Heterogeneous Intermediaries and Asset Prices: A Semiparametric Approach

Sai Ma*
Federal Reserve Board of Governors

First draft: October 2017
This version: February 2023

Abstract

This paper estimates a general heterogeneous intermediary stochastic discount factor (HI-SDF) that depends nonlinearly on aggregate leverage and net worth distribution among financial intermediaries, and evaluates its ability to fit asset return data. The HI-SDF exhibits substantial explanatory power for cross-sectional variation in expected returns across a wide range of assets, highlighting intermediaries' heterogeneity as an important source of risk. Through the lens of an analytical model, I show that the HI-SDF emerges in equilibrium due to the presence of financial frictions, its nonlinearity originates from heterogeneity in financing constraints, and its functional form hinges on the types of constraints.

JEL: E32, G11, G12, G21, G23. Keywords: intermediary asset pricing, heterogeneity

*I am indebted to Sydney C. Ludvigson for the encouragement and support for this project. I also would like to thank Jaroslav Borovička, Nina Boyarchenko, Tim Christensen, Mark Gertler, Simon Gilchrist, Boyan Jovanovic, Ralph Koijen, Alexi Savov, Tyler Muir, Stijn Van Nieuwerburgh, Venky Venkateswaran, and seminar participants at the CEA, NYU, NYU Stern, University of Guelph, Western University, Ryerson University, Toronto Rotman, UBC Sauder, BU Questrom, Rochester, Bank of Canada, NY Fed, Fed Board, and SF Fed for helpful comments. Sai: International Finance Division, Federal Reserve Board. Washington, D.C. Email: sai.ma@frb.gov. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Board or the Federal Reserve System. All errors are mine.

1 Introduction

A large and growing body of theoretical literature has given financial intermediaries a central role in business cycles and asset pricing.[1] The asset pricing literature has concentrated on the role of intermediaries as marginal investors for the pricing of risky securities. In all of these models, intermediaries are presumed to be more efficient than households at trading risky securities, but they face financial frictions that limit their access to external finance. The marginal utility of intermediaries hence depends on their leverage, which enters the stochastic discount factor (SDF) and determines risk premia in financial markets. Intriguing empirical findings suggest that an SDF that is linear in intermediary leverage is important for the pricing of risky securities (e.g., Adrian, Etula, and Muir (2014) (AEM) and He, Kelly, and Manela (2017) (HKM)).

In all of these papers, intermediaries are modeled as *representative* agents who specialize in trading risky securities and borrow to do so. Under this assumption, only the entire intermediary sector's aggregate leverage matters for the pricing of risky securities. In the data, however, there is considerable heterogeneity among financial intermediaries that is difficult to reconcile with a representative agent framework. The left panel of Figure 1 plots the leverage over time of two types of intermediaries that have been studied in the asset pricing literature: broker-dealers (BD) studied in AEM, and bank holding companies (BHC) studied in HKM. Figure 1 shows that BD leverage is *pro*cyclical and moves inversely with BHC leverage, which is *counter*cyclical. The former exhibits a positive 25% correlation with the innovation in GDP, while the latter exhibits a negative correlation of -18%.[2] The right panel of Figure 1 shows that the distribution of wealth between these two types of intermediaries varies considerably over time: the ratio of BHC equity to the sum of BHC and BD equity ("net worth share" hereafter) is countercyclical and highly volatile. By construction, models with a representative intermediary are silent on such variation's causes and economic consequences.

A natural question arises after the fact: does such heterogeneity matter in pricing assets? This paper proposes a general stochastic discount factor (SDF) model that accounts for intermediary heterogeneity featured in Figure 1 and takes the form,

$$m_t = \pi_t \Omega \left(\psi_t, \phi_t \right),$$

where π_t is the household marginal rate of substitution commonly studied in traditional consumption-

[1]See, for example, Adrian and Boyarchenko (2012), Brunnermeier and Pedersen (2008), Brunnermeier and Sannikov (2014), Gertler and Kiyotaki (2010), and He and Krishnamurthy (2013).

[2]The innovation is computed as the AR(1) residual. Both leverage and GDP are highly persistent in the data. The results are robust using the level of leverage: BD and BHC leverages are 43% and -37% correlated with GDP growth, respectively.

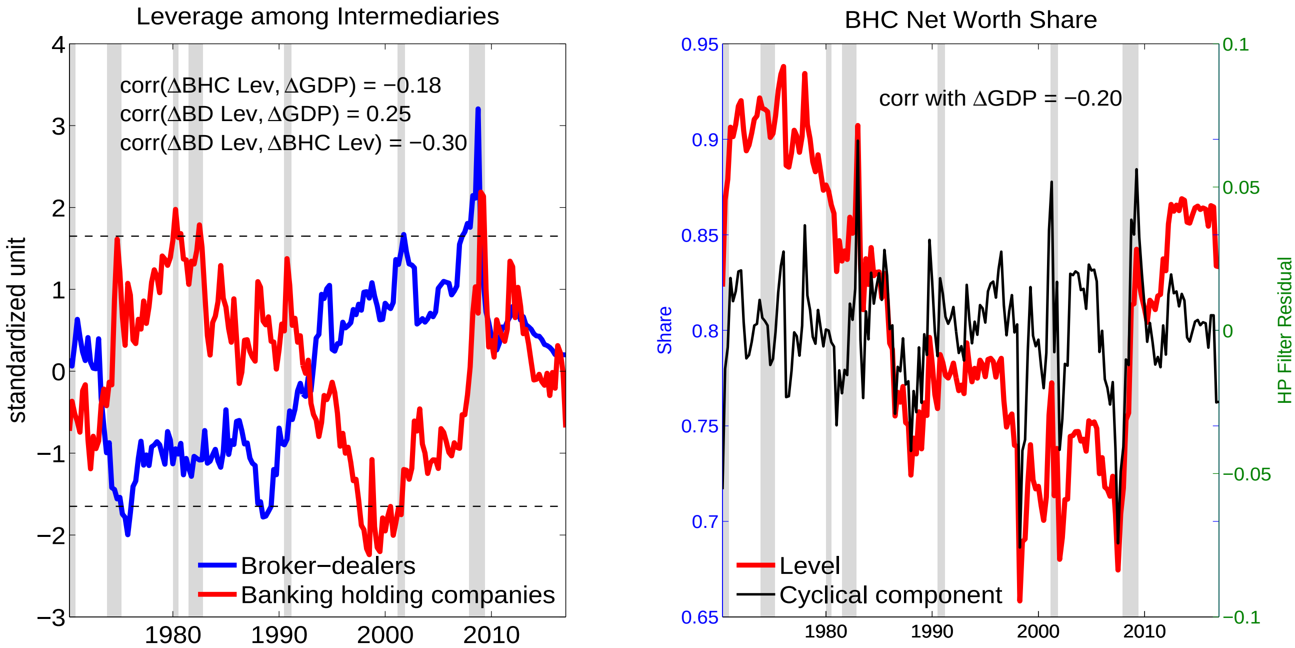

Figure 1: Heterogeneity among Intermediaries. The left panel reports the standardized log leverage of banking holding companies (red) and broker-dealers (blue). BHC leverage is the reciprocal of the capital ratio constructed in HKM and BD leverage is from AEM. The right panel reports the level (red) and HP Filter residual (black) of the net worth share of BHC relative to the sum of BHC and BD equity. Net Worth data of the BHC comes from Flow of Funds Table L.131 and BD comes from Flow of Funds Table L.130. Δ refers to the AR(1) residual. The sample spans 1970Q1 to 2019Q4.

based asset pricing literature (CCAPM), and $\Omega(\psi_t, \phi_t)$ is an additional *pricing wedge* that is a function of two variables attributable to the presence of intermediaries: aggregate intermediary leverage (ϕ_t) and the net worth share among intermediaries (ψ_t). I refer to m_t as the heterogeneous intermediary SDF, or HI-SDF for short.

While these two variables are observable, the functional form of the pricing wedge is generally unknown unless imposing additional parametric assumptions. I let the data speak and take a semiparametric approach to estimate the HI-SDF. More specifically, I impose a constant relative risk aversion (CRRA) function for the household's marginal rate of substitution π_t, but allow the data to dictate the appropriate functional form of the wedge $\Omega(\cdot)$ by estimating that component nonparametrically. I then evaluate the estimated HI-SDF's ability to fit cross-sections of expected returns using both linear and nonlinear tests with a wide range of test assets. These include equity portfolios sorted on the basis of characteristics such as size/book-market sorted portfolios and long-run reversal portfolios, and nonequity assets that include portfolios of corporate bonds, sovereign bonds, options on the market index, and CDS spreads. I further compare the HI-SDF to the classic Lucas-Breeden (Breeden (1979) and Lucas (1978)) consumption-based asset pricing model (CCAPM), and the representative intermediary asset pricing models studied by AEM and HKM.

The latter are special cases of the HI-SDF when the pricing wedge is (log) linear in intermediary's leverage.

The paper's main results can be stated as follows. First, using firm-level data, I show that wealth distribution among intermediaries is an important source of risk for pricing risky securities, which has been largely neglected in the literature. I sort CRSP stocks into portfolios based on their return exposures to the innovations in the net worth share of BHCs. I find that assets in the lowest quintile of net worth share betas have 5.8% higher annualized expected return than those in the highest quintile. Other competing asset-pricing models are not able to explain this spread.

Second, the semiparametrically estimated HI-SDF exhibits substantial explanatory power for the cross-section of expected returns in a wide range of test assets. A log-linearized version of the HI-SDF is able to explain 79% of cross-sectional variation of expected returns on a pooled portfolio that consists of 206 equity plus nonequity assets. The risk prices for the estimated pricing wedge are consistently negative and statistically significant across all individual portfolios, as well as the pooled (equity plus nonequity) test assets, and are of similar magnitudes across all assets. In comparison, a representative intermediary model from AEM or HKM explains 52% and 43% cross-sectional variation, respectively. In addition, the fully nonlinear SDF m_t delivers specification error – measured by Hansen-Jagannathan (HJ) distance – strictly smaller than competing specifications. Specifically, the HJ distance from the HI-SDF model is 23% smaller than the CCAPM and is up to 15% smaller than the SDF based on the representative intermediary model.

Third, the estimated HI-SDF is highly nonlinear in both net worth share and aggregate leverage, and this nonlinearity matters empirically. Relative to its linear counterpart, the nonlinearity improves the cross-sectional $\bar{R}^2$ by 34% and substantially reduces specification errors. The results are robust to a battery of checks and hold *out of sample*.

Next, I develop a parametric framework to provide economic interpretations of the HI-SDF. The main mechanism of the framework is illustrated in Figure 2. There are a continuum of households and two types of intermediaries: one is more financially constrained ("low type") than the other ("high type"). In good times, the high-type intermediary levers up by borrowing from both households and the low-type intermediary to invest in a risky asset. Because the high type's gains in the asset market are magnified by their high leverage, the wealth in the economy is concentrated in the hands of less constrained high types who are effectively less risk-averse. The equilibrium risk premium is thus low in good times (left panel). In bad times (right panel), however, the high types' net worth losses are amplified by their high leverage; they are forced to deleverage by selling the asset. As a result, the wealth in the economy shifts to the low types who are more constrained and effectively more risk-averse. However, the low type requires extra compensation for taking on the additional risk in the risky asset market; this is achieved by an increase in the risk premium. Without perfect risk sharing between agents, the heterogeneity of intermediates – captured by their relative wealth distribution – becomes a factor in risk pricing.

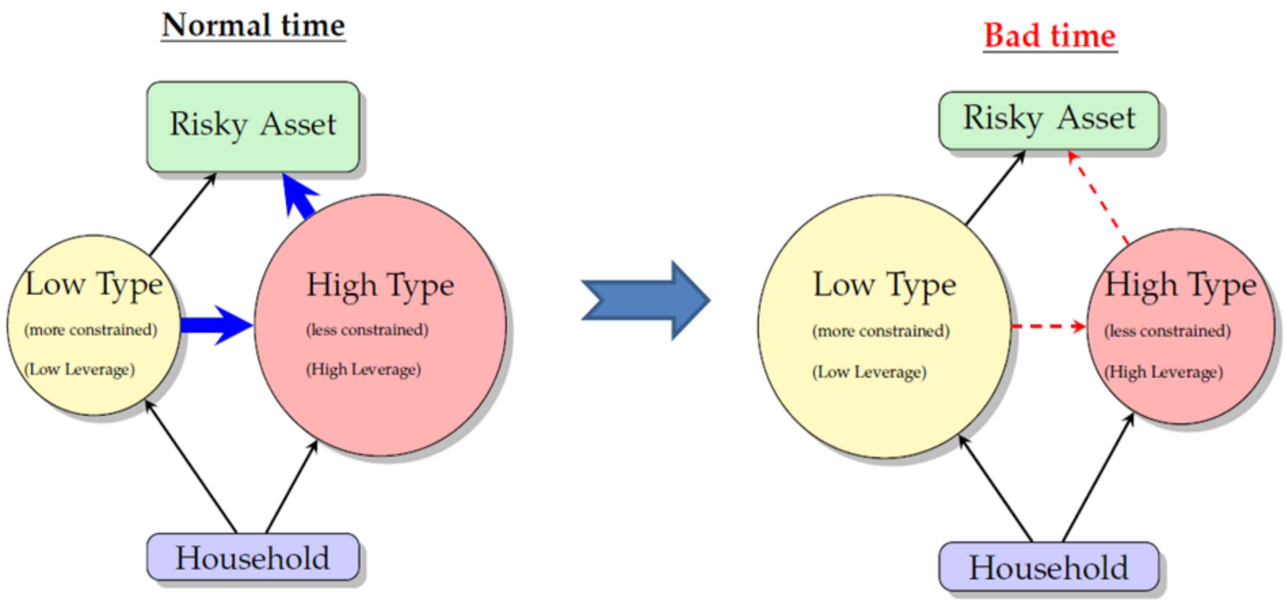

Figure 2: Schematic Illustration of HI-SDF Framework

I then analytically characterize HI-SDF m_t in equilibrium and investigate the following questions: why does the pricing wedge exist? And what determines the nonlinearity and, more importantly, governs the functional form of the wedge? I show that the pricing wedge arises in equilibrium due to the presence of financial friction, and its nonlinearity originates from the heterogeneity in financing constraints. When two types of intermediaries are *ex-ante* identical, the pricing wedge becomes a linear function of aggregate leverage.

Last, the equilibrium shows that the functional form of the HI-SDF crucially hinges on the *types* of constraints that intermediaries are subject to. I investigate five types of financing constraints commonly studied in the literature and show that equilibrium pricing wedge under these constraints differ substantially even using the same set of calibrated parameter values. Comparing these model-implied SDFs to the semiparametric estimate, I find that margin constraint is more supported by the data over other constraints when pricing equity portfolios, but the data favors the Value-at-Risk constraint for other asset classes. These results emphasize the role of financial constraints in improving explanatory power for cross-sectional variation in asset returns.

Related Literature

This paper contributes to three existing strands of the literature in finance, applied econometrics and macroeconomics.

First, this paper contributes to the growing *finance* literature on intermediary-based asset pricing models. Adrian, Etula, and Muir (2014) empirically shows that up to 70% of the cross-sectional variation in the characteristic-sorted equity portfolios can be explained by the broker-dealer leverage. He, Kelly, and Manela (2017) shows that the leverage of banking holding companies contains large pricing power for a wide range of assets. However, in these papers, the role of heterogeneity among intermediaries has been neglected. The paper is also closely related to works that consider the role of redistributive shocks among heterogeneous agents as a source of the priced risk. Examples include Danthine and Donaldson (2002), Gomez (2016), Greenwald, Lettau, and Ludvigson (2014) and Lettau, Ludvigson, and Ma (2019). Most of these works focus exclusively on the heterogeneity among households.

The importance of heterogeneity among intermediaries in pricing assets can be further emphasized by noting the disagreement over the implications of the sign of leverage risk price in the extant theoretical literature in finance. While many theories have been proposed in order to investigate the linkage between financial intermediary and asset pricing, different approaches to modeling the intermediary pricing kernel lead to *opposite* predictions on the sign of the price of leverage risk. On the one hand, Brunnermeier and Pedersen (2008) and Danıelsson, Shin, and Zigrand (2004) predict a *positive* price of leverage risk. In their model, the lower leverage corresponds to a tighter funding constraint. As a result, assets that are positively correlated with the leverage are considered to be risky and require larger risk premia.[3] On the other hand, He and Krishnamurthy (2013) predicts a *negative* price of leverage risk. In their model, intermediaries are subject to an equity constraint and the SDF is proportional to the equity. Consequently, assets that are positively correlated with the leverage require smaller risk premia because intermediaries with low leverage are more capitalized. Such disagreement originates from the different types of constraints intermediaries face. This paper contributes to this debate by investigating the model with different financial constraints and comparing the respective implied pricing kernel to the estimated HI-SDF using the asset return data.

After earlier versions of this paper were circulated, more papers started to pay attention to the heterogeneity among intermediaries and the role of their financing constraints in asset prices. In a contemporaneous work, Kargar (2021) shows balance sheet adjustments within the intermediary

[3] Adrian and Boyarchenko (2012) and Muir (2017) also predict a *positive* sign of leverage price risk when intermediaries are subject to a VaR constraint and a low probability rare disaster, respectively.

sector are quantitatively important for both the level of and variation in the risk premium. In contrast to Kargar (2021) where he focuses on margin constraints, this paper studies the role of different types of financial constraints and provides a semiparametric estimation and asset pricing evaluation of the HI-SDF whose functional form hinges on the specification of constraints. Using currency returns, Du, Hébert, and Huber (2022) uncovers that intermediary constraints are priced across various asset classes. This paper provides similar evidence using a semiparametric approach and information on intermediaries' balance sheets, and goes beyond it by investigating asset-pricing implications of different types of financing constraints.

Second, this paper contributes to the *applied econometrics* literature on the semi- and non-parametric estimation of stochastic discount factor models. While many macroeconomic models have been proposed to investigate the role of financial intermediaries, the evaluation of those models' fit with asset return data has received surprisingly less empirical attention. The empirical evaluation of the consumption-based asset pricing model, on the other hand, has already been conducted for both Habit-based utility (Chen and Ludvigson (2009)), and Epstein-Zin recursive preference (Chen, Favilukis, and Ludvigson (2014)). This paper follows the estimation procedure developed in Chen, Favilukis, and Ludvigson (2014) to estimate the HI-SDF semiparametrically and compares the model's fit with data against other commonly studied asset pricing models in terms of specification errors.[4]

Last, this paper is closely related to the large body of *macroeconomic* models of financial intermediations. The parametric framework in the paper combines insights from financial accelerator models of Gertler and Kiyotaki (2010), Maggiori (2017) and Chabakauri (2015), and heterogeneous agents asset pricing models of Brunnermeier and Pedersen (2008) and He and Krishnamurthy (2013). The former models show that financial constraints are important sources of the fluctuations in asset prices; latter models demonstrate the wealth share among different types of agents is a key state variable in the economy that drives both macroeconomic and asset price fluctuations. The importance of heterogeneity among intermediaries has already been emphasized in the literature. Coimbra and Rey (2017) studies how intermediary heterogeneity can jointly affect monetary expansion and financial stability. Gertler, Kiyotaki, and Prestipino (2016) emphasizes the distinct role of the wholesale banking sector in the breakdown of the financial system compared to the retail banks. In these papers, however, asset pricing implications of such heterogeneity have been largely neglected.

The rest of the paper is organized as follows. Section 2 discusses the data and details on

[4]Other works on nonparametric estimation of SDF models include Escanciano, Hoderlein, Lewbel, Linton, and Srisuma (2021) and Christensen (2017)

measuring the HI-SDF. Section 3 evaluates its ability to fit asset return data relative to other asset pricing models. Section 4 provides economic interpretations using a parametric example. Section 5 concludes.

2 Measuring HI-SDF

I start the analysis by measuring the heterogeneous intermediary stochastic discount factor (HI-SDF) m_t that takes the following form,

$$m_t\left(\psi_t, \phi_t\right) \equiv \beta\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} \Omega\left(\psi_t, \phi_t\right), \tag{1}$$

where $\beta\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}$ captures household marginal rate of substitution and $\Omega\left(\psi_t, \phi_t\right)$ is a pricing wedge that is a general function of two variables: ψ_t is the net worth share among intermediaries and ϕ_t is aggregate intermediary leverage.

While variables ψ_t and ϕ_t arise naturally as state variables in financial accelerator models with two types of intermediaries, such as Kargar (2021) and Gertler, Kiyotaki, and Prestipino (2016), the functional form of $\Omega(\cdot)$ is generally unknown in the literature or is often assumed to be linear for simplicity. For example, in both Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017), the pricing wedge Ω_t is a linear function of intermediary leverage.

In Section 4, I provide a parametric example that the equilibrium SDF satisfies equation (1) and more importantly, the wedge $\Omega(\cdot)$ is generally *nonlinear* due to the presence of financial frictions and its specific functional form crucially hinges on the types of financial constraints intermediaries are subject to.

2.1 Data and Preliminary Evidence

In this section, I describe the data source and then provide some preliminary empirical evidence on the importance of intermediaries' heterogeneity in asset pricing. Appendix C provides more detailed descriptions of all data. The sample is quarterly and spans the period from 1970:Q1 to 2019:Q4 unless noted otherwise.

Intermediary Balance Sheet data For the intermediaries' balance sheet information, I obtained data from various sources. For the BHCs, I obtained the data on total financial asset and liability from the Flow of Funds (FoF) Table L.131. To cross-check the results, I followed Avraham, Selvaggi, and Vickery (2012); I also used the Consolidated Report of Condition and

Income (FR Y-9C) to construct the leverage of U.S. banking holding companies. The FR Y-93 provides quarterly data on the balance sheet information of BHCs with at least \$500 million in total assets. One advantage of this data source is that it includes data on the financial condition of firms on each tier of the banking holding companies. For example, the domestic commercial banks file a separate *Call Report*, while the nonbank subsidiaries file a separate FR Y-11 report on their financial condition.

The aggregate broker-dealers data are obtained from the FoF Table L.130. This table contains the quarterly updated balance sheet information of both standalone BDs as well as nonbank subsidiaries in the market. The BDs largely engage in short-term collateralized lending and borrowing. In 2010, for example, 40.8% of their assets comprised security repo lending and 57.3% of their liabilities were covered by collateralized borrowing in the repo market.

Asset Return Data For test assets, I use the equity return data obtained from Kenneth French's Dartmouth data library on 25 size/book-market sorted (size/BM), 25 size/operating profitability (size/OP), 10 long-run reversal (REV), and 25 size/investment (size/INV) portfolios. I also use the portfolio data recently explored by HKM to investigate nonequity assets, including the 10 corporate bond portfolios from Nozawa (2017) ("bonds"), 6 sovereign bond portfolios from Borri and Verdelhan (2011) ("sovereign bonds"), 54 S&P 500 index options portfolios sorted on moneyness and maturity from Constantinides, Jackwerth, and Savov (2013) ("options"), and 20 CDS portfolios constructed by HKM. For the rest of the paper, I denote the gross one-period return on asset j from the end of $t-1$ to the end of t by $R_{j,t}$ and the gross risk-free rate by $R_{f,t}$. I use the three-month T-bill rate as a proxy for the risk-free rate. The gross excess return is denoted by $R^e_{j,t} \equiv R_{j,t} - R_{f,t}$.

2.1.1 Intermediary Leverage and Net Worth Share

Leverage is defined as the ratio of total assets to total equity,

$$\phi^i_t = \frac{\text{Total Financial Asset}^i_t}{\text{Total Financial Asset}^i_t - \text{Total Liability}^i_t}, \tag{2}$$

where $i \in \{BHC, BD\}$ is the type of intermediary during quarter t. Note that the equity series obtained from the FoF are of book equity so the constructed leverage reflects the book leverage. I further construct the market leverage for publicly traded banks using data from Compustat. Market leverage is defined as in equation (2), but the total financial asset is replaced by the Enterprise Value (EV), which is the sum of market capitalization and total debt. As pointed out by Adrian, Colla, and Shin (2013) and Nuño and Thomas (2017), book leverage is a more

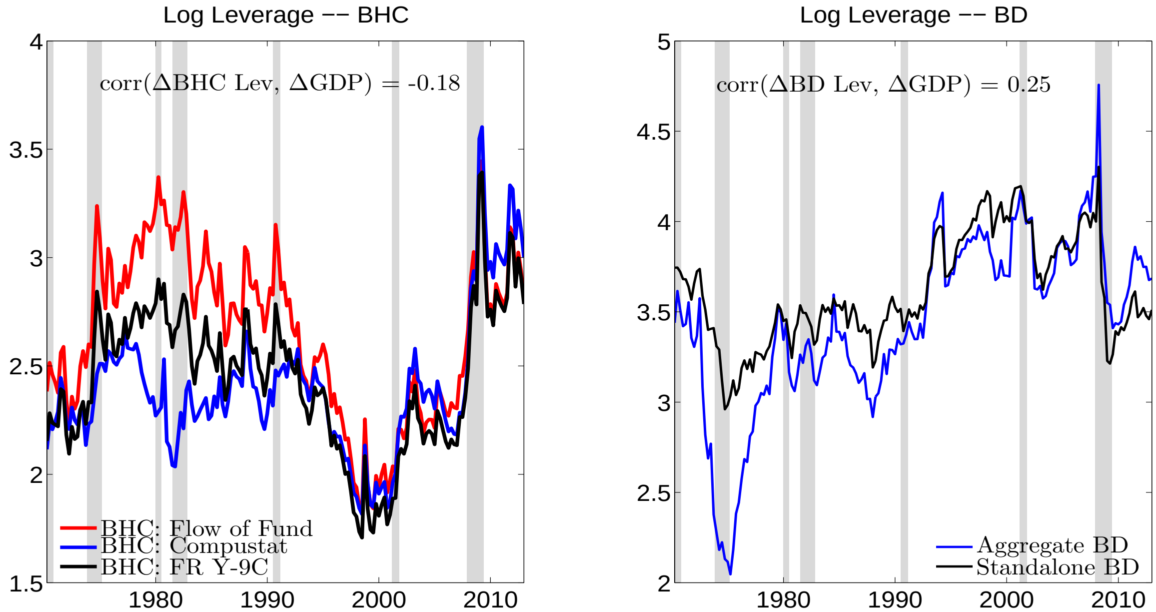

Figure 3: Leverage in BD and BHC. The left panel reports the log leverage of banking holding companies using Flow of Funds (red) data, Compustat (blue) and FR Y-9C (black). The right panel reports the log leverage of aggregate broker-dealer (blue) and proxy for standalone BD (black) described in section 3.6. The sample spans 1970Q1 to 2019Q4.

appropriate measure to study bank credit supply, as this paper does, whereas market leverage is more closely linked to new share issuance or merger and acquisition decisions.

Figure 3 displays the leverage of the BHCs (LHS panel) and BDs (RHS panel) from various sources. The LHS panel shows that the book leverage at the BHC level obtained from the FoF (red) is countercyclical and its innovation exhibits a -18% correlation with the innovation in GDP. The result is robust to either the regulatory filling data from FR Y-9C (black) or the market leverage data from Compustat (blue). The correlation between the BHC book leverage from the FoF and market leverage from Compustat is found to be 51%, which is consistent with the finding in HKM that market leverage is highly correlated with book leverage at the BHC level.

On the other hand, the BD sector features procyclical leverage, as depicted in the RHS panel of Figure 3. The result is similar when using the aggregate BD data from the FoF and a proxy for standalone BDs (for further discussion on the construction of standalone BDs, see section 3.3). The innovation in book leverage from the FoF in BHCs and BDs is negatively correlated (-30%).

The net worth share among intermediaries is defined as

$$\psi_t = \frac{\text{Total Asset}_t^{BHC} - \text{Total Liability}_t^{BHC}}{\text{Aggregate Asset}_t - \text{Aggregate Liability}_t},$$

where the numerator is the equity of the BHCs, defined as the difference between their total asset and total liability. The denominator is the aggregate of the BHC and BD equity. Appendix C specifies all the items included in the total assets and liabilities of the BHCs and BDs respectively.

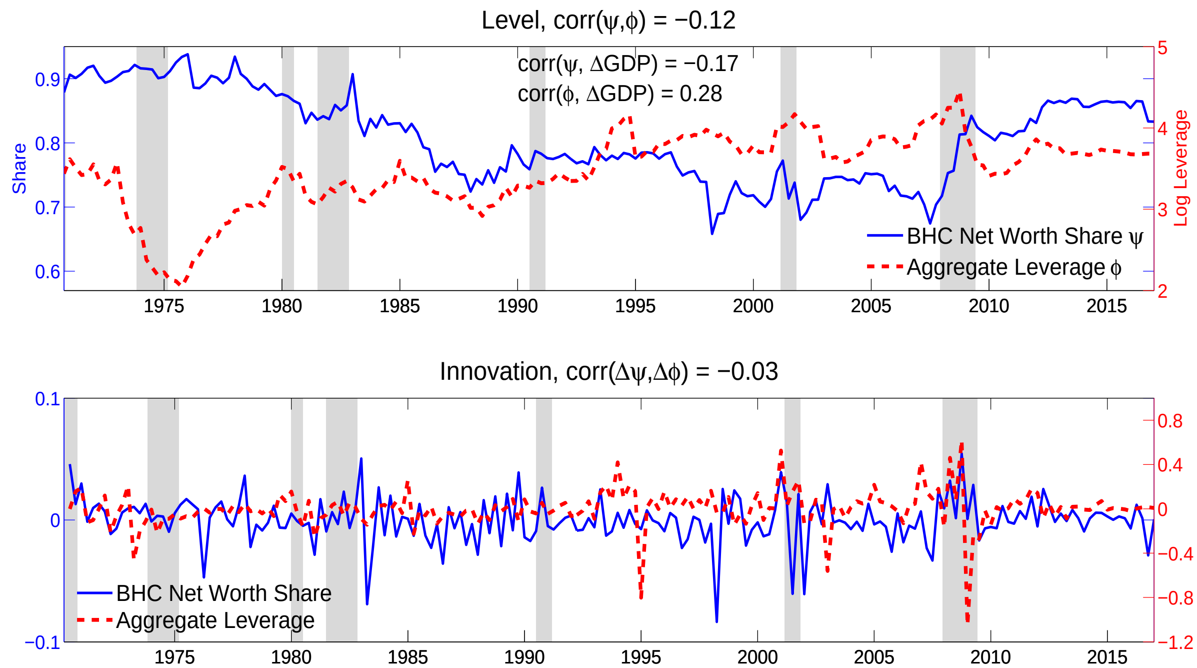

Figure 4: Time series of net worth share. Aggregate leverage is defined as the sum of total assets from BHC and BD over sum of book equity. Δ and Innovation refer to the AR(1) residual. The sample spans 1970Q1 to 2019Q4.

The top panel of Figure 4 shows that the net worth share is highly volatile, ranging from over 90% in the early 1970s to around 67% in the late 1990s. In addition, the net worth share is countercyclical: the net worth share level is −17% correlated with the innovation in GDP. Moreover, the share is negatively correlated with the aggregate leverage, with a correlation of −12%. The bottom panel of Figure 4 shows that the innovation of net worth share and aggregate leverage is only weakly correlated, suggesting that both variables contain some independent information from the financial intermediaries.

2.1.2 Intermediary Heterogeneity and Asset Prices: Firm-level Evidence

Before getting into the structural estimation of HI-SDF, I first investigate whether the heterogeneity among intermediaries —captured by the net worth share— matters in pricing risky securities. Following the literature, I form a stock portfolio based on the risk exposure to the innovation of net worth share. The stock portfolio includes all individual firm (with share codes 10 and 11) stock returns from the Center for Research in Security Prices (CRSP). For each stock j, I regress its excess return on the AR(1) residual of the net worth share, denoted by $\Delta\psi_t$,

$$R^e_{j,t} = a_j + \beta_{j,\psi}\Delta\psi_t + u_{j,t}, \quad \text{for } t = 1, 2...T,$$

where $\beta_{j,\psi}$ is the exposure to risk of net worth share. If the pricing wedge $\Delta\psi_t$ carries a negative price of risk, the assets with high $\beta_{j,\psi}$ should have lower expected returns.

Panel A of Table 1 reports the average annualized excess returns for firms at the top quintile of net worth share beta are 5.81% (equal-weighted) and 5.93% (value-weighted) lower than those at the bottom quintile. And these differences are statistically significant. This finding is also robust for portfolios *double sorted* by net worth share beta and intermediary leverage beta (Table A1 in Appendix A). Panel A thus provides preliminary empirical evidence that net worth share is an important risk factor that has been largely neglected in the representative intermediary framework.

Intermediary Heterogeneity and Stock Returns

Panel A: **Portfolios sorted by Net Worth Share Beta** $\beta_{j,\psi}$						
	Low	2	3	4	High	High-Low
Equal-weighted	10.59**	8.47***	6.89**	5.96**	4.78**	−5.81***
	(2.12)	(3.02)	(1.99)	(2.17)	(2.42)	(−3.74)
Value-weighted	11.21**	10.24***	8.92**	7.21**	5.28**	−5.93***
	(1.99)	(2.83)	(2.02)	(2.51)	(2.22)	(−3.17)

Panel B: Fama Macbeth Regression			
Model	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$	BIC
CCAPM	0.19	0.36	−193.52
FF Factors	0.31	0.28	−216.56
HKM	0.28	0.29	−212.91
AEM	0.41	0.24	−227.77
HI-SDF	0.72	0.17	−254.90

Table 1: This table reports the average annualized excess return for portfolios sorted by net worth share beta. The "Low" corresponds to stock returns at 0-20 percentile of the net worth beta. The "High" corresponds to stock returns at 80-100 percentile of net worth share beta. The NW t-statistics is reported in parenthesis. Panel B reports the adjusted R squared, pricing errors, and Bayesian information criterion (BIC) penalty from Fama-Macbeth regression using net worth share sorted portfolios and factors indicated in the first column. The sample spans the period 1970Q1 to 2019Q4. *sig. at 10%. **sig. at 5%. ***sig. at 1%.

Panel B of Table 1 shows that other competing asset pricing models are unable to capture this spread due to intermediary heterogeneity. The adjusted cross-sectional R^2 from the Fama-Macbeth regression ranges from 19% to 41% among alternative models, including Fama French (FF) five-factor model and representative intermediary asset pricing models. In contrast, the cross-sectional $\bar{R}^2$ is 72% when the HI-SDF is used to price the net worth share sorted portfolios. The pricing errors, summarized by root-mean-squared errors and the Bayesian information criterion

(BIC) penalty, are smaller for the HI-SDF model. These results provide preliminary evidence of the importance of incorporating the intermediaries' heterogeneity for pricing assets in the equity market.

2.2 Semiparametric Estimation of the HI-SDF

I now describe the semiparametric estimation of the HI-SDF. Using the formulation of HI-SDF given in (1), the Euler equation takes the form

$$E_t\left(\underbrace{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\Omega\left(\psi_{t+1},\phi_{t+1}\right)}_{m_{t+1}}R_{i,t+1}-1\right)=0,\ i=1,...,N, \tag{3}$$

where $\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$ is the intertemporal MRS in consumption, and $R_{i,t}$ is the gross return of test asset i at period t. The moment restriction (3) forms the basis of the empirical investigation.

2.2.1 Semiparametric Estimation Procedure

The key difference of the HI-SDF from the standard consumption-based asset pricing model is with regard to the pricing wedge $\Omega_t(\cdot)$ due to the presence of financial frictions. This leaves one challenge for the estimation: $\Omega_t(\cdot)$ is unobservable and must be inferred from observable data. In order to address this issue, I follow Chen, Favilukis, and Ludvigson (2014) (henceforth, CFL) and apply a two-step sieve minimum distance (SMD) procedure using the moment condition (3). In particular, I use a nonparametric specification of Ω_t and treat it as a general function of net worth share ψ_t and the aggregate leverage ϕ_t.

Let $\nu=(\beta,\gamma)'$ denote the vector of the finite dimensional preference parameters to be estimated, and for each test asset $i=1,...,N$,

$$\varkappa_i\left(\mathbf{z}_{t+1},\nu,F\right)\equiv\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\Omega\left(\phi_{t+1},\psi_{t+1}\right)R_{i,t+1}-1,$$

where $\mathbf{z}_{t+1}$ is the vector containing all observations, including consumption growth, asset return data, aggregate leverage, and net worth share.

The true value $\Omega_0(\cdot,\nu)$ minimizes the conditional moment restriction's squared distance from zero,

$$\Omega_0(\cdot,\nu)\equiv\arg\inf_{\Omega}E\left[\sum_{i=1}^{N}\left(E\left\{\varkappa_i\left(\mathbf{z}_{t+1},\nu,\Omega\right)|\mathcal{F}_t\right\}\right)^2\right],$$

where N is the total number of test assets and $\mathcal{F}_t$ is the set of information available at time t.

The true parameter value ν_0 is thus the minimizer of the conditional moment restriction's squared distance from zero evaluated at true value $\Omega_0(\cdot,\nu)$

$$\nu_0 = \min_{\nu} E\left[\sum_{i=1}^{N}\left(E\left\{\varkappa_i\left(\mathbf{z}_{t+1},\nu,\Omega_0(\cdot,\nu)\right)|\mathcal{F}_t\right\}\right)^2\right].$$

The model is correctly specified if

$$E\left\{\varkappa_i\left(\mathbf{z}_{t+1},\nu_0,\Omega_0(\cdot,\nu_0)\right)|\mathcal{F}_t\right\} = 0 \tag{4}$$

for all test assets $i \in \{1, 2...N\}$. If $\mathbf{w}_t$ denotes the instruments, an observable subset of $\mathcal{F}_t$, equation (4) implies that

$$E\left\{\varkappa_i\left(\mathbf{z}_{t+1},\nu_0,\Omega_0(\cdot,\nu_0)\right)|\mathbf{w}_t\right\} = 0 \tag{5}$$

Denote the conditional moment condition

$$M(\mathbf{w}_t,\nu,\Omega) \equiv E\left[\varkappa\left(\mathbf{z}_{t+1},\nu,\Omega\right)|w_t\right], \tag{6}$$

where $\varkappa = (\varkappa_i)_{i=1}^{N}$ is a vector of moment conditions $\varkappa_i$.

For any candidate value of preference parameters ν, I define Ω^* as

$$\Omega^* \equiv \inf_{\Omega} E\left[M(\mathbf{w}_t,\nu,\Omega)' M(\mathbf{w}_t,\nu,\Omega)\right]. \tag{7}$$

When the model is correctly specified, I have $\Omega^*(\mathbf{z}_t,\nu_0) = \Omega_0(\mathbf{z}_t,\nu_0)$. I estimate the model (5) using a profile semiparametric minimum distance procedure that consists of two steps. In the first step, for any finite-dimensional parameter candidate ν, the unknown function Ω^* is estimated using the SMD procedure.[5] In the second step, I conduct a Generalized Method of Moments (GMM) estimation of parameters ν using the optimal function Ω^* obtained from the first step.

First Step – Profile SMD Estimation of Ω For any preference parameter candidate ν, the unknown functions are estimated using an SMD procedure. Because Ω is an infinite-dimensional object, I approximate it by a sequence of finite-dimensional dimension-unknown sieve functions $B_{K_T}(\cdot,\nu)$, with dimension K_T and sample size T. For each ν, the sieve function is estimated by minimizing the weighted quadratic norm of nonparametrically estimated conditional expectation functions. The details on the choice of sieve function are discussed in Appendix D.

[5] The SMD procedure is developed in Newey and Powell (2003), Ai and Chen (2003), Ai and Chen (2007), and Chen, Favilukis, and Ludvigson (2014) applies it in the consumption-based asset pricing model with recursive preference. To my best knowledge, this is the first paper to apply this procedure for the estimation of intermediary-based asset pricing models

More specifically, the pricing wedge Ω is approximated as follows:

$$\begin{aligned}\Omega^{*}\left(\phi_{t},\psi_{t};\nu\right) &\approx \Omega_{K_{T}}\left(\cdot,\nu\right)\\ &= a_{0}\left(\nu\right)+\sum_{j=1}^{K_{T}}a_{j}\left(\nu\right)B_{j}\left(\phi_{t},\psi_{t}\right),\end{aligned}$$

where the sieve coefficients $\{a_j\}$ depend on the preference parameter ν whereas the basis function $\{B_j\}$ is independent of those parameters.[6]

For profile SMD estimation, I need a consistent estimate of $M(\mathbf{w}_t, \nu,\Omega)$, defined in equation (6), and consistency is achieved by applying the sieve least squares procedure. Following CFL, I form the first step profile SMD estimate $\widehat{\Omega}$ for Ω^* using (8) and the sample analog of (7)

$$\widehat{\Omega}\left(\cdot,\nu\right)=\arg\min_{\Omega_{K_{T}}}\frac{1}{T}\sum_{t=1}^{T}\widehat{M}\left(\mathbf{w}_{t},\nu,\Omega\right)'\widehat{M}\left(\mathbf{w}_{t},\nu,\Omega\right),$$

where

$$\widehat{M}\left(\mathbf{w}_{t},\nu,\Omega\right)=\left(\sum_{t=1}^{T}\varkappa\left(\mathbf{z}_{t+1},\nu,\Omega\right)p^{J_{T}}\left(\mathbf{w}_{t}\right)'\left(\mathbf{P}'\mathbf{P}\right)^{-1}\right)p^{J_{T}}\left(\mathbf{w}\right) \tag{8}$$

is the sieve least squares estimator of the condition mean vector $M(\mathbf{w}_t, \nu,\Omega)$ and

$$\mathbf{P}=\left(p^{J_{T}}\left(w_{1}\right),...,p^{J_{T}}\left(w_{T}\right)\right)$$

is the collection of sequences of known basis functions $p^{J_T}(\cdot) \equiv \left(p_{01}, ..., p_{0J_T'}\right)$ with order J_T.

As shown in CFL, this sieve least square procedure can be implemented as a GMM estimation with weighting matrix $\mathbf{I}_N \otimes (\mathbf{P}'\mathbf{P})^{-1}$. Thus, this particular weighting matrix gives greater weight to the moments that are more correlated with instruments $p^{J_T}(\cdot)$.

Second Step – GMM Estimation of ν Following the initial estimation of $\widehat{\Omega}(\cdot,\nu)$ in the first step, I estimate the preference parameters ν using the Hansen (1982) Generalized Method of Moments(GMM),

$$\begin{aligned}\widehat{\nu} &= \arg\min_{v}Q_{T}\left(\nu\right)\\ Q_{T}\left(\nu\right) &= \left[\mathbf{g}_{T}\left(\nu,\widehat{\Omega}\left(\cdot,\nu\right);\mathbf{y}^{T}\right)\right]'\mathbf{W}\left[\mathbf{g}_{T}\left(\nu,\widehat{\Omega}\left(\cdot,\nu\right);\mathbf{y}^{T}\right)\right],\end{aligned}$$

where $\mathbf{W}$ is a positive, semi-definite weighting matrix, and $\mathbf{y}^T$ is the vector of observables including $\mathbf{z}_t$ and any chosen measurable function of instrument set $\mathbf{w}_t$. The sample moment conditions are

[6] Details on sieve functions are in Appendix D.

expressed as

$$g_T\left(\gamma,\rho|\widehat{\Omega},\mathbf{y}^T\right)=\frac{1}{T}\sum_{t=1}^{T}\varkappa\left(\mathbf{z}_{t+1},\nu,\widehat{\Omega}\right)\otimes\mathbf{y}^T.$$

Note that $\widehat{\Omega}\left(\cdot,\nu\right)$ is not held fixed in the second step of the estimation, but depends on the finite-dimensional preference parameters ν. Consequently, the final estimate of $\widehat{\Omega}\left(\cdot,\nu\right)$ is determined in the second step along with the preference parameters ν.

Similar to CFL, I use two different weighting matrices $\mathbf{W}$ to obtain the second-step GMM estimation. The first one is the identity matrix $\mathbf{W}=\mathbf{I}$,and the second is the inverse of the sample second moment of asset returns, $\mathbf{W}=\mathbf{G}_T^{-1}$, where the element (i,j) of $\mathbf{G}_T$ is

$$\mathbf{G}_T^{(i,j)}=\frac{1}{T}\sum_{t=1}^{T}R_{i,t}R_{j,t}$$

for any $i,j\in\{1,2,...N\}$. Parameter estimates computed with $\mathbf{W}=\mathbf{G}_T^{-1}$ or $\mathbf{W}=\mathbf{I}$ have the advantage of being obtained by minimizing an objective function that is invariant to the initial choice of asset returns as suggested in Kandel and Stambaugh (1995). With the estimated $\widehat{\Omega}$ and $\widehat{\nu}=\left(\widehat{\beta},\widehat{\gamma}\right)'$ in hand, I estimate the HI-SDF,

$$\widehat{m}_t\left(\widehat{\nu},\widehat{\Omega}\right)=\widehat{\beta}\left(\frac{C_t}{C_{t-1}}\right)^{-\widehat{\gamma}}\widehat{\Omega}\left(\phi_t,\psi_t\right).$$

2.2.2 Comparison to a one-step procedure

In principle, all the parameters of the model, including the infinite-dimensional unknown function Ω_{K_T} and finite-dimensional parameters ν, can be estimated in a one-step SMD procedure, described in (9),

$$\min_{\Omega_{K_T},\nu}\frac{1}{T}\sum_{t=1}^{T}\widehat{M}\left(\mathbf{w}_t,\nu,\Omega\left(\cdot,\nu\right)\right)'\widehat{M}\left(\mathbf{w}_t,\nu,\Omega\left(\cdot,\nu\right)\right), \tag{9}$$

where $\widehat{M}\left(\mathbf{w}_t,\nu,\Omega\right)$ is as defined in equation (8).

There are two reasons for applying the two-step procedure instead. First, the preference parameter estimates, such as risk aversion, should reflect the values required to match the *unconditional* moments, as widely emphasized in the asset pricing literature (see Bansal and Yaron (2004)). The separate second procedure allows for estimating the preference parameters to match the *unconditional* risk premia, which is practically implausible in the one-step procedure in (9) because the unconditional estimation of the infinite-dimensional unknown function would require infinite unconditional moment restrictions.

The second reason is that both the weighting scheme and the use of instruments $p^{J_T}\left(\mathbf{w}\right)$ in the one-step procedure effectively changes the set of original test assets. Consequently, the preference parameters are estimated on a linear combination of the original portfolio returns, leading to a certain combination of short and long positions in the original test assets. Such a combination may contain a different set of information and asset spreads to the original test portfolios that are studied in the asset pricing literature. Breaking the estimation procedure into two steps allows for estimating parameters ν using an identity weighting matrix and retains the information from the original set of test assets.

2.2.3 Choice of Instruments

The empirical procedure's first step is to compute the instruments for estimating the conditional moment $\widehat{m}\left(\mathbf{w}_t, \nu, \widehat{\Omega}\right)$. The instruments are the known basis functions of conditioning variables $\mathbf{w}_t = \left[cay_t, RREL_t, SPEX_t, \frac{C_t}{C_{t-1}}, \phi_t\right]'$. For $\mathbf{w}_t$, I first include the proxy cay_t for the log consumption-wealth ratio, studied in Lettau and Ludvigson (2001), to forecast the asset returns. This proxy, cay_t, is measured as the cointegrating residual between log consumption, log asset wealth, and log labor income. The excess return on S&P 500 index, $SPEX$, and the relative T-bill rate, $RREL$, are also included because these two variables have been found to display forecasting power for various asset returns.[7] Last, I include the aggregate leverage ϕ_t, which captures the balance sheet information. In order to have the same units as asset returns, all the conditioning variables $\mathbf{w}_t$ are standardized, as recommended by Cochrane (2001). As noted in CFL, no formal test of instrument relevance has been developed so far for estimations involving an unknown function. Therefore, I select the common strong predictors of asset returns used in the asset pricing literature and undertake extensive robustness checks (more in section 3.3).

2.3 Estimation Results

For the empirical analysis, I consider individual test portfolios as well as the pooled estimations of all equity assets ("*equity portfolio*"), nonequity assets ("*nonequity portfolio*"), and all assets in the sample ("*all assets*").

2.3.1 Parameter Estimates

Figure 5 illustrates the estimated pricing wedge $\widehat{\Omega}(\phi_t, \psi_t)$ by using all assets as a function of BHC's net worth share ψ_t over the data range from 0.65 to 0.95, holding the aggregate leverage level ϕ_t fixed at its mean value (blue line) as well as at the 25th to 75th percentiles (grey shades). The estimation is obtained using a second-step weighting matrix equal to the identity matrix.[8]

[7]See Campbell (1991), Hodrick (1992), Lettau and Ludvigson (2004)

[8]Unreported results show that the estimated $\widehat{\Omega}$ is very similar if the weighting matrix $\mathbf{W} = \mathbf{G}_T^{-1}$.

To highlight the nonlinearity of the estimated wedge, I include a black dashed line that depicts a linear counterpart that connects the minimum to the maximum of the mean estimate. A large gap between the blue and the black dashed line indicates a large deviation from a linear model.

Figure 5 shows that the estimated pricing wedge is strictly increasing in ψ_t but decreasing in ϕ_t. Intuitively, intermediaries value their net worth more in the bad state of the economy where the wealth is concentrated in the hands of the more constrained intermediaries. Furthermore, the comparison between the blue and dashed black line highlights the nonlinearity of the wedge in net worth share and the deviation from its linear counterpart is the largest when ψ is in the range from 0.7 to 0.8, which contains more than 40% of observations in the sample (see Figure 4) and includes major events such as 1998 LTCM crisis and the 2007-09 global financial crisis. With regard to the aggregate leverage, the estimation shows that deleveraging has a larger impact on the pricing wedge when the net worth share is high. This suggests the nonlinearity of HI-SDF scales up the effect of changes in leverage on asset prices during the bad state of the economy but scales down the effect during the good time.

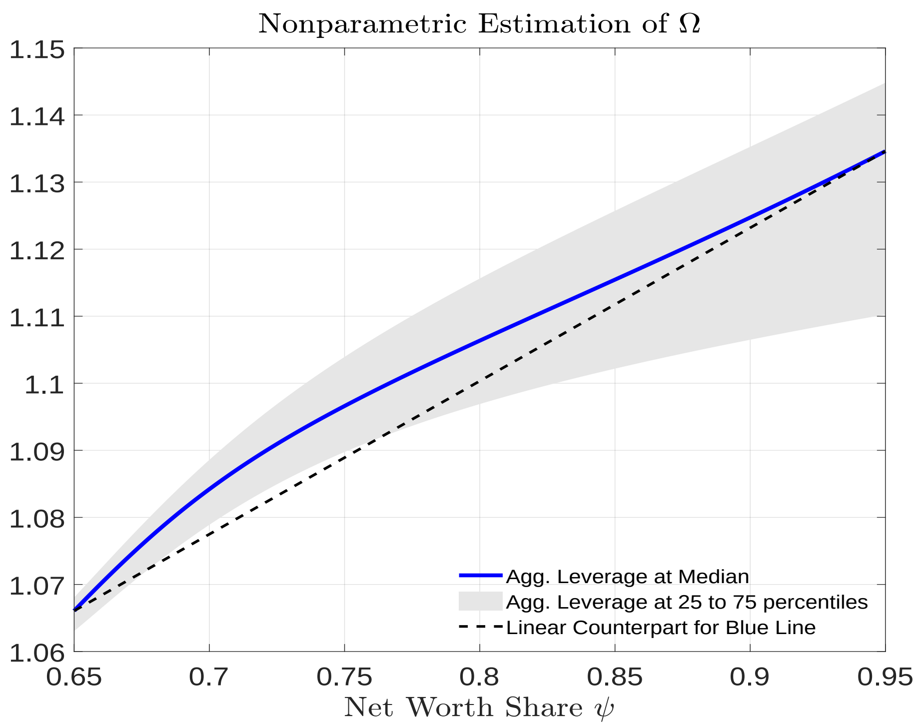

Figure 5: The solid blue line reports the non-parametrically estimated $\widehat{\Omega}\,(\psi, \phi)$ when the aggregate leverage ϕ is evaluated at its mean using weight matrix $\mathbf{W} = \mathbf{I}$. The black dashed line shows a linear counterpart that connects the minimum and maximum of the blue line. The shaded area represents $\widehat{\Omega}\,(\psi, \phi)$ when the aggregate leverage is evaluated at its 25 to 75 percentile values. The test assets are indicated by the panel title. See texts for the estimation procedure. The sample spans the period 1970Q1 to 2019Q4.

Figure 6 shows that the estimation result is quantitatively similar and robust when the HI-SDF is estimated using equity (left panel) and nonequity (right panel) portfolios, respectively. In both cases, the estimated pricing wedge is highly nonlinear in ψ_t and ϕ_t.

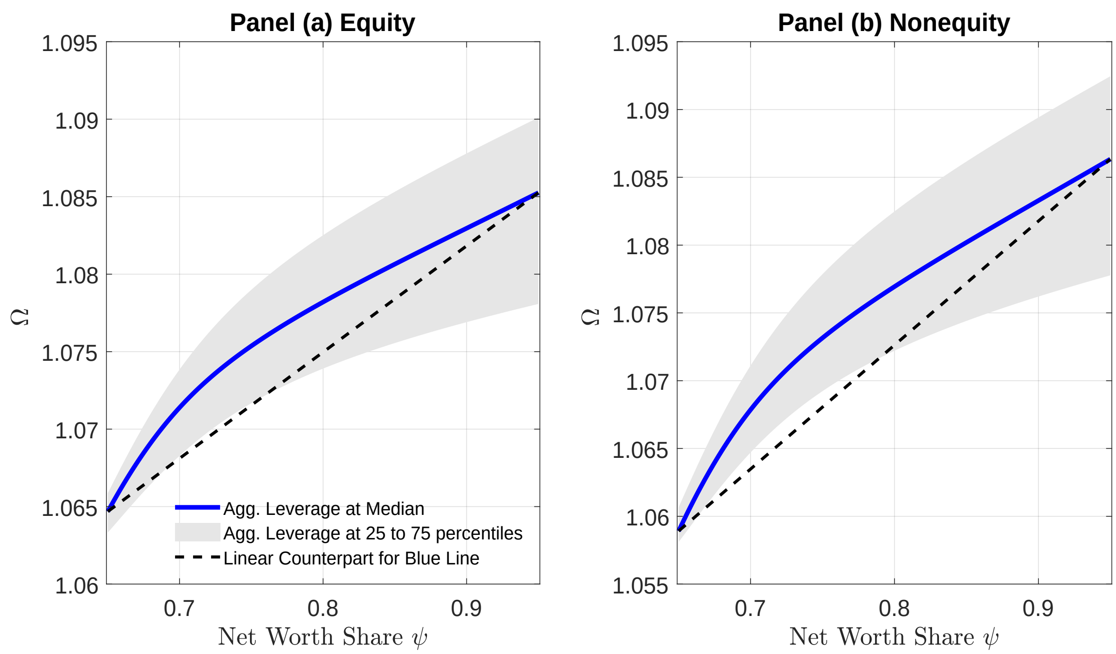

Figure 6: The solid blue line reports the non-parametrically estimated $\widehat{\Omega}(\psi, \phi)$ when the aggregate leverage ϕ is evaluated at its mean using weight matrix $\mathbf{W} = \mathbf{I}$. The black dashed line shows a linear counterpart that connects the minimum and maximum of the blue line. The shaded area represents $\widehat{\Omega}(\psi, \phi)$ when the aggregate leverage is evaluated at its 25 to 75 percentile values. The test assets are indicated by the panel title. See texts for the estimation procedure. The sample spans the period 1970Q1 to 2019Q4.

Table 2 reports the estimation results of finite-dimensional preference parameters ν using the SMD procedure. For comparison, an identical estimation using aggregate consumption data under recursive preference ("EZ Preferences") is reported.[9] The results using both the weighting matrixes specified in the second step ($\mathbf{W}_t = \mathbf{I}$ and $\mathbf{W}_t = \mathbf{G}_T^{-1}$) are included in the table. The top panel reports the summary statistics of the estimated wedge: $\widehat{\Omega}$ for HI-SDF and continuation value function for recursive preference. Compared to the consumption-based model (CCAPM) under recursive preferences, the wedge from HI-SDF is, on average twice, as large as the value function under CCAPM (2.8 v.s. 1.3) and twice more volatile (0.04 v.s. 0.02) when weighting matrix $\mathbf{W} = \mathbf{I}$. When weighting matrix $\mathbf{W} = \mathbf{G}_T^{-1}$ is used, $\widehat{\Omega}$ is almost three times more volatile than the wedge from the CCAPM model. As shown in Figure 4, the aggregate leverage is more volatile than the aggregate consumption growth and this plays a crucial role in the lower risk aversion estimates from the HI-SDF.

Panel B of Table 2 reports the estimated finite-dimensional parameters $\nu = (\gamma, \beta)'$. For both the weighting matrixes employed in the second step ($\mathbf{W}_t = \mathbf{I}$ and $\mathbf{W}_t = \mathbf{G}_T^{-1}$), the estimated

[9]This result is an updated estimation of Table 2 in CFL.

Estimation of Parameters: HI-SDF

Panel A: Summary Statistics of the SDF Wedge						
	Mean		Standard Deviation		Auto Corr.	
	HI-SDF	EZ Pref.	HI-SDF	EZ Pref.	HI-SDF	EZ Pref.
$\mathbf{W} = \mathbf{I}$	2.7518	1.2872	0.0422	0.0211	0.48	0.56
$\mathbf{W} = \mathbf{G}_T^{-1}$	2.5227	1.6873	0.0716	0.0238	0.51	0.59

Panel B: Preference Parameter Estimates				
	γ		β	
	HI-SDF	EZ Pref.	HI-SDF	EZ Pref.
$\mathbf{W} = \mathbf{I}$	31.5	53.6	0.989	0.989
	$[12.5, 42.0]$	$[25.8, 116.9]$	$[0.982, 0.999]$	$[0.983, 0.998]$
$\mathbf{W} = \mathbf{G}_T^{-1}$	26.0	60.8	0.997	0.999
	$[22.5, 58.5]$	$[43.2, 132.1]$	$[0.992, 0.999]$	$[0.993, 1.000]$

Table 2: This table reports parameter estimates based on the two-step SMD procedure described in the text for HI-SDF and consumption-based model under Epstein-Zin preferences ("EZ Pref."). The second-step parameters are obtained by minimizing the GMM criterion with the weighting matrix indicated in the table. The 95 percent confidence intervals are reported in the parenthesis. The test assets include all assets (equity plus nonequity portfolios). The sample spans the period 1970Q1 to 2019Q4.

time discount factor β is close to 1 and the 95% confidence interval ranges from 0.98 to 0.99. The estimated risk aversion is 32 using HI-SDF, which is lower than $\gamma = 54$ under CCAPM. The 95% confidence interval of γ with CCAPM widely ranges from 25 to 132. In contrast, the estimated γ using the HI-SDF ranges from 13 to 42 with the identity weighting matrix and 23 to 59 with $\mathbf{W}_t = \mathbf{G}_T^{-1}$.

2.3.2 Cyclical Properties of SDF

From equation (1), the estimated pricing kernel $\widehat{m}_t$ can be decomposed into products of two pieces $m_{1,t}$ and $m_{2,t}$,

$$\widehat{m}_t = \underbrace{\beta \left(\frac{C_t}{C_{t-1}} \right)^{-\gamma}}_{m_{1,t}} \underbrace{\widehat{\Omega} \left(\phi_t, \psi_t \right)}_{m_{2,t}}.$$

The first component $m_{1,t}$ is the traditional SDF from the consumption-based asset pricing model with constant risk aversion γ, and the second piece, the pricing wedge, is the multiplicative component existed because of the presence of intermediaries.

The top panel of Figure 7 plots the time series of the estimated pricing kernel $\widehat{m}_t$ using all assets. The middle and bottom panels show the time series of the estimated HI-SDF components $m_{1,t}$ and $m_{2,t}$, respectively. The top panel shows that the estimated SDF $\widehat{m}_t$ is countercyclical with a negative correlation -0.31 and GDP growth over the sample. The middle and bottom panels show that $m_{1,t}$ and $m_{2,t}$ are countercyclical with -0.17 and -0.48 correlation with GDP growth, respectively. However, because $m_{2,t} = \widehat{\Omega}_t$ is much more volatile than consumption-based $m_{1,t}$, the cyclical properties of the HI-SDF, which determine those of the risk premium, are mostly driven by the wedge.

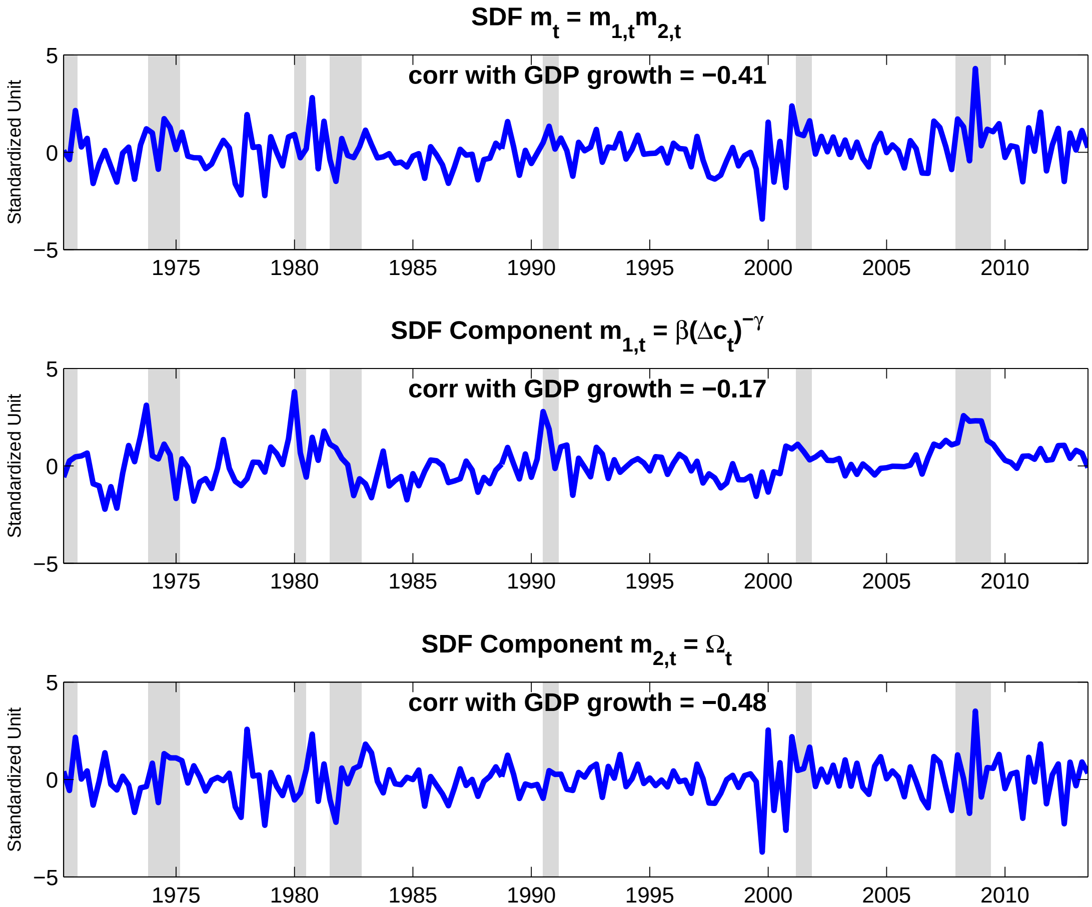

Figure 7: Time Series of HI-SDF. The top panel plots the time series of the estimated HI-SDF $m_t = m_{1,t}m_{2,t}$ from SMD procedure described in the text where $m_{1,t} = \beta\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}$ is the multiplicative piece from household's MRS and $m_{2,t} = \widehat{\Omega}\left(\psi_t, \phi_t\right)$ is the estimated pricing wedge. The middle and bottom panels correspond to the time series of estimated $m_{1,t}$ and $m_{2,t}$ respectively. The shaded areas correspond to the NBER recessions. The SDF are estimated using all assets including both equity and nonequity portfolios with identity weighting matrix $\mathbf{W} = \mathbf{I}$ employed in the second step of the SMD procedure. The sample spans the period 1970Q1 to 2019Q4.

3 Asset Pricing Tests and Results

This section describes the asset pricing tests and presents the results. Unless otherwise noted, the estimated HI-SDF $\widehat{m}_t$ in this section refers to the nonparametric estimates using all assets (equity plus nonequity portfolios) with identity weighting matrix $\mathbf{W} = \mathbf{I}$.

3.1 Linear Tests

I start the empirical investigation of the pricing power of the HI-SDF by log-linearizing the estimated $\widehat{m}_t$,

$$m_t \approx b_0 + b_1 \Delta c_t + b_2 \log \widehat{\Omega}\left(\psi_t, \phi_t\right). \tag{10}$$

Note that although I log-linearize the HI-SDF, the pricing wedge $\widehat{\Omega}\left(\psi_t, \phi_t\right)$ is still a nonlinear function in (ψ_t, ϕ_t).

There are two reasons to start the analysis using a log-linearized $\widehat{m}_t$. First, a (semi) linear SDF allows for estimating the unconditional price risk of consumption growth Δc_t, that of the pricing wedge $\widehat{\Omega}\left(\psi_t, \phi_t\right)$, and the explanation power for cross-sectional variation of excess returns using a two-step Fama–Macbeth regression. Second, the SDFs from representative intermediary asset pricing models (HKM and AEM) are (log) linear in the intermediary's leverage. Therefore, log-linearizing HI-SDF allows me to compare the pricing power using the *same* empirical strategy (Fama–Macbeth test) applied in HKM and AEM. The tests using nonlinear HI-SDF will be discussed in section 3.2.

3.1.1 Fama-Macbeth Regression

The log-linearized m_t in equation (10) suggests that unconditional price risks can be estimated in a cross-sectional regression of the expected excess returns on the risk exposure β of two series: log consumption growth Δc and the log of pricing wedge $\log \widehat{\Omega}\left(\psi_t, \phi_t\right)$.

More specifically, let N denote the number of portfolio returns in the cross-sectional investigation. Exposures to consumption growth risk and pricing wedge risk can be estimated via time-series regressions, one for each asset $j = 1, 2....N$,

$$R^e_{j,t} = a_j + \beta_{j,c} \Delta c_t + \beta_{j,\Omega} \log \widehat{\Omega}_t + u_{j,t}, \quad t = 1, 2...T, \tag{11}$$

I then run a cross-sectional regression, regressing the expected excess returns on estimated betas $(\widehat{\beta}_{j,c}, \widehat{\beta}_{j,\Omega})'$,

$$E_T\left(R^e_{j,t}\right) = \lambda_0 + \widehat{\beta}_{j,c}\lambda_c + \widehat{\beta}_{j,\Omega}\lambda_\Omega + \epsilon_j, \quad j = 1, 2....N, \tag{12}$$

where parameter λ_0 (same for all assets j) is included to account for the "zero beta" rate in case of no true risk-free rate. $\widehat{\lambda}_c$ and $\widehat{\lambda}_\Omega$ are respectively the consumption growth and pricing wedge risk prices to be estimated. $\widehat{\beta}_{j,c}$ and $\widehat{\beta}_{j,\Omega}$ denote the estimated risk exposure to Δc_t and $\log \widehat{\Omega}_t$, respectively, from the time-series regression. Equations (11) and (12) are estimated jointly in a GMM system.

One concern is that $\widehat{\Omega}$ is estimated from a separate semiparametric procedure that also minimizes the "distance" between the asset returns and the pricing wedge, which mechanically increases the pricing power of the HI-SDF. To alleviate this concern, I compute the HI-SDF using all pooled assets. In other words, I do not re-estimate $\widehat{\Omega}$ for different test assets used in the cross-sectional tests. In all cases, I also compute block bootstrap 95% confidence intervals for the estimated risk prices that account for the SMD estimation of the wedge (see Appendix D for more details). Last, I show the results hold out of sample (see section 3.3).

I report a cross-sectional $\overline{R}^2$ as a measure of how well the model explains the cross-section of quarterly returns. I also compute the pricing errors measured as the ratio of the root-mean-squared pricing errors (RMSE) to the root-mean-squared return (RMSR) on the portfolios being priced.[10] Last, because the number of factors varies widely across different asset pricing models, I rank the competing specifications according to a Bayesian Information Criterion (BIC) that adjusts for the number of free-factor risk prices λ chosen to minimize pricing errors,

$$\text{BIC} = 2N_K \ln RMSE + (N_F + 1) \ln N_K,$$

where N_K and N_F are the numbers of test assets and factors, respectively. A model is preferred more by data when its BIC is smaller relative to other models.

3.1.2 Linear Test Results

Table 3 reports the risk prices based on the cross-sectional regressions (12) on individual portfolios. In particular, Panels A–D report the risk price estimates on equity portfolios: size/BM, REV, Size/INV, and Size/OP. Panels E–H report the results for four nonequity asset classes: bonds, sovereign bonds, options, and CDS. For each portfolio group, I report the estimated risk prices of pricing wedge $\widehat{\lambda}_\Omega$ with and without the consumption growth risk factor Δc_t.

From Table 3, the risk price for pricing wedge $\log \widehat{\Omega}_t$ is negative and statistically significant in each of the portfolios, as indicated by the 95% bootstrap confidence interval. The result is quantitatively similar with and without controlling for the risks in consumption growth. The cross-sectional $\overline{R}^2$ ranges from 41% to 76% with tight confidence intervals in most cases. Turning to the nonequity asset classes, the risk prices for the pricing wedge betas are again negative and statistically significant in each case. The pricing wedge alone explains 85% of the cross-sectional variation in expected returns on corporate bonds and 99% on options. The pricing errors $RMSE/RMSR$ for both equity and nonequity portfolios are all below 20%, except for the case with CDS. The magnitudes of the risk prices of the pricing wedge are similar across various assets.

[10] $RMSE \equiv \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(E_T\left(R_j^e\right) - \widehat{R}_j^e\right)^2}$, $\quad RMSR \equiv \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(E_T\left(R_j^e\right)\right)^2}$

Expected Return-Beta Regressions

$E_T\left(R^e_{i,t}\right) = \lambda_0 + \lambda'\beta + \epsilon_i$, Estimates of Factor Risk Prices λ

Equity Portfolios

Panel A: **Size/BM**					Panel B: **REV**				
Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$	Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$
0.17		−1.72	0.71	0.14	0.31		−1.02	0.37	0.21
[0.01, 0.31]		[−2.08, −1.41]	[0.59, 0.87]		[0.23, 0.44]		[−1.31, −0.89]	[0.17, 0.52]	
−0.32	0.21	−1.54	0.76	0.12	0.44	0.17	−0.98	0.41	0.19
[−0.51, 0.14]	[0.01, 0.47]	[−1.73, −1.22]	[0.61, 0.89]		[0.13, 0.68]	[−0.07, 0.33]	[−1.32, −0.42]	[0.34, 0.51]	
Panel C: **Size/INV**					Panel D: **Size/OP**				
Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$	Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$
0.83		−1.41	0.54	0.19	−0.17		−0.97	0.61	0.17
[0.44, 1.17]		[−1.76, −1.08]	[0.51, 0.78]		[−0.33, 0.01]		[−1.33, −0.59]	[0.50, 0.77]	
0.34	0.28	−1.23	0.56	0.18	−0.01	0.31	−1.08	0.64	0.15
[0.25, 0.57]	[0.11, 0.51]	[−1.62, −0.96]	[0.49, 0.76]		[−0.41, 0.37]	[0.11, 0.76]	[−1.33, −0.78]	[0.57, 0.73]	
Nonequity Portfolios									
Panel E: **Bonds**					Panel F: **Sovereign Bonds**				
Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$	Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$
0.12		−1.27	0.85	0.18	−0.08		−1.78	0.71	0.19
[0.01, 0.32]		[−1.44, −1.08]	[0.81, 0.97]		[−0.31, 0.11]		[−2.03, −1.47]	[0.69, 0.87]	
0.01	0.47	−1.03	0.89	0.16	0.18	0.51	−1.66	0.72	0.19
[−0.05, 0.33]	[0.33, 0.58]	[−1.35, −0.84]	[0.79, 0.96]		[0.01, 0.41]	[0.27, 0.69]	[−1.88, −1.42]	[0.65, 0.83]	
Panel G: **Options**					Panel H: **CDS**				
Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$	Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$
−1.79		−2.13	0.99	0.09	−0.41		−1.07	0.51	0.49
[−2.13, −1.56]		[−2.37, −2.04]	[0.98, 0.99]		[−0.57, −0.28]		[−1.22, −0.79]	[0.42, 0.66]	
−1.39	0.77	−2.04	0.99	0.09	−0.07	0.34	−0.99	0.52	0.47
[−1.71, −1.18]	[0.52, 0.92]	[−2.21, −1.87]	[0.97, 0.99]		[−0.41, 0.32]	[0.22, 0.57]	[−1.32, −0.77]	[0.41, 0.68]	

Table 3: Expected return-beta regressions, individual portfolios. Bootstrap 95% confidence intervals are reported in square brackets. All coefficients are scaled by multiple of 100. The sample spans the period 1970Q1 to 2019Q4.

Table 4 reports the results when I pool all assets (both equity and nonequity) in the estimation. I find that the estimated price risk of $\log\widehat{\Omega}$ remains negative and statistically significant. In contrast, the representative intermediary models have trouble pricing all assets. The estimated price risk of the AEM leverage factor is negative and not statistically significant. The price risk of the HKM leverage factor remains statistically significant, but can only explain 43% of the cross-sectional $\bar{R}^2$.[11] Compared to these models, the HI-SDF can improve up to 27% cross-sectional $\bar{R}^2$ among these assets.

The results are similar if equity and nonequity assets are separately estimated. Panel A of Table 5 shows that the price risk of intermediary wedge $\log\widehat{\Omega}$ is statistically significant and also exhibits a large 77% cross-sectional $\bar{R}^2$ at a low 14% RMSE/RMSR pricing error for pooled equity portfolios. Similarly, Panel B shows that the HI-SDF can price nonequity portfolios well, with 81% $\bar{R}^2$ and 11% pricing error.

[11] I use the original AEM leverage factor and the original HKM two-factor model that includes market returns and intermediary capital ratio. For simplicity, the coefficient of market return is omitted in the table.

Expected Return-Beta Regressions, All Assets

$E_T\left(R^e_{i,t}\right) = \lambda_0 + \lambda'\beta + \epsilon_i$							
Estimates of Factor Risk Prices λ							
Constant	Δc_t	$\log\widehat{\Omega}\left(\psi_t,\phi_t\right)$	AEM	HKM	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$	BIC
-0.19^{*}	0.41	-0.98^{**}			0.79	0.12	-1259.78
(-1.66)	(1.55)	(-1.97)			$[0.68, 0.89]$		
$[-1.24, -0.07]$	$[0.21, 0.55]$	$[-1.04, -0.77]$					
0.97			-1.56		0.52	0.15	-1233.17
(1.00)			(-1.53)		$[0.41, 0.89]$		
$[0.02, 1.78]$			$[-2.77, 1.21]$				
1.20				1.55^{*}	0.43	0.20	-1241.32
(1.59)				(1.73)	$[0.32, 0.63]$		
$[0.53, 1.86]$				$[0.05, 2.01]$			

Table 4: Expected return-beta regressions, all assets portfolios. Bootstrap 95% confidence intervals are reported in square brackets. GMM t-statistics are reported in parenthesis. All coefficients are scaled by multiple of 100. "AEM" refers to the original leverage factor from Adrian et al.(2014). "HKM" refers to a two-factor model that includes the reciprocal of the intermediary capital ratio from He et al.(2017) and the market return. The coefficient of the market return is not reported for simplicity. The sample spans the period 1970Q1 to 2019Q4. *sig. at 10%. **sig. at 5%. ***sig. at 1% according to the GMM t-statistics.

HKM showed that the BD leverage can price equities but not nonequity portfolios. The opposite results hold for HKM's leverage factor based on BHC leverage. Table 5 shows that by incorporating heterogeneity among intermediaries, the HI-SDF can price *both* equity and nonequity portfolios. Panel A of Table 5 also shows that the intermediary wedge of the HI-SDF $\log\widehat{\Omega}_t$ performs better in terms of lower BIC than both HKM's and AEM's factors for equity portfolios. A similar finding holds for nonequity portfolios as shown in Panel B.

Figure 8 exhibit the visual impressions of these results. The left panel of Figure 8 focuses on equity portfolios and plots the observed quarterly return premia of each portfolio on the y-axis against the fitted return from the HI-SDF model. The right panel of Figure 8 reports the results from the same exercise with nonequity portfolios. Both panels show that most of the asset returns are close to the 45-degree line, with the exception of options. For equity portfolios, the left panel of Figure 8 shows that the HI-SDF is able to explain some of the large value spreads observed in the data. For example, the high-return S1B5 and low-return S1B1 are both close to the 45-degree line, and this result also holds for other size categories.

3.1.3 Is the HI-SDF just a proxy for other risk factors?

Table 5 shows that when the HI-SDF is included along with the betas from the intermediary factor of AEM (of HKM) for equity (nonequity) portfolios, the pricing power for these factors is substantially weakened. In contrast, the risk price of $\log\Omega_t$ remains negative and strongly significant in each case. In addition, for both portfolios, the magnitudes of the estimated risk prices

Expected Return-Beta Regressions

$E_T\left(R^e_{i,t}\right) = \lambda_0 + \lambda'\beta + \epsilon_i$, Estimates of Factor Risk Prices λ							
Panel A: **equity portfolios**							
Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	AEM	HKM	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$	BIC
0.39^{*}	0.35	-1.11^{**}			0.77	0.14	-987.44
(1.16)	(1.55)	(-2.04)			$[0.68, 0.98]$		
$[0.12, 0.51]$	$[0.01, 0.98]$	$[-1.31, -0.99]$					
0.98			2.15^{**}		0.68	0.18	-971.51
(1.07)			(2.63)		$[0.57, 0.72]$		
$[-0.03, 1.31]$			$[2.07, 2.43]$				
1.21	0.16	-1.16^{**}	2.37^{*}		0.71	0.17	-978.27
(1.19)	(1.41)	(-2.01)	(1.62)		$[0.67, 0.78]$		
$[1.01, 1.42]$	$[-0.03, 0.42]$	$[-1.51, -1.01]$	$[2.11, 2.45]$				
1.42	0.22	-1.08^{**}		-1.56^{*}	0.73	0.16	-981.78
(1.41)	(1.55)	(-1.98)		(-1.91)	$[0.61, 0.89]$		
$[0.98, 2.18]$	$[-0.06, 0.86]$	$[-1.27, -0.98]$		$[-2.27, -1.12]$			
Panel B: **nonequity portfolios**							
Constant	Δc_t	$\log\widehat{\Omega}(\psi_t,\phi_t)$	AEM	HKM	$\bar{R}^2$	$\frac{\text{RMSE}}{\text{RMSR}}$	BIC
0.33^{*}	0.31	-1.17^{**}			0.81	0.11	-661.18
(1.89)	(1.24)	(-2.37)			$[0.58, 0.91]$		
$[0.12, 0.57]$	$[-0.04, 0.89]$	$[-1.79, -1.01]$					
0.92	0.27^{*}	-1.11^{*}	-1.26		0.72	0.14	-632.47
(1.30)	(1.81)	(-1.93)	(-0.53)		$[0.68, 0.89]$		
$[0.21, 1.29]$	$[0.11, 0.48]$	$[-1.32, -1.02]$	$[-3.27, 0.79]$				
1.52				0.22^{**}	0.69	0.17	-621.11
(1.39)				(2.01)	$[0.61, 0.81]$		
$[0.71, 1.79]$				$[0.05, 0.42]$			
0.17	0.39	-1.23^{*}		0.16^{*}	0.77	0.12	-660.17
(0.99)	(1.22)	(-1.95)		(1.71)	$[0.70, 0.89]$		
$[-0.23, 0.92]$	$[0.02, 1.02]$	$[-1.42, -1.24]$		$[0.01, 0.28]$			

Table 5: Expected return-beta regressions, equity and nonequity portfolios. Bootstrap 95% confidence intervals are reported in square brackets. GMM t-statistics are reported in parenthesis. All coefficients are scaled by multiple of 100. 'AEM" refers to the original leverage factor from Adrian et al.(2014). "HKM" refers to a two-factor model that includes the reciprocal of the intermediary capital ratio from He et al.(2017) and the market return. The coefficient of the market return is not reported for simplicity. The sample spans the period 1970Q1 to 2019Q4. *sig. at 10%. **sig. at 5%. ***sig. at 1% according to the GMM t-statistics.

of $\log\Omega_t$ remain similar regardless of the inclusion of other intermediary factors. For example, the price risk of $\log\Omega_t$ changes from -1.11 to -1.16 when the AEM leverage factor is added and remains statistically significant at the five-percent level.

These findings suggest that the intermediary pricing wedge not only contains the balance sheet information from both leverage factors, but also can reconcile the relationship between different types of intermediaries and exhibit even higher pricing power for various types of financial assets.

3.1.4 Does the nonlinearity of Ω matter?

Figure 5 shows that the estimated pricing wedge Ω is highly nonlinear in both variables. Does this nonlinearity matter in explaining the cross-section variation in asset returns? Table 6 compares

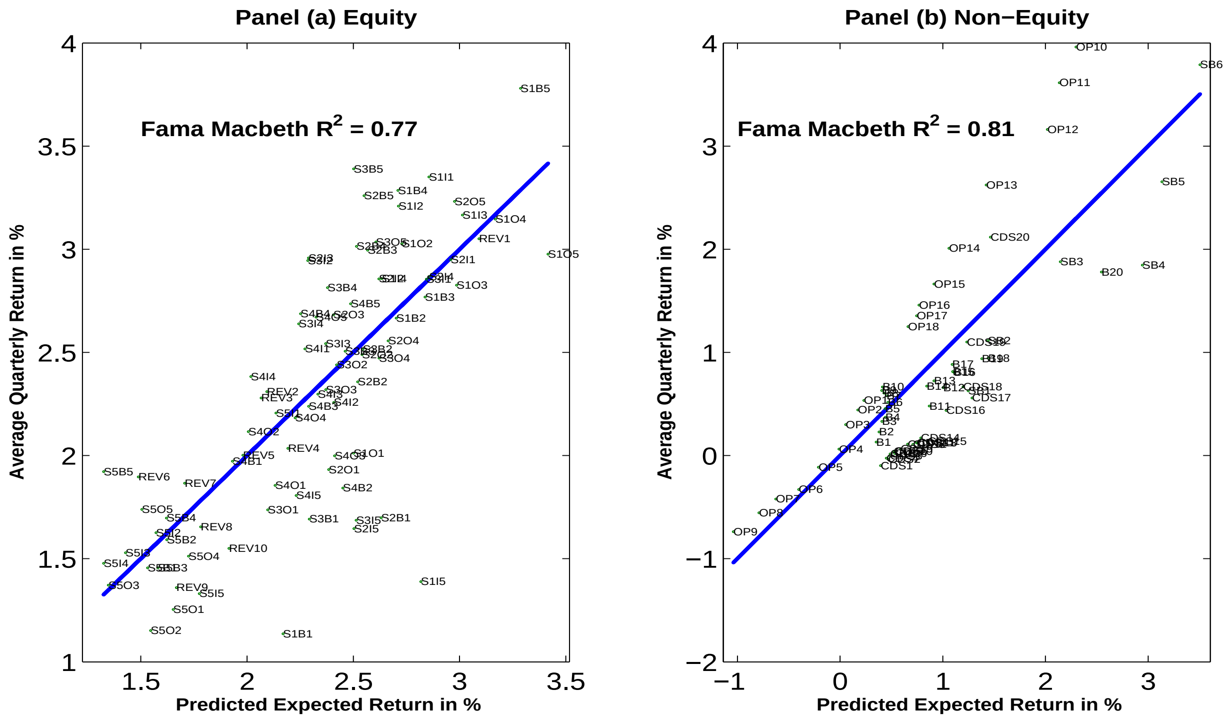

Figure 8: Realized vs Predicted Mean Returns: equity and non-equity. Predicted returns are constructed from Fama-MacBeth regressions of average returns on consumption growth beta and intermediary wedge beta using equity portfolios (including size/bm, REV, size/INV, size/OP) and nonequity portfolios (including corporate bonds, sovereign bonds, CDS and options). The sample spans the period 1970Q1 to 2019Q4.

the HI-SDF's explanatory power on asset return with two alternative asset pricing models. The first one is a linear asset pricing model with net worth share ψ_t and aggregate leverage ϕ_t. It is a special case of the HI-SDF when Ω is (log) linear in both variables. The second alternative model is a two-factor linear "naive" model with two leverage factors from AEM and HKM. Consumption growth is included in all models but is omitted in the table for simplicity of presentation. Panel A of Table 6 shows that, even in the linear model, the risk price of net worth share ψ_t is statistically significant and carries a negative price of risk. However, the adjusted $\bar{R}^2$ is 59%, thus explaining 20% lower cross-sectional variation than the HI-SDF. This suggests that the nonlinearity of Ω plays a non-trivial role in pricing the assets and improves the pricing power by 34%. These results are robust for equity and nonequity portfolios, as reported in Panels B and C, respectively. The cross-sectional $\bar{R}^2$ of the linear model is 22% lower than the HI-SDF for equity portfolios and 21% lower for nonequity portfolios. This result also holds out of sample (see section 3.3).

3.2 Nonlinear Tests

The previous section provides empirical evidence that the log-linearized HI-SDF exhibits substantial power for pricing risky securities. This section studies the pricing power of the estimated

Expected Return-Beta Regressions

$E_T\left(R^e_{i,t}\right) = \lambda_0 + \boldsymbol{\lambda}'\boldsymbol{\beta} + \epsilon_i$						
Estimates of Factor Risk Prices λ						
Constant	$\log\widehat{\Omega}\left(\psi_t,\phi_t\right)$	ψ_t	ϕ_t	AEM	HKM	$\bar{R}^2$
Panel A: **all assets**						
-0.19^{*}	-0.98^{**}					0.79
(-1.66)	(-1.97)					
0.37		-0.47^{**}	2.33^{*}			0.59
(0.27)		(-2.47)	(1.79)			
-0.27^{*}				3.17^{*}	0.27	0.52
(-2.37)				(1.77)	(1.12)	
Panel B: **equity portfolios**						
0.39^{*}	-1.11^{**}					0.77
(1.16)	(-2.04)					
-1.28		-0.69^{**}	2.13^{**}			0.55
(-0.58)		(-2.99)	(2.13)			
0.19				4.19^{**}	-0.21	0.49
(0.17)				(2.03)	(-0.29)	
Panel C: **nonequity portfolios**						
0.33^{*}	-1.17^{**}					0.81
(1.89)	(-2.37)					
1.27		-0.17^{**}	0.51^{*}			0.60
(1.03)		(-2.08)	(1.69)			
-0.08				1.08	-0.77^{*}	0.52
(-1.17)				(0.21)	(-1.88)	

Table 6: Expected return-beta regressions, linear models. Bootstrap 95% confidence intervals are reported in square brackets. GMM t-statistics are reported in parenthesis. Consumption growth is included in all models but is omitted. All coefficients are scaled by multiple of 100. 'AEM" refers to the original leverage factor from Adrian et al.(2014). "HKM" refers to a two-factor model that includes the reciprocal of the intermediary capital ratio from He et al.(2017) and the market return. The coefficient of the market return is not reported for simplicity. The sample spans the period 1970Q1 to 2019Q4. *sig. at 10%. **sig. at 5%. ***sig. at 1% according to the GMM t-statistics.

fully nonlinear $\widehat{m}_t$. The empirical strategy of the nonlinear test is based on GMM (GMM Hansen (1982)) and the basis of estimation is the Euler equations taking the form

$$E\left[\widehat{m}_{t+1}R^e_{t+1}\right] = 0. \tag{13}$$

3.2.1 Goodness-of-fit Measures

I study two goodness-of-fit measures. The first is the Hansen–Jagannathan (HJ) distance (Hansen and Jagannathan (1997)), and the second is the cross-sectional R^2 from the nonlinear GMM.

HJ-Distance To compare the nonlinear model's fitness of asset returns relative to other commonly studied models, I first use the methodology provided by Hansen and Jagannathan (1997) to compare the misspecification for the true unknown SDF for each model. Specifically, I compute the HJ distance for each SDF model with parameter ν and pricing kernel $m_t(\nu)$ as

$$d_T \equiv \sqrt{\mathbf{g}_T^{HJ}(\nu)}, \tag{14}$$

where

$$\mathbf{g}_T^{HJ}(\nu) = g_T(\nu)\,\mathbf{G}_T^{-1} g_T(\nu), \tag{15}$$

and $g_T(\nu)$ is a vector of the sample average of pricing errors defined as

$$g_{i,T}(\nu) = \frac{1}{T}\sum_{t=1}^{T} m_t(\nu)\,R_{i,t} - 1$$

for each test asset $i \in \{1, 2...N\}$.

Note that the HJ distance (14) for the HI-SDF is computed by using the parameter estimates obtained from the second-step SMD procedure with weighting matrix $\mathbf{W} = \mathbf{G}_T^{-1}$ (the sample second moment matrix of asset returns). This places a disadvantage for the HI-SDF because the sieve parameters of the unknown function Ω_t are not included in equation (14). On the other hand, parameters of other competing models are chosen to minimize the HJ criterion. In order to address this issue and rank the models' performance, I apply an Akaike information criterion (AIC) penalty to the HJ distance and report the adjusted HJ distance defined as

$$\text{HJ-Dist} = \sqrt{d_T^2 + \frac{N_P}{T}},$$

where N_P is the number of free parameters in ν chosen to minimize the HJ distance (15).

I also compute a conditional version of HJ distance:

$$d_T^{\text{Cont}} \equiv \sqrt{\mathbf{g}_T^{\text{Cont}}(\nu)\,\mathbf{G}_T^{-1}\mathbf{g}_T^{\text{Cont}}(\nu)},$$

where $g_T^{\text{Cont}}(\nu)$ is the sample average of pricing errors conditional on instrument Z_t, defined as

$$\mathbf{g}_T^{\text{Cont}}(\nu) = \frac{1}{T}\sum_{t=1}^{T}\left[\left(m_t(\nu)\,\mathbf{R}_t - 1\right)\otimes Z_t\right],$$

and the conditional second-moment weighting matrix is defined as

$$\mathbf{G}_T = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{R}_t \otimes Z_t\right)\left(\mathbf{R}_t \otimes Z_t\right)'.$$

Following CFL, I choose a single instrument $Z_t = cay_t$ for numerical tractability.

Cross-sectional R^2 For nonlinear GMM estimation with moment conditions (13), the cross-sectional R^2 can be computed as

$$R^2 = 1 - \frac{Var_c\left(E_T\left(R^e_j\right) - \widehat{R}^e_j\right)}{Var_c\left(E_T\left(R^e_j\right)\right)},$$

where Var_c stands for cross-sectional variance and $\widehat{R}^e_j$ is the average return premium predicted by the model for asset j.

3.2.2 Nonlinear Test Results

I analyze the pricing ability of the HI-SDF against the following competing models. First, representative intermediary asset pricing models with leverage of BDs as studied in AEM, and one with leverage of BHCs as in HKM. The SDFs for these models are estimated following the same two-step SMD procedure, except that the pricing wedge $\Omega_t(\phi)$ is nonparametrically estimated as a function of leverage only. Second, consumption-based asset pricing models under the Epstein-Zin (EZ) preferences and under CRRA power utility. I also include a linear version of HI-SDF in which the wedge is a linear combination of ψ and ϕ.

Tables 7 report the cross-sectional R^2, the unconditional and conditional HJ distance of different models. In all cases, the nonlinear HI-SDF displays smaller specification errors than other competing models.

First, for equity portfolios, Panel A of Table 7 shows that the unconditional HJ distance is 0.411 for the HI-SDF; this is 19.6% smaller than the consumption-based model with recursive preference and 22.7% smaller than the power utility model. Similarly, the conditional HJ distance for the HI-SDF is 0.416 and is smaller than both CCAPM models: 0.556 for recursive preference and 0.576 for the power utility model. Similar results hold for the nonequity portfolio reported in Panel B of Table 7.

Second, the HI-SDF outperforms the representative intermediary models with either BDs or BHCs. Panel A of Table 7 shows that, compared to the intermediary SDF with only BDs, as studied in Adrian, Etula, and Muir (2014), the HI-SDF has 11% smaller conditional HJ distance for equity portfolios. Panel B of Table 7 shows that the conditional HJ distance of the HI-SDF is 10% smaller than that of an intermediary SDF model with BHCs for pricing nonequity portfolios. The HI-SDF has the highest cross-sectional R^2 among all the competing models: the HI-SDF has 16% and 12% higher R^2 for equity portfolio and nonequity portfolios, respectively.

Last, compared with a linear version of the HI-SDF, the nonlinearity of the pricing wedge decreases the specification error up to 12% and increases the R^2 by 11%.

Specification Errors and Cross-Sectional R^2, Nonlinear Tests

	Panel A **Equity Portfolio**			Panel B **Nonequity Portfolios**		
Model	R^2	Unconditional HJ-Dist	Conditional HJ-Dist	R^2	Unconditional HJ-Dist	Conditional HJ-Dist
HI-SDF	0.70	0.411	0.416	0.68	0.427	0.512
HI-SDF (linear)	0.59	0.429	0.467	0.58	0.437	0.541
Rep. Int., BD Only	0.54	0.434	0.489	0.49	0.477	0.572
Rep. Int., BHC Only	0.49	0.452	0.501	0.56	0.441	0.561
EZ Pref., No intermediary	0.41	0.511	0.556	0.39	0.499	0.612
CRRA, No intermediary	0.27	0.532	0.576	0.24	0.512	0.674

Table 7: This table reports the cross-sectional R^2, adjusted unconditional and conditional HJ distance for each model specified in the first column. The conditional HJ distance is computed conditional on cay. "HI-SDF (linear)" refers to an SDF as a linear combination of the net worth share and aggregate leverage. "Rep. Int., BD Only" refers to an SDF as a function of Broker-dealer leverage. "Rep. Int., BHC Only" refers to an SDF as a function of BHC's leverage. "EZ Pref., No intermediary" refers to consumption-based model under Epstein-Zin preferences. "CRRA, No intermediary" refers to consumption-based model under CRRA utility. The sample spans the period 1970Q1 to 2019Q4.

3.3 Robustness and Out-of-Sample Evidence

This section investigates the robustness of the results, including the out-of-sample test. All the tables discussed in this section are reported in Appendix A.

Alternative Individual Test Assets Panel A of Tables A2 reports the specification errors with currency momentum portfolios (Menkhoff, Sarno, Schmeling, and Schrimpf (2012)) and industry equity portfolios, respectively. The HI-SDF again has a smaller adjusted HJ distance compared to other competing models. It outperforms the representative intermediary model for pricing currencies, with 10% smaller specification error. The result is similar using industry portfolios.

Control for Consumption of Wealthy Households Panel B of Table A2 presents the comparison of HJ distance by using a consumption proxy for households with the top 10% wealth instead of aggregate consumption in the estimation of the HI-SDF. I use a consumption proxy defined in Lettau, Ludvigson, and Ma (2019),

$$C_t^S = C_t \left(\widehat{Y_t^{>90\%}/Y_t} \right)^{\chi},$$

where $Y_t^{>90\%}/Y_t$ is the top 10 percentile wealth share obtained from the Survey of Consumer Finance (SCF) and the ordinary least square (OLS) fitted value is obtained by

$$\widehat{Y_t^{>90\%}/Y_t} = \widehat{\alpha} + \widehat{\beta}\,(1 - LS_t)\,,$$

where LS_t is the nonfarm sector labor share at time t and $\chi \geq 0$.

Panel B of Table A2 shows that the specification error of the HI-SDF with equity portfolio is still smaller than other alternative models. In addition, comparing to Table 7, the HI-SDF's specification error is strictly smaller when I replace the aggregate consumption with this consumption proxy. This result is in line with the findings in Lettau, Ludvigson, and Ma (2019), because the consumption proxy for wealthy households is much more volatile than the aggregate consumption. Results are robust if the HI-SDF is estimated using nonequity portfolios or all assets.

Different Instrument Variables Panel C of Table A2 shows the adjusted HJ distance with different sets of instruments in the first step of the SMD procedure. I consider three alternative instrument sets: (i) excluding cay_t, (ii) including a lag of net worth share ψ_t, and (iii) excluding the relative T-bill $RREL_t$. The results are highly robust: the HJ distance from the HI-SDF is the smallest among all models for these different instrument sets.

Pre-crisis Subsample Panel A Table A3 reports the estimated finite-dimensional parameters $\nu = (\gamma, \beta)'$ and the Fama–Macbeth regression result using a pre-crisis subsample, particularly from 1970Q1 to 2006Q4. The results remain robust: the estimated risk aversion is 39.7, ranging from 21 to 57 with the identity weighting matrix and 26 to 64 with $\mathbf{W}_t = \mathbf{G}_T^{-1}$. The estimated price risk of the pricing wedge remains negative and statistically significant (Panel C). The HI-SDF can explain 72% of the cross-sectional variation for all 206 assets using the pre-crisis sample.

Banks v.s. Nonbanks Adrian and Boyarchenko (2013) documented that the leverage of the banking sector is procyclical, whereas the leverage of the nonbanking financial sector is countercyclical. Unlike in this paper, the banking sector data include the aggregate balance sheet information of the BDs and commercial banks, whereas the nonbanking financial sector refers to the remaining balance sheets of the financial sector after subtracting the aggregate balance sheets of the banking sector. The results based on the estimation of the HI-SDF associated with these two sectors are robust. Panel B Table A3 shows that the price risk of the estimated $\widehat{\Omega}$ is still statistically significant with 70% adjusted cross-sectional $\bar{R}^2$.

Standalone Broker-dealers The baseline estimation treats BD and BHC leverage as two separate entities. One concern is that some BDs are subsidiaries of BHCs. I show that the main results still hold when the BD leverage is replaced by a *standalone* BD leverage proxy. Assuming that the subsidiaries are 100% equity-owned by their parent companies, I can calculate the standalone BD leverage as follows,

$$\phi_t^{BD} = 1 + \frac{\text{Total BD Liability - Total Liability from Parents Companies}}{\text{Total BD Equity}}.$$

Panel C Table A3 reports the estimated parameters $\nu = (\gamma, \beta)'$ using this measure along with the BHC leverage. The result is quantitatively similar to the baseline estimate. The estimated γ ranges from 14 to 44 with the identity weighting matrix, and Panel C shows that the price risk of the pricing wedge $\widehat{\Omega}$ remains negative and statistically significant, explaining 81% cross-sectional variations in all assets.

Out-of-sample Test Another concern is that HI-SDF's strong explanation power for cross-sectional variation in the test assets is due to the fact that HI-SDF itself is estimated using a subset of test assets. Note that the block bootstrap procedure already accounts for the SMD estimate of the SDF. To further alleviate this concern, Panel D of Table A3 reports results from an out-of-sample test, in which the HI-SDF is estimated semiparametrically using *equity* portfolios, but the price of risks are estimated using *nonequity* portfolios. The price risk of the HI-SDF remains statistically significantly negative, explaining 71% cross-sectional variation among nonequity portfolios despite $\widehat{\Omega}$ itself is estimated using equity returns.[12] It's worth highlighting that this out-of-sample nonlinear estimate also outperforms the in-sample linear model shown in Panel C of Table 6. These findings indicate that the strong pricing power of the nonlinear HI-SDF is more than a mere in-sample result.

4 Economic Interpretation

This section provides a parametric example in which the HI-SDF emerges in equilibrium. This framework builds on the financial accelerator models of Gertler and Kiyotaki (2010) and Nuguer (2018), and is intentionally over-simplified for analytical tractability of the SDF.

4.1 General Setup

Time is discrete; it is indexed by $t = 0, 1, 2...$ in this paper. The economy is endowed with a Lucas (1978) tree that yields an aggregate output of Y_t in each period t. I assume that the log growth

[12]Results are similar if I estimate the SDF using nonequity portfolios to price equity portfolios.

rate of output for the period from time t to $t+1$ follows the process

$$\log\left(\frac{Y_{t+1}}{Y_t}\right) = \mu_y + \sigma_y \varepsilon_{t+1},$$

where ε_{t+1} is an i.i.d. random variable drawn from normal distribution $N(0,1)$, and μ_y and σ_y represent the mean and volatility of the growth in endowment Y_t, respectively. The price of the Lucas tree is denoted by Q_t and the return of holding the tree is denoted by R_t; this can be defined as

$$R_{t+1} = \frac{Q_{t+1} + Y_{t+1}}{Q_t}.$$

A representative family in the economy consists of measure one continuum of members. The economy has fraction $f < 1$ households and $1 - f$ bankers. I assume that all family members have a constant relative risk aversion (CRRA) utility. Each banker i manages an intermediary i; I will discuss this later. I also assume that in each period, intermediaries exit the economy with probability $1 - \sigma > 0$. They pay their remaining net worth to households upon exit. They are immediately replaced by new intermediaries, who are endowed with ξ fraction of the total existing bankers' assets. This setup avoids the case of intermediaries strategically over-accumulate their net worth to become financially unconstrained. As in Gertler and Kiyotaki (2010), I assume that all intermediaries pay out their earnings equally across the entire household.

Three types of financial assets exist in the economy. Households can have deposits D_t with intermediaries at the rate $R_t^d > 0$. A zero-supply bond is traded at the price Q_t^b at each period t. The last type of asset is a claim on the aggregate output Y_t from the Lucas tree. Since Y_t is exposed to the aggregate shock ε_{t+1}, the claim can be considered a risky asset with gross return R_{t+1}. In agreement with He and Krishnamurthy (2013), I assume that only intermediaries have access to risky assets and bonds.

Households Households in the economy have a CRRA utility that maximizes their discounted lifetime utilities by choosing the consumption stream $\{C_t\}_{t=0}^{\infty}$ and deposits $\{D_t\}_{t=0}^{\infty}$. The value function of the households can be expressed as

$$\max_{\{C_t\}_{t=0}^{\infty}} J_0 = E_0\left(\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma}\right) \tag{16}$$

$$s.t\ C_t + D_t = D_{t-1}R_t^d + \Pi_t \text{ for all } t \geq 1$$

$$C_0 + D_0 = \Pi_0, \tag{17}$$

where $\beta \in (0,1)$ is the discount rate and Π_t is the endogenous stream of income transferred from intermediaries when they exit. In the model, the households' net worth is equal to the amount of deposits D_t. This depends on both the deposit rate R_t^d and bank transfer Π_t of each period t. At time $t = 0$, I assume that households are endowed with income $\Pi_0 > 0$ and that their budget constraint is $C_0 + D_0 = \Pi_0$.

Intermediaries All intermediaries raise deposits from households and trade the bond and risky assets to maximize their net worth n_t^i upon exit,

$$\max_{\left\{s_t^i, d_t^i, b_t^i\right\}} V_t^i = E_t \sum_{j=1}^{\infty} (1 - \sigma)\, \sigma^{j-1} \pi_{t+j} n_{t+j}^i, \tag{18}$$

where $\sigma > 0$ is the intermediary's survival probability, π_t is the households' MRS, and n_t^i is intermediary i's net worth. Let s_t^i denote the fraction of the Lucas tree that intermediary i chooses to invest. d_t^i and b_t^i are the intermediary's investments in deposits and bonds, respectively. Each intermediary maximizes the objective function stated in equation (18) subject to three constraints.

First, the balance sheet equation (19) for each intermediary restricts the sum of debt and equity, making it equal to the market value of the asset invested,

$$s_t^i Q_t = n_t^i + d_t^i + Q_t^b b_t^i. \tag{19}$$

The left-hand side (LHS) equation (19) is the total asset, specified by the investment in the risky asset $s_t^i Q$. The total liability on the right-hand side (RHS) includes the investment in bonds $Q_t^b b_t^i$ and deposits d_t^i, in addition to the net worth n_t^i. I restrict $s_t^i > 0$ to prevent short-selling of any fraction of the tree, but intermediaries can take both short and long positions in bond $b_t^i \in \mathbb{R}$.

Second, equation (20) specifies the law of motion of the net worth of individual intermediary i at time t,

$$n_t^i = s_{t-1}^i (Q_t + Y_t) - Q_t^b b_{t-1}^i - R_t^d d_{t-1}^i. \tag{20}$$

It depends on the payoff from the assets intermediary i invested at time $t - 1$. The first term is the time-t payoff from investing s_{t-1}^i fraction of the tree at time $t - 1$. The second and third terms capture the payoff or payment in the bond and deposits, respectively.

Third, all intermediaries are financially constrained. More specifically, intermediary i faces the following financing constraint

$$V_t^i \geq \kappa^i \left[s_t^i Q_t + Q_t^b \max\left(b_t^i, 0\right) \right]. \tag{21}$$

Constraint (21) restricts the value V_t^i to be at least equal to the κ^i fraction of the total assets consisting of the investment in both the tree $s_t^i Q_t$ and bond $Q_t^b b_t^i$. This constraint can be motivated by the limited enforcement argument in Gertler and Kiyotaki (2010): at each time t, intermediary i can divert a fraction κ^i of the total asset and default on its debt.

The heterogeneity among intermediaries is captured by the differences in the tightness of the constraint or κ^i. I assume two types of intermediaries with $\kappa^i \in \left\{\kappa^L, \kappa^H\right\}$ where $\kappa^H < \kappa^L$. The "high" type (κ^H) is less constrained than the "low" type $\left(\kappa^L\right)$ *ex-ante*, and has higher leverage *ex-post*, as will be shown in equilibrium.[13] For analytical tractability, I assume that financing constraints are always binding for both types of banks. The model with occasionally binding constraints can be numerically solved and provides similar economic insight for the pricing wedge. See Appendix B.2 for more discussions.

4.2 Equilibrium Characteristics

I characterize the equilibrium and present key results from the model in this section. The definition of equilibrium and all derivations are provided in Appendix B for simplicity of the main text.

4.2.1 State Variables

As in the financial accelerator models of Gertler and Kiyotaki (2010) and Maggiori (2017), the equilibrium with financial frictions hinges on the wealth distribution across all types of agents. First, it depends on the fraction of the total wealth in the economy held as equity in the intermediary sector; this is captured by the intermediary's aggregate leverage, defined as

$$\phi_t \equiv \frac{Q_t}{N_t^L + N_t^H}. \tag{22}$$

Second, since both types of intermediaries are financially constrained, the equilibrium also depends on the net worth share of the low-type intermediaries, defined as

$$\psi_t \equiv \frac{N_t^L}{N_t^L + N_t^H}. \tag{23}$$

where $N_t^i = (\sigma + \xi)\left[S_{t-1}^i (Q_t + Y_t) + Q_t^b B_{t-1}^i\right] - \sigma R_t^d D_{t-1}^i$ is type i's aggrgate net worth. The details on the verification of these state variables are in Appendix B.2.

4.2.2 Pricing Wedge $\Omega(\psi_t, \phi_t)$

The following proposition, derived in Appendix B.2, characterizes the first-order conditions for intermediaries and the associated pricing wedge $\Omega(\psi_t, \phi_t)$.

[13] Table A4 in Appendix B.2 shows that empirically low (high) type corresponds to BHCs (BDs) according to the volatility of their total assets.

Proposition 1 *In equilibrium, the first-order moment conditions are*

$$\nu_t = E_t \left(\pi_{t+1} \Omega \left(\psi_{t+1}, \phi_{t+1} \right) R^d_{t+1} \right) \quad (24)$$

$$\mu_t = E_t \left(\pi_{t+1} \Omega \left(\psi_{t+1}, \phi_{t+1} \right) \left[R_{t+1} - R^d_{t+1} \right] \right), \quad (25)$$

where the pricing wedge $\Omega(\psi_t, \phi_t)$ can be expressed as

$$\Omega \left(\psi_t, \phi_t \right) = 1 - \sigma + \sigma \left(\nu_t + \phi_t \left(\psi_t + (1 - \psi_t) \frac{\kappa^L - \mu_t}{\kappa^H - \mu_t} \right)^{-1} \mu_t \right). \quad (26)$$

If $1 > \kappa^L > \kappa^H$ and $\sigma \in (0,1)$, $\partial \Omega(\psi_t, \phi_t) / \partial \phi > 0$ and $\partial \Omega(\psi_t, \phi_t) / \partial \psi_t < 0$.

Proposition 1 states that HI-SDF emerges in equilibrium, and the pricing wedge, characterized in equation (26), is increasing in the leverage ϕ_t but decreasing (nonlinearly) in the low type's net worth share ψ_t. Intuitively, the marginal value of wealth increases when the economy is in a bad state where most of the wealth is in the hands of low-type intermediaries, who are more constrained and more effectively risk averse. The equilibrium risk premium is thus increasing in ψ_t so as to compensate the extra risk that the low-type intermediaries take on during bad times. As a result, an asset that pays off when the net worth share is high (bad times) is considered to be a hedge against the aggregate risk and earns a smaller risk premium – this is consistent with the empirical evidence (Panel A of Table 1).

Why does $\Omega(\psi_t, \phi_t)$ exist and why can it be nonlinear? Through the lens of this parametric example, I can investigate the source of the existence and the nonlinearity of the pricing wedge $\Omega(\cdot)$ based its analytical solution in equation (26) .

First, the pricing wedge arises in equilibrium due to the presence of financial frictions. Appendix B.2 shows that $\mu_t = 0$ when intermediaries are not subject to financial constraints ($\kappa^L = \kappa^H = 0$) or when neither of the financial constraints is binding. As a result, equation (26) implies that $\Omega(\cdot) = 1 - \sigma$ is constant over time and does not depend on ψ_t or ϕ_t in these cases.

Second, the nonlinearity of the $\Omega(\cdot)$ is due to the *heterogeneity* of intermediaries. To see this, in the special case where the two types of intermediaries are *ex-ante* identical $\kappa^L = \kappa^H = \kappa$, equation (26) becomes,

$$\Omega(\phi_t) = 1 - \sigma + \sigma \kappa \phi_t,$$

which is linear in aggregate leverage ϕ_t. Therefore, $\Omega(\psi_t, \phi_t)$ is nonlinear in ψ_t when heterogeneity presents ($\kappa^L > \kappa^H$) and when both financial constraints are binding.

4.2.3 Role of Financial Constraints

Proposition 1 shows that the nonlinear pricing wedge arises in equilibrium due to financial constraints. The natural question is, what determines the *functional form* of the wedge? In this section, I show that it hinges on the type of financial constraints.

Once I use the equilibrium value function and re-arrange the terms, the financial constraint (21) becomes

$$\frac{s_t^i Q_t}{n_t} \leq \frac{\Omega_t}{\kappa^i}. \tag{27}$$

Depending on the value of κ^i for both types of intermediaries, constraint (27) reflects the different types of constraints used in the literature.

1. **Frictionless economy**. $\kappa^i = 0$ corresponds to the frictionless economy, where both types of intermediaries are unconstrained.

2. **Margin constraint**. The case of $0 < \kappa^i < \Omega(0, \phi)$ corresponds to the margin constraint as in Brunnermeier and Pedersen (2008). To see this, if I re-arrange the terms in constraint (27), the financial constraint can be re-stated as

$$\underbrace{d_t^i + Q_t^b b_t^i}_{\text{Debt}} \leq \underbrace{\left(\frac{\Omega_t}{\kappa^i} - 1\right)}_{\text{Margin Rsq.}} n_t^i. \tag{28}$$

The RHS corresponds to the margin requirement for collateralized borrowing, when the intermediary can borrow only up to a fraction $\frac{\Omega_t}{\kappa^i} - 1$ of the current net worth in the book. Condition $\kappa^i < \Omega(0, \phi)$ ensures that margin requirement is feasible ($\frac{\Omega_t}{\kappa^i} - 1 > 0$) for all t.

3. **Limited participation constraint**. The case of $\infty > \kappa^i > \Omega(1, \phi)$ corresponds to the limited participation constraint, where the intermediary cannot take any leverage but can instead invest $\frac{\Omega_t}{\kappa^i} < 1$ fraction of own net worth n_t. The extreme case $\kappa^i = \infty$ corresponds to the market inactivity, where the intermediary is not allowed to invest in any risky assets.

In addition to the constraints mentioned above, I study two other types of constraints that cannot be generally represented by equation (27).

4. **Value-at-Risk (VaR) constraint.** The VaR constraint is expressed as follows,

$$s_t^i \sigma_y \leq \kappa^i V_t^i, \tag{29}$$

which specifies that the maximum volatility of the asset, captured by $s_t^i \sigma_y$, cannot exceed the κ^i fraction of the franchise value V_t^i.[14]

[14] VaR constraint has been studied in intermediary models, mostly in the continuous-time setting. Examples include Adrian and Shin (2010), Danielsson, Shin, and Zigrand (2004) and Etula (2013).

5. **Skin-in-the-game constraint**. The constraint is expressed as follows,

$$d_t^i \leq \kappa^i n_t^i, \tag{30}$$

which specifies that the total deposit d_t^i from households cannot exceed the $\kappa^i n_t^i$. This latter constraint is a variant of the equity constraint studied in He and Krishnamurthy (2013). It specifies that households' wealth allocated to intermediary i, captured by deposits d_t^i in this paper, cannot exceed the κ^i fraction of the intermediary wealth n_t^i at t.

While the wedge cannot be analytically characterized under VaR and Skin-in-the-game constraints, the HI-SDF can be numerically computed according to the computation procedure described in Appendix B.3. The parameter values are specified in Table 8. The preference parameters (β, γ) are from the empirical analysis (Table 2), and values of parameters (σ, ξ) are from Gertler and Kiyotaki (2010). The parameters (μ_y, σ_y) are set to be $(0.21, 1)$ such that the equilibrium return R_t matches the average CRSP value-weighted stock return, and the parameters of constraints $\left(\kappa^L, \kappa^H\right)$ are set to match the volatility of total assets for BHCs and BDs.[15]

Parameter Values of the Model

β	γ	σ	ξ	μ_y	σ_y	κ^L	κ^H
0.99	26	0.972	0.003	0.21	1	0.41	0.05

Table 8: This table reports parameter values for computing the pricing wedge in the model with margin constraint.

Figure 9 plots the model-implied functional form of Ω_t over the net worth share ψ for different constraints. It's clear that the functional form varies in other cases, and the convexity or concavity of Ω_t depends on the *type* of constraints specified in the model.

4.2.4 From Model to Empirical Analysis: Pricing Constraints

One question remains: *which type* of constraint used in the literature is more supported by the data? To answer this question, I evaluate the model's pricing ability following the same two-step procedure used in the empirical analysis (Section 2.2.1), except that the functional form Ω_t is numerically estimated by the model as depicted in Figure 9. For each constraint, I compute the Hansen–Jagannathan (HJ) distance, and the cross-sectional R^2 from the nonlinear GMM. Table 9 reports the results. Panel A shows that the model with margin constraint is more supported by the data over other constraints when pricing equity portfolios. When pricing nonequity portfolios, interestingly, the data favors the VaR constraint, with the largest R^2 and smallest HJ-distance

[15]The values of $\left(\kappa^L, \kappa^H\right)$ depend on the type of constraints and they are set to match the volatility of assets.

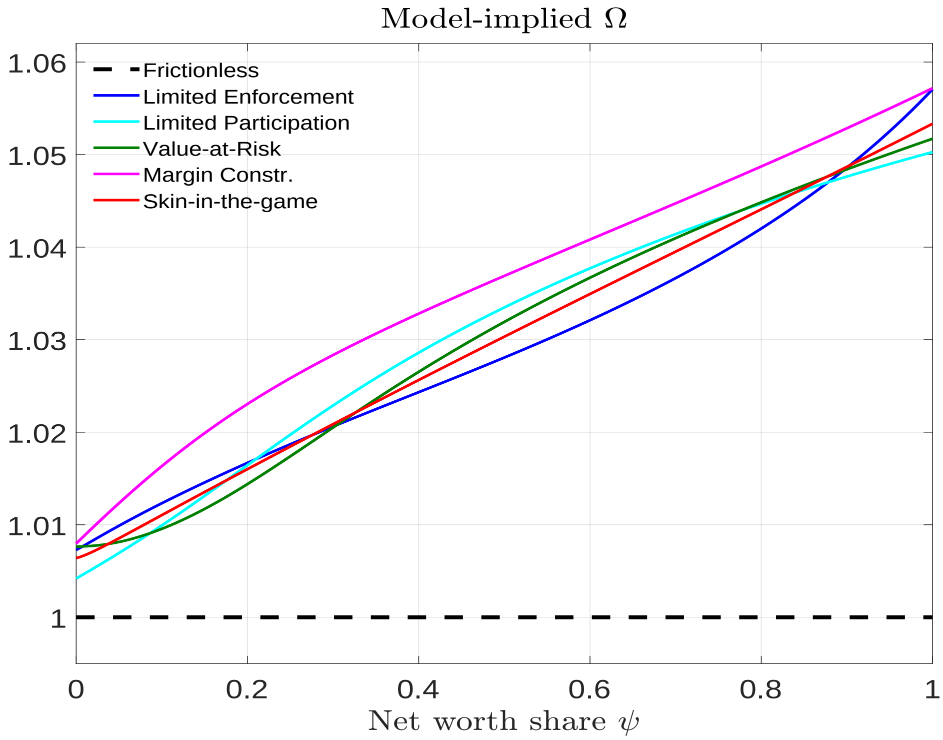

Figure 9: This figure plots the model's implied pricing wedge $\Omega(\psi, \phi)$ over ψ with the frictionless economy (black), limited enforcement constraint (dark blue), margin constraint (pink), VaR constraint (green), equity constraint (red) and limited participation constraint (light blue). The aggregate intermediary leverage ψ is fixed to be its mean value. The parameter values are set according to Table 8.

over other constraints (Panel B). While the HI-SDF with a nonparametric pricing wedge still delivers the highest pricing power in both cases, the R^2s produced from the model subject to the best-fitting financial constraint are fairly close to the R^2 from the semiparametric estimate: 62% versus 70% for equity and 61% versus 68% for nonequity returns. These results again emphasize the role of financial constraints in improving explanatory power for cross-sectional variation in expected returns across a wide range of test assets.

5 Conclusion

This paper uncovers an important linkage between the heterogeneity among intermediaries and asset prices. I propose an empirically testable HI-SDF that depends on not only the household's consumption growth but also an additional intermediary pricing wedge, a nonlinear function of the aggregate leverage and net worth share of individual intermediaries. I find that the net worth-share-beta sorted portfolios have a negative per-annum return spread of 5.8%; a spread cannot be explained by other alternative asset pricing models.

I let the data speak and estimate the HI-SDF semiparametrically and show that it exhibits sig-

Specification Errors and Cross-Sectional R^2, Nonlinear Tests

	Panel A **Equity Portfolio**			Panel B **Nonequity Portfolios**		
Constraint	R^2	Unconditional HJ-Dist	Conditional HJ-Dist	R^2	Unconditional HJ-Dist	Conditional HJ-Dist
Nonparametric Ω_t	0.70	0.411	0.416	0.68	0.427	0.512
Margin Constraint	**0.62**	0.432	0.421	0.47	0.472	0.592
Value-at-Risk	0.42	0.471	0.497	**0.61**	0.435	0.548
Limited Participation	0.31	0.499	0.552	0.29	0.501	0.627
Frictionless	0.27	0.532	0.576	0.24	0.512	0.674

Table 9: This table reports the cross-sectional R^2, adjusted unconditional and conditional HJ distance for estimated SDF of model under financial constraint specified in the first column. The conditional HJ distance is computed conditional on cay. "Nonparametric" refers to the empirical HI-SDF with nonparametrically estimated pricing wedge. Other types of constrains are defined in the text (section 4.2.3). The sample spans the period 1970Q1 to 2019Q4.

nificant pricing power to explain the cross-section variation in multiple test assets. The estimated HI-SDF outperforms other commonly studied asset pricing models in terms of smaller specification errors, such as cross-sectional R^2 and adjusted HJ distance which penalizes the model with more free parameters. The nonlinearity of the pricing wedge substantially improves HI-SDF's fit to asset return data. The results are robust with a host of additional checks.

I provide a parametric example that the HI-SDF emerges in equilibrium. Through the lens of this framework, I show analytically that the pricing wedge exists due to the presence of financial frictions, its nonlinearity originates from the heterogeneity in financing constraints, and the functional form of the HI-SDF hinges on the types of constraints that intermediaries are facing. While the calibrated model delivers slightly lower pricing power than the semiparametric estimate, the data favors margin (VaR) constraint when pricing equity (nonequity) portfolios, which highlights the role of financial constraints in improving explanatory power for cross-sectional variation in expected returns across a wide range of test assets. Although the calibrated model produces the pricing wedge reasonably close to the semiparametric estimate, the current setup is over-simplified intentionally for analytical tractability and is not able to simulate the dynamics of asset prices in the presence of multiple shocks or other types of intermediaries (e.g., mutual funds). I leave this challenge for future research.

References

Adrian, T., and N. Boyarchenko (2012): “Intermediary leverage cycles and financial stability,” Unpublished paper, Federal Reserve Bank of New York.

——— (2013): “Intermediary balance sheets,” *FRB of New York Staff Report*, (651).

Adrian, T., P. Colla, and H. S. Shin (2013): “Which financial frictions? Parsing the evidence from the financial crisis of 2007 to 2009,” *NBER Macroeconomics Annual*, 27(1), 159–214.

Adrian, T., E. Etula, and T. Muir (2014): “Financial intermediaries and the cross-section of asset returns,” *The Journal of Finance*, 69(6), 2557–2596.

Adrian, T., and H. S. Shin (2010): “Liquidity and leverage,” *Journal of financial intermediation*, 19(3), 418–437.

Ai, C., and X. Chen (2003): “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions,” *Econometrica*, 71, 1795–1843.

——— (2007): “Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables,” *Journal of Econometrics*, 141, 5–43.

Ai, H., R. Bansal, and K. Li (2012): “Financial intermediary, asset prices and the real economy.,” Unpublished paper.

Avraham, D., P. Selvaggi, and J. Vickery (2012): “A Structural View of US Bank Holding Companies,” *Economic Policy Review*, p. 65.

Bansal, R., and A. Yaron (2004): “Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles,” *Journal of Finance*, 59(4), 1481–1509.

Borri, N., and A. Verdelhan (2011): “Sovereign risk premia,” Unpublished manuscript, MIT Sloan School.

Breeden, D. (1979): “An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities,” *Journal of Financial Economics*, 7, 265–296.

Brunnermeier, M., and Y. Sannikov (2014): “A Macroeconomic Model with a Financial Sector,” *American Economic Review*, 104, 379–421.

Brunnermeier, M. K., and L. H. Pedersen (2008): “Market liquidity and funding liquidity,” *The review of financial studies*, 22(6), 2201–2238.

Campbell, J. Y. (1991): “A Variance Decomposition for Stock Returns,” *Economic Journal*, 101, 157–179.

Chabakauri, G. (2015): “Asset pricing with heterogeneous preferences, beliefs, and portfolio constraints,” *Journal of Monetary Economics*, 75, 21–34.

Chan, Y. L., and L. Kogan (2002): “Catching Up With the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices,” *Journal of Political Economy*, 110, 1255–1285.

Chen, X., J. Favilukis, and S. C. Ludvigson (2014): “An Estimation of Economic Models with Recursive Preferences,” *Quantitative Economics*, 4(1), 39–83.

Chen, X., O. Linton, and I. van Keilegom (2003): “Estmiation of Semiparametric Models when the Criterion Function is Not Smooth,” *Econometrica*, 71, 1591–1608.

Chen, X., and S. C. Ludvigson (2009): “Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models,” *Journal of Applied Econometrics*, 24(7), 1057–1093.

Christensen, T. M. (2017): “Nonparametric stochastic discount factor decomposition,” *Econometrica*, 85(5), 1501–1536.

Cochrane, J. H. (2001): *Asset Pricing*. Princeton University Press, Princeton, NJ.

Coimbra, N., and H. Rey (2017): “Financial cycles with heterogeneous intermediaries,” Discussion paper, National Bureau of Economic Research.

Constantinides, G. M., J. C. Jackwerth, and A. Savov (2013): “The puzzle of index option returns,” *Review of Asset Pricing Studies*, p. rat004.

Danielsson, J., H. S. Shin, and J.-P. Zigrand (2004): “The impact of risk regulation on price dynamics,” *Journal of Banking & Finance*, 28(5), 1069–1087.

Danthine, J.-P., and J. B. Donaldson (2002): “Labour Relations and Asset Returns,” *Review of Economic Studies*, 69(1), 41–64.

Diamond, D. W., and P. H. Dybvig (1983): “Bank runs, deposit insurance, and liquidity,” *Journal of political economy*, 91(3), 401–419.

Drechsler, I., A. Savov, and P. Schnabl (2018): “A model of monetary policy and risk premia,” *The Journal of Finance*, 73(1), 317–373.

Du, W., B. Hébert, and A. W. Huber (2022): “Are Intermediary Constraints Priced?,” *The Review of Financial Studies*.

Escanciano, J. C., S. Hoderlein, A. Lewbel, O. Linton, and S. Srisuma (2021): “Nonparametric Euler equation identification and estimation,” *Econometric Theory*, 37(5), 851–891.

Etula, E. (2013): “Broker-dealer risk appetite and commodity returns,” *Journal of Financial Econometrics*, 11(3), 486–521.

Gertler, M., and N. Kiyotaki (2010): “Financial intermediation and credit policy in business cycle analysis,” *Handbook of monetary economics*, 3(3), 547–599.

Gertler, M., N. Kiyotaki, and A. Prestipino (2016): “Wholesale banking and bank runs in macroeconomic modeling of financial crises,” *Handbook of Macroeconomics*, 2, 1345–1425.

Gomez, M. (2016): “Asset prices and wealth inequality,” .

Greenwald, D., M. Lettau, and S. C. Ludvigson (2014): “Origins of Stock Market Fluctuations,” National Bureau of Economic Research Working Paper No. 19818.

Hall, P., J. L. Horowitz, and B. Y. Jing (1995): “On Blocking Rules for the Bootstrap with Dependent Data,” *Biometrika*, 82, 561–574.

Hansen, L. P. (1982): “Large Sample Properties of Generalized Methods of Moments Estimators,” *Econometrica*, 50, 1029–54.

Hansen, L. P., and R. Jagannathan (1997): “Assessing Specification Errors in Stochastic Discount Factor Models,” *Journal of Finance*, 52, 557–590.

He, Z., B. Kelly, and A. Manela (2017): "Intermediary asset pricing: New evidence from many asset classes," *Journal of Financial Economics*, 126(1), 1–35.

He, Z., and A. Krishnamurthy (2012): "A Model of Capital and Crises," *Review of Economic Studies*, 79, 735–777.

——— (2013): "Intermediary Asset Pricing," *American Economic Review*, 103(2), 732–770.

Hodrick, R. (1992): "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 357–386.

Horowitz, J. L. (2003): "The Bootstrap," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5. Elsevier Science B.V., North Holland.

Kandel, S., and R. F. Stambaugh (1995): "Portfolio Inefficiency and the Cross-Section of Expected Returns," *Journal of Finance*, 50, 157–184.

Kargar, M. (2021): "Heterogeneous intermediary asset pricing," *Journal of Financial Economics*, 141(2), 505–532.

Lettau, M., and S. C. Ludvigson (2001): "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

——— (2004): "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption," *American Economic Review*, 94(1), 276–299.

Lettau, M., S. C. Ludvigson, and S. Ma (2019): "Capital share risk in US asset pricing," *The Journal of Finance*, 74(4), 1753–1792.

Li, K., and C. Xu (2022): "Asset pricing with a financial sector," *Financial Management*.

Lucas, R. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–1446.

Maggiori, M. (2017): "Financial intermediation, international risk sharing, and reserve currencies," *American Economic Review*, 107(10), 3038–3071.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012): "Currency momentum strategies," *Journal of Financial Economics*, 106(3), 660–684.

Miranda, M. J., and P. L. Fackler (2004): *Applied computational economics and finance*. MIT press.

Muir, T. (2017): "Financial crises and risk premia," *The Quarterly Journal of Economics*, 132(2), 765–809.

Newey, W. K., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1557–1569.

Nozawa, Y. (2017): "What Drives the Cross-Section of Credit Spreads?: A Variance Decomposition Approach," *The Journal of Finance*, 72(5), 2045–2072.

Nuguer, V. (2018): "Financial intermediation in a global environment," *45th issue (September 2016) of the International Journal of Central Banking*.

Nuño, G., and C. Thomas (2017): "Bank leverage cycles," *American Economic Journal: Macroeconomics*, 9(2), 32–72.

Appendix A : Additional Tables and Figures

Portfolios Double-sorted by ψ beta and ϕ beta

	Low ψ	Mid ψ	High ψ	High-Low
Low ϕ	7.23	5.24	2.85	-4.38^{*}
	(1.78)	(1.97)	(1.86)	(-1.89)
Mid ϕ	8.24	6.99	4.24	-4.00^{**}
	(2.18)	(2.01)	(1.87)	(-1.99)
High ϕ	10.28	8.22	6.40	-3.88^{**}
	(2.22)	(2.10)	(2.05)	(-2.12)

Table A1: This table reports the average annualized excess return for portfolios double sorted by net worth share beta and aggregate leverage beta. The Low corresponds to assets at 0-33 percentile. The Mid corresponds to assets at 34-66 percentile. The High corresponds to assets at 67-100 percentile.The NW t-statistics is reported in parenthesis. The sample spans the period 1970Q1 to 2019Q4.

Specification Error Robustness Checks: HJ Distance

	Panel A Alternative Test Assets		Panel B: Consumption for Wealthy Households			Panel C: Alternative Instruments		
Model	**Currencies**	**Industries**	**Equity**	**Nonequity**	**All Asset**	**No** cay	**Add** ψ_{t-1}	**No**$RREL$
HI-SDF	0.391	0.442	0.513	0.476	0.577	0.499	0.431	0.457
Rep. Int. SDF, BD Only	0.458	0.476	0.544	0.553	0.598	0.524	0.481	0.492
Rep. Int. SDF, BHC Only	0.431	0.499	0.587	0.501	0.611	0.531	0.479	0.487
Recursive, No intermediary	0.512	0.598	0.611	0.532	0.678	0.565	0.555	0.533
CRRA Utility, No intermediary	0.523	0.672	0.651	0.598	0.728	0.622	0.587	0.581

Table A2: The table reports adjusted HJ distance for each specification. Panel A uses alternative test assets include currency (FX) momentum portfolios from Menkhoff et al. (2012) and 49 industry equity portfolios. Panel B uses proxy for consumption for household of top 10 percent wealth distribution. Panel C uses different set of instruments in the semiparametric estimation. See text for more details. In all cases, the sample spans the period 1970Q1 to 2019Q4.

Robustness Checks

Panel A: Pre-crisis Sample			**Panel B: Banks v.s. Nonbanks**		
Preference Parameter Estimates, All Assets			Preference Parameter Estimates, All Assets		
	γ	β		γ	β
$\mathbf{W}=\mathbf{I}$	39.7	0.987	$\mathbf{W}=\mathbf{I}$	47.1	0.989
	$(21.3, 56.9)$	$(0.976, 0.999)$		$(31.4, 66.7)$	$(0.972, 0.999)$
$\mathbf{W}=\mathbf{G}_T^{-1}$	29.1	0.995	$\mathbf{W}=\mathbf{G}_T^{-1}$	36.7	0.992
	$(25.8, 64.3)$	$(0.991, 0.999)$		$(28.2, 59.7)$	$(0.988, 0.998)$
Estimates of Factor Risk Prices λ , All Assets			Estimates of Factor Risk Prices λ , All Assets		
Δc_t	$\log\widehat{\Omega}(\psi_t, \phi_t)$	$\bar{R}^2$	Δc_t	$\log\widehat{\Omega}(\psi_t, \phi_t)$	$\bar{R}^2$
0.44^{*}	-0.51^{**}	0.72	0.48^{**}	-0.59^{**}	0.70
(1.77)	(-1.97)	$[0.61, 0.87]$	(1.99)	(-2.41)	$[0.63, 0.81]$
$[0.17, 0.61]$	$[-1.66, -0.41]$		$[0.03, 0.72]$	$[-1.87, -0.41]$	
Panel C: **Standalone Broker-dealers**			Panel D: **Out-of-sample Tests**		
Preference Parameter Estimates, All Assets			Preference Parameter Estimates, Equity		
	γ	β		γ	β
$\mathbf{W}=\mathbf{I}$	34.2	0.978	$\mathbf{W}=\mathbf{I}$	27.4	0.991
	$(14.1, 44.2)$	$(0.972, 0.999)$		$(11.2, 48.7)$	$(0.981, 0.999)$
$\mathbf{W}=\mathbf{G}_T^{-1}$	27.5	0.998	$\mathbf{W}=\mathbf{G}_T^{-1}$	29.3	0.994
	$(21.1, 62.3)$	$(0.993, 0.999)$		$(15.7, 41.2)$	$(0.987, 0.999)$
Estimates of Factor Risk Prices λ , All Assets			Estimates of Factor Risk Prices λ , All Assets		
Δc_t	$\log\widehat{\Omega}(\psi_t, \phi_t)$	$\bar{R}^2$	Δc_t	$\log\widehat{\Omega}(\psi_t, \phi_t)$	$\bar{R}^2$
0.33	-0.79^{**}	0.81	0.19^{*}	-1.07^{**}	0.71
(1.24)	(-2.09)	$[0.72, 0.92]$	(1.78)	(-2.03)	$[0.59, 0.89]$
$[0.27, 0.41]$	$[-1.21, -0.71]$		$[0.02, 2.38]$	$[-1.91, -0.73]$	

Table A3: The table reports the parameter estimates and Fama-Macbeth regression estimates. The preference parameters are estimated based on the two-step SMD procedure described in the text. The second-step parameters are obtained by minimizing the GMM criterion with weighting matrix indicated in the first column. The 95 percent confidence intervals are reported in the parenthesis. The price of risks are estimated using Fama-Macbeth regressions. GMM t-statistics are reported in parenthesis and bootstrap confidence intervals are reported in brackets. See texts for more details. The sample spans the period 1970Q1 to 2019Q4. *sig. at 10%. **sig. at 5%. ***sig. at 1% according to the GMM t-statistics.

Appendix B – Details on Theoretical Framework

In this section, I characterize the solution to the problem of low type intermediary, and verify (ψ_t, ϕ_t) are state variables of the economy which are omitted in the main text. The problem of high type intermediary is analogous.

B.1 Equilibrium Definition

By aggregating across the individual type-i intermediary's net worth from equation (20), the aggregate type-i's net worth N_t^i can be expressed as

$$N_t^i = (\sigma + \xi) \left[S_{t-1}^i (Q_t + Y_t) + Q_t^b B_{t-1}^i \right] - \sigma R_t^d D_{t-1}^i, \tag{31}$$

where S_{t-1}^i is the aggregate investment in the Lucas tree for type i intermediaries, and B_{t-1}^i and D_{t-1}^i are the type-i intermediaries' aggregate investment in bonds and deposits, respectively. The first term on the RHS of (31) captures the value of the tree invested in the last period among the surviving intermediaries plus the start-up equity of the new intermediaries. The second term is associated with the return on deposits that the intermediaries need to pay back to households.

The aggregate net transfers from intermediaries i to households are

$$\Pi_t^i = (1 - \sigma) N_t^i - \xi S_t^i Q_t, \tag{32}$$

where the first term captures the transfer of the exiting intermediaries' net worth and the second term captures the start-up net worth of the new intermediaries, which is fraction ξ of the total asset $S_t^i Q_t$ at time t.

An equilibrium is a collection of asset prices $\left\{Q_t, R_t^d, Q_t^b\right\}_{t=0}^{\infty}$ and set of quantities $\{C_t, s_t^i, b_t^i, d_t^i, n_t^i\}$ and $\{N_t^i, D_t, B_t^i\}$ for all t, $i \in \{L, H\}$, such that at given prices, (a) households maximize their utility according to (16), (b) each bank maximizes its equity upon exit, and (c) all markets are clear:

$$C_t = Y_t, \text{ for all } t \tag{33}$$

$$\sum_{i \in \{L,H\}} S_t^i = 1, \quad \sum_{i \in \{L,H\}} B_t^i = 0, \quad \sum_{i \in \{L,H\}} d_t^i = D_t \text{ for all } t \tag{34}$$

$$\begin{aligned} N_t^i &= (\sigma + \xi) \left[S_{t-1}^i (Q_t + Y_t) + Q_t^b B_{t-1}^i \right] - \sigma R_t^d D_{t-1}^i \text{ , for all } i \text{ and } t \\ S_t^i Q_t^i &= N_t^i + D_t^i + Q_t^b B_t^i \text{ , for all } i \text{ and } t \\ \Pi_t^i &= (1 - \sigma) N_t^i - \xi S_t^i Q_t \text{ , for all } i \text{ and } t. \end{aligned} \tag{35}$$

Equation (33) is the market clearing condition for the consumption good. It states that the households' consumption should be equal to the Lucas tree's output at each period t. Equation (34) gives the market clearing condition for all types of assets. More specifically, the sum of all shares invested in the Lucas tree should be equal to 1, and the total demand for bonds and deposits should be equal to the supply, that is, 0 and D_t, respectively. Equation (35) is the accounting identity for the balance sheet of type-i intermediaries. Finally, the marking clearing conditions include the evolution of the aggregate net worth N_t^i expressed in equation (31) and the net transfer from the exiting intermediaries stated in equation (32) for each type i.

B.2 Derivation of Intermediary's Problem and Proof of Proposition 1

For the low type intermediary, since they are more constrained and they fund their investment in both risky asset (tree) and interbank bonds using deposits from households. Their problem can be recursively expressed as follows,

$$
\begin{gathered}
V_t^L = \max_{\left\{s_t^i, d_t^i, b_t^i\right\}} E_t \pi_{t+1}^L \left\{(1-\sigma) n_t^L + \sigma V_{t+1}^L\right\} \\
s.t\ s_t^L Q_t + Q_t^b b_t^L = n_t^L + d_t^L \\
n_t^i = s_{t-1}^L (Q_t + Y_t) + Q_t^b b_{t-1}^L - R_t d_{t-1}^L \\
V_t^L \geq \kappa^L \left(s_t^L Q_t + Q_t^b b_t^L\right)
\end{gathered}
$$

For notation simplicity, I drop the superscript i unless otherwise stated. I first guess and express the franchise value V is linear in assets and liabilities. The guess is also used in Nuguer (2018).

$$
V_t = \nu_{s,t} s_t + \nu_{b,t} b_t - \nu_t d_t \tag{36}
$$

where $\nu_{s,t}$, $\nu_{b,t}$ and $-\nu_t$ is the marginal value of holding the tree, interbank bond, and deposit respectively. Let χ_t denote the Lagrangian multiplier of the financial constraint, the first-order conditions with respective to $\{s, d, b, \chi\}$ are

$$
\begin{aligned}
\nu_{s,t} &= \chi_t \left(\nu_{s,t} - \kappa Q_t\right) \\
\nu_{b,t} &= \chi_t \left(\nu_{b,t} - \kappa Q_{bt}\right) \\
\nu_t - \chi_t (1 - R_t) &= 0 \\
\kappa \left(Q_t s_t + Q_t^b b_t\right) &= \nu_{s,t} s_t + \nu_{b,t} b_t - \nu_t d_t
\end{aligned} \tag{37}
$$

After some tedious terms re-arranging, I obtain the following equations

$$
\nu_{b,t} - \nu_t = \frac{\chi_t}{1+\chi_t} \kappa Q_{bt} \tag{38}
$$

$$
\frac{\nu_{s,t}}{Q_t} = \frac{\nu_{b,t}}{Q_t^b} \tag{39}
$$

$$
\left[\kappa - \left(\frac{\nu_{s,t}}{Q_t} - \nu_t\right)\right] Q_t s_t + \left[\kappa - \left(\frac{\nu_{b,t}}{Q_t^b} - \nu_t\right)\right] Q_t^b b_t = \nu_t n_t \tag{40}
$$

Combining equations (39) to (40), the expected excess return of holding the tree μ_t can be expressed as

$$
\mu_t = \frac{\nu_{s,t}}{Q_t} - \nu_t \tag{41}
$$

I assume that financing constraints are always binding for both types of banks for analytical tractability. For the case where the constraint is occasionally binding, Ω_t cannot be analytically characterized (see Li and Xu (2022)) and numerically it can be solved using the computation procedure described in the next section. Note that if the financial constraint is not binding, I have $\chi_t = 0$, which implies that $\mu_t = 0$ according to equation (41) and (37).

Finally plug in equation (41) into equations (38) to (40), I obtain (24) to (25) reported in the main text.

Next, the value function $V_t^L = \max_{\left\{s_t^i, d_t^i, b_t^i\right\}} E_t \pi_{t+1} \left\{(1-\sigma) n_t + \sigma V_{t+1}\right\}$ can be written as

$$V(\psi_t, \phi_t) = \max_{\left\{s_t^i, d_t^i, b_t^i\right\}} E_t \pi_{t+1} \left\{\Omega(\psi_{t+1}, \phi_{t+1}) n_{t+1}\right\}$$

where

$$\Omega(\psi_t, \phi_t) = 1 - \sigma + \sigma \left(\frac{V(\psi_t, \phi_t)}{n_t}\right) \tag{42}$$

Using the balance sheet identity (19) and constraint (21), I have that when the financial constraint is binding,

$$\frac{V(\psi_t, \phi_t)}{n_t} = \kappa \left(\frac{\left(Q_t s_t + Q_t^b b_t\right)}{n_t}\right) = \kappa \widetilde{\phi}_t$$

where $\widetilde{\phi}$ is the leverage of individual intermediary, which can be expressed as the ratio of the total asset over the equity

$$\widetilde{\phi}_t \equiv \frac{s_t Q_t + Q_t^b b_t^i}{n_t}$$

This implies that Ω_t can be determined by

$$\Omega_t = 1 - \sigma + \sigma \kappa \widetilde{\phi}_t \tag{43}$$

To derive $\widetilde{\phi}_t$, I first combine equation (40) and (41),

$$[\kappa - \mu_t] Q_t s_t + [\kappa - \mu_t] Q_t^b b_t = \nu_t n_t$$

Thus the individual intermediary leverage can be expressed as

$$\widetilde{\phi}_t \equiv \frac{Q_t s_t + Q_t^b b_t}{n_t} = \frac{\nu_t}{\kappa - \mu_t} \tag{44}$$

Using equation (44), I can re-write equation (43)

$$\Omega_t = 1 - \sigma + \sigma \left(\nu_t + \mu_t \widetilde{\phi}_t\right) \tag{45}$$

Note that equation (44) holds for all intermediaries (low and high types), thus I have

$$\widetilde{\phi}_t^H = \frac{Q_t s_t + Q_t^b b_t}{n_t} = \frac{\nu_t}{\kappa^H - \mu_t} \tag{46}$$

$$\phi_t^L = \frac{Q_t s_t + Q_t^b b_t}{n_t} = \frac{\nu_t}{\kappa^L - \mu_t} \tag{47}$$

note that the higher is κ (less constrained), the higher is the leverage.

Combining equations (46) to (47), I have

$$\widetilde{\phi}_t^H = \frac{\kappa^L - \mu_t}{\kappa^H - \mu_t} \widetilde{\phi}_t^L \tag{48}$$

It's worth noting that the aggregate leverage

$$\phi_t = \frac{Q_t}{n_t^H + n_t^L} = \psi_t \widetilde{\phi}_t^L + (1 - \psi_t) \phi_t^H$$

where the second equality combines the market clearing condition (35) and definition of net worth share (23)

Thus using equation (48), the aggregate leverage can be expressed as

$$\phi_t = \psi_t \widetilde{\phi}_t^L + (1 - \psi_t) \frac{\kappa^L - \mu_t}{\kappa^H - \mu_t} \widetilde{\phi}_t^L$$

This implies that

$$\widetilde{\phi}_t^L = \phi_t \left(\psi_t + (1 - \psi_t) \frac{\kappa^L - \mu_t}{\kappa^H - \mu_t} \right)^{-1}$$

In the special case where $\kappa^L = \kappa^H$, the aggregate leverage is equal to the low type leverage

$$\phi_t = \widetilde{\phi}_t^L.$$

It's also worth noting that since $\kappa^L > \kappa^H$, this implies that $\widetilde{\phi}_t^L$ is increasing in ψ_t.

As a result, equation (45) becomes

$$\Omega_t = 1 - \sigma + \sigma \left(\nu_t + \phi_t \left(\psi_t + (1 - \psi_t) \frac{\kappa^L - \mu_t}{\kappa^H - \mu_t} \right)^{-1} \mu_t \right) \tag{49}$$

as expressed in the main text. Furthermore, equation (49) states that Ω_t only depends on (ϕ_t, ψ_t) at current t (the price ν_t and μ_t are functions of $\Omega_{t+1}(\phi_{t+1}, \psi_{t+1})$ by definition) and therefore proposition 1 is proved as stated.

In addition, it's easy to show that the price of the tree can be recursively defined as

$$Q_t = \frac{E_t \left[\pi_{t+1} \Omega_{t+1} \left(Q_{t+1} + Y_{t+1} \right) \right]}{\Omega_t}. \tag{50}$$

It shows the price of the tree Q_t only depends on (ϕ_t, ψ_t) through $\Omega(\phi_t, \psi_t)$ in the denominator. The value function V_t only depends on (ϕ_t, ψ_t) through $\Omega(\phi_t, \psi_t)$ according to equation (42). This confirms the guess that the state variables in the economy are (ϕ_t, ψ_t).

The problem for high type intermediary is analogous but they fund their investment using both deposits from households and interbank bonds. Their problem can be recursively expressed as follows,

$$\begin{gathered} V_t^H = \max_{\left\{ s_t^i, d_t^i, b_t^i \right\}} E_t \pi_{t+1} \left\{ (1 - \sigma) n_t^H + \sigma V_{t+1}^H \right\} \\ s.t \; s_t^H Q_t = n_t^H + d_t^H + b_t^H \\ n_t^H = s_{t-1}^H (Q_t + Y_t) + Q_t^b b_{t-1}^H - R_t d_{t-1}^H \\ V_t^H \geq \kappa^H s_t^H Q_t \end{gathered}$$

The first-order conditions are the same as (38) to (40) and the pricing wedge Ω_t is also the same as stated in equation (49), depending on both aggregate leverage ϕ_t and net worth share ψ_t.

Comment on Model Assumptions I assume that the heterogeneity among intermediaries is captured by the difference in tightness of financing constraints or κ^i. This assumption can be rationalized as follows. First, intermediaries with higher (lower) κ^i have less (more) incentive to invest in risky assets and hold less (more) volatile assets. This is consistent with the following finding in the data: From Table A4, broker-dealers (BD) hold more volatile assets than banking holding companies (BHC), who match the low-type intermediaries in the sample. This result is robust to the different subsamples investigated except the post-2007 sample; that is, BHCs were found to hold relatively more volatile assets during and after the recent financial crisis. This finding is consistent with the prediction of the HI-SDF framework that wealth is transferred to low-type intermediaries during financial distress.

Volatility of Total Financial Assets

	Full Sample	Pre-2000	Pre-2007	Post-2007
BHC	1.4071	0.2110	0.4803	0.8217
BD	1.4952	0.5401	1.0795	0.5218

Table A4: Time series standard deviation of total assets for each sector are reported. All volatilities are scaled by 10^{-6}. BHC total assets and BD assets are from Flow of Funds. Full sample spans 1970Q1 to 2019Q4.

Second, the differences in κ^i could also correspond to various types of VaR management or differentiated implementation of regulatory requirements among intermediaries. According to Coimbra and Rey (2017) and the Basel Committee on Banking Supervision (2013), when global banks are given a diversified test portfolio, they produced a wide range of results in terms of modeled VaR, with responses ranging from 13 million to 33 million euros on capital that they put aside for regulatory requirement purpose. Third, under some settings (for example, see Etula (2013)), the tightness of constraints is directly related to the intermediary's effective risk aversion. In these cases, the differences in financial constraints κ^i could also be interpreted as the variation in risk appetites.

Note that in the literature, the heterogeneity among agents can also be modeled as differences in risk aversion (Chan and Kogan (2002), Drechsler, Savov, and Schnabl (2018)), survival rate (Nuguer (2018)), and subjective discount rate (Diamond and Dybvig (1983)). Each heterogeneity modeling strategy can lead to different functional forms of the equilibrium pricing wedge Ω_t. This motivates the model-free nonparametric estimation of Ω_t in the next section.[16]

B.3 Computation Procedure

For each type of financial constraint, the model is solved numerically. More specifically, I solve the model by recursively solving the numerical functions $\{\Omega(\phi,\psi), Q(\phi,\psi)\}$ on grids of $[\psi,\phi]$. The map

$$\mathcal{T}\left(\Omega\left(\phi_{t+1},\psi_{t+1}\right), Q\left(\phi_{t+1},\psi_{t+1}\right)\right) = \left(\Omega\left(\phi_t,\psi_t\right), Q\left(\phi_t,\psi_t\right)\right) \tag{51}$$

[16] I don't take a parametric stand on the functional form of the pricing wedge Ω, but I restrict the pricing wedge to be a function of net worth share and aggregate leverage as motivated in the theoretical framework. There are three main reasons for doing this. First, as in this paper, the net worth shares among different types of agents are common state variables used in heterogeneous agent models such as Brunnermeier and Sannikov (2014) and He and Krishnamurthy (2012). Second, net worth shares are available in the data for both types of intermediaries. Third, the empirical representative intermediary asset pricing models construct pricing factor based on the capital ratio (such as HKM and AEM). Therefore, HI-SDF can nest those models' asset pricing implications.

is recursively defined in collection of equations (49) and (50). I solve the model imposing the assumption that the constraints are always binding around the steady state.[17] For simplicity of the text, any variable with superscript x' is the value in the next period.

There are several numerical challenges to solve the model. First, models with financial constraints, as in this paper, feature an incomplete market and thus the equilibrium cannot be solved using a social planner's solution. Therefore, the equilibrium needs to be solved directly. In addition, the local approximation method that is widely used to solve these types of model converges rather slowly with two types of intermediaries and two state variables (instead of one state variable in the representative intermediary case) for present-day computers even with relatively coarse grids on $[\psi, \phi]$.

Therefore, I use a recursive method developed in Ai, Bansal, and Li (2012) and Li and Xu (2022) that is able to solve the model on a relatively large number of girds on state variable space $[\psi, \phi]$ faster. For example, the convergence of the recursive method on a 10×8 grids on $[\psi, \phi]$ took $5,789$ seconds whereas local approximation method applied on the same grids took $17,274$ seconds.

I solve the mapping $\mathcal{T}$ defined in (51) by iterating functions of $\{\Omega(\phi, \psi), Q(\phi, \psi)\}$ that satisfies equations (49) and (50). More specifically, for any given pair (ψ, ϕ) and i th iteration, I guess a functional form on $\left\{\Omega^{(i)}(\phi, \psi), Q^{(i)}(\phi, \psi)\right\}$ and solve the (ψ', ϕ') using law of motions specified above, and then compute $\left\{\Omega^{(i)}(\phi', \psi'), Q^{(i)}(\phi', \psi')\right\}$. The algorithm stops if

$$
\begin{gathered}
\sqrt{\mathcal{D}'\mathcal{D}} < \varepsilon \\
s.t\ \mathcal{D} = \max_{\{\phi,\psi\} \in \mathcal{G}} \left\{ \left(\Omega^{(i)}(\phi', \psi'), Q^{(i)}(\phi', \psi')\right) - \mathcal{T}^{-1}\left(\Omega(\phi, \psi), Q(\phi, \psi)\right) \right\}
\end{gathered}
$$

where $\mathcal{G}$ is the grid on (ϕ, ψ) and $\mathcal{D}$ is the vector of distance between guessed functional form (at ith iteration) $\left\{\Omega^{(i)}(\phi', \psi'), Q^{(i)}(\phi', \psi')\right\}$ and model-implied functional form $\mathcal{T}^{-1}\left(\Omega(\phi_t, \psi_t), Q(\phi_t, \psi_t)\right)$ at the grid that the distance is maximized. $\sqrt{\mathcal{D}'\mathcal{D}}$ hence captures the root of the mean square of $\mathcal{D}$. The smaller is $\sqrt{\mathcal{D}'\mathcal{D}}$, the closer between the guessed and model implied functions. In the practice, I set the tolerance $\varepsilon = 10^{-3}$ and the function is guessed with Chebyshev Polynomials using approximation toolkit in the CompEcon Toolbox of Miranda and Fackler (2004).

Appendix C – Additional Data Details

CONSUMPTION

Consumption is measured as either total personal consumption expenditure (PCE) or expenditure on nondurables and services, excluding shoes and clothing. The quarterly data are seasonally adjusted at annual rates, in billions of chain-weighted 2005 dollars. The components are chain-weighted together, and this series is scaled up so that the sample mean matches the sample mean of total personal consumption expenditures. The source is the U.S. Department of Commerce, Bureau of Economic Analysis.

BALANCE SHEET DATA

The Balance sheet data of both BHC and BD come from Flow of Funds. For BHCs, the data come from Table L.130. The net worth is computed as total asset minus total liability. The

[17] For occasionally binding constraints, I used the procedure suggested by Li and Xu (2022). I am able to achieve convergence of the mapping $\mathcal{T}$ on coarse grids. The resulting wedge is numerically similar to the always-binding case under the same parameter values but the difference is substantially larger when the parameter γ is smaller.

total asset includes following items: (i) total time and savings deposits (ii) security repurchase agreements (iii) long-term debt securities (iv) Treasury securities (v) agency- and GSE-backed securities (vi) corporate and foreign bonds (vii) other loans and advances (viii) life insurance reserves and (viiii) total miscellaneous assets. The total liability include: (i) security repurchase agreements (ii) debt securities including both commercial paper and corporate foreign bonds (iii) depository institution loans (iv) foreign direct investment in U.S.

For BDs, the data come from Table L.131. I construct the equity following Gertler, Kiyotaki, and Prestipino (2016). The total asset includes: (i) checkable deposits and currency (ii) security repurchase agreements (iii) debt securities including commercial paper, treasury securities, agency- and GSE-backed securities, municipal securities, corporate and foreign bonds (iv) other loans and advances (v) corporate equities (vi) U.S. direct investment abroad (vii) total miscellaneous assets. The total liability includes (i) security repurchase agreements (ii) corporate and foreign bonds (iii) loans including depository institution loans and household cash accounts (vi) trade payables (vii) taxes payable (viii) foreign direct investment in U.S.

TEST PORTFOLIOS

All returns of test equity portfolios used in the paper are obtained from professor French's online data library.[18] All original returns are monthly data and I compounded them into quarterly data. The return in quarter Q of year Y , is the compounded monthly return over the three months in the quarter, $m1$,...,$m3$:

$$1 + R_{Q,Y} = \left(1 + \frac{R^{m1}_{Q,Y}}{100}\right)\left(1 + \frac{R^{m2}_{Q,Y}}{100}\right)\left(1 + \frac{R^{m3}_{Q1,Y}}{100}\right)$$

As test portfolios, I use the excess return constructed by subtracting the quarterly 3-month Treasury bill rate from the above.

The nonequity assets are obtained from Asaf Manela's website[19]. The sample spans from 1970Q1 to 2019Q4.

FAMA FRENCH PRICING FACTORS

I obtain quarterly Fama French pricing factor HML, SMB, Rm, and risk free rates from professor French's online data library http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Benchmark_Factors_Quarterly.zip.

AEM AND HKM LEVERAGE FACTORS

The broker-dealer leverage factor for AEM $\phi_t\,(BD)$ is constructed as follows

$$Leverage^{BD}_t = \frac{\text{Total Financial Assets}^{BD}_t}{\text{Total Financial Assets}^{BD}_t - \text{ Total Liabilities}^{BD}_t}$$

The leverage factor is constructed as seasonally adjusted log changes

$$LevFac_t = \left[\Delta \log\left(Leverage^{BD}_t\right)\right]^{SA}.$$

To extend the sample to 2019Q4 I use the original data on the total financial asset and liability of brokers and dealers data from flow of funds, Table L.128 and seasonally adjust $\Delta \log\left(Leverage^{BD}_t\right)$ by computing an expanding window regression of $\Delta \log\left(Leverage^{BD}_t\right)$ on dummies for three of the four quarters in the year at each date using the data up to that date.

The leverage factor for HKM is obtained from Asaf Manela's website.

[18]Link: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[19]Link: http://apps.olin.wustl.edu/faculty/manela/data.html

Appendix D – Estimation Details

Detail on Seive Functions

I use cubic B-splines to approximate the unknown function Ω. The multivariate sieve functions $\{B_j : j = 1, ..., K_T\}$ are implemented as a tensor product cubic B-spline taking the form:

$$F(\psi, \phi) = \alpha_0 + \sum_{i=1}^{K_{1T}} \sum_{j=1}^{K_{2T}} a_{ij} B_m \left(\psi - i + \frac{m}{2} \right) B_m \left(\frac{\phi}{\Delta_2} + \zeta - j \right)$$

where B_m is a B-spline of degree m. The term $\frac{m}{2}$ recenter the function, insuring the function preserving nonnegativity. Δ_2 and ζ are set to guarantee B_m stays within the bounds $[1.27, 26.71]$ which is the range for which I observe variation in aggregate leverage. I use a cardinal B-spline given by

$$B_m(y) = \frac{1}{(m-1)!} \sum_{k=0}^{m} (-1) \left(\begin{array}{c} m \\ k \end{array} \right) [\max(0, y-k)]^{m-1}$$

I set the order of the spline $m = 3$ and sieve dimension $K_{1T} = K_{2T} = 5$.

Detail on GMM Estimation in Linear Factor Models

Denote the factors together as

$$\mathbf{f}_t = \left[\Delta c_t, \log \widehat{\Omega}_t \right]'$$

and let K generically denote the number of factors (two here). Denote the $K \times 1$ vector $\beta_i = \left[\widehat{\beta}_{i,c}, \widehat{\beta}_{i,\Omega} \right]'$. The moment conditions for the expected return-beta representations are

$$g_T(\boldsymbol{b}) = \left[\begin{array}{c} E_T \left(\underbrace{\mathbf{R}_t^e}_{N\times 1} - \underbrace{\mathbf{a}}_{N\times 1} - \underbrace{\beta}_{(N\times K)} \underbrace{\mathbf{f}_t}_{(K\times 1)} \right) \\ E_T \left((\mathbf{R}_t^e - \mathbf{a} - \beta \mathbf{f}_t) \otimes \mathbf{f}_t \right) \\ E_T \left(\underbrace{\mathbf{R}_t^e}_{N\times 1} - \lambda_0 - \underbrace{\beta}_{(N\times K)} \underbrace{\lambda}_{(K\times 1)} \right) \end{array} \right] = \left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{array} \right] \tag{52}$$

where $\mathbf{a} = [a_1 ... a_N]'$ and $\beta = [\beta_1 ... \beta_N]'$, with parameter vector $\boldsymbol{b}' = [\mathbf{a}, \beta, \lambda_0, \lambda]'$. To obtain OLS time-series estimates of $\mathbf{a}$ and β and OLS cross sectional estimates of λ_0 and λ, I choose parameters $\boldsymbol{b}$ to set the following linear combination of moments to zero

$$\boldsymbol{a}_T g_T(\boldsymbol{b}) = 0,$$

where

$$\boldsymbol{a}_T = \left[\begin{array}{cc} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & [\mathbf{1}_N, \beta]' \end{array} \right].$$

The point estimates from GMM are identical to those from Fama-Macbeth regressions. To see this, in order to do OLS cross sectional regression of $E\left(R_{i,t}\right)$ on β, recall that the first-order necessary condition for minimizing the sum of squared residual is

$$\begin{aligned}\widetilde{\beta}\left(E\left(R_{i,t}\right)-\widetilde{\beta}\left[\lambda_0,\lambda\right]\right) &= 0 \Longrightarrow \\ \left[\lambda_0,\lambda\right] &= \left(\widetilde{\beta}'\widetilde{\beta}\right)^{-1}\widetilde{\beta}E\left(R_{i,t}\right)\end{aligned}$$

where $\widetilde{\beta}=\left[\mathbf{1}_N,\beta\right]$ to account for the intercept. If I multiply the first moment conditions with the identity matrix and the last moment condition with $(K+1)\times N$ vector $\widetilde{\beta}'$, I will then have OLS time-series estimates of $\mathbf{a}$ and β and OLS cross sectional estimates of λ. To estimate the parameter vector $\mathbf{b}$, I set

$$\boldsymbol{a}_T g_T\left(\mathbf{b}\right)=0$$

where

$$\underbrace{\boldsymbol{a}_T}_{\#Params\times\#Moments}=\left[\begin{array}{cc}\underbrace{\mathbf{I}_{(K+1)N}}_{(K+1)N\times(K+1)N} & \underbrace{\mathbf{0}}_{(K+1)N\times N}\\ \underbrace{\mathbf{0}}_{(K+1)\times(K+1)N} & \underbrace{\left[\mathbf{1}_N,\beta\right]'}_{(K+1)\times N}\end{array}\right]$$

In order to use Hansen's formulas for standard errors, I compute the $\mathbf{d}$ matrix of derivatives

$$\begin{aligned}\underbrace{\mathbf{d}}_{(K+2)N\times[(K+1)N+K+1]} &= \frac{\partial g_T}{\partial\mathbf{b}'}\\ &= \left[\begin{array}{ccc}\underbrace{-\mathbf{I}_N}_{N\times N} & \underbrace{-\mathbf{I}_N\otimes E_T\left(f_1\right)\ \cdots\ -\mathbf{I}_N\otimes E_T\left(f_K\right)}_{N\times KN} & \underbrace{\mathbf{0}}_{N\times(K+1)}\\ \underbrace{\begin{array}{c}-\mathbf{I}_N\otimes E_T\left(f_1\right)\\ \vdots\\ -\mathbf{I}_N\otimes E_T\left(f_K\right)\end{array}}_{KN\times N} & \underbrace{\begin{array}{ccc}-\mathbf{I}_N\otimes E_T\left(f_1^2\right) & \cdots & -\mathbf{I}_N\otimes E_T\left(f_K f_1\right)\\ \vdots & \ddots & \vdots\\ -\mathbf{I}_N\otimes E_T\left(f_1 f_K\right) & \cdots & -\mathbf{I}_N\otimes E_T\left(f_K^2\right)\end{array}}_{KN\times KN} & \underbrace{\mathbf{0}}_{KN\times(K+1)}\\ \underbrace{\mathbf{0}}_{N\times N} & \underbrace{-\mathbf{I}_N\otimes\lambda_1'\ \cdots\ -\mathbf{I}_N\otimes\lambda_K'}_{N\times KN} & -\underbrace{\left[\mathbf{1}_N,\beta\right]}_{N\times(K+1)}\end{array}\right]\end{aligned}$$

I also need $\mathbf{S}$ matrix, the spectral density matrix at frequency zero of the moment conditions

$$\mathbf{S}=\sum_{j=-\infty}^{\infty}E\left(\left[\begin{array}{c}\mathbf{R}_t^e-\mathbf{a}-\beta\mathbf{f}_t\\ \left(\mathbf{R}_t^e-\mathbf{a}-\beta\mathbf{f}_t\right)\otimes\mathbf{f}_t\\ \mathbf{R}_t^e-\ \lambda_0-\beta\lambda\end{array}\right]\left[\begin{array}{c}\mathbf{R}_{t-j}^e-\mathbf{a}-\beta\mathbf{f}_{t-j}\\ \left(\mathbf{R}_{t-j}^e-\mathbf{a}-\beta\mathbf{f}_{t-j}\right)\otimes\mathbf{f}_{t-j}\\ \mathbf{R}_{t-j}^e-\ \lambda_0-\beta\lambda\end{array}\right]\right).$$

Denote

$$h_t\left(\mathbf{b}\right)=\left[\begin{array}{c}\mathbf{R}_t^e-\mathbf{a}-\beta\mathbf{f}_t\\ \left(\mathbf{R}_t^e-\mathbf{a}-\beta\mathbf{f}_t\right)\otimes\mathbf{f}_t\\ \mathbf{R}_t^e-\ \lambda_0-\beta\lambda\end{array}\right].$$

I employ a Newey west correction to the standard errors with lag L by using the estimate

$$\mathbf{S}_T=\sum_{j=-L}^{L}\left(\frac{L-|j|}{L}\right)\frac{1}{T}\sum_{t=1}^{T}h_t\left(\widehat{\mathbf{b}}\right)h_{t-j}\left(\widehat{\mathbf{b}}\right)'$$

To get standard errors for the factor risk price estimates, λ, I use Hansen's formula for the sampling distribution of the parameter estimates

$$\underbrace{Var\left(\widehat{\mathbf{b}}\right)}_{[(K+1)N+K+1]\times[(K+1)N+K+1]} = \frac{1}{T}\left(\boldsymbol{a}_T\mathbf{d}\right)^{-1}\boldsymbol{a}_T\mathbf{S}_T\boldsymbol{a}_T'\left(\boldsymbol{a}_T\mathbf{d}\right)\prime^{-1}.$$

Bootstrap Procedure

This section describes the block bootstrap procedure applied in the estimation. As noted in CFL, under standard regularity condition imposed in semiparametric models, the two-step estimator $\widehat{\nu}$ is $\sqrt{T}$ asymptotically normally distributed but the asymptotic variance-covariance matrix is very of complicated form. Therefore, I follow Chen, Linton, and van Keilegom (2003) and compute block bootstrap estimates of sieve parameters as well as finite dimensional parameters ν.

The block bootstrap consists of following steps to ensure that the bootstrap estimates of price risks λ that accounts for (a) SMD estimates of $\widehat{\Omega}_t$ (b) time-series estimation of β (c) cross-sectional correlation among test assets.

1. For each test asset j, I estimate the time-series regressions on historical data:

$$R^e_{j,t} = a_j + \beta_{j,c}\Delta c_t + \beta_{j,\Omega}\log\widehat{\Omega}_t\left(\psi_t,\phi_t\right) + u_{j,t}, \tag{53}$$

where $\widehat{\Omega}_t$ is the SMD estimates based on historical data (ψ_t, ϕ_t). I obtain the full-sample estimates of the parameters of a_j and $\beta_{j,\Omega}$, which I denote $\widehat{a}_j$ and $\widehat{\beta}_{j,\boldsymbol{\Omega}}$.

2. I estimate λ_0, λ_c and λ_Ω using sampled data from cross-sectional regressions

$$E_T\left(R^e_{j,t}\right) = \lambda_0 + \widehat{\beta}_{j,c}\lambda_c + \widehat{\beta}_{j,\Omega}\lambda_\Omega + \epsilon_j$$

where $R^e_{j,t}$ is the quarterly excess return. From this regression I obtain the cross sectional fitted errors $\left\{\widehat{\epsilon}_j\right\}_j$ and historical sample estimates $\widehat{\lambda}_0$, $\widehat{\lambda}_c$ and $\widehat{\lambda}_\Omega$.

3. I draw randomly with replacement from blocks of the fitted residuals $u_{j,t}$ from the above time-series regressions along with sample data of (c_t, ψ_t, ϕ_t):

$$\begin{bmatrix} \widehat{u}_{1,1} & \cdots & \widehat{u}_{N,1} & c_1 & \psi_1 & \phi_1 \\ \widehat{u}_{1,2} & \cdots & \widehat{u}_{N,2} & c_2 & \psi_2 & \phi_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \widehat{u}_{1,T} & \cdots & \widehat{u}_{N,T} & c_T & \psi_T & \phi_T \end{bmatrix} \tag{54}$$

The mth bootstrap sample $\left\{\left\{u^{(m)}_{j,t}\right\}_{j=1}^{N}, c^{(m)}_t, \phi^{(m)}_t, \psi^{(m)}_t\right\}$ is obtained by sampling blocks of the raw data randomly with replacement and laying them end-to-end in the order sampled until a new sample of observations of length equal to the historical dataset is obtained. To choose the block length, I follow the recommendation of Hall, Horowitz, and Jing (1995) who show that the asymptotically optimal block length for estimating a symmetrical distribution function is $l \propto T^{1/5}$; also see Horowitz (2003).

4. I conduct the two-step SMD estimation of SDF m_t based on sampled data $\left(c^{(m)}_t, \phi^{(m)}_t, \psi^{(m)}_t\right)$ and generate new data series for $\left\{\Delta c_t, \log\widehat{\Omega}_t\right\}_t^{(m)}$.

5. I generate new samples of observations on excess returns returns $\left\{R_{j,t}^{(m)}\right\}_t$ from new data on $\left\{u_{j,t}^{(m)}\right\}_t$ and $\left\{\Delta c_t, \log \widehat{\Omega}_t\right\}_t^{(m)}$ and the sample estimates $\widehat{a}_j$, $\beta_{j,c}$ and $\beta_{j,\Omega}$.

6. I generate mth observation $\beta_{j,c}^{(m)}$ and $\beta_{j,\Omega}^{(m)}$. from regression of $\left\{R_{j,t}^{(m)}\right\}_t$ on $\left\{\Delta c_t, \log \widehat{\Omega}_t\right\}_t^{(m)}$ and a constant.

7. I obtain an mth bootstrap sample $\left\{\epsilon_j^{(m)}\right\}_j$ by sampling the fitted errors $\{\widehat{\epsilon}_j\}_j$ randomly with replacement and laying them end-to-end in the order sampled until a new sample of observations of length N equal to the historical cross-sectional sample is obtained. I then generate new samples of observations on quarterly average excess returns $\left\{E\left(R_{j,t}^{e(m)}\right)\right\}_j$ from new data on $\left\{\epsilon_j^{(m)}\right\}_j$ and $\left\{\beta_{j,c}^{(m)}\right\}_j$ and $\left\{\beta_{j,\Omega}^{(m)}\right\}_j$ and the sample estimates $\widehat{\lambda}_0$, $\widehat{\lambda}_c$ and $\widehat{\lambda}_\Omega$.

8. I form the mth estimates $\lambda_0^{(m)}$, $\lambda_c^{(m)}$ and $\lambda_\Omega^{(m)}$ by regressing $\left\{E\left(R_{j,t}^{e(m)}\right)\right\}_j$ on the mth observation $\left\{\beta_{j,c}^{(m)}\right\}_j$, $\left\{\beta_{j,\Omega}^{(m)}\right\}$ and a constant. I store the mth sample cross-sectional $\overline{R}^2$, $\overline{R}^{(m)2}$ along with the mth values of $\lambda_0^{(m)}$, $\lambda_c^{(m)}$ and $\lambda_\Omega^{(m)}$.

9. I repeat steps 1-8 500 times, and report the 95% confidence intervals for $\left\{\nu^{(m)}, \overline{R}^{(m)2}, \lambda_c^{(m)}, \lambda_\Omega^{(m)}\right\}_m$.

It's worth noting that the computation procedure of this block bootstrap is highly computationally intensive and limit the number of bootstrap replication that can be feasibly implemented. In the practice, I choose 500 blocks of bootstrap samples. The results are robust if we only repeat step 1-7 $10,000$ times and use historical data on $\left(c_t, \widehat{\Omega}_t\right)$ (i.e, $\widehat{\Omega}_t$ is not SMD re-estimated at each bootstrap sample).