# Financial Regulation and AI: A Faustian Bargain?

Christopher Clayton* Antonio Coppola†

October 2025

**Abstract**

We examine whether and how granular, high-dimensional predictive models should be integrated into central banks' macroprudential toolkit. First, we develop a tractable framework that formalizes the tradeoff regulators face when choosing between implementing models that forecast systemic risk accurately but have uncertain causal content and models with the opposite profile. We derive the regulator's optimal policy in an equilibrium setting in which private portfolios react endogenously to the regulator's model choice and policy rule. We show that even purely predictive models can generate welfare gains for a regulator, and that predictive precision and knowledge of causal impacts of policy interventions are complementary. Second, we introduce and train a deep learning architecture tailored to financial holdings data—a graph transformer—and we discuss why it is optimally suited to this problem. The model learns vector embedding representations for both assets and investors by explicitly modeling the relational structure of portfolio holdings. We show that this architecture attains state-of-the-art predictive accuracy in relevant out-of-sample forecasting tasks, including trading behavior prediction both in regular times and during crises.

**Keywords:** Fire Sales, Macroprudential Policy, Artificial Intelligence, Graph Neural Networks, Holdings Data, Embeddings, Stress Testing, Deep Learning, Information Design.

**JEL Codes:** C4, G1.

*Yale School of Management and NBER; christopher.clayton@yale.edu.

†Stanford University Graduate School of Business and CEPR; acoppola@stanford.edu.

We thank Adrien Auclert, Hui Chen, Paul Fontanier, Xavier Gabaix, Stefano Giglio, Tarek Hassan, Arvind Krishnamurthy, Hanno Lustig, Matteo Maggiori, Pablo Ottonello, Ashesh Rambachan, Ricardo Reis, Jesse Schreger, Jann Spiess, Stijn Van Nieuwerburgh, and Yuyan Wang for helpful comments. We thank Oliver Wang and Zhongyu Yin for outstanding research assistance. Financial support from the Stanford GSB Business, Government, and Society (BGS) Initiative is gratefully acknowledged. First draft: July 2025.

# 1 Introduction

A central concern of macroprudential policy is identifying and mitigating the amplification of shocks through the financial system. Traditional macroprudential frameworks focus on ex ante regulation of well-established sources of financial fragility such as leverage accumulation or maturity mismatches. Yet over the past two decades, the data environment for regulators has changed dramatically. Supervisory filings and rapid data collection now give regulators the ability to observe granular, high-frequency information on investor portfolios. At the same time, predictive technologies such as deep learning have achieved significant gains in out-of-sample performance across domains. These advances raise the question: can real-time, high-dimensional predictive models be used productively in financial regulation and interventions? On the one hand, predictive models can help effectively detect signals of fragility that are not captured by canonical macroprudential metrics: where exactly fire sales may erupt, how crowded trades may unwind, or which asset classes are most exposed to redemption risk. On the other hand, these models are often reduced-form and highly non-linear, with no guarantee that they might recover deep underlying structural forces that are invariant to the regulators' own use of the models.

This paper develops a theoretical and empirical framework to address this question. Our central object of analysis is the regulator's model choice and how it informs intervention. We ask whether regulators should deploy high-performing predictive models when they face uncertainty about the causal consequences of acting on the model's output. In this context, we show that the answer depends on the interaction between the predictive model's informational content as well as forecast accuracy, and the regulator's ability to target interventions and estimates of causal impact of those interventions. Empirically, we introduce and build new predictive architectures using deep learning models tailored to the relational structure of financial holdings data, and we show that these can achieve benchmark-setting predictive performance in the dimensions relevant to the macroprudential task, providing a blueprint for practical implementation by central banks and other regulators.

In our theory, we model a three-period economy with intermediaries and a government regulator. The model builds upon canonical fire sales environments. In the first period, financial intermediaries choose portfolios. In the second, intermediaries face constraints (e.g., pledgeability and collateral constraints) that force them to sell some assets prior to maturity. The key fire sale externalities arise because these assets are sold to second-best users whose productivity depends on the amount purchased. Finally, any assets held to maturity pay off in the third period.

The regulator can intervene in the second period to try to manage fire sales. We allow the regulator to employ a (potentially incomplete) set of liquidation wedges that affect intermediaries' choice of which assets to liquidate. Formally, liquidation wedges take the form of revenue-neutral taxes on selling an asset, although we later show that results extend to applying subsidies for retaining an asset. The key informational friction is that the regulator faces uncertainty over the mapping from intermediaries' asset positions and the policy intervention to realized liquidations and fire sale prices. The regulator has a prior over this mapping. Before undertaking the intervention,

the regulator can choose to deploy a model that delivers a signal about the latent fundamentals (e.g., predicted sales volumes or liquidation discounts). The regulator can then design its intervention based on the model used and the signal produced. Crucially, models can differ in what aspects of the system they inform. A model may provide strong predictive content (about liquidations and prices) but little information about the causal effects of interventions, and vice versa.

We characterize the regulator's optimal ex post policy rule as a function of its Bayesian posterior over key primitives. The regulator's optimal rule depends on the product between the predicted causal impact of the policy and the predicted social benefit of the policy. The social benefit of the policy, deriving from raising the fire sale price, is determined by the amount of each type of asset being sold. As a consequence, even a purely predictive model can improve welfare if its forecasts are aligned with dimensions of the system where the regulator has prior knowledge about the causal impact of the policy intervention. For example, if the regulator knows that its intervention will have a strong causal effect on fire sale prices in certain markets, then predictive models that are informative precisely about forced sales in those markets have particular potential to generate welfare gains. This suggests a complementarity between causal knowledge of policy interventions and predictive models that inform the social benefit of intervention.

We then characterize the expected welfare gains associated with the choice of a model by the regulator as well as optimal model choice. The welfare gains depend on the prior expectation of the size of the intervention, and also on the covariance matrix of the policy intervention (assessed from the prior perspective). For purely predictive models that are uninformative as to the causal structure, only this latter term depends on the choice of model. Hence choosing an optimal predictive model amounts to a choice of this covariance matrix. The intuition comes from the law of total variance. The regulator benefits from acquiring precise information on ex-post liquidations to better target its policies. However, not all predictive information is equally valuable. First, predictive precision is more valuable on margins where the regulator has prior knowledge that the policy intervention will have a strong causal impact, reinforcing the complementarity between causal knowledge and the value of predictive information. Second, the value of interventions (and hence of predictive precision) scales with the intervention's causal impact and with intermediaries' ex-post costs of portfolio adjustment. The regulator's optimal model choice thus tends to focus predictive precision on those margins with stronger predicted causal impacts and higher benefits of intervention. As a corollary to these results, we show that the predictive $R$-squared achieved by a predictive model in forecasting liquidations, weighted appropriately according to the value of the intervention for each asset, is a sufficient statistic for the welfare gain induced by the model.

To understand the implications of model deployment on intermediaries' decision making, we turn to the endogenous ex-ante portfolio response of private agents that anticipate the regulatory intervention. Intermediaries internalize the mapping from their portfolios to expected liquidation costs under the regulator's anticipated policy, but take as given both what model the regulator will adopt and the intervention and fire sale prices. As a result, the regulator's model choice affects

portfolio selection. We show that the regulator's model choice and intervention can both discourage intermediaries from holding assets associated with fire sales ex post, but also potentially lead to moral hazard. Intuitively, the ex post intervention has two competing effects on the initial asset allocation. First, intermediaries directly perceive assets on which the regulator will impose liquidation wedges as more expensive to hold, and so shift portfolios away from these assets. This can create a virtuous effect of the ex-post intervention, potentially substituting for the need to regulate these assets ex ante. However, the countervailing effect is that by using the more informed intervention to raise liquidation prices ex post, the regulator reduces fire sale discounts and so encourages intermediaries to hold more of these assets. This latter effect can be particularly pronounced if the regulator is not able to regulate all holders of certain assets. All of these effects, both virtuous and moral hazard, are limited by the predictability and granularity of the model—that is, the precision with which the regulator can target fire sales in real time.

Further, we explicitly model a regulator's optimal ex-ante macroprudential intervention in the portfolio holdings of intermediaries. We allow the regulator to employ a set of wedges on asset holdings, modeled as revenue-neutral taxes. The regulator's optimal tax ex ante starts from the baseline regulation it would apply if it did not intervene ex post, and then makes three adjustments reflecting model choice and the ex post intervention. First, the regulator discourages holdings of assets that make it more costly to acquire precise predictive information ex post. Second, the regulator encourages relative holdings of assets that will cause it to increase the size of its ex-post intervention. Intuitively, the regulator recognizes for these assets that the larger ex-post regulation serves as a substitute for the ex-ante intervention, so the need to intervene ex ante is muted. Finally, the regulator reduces regulation of assets that are expected to be subject to high ex post taxes, since the ex post taxes help discourage ex ante purchases of those assets. Conversely, the regulator increases regulation of assets for which moral hazard is induced by ex post bolstering of asset prices.

Although we illustrate our key insights for policy interventions in the form of liquidation wedges, in practice ex post interventions often resemble "bailout" measures with a subsidy component. We accommodate this policy in an extension in which we assume the regulator ex post imposes an asset retention wedge, formally modeled as a revenue-neutral subsidy for holding an asset to maturity. These subsidies imply a de facto tax on liquidating an asset, and accordingly the optimal intervention and model design remain the same. However, these subsidies introduce an additional moral hazard channel, since assets that are subsidized ex post become more attractive to purchase ex ante. Because this moral hazard effect depends on the expected subsidy size, models that increase predictive precision but do not change the expected tax rate are not subject to it (akin to Laffont and Tirole 1986). Since in our setup purely predictive models increase precision without changing the expected intervention size, the moral hazard they generate interestingly does not differ relative to the case with liquidation taxes instead of asset retention subsidies.

To empirically assess whether predictive models can in fact deliver useful signals for these macroprudential purposes, we develop a deep learning architecture tailored to the structure of financial

holdings data. Much of the modern deep learning toolkit—including models that have been applied in asset pricing contexts—is tailored toward grid and sequence inputs such as images and text. Yet in domains where the data has important graph structure, architectures tailored specifically to learn from graph structures have achieved breakthrough performance, including for instance when applied to problems of protein folding prediction and drug discovery (Jumper et al. 2021). In our setting, the data naturally form a graph: investors are connected to assets through their positions. We therefore design our empirical deep learning models to explicitly capture and make use of this crucial dimension of the data.

To exploit the relational structure of holdings data, we implement and train a *graph transformer* model, a form of graph neural network (GNN) augmented with an attention mechanism, to learn latent representations of both investors and assets (i.e., investor and asset embeddings into latent vector spaces). The architecture features two properties that are essential for this setting. First, it is permutation-invariant: predictions do not depend on the arbitrary ordering of investors or assets. This is a key distinction with sequence-based architectures (such as sequence transformers which underpin modern large language models) commonly used for text, in which the order of words in a document is essential to their contextual meaning. Second, the graph transformer model is inductive: all model parameters are fully shared across nodes, allowing the model to generalize to new investors or assets without retraining, while at the same time enforcing a strong form of regularization that leads to good generalization out of the model's training sample. The model is optimally sample-efficient in the sense of requiring minimal degrees of freedom to learn functions of the holdings data that are themselves invariant to arbitrary relabelings of investors and assets—a restriction that reflects the economics of the problem. Intuitively, this occurs because graph-based architectures do not need to use parameters to relearn permutation invariance from the data, but rather they enforce it explicitly through their structure.

Our empirical application is in the context of non-bank financial intermediation, which is an increasingly important source of financial fragility, partly as a result of regulatory reform following the Great Financial Crisis.[1] Using granular data on the securities holdings of non-bank financial institutions, we train the model jointly on two tasks: a masked autoencoder task, in which the model learns to reconstruct partially masked holdings, and a supervised prediction task, in which the model forecasts the cross-sectional pattern of future trades in the data. Both tasks are designed to inform the model about latent economic relationships governing the portfolio choice and rebalancing of these financial intermediaries. The training generates embeddings for both assets and investors which are general-purpose and can be used for multiple objectives. We do not train the model explicitly to predict fire sales in a supervised fashion, but rather turn to models which can learn general-purpose embeddings, precisely to avoid in-sample over-fitting and thus avoid a sharp degradation in out-of-sample performance. The quarterly holdings data are from Factset and cover a broad range of

[1] In particular, the Dodd-Frank Act of 2010, including its Volcker Rule provision, led to a migration of risk away from the banking system and towards other, less levered intermediaries.

institutional investors and assets.

We find that the model performs well on both tasks. It accurately reconstructs positions data with correlations exceeding 90% between predictions and targets, and it achieves strong out-of-sample predictive accuracy of nearly 30% in forecasting trade patterns. Notably, the model's performance is indeed stable out of sample, due to its inductive structure and shared-parameter design. The high holdings reconstruction fidelity on the autoencoder task should naturally be interpreted in light of the model's parameter-to-data ratio: with roughly 3.6 million parameters—representing less than 1% of the possible asset-investor pairs in the holdings data—the model extracts economically relevant patterns from high-dimensional data. Further, we show that the model's performance in forecasting trading behavior (our second task) remains high even when restricting the sample to the set of active investment managers only and to stress periods, both of which are especially relevant to the macroprudential application we are studying. For example, we show that the model is able to forecast the pattern of asset trades during the market crash of 2020 induced by the Covid pandemic with good accuracy, having been trained only on pre-2020 data. Importantly, we also show that the representations learned by the model explain the cross-section of liquidation prices in market stress episodes, despite not having been trained on price or returns data.

The predictive success of our deep learning architecture provides empirical support for the potential of model-informed ex post regulation. Moreover, we view our model as a blueprint for real-time regulatory approaches: a framework to transform high-frequency, granular data into relatively low-dimensional representations suitable for intervention design. Although the model is not structurally interpretable in the traditional sense, our theory shows that it can nonetheless play a useful role when paired with prior knowledge of structural effects. Going beyond the closed-economy macroprudential context studied here, we also see the theoretical and empirical approaches in this paper as similarly applicable in other macro-regulatory environments, such as in the setting of capital controls in open economies.

**Related Literature.** At their heart, our questions are situated within a longstanding intellectual debate concerning whether models that are primarily predictive in nature should be used to inform macroeconomic policy and financial regulation. Koopmans (1947), representative of the viewpoints then prevalent at the Cowles Commission, introduced his landmark "measurement without theory" critique of the earlier empirical studies of business cycles by Burns and Mitchell (1946). These early debates set the stage for the methodological ideas of Haavelmo (1944) and Friedman (1953), and for Lucas's (1976) eventual critique of policy evaluation carried out without individual-optimization microfoundations. While the qualitative outline of this debate has remained the same, its quantitative content has evolved, as the performance and generalization ability of predictive models has increased sharply. In this context, we show that sufficiently capable predictive models can have a role in the macro-regulatory toolbox, as a complement (rather than a substitute) to traditional structural approaches.

Our analysis relates to four broad areas of the literature. First, we connect to the literature on fire sales, and macroprudential regulation and ex-post interventions. Theoretical contributions include Bernanke and Gertler (1986), Williamson (1988), Kiyotaki and Moore (1997), Caballero and Krishnamurthy (2001), Lorenzoni (2008), Bianchi (2011, 2016), Stein (2012), Farhi and Werning (2016), Chari and Kehoe (2016), Bianchi and Mendoza (2018), Dávila and Korinek (2017), and Clayton and Schaab (2022, 2025). On the empirical side, prior work has focused on particular variables to explain and predict the occurrence and magnitude of fire sales, such as leverage and maturity mismatches (Fisher 1933; Kindleberger and Aliber 1978; Brunnermeier and Oehmke 2013; Schularick and Taylor 2012; Adrian and Shin 2014; Tobias and Brunnermeier 2016; Acharya, Pedersen, Philippon and Richardson 2017; Krishnamurthy and Muir 2025) as well as investor composition (Brainard and Tobin 1968; Coval and Stafford 2007; Haddad, Moreira and Muir 2021; Coppola 2025; Fang, Hardy and Lewis 2025): the empirical contribution of this paper asks whether there is additional value from a more agnostic approach that does not impose an ex-ante focus on particular dimensions of the data.

Second, we connect to the literature on applications of machine learning and deep learning to finance. Gu, Kelly and Xiu (2020, 2021), Kozak, Nagel and Santosh (2020), Nagel (2021), and Bryzgalova, DeMiguel, Li and Pelger (2023) apply machine learning techniques to the canonical empirical asset pricing problem of valuation in the cross-section of assets. Similarly, Chen, Pelger and Zhu (2024) introduce a deep learning architecture for modeling stock returns. Giglio, Kelly and Xiu (2022) review the intersection of empirical asset pricing and machine learning. Gabaix, Koijen, Richmond and Yogo (2024) use holdings data together with a variety of deep learning sequence models (e.g., Word2Vec and BERT) and recommender systems to estimate asset embeddings, and Gabaix, Koijen, Richmond and Yogo (2025) apply these to explain variation and volatility in credit spreads. Dolphin, Smyth and Dong (2022) and Sarkar (2025) also construct asset embeddings, respectively from return data and textual sources. Methodologically, our paper employs graph neural network architectures (Scarselli, Gori, Tsoi, Hagenbuchner and Monfardini 2008; Hamilton, Ying and Leskovec 2017; Xu, Hu, Leskovec and Jegelka 2018; Wu, Pan, Chen, Long, Zhang and Yu 2020) to learn representations from the relational structure of asset holdings data.[2]

Third, we relate to literature studying the deployment of machine learning and large-scale data for economic analysis and policy problems more broadly, including Einav and Levin (2014b,a), Kleinberg, Ludwig, Mullainathan and Obermeyer (2015), Athey (2017, 2018), Mullainathan and Spiess (2017), Kleinberg, Lakkaraju, Leskovec, Ludwig and Mullainathan (2018), Athey and Imbens (2019), Gillis and Spiess (2019), Farboodi and Veldkamp (2020), and Athey and Wager (2021). Relatedly, a recent literature in macroeconomics has highlighted how empirically identified causal effect estimates can be used to inform policy (Nakamura and Steinsson 2018, McKay and Wolf 2023). The question of how non-policy invariant relationships should be exploited by regulators

[2]See also Elliott, Golub and Jackson (2014) and Acemoglu, Ozdaglar and Tahbaz-Salehi (2015) for theory highlighting the importance of the network structure of positions for financial stability.

also has earlier conceptual parallels, for instance by Barro and Gordon (1983) in the context of the Phillips curve. Lastly, the model and information design aspect of our theoretical approach connects to work on stress testing (Shapiro and Skeie 2015; Faria-e Castro, Martinez and Philippon 2017; Goldstein and Leitner 2018; Leitner and Williams 2023; Orlov, Zryumov and Skrzypacz 2023; Parlatore and Philippon 2025).

# 2 A Framework for Regulatory Model Choice

There are $N$ assets. There are $I$ intermediary types (each of equal measure), with a representative intermediary of each type $i$. There is also a representative arbitrageur. The model has a Beginning-Middle-End structure. In the Beginning, initial asset positions are undertaken. In the Middle, intermediaries may be forced to sell assets prior to maturity to arbitrageurs, who are second best users. In the End, payoffs are distributed and consumption occurs.

**Intermediaries.** Intermediaries are risk neutral. In the Beginning, intermediary $i$ invests in a vector $q_i = (q_{i1}, \ldots, q_{iN})^T$ of assets. If $q_{in} < 0$, then intermediary $i$ has undertaken a negative investment (shorting) asset $n$.[3] The cost to intermediary $i$ of producing the asset vector $q_i$ is $C_i(q_i) = p_i^T q_i - \frac{1}{2} q_i^T H_i^q q_i$, where $p_{in}$ is the per-unit cost to $i$ of producing asset $n$ and where $p_i = (p_{i1}, \ldots, p_{iN})^T$. The cost component $q_i^T H_i^q q_i$ is a quadratic adjustment/holding cost, where $H_i^q$ is an $N \times N$ matrix. If held to maturity, intermediary $i$'s holdings of asset $n$ will produce a per-unit return $R_{in}$ in the End, with $R_i = (R_{i1}, \ldots, R_{iN})^T$.

In the Middle, intermediary $i$ can sell assets prior to maturity. We denote $\ell_{in} \geq 0$ to be sales by intermediary $i$ at endogenous price $\gamma_n$, with $\ell_i = (\ell_{i1}, \ldots, \ell_{iN})^T$ and $\gamma = (\gamma_1, \ldots, \gamma_N)^T$. Intermediary $i$ faces an adjustment cost $\frac{1}{2} \ell_i^T H_i^\ell \ell_i$ on asset sales, where $H_i^\ell$ is $N \times N$. Assets will be sold at a discount on their fundamental value ($\gamma < R_i$, see below), and intermediary $i$ faces a set of "rollover constraints" that force asset sales. This set of $M$ constraints is given by

$$A_i^q q_i + \rho_i \leq A_i^\ell \ell_i, \tag{1}$$

where $A_i^q, A_i^\ell$ are $M \times N$ and $\rho_i$ is $M \times 1$. For example, equation 1 can capture constraints on asset positions (e.g., a limit on debt) or a requirement to raise funds based on asset holdings (e.g., a cost of maintaining the project).[4]

In the End, intermediary $i$ realizes payoff on assets held to maturity and consumes. Intermediary

[3] We could endow intermediaries with a stock of assets, but for expositional convenience we set that stock to 0.

[4] We simplify analysis by not having equation 1 depend on the liquidation price $\gamma$, as for example in price-dependent collateral constraints. This enables us to maintain a linear-quadratic structure throughout the paper. It is straightforward to extend analysis to include prices in constraints, but the characterization of the ex-ante optimal portfolio would no longer admit a closed-form solution.

$i$'s total payoff (consumption) in the End, inclusive of adjustment costs, is

$$U_i = q_i^T(R_i - p_i) - \ell_i^T(R_i - \gamma) - \frac{1}{2}q_i^T H_i^q q_i - \frac{1}{2}\ell_i^T H_i^\ell \ell_i. \tag{2}$$

**Arbitrageurs.** A representative arbitrageur is a second-best user of intermediary assets. If the arbitrageur purchases a vector $L = (L_1, \ldots, L_N)^T$ of intermediary assets in the Middle, the arbitrageur can use them in a production technology to produce $\mathcal{F}(L) = L^T\overline{\gamma} - \frac{1}{2}L^T\Gamma L$ units of the consumption good in the End, where $\overline{\gamma}$ is $N \times 1$ and $\Gamma$ is $N \times N$. The representative arbitrageur's payoff is

$$U_A = L^T(\overline{\gamma} - \gamma) - \frac{1}{2}L^T\Gamma L. \tag{3}$$

**Information Structure and Timing.** Although all model parameters are determined in the Beginning, all agents in the Beginning are uncertain about the true values of the parameters $\Phi = \{A_i^q, A_i^\ell, H_i^\ell, \rho_i, R_i, \overline{\gamma}\}$. That is, agents are uncertain as to the true parameters underlying the rollover constraint (equation 1), the asset return $R_i$, the arbitrageurs' baseline productivity $\overline{\gamma}$, and the liquidation adjustment cost $H_i^\ell$. All agents have a common prior $\Phi \sim \mu_0$ in the Beginning. In the Middle before asset sales and purchases are chosen, the true model parameters become common knowledge of private agents. After becoming common knowledge, intermediaries choose asset liquidations and arbitrageurs choose asset purchases.[5]

**Market Clearing.** Markets must clear for liquidations in the Middle, that is

$$L = \sum_i \ell_i. \tag{4}$$

**Competitive Equilibrium.** We define a competitive equilibrium as follows.

**Definition 1.** *A competitive equilibrium of the model, given true model parameters $\Phi$, is a vector of prices $\gamma$ and a set of allocations $\{q_i, \ell_i, L\}$ such that:*

1. *In the Middle, taking as given asset allocations $\{q_i\}$ and model parameters $\Phi$:*
   (a) *Intermediary $i$ chooses $\ell_i$ to maximize utility (equation 2) subject to the rollover constraint (1), taking prices $\gamma$ as given.*

[5] Assuming that private agents and the regulator (see below) have a common prior in the Beginning simplifies analysis because it prevents the regulator from learning information about these parameters from inference about private sector beliefs based on the asset allocation. It also eliminates a regulatory incentive based on different beliefs the regulator and agents (e.g., Fontanier 2025.)

Assuming $\Gamma$ is known to all agents ex ante simplifies analysis by eliminating the ability of a regulator to learn about the price impact of liquidations from observing a signal of $\gamma$ or $\ell$. This will allow us to fully separate predictive and causal channels in our framework. Absent this assumption, predictive models that inform a regulator about $\gamma$ and $\ell$ might have even more value ex post because they would also allow the regulator to learn about these structural parameters that determine the causal impact of a policy intervention (even if very little information is learned).

*(b) Arbitrageurs choose L maximize utility (equation 3), taking prices $\gamma$ as given.*

*(c) The liquidation markets clear (equation 4).*

*2. In the Beginning:*

*(a) Intermediary $i$ chooses $q_i$ to maximize expected utility $\mathbb{E}_0[U_i]$, where $\mathbb{E}_0$ denotes the expectation given the prior $\mu_0$ over model parameters $\Phi$.*

## 2.1 Regulator's Model Design and Intervention in the Middle

In the Middle, a regulator is able to intervene in order to try to manage the fire sale price impact of liquidations. Formally, the regulator can impose a vector $\tau_i = (\tau_{i1}, \ldots, \tau_{iN})^T$ of revenue-neutral liquidation wedges on each intermediary $i$, which alter the intermediary's perceived price for selling the asset.[6] $\tau_{in}$ represents a tax on selling asset $n$, with $\tau_{in} < 0$ being a subsidy for sale. As a result, intermediary $i$'s payoff in the End is modified to be

$$U_i - (\ell_i^T - \ell_i^{*T})\tau_i \tag{5}$$

where $\ell_i^{*T}\tau_i$ is equilibrium revenue remissions based on the equilibrium asset liquidations $\ell_i^*$ of intermediary $i$. Intermediaries take revenue remissions as given. We allow for the possibility that the regulator has potentially incomplete instruments, represented by a regulatory cost $\delta(\tau) = \frac{1}{2}\tau^T\Delta\tau$, where $\Delta$ is $NI \times NI$.[7]

Unlike private agents, the regulator does not learn the true model parameters $\Phi$. Instead, the regulator has to use a *model* to make an inference about the true parameters. The regulator's *model* $M \in \mathcal{M}$ formally is a process of drawing a signal $s$ of the parameters. The signal updates the regulator's posterior distribution to $\Phi \sim \mu_p|s, M$. The regulator must choose the wedges $\tau$ before private agents move, and so the regulator cannot obtain any new information from observing the market before setting regulation (apart from the signal drawn).

To simplify exposition, we introduce an assumption of matrix symmetry that we maintain throughout the paper.

**Assumption 1.** *The matrices $H_i^q, H_i^\ell, \Gamma, \Delta$ are symmetric.*

**Impact of Regulatory Intervention.** To solve the regulator's optimum, we begin by characterizing the impact of the regulator's wedges $\tau = (\tau_1^T, \ldots, \tau_I^T)^T$ on the equilibrium in the Middle. Our model is substantially simplified by the information structure: because private agents directly

[6] In Section 2.7, we instead assume the regulator must use asset holding subsidies rather than liquidation taxes. The results in the Middle are identical, but the asset holding subsidies introduce additional moral hazard in the Beginning.

[7] One could instead model incompleteness as a restriction $\Delta(\tau) \leq 0$ in which case the regulator's Lagrangian will be $\mathbb{E}[\sum_i U_i|\Gamma, M] - \lambda\Delta(\tau)$ where $\lambda$ is the Lagrange multiplier. This is akin to the reduced-form cost $\delta(\tau) = \lambda\Delta(\tau)$ except that $\lambda$ is endogenous.

observe $\Phi$ in the Middle, for a given vector of asset allocations $q = (q_1^T, \ldots, q_I^T)^T$ the equilibrium in the Middle does not depend on the regulator's choice of model $M$ or the realized signal $s$ except through the choice of wedges $\tau$.

The linear-quadratic structure of preferences and the rollover constraint leads to a characterization of the equilibrium as a simple linear system of equations. The following Lemma characterizes this equilibrium.

**Lemma 1.** *Given asset allocations $q$, true parameters $\Phi$, and regulatory wedges $\tau$, the equilibrium in the Middle is given by*

$$\ell_i = \overline{\ell}_i + \Lambda_i^q q_i - \Lambda_i^\tau \tau_i - \sum_{j=1}^{N} \left[ \Lambda_j^{q,e} q_j^* - \Lambda_j^{\tau,e} \tau_j^* \right] \tag{6}$$

$$\gamma = \overline{\gamma} - \Gamma \sum_i \ell_i \tag{7}$$

*where $\overline{\ell}_i$ $(N \times 1)$ and $\Lambda_i^q, \Lambda_i^\tau, \Lambda_i^{q,e}, \Lambda_i^{\tau,e}$ $(N \times N)$ are defined in the proof.*

The characterization of the equilibrium in Lemma 1 is intuitive. Liquidations start from a benchmark $\overline{\ell}_i$ and increase (for positive $\Lambda_i^q$) in asset holdings while decreasing (for positive $\Lambda_i^\tau$) in the liquidation wedge. Higher liquidations result in a lower liquidation price (for positive $\Gamma$) due to more assets having to be absorbed by arbitrageurs (equation 7). These means that liquidations by intermediary $i$ decrease in asset holdings and increase in the liquidation wedges applied to other intermediaries within the same sector and across different sectors (for positive matrix elements), a result of a substitution effect resulting from the increase in equilibrium price. In a model without fire sales, we have $\Lambda_j^{q,e} = \Lambda_j^{\tau,e} = 0$. We distinguish in equation 6 between the asset choices of intermediary $i$ and the wedges applied to intermediary $i$'s liquidations (denoted with no $*$) as opposed to equilibrium objects (denoted with $*$) that enter because they determine the equilibrium liquidation price.

Crucially, under the maintained information structure that private agents have a finer information partition than the regulator, the regulator's model choice can alter the endogenous portfolios $q_i$ that feed into the determination of the equilibrium liquidations and liquidation prices through choice of the regulatory wedges, but not by changing the structural loadings (or agents' beliefs about them) $\Lambda_i = \{\Lambda_i^q, \Lambda_i^\tau, \Lambda_i^{q,e}, \Lambda_i^{\tau,e}\}$ that translate portfolio quantities and wedges into outcomes. In this sense, the equilibrium determination and the regulator's forecasting problem satisfy a key form of policy invariance.

## 2.2 Regulator's Optimal Wedges

We solve the regulator's problem by backward induction: first, we characterize the regulator's optimal intervention $\tau$ given a choice of model $M$ and signal $s$. We then characterize the regulator's

optimal choice of model $M$. To simplify analysis, we assume that the regulator places equal weight on all intermediaries but places a welfare weight of zero on arbitrageurs.[8]

Given that liquidation wedges are revenue-neutral, the regulator's optimal choice of $\tau$ solves

$$\max_{\tau} \mathbb{E}\left[\sum_i U_i \,|\, s, M\right] - \delta(\tau),$$

subject to equilibrium determination (Lemma 1).

As preliminaries to the proposition below, we define $\overline{\ell}_i(q) = \overline{\ell}_i + \Lambda_i^q q_i - \sum_i \Lambda_i^{q,e} q_i$ to be the liquidations of intermediary $i$ if there is no regulatory intervention ($\tau = 0$). We define $L(q) = \sum_i \overline{\ell}_i(q)$ to be the $N \times 1$ vector capturing total liquidations of each asset if there is no intervention. The following proposition characterizes optimal liquidation wedges in terms of these objects and model parameters.

**Proposition 1.** *Given a model $M$ and signal $s$, the regulator's optimal policy in the Middle is*

$$\tau^* = \mathbb{E}\left[\Xi \middle| s, M\right]^{-1} \mathbb{E}\left[(\sum_i \overline{\Lambda}_i^{\tau})^T \Gamma L(q) \middle| s, M\right] \tag{8}$$

*where $\Xi = \overline{\Lambda}^{\tau T} + (\sum_i \overline{\Lambda}_i^{\tau})^T \Gamma (\sum_i \overline{\Lambda}_i^{\tau}) + \Delta$, where $\overline{\Lambda}_i^{\tau} = (e_i \otimes I_N)^T \Lambda_i^{\tau} - (\Lambda_1^{\tau,e}, \ldots, \Lambda_I^{\tau,e})$ is $N \times NI$ (where $e_i$ is the standard basis vector whose $i^{th}$ element is 1 and $\otimes$ is the Kronecker product), and where $\overline{\Lambda}^{\tau} = (\overline{\Lambda}_1^{\tau T}, \ldots, \overline{\Lambda}_I^{\tau T})^T$ is $NI \times NI$.*

The optimal wedges of Proposition 1 are familiar from the macroprudential policy literature on fire sales, and intuitively encode an expected marginal cost-marginal benefit trade-off. The marginal cost of the policy is captured in the (conditional expectation of the) inverse matrix $\Xi$, while the marginal benefit is captured in the expectation.

There are three components to the private and regulatory marginal cost of liquidations, reflecting the distortion of intermediaries' activities away from their private optimum. First, increasing liquidation wedges directly distorts the intermediaries' activities (the first term of $\Xi$). Second, by using wedges to alter equilibrium prices, the regulator changes the incentives for intermediaries to sell different assets (the second term). Finally, there is the regulatory cost $\Delta$.

The social marginal benefit of regulation arises from the mitigation of the fire sale. This has two components that are central to our analysis. First is the causal effect of the policy intervention on equilibrium liquidation prices, captured by the term $(\sum_i \overline{\Lambda}_i^{\tau})^T \Gamma$. The utility consequence of these price changes is proportional to the total value of assets being sold by intermediaries, $L(q)$.

[8] This focuses attention on a distributive externality from shifting wealth between arbitrageurs and intermediaries. This can be incorporated by assuming that arbitrageurs have a high marginal value of wealth in the Beginning but cannot borrow, meaning that Pareto improvements are achieved by raising the liquidation price in the Middle and having a lump sum transfer from intermediaries to arbitrageurs in the Beginning (see e.g., Clayton and Schaab 2025).

Importantly, this means the social benefit of regulation derives from a product of the causal impact of the policy intervention, $(\sum_i \overline{\Lambda}_i^\tau)^T \Gamma^T$, and the social value of changing the price, here given by $L(q)$. This means that in designing regulation, there is value not only to identifying causal policy impacts, but also to predicting the magnitude of liquidations $L(q)$. To fully disentangle these two mechanisms, we can assume independence of the predictive and causal components of the system.[9]

**Definition 2** (Predictive-Causal Independence)**.** *We have predictive-causal independence if* $\{\Lambda_i^\tau\}$ *and* $\{\overline{\ell}_i, \Lambda_i^q, \Lambda_i^{q,e}\}$ *are independent of one another.*

Under predictive-causal independence, we obtain the following trivial corollary to Proposition 1

**Corollary 1.** *Under predictive-causal independence, the regulator's optimal wedges are*

$$\tau^* = \mathbb{E}\left[\Xi \middle| s, M\right]^{-1} \underbrace{\mathbb{E}\left[(\sum_i \overline{\Lambda}_i^\tau)^T \Gamma \middle| s, M\right]}_{\text{Causal: Policy Impact on Liquidation Price}} \overbrace{\mathbb{E}\left[L(q) \middle| s, M\right]}^{\text{Predictive: Total Liquidations}} \tag{9}$$

An important implication of Proposition 1 and Corollary 1 is that a model that is purely predictive—that is, that can predict liquidations $L(q)$—can be valuable to a regulator, provided that the regulator has some prior or posterior knowledge over the causal impact of the policy. This causal component of the optimal policy rule could be obtained through a fully-specified equilibrium structural model (e.g., as reviewed in Christiano et al. 2005) or, under certain conditions, from causal effects that are estimated empirically in reduced form (Guren et al. 2021; McKay and Wolf 2023; Caravello et al. 2025). The predictive information informs the regulator as to the magnitude and margins for intervention by informing the regulator about the social benefit of the intervention. Corollary 1 implies that, holding fixed the causal impact of the policy, the regulator would design larger-magnitude interventions when the regulator's model predicted that total liquidations would be larger.

## 2.3 Expected Welfare from a Model and Optimal Model Choice

We next ask how a regulator in the Middle would choose a model $M \in \mathcal{M}$ under discretion, taking as given the asset holdings $q$ of intermediaries. Because choice of model in turn impacts intermediaries' optimal portfolio allocations (see Section 2.5), we later consider model choice under commitment.

We begin by characterizing the expected utility to the regulator from a choice $M$ of model, and then characterize optimal model choice. As a preliminary to the proposition below, we define $\theta_i(q) = R_i - (\overline{\gamma} - \Gamma L(q))$ to be the fire sale losses to intermediary $i$ from selling assets prior to

[9] Naturally absent this assumption, we could perform an analogous decomposition to Corollary 1, but also include the covariance between the two terms.

maturity in the event of no regulatory intervention. We define $\hat{L}(q) = \mathbb{E}[L(q)|s, M]$ to be predicted total liquidations, given a model $M$ and signal $s$. For the presentation in the main text, we focus on the case assuming predictive-causal independence (Definition 2). The proof of Proposition 2 characterizes the general case.

**Proposition 2.** *Under predictive-causal independence, the regulator's ex-ante expected welfare given positions $q$ and a model $M$ is*

$$\begin{aligned} V(q, M) =& \mathbb{E}_0\left[q^T(R-p) - \ell(q)^T\theta(q) - \frac{1}{2}q^T H^q q - \frac{1}{2}\ell(q)^T H^\ell \ell(q)\right] \\ &+ \frac{1}{2}L_0(q)^T\Theta_0 L_0(q) + \frac{1}{2}\text{tr}\left(\Theta_0\, cov_0(\hat{L}(q))\right), \end{aligned} \tag{10}$$

*where* $\Theta_0 = \mathbb{E}_0[\Upsilon^T\Psi\Upsilon|M]$, *where* $\Psi = \mathbb{E}\left[2(\sum_i \overline{\Lambda}_i^\tau)^T\Gamma(\sum_i \overline{\Lambda}_i^\tau) + \Delta + \overline{\Lambda}^{\tau T} H^\ell \overline{\Lambda}^\tau \middle| s, M\right]$, *and where* $\Upsilon = \mathbb{E}\left[\Xi \middle| s, M\right]^{-1}\mathbb{E}\left[(\sum_i \overline{\Lambda}_i^\tau)^T\Gamma \middle| s, M\right]$.

Proposition 2 shows that the regulator's value function over assets $q$ and the model $M$ depends on two sets of terms.

The first line is the regulator's baseline welfare in the absence of intervention. Baseline welfare starts from the return on assets, $q^T(R-p)$, net of losses on assets sold prior to maturity, $-\ell(q)^T\theta(q)$. It then nets out the adjustment costs from the initial portfolio and from liquidations in the Middle. All of these terms are evaluated assuming no intervention, that is $\tau = 0$. As a result, they do not depend on the regulator's choice of model $M$.

The second line is the welfare gains resulting from the regulator's optimally chosen intervention, which comprises two terms. The first term reflects the expected magnitude and impact of the regulator's intervention. From Proposition 1 and Corollary 1, the intervention is determined by the predicted liquidations $L(q)$ and the causal impact of the policy intervention. Because interventions are targeted to the social benefit of intervening (equation 8), this term is therefore quadratic in the expected intervention and hence in ex-ante predicted liquidations $L_0(q)$. It is weighted by the consequences of intervention $\Theta_0$, reflected by the extent to which interventions move prices (the first term in $\Psi_0$), the regulatory costs (the second term), and the movement in holding costs from liquidations.

The welfare also depends on the accuracy of the regulator's model and intervention, captured in the second term that depends on the covariance matrix of predicted liquidations, $cov_0(\hat{L}(q))$. The intuition comes from the law of total variance: a perfectly informed regulator would get an exact prediction of liquidations from observing the true parameters of the system, $\hat{L}(q) = L(q)$. A regulator that learned no information relative to the prior would have no covariance, so that all benefits would be reflected in the prior beliefs that are used to design the intervention.

In general, both of these terms depend on the choice of model $M$. The latter term depends on the choice of model because the choice of model determines the covariance matrix of the policy

intervention. The former term also depends on the model, because the anticipated choice of model can affect the expected intervention. In particular, even under predictive-causal independence, from Proposition 1 a model that informs about the causal impact of the policy intervention $\tau$ will inform about both the costs $\Xi$ of regulation and also about the causal impact $(\sum_i \overline{\Lambda}_i^\tau)^T$ of the policy intervention. We define a *predictive* model as one that is uninformative about the causal structure of the policy intervention.

**Definition 3.** *We say a model $M$ is* predictive *if $\mathbb{E}[\Xi|s,M] = \mathbb{E}_0[\Xi]$ and $\mathbb{E}[\overline{\Lambda}_i^\tau|s,M] = \mathbb{E}_0[\overline{\Lambda}_i^\tau]$ for all signals $s$.*

## 2.4 Optimal Predictive Model Choice

We can now characterize the optimal choice of a model. We maintain simplicity by assuming predictive-causal independence for the main text (Definition 2) and by focusing on predictive models (Definition 3).

We assume that the regulator faces a separable utility cost $\mathcal{C}(M,q)$ from adopting (predictive) model $M \in \mathcal{M}$. In Proposition 2, for a predictive model (Definition 3) the key sufficient statistic of the model from a welfare perspective is the prior covariance matrix over the ex-post prediction of liquidations, $\Sigma_0^{\hat{L}} = cov_0(\hat{L}(q))$. The matrix $\Theta_0$ and the averaged expected liquidations $L_0(q)$ are invariant to model choice. For symmetric matrices $\Sigma_0^{\hat{L}}$ that are attainable from some model $M$, we can therefore re-represent costs as $C(\Sigma_0^{\hat{L}}, q) = \inf_{M \in \mathcal{M} | cov_0(\hat{L}(q)) = \Sigma_0^{\hat{L}}} \mathcal{C}(M,q)$. We assume that $C$ is differentiable in $(\Sigma_0^{\hat{L}}, q)$ over an open ball that contains its optimal value.

As a result, for a predictive model we can write the regulator's optimization problem as

$$\max_{\Sigma_0^{\hat{L}}} \frac{1}{2}\text{tr}\left(\Theta_0 \Sigma_0^{\hat{L}}\right) - C(\Sigma_0^{\hat{L}}, q)$$

subject to the constraint that $\Sigma_0^{\hat{L}}$ is a symmetric matrix. We obtain the following result on the optimal predictive model.

**Proposition 3.** *The regulator's optimal predictive model solves*

$$\frac{\partial C(\Sigma_0^{\hat{L}}, q)}{\partial \Sigma_0^{\hat{L}}} + \left(\frac{\partial C(\Sigma_0^{\hat{L}}, q)}{\partial \Sigma_0^{\hat{L}}}\right)^T = \Theta_0 \tag{11}$$

*where $\frac{\partial C(\Sigma_0^{\hat{L}}, q)}{\partial \Sigma_0^{\hat{L}}}$ is a square matrix whose ij-th element is $\frac{\partial C(\Sigma_0^{\hat{L}}, q)}{\partial (\Sigma_0^{\hat{L}})_{ij}}$.*

Proposition 3 yields an intuitive trade-off on optimal predictive model choice in terms of choice of covariance matrix of the policy intervention. The regulator is willing to pay a higher marginal cost to increase precision on dimensions where $\Theta_0$ is larger, that is when the policy has greater impact on aggregate liquidations and prices, when the regulatory cost is higher, or when the impact

through holding costs is higher. In this sense, there is a complementarity between predictive power and knowledge of the causal policy impact: the regulator is willing to pay larger costs to acquire precise predictions of aggregate liquidations precisely on dimensions on which the policy impact on liquidations and liquidation prices is anticipated to be largest.

## 2.5 Privately Optimal Asset Allocations

We now turn to studying how model choice and ex-post intervention shape the ex-ante asset allocations of intermediaries. We begin by studying the private optimum of individual intermediaries, and then study socially optimal interventions.

Intermediary $i$ in the Beginning takes as given the model choice $M^*$ of the regulator and the resulting possible equilibria, and solves

$$\max_{q_i} \mathbb{E}_0\left[q_i^T(R_i - p_i) - \ell_i^T(R_i + \tau_i^* - \gamma) - \frac{1}{2}q_i^T H_i^q q_i - \frac{1}{2}\ell_i^T H_i^\ell \ell_i\right].$$

Intermediary $i$'s asset allocation is therefore affected both through the specific intervention $\tau_i^*$ anticipated, and also through the liquidation price (which in turn also affects equilibrium liquidations). Intermediary $i$ knows that equilibrium liquidations are determined as in Lemma 1, but only internalizes the effect of its own $q_i$ on its own liquidations (and not on equilibrium liquidation price or equilibrium liquidations, and not on the optimal model choice or intervention).

We next turn to studying the effect of the regulator's model choice on ex-ante asset allocations of intermediaries. Note that Proposition 4 does not rely on predictive-causal independence or a predictive model.

**Proposition 4.** *The privately optimal asset allocation satisfies*

$$\mathbb{E}_0\left[H_i^q q_i^* + \Lambda_i^{qT} H_i^\ell \overline{\ell}_i(q^*)\right] = \mathbb{E}_0\left[R_i - p_i - \Lambda_i^{qT}\theta_i(q^*)\right] - \mathbb{E}_0\left[\Lambda_i^{qT}\left(\tau_i^* - \left(H_i^\ell \overline{\Lambda}_i^\tau + \Gamma(\sum_i \overline{\Lambda}_i^\tau)\right)\tau^*\right]\right. \tag{12}$$

Proposition 4 expresses the optimal portfolio choice in the form of a marginal cost-marginal benefit trade-off. The left hand side captures the marginal cost of increasing holdings of an asset, which includes both the ex-ante and ex-post adjustment costs of holding more and liquidating more of an asset. This term is evaluated at the no-intervention benchmark, that is as-if we had $\tau^* = 0$.

The right hand side captures the marginal benefit, which is decomposed into a marginal benefit absent the ex-post regulatory intervention (the first term) plus the marginal benefit arising from the impact of intervention (the second term). The first term on the RHS is the baseline expected asset return, $R_i - p_i$, net of costs of liquidations. The liquidation costs are the amount liquidations are changed by increasing holdings, $\Lambda_i^{qT}$, times the fire sale loss in liquidation, $\theta_i(q^*)$.

The final term on the RHS captures the impacts of model choice and the ex-post intervention.

This is the only term in equation 12 that depends on model choice and intervention. It captures the cost of holding an asset induced through regulation. The expectation of ex-post policy intervention has two impacts. First, a higher wedge $\tau_{in}$ on intermediary $i$ increases the cost of liquidating asset $n$, discouraging the intermediary from holding portfolios that result in it liquidating that asset ex post. There are also equilibrium costs of the vector of wedges $\tau$ through the equilibrium price. In contrast, here a higher wedge $\tau_{in}$ encourages the intermediary $i$ to hold more of an asset when the asset price rises as a result, which partially offsets the benefit from raising the price. This is a standard channel of moral hazard.

It is clear that the ex-post intervention affects the optimal asset allocation: a higher expected tax on asset $n$ ex post directly discourages its purchase ex ante, but a higher liquidation price ex post encourages its purchase ex ante. As a result, even a purely predictive model can impact the asset allocation ex ante. In particular, even though under a purely predictive model (Definition 3) the expectation of $\tau^*$ is that same as if the regulator ran no model, there is a covariance induced between the tax itself, $\tau^*$, and the impact on liquidations, $\Lambda_i^q$, as long as the predictive model is loading at least some on Bayesian inference on the impact of asset holdings on liquidations $\Lambda_i^q$. That is to say, focusing on the direct term $\mathbb{E}_0[\Lambda_i^{qT}\tau_i^*]$, for a purely predictive model we can write

$$\mathbb{E}_0\left[\Lambda_i^{qT}\tau_i^*\right] = \mathbb{E}_0\left[\Lambda_i^{qT}\right]\mathbb{E}_0\left[\tau_i\right] + \text{cov}_0\left(\Lambda_i^{qT}, \mathbb{E}_0[\Xi]^{-1}\mathbb{E}_0\left[(\sum_i \overline{\Lambda}_i^{\tau})^T\Gamma\right]\mathbb{E}[L(q)|s,M]\right),$$

where the second term reflects how the true value of $\Lambda_i^q$ varies with the prediction of liquidations. In a similar fashion, moral hazard can also be exacerbated if the impact of asset allocations on liquidations $\Lambda_i^q$ is what a key driver of the predictability of ex post liquidations.

## 2.6 Socially Optimal Asset Allocation

Finally, we study the constrained efficient asset allocation $q$ of the regulator. Formally, we assume that the regulator can impose a wedge on asset holdings, that is a revenue-neutral tax $t_i = (t_{i1}, \ldots, t_{iN})^T$. Because the regulator has a complete set of wedges ex ante, this is equivalent to the regulator directly picking portfolio allocations ex ante.[10] The following proposition characterizes the optimal wedges $t = (t_1^T, \ldots, t_I^T)^T$ when the regulator chooses an optimal predictive model ex post.

[10]It is straightforward to extend results to incomplete ex-ante instruments, but for brevity we focus on complete instruments.

**Proposition 5.** *For a predictive model, the social planer's optimal ex-ante portfolio wedges $t$ are*

$$\begin{aligned} t =& \mathbb{E}_0\left[(\sum_j \overline{\Lambda}_j^q)^T \Gamma L(q) - \Lambda^{q,eT} \sum_i \left(\theta_i(q) + H_i^\ell \ell_i(q)\right)\right] \\ &+ \frac{dC(\Sigma_0^{\hat{L}}, q)}{dq} - \mathbb{E}_0\left[(\sum_i \overline{\Lambda}_i^q)^T\right]\Theta_0 \mathbb{E}_0[L(q)] - \mathbb{E}_0\left[\Psi\tau\right] \end{aligned} \tag{13}$$

*where $\mathbb{E}_0[\Psi\tau]$ is a stacked vector whose $i^{th}$ block is $\mathbb{E}_0\left[\Lambda_i^{qT}\left(\tau_i - \left(\Gamma(\sum_i \overline{\Lambda}_i^\tau) + H_i^\ell \overline{\Lambda}_i^\tau\right)\tau\right)\right]$.*

The optimal ex-ante asset tax formula is intuitive and is decomposed into two lines. The first line captures the uninternalized effects of intermediaries increasing holdings of an asset in absence of regulatory intervention. First, as $q$ changes, total liquidations change across intermediaries, resulting in changes in the liquidation price and hence changes in revenue proceeds from total liquidations $L(q)$. Second, changes in forced liquidations through changes in the equilibrium price result in losses due to the size of the discount $\theta_i(q)$ and due to changes in marginal holding costs $H_i^\ell \ell_i(q)$. In absence of an ex-post intervention, this first line would capture the standard macroprudential regulation of asset positions.[11]

The second line captures the interaction of the ex-ante asset regulation with the ex-post model design and intervention. The first term, $dC/dq$, reflects how the changing assets held by intermediaries affects the cost of acquiring information through the model ex post. The regulator imposes larger holding taxes on asset positions that make running the model more costly ex post. For example, the regulator might discourage holdings of non-transparent or hard-to-assess asssets.

The second term on the second line captures the effect on the expected intervention size (the term $L_0(q)^T \Theta_0 L_0(q)$ in equation 10). Intuitively, concentrations into a position $q$ increase total liquidations on the margin by $\sum_i \overline{\Lambda}_i^q$, which prompts a larger ex post intervention by the regulator. This increase in ex-post intervention helps to manage the fire sale and so mitigates the impact of the ex-ante increase in asset allocation. This gives a first dimension whereby the ex-post intervention provides a partial substitute for the ex-ante asset tax, and leads the regulator to require potentially smaller interventions ex ante, knowing that fire sales will be managed ex post.

Finally, there are the direct and moral hazard effects on intermediaries, $\mathbb{E}_0[\Psi\tau^*]$, are those in the final term in equation 12 of Proposition 4. First, the prospect of the ex-post tax on liquidations directly discourages holdings of the asset ex ante when holding more promotes liquidations ex post. Second, there is the moral hazard effect: as the liquidation price rises, intermediaries' perceived cost of liquidations $\theta_i(q)$ falls, and so they become more willing to hold more of an asset even if that forces them to liquidate. It is interesting again to observe that for assets for which the direct effect dominates the indirect price effect, the ex-post tax actually serves as a substitute for ex-ante regulation, leading to a smaller intervention $t$ ex ante.

[11] Since our framework allows negative positions $q_{in} < 0$, we can think of these negative positions as liabilities. As such, it also captures regulation of such liabilities.

## 2.7 Ex-Post Subsidy-Based Interventions

Our baseline analysis assumes that the regulator directly manages liquidations through a wedge/tax on liquidations. In practice, ex-post interventions can also involve "bailout" interventions that subsidize retaining assets rather than taxing selling them (e.g. lender of last resort, debt guarantees). Suppose that rather than taxing liquidations $\ell_i$, the regulator instead applies a subsidy on the amount $q_i - \ell_i$ of the asset that is held to maturity. Formally, the payoff to intermediary $i$ in the End is now

$$U_i = U_i + (q_i^T - q_i^{*T})\tau_i - (\ell_i^T - \ell_i^{*T})\tau_i,$$

so that the regulator continues to apply revenue-neutral interventions.[12]

Because the regulator in the Middle takes the asset allocations as given, the regulator's optimization problem over both model choice and the ex-post intervention are formally the same as before. Thus, Lemma 1 and Propsitions 1-3 continue to apply. However, the privately optimal asset allocation is affected by the subsidy on asset retention. In particular, in Proposition 4, the expected return on holding asset $q_i$ is raised from $R_i$ to $R_i + \tau_i^*$. Intuitively an ex-post subsidy promotes overinvestment in that asset. This gives rise to a familiar channel of moral hazard.

Relative to the case of a liquidation tax, it is worthwhile to note that the only new term for private intermediaries in their optimization problem is the revenue benefits $q_i^T \mathbb{E}_0[\tau_i^*]$ of their asset position. Therefore, relative to the liquidation tax, the regulator's model choice only affects intermediaries' asset allocation ex ante to the extent it changes the expected size of the intervention, $\mathbb{E}_0[\tau_i^*]$. In particular for a predictive model under predictive-causal independence (Definitions 2 and 3), we know that $\mathbb{E}_0[\tau_i^*]$ does not depend on the model used; instead, it is only the covariance matrix of the intervention (its precision) that deepends on the model. It is thus interesting and surprising that purely predictive models that help with prediction but not with causal inference do not exacerbate moral hazard relative to the case of the liquidation tax. In contrast, models that are informative about the causal impact of the policy intervention change expectations of the policy intervention, and are potentially associated with moral hazard. The logic is reminiscent of that of Laffont and Tirole (1986) in the context of regulation of a firm with unobservable effort and uncertain costs.

## 2.8 Predictive Correlations as Sufficient Statistics for Welfare

In Section 2.3, we showed that the welfare-relevant information for a predictive model is summarized by the prior covariance matrix over the ex-post policy interventions induced by the choice of that particular model. It is intuitive that this covariance matrix should relate to the predictive accuracy of the model, since it characterizes how well the intervention is targeted. In this section, we provide a corollary to Proposition 2 which formalizes this intuition, showing that the predictive correlation

[12] Since individual intermediaries take revenue remissions as given, there is still moral hazard from the subsidy.

of a predictive model's forecasts—weighted appropriately according to the causal leverage of the policy—is a sufficient statistic for the welfare gains from the model.

To derive this corollary, we define a few objects. Given a square matrix $W$, we define variance and covariance operators that use $W$ as a weighting matrix:

$$\mathrm{Cov}_W(X,Y) = \mathbb{E}\left[(X - E[X])^T W (Y - E[Y])\right], \quad \mathrm{Var}_W(X) = \mathrm{Cov}_W(X,X).$$

The maximum ex-ante expected welfare gain (relative to the no-intervention baseline) within the class of predictive models, $\Delta V_{\max}$, is attained by a perfect predictive signal—that is, a predictive model $M$ whose forecasts $\widehat{L}(q) = \mathbb{E}[L(q) \mid s, M]$ satisfy $\widehat{L}(q) = L(q)$. The following corollary characterizes welfare losses relative to the perfect forecast benchmark $\Delta V_{\max}$ in terms of the model's predictive accuracy.

**Corollary 2.** *Under predictive-causal independence, the regulator's ex-ante expected welfare gain from a predictive model relative to the no-intervention baseline, given positions $q$ and a model $M$ is*

$$\Delta V(M) = \Delta V_{\max} \cdot \rho^2_{\Theta_0}(M), \tag{14}$$

*where*

$$\rho_{\Theta_0}(M) \equiv \frac{\mathrm{Cov}_{\Theta_0}(L, \widehat{L})}{\sqrt{\mathrm{Var}_{\Theta_0}(L)\,\mathrm{Var}_{\Theta_0}(\widehat{L})}}, .$$

This result shows that the welfare gains from a predictive model scale linearly with the $\Theta_0$-weighted $R$-squared achieved by the model in predicting liquidations. The weighting matrix $\Theta_0$ has a natural interpretation as the metric that converts variation in predicted liquidations $\hat{L}(q)$ into variation in welfare after the regulator applies the optimal wedges $\tau^*$: therefore, assets that generate larger externalities or for which the intervention has a higher causal leverage on outcomes receive a larger weight in the statistic $\rho_{\Theta_0}(M)$. The predictive $R$-squared has in fact an exact interpretation as the fraction of the maximum achievable welfare gain within the class of predictive models that is actually attained. This sufficient statistic provides a simple but direct way to quantify map the empirical performance of predictive models to welfare.

## 2.9 Extensions

**Better Informed Intermediaries.** Our model assumes that in the Beginning, intermediaries and regulators share a common prior over the model parameters. In practice, intermediaries may be better informed about model parameters. One could extend the model by assuming that intermediaries had a more precise prior than regulators. This would lead a regulator to also infer information about the structural parameters of the economy from observing portfolio holdings directly. One possible method of incorporating is to interpret $C(M, q)$ as a cost of the inference the regulator is undertaking, and so interpret the model as both a processing of information (including from such

Bayesian inference). One possible advantage of the ex-post intervention would be its ability to react to information discovered from observing intermediaries' choices, which might reduce risks of moral hazard or imperfectly calibrated regulation ex ante. That is, better-informed intermediaries who saw a regulator was under-regulating an asset $n$ ex ante would also know that the regulator would discover the mistake ex post, and so intervene more strongly upon it. Exploring the effects of differential information between the regulator and intermediaries is an interesting direction.

**Dynamic Learning.** We embed a one-shot learning problem. One could extend our framework to embed dynamic information acquisition by assuming that our baseline Beginning-Middle-End model was a stage game played at each date $t = 0, 1, \ldots$. In this environment, one could think of there being a true distribution $\mu^*$ from which model parameters are drawn each period, so that the regulator and intermediaries learn about this true distribution over time. The regulator and intermediaries would carry information forward at each date, and so the regulator would consider both how a model acquired information on the current crisis and also how it informed about the underlying variables. We conjecture that this would make predictive models relatively myopically useful: they would give potentially substantial information for intervening in the current crisis, but relatively little information about the underlying structural parameters, and so might be of limited use in updating beliefs about the true distribution of parameters. This could face the regulator with an interesting dynamic trade-off between myopically acquiring more predictive information, and trying to uncover the true structural parameters that would also be useful for designing regulation and interventions during the next crisis.

**Commitment vs. Discretion in Model Choice.** We have assumed the regulator chooses the model and ex-post policy intervention with discretion, after asset allocations are chosen. As highlighted by the results in Section 2.5, this leads to a potential time consistency problem in model choice since the model choice affects ex-ante asset allocations. This time consistency problem will likely be more pronounced the more incomplete the regulator's ability to regulate portfolio positions is ex ante, for example if the regulator cannot regulate all intermediaries or all assets. In our ongoing work, we are exploring the implications of commitment versus discretion for optimal model design and welfare gains from use of a predictive model.

# 3 Graph Representation Learning for Holdings Data

We now move to an empirical examination of whether high-dimensional forecasting models can in fact be used successfully for macroprudential applications. This analysis establishes several points. First, it provides a blueprint for the practical implementation of deep learning models by central banks and other financial supervisors. Second, it demonstrates that such models do in fact have significant predictive power—which is a crucial motivating fact for the theory. Third, motivated

by the theoretical setting, it lays out several key design principles that deep learning architectures should follow when applied to financial holdings data.

**GNN Architectures and Holdings Data.** We begin by introducing a deep learning architecture tailored to holdings data, and we discuss why it is optimally suited to this setting. Much of the modern deep learning toolkit is optimized for grid inputs such as images and sequence inputs such as text. Indeed, architectures that have proved successful are those that exploit the particular structure of the data they are modeling: examples include convolutional neural networks for images and other grid-structured data, and text transformers—which underpin modern large language models—for textual data. Financial holdings data does not fit neatly into either of these categories. Instead, its defining feature is that it has very rich *relational structure*: the data can be naturally thought of as representing a graph connecting investors and assets, with the information relevant to the learning task contained in the graph's edges, i.e. the positions connecting investors to assets.

Graph neural networks (GNNs) are a class of deep learning models that specifically models and exploits relational structure in the data (Scarselli et al. 2008; Hamilton et al. 2017; Wu et al. 2020). By iteratively propagating and aggregating information along the edges that connect graph nodes (in this case, assets and investors), GNNs learn embeddings for each node in the graph—i.e., representations of the nodes in a latent vector space that effectively capture the characteristics relevant to the tasks the model is trained against. In practice, in our implementation an investor's embedding evolves based on the embeddings of the assets they hold, and those asset embeddings, in turn, adjust in light of the embeddings of the investors that include them. Through iterated rounds of this neighbor-aggregation mechanism (also known as *message passing*), information flows along the network of positions, allowing the model to learn explicitly from the relational structure of the holdings data.

GNNs have state-of-the-art empirical success in domains where relational structure is paramount. The protein-folding model AlphaFold, for example, uses a graph-based architecture to encode the three-dimensional interactions among amino acids, and it has dramatically advanced the field of protein structure prediction (Jumper et al. 2021). Traffic-forecasting systems used for products such as Google Maps represent road networks as graphs and learn congestion and travel-time patterns directly from the topology of streets and highways (Derrow-Pinion et al. 2021). Similarly, in frontier drug discovery models, molecules are treated as graphs of atoms and bonds, and GNNs have revolutionized the prediction of chemical properties and binding affinities (Jiang et al. 2021). Each of these breakthroughs rests on the fundamental ability of GNNs to natively handle and learn from graph-structured data.

In the context of asset holdings data, GNNs have two key properties that other deep architectures lack: permutation invariance and inductive learning. First, the models are *permutation invariant*, meaning that the learning process and estimation results do not depend on any arbitrary relabeling or reshuffling of investor and asset identifiers. The architecture attains permutation invariance

because aggregation operates over unordered sets of neighbors, and by doing so they respect a fundamental symmetry of the problem.

Second, the GNN architecture features *inductive learning*. Crucially, all of the model's parameters are shared, in the sense that there are no parameters specific to particular assets or investors. Therefore the same learned representation rules apply across all nodes and edges: given a graph $\mathcal{G}$, regardless of the number and identity of the nodes, the trained GNN architecture is able to construct asset and investor embeddings simply from the graph structure and the node characteristics. Inductive generalization thus follows naturally from parameter sharing: the model can immediately generate embeddings and forecasts for new, unseen investors or assets without any retraining. This feature is not only valuable for real-time regulatory applications, but also it enforces a strong form of regularization upon the model, preventing overfitting of the training data and leading to good generalization to unseen, out-of-sample data.

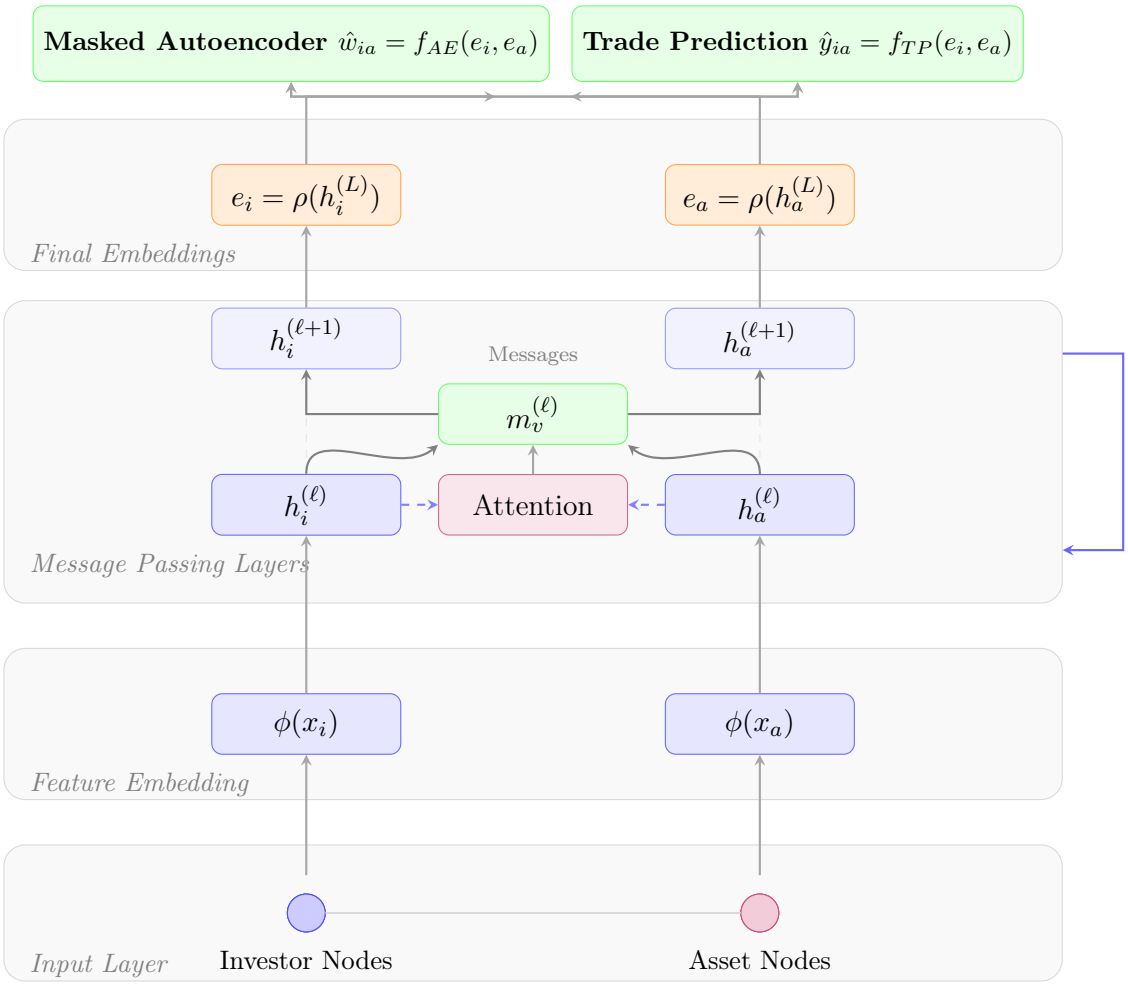
This combination of relational representation learning, proven empirical performance, permutation invariance, and inductive design makes GNNs a particularly well suited deep learning architecture for the holdings data setting and for macroprudential applications. Our specific implementation is a *graph transformer*, which incorporates an attention mechanism in the basic GNN architecture, as we discuss below.

**Core Architecture Specification.** To formally specify the GNN graph transformer architecture, we start by modeling the holdings data as a bipartite graph—meaning a graph with two distinct types of nodes, and whose edges only connect nodes of the two different types. The holdings data forms a bipartite graph in which investors connect to assets via position edges. Formally, we let a given cross-section of the data be represented as $\mathcal{G} = (\mathcal{I}, \mathcal{A}, \mathcal{E})$, where $\mathcal{I} = \{1, \ldots, I\}$ indexes investors, $\mathcal{A} = \{1, \ldots, N\}$ is the set of assets, and $\mathcal{E}$ is the set of edges (i.e., positions). A certain position exists in the graph if the corresponding investor-asset pair is in the set of edges: whenever investor $i$ holds asset $a$, letting $w_{ia}$ be the size of the position, we have that $(i, a, w_{ia}) \in \mathcal{E}$.[13] We write $\mathcal{V} = I \cup \mathcal{A}$ as the set of all nodes, both assets and investors. We let $N(v)$ be the neighborhood set for a given node $v$. This is the set of all nodes that are connected to $v$: for assets, $N(v)$ corresponds to the investors holding the asset, while for investors this corresponds to the set of assets in the investor's portfolio.

Each node $v \in \mathcal{V}$ has an associated set of characteristics $x_v$. The characteristics can be either numerical (e.g., the total amount outstanding of a given security) or categorical (e.g., the type of institutional investor). For categorical features, these are pre-embedded to a vector space of dimension $d_c$ using maps that are learned jointly with the rest of the model's parameters. The node characteristics vector $x_v$ concatenates over all individual features, both scalar-valued ones and vector-valued categorical ones.

[13] The positions can alternatively be written as $w_{ai} = w_{ia}$. We allow for both notations so as to keep the rest of the formal architecture specification symmetric.

Figure 1: **Model architecture**




*Notes:* We visualize the architecture of our graph transformer model diagrammatically.

The graph transformer learns node embeddings (both asset embeddings and investor embeddings) at various hierarchical levels of information aggregation. The first-level embeddings, which we denote as $h_v^{(0)}$, are obtained simply by embedding the node characteristics vectors $x_v$ into a $d_h$–dimensional hidden space via a learnable map $\phi$:

$$h_v^{(0)} \;=\; \phi(x_v)\,. \tag{15}$$

Next, the architecture passes these layer-zero node embeddings through several successive layers of message passing which aggregate information over the graph: this message passing stage is the core of the GNN model and is what allows the model to learn from the data's relational structure. To increase the expressivity and learning capability of the model, we integrate attention mechanisms in each of the GNN message passing layers, which allow the architecture to learn from the data which positions should be given more or less weight in any given information aggregation steps. This incorporation of attention layers is what makes our architecture a graph transformer.

We perform $L$ distinct layers of message passing. The message passing at layer $\ell$ unfolds in several distinct steps. First, the prior-layer node embeddings are passed through a feed-forward layer $M^{(\ell)} : \mathbb{R}^{d_h} \to \mathbb{R}^{d_h}$ to construct *node messages*:[14]

$$M_v^{\ell} = M^{(\ell)}(h_v^{(\ell-1)}). \tag{16}$$

Next, the attention mechanism computes aggregation weights that dictate how the individual node messages will be weighted when passing information over the graph. We allow for $S_A$ distinct and independent attention heads. Each attention head $s = 1, \ldots, S_A$ forms attention weights using learnable projection matrices $W_{qs}, W_{ks} \in \mathbb{R}^{d_h \times d_h}$ which project prior-layer embeddings into query and key spaces, respectively—akin to how text transformers form attention values for the interactions of tokens in a text sequence (Vaswani et al. 2017). The attention weight $\alpha_{vu}^{(\ell)}$ indicates the importance of messages from each node $u$ in the neighborhood $N(v)$ of node $v$ at layer $\ell$, and it is formed by averaging over the individual attention heads:

$$\alpha_{vu}^{(\ell)} = \frac{1}{S_A} \sum_{s=1}^{S_A} \frac{\exp\{(W_{qs}\, h_v^{(\ell-1)})^T (W_{ks}\, h_u^{(\ell-1)})\}}{\sum_{u' \in \mathcal{N}(v)} \exp\{(W_{qs}\, h_v^{(\ell-1)})^\top (W_{ks}\, h_{u'}^{(\ell-1)})\}}. \tag{17}$$

Having established attention weights $\alpha_{vu}^{(\ell)}$, the message passing algorithm then constructs an aggregated message $m_v^{(\ell)}$ for each node $v$ by averaging over the messages received from each of its neighbors $u \in \mathcal{N}(v)$, according to:

$$m_v^{(\ell)} = \sum_{u \in \mathcal{N}(v)} w_{vu}\, \alpha_{vu}^{(\ell)}\, M_u^{(\ell)}, \tag{18}$$

[14] All feed-forward layers in our architecture use GELU non-linearities (Hendrycks and Gimpel 2016). These help avoid dead gradients during training.

The node embeddings at the next hierarchical stage $\ell$ are then updated using another feed-forward layer $U^{(\ell)} : \mathbb{R}^{2d_h} \to \mathbb{R}^{d_h}$ which maps the current embeddings and the current aggregated messages into the next-stage representations:

$$h_v^{(\ell)} = U^{(\ell)}(h_v^{(\ell-1)}, m_v^{(\ell)}). \tag{19}$$

After $L$ rounds of message-passing, a readout feed-forward layer $\rho : \mathbb{R}^{d_h} \to \mathbb{R}^{d_e}$ produces final asset and investor embeddings:

$$e_v = \rho(h_v^{(L)}). \tag{20}$$

When referring to node embeddings in subsequent sections, unless otherwise specified we refer to these final representations $e_v$. Because all parameters are shared across nodes and layers, this architecture is both permutation invariant and inductive, allowing predictions on unseen investors or assets without retraining.

**Prediction Heads.** To train and leverage the learned embeddings $e_v$, we attach two task-specific heads to the graph transformer. The first task uses a masked autoencoder (MAE) objective, where we randomly mask a subset of edges (including non-existent edges drawn at random, which we represent using $w_v = 0$) and ask the model to predict whether the edge exists ($w_v \neq 0$) and the associated position size $w_v$. The second task uses a supervised trade prediction objective, where we ask the model to engage in a pure forecasting task, using embeddings computed using time $t$ holdings information to predict the cross-sectional pattern of investor trades in the future, between time $t$ and time $t+1$.

For the masked autoencoder (MAE) objective, we randomly mask a subset of edges and predict masked weights $\hat{w}_{ia} = f_{\text{AE}}(e_i, e_a)$ using a feed-forward head layer $f_{\text{AE}}$ which takes as inputs the embeddings for the given masked investor-asset pair $(i, a)$. The autoencoder objective minimizes a mean squared error loss defined over the divergence between the true edge weights $w_v$ and the predicted edge weights $\hat{w}_v$:

$$\mathcal{L}_{\text{AE}} = \sum_{(i,a)\in\mathcal{V}_{\text{masked}}} (w_{ia} - \hat{w}_{ia})^2, \tag{21}$$

where $\mathcal{V}_{\text{masked}}$ denotes the set of masked edges.

For trade prediction, we construct the targets by first defining the percent changes in holdings for a given position in asset $a$ by investor $i$ between time $t$ and time $t+1$ (in practice, quarters) as:

$$\Delta\%q_{ia,t} = \frac{q_{ia,t} - q_{ia,t-1}}{q_{ia,t-1}}. \tag{22}$$

We strip away all common movement in trades for a particular asset $a$, such as for example movement induced by changes in the asset's valuation which affect all investors. To do this, we construct cross-sectional z-scores of the trades $\Delta\%q_{ia,t}$, which subtract the average percentage position change for asset $a$ in the given time period ($\overline{\Delta\%q_{a,t}}$) and divide by the standard deviation of the same position

changes ($\sigma(\Delta\%q_{a,t})$), placing all assets in all time periods on the same scale. The cross-sectional trade z-scores are thus defined as:

$$y_{ia,t} = \frac{\Delta\%q_{ia,t} - \overline{\Delta\%q_{a,t}}}{\sigma(\Delta\%q_{a,t})}. \tag{23}$$

The cross-sectional trade patterns captured by $y_{ia,t}$ are the targets for the supervised prediction head. Specifically, we construct predicted trades $\hat{y}_{ia} = f_{\text{TP}}(e_i, e_a)$ using a feed-forward layer head $f_{\text{TP}}$ which acts on the relevant pair of asset and investor embeddings.[15] The supervised objective is again defined over the mean squared error between the actual trades and the predicted trades:

$$\mathcal{L}_{\text{TP}} = \sum_{(i,a)\in\mathcal{V}} (y_{ia} - \hat{y}_{ia})^2. \tag{24}$$

**Model Training.** To recap, the model contains several learnable components, all of which are parameterized using a high-dimensional set of parameters. The trainable components include the feature map $\phi$, the pre-embedding functions for categorical characteristics, the message feed-forward layers $M^{(\ell)}$, the attention mechanism projection matrices $W_{qs}$ and $W_{ks}$, the node update function $U^{(\ell)}$, the embeddings projection layer $\rho$, and the task-specific prediction heads $f_{\text{AE}}$ and $f_{\text{TP}}$. The model architecture is summarized visually in Figure 1. We collect the set of parameters in all these learnable components in the vector $\Theta$.

The model is trained end-to-end by minimizing a joint loss which combines the mean squared error losses from the two training tasks:

$$\min_{\Theta} \; \mathcal{L}(\Theta) = \mathcal{L}_{\text{AE}} + \kappa\, \mathcal{L}_{\text{TP}}, \tag{25}$$

where $\kappa > 0$ determines the relative weight of the two training tasks. We optimize the model parameters $\Theta$ using the Adam optimizer (Kingma 2014).

**Optimality of Graph-Based Architectures for Holdings Data.** Before moving on the empirical implementation, we discuss more precisely the sense in which message-passing, graph-based architectures are optimal in the context of holdings-based problems. To do this, we lay out a few definitions. Let $W \in \mathbb{R}^{I\times N}$ be the full holdings matrix with entries $w_{ia}$, and let $f\colon \mathbb{R}^{I\times N} \to \mathbb{R}^d$ be a functional acting on the graph $\mathcal{G}$ represented by $W$. A continuous graph functional is permutation-invariant if, for all permutation matrices $P_1 \in \mathbb{R}^{I\times I}$, $P_2 \in \mathbb{R}^{N\times N}$, it satisfies $f(P_1 W P_2^T) = f(W)$. Informally, permutation invariance means that if we were to arbitrarily relabel columns and rows of the holdings matrix (i.e., investors and assets), the output of $f$ would remain the same. All regulatory or prediction targets we care about (future trading patterns, systemic risk scores, etc.)

[15] While for compactness we are not carrying through time subscripts on the embeddings $e_v$, naturally for the trade prediction task we use embeddings estimated using the graph $\mathcal{G}$ at time $t-1$ to construct the predictions for time $t$, $y_{ia,t}$.

Table 1: **Model hyperparameters**

| | |
|---|---|
| Hidden dimension ($d_h$) | 256 |
| Embedding dimension ($d_e$) | 128 |
| Layers ($L$) | 3 |
| Attention heads | 4 |
| Dropout rate | 0.1 |
| Learning rate | $10^{-2}$ |
| Weight decay | $10^{-5}$ |
| Loss weight ($\kappa$) | 1 |

*Notes:* We list the hyperparameters used for our graph transformer architecture and for training.

are assumed to lie in this family, reflecting the economics of the problem.

A well-known idea in the literature on graph deep learning is that in order to represent a permutation-invariant mapping with the fewest parameters, one should enforce permutation invariance via shared (message-passing) parameters (Zaheer et al. 2017, Maron et al. 2018, Xu et al. 2018). This is the sense in which the models are optimally sample-efficient. For illustration, compare two classes of models. First, consider the class of GNNs which implement permutation-invariant message-passing layers as described above, with shared parameters. Second, consider the class of sequence or grid networks that act on the flattened matrix vec($W$) under a fixed but arbitrary ordering of rows and columns (this class includes recurrent neural networks, convolutional neural networks, sequence transformers, and so on). Intuitively, GNNs can approximate permutation-invariant graph functionals without carrying superfluous degrees of freedom that sequence/grid networks would have to "use up" to relearn permutation invariance from data. Message-passing GNNs are efficient because the architecture is itself permutation-invariant, avoiding the need to use additional parameters to learn and enforce it.

# 4 Empirical Implementation

**Setting and Data.** We implement our empirical approach in the setting of non-bank financial intermediation. Since the Great Financial Crisis, tighter bank regulation (e.g., leverage and proprietary-trading constraints) has reduced dealer balance-sheet intermediation and pushed a larger share of risk-taking and liquidity transformation into non-banks, such as money market funds, funds holding bonds and collateralized loan obligations, ETFs, insurers, and pensions. These institutions now intermediate a substantial fraction of tradable credit and duration risk and are subject to fire-sale amplification mechanisms such as runnable or flow-sensitive funding and mark-to-market constraints (Chen et al. 2010; Goldstein et al. 2017; Aramonte et al. 2023; Coppola 2025). Therefore they have become central to episodes where widespread price dislocations occur: for instance, acute liquidity needs by bond mutual funds played a key role in debt market disruptions in the first

quarter of 2020 (Haddad, Moreira and Muir 2021; Ma, Xiao and Zeng 2022).

We train our deep learning model on quarterly institutional holdings from Factset, starting in 2005Q1, which cover a broad range of non-bank intermediaries. The training-validation split is done at the level of quarters: we sort quarters into training and validation sets. The holdings data gives us observations of the positions graph $\mathcal{G}$ for each quarter's cross-section, and we also use it to construct the standardized trade indicators $y_{ia,t}$. We do not include any quarters following 2019Q3 in the training set, so that we can use the Covid crisis of 2020 as a particularly strict test episode, in the sense that the Covid quarters are not just out of the training sample, but also the model is only trained with data prior to the start of the crisis, mimicking the way in which the model would be deployed in an actual regulatory scenario.

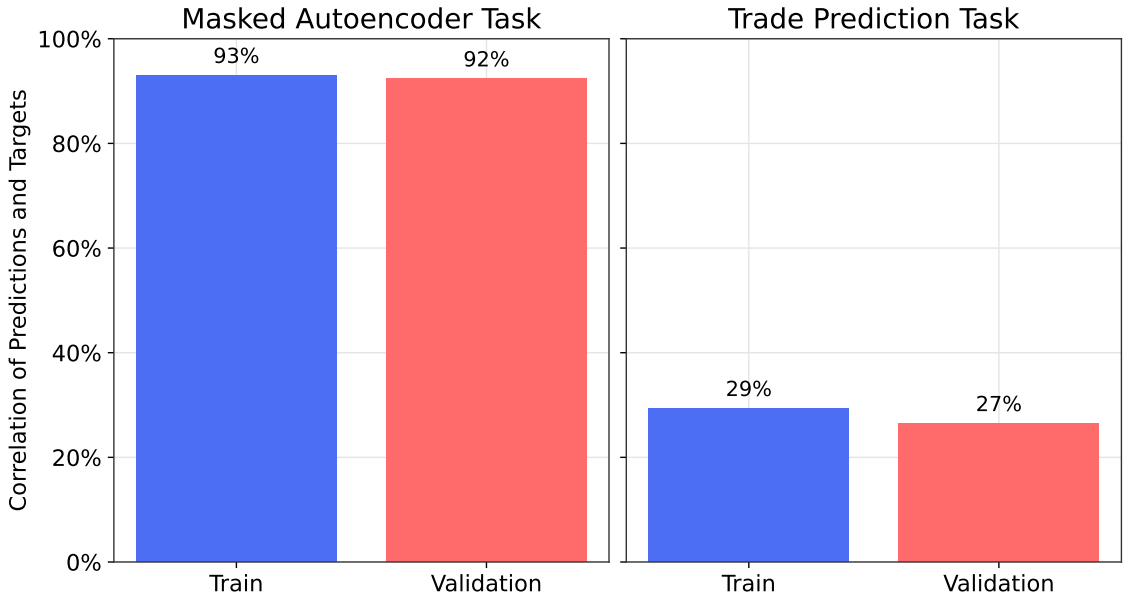
We construct the node feature vectors $x_v$ using reference information from Factset as well as from the Global Capital Allocation Project (GCAP) security master file (Coppola et al. 2021, Coppola 2025). For assets, the feature vectors $x_v$ include asset class, currency of denomination, amount outstanding, number of holders, average position size, standard deviation of position size, as well as bond sub-class and coupon for debt securities. For investors, they include institution type (such as open-end mutual funds, ETFs, separate accounts, etc.), manager style (including flags for active vs. passive portfolio management and strategy types), total AUM, number of positions, average position size, and standard deviation of position size. In principle, our architecture also allows for the use of global features that vary over time but not across nodes: these can be introduced as vectors which enter message-passing in the same way for all nodes in a given time period. In ongoing work, we are integrating global features and assessing the impact on the model's performance.[16] Similarly, we are exploring the use of price data both as predictive features and targets for the model.

**Parameterization and Training.** Hyperparameters are chosen as in Table 1. In particular, we set the hidden dimension to $d_h = 256$ and the final embeddings dimensionality to $d_e = 128$. We use $L = 3$ layers of message-passing: we do not increase $L$ beyond this number to prevent over-smoothing problem, whereby the node embeddings would converge to similar values: intuitively, over-smoothing occurs for higher numbers of message-passing iterations since each consecutive iteration increases the *receptive field* of each node, i.e. the set of nodes that the final-layer embeddings attend to, and for high $L$ values the receptive fields of all nodes converge to the largest possible field, which is the set of all nodes in the graph. We allow for four attention heads and we give equal weight to the two training tasks by setting $\kappa = 1$. Altogether, the model has a total of 3,640,465 parameters.

We introduce dropout during training for additional regularization, with a dropout rate of 0.1. The Adam optimizers uses a starting learning rate of $10^{-2}$ with progressive weight decay. We train the model using a compute node with four NVIDIA H100 GPUs. We train for up to 500 epochs, with early stopping based on the loss on the validation sample.

[16]The weight to be placed on global features can be made learnable by the model. Global features may

Figure 2: **Performance metrics**




*Notes:* We plot the correlations between the trained model's predictions and the targets for the two tasks (masked autoencoder and cross-sectional trade prediction). The blue bars show performance on the training set, while the red bars show out-of-sample performance on the validation set.

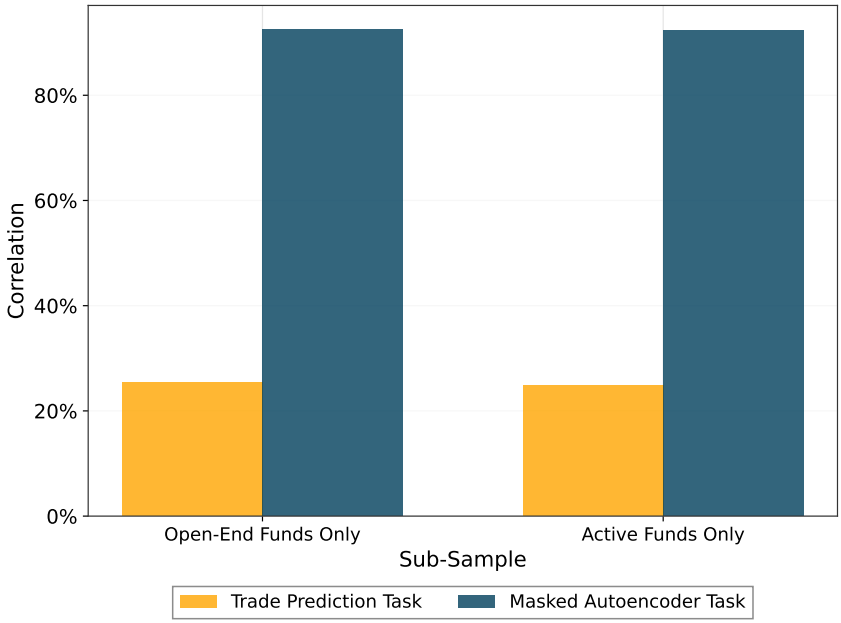
**Predictive Performance of the Model.** Out-of-sample, the MAE head achieves a correlation above 0.90 between reconstructed and true positions, indicating that the GNN captures structural regularities in holdings. As mentioned in the introduction, the autoencoder prediction is best interpreted in the context of the model's parameter-to-data ratio: in this case, the model's roughly 3.6 million parameters represent less than 1% of possible investor-asset pairs in the data. The trade-prediction head yields an average correlation of just under 0.30 between predicted and realized trade indicators, with minimal degradation from training to validation sets. Figure 2 plots the correlation between the trained model's predictions and the targets for both tasks: the blue bars show the performance on the training sample, while the red bars show the out-of-sample performance on the validation set. The fully inductive design, sharing parameters across nodes and layers, ensures stable performance on unseen investors and assets, such that the model performs very similarly out-of-sample as it does on the training data.

The performance of the model on out-of-sample data is significantly higher than that achieved by alternative benchmark architectures: for example, a ridge regression model using the same asset and investor characteristics as regressors achieves 3% out-of-sample correlation on the trading prediction task, while a recommender system model using factorizations of the holdings data achieves a correlation of 8%. In ongoing work, we are performing a descriptive analysis of the asset and investor embeddings produced by the graph transformer model, so as to provide greater interpretability of

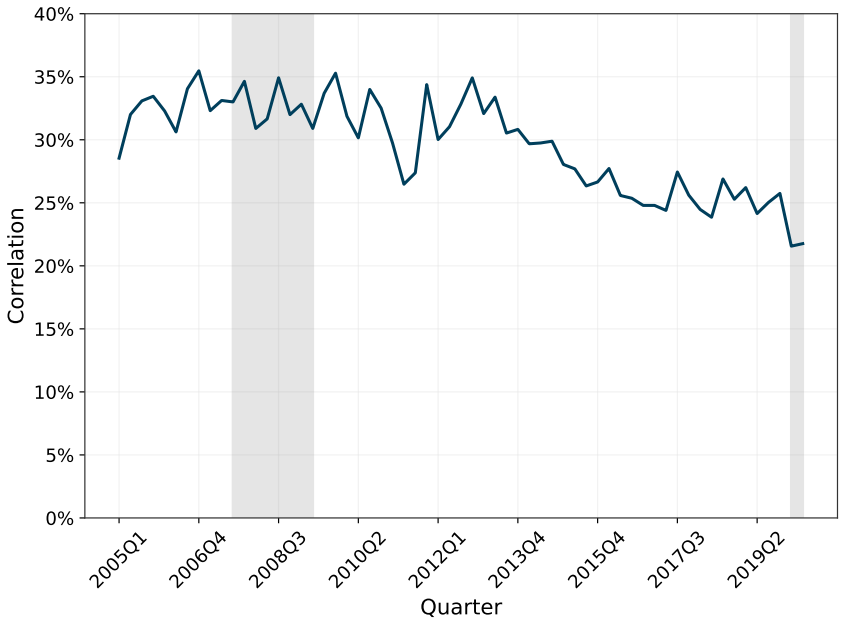
include time series measures such as aggregate credit spreads and other macro series.

Figure 3: **Performance metrics: heterogeneity**

(a) Validation set performance, by subsample




(b) Performance for trade prediction, by date




*Notes:* Panel A shows the out-of-sample correlation on the validation set between the trained model's predictions and the targets, for both tasks (trade prediction and masked autoencoder). We show this in the subsamples consisting of open-end mutual funds only and of active funds only. Panel B shows the correlation between trade predictions and targets on the full training and validation data, by quarter. Shaded gray areas correspond to periods where the St. Louis Fed Financial Stress Index is above 1.5.

the model's predictions.

A natural question is whether the predictive ability of the model comes primarily from relatively more mechanical aspects of the data, such as by correctly assessing the trades of passive index-tracking investors. In Figure 3a, we show that this is not the case by reporting the out-of-sample performance of the model on a sub-sample consisting only of active investors. We also show performance on a sub-sample that only includes open-end mutual funds, as these are the investor category with the highest degree of coverage within the Factset holdings data. In both cases, the forecasting performance of the model is quantitatively similar to that on the full validation sample.

An additional possible concern is that the model's predictive ability may be concentrated in calm periods rather than the market stress episodes where macroprudential interventions are most relevant. To rule this out, in Figure 3b we also show the performance on the trade prediction task separately for each quarter in the sample, as a time series: since the train-validation split occurs at the level of quarters, this time series naturally combines both training and validation data. The shaded gray areas correspond to market stress periods, defined as those when the St. Louis Fed Financial Stress Index is above 1.5. The predictive performance is consistently high both in stress periods and in non-stress periods.[17] In particular, the model's performance during the Covid crisis yields a particularly stringent test, since none of the quarters following 2019Q3 are included in the training set: the model is only trained with pre-crisis data, precisely as it would be deployed in an actual regulatory scenario, and nonetheless it displays high accuracy in predicting the patterns of trading during the course of the crisis quarters.

**Explaining Liquidations in a Stress Episode.** Finally, we study if the model's learned embeddings are in fact capturing economically meaningful risk dimensions. We trained the model to understand and predict portfolio positions and trading behavior in both calm times and market stress periods, which are tasks for which training data is relatively abundant. We now investigate whether the trained embeddings can explain the cross-section of liquidation prices in a particular out-of-sample market event where price dislocations were present: since these are rare events, this task has far less data available, and hence our approach is not to fit the model to it at training time, but rather ask whether the structure learned by the model generalizes to it.

Specifically, we focus on the dynamics of bond markets in the first quarter of 2020, during the Covid crisis: as discussed above, this is an episode that featured rapid and widespread disruptions in debt markets, partly driven by acute liquidity needs of investment funds and other financial intermediaries (Haddad et al. 2021; Ma et al. 2022). The fact that this event is also fully outside of our training sample makes it ideal to study our question of interest. We consider the cross-section of bond returns $r_{a,t}$ in 2020Q1, and for each bond we use the trained graph transformer model

[17] The FRED ticker for the St. Louis Fed Financial Stress Index is STLFSI4. We also note that the model's performance exhibits a slight downtrend occurring between 2012 and 2017, stabilizing by the end of the period.

Table 2: **Explaining bond price dislocations**

| | (1) | (2) |
|---|---|---|
| **Cross-Validated $R^2$** | | |
| Total | .08 | .42 |
| Incremental from $e_{a,t-1}$ | — | .18 |
| Bond Characteristics ($X_{a,t-1}$) | No | Yes |
| Observations ($N$) | 4,771 | 4,771 |

*Notes:* We report the out-of-fold $R$-squared from gradient boosting models applied to the specification in equation (26), which models returns in the cross-section of bonds in 2020Q1 as a function of the GNN's learned prior-quarter embeddings $e_{a,t-1}$ and security characteristics.

applied to holdings data from 2019 to construct the prior quarter embeddings $e_{a,t-1}$.[18] We estimate empirical models of the form

$$r_{a,t} = f(e_{a,t-1},\, X_{a,t-1}) + \varepsilon_{a,t}, \tag{26}$$

where $X_a$ is a bond characteristics vector including bond rating and residual maturity. Given the high-dimensional nature of the embeddings $e_{a,t-1}$, we estimate a non-linear, nonparametric model. Specifically, we train a gradient boosting model using five-fold cross-validation. Table 2 reports the $R$-squared values from out-of-fold prediction in the cross-section of bonds using the specification in equation (26). We report the out-of-sample $R$-squared both from a model which includes only the embeddings (column 1) and a model which includes both the embeddings and bond characteristics (column 2). For the latter, we also report the incremental $R$-squared from the embeddings, which is the difference in the $R$-squared between the full model (including both $X_{a,t-1}$ and $e_{a,t-1}$) and a model including only characteristics. The incremental $R$-squared coming from the embeddings in the full model is 18%, and it is in fact higher than the value from the model which includes only the embeddings (as the full model, by virtue of being non-linear, also allows for interactions between the embeddings and the characteristics). Hence the GGN's learned representations explain a large portion of the variation in the cross-section of bond liquidation prices in this market stress episode, even after accounting for bond characteristics, despite these not being a target of the model's training. Altogether, these results demonstrate the efficacy of graph-based predictive models for real-time macroprudential surveillance.

[18] We use the sample of bonds that are reportable to TRACE, the transactions database maintained by FINRA. We apply to TRACE the data cleaning steps in Dick-Nielsen (2014), which are common in the literature. TRACE lets us observe the prices at which bonds transact. Quarterly bond returns account for price changes, coupons, and accrued interest, and they are constructed as in Coppola (2025).

# 5 Conclusion

This paper develops a theoretical and empirical framework for understanding the role of high-dimensional predictive models in financial regulation. We formalize the tradeoffs regulators face when deploying models that deliver precise forecasts but limited causal insight, and we characterize when and how such models can improve welfare. We introduce a graph-based deep learning architecture tailored to holdings data, which we use to learn representations of assets and investors which achieve state-of-the-art results in forecasting trading patterns with minimal out-of-sample performance loss. Our empirical analysis demonstrates that real-time prediction of portfolio dynamics is feasible and provides a blueprint for practical implementation. While predictive models are not substitutes for structural content or causal inference, our results suggest that they can meaningfully complement structural knowledge, particularly when regulators have the ability to target ex post interventions.

# References


**Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi**, "Systemic risk and stability in financial networks," *American Economic Review*, 2015, *105* (2), 564–608.

**Acharya, Viral V, Lasse H Pedersen, Thomas Philippon, and Matthew Richardson**, "Measuring systemic risk," *The review of financial studies*, 2017, *30* (1), 2–47.

**Adrian, Tobias and Hyun Song Shin**, "Procyclical leverage and value-at-risk," *The Review of Financial Studies*, 2014, *27* (2), 373–403.

**Aramonte, Sirio, Andreas Schrimpf, and Hyun Song Shin**, "Non-bank financial intermediaries and financial stability," *BIS Working Paper*, 2023.

**Athey, Susan**, "Beyond prediction: Using big data for policy problems," *Science*, 2017, *355* (6324), 483–485.

**Athey, Susan**, "The impact of machine learning on economics," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 507–547.

**Athey, Susan and Guido W Imbens**, "Machine learning methods that economists should know about," *Annual Review of Economics*, 2019, *11* (1), 685–725.

**Athey, Susan and Stefan Wager**, "Policy learning with observational data," *Econometrica*, 2021, *89* (1), 133–161.

**Barro, Robert J and David B Gordon**, "A positive theory of monetary policy in a natural rate model," *Journal of political economy*, 1983, *91* (4), 589–610.

**Bernanke, Ben S and Mark Gertler**, "Agency costs, collateral, and business fluctuations," 1986.

**Bianchi, Javier**, "Overborrowing and Systemic Externalities in the Business Cycle," *American Economic Review*, December 2011, *101* (7), 3400–3426.

**Bianchi, Javier**, "Efficient Bailouts?," *American Economic Review*, December 2016, *106* (12), 3607–59.

**Bianchi, Javier and Enrique G. Mendoza**, "Optimal Time-Consistent Macroprudential Policy," *Journal of Political Economy*, 2018, *126* (2), 588–634.

**Brainard, William C and James Tobin**, "Pitfalls in financial model building," *The American economic review*, 1968, *58* (2), 99–122.

**Brunnermeier, Markus K and Martin Oehmke**, "Bubbles, financial crises, and systemic risk," *Handbook of the Economics of Finance*, 2013, *2*, 1221–1288.

**Bryzgalova, Svetlana, Victor DeMiguel, Sicong Li, and Markus Pelger**, "Asset-pricing factors with economic targets," *Available at SSRN*, 2023, *4344837*.

**Burns, Arthur F and Wesley C Mitchell**, *Measuring business cycles*, National bureau of economic research, 1946.

**Caballero, Ricardo J. and Arvind Krishnamurthy**, "International and domestic collateral constraints in a model of emerging market crises," *Journal of Monetary Economics*, December 2001, *48* (3), 513–548.

**Caravello, Tomás, Alisdair McKay, and Christian Wolf**, "Evaluating Monetary Policy Counterfactuals:(When) Do We Need Structural Models?," *Working Paper*, 2025.

**Chari, V. V. and Patrick J. Kehoe**, "Bailouts, Time Inconsistency, and Optimal Regulation: A Macroeconomic View," *American Economic Review*, September 2016, *106* (9), 2458–93.

**Chen, Luyang, Markus Pelger, and Jason Zhu**, "Deep learning in asset pricing," *Management Science*, 2024, *70* (2), 714–750.

**Chen, Qi, Itay Goldstein, and Wei Jiang**, "Payoff complementarities and financial fragility: Evidence from mutual fund outflows," *Journal of financial economics*, 2010, *97* (2), 239–262.

**Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans**, "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of political Economy*, 2005, *113* (1), 1–45.

**Clayton, Christopher and Andreas Schaab**, "Multinational Banks and Financial Stability," *The Quarterly Journal of Economics*, 01 2022, *137* (3), 1681–1736.

**Clayton, Christopher and Andreas Schaab**, "Bail-Ins, Optimal Regulation, and Crisis Resolution," *The Review of Financial Studies*, 03 2025, p. hhaf002.

**Coppola, Antonio**, "In safe hands: The financial and real impact of investor composition over the credit cycle," *The Review of Financial Studies*, 2025, p. hhaf017.

**Coppola, Antonio, Matteo Maggiori, Brent Neiman, and Jesse Schreger**, "Redrawing the Map of Global Capital Flows: The Role of Cross-Border Financing and Tax Havens," *The Quarterly Journal of Economics*, 2021, *136* (3), 1499–1556.

**Coval, Joshua and Erik Stafford**, "Asset fire sales (and purchases) in equity markets," *Journal of Financial Economics*, 2007, *86* (2), 479–512.

**Derrow-Pinion, Austin, Jennifer She, David Wong, Oliver Lange, Todd Hester, Luis Perez, Marc Nunkesser, Seongjae Lee, Xueying Guo, Brett Wiltshire et al.**, "Eta prediction with graph neural networks in google maps," in "Proceedings of the 30th ACM international conference on information & knowledge management" 2021, pp. 3767–3776.

**Dick-Nielsen, Jens**, "How to clean enhanced TRACE data," *Available at SSRN 2337908*, 2014.

**Dolphin, Rian, Barry Smyth, and Ruihai Dong**, "Stock embeddings: Learning distributed representations for financial assets," *arXiv preprint arXiv:2202.08968*, 2022.

**Dávila, Eduardo and Anton Korinek**, "Pecuniary Externalities in Economies with Financial Frictions," *The Review of Economic Studies*, 02 2017, *85* (1), 352–395.

**Einav, Liran and Jonathan Levin**, "The data revolution and economic analysis," *Innovation Policy and the Economy*, 2014, *14* (1), 1–24.

**Einav, Liran and Jonathan Levin**, "Economics in the age of big data," *Science*, 2014, *346* (6210), 1243089.

**Elliott, Matthew, Benjamin Golub, and Matthew O Jackson**, "Financial networks and contagion," *American Economic Review*, 2014, *104* (10), 3115–3153.

**Fang, Xiang, Bryan Hardy, and Karen K Lewis**, "Who holds sovereign debt and why it matters," *The Review of Financial Studies*, 2025, p. hhaf031.

**Farboodi, Maryam and Laura Veldkamp**, "Long-run growth of financial data technology," *American Economic Review*, 2020, *110* (8), 2485–2523.

**Farhi, Emmanuel and Iván Werning**, "A theory of macroprudential policies in the presence of nominal rigidities," *Econometrica*, 2016, *84* (5), 1645–1704.

**Faria-e Castro, Miguel, Joseba Martinez, and Thomas Philippon**, "Runs versus lemons: Information disclosure and fiscal capacity," *The Review of Economic Studies*, 2017, *84* (4), 1683–1707.

**Fisher, Irving**, "The debt-deflation theory of great depressions," *Econometrica: Journal of the Econometric Society*, 1933, pp. 337–357.

**Fontanier, Paul**, "Optimal policy for behavioral financial crises," *Journal of Financial Economics*,

2025, *166*, 104005.

**Friedman, Milton**, "The methodology of positive economics," *Essays In Positive Economics*, 1953.

**Gabaix, Xavier, Ralph SJ Koijen, Robert Richmond, and Motohiro Yogo**, "Asset embeddings," *Available at SSRN 4507511*, 2024.

**Gabaix, Xavier, Ralph SJ Koijen, Robert Richmond, and Motohiro Yogo**, "Upgrading Credit Pricing and Risk Assessment through Embeddings," *Available at SSRN*, 2025.

**Giglio, Stefano, Bryan Kelly, and Dacheng Xiu**, "Factor models, machine learning, and asset pricing," *Annual Review of Financial Economics*, 2022, *14* (1), 337–368.

**Gillis, Talia B and Jann L Spiess**, "Big data and discrimination," *The University of Chicago Law Review*, 2019, *86* (2), 459–488.

**Goldstein, Itay and Yaron Leitner**, "Stress tests and information disclosure," *Journal of Economic Theory*, 2018, *177*, 34–69.

**Goldstein, Itay, Hao Jiang, and David T Ng**, "Investor flows and fragility in corporate bond funds," *Journal of Financial Economics*, 2017, *126* (3), 592–613.

**Gu, Shihao, Bryan Kelly, and Dacheng Xiu**, "Empirical asset pricing via machine learning," *The Review of Financial Studies*, 2020, *33* (5), 2223–2273.

**Gu, Shihao, Bryan Kelly, and Dacheng Xiu**, "Autoencoder asset pricing models," *Journal of Econometrics*, 2021, *222* (1), 429–450.

**Guren, Adam, Alisdair McKay, Emi Nakamura, and Jón Steinsson**, "What do we learn from cross-regional empirical estimates in macroeconomics?," *NBER Macroeconomics Annual*, 2021, *35* (1), 175–223.

**Haavelmo, Trygve**, "The probability approach in econometrics," *Econometrica: Journal of the Econometric Society*, 1944, pp. iii–115.

**Haddad, Valentin, Alan Moreira, and Tyler Muir**, "When selling becomes viral: Disruptions in debt markets in the COVID-19 crisis and the Fed's response," *The Review of Financial Studies*, 2021, *34* (11), 5309–5351.

**Hamilton, Will, Zhitao Ying, and Jure Leskovec**, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, 2017, *30*.

**Hendrycks, Dan and Kevin Gimpel**, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

**Jiang, Dejun, Zhenxing Wu, Chang-Yu Hsieh, Guangyong Chen, Ben Liao, Zhe Wang, Chao Shen, Dongsheng Cao, Jian Wu, and Tingjun Hou**, "Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models," *Journal of cheminformatics*, 2021, *13*, 1–23.

**Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko et al.**, "Highly accurate protein structure prediction with AlphaFold," *nature*, 2021, *596* (7873), 583–589.

**Kindleberger, Charles and Robert Aliber**, *Manias, panics and crashes: a history of financial crises*, Springer, 1978.

**Kingma, Diederik P**, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

**Kiyotaki, Nobuhiro and John Moore**, "Credit cycles," *Journal of political economy*, 1997, *105* (2), 211–248.

**Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan**, "Human decisions and machine predictions," *The quarterly journal of economics*, 2018, *133* (1), 237–293.

**Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer**, "Prediction policy problems," *American Economic Review*, 2015, *105* (5), 491–495.

**Koopmans, Tjalling C**, "Measurement without theory," *The Review of Economics and Statistics*, 1947, *29* (3), 161–172.

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

**Krishnamurthy, Arvind and Tyler Muir**, "How credit cycles across a financial crisis," *The Journal of Finance*, 2025, *80* (3), 1339–1378.

**Laffont, Jean-Jacques and Jean Tirole**, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 1986, *94* (3), 614–641.

**Leitner, Yaron and Basil Williams**, "Model secrecy and stress tests," *The Journal of Finance*, 2023, *78* (2), 1055–1095.

**Lorenzoni, Guido**, "Inefficient Credit Booms," *The Review of Economic Studies*, 07 2008, *75* (3), 809–833.

**Lucas, Robert E**, "Econometric policy evaluation: A critique," in "Carnegie-Rochester conference series on public policy," Vol. 1 North-Holland 1976, pp. 19–46.

**Ma, Yiming, Kairong Xiao, and Yao Zeng**, "Mutual fund liquidity transformation and reverse flight to liquidity," *The Review of Financial Studies*, 2022, *35* (10), 4674–4711.

**Maron, Haggai, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman**, "Invariant and equivariant graph networks," *arXiv preprint arXiv:1812.09902*, 2018.

**McKay, Alisdair and Christian K Wolf**, "What can time-series regressions tell us about policy counterfactuals?," *Econometrica*, 2023, *91* (5), 1695–1725.

**Mullainathan, Sendhil and Jann Spiess**, "Machine learning: an applied econometric approach," *Journal of Economic Perspectives*, 2017, *31* (2), 87–106.

**Nagel, Stefan**, *Machine learning in asset pricing*, Princeton University Press, 2021.

**Nakamura, Emi and Jón Steinsson**, "Identification in macroeconomics," *Journal of Economic Perspectives*, 2018, *32* (3), 59–86.

**Orlov, Dmitry, Pavel Zryumov, and Andrzej Skrzypacz**, "The design of macroprudential stress tests," *The Review of Financial Studies*, 2023, *36* (11), 4460–4501.

**Parlatore, Cecilia and Thomas Philippon**, "Designing stress scenarios," *The Journal of Finance*, 2025, *80* (2), 833–873.

**Sarkar, Suproteem K**, "Economic representations," 2025.

**Scarselli, Franco, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini**, "The graph neural network model," *IEEE transactions on neural networks*, 2008, *20* (1), 61–80.

**Schularick, Moritz and Alan M Taylor**, "Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008," *American Economic Review*, 2012, *102* (2), 1029–1061.

**Shapiro, Joel and David Skeie**, "Information management in banking crises," *The Review of*

*Financial Studies*, 2015, *28* (8), 2322–2363.

**Stein, Jeremy C.**, "Monetary Policy as Financial Stability Regulation*," *The Quarterly Journal of Economics*, 01 2012, *127* (1), 57–95.

**Tobias, Adrian and Markus K Brunnermeier**, "CoVaR," *The American Economic Review*, 2016, *106* (7), 1705.

**Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin**, "Attention is all you need," *Advances in neural information processing systems*, 2017, *30*.

**Williamson, Oliver E**, "Corporate finance and corporate governance," *The journal of finance*, 1988, *43* (3), 567–591.

**Wu, Zonghan, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu**, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, 2020, *32* (1), 4–24.

**Xu, Keyulu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka**, "How powerful are graph neural networks?," *arXiv preprint arXiv:1810.00826*, 2018.

**Zaheer, Manzil, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola**, "Deep sets," *Advances in neural information processing systems*, 2017, *30*.

Online Appendix for
"Financial Regulation and AI: A Faustian Bargain?"

Christopher Clayton  Antonio Coppola

October 2025

# A Proofs

## A.1 Proof of Lemma 1

Intermediary $i$'s Lagrangian over $\ell_i$ is

$$\mathcal{L}_i = q_i^T(R_i - p_i) - \ell_i^T(R_i + \tau_i - \gamma) - \frac{1}{2}q_i^T H_i^q q_i - \frac{1}{2}\ell_i^T H_i^\ell \ell_i - \lambda_i^T(A_i^\ell \ell_i - A_i^q q_i - \rho_i),$$

yielding FOCs

$$0 = -(R_i + \tau_i - \gamma) - H_i^\ell \ell_i - A_i^{\ell T}\lambda_i,$$

$$\ell_i = -H_i^{\ell-1}(R_i + \tau_i - \gamma) - H_i^{\ell-1}A_i^{\ell T}\lambda_i.$$

Assuming the rollover constraint binds,

$$A_i^\ell\Big[ - H_i^{\ell-1}(R_i + \tau_i - \gamma) - H_i^{\ell-1}A_i^{\ell T}\lambda_i\Big] = A_i^q q_i + \rho_i,$$

$$-\lambda_i = (A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}A_i^\ell H_i^{\ell-1}(R_i + \tau_i - \gamma) + (A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}A_i^q q_i + (A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}\rho_i,$$

and so we have

$$\begin{aligned}\ell_i = {} & H_i^{\ell-1}A_i^{\ell T}(A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}\rho_i + H_i^{\ell-1}A_i^{\ell T}(A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}A_i^q q_i \\ & + \Big[H_i^{\ell-1}A_i^{\ell T}(A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}A_i^\ell H_i^{\ell-1} - H_i^{\ell-1}\Big](R_i + \tau_i - \gamma).\end{aligned}$$

We therefore write

$$\ell_i = \overline{\ell}_i^0 + \Lambda_i^q q_i - \Lambda_i^\tau(\tau_i - \gamma),$$

where

$$\overline{\ell}_i^0 = H_i^{\ell-1}A_i^{\ell T}(A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}\rho_i - \Lambda_i^\tau R_i,$$

$$\Lambda_i^q = H_i^{\ell-1}A_i^{\ell T}(A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}A_i^q,$$

$$\Lambda_i^\tau = H_i^{\ell-1} - H_i^{\ell-1}A_i^{\ell T}(A_i^\ell H_i^{\ell-1}A_i^{\ell T})^{-1}A_i^\ell H_i^{\ell-1}.$$

Arbitrageur maximization yields

$$\gamma = \overline{\gamma} - \Gamma L,$$

$$\gamma = \overline{\gamma} - \Gamma \sum_i \overline{\ell}_i^0 - \Gamma \sum_i \Lambda_i^q q_i + \Gamma \sum_i \Lambda_i^\tau (\tau_i - \gamma),$$

$$\gamma = \left(\mathbb{I} + \Gamma \sum_i \Lambda_i^\tau\right)^{-1} \left(\overline{\gamma} - \Gamma \sum_i \overline{\ell}_i^0\right) - \sum_i \left[\left(\mathbb{I} + \Gamma \sum_i \Lambda_i^\tau\right)^{-1} \Gamma \Lambda_i^q q_i - \left(\mathbb{I} + \Gamma \sum_i \Lambda_i^\tau\right)^{-1} \Gamma \Lambda_i^\tau \tau_i\right],$$

$$\gamma = \overline{\gamma}^e - \Gamma^e \sum_i \left[\Lambda_i^q q_i^* - \Lambda_i^\tau \tau_i^*\right],$$

where we have defined $\overline{\gamma}^e = \left(\mathbb{I} + \Gamma \sum_i \Lambda_i^\tau\right)^{-1} \left(\overline{\gamma} - \Gamma \sum_i \overline{\ell}_i^0\right)$ and $\Gamma^e = \left(\mathbb{I} + \Gamma \sum_i \Lambda_i^\tau\right)^{-1} \Gamma$. Finally substituting back in,

$$\ell_i = \overline{\ell}_i + \Lambda_i^q q_i - \Lambda_i^\tau \tau_i - \sum_i \left[\Lambda_i^{q,e} q_i^* - \Lambda_i^{\tau,e} \tau_i^*\right],$$

where $\overline{\ell}_i = \overline{\ell}_i^0 + \Lambda_i^\tau \overline{\gamma}^e$, $\Lambda_i^{q,e} = \Lambda_i^\tau \Gamma^e \Lambda_i^q$, and $\Lambda_i^{\tau,e} = \Lambda_i^\tau \Gamma^e \Lambda_i^\tau$.

## A.2 Proof of Proposition 1

From Lemma 1, equilibrium liquidations given $(q, \tau)$ are

$$\ell_i = \overline{\ell}_i + \Lambda_i^q q_i - \Lambda_i^\tau \tau_i - \sum_i \left[\Lambda_i^{q,e} q_i - \Lambda_i^{\tau,e} \tau_i\right].$$

We define $\overline{\Lambda}_i^\tau = (e_i \otimes I_N)^T \Lambda_i^\tau - (\Lambda_1^{\tau,e}, \ldots, \Lambda_I^{\tau,e})$ (where $e_i$ is the standard basis vector whose $i^{th}$ element is 1 and $\otimes$ is the Kronecker product) and similarly define $\overline{\Lambda}_i^q = (e_i \otimes I_N)^T \Lambda_i^q q_i - (\Lambda_1^{q,e}, \ldots, \Lambda_I^{q,e})$. We then write

$$\ell_i = \overline{\ell}_i + \overline{\Lambda}_i^q q - \overline{\Lambda}_i^\tau \tau.$$

Stacking over $i$, we have

$$\ell = \overline{\ell} + \overline{\Lambda}^q q - \overline{\Lambda}^\tau \tau.$$

Next, we can write the regulator's objective as

$$\mathbb{E}\left[\sum U_i - \delta(\tau) \middle| s, M\right] = \mathbb{E}\left[\sum_i \mathcal{L}_i \middle| s, M\right] + \mathbb{E}\left[\tau^T \ell \middle| s, M\right] - \frac{1}{2} \tau^T \Delta \tau,$$

where as in the proof of Lemma 1 we have defined the intermediary $i$ Lagrangian by

$$\mathcal{L}_i = q_i^T (R_i - p_i) - \ell_i^T (R_i + \tau_i - \gamma) - \frac{1}{2} q_i^T H_i^q q_i - \frac{1}{2} \ell_i^T H_i^\ell \ell_i - \lambda_i^T (A_i^\ell \ell_i - A_i^q q_i - \rho_i).$$

The regulator thus solves

$$\max_\tau \mathbb{E}\left[\sum_i \mathcal{L}_i \middle| s, M\right] + \mathbb{E}\left[\tau^T \ell \middle| s, M\right] - \frac{1}{2} \tau^T \Delta \tau$$

subject to

$$\ell = \overline{\ell} + \overline{\Lambda}^q q - \overline{\Lambda}^\tau \tau,$$

$$\gamma = \overline{\gamma} - \Gamma L.$$

By the Envelope Theorem, we have the regulator's FOC given by

$$0 = \mathbb{E}\Big[ -\ell + \sum_i \ell_i \frac{\partial \gamma}{\partial \tau} \Big| s, M \Big] + \mathbb{E}\Big[ \frac{\partial \tau^T \ell}{\partial \tau} \Big| s, M \Big] - \Delta\tau,$$

$$0 = \mathbb{E}\Big[ \sum_i (\overline{\Lambda}_i^\tau)^T \Gamma^T L \Big| s, M \Big] + \mathbb{E}\Big[ -\overline{\Lambda}^{\tau T} \tau \Big| s, M \Big] - \Delta\tau.$$

Writing $L = L(q) - \sum_i (\overline{\Lambda}_i^\tau)\tau$ and using the assumption that $\Gamma$ is symmetric,

$$\mathbb{E}[\overline{\Lambda}^{\tau T} | s, M] + \Delta\tau = \mathbb{E}\Big[ \sum_i (\overline{\Lambda}_i^\tau)^T \Gamma \Big( L(q) - \sum_i (\overline{\Lambda}_i^\tau)\tau \Big) \Big| s, M \Big],$$

$$\tau = \mathbb{E}\Big[ \Xi \Big| s, M \Big]^{-1} \mathbb{E}\Big[ (\sum_i \overline{\Lambda}_i^\tau)^T \Gamma L(q) \Big| s, M \Big],$$

where $\Xi = \overline{\Lambda}^{\tau T} + (\sum_i \overline{\Lambda}_i^\tau)^T \Gamma^T (\sum_i \overline{\Lambda}_i^\tau) + \Delta$.

## A.3 Proof of Proposition 2

As in the proof of Proposition 1, the regulator's ex-ante expected welfare given positions $q$, a model $M$, and a signal $s$ is

$$V(q, M, s) = \mathbb{E}\Big[ q^T (R - p) - \frac{1}{2} q^T H^q q \Big| s, M \Big] + \mathbb{E}\Big[ \sum_i \Big[ -\ell_i^T (R_i - \gamma) - \frac{1}{2} \ell_i^T H_i^\ell \ell_i \Big] - \frac{1}{2} \tau^T \Delta \tau \Big| s, M \Big],$$

where we define $H^q$ to be the block diagonal matrix whose diagonal elements are $H_i^q$. We define $\theta_i(q) = R_i - \overline{\gamma} + \Gamma L(q)$ to obtain:

$$V(q, M, s) = \mathbb{E}\Big[ q^T (R - p) - \frac{1}{2} q^T H^q q \Big| s, M \Big] + \mathbb{E}\Big[ -\ell^T \theta(q) + L^T \Gamma \sum_i (\overline{\Lambda}_i^\tau)\tau - \frac{1}{2} \ell_i^T H_i^\ell \ell_i - \frac{1}{2} \tau^T \Delta \tau \Big| s, M \Big],$$

$$\begin{aligned} V(q, M, s) = \mathbb{E}\Big[ & q^T (R - p) - \frac{1}{2} q^T H^q q - \ell(q)^T \theta(q) \Big| s, M \Big] + \mathbb{E}\Big[ \tau^T \overline{\Lambda}^{\tau T} \theta(q) \\ & + \tau^T (\sum_i \overline{\Lambda}_i^\tau)^T \Gamma \Big( L(q) - (\sum_i \overline{\Lambda}_i^\tau)\tau \Big) - \frac{1}{2} \ell_i^T H_i^\ell \ell_i - \frac{1}{2} \tau^T \Delta \tau \Big| s, M \Big], \end{aligned}$$

$$\begin{aligned} V(q, M, s) = \mathbb{E}\Big[ & q^T (R - p) - \frac{1}{2} q^T H^q q - \ell(q)^T \theta(q) \Big| s, M \Big] + \mathbb{E}\Big[ \tau^T \Big( \overline{\Lambda}^{\tau T} \theta(q) \\ & + (\sum_i \overline{\Lambda}_i^\tau)^T \Gamma L(q) \Big) - \tau^T \Big( (\sum_i \overline{\Lambda}_i^\tau)^T \Gamma (\sum_i \overline{\Lambda}_i^\tau) + \frac{1}{2} \Delta \Big) \tau - \frac{1}{2} \ell^T H^\ell \ell \Big| s, M \Big]. \end{aligned}$$

From here, we can expand out

$$\ell^T H^\ell \ell = \left(\ell(q)^T - \tau^T \overline{\Lambda}^{\tau T}\right) H^\ell \left(\ell(q) - \overline{\Lambda}^\tau \tau\right) = \ell(q)^T H^\ell \ell(q) - 2\tau^T \overline{\Lambda}^{\tau T} H^\ell \ell(q) + \tau^T \overline{\Lambda}^{\tau T} H^\ell \overline{\Lambda}^\tau \tau,$$

and substitute back in to obtain

$$\begin{aligned} V(q, M, s) = \mathbb{E}\Bigg[ & q^T(R-p) - \frac{1}{2} q^T H^q q - \ell(q)^T \theta(q) - \frac{1}{2}\ell(q)^T H^\ell \ell(q) \Bigg| s, M \Bigg] \\ & + \mathbb{E}\Bigg[ \tau^T \left( \overline{\Lambda}^{\tau T} \theta(q) + (\sum_i \overline{\Lambda}_i^\tau)^T \Gamma L(q) + \overline{\Lambda}^{\tau T} H^\ell \ell(q) \right) \\ & - \frac{1}{2}\tau^T \left( 2(\sum_i \overline{\Lambda}_i^\tau)^T \Gamma (\sum_i \overline{\Lambda}_i^\tau) + \Delta + \overline{\Lambda}^{\tau T} H^\ell \overline{\Lambda}^\tau \right) \tau \Bigg| s, M \Bigg]. \end{aligned}$$

Note that this welfare function is necessarily equivalent to that obtained in Proposition 1. Therefore, an optimal tax formula that is equivalent to that in Proposition 1 yields:

$$\tau = \mathbb{E}\left[ 2(\sum_i \overline{\Lambda}_i^\tau)^T \Gamma (\sum_i \overline{\Lambda}_i^\tau) + \Delta + \overline{\Lambda}^{\tau T} H^\ell \overline{\Lambda}^\tau \Big| s, M \right]^{-1} \mathbb{E}\left[ \overline{\Lambda}^{\tau T} \theta(q) + (\sum_i \overline{\Lambda}_i^\tau)^T \Gamma L(q) + \overline{\Lambda}^{\tau T} H^\ell \ell(q) \Big| s, M \right].$$

Accordingly substituting back into welfare, we have

$$\begin{aligned} V(q, M, s) = & \mathbb{E}\Bigg[ q^T(R-p) - \frac{1}{2} q^T H^q q - \ell(q)^T \theta(q) - \frac{1}{2}\ell(q)^T H^\ell \ell(q) \Bigg| s, M \Bigg] \\ & + \frac{1}{2}\tau^T \mathbb{E}\left[ 2(\sum_i \overline{\Lambda}_i^\tau)^T \Gamma (\sum_i \overline{\Lambda}_i^\tau) + \Delta + \overline{\Lambda}^{\tau T} H^\ell \overline{\Lambda}^\tau \Big| s, M \right] \tau. \end{aligned}$$

We can then characterize the expected welfare gains by

$$\begin{aligned} V(q, M) = & \mathbb{E}_0\Bigg[ q^T(R-p) - \frac{1}{2} q^T H^q q - \ell(q)^T \theta(q) - \frac{1}{2}\ell(q)^T H^\ell \ell(q) \Bigg| M \Bigg] \\ & + \frac{1}{2}\mathbb{E}_0\left[ \tau^T \mathbb{E}\left[ 2(\sum_i \overline{\Lambda}_i^\tau)^T \Gamma (\sum_i \overline{\Lambda}_i^\tau) + \Delta + \overline{\Lambda}^{\tau T} H^\ell \overline{\Lambda}^\tau \Big| s, M \right] \tau \Big| M \right], \end{aligned}$$

where $\tau$ is given as in Proposition 1.

From here, we define $\Psi = \mathbb{E}\left[ 2(\sum_i \overline{\Lambda}_i^\tau)^T \Gamma (\sum_i \overline{\Lambda}_i^\tau) + \Delta + \overline{\Lambda}^{\tau T} H^\ell \overline{\Lambda}^\tau \Big| s, M \right]$, and note that $\Psi^T = \Psi$. We can then write more compactly,

$$\begin{aligned} V(q, M) = & \mathbb{E}_0\Bigg[ q^T(R-p) - \frac{1}{2} q^T H^q q - \ell(q)^T \theta(q) - \frac{1}{2}\ell(q)^T H^\ell \ell(q) \Bigg| M \Bigg] \\ & + \frac{1}{2}\mathbb{E}_0\left[ \tau^T \Psi \tau \Big| M \right]. \end{aligned}$$

**Predictive-Causal Independence.** Under predictive-causal independence, we have

$$\tau^* = \mathbb{E}\left[\Xi\middle|s, M\right]^{-1}\mathbb{E}\left[(\sum_i \overline{\Lambda}_i^\tau)^T\Gamma\middle|s, M\right]\mathbb{E}\left[L(q)\middle|s, M\right].$$

We therefore define $\Upsilon = \mathbb{E}\left[\Xi\middle|s, M\right]^{-1}\mathbb{E}\left[(\sum_i \overline{\Lambda}_i^\tau)^T\Gamma\middle|s, M\right]$ and define $\Theta = \Upsilon^T\Psi\Upsilon$. Then, defining $\hat{L}(q) = \mathbb{E}[L(q)|s, M]$ to be predicted liquidations, we have:

$$\begin{aligned}V(q, M) =&\mathbb{E}_0\left[q^T(R-p) - \frac{1}{2}q^T H^q q - \ell(q)^T\theta(q) - \frac{1}{2}\ell(q)^T H^\ell \ell(q)\middle|M\right]\\&+\frac{1}{2}\mathbb{E}_0\left[\hat{L}(q)^T\Theta\hat{L}(q)\middle|M\right].\end{aligned}$$

Using predictive-causal independence, we have:

$$\begin{aligned}V(q, M) =&\mathbb{E}_0\left[q^T(R-p) - \frac{1}{2}q^T H^q q - \ell(q)^T\theta(q) - \frac{1}{2}\ell(q)^T H^\ell \ell(q)\middle|M\right]\\&+\frac{1}{2}\mathbb{E}_0\left[\hat{L}(q)^T\Theta_0\hat{L}(q)\middle|M\right].\end{aligned}$$

Finally, expanding the second line, we obtain

$$\begin{aligned}V(q, M) =&\mathbb{E}_0\left[q^T(R-p) - \frac{1}{2}q^T H^q q - \ell(q)^T\theta(q) - \frac{1}{2}\ell(q)^T H^\ell \ell(q)\middle|M\right]\\&+\frac{1}{2}L_0(q)^T\Theta_0 L(q) + \frac{1}{2}\text{tr}(\Theta_0 cov_0(\hat{L}(q))).\end{aligned}$$

## A.4 Proof of Proposition 3

The optimal model choice solves

$$\max_{\Sigma_0^{\hat{L}}}\frac{1}{2}\text{tr}\left(\Theta_0\Sigma_0^{\hat{L}}\right) - C(\Sigma_0^{\hat{L}}, q)$$

subject to the restriction that $\Sigma_0^{\hat{L}}$ is symmetric. Taking the first order condition, we have

$$\frac{\partial C(\Sigma_0^{\hat{L}}, q)}{\partial\Sigma_0^{\hat{L}}} + \left(\frac{\partial C(\Sigma_0^{\hat{L}}, q)}{\partial\Sigma_0^{\hat{L}}}\right)^T = \Theta_0$$

since $\Theta_0$ is symmetric.

## A.5 Proof of Proposition 4

The intermediary $i$ ex ante optimization problem is

$$\max_{q_i} \mathbb{E}_0\Big[\sum_i \Big[q_i^T(R_i - p_i) - \ell_i^T(R_i + \tau_i^* - \gamma) - \frac{1}{2}q_i^T H_i^q q_i - \frac{1}{2}\ell_i^T H_i^\ell \ell_i\Big],$$

where we have

$$\ell_i = \overline{\ell}_i + \Lambda_i^q q_i - \sum_j \Lambda_j^{q,e} q_j^* - \overline{\Lambda}_i^\tau \tau^*.$$

So, we have the FOC

$$0 = \mathbb{E}_0\Big[R_i - p_i - \Lambda_i^{qT}(R_i + \tau_i^* - \gamma) - H_i^q q_i - \Lambda_i^{qT} H_i^\ell \ell_i\Big].$$

Substituting in for $\ell_i$, we have

$$\mathbb{E}_0\Big[H_i^q\Big]q_i = \mathbb{E}_0\Big[R_i - p_i - \Lambda_i^{qT}(R_i + \tau_i^* - \gamma) - \Lambda_i^{qT} H_i^\ell\Big(\overline{\ell}_i + \Lambda_i^q q_i - \sum_j \Lambda_j^{q,e} q_j^* - \overline{\Lambda}_i^\tau \tau^*\Big)\Big],$$

$$\mathbb{E}_0\Big[H_i^q + \Lambda_i^{qT} H_i^\ell \Lambda_i^q\Big]q_i = \mathbb{E}_0\Big[R_i - p_i - \Lambda_i^{qT}(R_i - \gamma) - \Lambda_i^{qT} H_i^\ell\Big(\overline{\ell}_i - \sum_j \Lambda_j^{q,e} q_j^*\Big)\Big] + \mathbb{E}_0\Big[-\Lambda_i^{qT}\tau_i^* + \Lambda_i^{qT} H_i^\ell \overline{\Lambda}_i^\tau \tau^*\Big].$$

We use that $R_i - \gamma = \theta_i(q^*) - \Gamma(\sum_i \overline{\Lambda}_i^\tau)\tau^*$ to obtain

$$\mathbb{E}_0\Big[H_i^q q_i^* + \Lambda_i^{qT} H_i^\ell \overline{\ell}_i(q^*)\Big] = \mathbb{E}_0\Big[R_i - p_i - \Lambda_i^{qT}\theta_i(q^*)\Big] - \mathbb{E}_0\Big[\Lambda_i^{qT}\Big(\tau_i^* - \Big(H_i^\ell \overline{\Lambda}_i^\tau + \Lambda_i^{qT}\Gamma(\sum_i \overline{\Lambda}_i^\tau)\Big)\tau^*\Big].$$

## A.6 Proof of Proposition 5

We have that:

$$V(q, M) = \mathbb{E}\Big[q^T(R - p) - \frac{1}{2}q^T H^q q - \ell(q)^T\theta(q) - \frac{1}{2}\ell(q)^T H^\ell \ell(q)\Big] + \frac{1}{2}L_0(q)\Theta_0 L_0(q) + \frac{1}{2}\mathrm{tr}(\Theta_0 cov_0(\hat{L}(q)).$$

Recalling that $\ell_i(q) = \overline{\ell}_i + \overline{\Lambda}_i^q q$, $\theta_i(q) = R_i - \overline{\gamma} + \Gamma L(q)$, and $L(q) = \sum_i \overline{\ell}_i + \sum_i \overline{\Lambda}_i^q q$, then we can expand:

$$\sum_i \ell_i(q)^T\theta_i(q) = \sum_i \overline{\ell}_i^T\theta_i(0) + q^T\sum_i \overline{\Lambda}_i^{qT}\theta_i(0) + q^T(\sum_j \overline{\Lambda}_j^q)^T\Gamma L(0) + q^T\sum_i\Big(\overline{\Lambda}_i^{qT}\Big)\Big(\Gamma\sum_j \overline{\Lambda}_j^q\Big)q,$$

$$\ell(q)H^\ell\ell(q) = \sum_i \overline{\ell}_i^T H_i^\ell \overline{\ell}_i + 2q^T\Big(\sum_i \overline{\Lambda}_i^{qT} H_i^\ell \overline{\ell}_i\Big) + q^T\Big(\sum_i \overline{\Lambda}_i^{qT} H_i^\ell \overline{\Lambda}_i^q\Big)q.$$

For a predictive model, $\Theta_0$ is invariant to the model choice $M$ and to $q$. Therefore by the

Envelope Theorem,

$$0 = \mathbb{E}_0\Bigg[R - p - H^q q - \Bigg[\sum_i \overline{\Lambda}_i^{qT}\theta_i(0) + (\sum_j \overline{\Lambda}_j^q)^T \Gamma L(0) + 2\sum_i \Big(\overline{\Lambda}_i^{qT}\Big)\Big(\Gamma \sum_j \overline{\Lambda}_j^q\Big)q\Bigg] - \Bigg[\Big(\sum_i \overline{\Lambda}_i^{qT} H_i^\ell \overline{\ell}_i\Big) + \Big(\sum_i \overline{\Lambda}_i^{qT} H_i^\ell \overline{\Lambda}_i^q\Big)q\Bigg]\Bigg] - \frac{dC(\Sigma_0^{\hat{L}}, q)}{dq} + \mathbb{E}_0\Bigg[(\sum_i \overline{\Lambda}_i^q)^T\Bigg]\Theta_0 \mathbb{E}_0[L(q)],$$

$$0 = \mathbb{E}_0\Bigg[R - p - H^q q - \Bigg[\sum_i \overline{\Lambda}_i^{qT}\Big(\theta_i(0) + \Big(\Gamma \sum_j \overline{\Lambda}_j^q\Big)q\Big) + (\sum_j \overline{\Lambda}_j^q)^T \Gamma\Big(L(0) + \sum_j \overline{\Lambda}_j^q q\Big)\Bigg] - \Bigg[\Big(\sum_i \overline{\Lambda}_i^{qT} H_i^\ell \overline{\ell}_i\Big) + \Big(\sum_i \overline{\Lambda}_i^{qT} H_i^\ell \overline{\Lambda}_i^q\Big)q\Bigg]\Bigg] - \frac{dC(\Sigma_0^{\hat{L}}, q)}{dq} + \mathbb{E}_0\Bigg[(\sum_i \overline{\Lambda}_i^q)^T\Bigg]\Theta_0 \mathbb{E}_0[L(q)]\,.$$

We have

$$\theta_i(0) + \Gamma\Big(\sum_j \overline{\Lambda}_j^q\Big)q = \theta_i(q),$$

$$L(0) + \Big(\sum_j \overline{\Lambda}_j^q\Big)q = L(q),$$

so we get:

$$\mathbb{E}_0\Big[H^q q\Big] = \mathbb{E}_0\Big[R - p - \sum_i \overline{\Lambda}_i^{qT}\Big(\theta_i(q) + H_i^\ell \ell_i(q)\Big)\Big] - \mathbb{E}_0\Big[(\sum_j \overline{\Lambda}_j^q)^T \Gamma L(q)\Big] - \frac{dC(\Sigma_0^{\hat{L}}, q)}{dq} + \mathbb{E}_0\Bigg[(\sum_i \overline{\Lambda}_i^q)^T\Bigg]\Theta_0 \mathbb{E}_0[L(q)].$$

Recall that we have

$$\ell_i(q) = \overline{\ell}_i + \Lambda_i^q q_i - \sum_j \Lambda_j^{q,e} q_{ij},$$

then

$$\sum_i \overline{\Lambda}_i^{qT}\Big(\theta_i(q) + H_i^\ell \ell_i(q)\Big) = \Lambda^{qT}\Big(\theta(q) + H^\ell \ell(q)\Big) - \sum_i (\Lambda_1^{q,e}, \ldots, \Lambda_I^{q,e})^T\Big(\theta_i(q) + H_i^\ell \ell_i(q)\Big),$$

so we can then write

$$\mathbb{E}_0\Big[H^q q\Big] = \mathbb{E}_0\Big[R - p - \Lambda^{qT}\Big(\theta(q) + H^\ell \ell(q)\Big)\Big] + \mathbb{E}_0\Big[\sum_i (\Lambda_1^{q,e}, \ldots, \Lambda_I^{q,e})^T\Big(\theta_i(q) + H_i^\ell \ell_i(q)\Big) - (\sum_j \overline{\Lambda}_j^q)^T \Gamma L(q)\Big] - \frac{dC(\Sigma_0^{\hat{L}}, q)}{dq} + \mathbb{E}_0\Big[(\sum_i \overline{\Lambda}_i^q)^T\Big]\Theta_0 \mathbb{E}_0[L(q)].$$

Now, recall that we can write the private FOC, subject to an ex-ante asset tax $t_i$, as

$$\mathbb{E}_0\Big[H_i^q\Big]q_i = \mathbb{E}_0\Big[R_i - (p_i + t_i) - \Lambda_i^{qT}\Big(\theta_i(q^*) + H_i^\ell \ell_i(q)\Big)\Big] - \mathbb{E}_0\Big[\Lambda_i^{qT}\Big(\tau_i - \Big(\Gamma(\sum_i \overline{\Lambda}_i^\tau) + H_i^\ell \overline{\Lambda}_i^\tau\Big)\tau\Big)\Big].$$

Therefore, stacking and combining the two equations we obtain:

$$\begin{aligned} -t - \mathbb{E}_0\Big[\Psi\tau\Big] = \mathbb{E}_0\Big[\sum_i (\Lambda_1^{q,e}, \ldots, \Lambda_I^{q,e})^T\Big(\theta_i(q) + H_i^\ell \ell_i(q)\Big) \\ - (\sum_j \overline{\Lambda}_j^q)^T \Gamma L(q)\Big] - \frac{dC(\Sigma_0^{\hat{L}}, q)}{dq} + \mathbb{E}_0\Big[(\sum_i \overline{\Lambda}_i^q)^T\Big]\Theta_0 \mathbb{E}_0[L(q)], \end{aligned}$$

where $\mathbb{E}_0\Big[\Psi\tau\Big]$ is the stacking of $\mathbb{E}_0\Big[\Lambda_i^{qT}\Big(\tau_i - \Big(\Gamma(\sum_i \overline{\Lambda}_i^\tau) + H_i^\ell \overline{\Lambda}_i^\tau\Big)\tau\Big)\Big]$. Thus defining $\Lambda^{q,e} = (\Lambda_1^{q,e}, \ldots, \Lambda_I^{q,e})$, we have

$$t = \mathbb{E}_0\Big[(\sum_j \overline{\Lambda}_j^q)^T \Gamma L(q) - \Lambda^{q,eT} \sum_i \Big(\theta_i(q) + H_i^\ell \ell_i(q)\Big)\Big].$$

## A.7 Proof of Corollary 2

From Proposition 2, the welfare gain of a predictive model relative to using just the prior is:

$$\Delta V = \frac{1}{2}\operatorname{tr}(\Theta_0 \operatorname{cov}(\widehat{L})).$$

The maximum welfare gain within the class of predictive models is attained by a perfect predictive signal ($\widehat{L} = L$), so:

$$\Delta V_{\max} = \frac{1}{2}\operatorname{tr}(\Theta_0 \operatorname{cov}(L)).$$

Define variance and covariance operators that use $S$ as a weighting matrix:

$$\operatorname{Var}_S(X) := \mathbb{E}\left[(X - E[X])^T S(X - E[X])\right],$$

$$\operatorname{Cov}_S(X, Y) := \mathbb{E}\left[(X - E[X])^T S(Y - E[Y])\right].$$

Then the $S$-weighted correlation between the predictions $\widehat{L}$ and the true liquidations $L$ is:

$$\rho_S \equiv \frac{\operatorname{Cov}_S(L, \widehat{L})}{\sqrt{\operatorname{Var}_S(L)\operatorname{Var}_S(\widehat{L})}}.$$

For $S$ deterministic and since $\mathbb{E}L = \mathbb{E}\widehat{L}$ by the law of iterated expectations, we have that:

$$\operatorname{Cov}_S(L, \widehat{L}) := \mathbb{E}\Big[(L - \mathbb{E}L)^T S(\widehat{L} - \mathbb{E}L)\Big] = \mathbb{E}\left[(\mathbb{E}[L \mid s] - \mathbb{E}L)^T S(\mathbb{E}[L \mid s] - \mathbb{E}L)\right] = \operatorname{Var}_S(\widehat{L}).$$

Therefore:

$$\rho_S^2 = \frac{\mathrm{Var}_S(\widehat{L})^2}{\mathrm{Var}_S(\widehat{L})\,\mathrm{Var}_S(L)} = \frac{\mathrm{Var}_S(\widehat{L})}{\mathrm{Var}_S(L)}.$$

Using the fact that $\mathrm{Var}_S(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^T S (X - \mathbb{E}[X])\right] = \mathrm{tr}(S\,\mathrm{cov}(X))$, we then have that:

$$\rho_S^2 = \frac{\mathrm{tr}(S\,\mathrm{cov}(\widehat{L}))}{\mathrm{tr}(S\,\mathrm{cov}(L))}.$$

Therefore, $\Delta V = \Delta V_{\max} \cdot \rho^2_{\Theta_0}$.