asset Class		0.03***		0.05 ***		−0.10 ***
		(3.28)		(4.30)		(−9.18)
Same Cluster			22.82***		22.10***	22.18 ***
			(87.84)		(76.27)	(76.47)
Same Bank × Same AC				−0.07 ***		−0.17 ***
				(−3.12)		(−8.08)
Same Bank × Same Cluster					2.40***	2.55 ***
					(5.70)	(6.02)
Constant	0.05***	0.05***	0.03***	0.05 ***	0.03***	0.05 ***
	(9.71)	(9.83)	(6.26)	(8.32)	(6.00)	(8.32)
N	34,582,464	34,582,464	34,582,464	34,582,464	34,582,464	34,582,464
R^2	0.000	0.000	0.002	0.000	0.002	0.002

Table IA.1: Bank Specialization in Risk Factors when clustering on signed returns. Table reports regressions of risk factor pair correlations on dummies for whether the risk factors are in the same bank, same asset class, or same cluster. Identical to Table 7 except the clustering here is calculated by using signed returns rather than the absolute value of returns. Sample includes all pairs that have at least one year of overlapping daily observations. Standard errors are two way clustered on the two factors in the pair. Coefficients are multiplied by 100 for readability.

IA.C Model Proofs

This section presents proofs for the model.

IA.C.1 Pricing Equilibrium Conditional on Hiring Choices

Proof of Proposition 1

Proof. Consider factor f in cluster c. We conjecture the pricing function is of the form

$$p_f = A_c\theta_f - B_c D_f.$$

Denote τ_c^p the precision of the signal about θ that each bank extracts from prices, which, given our conjecture, can be written as

$$\tau_c^p = \frac{A_c^2}{B_c^2\sigma_D^2}.$$

We can then write bank b's posterior precision of θ_f (for all factors in cluster c) as

$$T_{b,c} = \sigma_\theta^{-2} + \tau_c^p + x_{b,c},$$

and write bank b's expectation of θ_f given their information set (including prices) as

$$\mathbb{E}_b(\theta_f|\mathcal{I}_b) = \frac{x_{b,c}}{T_{b,c}}q_{b,f} + \frac{\tau_c^p}{T_{b,c}}\frac{p_f}{A_c}.$$

Denoting $\mathcal{I}_b$ as b's information set, and noting that bank b's posterior variance is $\frac{1}{T_{b,c}} + \sigma_\eta^2$, we can write bank b's first order condition as

$$z_{b,f} = \frac{\mathbb{E}_b(\theta_f|\mathcal{I}_b) - p_f}{\gamma(T_{b,c}^{-1} + \sigma_\eta^2)}. \tag{6}$$

The market clearing condition yields:

$$\begin{aligned} D_f =& \int_0^1 z_{b,f} \\ =& \frac{1}{\gamma}\int_0^1 \frac{\mathbb{E}_b(\theta_f|\mathcal{I}_b) - p_f}{(T_{b,c}^{-1} + \sigma_\eta^2)}db. \end{aligned}$$

Defining $\kappa_{b,c} \equiv \frac{T_{b,c}}{1+T_{b,c}\sigma_\eta^2}$, $\hat{\kappa}_c \equiv \int_0^1 \kappa_{b,c}db$, and $\hat{\theta}_f \equiv \frac{\int_0^1 \kappa_{b,c}\mathbb{E}_b(\theta_f|\mathcal{I}_b)}{\hat{\kappa}_c}$, we can re-arrange the market

clearing condition as

$$\begin{aligned}
p_f =& \hat{\theta}_f - \frac{\gamma}{\hat{\kappa}_c} D_f \\
=& \int_0^1 \frac{\kappa_{b,c}}{\hat{\kappa}_c} \left(\frac{x_{b,c}}{T_{b,c}} q_{b,f} + \frac{\tau_c^p}{T_{b,c}} \frac{p_f}{A_c} \right) db - \frac{\gamma}{\hat{\kappa}_c} D_f \\
=& \frac{\tau_c^p}{\hat{\kappa}_c A_c} p_f \int_0^1 \left(1 + T_{b,c}\sigma_\eta^2\right)^{-1} db + \frac{\int_0^1 x_{b,c}(1 + T_{b,c}\sigma_\eta^2)^{-1} db}{\hat{\kappa}_c} \theta_f - \frac{\gamma}{\hat{\kappa}_c} D_f.
\end{aligned}$$

Then, we can write

$$p_f = \frac{A_c \int_0^1 x_{b,c}(1 + T_{b,c}\sigma_\eta^2)^{-1} db}{\hat{\kappa}_c A_c - \tau_c^p \int_0^1 (1 + T_{b,c}\sigma_\eta^2)^{-1}) db} \theta_f - \frac{\gamma A_c}{\hat{\kappa}_c A_c - \tau_c^p \int_0^1 (1 + T_{b,c}\sigma_\eta^2)^{-1}) db} D_f. \tag{7}$$

Matching coefficients yields the system of equations:

$$\begin{aligned}
A_c =& \frac{A_c \int_0^1 x_{b,c}(1 + T_{b,c}\sigma_\eta^2)^{-1} db}{\hat{\kappa}_c A_c - \tau_c^p \int_0^1 (1 + T_{b,c}\sigma_\eta^2)^{-1} db}, \text{ and} \\
B_c =& \frac{\gamma A_c}{\hat{\kappa}_c A_c - \tau_c^p \int_0^1 (1 + T_{b,c}\sigma_\eta^2)^{-1} db}.
\end{aligned}$$

Solving the system, we get:

$$A_c = \frac{\int_0^1 x_{b,c}(1 + T_{b,c}\sigma_\eta^2)^{-1} db + \tau_c^p \int_0^1 (1 + T_{b,c}\sigma_\eta^2)^{-1} db}{\hat{\kappa}_c}, \text{ and} \tag{8}$$

$$B_c = \frac{\gamma(\int_0^1 x_{b,c}(1 + T_{b,c}\sigma_\eta^2)^{-1} db + \tau_c^p \int_0^1 (1 + T_{b,c}\sigma_\eta^2)^{-1} db)}{\hat{\kappa}_c \int_0^1 x_{b,c}(1 + T_{b,c}\sigma_\eta^2)^{-1} db}. \tag{9}$$

Lastly, we can solve for price informativeness, τ_c^p, which is the solution to the fixed point problem

$$\begin{aligned}
\tau_c^p =& \left(\frac{A_c}{B_c}\right)^2 \sigma_D^{-2} \\
=& \left(\int_0^1 x_{b,c}(1 + T_{b,c}\sigma_\eta^2)^{-1} db \right)^2 \frac{1}{\gamma^2 \sigma_D^2} \\
=& \frac{1}{\gamma^2 \sigma_D^2} \left(\int_0^1 \frac{x_{b,c}}{(1 + \sigma_\eta^2(\sigma_\theta^{-2} + \tau_c^p + x_{b,c}))} db \right)^2,
\end{aligned}$$

where uniqueness comes from the fact that the left hand side is strictly increasing in τ_c^p and the right hand side is weakly decreasing in τ_c^p on $[0, \infty)$. □

IA.C.2 Hiring Equilibrium and Specialization

To prove Proposition 2, we proceed in three steps. First, in Lemma 1, we rewrite bank b's period zero problem in terms of the utility it gets from hiring in cluster c, denoted by Ψ_c. Second, in

Lemma 2, we provides a constructive condition under which there is an equilibrium where banks specialize by hiring only in one cluster. Third, in Lemma 3, we provide a condition under which there is no equilibrium where all banks choose the same hiring strategy across clusters. Finally, in Proposition 2, we show that when the bank-level hiring constraint is sufficiently binding, there is a non-empty interval of values of β for which both lemmas apply at the same time.

Lemma 1 (Banks' Hiring Problem)**.** *We can represent bank b's period zero problem as*

$$\max_{n_{b,c}} \sum_{c=1}^{C} \Psi_c(n_{b,c}; \tau_c^p, \hat{\kappa}_c) \qquad \textit{subject to } \sum_{c=1}^{C} n_{b,c} \leq N,$$

where

$$\begin{aligned}
\Psi_c(n; \tau_c^p, \hat{\kappa}_c) \equiv & \frac{L\, x(n)}{2\gamma S_c \left[1 + \sigma_\eta^2 (S_c + x(n))\right]} + \frac{L\gamma\sigma_D^2}{2\hat{\kappa}_c^2} \left[\frac{S_c + x(n)}{1 + \sigma_\eta^2 (S_c + x(n))} - \frac{S_c}{1 + \sigma_\eta^2 S_c}\right], \\
x(n) \equiv & \alpha n + \frac{\beta n^2}{1 + \lambda n}, \textit{ and} \\
S_c \equiv & \sigma_\theta^{-2} + \tau_c^p.
\end{aligned}$$

Additionally, we can write

$$\begin{aligned}
\Psi_c'(n; \tau_c^p, \hat{\kappa}_c) = & \frac{L\, x'(n)}{2\left[1 + \sigma_\eta^2 (S_c + x(n))\right]^2} \left[\frac{1 + \sigma_\eta^2 S_c}{\gamma S_c} + \frac{\gamma\sigma_D^2}{\hat{\kappa}_c^2}\right], \textit{ and} \\
\Psi_c''(n; \tau_c^p, \hat{\kappa}_c) = & \frac{L}{2\left[1 + \sigma_\eta^2 (S_c + x(n))\right]^3} \left[\frac{1 + \sigma_\eta^2 S_c}{\gamma S_c} + \frac{\gamma\sigma_D^2}{\hat{\kappa}_c^2}\right] \mathcal{H}_c(n),
\end{aligned}$$

where

$$\mathcal{H}_c(n) = \frac{2\beta}{(1 + \lambda n)^3} \left[1 + \sigma_\eta^2 \left(S_c + x(n)\right)\right] - 2\sigma_\eta^2 \left(\alpha + \frac{\beta n(2 + \lambda n)}{(1 + \lambda n)^2}\right)^2$$

is strictly decreasing in n.

Proof. First, we compute quantities corresponding with bank b's choice after their hiring choices.

Plugging bank b's first order condition from Equation (6) into its objective function, we can observe that b's utility from their factor f position is

$$\frac{T_{b,c}}{2\gamma(1 + \sigma_\eta^2 T_{b,c})} \left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right)^2. \tag{10}$$

To compute $\mathbb{E}_b[\theta_f \mid \mathcal{I}_b]$, we note that $T_{b,c} = S_c + x_{b,c}$, and that we can write the signal extracted from prices as

$$\mathbb{E}(\theta_f \mid p_f) = \frac{\tau_c^p}{S_c} \frac{p_f}{A_c}.$$

Doing so, we can observe that

$$\begin{aligned}\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] =& \frac{S_c}{T_{b,c}}\,\mathbb{E}[\theta_f \mid p_f] + \frac{x_{b,c}}{T_{b,c}}\,q_{b,f} \\ =& \frac{S_c}{T_{b,c}}\left(\frac{\tau_c^p}{S_c}\frac{p_f}{A_c}\right) + \frac{x_{b,c}}{T_{b,c}} q_{b,f} \\ =& \hat{\theta}_f + \frac{x_{b,c}}{T_{b,c}}(q_{b,f} - \hat{\theta}_f),\end{aligned}$$

where the last step comes from noting that along the linear REE, the public posterior mean from prices coincides with the market posterior $\hat{\theta}_f$. We then compute $\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f$ by substituting in $p_f = \hat{\theta}_f - \frac{\gamma D_f}{\hat{\kappa}_c}$ from Proposition 1, which yields

$$\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f = \frac{x_{b,c}}{T_{b,c}}(q_{b,f} - \hat{\theta}_f) + \frac{\gamma D_f}{\hat{\kappa}_c}. \tag{11}$$

Next, we compute the ex-ante second moment of

$$\mathbb{E}\left((\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f)^2\right),$$

i.e., after hiring choices but before signals are realized. We denote

$$\begin{aligned}\Delta_{b,f} \equiv& q_{b,f} - \hat{\theta}_f \\ =& (\theta_f - \hat{\theta}_f) + u_{b,f}.\end{aligned}$$

We observe that conditioning only on prices (and not private information), $\theta_f - \hat{\theta}_f$ has mean zero and variance $1/S_c$, while $u_{b,f}$ has mean zero and variance $1/x_{b,c}$. These two terms are independent, hence

$$\begin{aligned}\mathbb{E}[\Delta_{b,f} \mid p_f] =& 0, \text{ and} \\ \mathrm{Var}(\Delta_{b,f} \mid p_f) =& \frac{1}{S_c} + \frac{1}{x_{b,c}}.\end{aligned}$$

Hence, we have that

$$(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f)^2 = \left(\frac{x_{b,c}}{T_{b,c}}\Delta_{b,f}\right)^2 + \left(\frac{\gamma D_f}{\hat{\kappa}_c}\right)^2 + 2\frac{x_{b,c}}{T_{b,c}}\Delta_{b,f}\frac{\gamma D_f}{\hat{\kappa}_c}.$$

Taking expectations, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{x_{b,c}}{T_{b,c}}\Delta_{b,f}\right)^2\right] &= \left(\frac{x_{b,c}}{T_{b,c}}\right)^2 \mathbb{E}[\Delta_{b,f}^2] \\
&= \left(\frac{x_{b,c}}{T_{b,c}}\right)^2 \left(\frac{1}{S_c} + \frac{1}{x_{b,c}}\right) \\
&= \frac{x_{b,c}}{S_c T_{b,c}}, \\
\mathbb{E}\left[\left(\frac{\gamma D_f}{\hat{\kappa}_c}\right)^2\right] &= \frac{\gamma^2 \sigma_D^2}{\hat{\kappa}_c^2},
\end{aligned}$$

and the last term is zero since D_f is independent of $\theta_f - \hat{\theta}_f$ and $u_{b,f}$. Therefore,

$$\mathbb{E}\Big[(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f)^2\Big] = \frac{x_{b,c}}{S_c T_{b,c}} + \frac{\gamma^2 \sigma_D^2}{\hat{\kappa}_c^2}. \tag{12}$$

Then, we can take expectations of Equation (10) and observe that

$$\begin{aligned}
\mathbb{E}\left(\frac{T_{b,c}}{2\gamma(1+\sigma_\eta^2 T_{b,c})}\left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right)^2\right) =& \frac{T_{b,c}}{2\gamma(1+\sigma_\eta^2 T_{b,c})}\left[\frac{x_{b,c}}{S_c T_{b,c}} + \frac{\gamma^2\sigma_D^2}{\hat{\kappa}_c^2}\right] \\
=& \frac{x_{b,c}}{2\gamma S_c(1+\sigma_\eta^2 T_{b,c})} + \frac{\gamma\sigma_D^2}{2\hat{\kappa}_c^2}\frac{T_{b,c}}{1+\sigma_\eta^2 T_{b,c}}.
\end{aligned}$$

Since there are L independent factors per cluster, b's utility from their optimal hiring choice in cluster c can be expressed as

$$\frac{L\, x_{b,c}}{2\gamma S_c(1+\sigma_\eta^2 T_{b,c})} + \frac{L\gamma\sigma_D^2}{2\hat{\kappa}_c^2}\frac{T_{b,c}}{1+\sigma_\eta^2 T_{b,c}}. \tag{13}$$

To compute Ψ_c, we also subtract off bank b's utility at zero hiring, which is a constant with respect to bank b's choice of $n_{b,c}$, and does not affect their optimization problem, but yields the convenient property that $\Psi_c(0) = 0$. The zero-hiring utility is

$$\frac{L\gamma\sigma_D^2}{2\hat{\kappa}_c^2}\frac{S_c}{1+\sigma_\eta^2 S_c},$$

and subtracting it off yields

$$\Psi_c = \frac{L\, x_{b,c}}{2\gamma S_c(1+\sigma_\eta^2 T_{b,c})} + \frac{L\gamma\sigma_D^2}{2\hat{\kappa}_c^2}\left[\frac{T_{b,c}}{1+\sigma_\eta^2 T_{b,c}} - \frac{S_c}{1+\sigma_\eta^2 S_c}\right].$$

The expression in the proposition comes from substituting $x_{b,c} = x(n)$ and $T_{b,c} = S_c + x(n)$.

Next, we compute the first derivative of Ψ_c. We can write

$$\Psi_c(n; \tau_c^p, \hat{\kappa}_c) = \frac{L}{2}\left[\frac{x(n)}{\gamma S_c\left(1+\sigma_\eta^2(S_c+x(n))\right)} + \frac{\gamma\sigma_D^2}{\hat{\kappa}_c^2}\frac{S_c+x(n)}{1+\sigma_\eta^2(S_c+x(n))}\right] \;+\; \text{constant},$$

where the constant is the zero-hiring benchmark. Differentiating the first term, we have

$$\begin{aligned}\frac{d}{dn}\left(\frac{x(n)}{1+\sigma_\eta^2(S_c+x(n))}\right)=&\frac{x'(n)\left[1+\sigma_\eta^2(S_c+x(n))\right]-x(n)\sigma_\eta^2x'(n)}{\left[1+\sigma_\eta^2(S_c+x(n))\right]^2}\\=&\frac{x'(n)(1+\sigma_\eta^2S_c)}{\left[1+\sigma_\eta^2(S_c+x(n))\right]^2},\end{aligned}$$

and differentiating the second term, we have

$$\begin{aligned}\frac{d}{dn}\left(\frac{S_c+x(n)}{1+\sigma_\eta^2(S_c+x(n))}\right)=&\frac{x'(n)\left[1+\sigma_\eta^2(S_c+x(n))\right]-(S_c+x(n))\sigma_\eta^2x'(n)}{\left[1+\sigma_\eta^2(S_c+x(n))\right]^2}\\=&\frac{x'(n)}{\left[1+\sigma_\eta^2(S_c+x(n))\right]^2}.\end{aligned}$$

Combining these two expressions yields the expression for $\Psi_c'(n;\tau_c^p,\hat{\kappa}_c)$ in the lemma.

Lastly, we compute the second derivative of Ψ_c. Using the expression for $\Psi_c'(n;\tau_c^p,\hat{\kappa}_c)$, we have

$$\Psi_c'(n;\tau_c^p,\hat{\kappa}_c)=\frac{L}{2}\left[\frac{1+\sigma_\eta^2S_c}{\gamma S_c}+\frac{\gamma\sigma_D^2}{\hat{\kappa}_c^2}\right]\frac{x'(n)}{\left[1+\sigma_\eta^2(S_c+x(n))\right]^2}.$$

The bracketed term is constant in n, so differentiating once more gives

$$\begin{aligned}\Psi_c''(n;\tau_c^p,\hat{\kappa}_c)=&\frac{L}{2}\left[\frac{1+\sigma_\eta^2S_c}{\gamma S_c}+\frac{\gamma\sigma_D^2}{\hat{\kappa}_c^2}\right]\frac{d}{dn}\left(\frac{x'(n)}{\left[1+\sigma_\eta^2(S_c+x(n))\right]^2}\right)\\=&\frac{L}{2}\left[\frac{1+\sigma_\eta^2S_c}{\gamma S_c}+\frac{\gamma\sigma_D^2}{\hat{\kappa}_c^2}\right]\frac{x''(n)\left[1+\sigma_\eta^2(S_c+x(n))\right]-2\sigma_\eta^2(x'(n))^2}{\left[1+\sigma_\eta^2(S_c+x(n))\right]^3}.\end{aligned}$$

This gives the expression for $\Psi_c''(n;\tau_c^p,\hat{\kappa}_c)$ in the lemma, where

$$\mathcal{H}_c(n)=x''(n)\left[1+\sigma_\eta^2(S_c+x(n))\right]-2\sigma_\eta^2(x'(n))^2.$$

Finally, using

$$x(n)=\alpha n+\frac{\beta n^2}{1+\lambda n},$$

we have

$$x'(n)=\alpha+\frac{\beta n(2+\lambda n)}{(1+\lambda n)^2},\qquad x''(n)=\frac{2\beta}{(1+\lambda n)^3},\qquad x'''(n)=-\frac{6\beta\lambda}{(1+\lambda n)^4}.$$

Substituting these expressions into $\mathcal{H}_c(n)$ yields the expression in the lemma. Differentiating $\mathcal{H}_c(n)$,

we have

$$\mathcal{H}_c'(n) = x'''(n)\left[1 + \sigma_\eta^2(S_c + x(n))\right] - 3\sigma_\eta^2 x'(n)x''(n),$$

which is strictly negative since $x'(n) > 0$, $x''(n) > 0$, and $x'''(n) < 0$. Hence $\mathcal{H}_c(n)$ is strictly decreasing in n. □

Lemma 2 (Existence of a specializing equilibrium)**.** *Under the conjecture that each bank hires only in one cluster, let τ_1^p denote the unique nonnegative solution to*

$$\gamma\sigma_D\sqrt{\tau_1^p} = \frac{1}{C}\frac{x(N)}{1 + \sigma_\eta^2\left(\sigma_\theta^{-2} + \tau_1^p + x(N)\right)}.$$

Let the corresponding aggregate loading be

$$\hat{\kappa}_1 = \frac{1}{C}\frac{\sigma_\theta^{-2} + \tau_1^p + x(N)}{1 + \sigma_\eta^2\left(\sigma_\theta^{-2} + \tau_1^p + x(N)\right)} + \frac{C-1}{C}\frac{\sigma_\theta^{-2} + \tau_1^p}{1 + \sigma_\eta^2\left(\sigma_\theta^{-2} + \tau_1^p\right)}.$$

Additionally, let

$$G_1(n) \equiv \Psi(n; \tau_1^p, \hat{\kappa}_1) \qquad \textit{and} \qquad A_1(n) \equiv \frac{G_1(n)}{n}.$$

If

$$A_1(N) \geq A_1(n) \qquad \textit{for all } n \in (0, N],$$

then there is an equilibrium in which banks specialize by hiring only in one cluster.

Proof. First, under the conjecture that each bank hires only in one cluster, a bank is active in a given cluster with probability $1/C$ and inactive otherwise. Proposition 1 then implies that the common price informativeness must satisfy

$$\gamma\sigma_D\sqrt{\tau_1^p} = \frac{1}{C}\frac{x(N)}{1 + \sigma_\eta^2\left(\sigma_\theta^{-2} + \tau_1^p + x(N)\right)}.$$

This gives the expression above for τ_1^p. The corresponding aggregate loading is the cross-sectional average of the active and inactive loadings, which gives the formula above for $\hat{\kappa}_1$.

Next, given $(\tau_1^p, \hat{\kappa}_1)$, bank b solves

$$\max_{\{n_{b,c}\}_{c=1}^C} \sum_{c=1}^C G_1(n_{b,c}) \qquad \text{subject to} \qquad \sum_{c=1}^C n_{b,c} \leq N.$$

Since $\Psi_c'(0) > 0$, bank b uses its full hiring capacity, so $\sum_{c=1}^C n_{b,c} = N$. Then, for any feasible hiring

vector, we can write

$$\begin{aligned}\sum_{c=1}^{C} G_1(n_{b,c}) =& \sum_{c=1}^{C} n_{b,c} A_1(n_{b,c}) \\ \leq& \sum_{c=1}^{C} n_{b,c} \sup_{m\in(0,N]} A_1(m) \\ =& N \sup_{m\in(0,N]} A_1(m).\end{aligned}$$

Lastly, if $A_1(N) \geq A_1(n)$ for all $n \in (0, N]$, then the right hand side is attained by setting one cluster equal to N and all remaining clusters equal to 0. Hence hiring only in one cluster is a best response. Since banks are ex-ante identical, assigning banks symmetrically across clusters yields an equilibrium in which banks specialize by hiring only in one cluster. □

Lemma 3 (Non-existence of a symmetric equilibrium)**.** *For each $n \in [0, N]$, let $\tau^p(n)$ denote the unique nonnegative solution to*

$$\gamma\sigma_D\sqrt{\tau^p(n)} = \frac{x(n)}{1+\sigma_\eta^2\left(\sigma_\theta^{-2}+\tau^p(n)+x(n)\right)},$$

and let

$$\hat{\kappa}(n) = \frac{\sigma_\theta^{-2}+\tau^p(n)+x(n)}{1+\sigma_\eta^2\left(\sigma_\theta^{-2}+\tau^p(n)+x(n)\right)}$$

denote the corresponding aggregate loading under the symmetric hiring profile. Additionally, define

$$\mathcal{M}(n) \equiv \partial_n \Psi(n; \tau^p(n), \hat{\kappa}(n)).$$

If the common hiring vector that assigns N/C workers to every cluster is not optimal under its induced aggregate state and $\mathcal{M}(n)$ is strictly decreasing on $[0, N]$, then there is no equilibrium in which all banks choose the same hiring strategy across clusters.

Proof. Suppose, towards a contradiction, that there is an equilibrium in which every bank chooses the same deterministic vector

$$(n_1, \ldots, n_C), \qquad \sum_{c=1}^{C} n_c \leq N.$$

If this common vector is flat, so that $n_c = N/C$ for every c, then it is ruled out by assumption.

Next, suppose the common vector is non-flat. Then there exist two clusters, say h and ℓ, such that $n_h > n_\ell$. Consider a deviation that moves $\varepsilon > 0$ units of hiring from cluster h to cluster ℓ. The first-order gain from this deviation is

$$\varepsilon\left[\mathcal{M}(n_\ell) - \mathcal{M}(n_h)\right] + o(\varepsilon).$$

Since $\mathcal{M}(n)$ is strictly decreasing, we have $\mathcal{M}(n_\ell) > \mathcal{M}(n_h)$, so this gain is strictly positive for ε small enough.

Lastly, this implies that the common hiring vector cannot be an equilibrium. Therefore, there is no equilibrium in which all banks choose the same hiring strategy across clusters. □

Proof of Proposition 2

Proof. First, we define the lower and upper bounds on β that will be used in the proof. Write

$$B(N) \equiv \frac{1+\sigma_\eta^2(\sigma_\theta^{-2}+\alpha N)}{(1+\lambda N)^3} - \frac{2\sigma_\eta^2\alpha N(2+\lambda N)}{(1+\lambda N)^2},$$

and

$$C(N) \equiv \sigma_\eta^2 N^2 \frac{(2+\lambda N)^2-1}{(1+\lambda N)^4}.$$

Whenever

$$B(N)^2 - 4\sigma_\eta^2\alpha^2 C(N) > 0,$$

define

$$\underline{\beta}(N) \equiv \frac{B(N) - \sqrt{B(N)^2 - 4\sigma_\eta^2\alpha^2 C(N)}}{2C(N)}.$$

Also, define

$$y_N(\beta) \equiv 1 + \sigma_\eta^2\left(\sigma_\theta^{-2} + \alpha N + \frac{\beta N^2}{1+\lambda N} + \frac{\left(\alpha N + \frac{\beta N^2}{1+\lambda N}\right)^2}{\gamma^2\sigma_D^2}\right),$$

and let $\beta^\star(N)$ denote the unique positive solution to

$$\beta = \frac{\sigma_\eta^2\alpha^2}{y_N(\beta)} + \frac{\gamma\sigma_D^2\alpha^2(\sigma_\eta^2)^3}{y_N(\beta)^2\left[\frac{y_N(\beta)}{\gamma\sigma_\theta^{-2}} + \frac{\gamma\sigma_D^2 y_N(\beta)^2}{\sigma_\theta^{-4}}\right]}.$$

Lastly, define

$$\overline{\beta}(N) \equiv \min\left\{\frac{B(N) + \sqrt{B(N)^2 - 4\sigma_\eta^2\alpha^2 C(N)}}{2C(N)}, \beta^\star(N)\right\}.$$

Next, we show that every $\beta \in (\underline{\beta}(N), \overline{\beta}(N))$ satisfies Lemma 2 and Lemma 3. Fix $\beta \in (\underline{\beta}(N), \overline{\beta}(N))$.

Because β lies between the two roots of

$$-\sigma_\eta^2\alpha^2 + B(N)\beta - C(N)\beta^2 = 0,$$

we have

$$-\sigma_\eta^2\alpha^2 + B(N)\beta - C(N)\beta^2 \geq 0.$$

Since the expression multiplying Ψ_c'' in Lemma 1 is strictly decreasing in n, this inequality implies that it is positive for every $n \in [0, N]$ when a bank puts all of its hiring in one cluster. Therefore bank utility from hiring in one cluster is strictly convex on $[0, N]$. Hence

$$A_1(n) = \frac{G_1(n)}{n}$$

is strictly increasing on $(0, N]$, so

$$A_1(N) \geq A_1(n) \qquad \text{for all } n \in (0, N].$$

Lemma 2 then implies that there is an equilibrium in which banks specialize by hiring only in one cluster.

Next, because $\beta < \beta^\star(N)$, the defining equation for $\beta^\star(N)$ implies

$$\beta < \frac{\sigma_\eta^2\alpha^2}{y_N(\beta)} + \frac{\gamma\sigma_D^2\alpha^2(\sigma_\eta^2)^3}{y_N(\beta)^2\left[\frac{y_N(\beta)}{\gamma\sigma_\theta^{-2}} + \frac{\gamma\sigma_D^2 y_N(\beta)^2}{\sigma_\theta^{-4}}\right]}.$$

This inequality implies that $\mathcal{M}(n)$ is strictly decreasing in n when all banks choose the same hiring profile. At the same time, the inequality above also implies that the common hiring vector that assigns N/C workers to every cluster is not optimal. Hence the hypotheses of Lemma 3 are satisfied, and so there is no equilibrium in which all banks choose the same hiring strategy across clusters.

Lastly, we show that the interval $(\underline{\beta}(N), \overline{\beta}(N))$ is non-empty for sufficiently small N. As $N \downarrow 0$, we have

$$B(N) \to 1 + \sigma_\eta^2\sigma_\theta^{-2} \qquad \text{and} \qquad C(N) \sim 3\sigma_\eta^2 N^2.$$

Hence the discriminant is positive for all sufficiently small N, and

$$\underline{\beta}(N) \to \frac{\sigma_\eta^2\alpha^2}{1 + \sigma_\eta^2\sigma_\theta^{-2}}.$$

At the same time, the larger root in the definition of $\overline{\beta}(N)$ diverges to $+\infty$ as $N \downarrow 0$.

Also, $y_N(\beta) \to 1 + \sigma_\eta^2\sigma_\theta^{-2}$ uniformly on bounded sets, so the right hand side of the equation defining $\beta^\star(N)$ converges uniformly on bounded sets to

$$\frac{\sigma_\eta^2\alpha^2}{1 + \sigma_\eta^2\sigma_\theta^{-2}} + \frac{\gamma\sigma_D^2\alpha^2(\sigma_\eta^2)^3}{\left(1 + \sigma_\eta^2\sigma_\theta^{-2}\right)^2\left[\frac{1+\sigma_\eta^2\sigma_\theta^{-2}}{\gamma\sigma_\theta^{-2}} + \frac{\gamma\sigma_D^2\left(1+\sigma_\eta^2\sigma_\theta^{-2}\right)^2}{\sigma_\theta^{-4}}\right]},$$

which is strictly larger than

$$\frac{\sigma_\eta^2 \alpha^2}{1 + \sigma_\eta^2 \sigma_\theta^{-2}}.$$

Since the fixed point is unique, this implies that $\beta^\star(N)$ converges to a limit that is strictly larger than the limit of $\underline{\beta}(N)$.

Therefore, for all sufficiently small N, we have

$$\underline{\beta}(N) < \overline{\beta}(N).$$

□

IA.C.3 Model Predictions

For the empirical predictions, we first prove a lemma linking expertise with trading activity within banks across clusters, and across banks by cluster, which our empirical analysis of specialization is based on.

Lemma 4 (Expertise and Specialization)**.** *Observed positions measure specialization in two ways.*

1. *Consider a factor f belonging to cluster c, and factor g belonging to cluster d. Suppose $\tau_c^p = \tau_d^p$ and $\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f = \mathbb{E}_b[\theta_g \mid \mathcal{I}_b] - p_g \neq 0$. If*

   $$|z_{b,f}| > |z_{b,g}|,$$

   then

   $$x_{b,c} > x_{b,d}.$$

2. *Consider two banks b and b' satisfing $\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f = \mathbb{E}_{b'}[\theta_f \mid \mathcal{I}_{b'}] - p_f \neq 0$. If*

   $$|z_{b,f}| > |z_{b',f}|,$$

   then

   $$x_{b,c} > x_{b',c}.$$

Thus, in the ex-ante identical benchmark of the model, observed positions provide an empirical measure of specialization both within a bank across clusters and across banks within a cluster.

Proof. First, to prove part (1), the bank b's first order condition implies

$$z_{b,f} = \frac{\kappa_{b,c}}{\gamma} \left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f \right), \qquad z_{b,g} = \frac{\kappa_{b,d}}{\gamma} \left(\mathbb{E}_b[\theta_g \mid \mathcal{I}_b] - p_g \right).$$

Since the conditional expected payoff net of price is the same for the two factors and is nonzero, the inequality $|z_{b,f}| > |z_{b,g}|$ implies that

$$\kappa_{b,c} > \kappa_{b,d}.$$

Next, since

$$\kappa_{b,c} = \frac{T_{b,c}}{1 + \sigma_\eta^2 T_{b,c}},$$

and this expression is strictly increasing in $T_{b,c}$, we have

$$T_{b,c} > T_{b,d}.$$

Since $\tau_c^p = \tau_d^p$, it follows from

$$T_{b,c} = \sigma_\theta^{-2} + \tau_c^p + x_{b,c}, \qquad T_{b,d} = \sigma_\theta^{-2} + \tau_d^p + x_{b,d},$$

that

$$x_{b,c} > x_{b,d}.$$

Next, to prove part (2), the banks' first order conditions imply

$$z_{b,f} = \frac{\kappa_{b,c}}{\gamma} \left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f \right), \qquad z_{b',f} = \frac{\kappa_{b',c}}{\gamma} \left(\mathbb{E}_{b'}[\theta_f \mid \mathcal{I}_{b'}] - p_f \right).$$

Since the conditional expected payoff net of price is the same for the two banks and is nonzero, the inequality $|z_{b,f}| > |z_{b',f}|$ implies that

$$\kappa_{b,c} > \kappa_{b',c}.$$

Next, since

$$\kappa_{b,c} = \frac{T_{b,c}}{1 + \sigma_\eta^2 T_{b,c}},$$

and this expression is strictly increasing in $T_{b,c}$, we have

$$T_{b,c} > T_{b',c}.$$

Finally,

$$T_{b,c} = \sigma_\theta^{-2} + \tau_c^p + x_{b,c},$$

so $T_{b,c} > T_{b',c}$ implies that

$$x_{b,c} > x_{b',c}.$$

□

Proof of Prediction 1

Proof. This follows immediately from Proposition 2. □

Proof of Prediction 2

Proof. Formally, we show the following:

1. Consider a factor f belonging to cluster c, and a factor g belonging to cluster d. Suppose $\tau_c^p = \tau_d^p$ and

   $$\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f = \mathbb{E}_b[\theta_g \mid \mathcal{I}_b] - p_g \neq 0.$$

   If $|z_{b,f}| > |z_{b,g}|$, then the risk-adjusted return from factor f is larger than the risk-adjusted return from factor g.

2. Consider a factor f in cluster c, and suppose two banks b and b' satisfy

   $$\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f = \mathbb{E}_{b'}[\theta_f \mid \mathcal{I}_{b'}] - p_f \neq 0.$$

   If $|z_{b,f}| > |z_{b',f}|$, then the risk-adjusted return from factor f is larger for bank b than for bank b'.

First, define bank b's factor-level profit and loss as

$$\pi_{b,f} = z_{b,f}(\theta_f + \eta_f - p_f).$$

Conditional on bank b's information set, we can write

$$\mathbb{E}_b[\pi_{b,f} \mid \mathcal{I}_b] = z_{b,f}\left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right),$$

and

$$\mathrm{Var}_b(\pi_{b,f} \mid \mathcal{I}_b) = z_{b,f}^2\left(T_{b,c}^{-1} + \sigma_\eta^2\right).$$

Using b's first order condition,

$$z_{b,f} = \frac{\kappa_{b,c}}{\gamma}\left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right),$$

we therefore have

$$\mathbb{E}_b[\pi_{b,f} \mid \mathcal{I}_b] = \frac{\kappa_{b,c}}{\gamma}\left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right)^2,$$

and

$$\mathrm{Var}_b(\pi_{b,f} \mid \mathcal{I}_b) = \frac{\kappa_{b,c}^2}{\gamma^2} \left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right)^2 \left(T_{b,c}^{-1} + \sigma_\eta^2\right).$$

Since

$$\kappa_{b,c} = \frac{1}{T_{b,c}^{-1} + \sigma_\eta^2},$$

this simplifies to

$$\mathrm{Var}_b(\pi_{b,f} \mid \mathcal{I}_b) = \frac{\kappa_{b,c}}{\gamma^2} \left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right)^2.$$

Therefore bank b's risk-adjusted return from factor f is

$$\frac{\mathbb{E}_b[\pi_{b,f} \mid \mathcal{I}_b]}{\sqrt{\mathrm{Var}_b(\pi_{b,f} \mid \mathcal{I}_b)}} = \sqrt{\kappa_{b,c}} \left|\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right|.$$

The same expression holds for bank b'. Since the conditional expected payoff net of price is common across the relevant comparisons, the ranking of risk-adjusted returns is the same as the ranking of κ.

Next, by part (1) of Lemma 4, the within-bank inequality $|z_{b,f}| > |z_{b,g}|$ implies that $x_{b,c} > x_{b,d}$. Since $\tau_c^p = \tau_d^p$, we have $T_{b,c} > T_{b,d}$ and $\kappa_{b,c} > \kappa_{b,d}$. Therefore the risk-adjusted return from factor f is larger than the risk-adjusted return from factor g.

Lastly, by part (2) of Lemma 4, the cross-bank inequality $|z_{b,f}| > |z_{b',f}|$ implies that $x_{b,c} > x_{b',c}$. Therefore $T_{b,c} > T_{b',c}$ and $\kappa_{b,c} > \kappa_{b',c}$. Hence bank b's risk-adjusted return from factor f is strictly larger than bank b''s. □

Proof of Prediction 3

Proof. Formally, we show:

1. let factor f belong to cluster c, and let factor g belong to cluster d. Suppose $\tau_c^p = \tau_d^p$ and

   $$\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f = \mathbb{E}_b[\theta_g \mid \mathcal{I}_b] - p_g \neq 0.$$

   If $|z_{b,f}| > |z_{b,g}|$, then $\mathrm{Var}_b(\theta_f \mid \mathcal{I}_b) < \mathrm{Var}_b(\theta_g \mid \mathcal{I}_b)$.

2. suppose two banks b and b' satisfy

   $$\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f = \mathbb{E}_{b'}[\theta_f \mid \mathcal{I}_{b'}] - p_f \neq 0.$$

   If $|z_{b,f}| > |z_{b',f}|$, then $\mathrm{Var}_b(\theta_f \mid \mathcal{I}_b) < \mathrm{Var}_{b'}(\theta_f \mid \mathcal{I}_{b'})$.

First, by part (1) of Lemma 4, the within-bank inequality $|z_{b,f}| > |z_{b,g}|$ implies that

$x_{b,c} > x_{b,d}$.

The posterior variances of bank b about factors f and g are

$$\mathrm{Var}_b(\theta_f \mid \mathcal{I}_b) = \frac{1}{T_{b,c}}, \qquad \mathrm{Var}_b(\theta_g \mid \mathcal{I}_b) = \frac{1}{T_{b,d}}.$$

Since $\tau_c^p = \tau_d^p$ and $x_{b,c} > x_{b,d}$, we have $T_{b,c} > T_{b,d}$, and therefore $\mathrm{Var}_b(\theta_f \mid \mathcal{I}_b) < \mathrm{Var}_b(\theta_g \mid \mathcal{I}_b)$.

Next, by part (2) of Lemma 4, the cross-bank inequality $|z_{b,f}| > |z_{b',f}|$ implies that

$$x_{b,c} > x_{b',c}.$$

The posterior variances of banks b and b' about factor f are

$$\mathrm{Var}_b(\theta_f \mid \mathcal{I}_b) = \frac{1}{T_{b,c}}, \qquad \mathrm{Var}_{b'}(\theta_f \mid \mathcal{I}_{b'}) = \frac{1}{T_{b',c}}.$$

Since $x_{b,c} > x_{b',c}$, we again have $T_{b,c} > T_{b',c}$, and therefore $\mathrm{Var}_b(\theta_f \mid \mathcal{I}_b) < \mathrm{Var}_{b'}(\theta_f \mid \mathcal{I}_{b'})$. □

Proof of Prediction 4

Proof. Formally, we consider a hiring equilibrium, and consider a perturbation in which cluster c has customer-demand variance $\sigma^2_{D,c}$ and cluster d has customer-demand variance $\sigma^2_{D,d}$, while all other primitives are the same in the two clusters. Suppose also that $\Psi''_c(n) < 0$ and $\Psi''_d(n) < 0$ on the relevant range. If $\sigma^2_{D,c} > \sigma^2_{D,d}$, then any optimal hiring vector satisfies $n_{b,c} \geq n_{b,d}$.

First, writing the bank's utility from hiring in cluster c with cluster-specific demand variance gives

$$\Psi_c(n) = \frac{L\,x(n)}{2\gamma S_c\left[1+\sigma_\eta^2(S_c+x(n))\right]} + \frac{L\gamma\sigma^2_{D,c}}{2\hat{\kappa}_c^2}\left[\frac{S_c+x(n)}{1+\sigma_\eta^2(S_c+x(n))} - \frac{S_c}{1+\sigma_\eta^2 S_c}\right].$$

Differentiating with respect to n, we have

$$\Psi'_c(n) = \frac{L\,x'(n)}{2\left[1+\sigma_\eta^2(S_c+x(n))\right]^2}\left[\frac{1+\sigma_\eta^2 S_c}{\gamma S_c} + \frac{\gamma\sigma^2_{D,c}}{\hat{\kappa}_c^2}\right].$$

Therefore $\Psi'_c(n)$ is strictly increasing in $\sigma^2_{D,c}$ for every $n > 0$.

Next, consider two otherwise identical clusters c and d with $\sigma^2_{D,c} > \sigma^2_{D,d}$. Then $\Psi'_c(n) > \Psi'_d(n)$ for every $n > 0$. Suppose, towards a contradiction, that an optimal hiring vector satisfies $n_{b,c} < n_{b,d}$. Since $\Psi''_d(n) < 0$, the derivative $\Psi'_d(n)$ is decreasing in n, and so $\Psi'_d(n_{b,c}) \geq \Psi'_d(n_{b,d})$. Combining this with $\Psi'_c(n_{b,c}) > \Psi'_d(n_{b,c})$, we obtain $\Psi'_c(n_{b,c}) > \Psi'_d(n_{b,d})$. Therefore moving a small amount of hiring from cluster d to cluster c raises the objective, which contradicts optimality. Hence any optimal hiring vector must satisfy $n_{b,c} \geq n_{b,d}$.

Lastly, since $x(n)$ is strictly increasing in n, we have $x(n_{b,c}) \geq x(n_{b,d})$.

□

Proof of Prediction 5

Proof. Formally, we consider a bank b after hiring has been chosen, and signals and prices have been realized, which then solves its portfolio choice problem, but faces a binding risk-weighted position constraint

$$\sum_{c=1}^{C}\sum_{f\in c}\sqrt{T_{b,c}^{-1}+\sigma_\eta^2}\,|z_{b,f}| \le \bar{Z}.$$

Consider factor f belonging to cluster c, and factor g belonging to cluster d. Suppose $x_{b,c} > x_{b,d}$ and

$$|\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f| = |\mathbb{E}_b[\theta_g \mid \mathcal{I}_b] - p_g| > 0.$$

Then, we show that as $\bar{Z}$ is reduced, the position in factor g reaches zero before the position in factor f.

First, conditional on hiring and on the realized information set, the bank chooses positions to solve

$$\max_{\{z_{b,f}\}} \sum_{c=1}^{C}\sum_{f\in c} z_{b,f}\left(\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f\right) \;-\; \frac{\gamma}{2}\sum_{c=1}^{C}\sum_{f\in c}\left(T_{b,c}^{-1}+\sigma_\eta^2\right) z_{b,f}^2$$

subject to

$$\sum_{c=1}^{C}\sum_{f\in c}\sqrt{T_{b,c}^{-1}+\sigma_\eta^2}\,|z_{b,f}| \le \bar{Z}.$$

For each factor, the bank chooses the sign of $z_{b,f}$ to match the sign of $\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f$. Therefore it is enough to study position magnitudes.

Next, let $\nu \ge 0$ denote the multiplier on the position constraint. For an active factor f in cluster c, the first order condition is

$$|\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f| \;-\; \gamma\left(T_{b,c}^{-1}+\sigma_\eta^2\right)|z_{b,f}| \;-\; \nu\sqrt{T_{b,c}^{-1}+\sigma_\eta^2} = 0.$$

Hence

$$|z_{b,f}| = \frac{|\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f| \;-\; \nu\sqrt{T_{b,c}^{-1}+\sigma_\eta^2}}{\gamma\left(T_{b,c}^{-1}+\sigma_\eta^2\right)}$$

whenever the numerator is positive, and otherwise $|z_{b,f}| = 0$. The same expression holds for factor g in cluster d.

Then factor f remains active if and only if

$$\nu < \frac{|\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f|}{\sqrt{T_{b,c}^{-1}+\sigma_\eta^2}},$$

and factor g remains active if and only if

$$\nu < \frac{|\mathbb{E}_b[\theta_g \mid \mathcal{I}_b] - p_g|}{\sqrt{T_{b,d}^{-1} + \sigma_\eta^2}}.$$

Since $x_{b,c} > x_{b,d}$, we have $T_{b,c} > T_{b,d}$. Therefore

$$\sqrt{T_{b,c}^{-1} + \sigma_\eta^2} < \sqrt{T_{b,d}^{-1} + \sigma_\eta^2}.$$

Since the two factors have the same absolute conditional expected payoff net of price, it follows that

$$\frac{|\mathbb{E}_b[\theta_f \mid \mathcal{I}_b] - p_f|}{\sqrt{T_{b,c}^{-1} + \sigma_\eta^2}} > \frac{|\mathbb{E}_b[\theta_g \mid \mathcal{I}_b] - p_g|}{\sqrt{T_{b,d}^{-1} + \sigma_\eta^2}}.$$

Lastly, as the bank-level position constraint is tightened, the multiplier ν rises. The cutoff for factor g is therefore reached before the cutoff for factor f. Hence the position in the peripheral cluster d is set to zero before the position in the core cluster c. Therefore, tighter position constraints lead banks to cut peripheral positions before core positions. □

Proof of Prediction 6

Proof. Formally, we consider a hiring equilibrium, and consider a local perturbation in which expected demand in cluster c rises through an increase in $\sigma_{D,c}^2$. We consider two banks, b and b', such that $n_{b,c} > n_{b',c} > 0$, and assume that that $\Psi_c''(n) > 0$ on $[n_{b',c}, n_{b,c}]$. Then the increase in the marginal utility of hiring in cluster c is larger at bank b than at bank b'.

First, writing the bank's utility from hiring in cluster c with cluster-specific expected demand gives

$$\Psi_c'(n) = \frac{L\, x'(n)}{2\left[1 + \sigma_\eta^2(S_c + x(n))\right]^2}\left[\frac{1 + \sigma_\eta^2 S_c}{\gamma S_c} \; + \; \frac{\gamma \sigma_{D,c}^2}{\hat{\kappa}_c^2}\right].$$

Differentiating with respect to $\sigma_{D,c}^2$, we have

$$\frac{\partial \Psi_c'(n)}{\partial \sigma_{D,c}^2} = \frac{L\gamma}{2\hat{\kappa}_c^2}\frac{x'(n)}{\left[1 + \sigma_\eta^2(S_c + x(n))\right]^2}.$$

Next, differentiating this expression with respect to n, we obtain

$$\frac{\partial^2 \Psi_c'(n)}{\partial n\, \partial \sigma_{D,c}^2} = \frac{L\gamma}{2\hat{\kappa}_c^2}\frac{\mathcal{H}_c(n)}{\left[1 + \sigma_\eta^2(S_c + x(n))\right]^3}.$$

Since

$$\Psi_c''(n) = \frac{L}{2\left[1 + \sigma_\eta^2(S_c + x(n))\right]^3}\left[\frac{1 + \sigma_\eta^2 S_c}{\gamma S_c} \; + \; \frac{\gamma \sigma_{D,c}^2}{\hat{\kappa}_c^2}\right]\mathcal{H}_c(n),$$

the maintained condition $\Psi_c''(n) > 0$ on $[n_{b',c}, n_{b,c}]$ implies that $\mathcal{H}_c(n) > 0$ on that interval. Therefore

$$\frac{\partial^2 \Psi_c'(n)}{\partial n \, \partial \sigma_{D,c}^2} > 0 \qquad \text{for all } n \in [n_{b',c}, n_{b,c}].$$

It follows that

$$\frac{\partial \Psi_c'(n_{b,c})}{\partial \sigma_{D,c}^2} > \frac{\partial \Psi_c'(n_{b',c})}{\partial \sigma_{D,c}^2}.$$

□