Internet Appendix for "Heterogeneous Intermediary Asset Pricing"

Mahyar Kargar

September 27, 2020

A Proof of propositions

Proof of proposition 1

Proof. State variable x is the wealth share of the financial sector: $x_t = \frac{W_A+W_B}{W}$ and state variable y is the wealth share of type-A agents in the financial sector: $y_t = \frac{W_A}{W_A+W_B}$, where W, W_A, and W_B are the aggregate wealth, wealth of A-, and B-type agents, respectively.

Dynamics of x_t: From Eq. (5), W_A has the following law of motion

$$\frac{dW_A}{W_A} = \left(r + w_s^A(\mu - r) - c_A\right)\, dt + w_s^A \sigma\, dZ.$$

The law of motion for W_B can be written similarly.

The law of motion for the numerator, $W_A + W_B$, will be

$$\begin{aligned}\frac{d\left(W_A + W_B\right)}{W_A + W_B} &= \left[r + \left(y w_s^A + (1-y) w_s^B\right)(\mu - r) - (y c_A + (1-y) c_B)\right] dt \\ &\quad + \left(y w_s^A + (1-y) w_s^B\right) \sigma dZ_t\end{aligned}$$

Define wealth share of agents of type A and B as $u \equiv W_A/W = xy$, and $v \equiv W_B/W = x(1-y)$, respectively. C types wealth share will then be $1 - u - v$. As the aggregate wealth is

$W = W_A + W_B + W_C$, the law of motion for the denominator is

$$\begin{aligned}\frac{dW}{W} &= \left[r + \underbrace{\left(xyw_s^A + x(1-y)w_s^B + (1-x)w_s^C\right)}_{=1 \text{ (by stock market-clearing)}}(\mu - r) - \underbrace{\left(xyc_A + x(1-y)c_B + (1-x)c_C\right)}_{=F \text{ (by goods market-clearing)}} \right] dt \\ &+ \left[\underbrace{xyw_s^A + x(1-y)w_s^B + (1-x)w_s^C}_{=1} \right] \sigma\, dZ_t \\ &= \left[r + (\mu - r) - F\right] dt + \sigma\, dZ_t\end{aligned}$$

From Ito's lemma for the ratio of two stochastic processes,

$$\begin{aligned}\frac{dx}{x} &= \kappa(\bar{x} - x)dt + \left[\left(yw_s^A + (1-y)w_s^B - 1\right)\left(\mu - r - \sigma^2\right) - yc_A - (1-y)c_B + F\right] dt \\ &+ \left(yw_s^A + (1-y)w_s^B - 1\right)\sigma\, dZ_t\end{aligned}$$

Thus from the dynamics of x in Eq. (12) we have

$$\begin{aligned}\mu_x &= \left(yw_s^A + (1-y)w_s^B - 1\right)\left(\mu - r - \sigma^2\right) - yc_A - (1-y)c_B + F \\ \sigma_x &= \left(yw_s^A + (1-y)w_s^B - 1\right)\sigma\end{aligned}$$

Dynamics of $\boldsymbol{y_t}$: The numerator of y is W^A, and its denominator is $\left(W^A + W^B\right)$ for which the law of motion is calculated above. So, from Ito's lemma for a ratio, we get

$$\begin{aligned}\frac{dy}{y} &= \kappa(\bar{y} - y)dt + (1-y)\left[\left(w_s^A - w_s^B\right)(\mu - r) - c_A + c_B - \left(yw_s^A + (1-y)w_s^B\right)\left(w_s^A - w_s^B\right)\sigma^2\right] dt \\ &+ (1-y)\left(w_s^A - w_s^B\right)\sigma\, dZ_t\end{aligned}$$

Thus, from the dynamics of y in Eq. (13), we have

$$\begin{aligned}\mu_y &= \left(w_s^A - w_s^B\right)(\mu - r) - c_A + c_B - \left(yw_s^A + (1-y)w_s^B\right)\left(w_s^A - w_s^B\right)\sigma^2 \\ \sigma_y &= \left(w_s^A - w_s^B\right)\sigma\end{aligned}$$

$\square$

Proof of proposition 2

Proof. Give than σ_x and σ_y are finite, we trivially get

$$\lim_{x\to 0} x\,\sigma_x = 0,\ \forall y \quad \text{and} \quad \lim_{y\to 0} y(1-y)\,\sigma_y = \lim_{y\to 1} y(1-y)\,\sigma_y = 0,\ \forall x.$$

We need only to show $\lim_{x\to 1} x\,\sigma_x = 0\ \forall y$. The market clearing condition for the risky asset when $x \to 1$ becomes $yw_s^A + (1-y)w_s^B = 1$.

So, from the expression for σ_x in Eq. (16), we have:

$$x\,\sigma_x = x\left[yw_s^A + (1-y)w_s^B - 1\right]\sigma$$

which goes to zero as $x \to 1$ for all y from the stock market-clearing. □

Proof of proposition 3

Proof. We can write agent i's optimization problem in Eq. (8) as

$$0 = \max_{c_i, w_s^i} \left\{ f_i\left(c_{i,t}, V_{i,t}\right) dt + \mathbb{E}_t\left[dV_{i,t}\right] \right\}$$

Using Ito's lemma we have

$$\mathbb{E}_t[dV_i] = V_{i,W_i}\mathbb{E}_t[dW_i] + \frac{1}{2}V_{i,W_iW_i}\mathbb{E}_t\left[dW_i^2\right] + V_{i,J_i}\mathbb{E}_t[dJ_i] + \frac{1}{2}V_{i,J_iJ_i}\mathbb{E}_t\left[dJ_i^2\right] + V_{i,W_iJ_i}\mathbb{E}_t[dW_i dJ_i]$$

where V_{i,W_i} and V_{i,W_iW_i} are the first and second partial derivatives of V_i with respect to W_i (similarly for V_{i,J_i}, V_{i,J_iJ_i}, and V_{i,W_iJ_i}.) Also posit the following Ito process for marginal value of wealth J_i:

$$\frac{dJ_i}{J_i} = \mu_{J_{i,t}} dt + \sigma_{J_{i,t}} dZ,$$

with adapted processes $\mu_{J_{i,t}} = \mu_{J_i}(x_t, y_t)$ and $\sigma_{J_{i,t}} = \sigma_{J_i}(x_t, y_t)$. I will drop t subscripts for notational simplicity.

Using Ito's lemma, we can find the drift and diffusions μ_{J_i} and σ_{J_i} as follows:

$$\begin{aligned}
\mu_{J_i} &= \frac{J_{i,x}}{J_i}\left[\kappa(\bar{x}-x)+x\,\mu_x\right]+\frac{J_{i,y}}{J_i}\left[\kappa(\bar{y}-y)+y(1-y)\mu_y\right] \\
&\quad +\frac{1}{2}\frac{J_{i,xx}}{J_i}x^2\sigma_x^2+\frac{J_{i,xy}}{J_i}xy(1-y)\sigma_x\sigma_y+\frac{1}{2}\frac{J_{i,yy}}{J_i}y^2(1-y)^2\sigma_y^2 && \text{(A.1)} \\
\sigma_{J_i} &= \frac{J_{i,x}}{J_i}x\,\sigma_x+\frac{J_{i,y}}{J_i}y(1-y)\sigma_y && \text{(A.2)}
\end{aligned}$$

Plugging in the felicity function $f(C,U)$ in Eq. (3) and the conjecture for value function V_i in (21) into the HJB equation above, using the budget constraint in (5) and the law of motion for J_i in (A.1) and (A.2), and dropping the $\left(W_i^{1-\gamma_i}J_i^{\frac{1-\gamma_i}{1-\psi_i}}\right)$ and dt terms yields

$$\begin{aligned}
0 = \max_{c_i,w_s^i}\ & \frac{1}{1-1/\psi_i}\left[\left(\frac{c_i}{J_i^{1/(1-\psi_i)}}\right)^{1-1/\psi_i}-(\rho+\kappa)\right]+\left[r-c_i+\kappa+w_s^i(\mu-r)-\frac{\gamma_i}{2}\left(w_s^i\right)^2\sigma^2\right] \\
&+\left(\frac{1}{1-\psi_i}\right)\left\{\frac{J_{i,x}}{J_i}\left[\kappa(\bar{x}-x)+x\mu_x\right]+\frac{J_{i,y}}{J_i}\left[\kappa(\bar{y}-y)+y(1-y)\mu_y\right]\right. \\
&+(1-\gamma_i)\left.\left(\frac{J_{i,x}}{J_i}x\sigma_x+\frac{J_{i,y}}{J_i}y(1-y)\sigma_y\right)w_s^i\sigma\right\}+\frac{1}{2}\left(\frac{1}{1-\psi_i}\right)\left[\left(\frac{\psi_i-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i}x\sigma_x+\frac{J_{i,y}}{J_i}y(1-y)\mu_y\right)^2\right. \\
&\left.+\frac{J_{i,xx}}{J_i}x^2\sigma_x^2+2\frac{J_{i,xy}}{J_i}xy(1-y)\sigma_x\sigma_y+\frac{J_{i,yy}}{J_i}y^2(1-y)^2\sigma_y^2\right]+\lambda_i\left(\bar{\theta}_t-w_s^i\right),
\end{aligned}$$

where λ_i is proportional to the Lagrange multiplier on the time-varying margin constraint. The first-order condition for the consumption-wealth ratio and portfolio share will lead to Eq. (23) and (25):

$$\begin{aligned}
c_i &= J_i && \text{(A.3)} \\
w_s^i &= \frac{1}{\gamma_i}\left[\frac{\mu-r}{\sigma^2}+\left(\frac{1-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i}x\,\frac{\sigma_x}{\sigma}+\frac{J_{i,y}}{J_i}y(1-y)\frac{\sigma_y}{\sigma}\right)\right]-\frac{1}{\gamma_i\sigma^2}\lambda_i && \text{(A.4)}
\end{aligned}$$

When the margin constraint for agent i is slack, $\lambda_i=0$ and we have

$$w_s^{i,*}=\frac{\mu-r}{\gamma_i\sigma^2}+\frac{1}{\gamma_i}\left(\frac{1-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i}x\,\frac{\sigma_x}{\sigma}+\frac{J_{i,y}}{J_i}y(1-y)\frac{\sigma_y}{\sigma}\right)$$

When the margin constraint for agent i is binding, λ_i is strictly positive and $w_s^{i,const}=\bar{\theta}_t$.

Plugging in the $w_s^{i,const}$ into (A.4), we derive the expression for the multiplier on the time-varying margin constraint:

$$\lambda_i = (\mu - r) + \left(\frac{1-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i}x\sigma_x + \frac{J_{i,y}}{J_i}y(1-y)\sigma_y\right)\sigma - \gamma_i\sigma^2\bar{\theta}_t \tag{A.5}$$

$\square$

B Numerical procedure

The computation of equilibrium is reduced to solving three second-order PDEs for functions J_i for $i \in \{A, B, C\}$.[1] I use the Chebyshev orthogonal collocation method to solve the model.[2] The HJB equation for agent i can be written as the following functional equation:

$$\mathcal{H}_i\left(J_i\right) = \mathbf{0}.$$

I express marginal value of wealth functions $J_A(x,y)$, $J_B(x,y)$ and $J_C(x,y)$ as bivariate Chebyshev polynomials of order N (I use $N = 20$), that is, I approximate J_i with the tensor product of Chebyshev polynomials of order N:

$$\widehat{J}_i(x,y) = \sum_{j=0}^{N}\sum_{k=0}^{N} a^i_{jk}\psi_j(\omega_j(x))\psi_k(\omega_k(y)), \; i \in \{A, B, C\}. \tag{B.1}$$

where ψ_j is the Chebyshev polynomial of degree $j = 0, 1, \ldots, N$, called the *basis function*, $\Psi_{jk}(x,y) = \psi_j(x)\psi_k(y)$ is a tensor product basis, $\{a^i_{jk}\}^N_{j,k=1}$ are unknown coefficients for agent i, and ω_j's are the Chebyshev nodes (collocation points) defined below.

I then plug $\widehat{J}_i$ into the HJB equation for agent i to form the *residual equation*:

$$\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right) = \mathcal{H}_i\left(\widehat{J}_i\right),$$

and find the vector of coefficients $\boldsymbol{a}^i$ that makes the residual equation as close to $\mathbf{0}$ as possible given

[1] Duffie and Lions (1992) show existence and uniqueness of infinite-horizon stochastic differential utility by partial differential equation techniques in a Markov diffusion setting.

[2] For more details, see Judd (1992, 1998) and Computational Tools & Macroeconomic Applications, NBER Summer Institute 2011 Methods Lectures, Lawrence Christiano and Jesus Fernandez-Villaverde, Organizers.

some objective function $\rho\left(\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right), \mathbf{0}\right)$:

$$\boldsymbol{a}^i = \arg\min_{\boldsymbol{a}^i} \; \rho\left(\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right), \mathbf{0}\right)$$

The most common objective function is a *weighted residual* given some weight functions $\phi_j : \Omega \to \mathbb{R}^m$:

$$\rho\left(\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right), \mathbf{0}\right) = \begin{cases} 0 & \text{if} \iint\limits_{\Omega\times\Omega} \phi_j(x)\phi_k(y)\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right) dx\, dy = \mathbf{0}, \text{ for } j, k = 1, \ldots, N \\ \\ 1 & \text{otherwise} \end{cases}$$

In the pseudo-spectral (or collocation) method, the weight functions are chosen as $\phi_j(x) = \delta(x - x_i)$, where δ is the dirac delta function and x_is are the collocation points. In the *orthogonal collocation* method, which I use to solve the model, the basis functions are a set of orthogonal Chebyshev polynomials and collocation points are given by the roots of the N^{th} polynomial.

Chebyshev polynomials of degree n can be easily defined recursively:

$$\begin{aligned} \psi_0(\omega) &= 1 \\ \psi_1(\omega) &= x \\ \psi_{n+1}(\omega) &= 2\omega\psi_n(\omega) - \psi_{n-1}(\omega) \end{aligned} \tag{B.2}$$

As mentioned above, the collocation points are the N zeros of the Chebyshev polynomial of order N, $(\psi_N(\omega_j) = 0)$, and are given by the following expression:

$$\omega_j = \cos\left(\frac{2j-1}{2n}\pi\right), \; j = 1, \ldots, N.$$

These roots are clustered quadratically towards ± 1. Chebyshev polynomials are defined on $\omega_i \in [-1, 1]$. As the state variables $x, y \in [0, 1]$ in my model, I use the linear transformation $x_j = (1 + \omega_j)/2$.[3]

[3]For a general state space $x \in [x_L, x_H]$, one can use a linear transformation $x_j = x_L + 0.5(x_H - x_L)(1 + \omega_j)$.

I calculate the derivatives of these functions as well as the state variable dynamics, agents' portfolio choice, risky asset return and volatility using the relevant equilibrium expressions. I then plug these quantities into the HJB equations (22) and project the resulting residuals onto the complete set of Chebyshev polynomials up to order N. I use the built-in MATLAB function `fsolve` to find the coefficients of J_i polynomials that make the projected residuals equal to zero. This results in a highly accurate solution for coefficients in the $\widehat{J}_i$ functions with errors on the order of 10^{-20}.

The numerical algorithm is summarized below.

1. From goods market-clearing conditions and differentiating it with respect to the state variable, we get expressions for dividend yield F and its derivatives with respect to x and y.

$$
\begin{aligned}
&F = xyJ_A + x(1-y)J_B + (1-x)J_C, \\
&F_x = yJ_A + (1-y)J_B - J_C + xyJ_{A,x} + x(1-y)J_{B,x} + (1-x)J_{C,x}, \\
&F_y = xJ_A - xJ_B + xyJ_{A,y} + x(1-y)J_{B,y} + (1-x)J_{C,y}, \\
&F_{xx} = 2yJ_{A,x} + 2(1-y)J_{B,x} - 2J_{C,x} + xyJ_{A,xx} + x(1-y)J_{B,xx} + (1-x)J_{C,xx}, \\
&F_{yy} = 2xJ_{A,y} - 2xJ_{B,y} + xyJ_{A,yy} + x(1-y)J_{B,yy} + (1-x)J_{C,yy}, \\
&F_{xy} = J_A - J_B + xJ_{A,x} - xJ_{B,x} + yJ_{A,y} + (1-y)J_{B,y} - J_{C,y} + xyJ_{A,xy} \\
&\quad + x(1-y)J_{B,xy} + (1-x)J_{C,xy},
\end{aligned}
$$

   where $J_{i,x}$ and $J_{i,xx}$ are the first and second partial derivative of J_i with respect to x, respectively, and similarly for $J_{i,y}$, $J_{i,yy}$ and $J_{i,xy}$.

2. Using the market-clearing condition for the endowment claim, plugging in the expression for agent C's optimal portfolio choice w_s^C from (24), and substituting for $(\mu - r)/\sigma^2$ from the expression for $w_s^{A,*}$, we will derive the first of the two equations that $w_s^{A,*}$ and w_s^B have to

satisfy:

$$
\begin{aligned}
1 &= xyw_s^{A,*} + x(1-y)w_s^B + (1-x)w_s^C \\
&= xyw_s^{A,*} + x(1-y)w_s^B \\
&+ (1-x)\frac{1}{\gamma_C}\left\{\frac{\mu - r}{\sigma^2} + \left(\frac{1-\gamma_C}{1-\psi_C}\right)\left[\frac{J_{C,x}}{J_C}x\left(yw_s^{A,*} + (1-y)w_s^B - 1\right) + \frac{J_{C,y}}{J_C}y(1-y)\left(w_s^{A,*} - w_s^B\right)\right]\right\} \\
&= xw_s^{A,*} + yw_s^B + (1-x)\frac{1}{\gamma_C}\left\{\gamma_A w_s^{A,*} - \left(\frac{1-\gamma_A}{1-\psi_A}\right)\left[\frac{J_{A,x}}{J_A}x\left(yw_s^{A,*} + (1-y)w_s^B - 1\right)\right.\right. \\
&\left. + \frac{J_{A,y}}{J_A}y(1-y)\left(w_s^{A,*} - w_s^B\right)\right] \\
&\left. + \left(\frac{1-\gamma_C}{1-\psi_C}\right)\left[\frac{J_{C,x}}{J_C}x\left(yw_s^{A,*} + (1-y)w_s^B - 1\right) + \frac{J_{C,y}}{J_C}y(1-y)\left(w_s^{A,*} - w_s^B\right)\right]\right\}
\end{aligned}
$$

To derive the second equation, I plug in the expression for $(\mu - r)/\sigma^2$ from the A-type agent's optimal portfolio $w_s^{A,*}$ in the expression for w_s^B:

$$
\begin{aligned}
w_s^B &= \frac{1}{\gamma_B}\left[\frac{\mu - r}{\sigma^2} + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\left(\frac{J_{B,x}}{J_B}x\left(yw_s^A + (1-y)w_s^B - 1\right) + \frac{J_{B,y}}{J_B}y(1-y)(w_s^A - w_s^B)\right)\right] \\
&= \frac{1}{\gamma_B}\left[\gamma_A w_s^A - \left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}x\left(yw_s^A + (1-y)w_s^B - 1\right) + \frac{J_{A,y}}{J_A}y(1-y)(w_s^A - w_s^B)\right)\right. \\
&\left. + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\left(\frac{J_{B,x}}{J_B}x\left(yw_s^A + (1-y)w_s^B - 1\right) + \frac{J_{B,y}}{J_B}y(1-y)(w_s^A - w_s^B)\right)\right]
\end{aligned}
$$

We can rewrite the systems of equation as

$$
\begin{aligned}
a_{11}w_s^{A,*} + a_{12}w_s^B &= b_1 \\
a_{21}w_s^{A,*} + a_{22}w_s^B &= b_2
\end{aligned}
$$

where

$$
\begin{aligned}
a_{11} &= xy + (1-x)\frac{1}{\gamma_C}\left[\gamma_A - \left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}xy + \frac{J_{A,y}}{J_A}y(1-y)\right)\right.\\
&\quad\left. + \left(\frac{1-\gamma_C}{1-\psi_C}\right)\left(\frac{J_{C,x}}{J_C}xy + \frac{J_{C,y}}{J_C}y(1-y)\right)\right],\\
a_{12} &= x(1-y) + (1-x)\frac{1}{\gamma_C}\left[-\left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}x(1-y) - \frac{J_{A,y}}{J_A}y(1-y)\right)\right.\\
&\quad\left. + \left(\frac{1-\gamma_C}{1-\psi_C}\right)\left(\frac{J_{C,x}}{J_x}x(1-y) - \frac{J_{C,y}}{J_C}y(1-y)\right)\right],\\
a_{21} &= \frac{1}{\gamma_B}\left[\gamma_A - \left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}xy + \frac{J_{A,y}}{J_A}y(1-y)\right) + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\left(\frac{J_{B,x}}{J_B}xy + \frac{J_{B,y}}{J_B}y(1-y)\right)\right],\\
a_{22} &= -1 + \frac{1}{\gamma_B}\left[-\left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}x(1-y) + \frac{J_{A,y}}{J_A}y(1-y)\right)\right.\\
&\quad\left. + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\left(\frac{J_{B,x}}{J_B}x(1-y) - \frac{J_{B,y}}{J_B}y(1-y)\right)\right],\\
b_1 &= 1 + (1-x)\frac{1}{\gamma_C}\left[-\left(\frac{1-\gamma_A}{1-\psi_A}\right)\frac{J_{A,x}}{J_A}x + \left(\frac{1-\gamma_C}{1-\psi_C}\right)\frac{J_{C,x}}{J_C}x\right],\\
b_2 &= \frac{1}{\gamma_B}\left[-\left(\frac{1-\gamma_A}{1-\psi_A}\right)\frac{J_{A,x}}{J_A}x + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\frac{J_{B,x}}{J_B}x\right].
\end{aligned}
$$

The system of equations above can be solved easily to find $w_s^{A,*}$ and w_s^B.

3. Because the return volatility can be written as

$$
\sigma = \frac{\sigma_D}{1 + \frac{F_x}{F}x\left[yw_s^A + (1-y)w_s^B - 1\right] + \frac{F_y}{F}y(1-y)\left(w_s^A - w_s^B\right)}, \tag{B.3}
$$

when the margin constrains for agent A bind, from Eq. (7) we must have

$$
w_s^{A,const} = \frac{1 - \frac{F_x}{F}x + \left(\frac{F_x}{F}x - \frac{F_y}{F}y\right)(1-y)w_s^B}{\alpha\sigma_D - \left[\frac{F_x}{F}x + \frac{F_y}{F}(1-y)\right]y} \tag{B.4}
$$

So, we have $w_s^A \leq w_s^{A,const}$. Then from Eq. (25) we can find the A-type's and B-type's portfolio weights in the risky asset

$$
w_s^A = \min\left(w_s^{A,*}, w_s^{A,const}\right),
$$

where $w_s^{A,const}$ is given in Eq. (B.4).

4. From stock market clearing, we can derive the C-type's optimal portfolio weight

$$w_s^C = \frac{1 - xy\, w_s^A - x(1-y)\, w_s^B}{1-x}$$

5. Using the expression for the return volatility in Eq. (19) and plugging in expressions for σ_x and σ_y from Eq. (16) and (17), the expression for return volatility is

$$\sigma = \frac{\sigma_D}{1 + \frac{F_x}{F} x\left[y w_s^A + (1-y) w_s^B - 1\right] + \frac{F_y}{F} y(1-y)\left(w_s^A - w_s^B\right)}.$$

6. Using the expression for σ above, state variable diffusions (σ_x and σ_y) can be found from Eq. (16) and (17):

$$\sigma_x = \left[y w_s^A + (1-y) w_s^B - 1\right]\sigma, \quad \text{and} \quad \sigma_y = \left(w_s^A - w_s^B\right)\sigma.$$

7. From the expression for w_s^C, σ, σ_x, and σ_y, the expected excess return (risk premium) on the risky asset is

$$\mu - r = \gamma_C w_s^C \sigma^2 - \left(\frac{1-\gamma_C}{1-\psi_C}\right)\left(\frac{J_{C,x}}{J_C} x\sigma_x + \frac{J_{C,y}}{J_C} y(1-y)\sigma_y\right)\sigma.$$

8. Using the optimal consumption-wealth ratios $c_i = J_i$, we can then compute drifts of the state variables μ_x and μ_y as

$$\begin{aligned}
\mu_x &= \left[y w_s^A + (1-y) w_s^B - 1\right](\mu - r - \sigma^2) - y J_A - (1-y) J_B + F, \\
\mu_y &= \left(w_s^A - w_s^B\right)(\mu - r) - J_A + J_B - \left[y w_s^A + (1-y) w_s^B\right]\left(w_s^A - w_s^B\right)\sigma^2.
\end{aligned}$$

9. From Eq. (18) the expected return on the risky asset can be calculated

$$\begin{aligned}
\mu = \mu_D + F - \frac{F_x}{F}\left[\kappa(\bar{x} - x) + x(\mu_x + \sigma_D \sigma_x)\right] - \frac{F_y}{F}\left[\kappa(\bar{y} - y) + y(1-y)(\mu_y + \sigma_D \sigma_y)\right] \\
+ \left[\left(\frac{F_x}{F}\right)^2 - \frac{1}{2}\frac{F_{xx}}{F}\right] x^2 \sigma_x^2 + \left[\left(\frac{F_y}{F}\right)^2 - \frac{1}{2}\frac{F_{yy}}{F}\right] y^2 (1-y)^2 \sigma_y^2 + \left[2\left(\frac{F_x}{F}\right)\left(\frac{F_y}{F}\right) - \frac{F_{xy}}{F}\right] xy(1-y)\sigma_x \sigma_y.
\end{aligned}$$

10. The real interest rate is

$$r = \mu - (\mu - r).$$

11. Plugging the above expressions into agent i's HJB equations in (22), we derive the residual functions for agent i:

$$\begin{aligned}
0 = & -(\rho+\kappa) + \frac{1}{\psi_i} J_i + \left(1 - \frac{1}{\psi_i}\right) \left[r + w_s^i(\mu - r) - \frac{\gamma_i}{2} \left(w_s^i\right)^2 \sigma^2 \right] \\
& - \frac{1}{\psi_i} \left\{ \frac{J_{i,x}}{J_i} \left[\kappa(\bar{x} - x) + x\mu_x\right] + \frac{J_{i,y}}{J_i} \left[\kappa(\bar{y} - y) + y(1-y)\mu_y\right] + (1-\gamma_i) \left(\frac{J_{i,x}}{J_i} x\sigma_x + \frac{J_{i,y}}{J_i} y(1-y)\sigma_y \right) w_s^i \sigma \right\} \\
& - \frac{1}{2\psi_i} \left[\left(\frac{\psi_i - \gamma_i}{1 - \psi_i} \right) \left(\frac{J_{i,x}}{J_i} x\sigma_x + \frac{J_{i,y}}{J_i} y(1-y)\mu_y \right)^2 + \frac{J_{i,xx}}{J_i} x^2 \sigma_x^2 + 2\frac{J_{i,xy}}{J_i} xy(1-y)\sigma_x\sigma_y + \frac{J_{i,yy}}{J_i} y^2(1-y)^2\sigma_y^2 \right].
\end{aligned}$$

C Alternative modeling of intermediary heterogeneity

In my model, I assume intermediaries are heterogeneous in their risk-bearing capacity. Both A and B agents face a leverage constraint that, in equilibrium, only (occasionally) binds for less risk-averse A-type intermediary (a proxy for broker-dealers). This modeling choice seems to be in line with the empirical evidence that leverage constraints for large banks did not bind even during the height of financial crisis.

If, instead, I assume B-types intermediaries face a *tighter* financial constraint instead of modeling them as more risk-averse, my model's predictions for asset reallocation and opposite leverage dynamics within the financial sector remain unchanged. As I discuss below, my model's *qualitative* results with intermediaries heterogeneous in risk aversions facing the same financial constraints are similar to a model with *risk-neutral* intermediaries facing heterogeneous financial constraints: the intermediary with a *lower* risk-aversion (A-type in my model) or *looser* constraint deleverages in bad times while the more risk-averse intermediary (B-type in my model) or the one with a tighter constraint levers up to clear the risky asset market.

Consider a model that features two types of intermediaries who differ in the tightness of their funding constraints: a *low-type* that faces tighter financing constraints and a *high-type* that faces looser constraints. In equilibrium, like my model, the high-type has higher leverage in normal time. Similar to my model, following a negative shock, high-type (low-type) intermediaries decrease (increase) their leverage, i.e., high-type (low-type) intermediaries exhibit procyclical (countercyclical) leverage.[4] High- and low-type intermediaries in this alternative setting play the same role as less risk-averse A-type and more risk-averse B-type intermediaries in my model, respectively. Therefore, the less constrained (or less risk-averse) intermediaries can act as proxies for broker-dealers, and the ones facing tighter constraints (or with higher risk aversion) can correspond to banks.

Both methods produce higher risk premiums in high marginal utility states where the less risk-averse or less-constrained intermediary has a lower share of the financial sector wealth. Consistent with the empirical evidence, both approaches also lead to higher leverage for one intermediary type (the less risk-averse in my model or the less-constrained in the alternative model) and countercyclical leverage for the other type (the more risk-averse in my model or the one with a tighter constraint in the alternative model).

The economic mechanism of this alternative modeling approach, however, is slightly different from my model. Following a series of negative shocks in my model, both intermediaries suffer losses, and their leverage will rise. However, since the less risk-averse intermediary has higher initial leverage, it hits the borrowing constraint and has to sell assets. To clear the market, the more risk-averse intermediary need to hold a larger supply of the risky asset. To be enticed to do so, the risk premium has to increase. In the alternative approach, the less-constrained intermediary, who has higher leverage, suffers larger losses than the more-constrained one following a negative shock. Thus, the wealth share of the more-constrained intermediaries in the financial sector rises, therefore they are willing to buy the assets that the less-constrained intermediaries are trying to sell.

My model delivers intermediary asset reallocation and leverage dynamics consistent with empirical evidence: procyclical for broker-dealers and countercyclical for banks. I model intermediaries as high and low risk-tolerance agents who face occasionally binding constraints. One can obtain similar *qualitative* patterns for intermediary leverage by modeling broker-dealers and banks as risk-neutral agents who face looser and tighter funding constraints, respectively.

[4] See Coimbra and Rey (2017) and Ma (2017) for more details of such models. I discuss several dimensions along which my paper contributes to the literature in Section 1.

D Additional model results

D.1 Intermediary leverage

From the left panel of Fig. 4, we see that in the constrained model, A-types can have lower leverage than B-type intermediaries. In Fig. D.1 below, I plot the ratio of A-type intermediary's leverage to B-type's, w_s^A/w_s^B, for the baseline parameters in Table 1. In states where the leverage constraint binds, A-type intermediary can have *lower* leverage than the B-type. The relationship between portfolio weights and the state variables in Fig. D.1 is the result of market-clearing in the risky asset: As A-type intermediaries who face binding constraint deleverage and B-types lever up to clear the market, the ratio w_s^A/w_s^B declines and eventually goes below one.

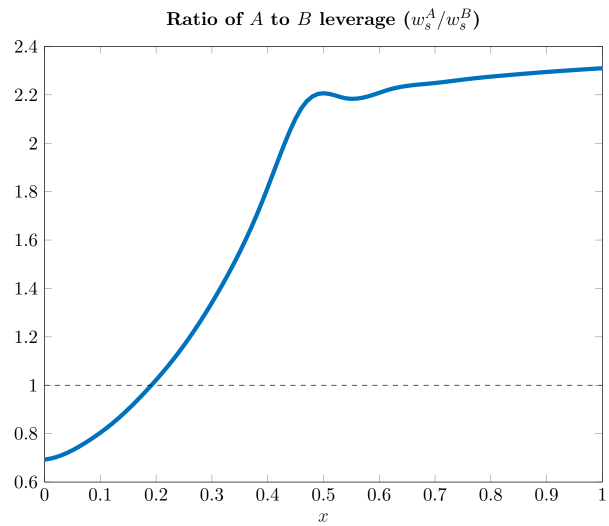


Fig. D.1. Ratio of A-type intermediary's leverage to B-type's in the model.
This figure presents the ratio of leverage for two financial intermediaries in the model: the less risk-averse A-types and the more risk-averse B-types. Each quantity is plotted against state variable x_t (wealth share of the financial sector, i.e., A and B types) while the value of the second state variable y_t (wealth share of type A-type intermediaries in the financial sector) is held fixed at its stochastic steady-state value.

To see whether the same relationship holds in the data, in Fig. D.2 I plot the ratio of BD's leverage to BHC's (empirical proxies I consider for A- and B-type intermediaries in my model) using data from the Flow of Funds, CRSP/Compustat, and Datastream. In the pre-crisis years, this ratio is well above one (between 2 and 2.5): BDs have higher leverage than BHCs. We observe that consistent with the leverage ratio in the model from Fig. D.1, during the 2008 financial crisis, however, the ratio of leverages also went well below one (to 0.64). In the data, the ratio of BD leverage to BHC leverage declines by approximately 72% (from 2.3 pre-crisis to 0.64 in 2009). In the model, from Fig. D.1, the ratio of A-type's to B-types declines by a very similar magnitude: from 2.3 to 0.69 (or by approximately 70%).

Therefore the results from the model that more risk-tolerant A-type intermediary can exhibit lower leverage than more risk-averse B-type is consistent with empirical evidence from the leverage of broker-dealers and banks during the height of the 2008 financial crisis.

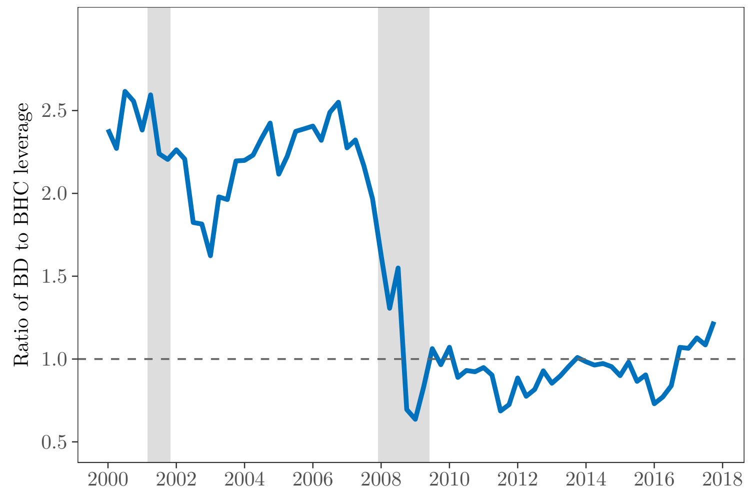


Fig. D.2. Ratio of BD to BHC Leverage in the data.
This figure presents time-series of the ratio of leverage for two financial intermediaries: security broker-dealers (BDs) and bank holding companies (BHCs). Leverage for broker-dealers is defined as the ratio total financial assets to total equity (total financial assets minus total liabilities) from Table L.130 of the Flow of Funds. BHC leverage is defined as the ratio of total market assets (book debt plus market equity) to total market equity constructed for publicly-traded holding companies of primary dealers of the New York Fed using CRSP/Compustat and Datastream. Data are quarterly from 1970Q1 to 2017Q4. The vertical shaded bars indicate NBER recessions.

D.2 Heterogeneous vs. representative intermediaries

I simulate representative- and heterogeneous-intermediary models for 3,000 quarters 20,000 times and examine the distribution of risk premium volatility. Fig. D.3 shows these distributions. As expected, the model with heterogeneous intermediaries exhibits greater variation in risk premiums than the model with a representative financial sector. Because the aggregate intermediary sectors in both models are (almost) identical, any excess variation in risk premiums in the heterogeneous intermediary model has to be attributable to state variable y. In my calibration, approximately 20% of the variation in risk premiums can be attributed to heterogeneity in the financial sector (state variable y). Therefore, failing to account for heterogeneity among intermediaries can result in missing a substantial portion of the variation in risk premiums.

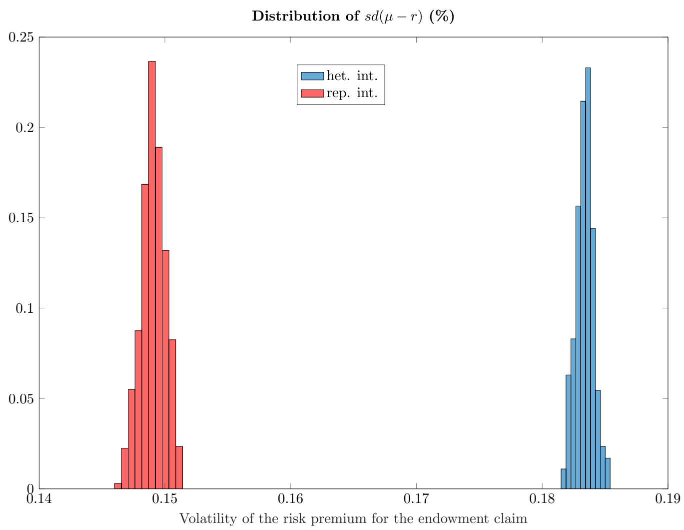


Fig. D.3. Heterogeneous and representative intermediaries.
This figure presents distributions of risk premiums volatility in models with representative (red) and heterogeneous (blue) intermediaries. I simulate each model 20,000 times for 3,000 quarters. Notice that the horizontal axis is the volatility of the risk premium for the *endowment claim,* which has a low volatility ($\sigma_D = 3.5\%$) relative to the market (approximately 16%). Therefore, the equity premium volatility implied by the model is about five to six times larger than that of the endowment claim (approximately 1% for the heterogeneous intermediary model, for example).

E Comparison with the model in HKM's appendix

In this section, I provide more details about the model in HKM's appendix and discuss several dimensions along which this paper is different. As stated by the authors in the appendix on page 32 of HKM:

> "The purpose of this appendix is to offer a general equilibrium framework to reconcile the empirical regularities documented in [AEM] and our paper."

In contrast, in my paper, I do not exclusively focus on reconciling AEM's and HKM's findings and provide additional quantitative and empirical results that I discuss in more detail after summarizing the model in HKM's appendix.

E.1 Model in HKM's appendix

I first summarize the "toy" model in HKM's appendix. This simple framework aims to reconcile the contradiction in intermediary leverage patterns between their results and the ones documented in AEM. Like my model, there are three types of agents; Households and two intermediaries: Hedge Funds and Banks. But unlike my model, households cannot participate in the risky asset market, and intermediaries are myopic with risk-neutral hedge funds/dealers and mean-variance banks. HKM's model also does not solve for asset prices nor study risk premium dynamics. More importantly, HKM's static framework is unable to capture asset pricing implications of balance sheet adjustments within the financial sector presented in Sections 6.3 and 6.4. I present a fully dynamic model that allows me to explore the quantitative implications of heterogeneity in the intermediary sector.

In HKM's appendix, households are assumed to be infinitely risk-averse (or unsophisticated) and do not participate in the risky asset market. Intermediaries consists of risk-neutral hedge funds (HFs) who face a leverage constraint that *always binds*, and risk-averse banks with mean-variance preferences and absolute risk aversion coefficient γ.

Risk-neutral HFs choose portfolio weight α_{HF} to maximize their expected wealth subject to a VaR constraint:

$$\max_{\alpha_{HF}} \mathbb{E}_t \left[W_{t+1}^{HF}\right] \quad \text{s.t.} \quad \text{Var}_t \left(R_{t+1}^{HF}\right) = \alpha_{HF}^2 \sigma_{R,t}^2 \leq \overline{\sigma}^2,$$

where $\sigma_{R,t}^2$ is the conditional variance of the risky asset return and $\overline{\sigma}^2$ is the maximum allowable risk. Since hedge funds are risk-neutral, assuming the risky asset provides a positive expected excess return, they will lever up to their constraint. Thus, we get $\alpha_{HF} = \overline{\sigma}/\sigma_{R,t}$. Bankers solve the mean-variance objective

$$\max_{\alpha_B} \mathbb{E}_t \left[W_{t+1}^B\right] - \frac{\gamma}{2} \text{Var}_t \left(R_{t+1}^B\right)$$

which leads to

$$\alpha_B = \frac{\mathbb{E}_t \left(R_{t+1} - r_{t+1}^f\right)}{\gamma \sigma_{R,t}^2}.$$

Since households cannot participate in the risky asset market, the market clearing condition for the risky asset is

$$w_B \, \alpha_B + w_{HF} \, \alpha_{HF} = 1,$$

where w_{HF} and w_B denote the wealth shares of hedge funds and banks, respectively.

As noted by the authors, the model in HKM's appendix is meant only to generate intermediary leverage patterns consistent with the data. HF leverage α_{HF} is procyclical: in bad time when $\sigma_{R,t}$ tends to be high, α_{HF} will be lower. It is important to point out that in the appendix model, HKM *assume* that $\sigma_{R,t}$ is countercyclical, and they do not solve for equilibrium asset prices and volatility. Also, because both hedge funds and banks take on leverage (by borrowing from households), their equilibrium wealth shares w_B and w_{HF} are procyclical and go down following a negative aggregate shock. These results and the market clearing condition above imply that bank leverage $\alpha_B = (1 - w_{HF}\,\alpha_{HF})\,/w_B$, increases in bad states, so banks have countercyclical leverage.

E.2 Additional results that this paper can provide

I now compare my model with the one from HKM's appendix and discuss dimensions along which my model allows me to provide additional results. Like their model, in my framework, I reconcile seemingly contradictory evidence in AEM and HKM but do so by proposing a *dynamic* general equilibrium framework with heterogeneous intermediaries and financial constraints. My model allows me to provide additional results beyond HKM's appendix model that I highlight below.

i. The dynamic model allows me to study the impact of heterogeneity among intermediaries *quantitatively*. I do so by considering a counterfactual scenario in which the entire wealth of the two heterogeneous intermediates is pulled into one representative financial sector. I then answer the following question: What fraction of the variation in risk premiums can be attributed to the state variable y, the measure of the composition of the financial sector I construct in my model? As described in Sections 6.3, I show that, in my calibration, approximately 20% of the variation in risk premiums can be attributed to this measure of the composition of the intermediary sector. I simulate the representative- and heterogeneous-intermediary models for 3,000 quarters 20,000 times and examine the distribution of the volatility of the risk premium. As expected, the model with heterogeneous intermediaries exhibits greater variation in risk premiums than the model with a representative financial sector. Since the aggregate intermediary sectors in both models are (almost) identical, any excess variation in risk premiums in the heterogeneous intermediary model has to be a result of the state variable y. The static framework in HKM's appendix is not suited to perform such counterfactual and quantitative exercises.

ii. I use my model to quantify the asset pricing implications of massive asset flows between intermediaries observed during the 2008 global financial crisis (GFC) shown in Fig. 1B. In

Section 6.4, I show that, in my calibration, a dealer deleveraging episode comparable in magnitude to the one observed during the GFC leads to an approximately 55% increase in the risk premiums and a 5% increase in endogenous risk. The static one-period framework in HKM's appendix is unable to capture the quantitative implications of balance sheet adjustments within the financial sector for the risk premium and volatility discussed in my paper.

iii. In the model in HKM's appendix, leverage constraints *always* bind for risk-neutral HFs. In my model, in contrast, intermediaries face constraints that *occasionally bind* for less risk-averse A-types. Although occasionally binding financial constraints make solving the model considerably more challenging, this is a necessary step for studying the implications of asset reallocation within the financial sector during financial crises. The reason is that financial crises are rare, and in most cases, we are interested in understanding the transitional dynamics of the economy from non-crisis to crisis states. As emphasized in He and Krishnamurthy (2019), occasionally binding constraint generate substantial non-linearity and *conditional amplification*: the same size negative shock triggers much larger declines in intermediary wealth share and asset prices when constraints are binding (or likely to bind in the near future) than when they are slack.

    The model in HKM's appendix is unable to capture these nonlinear dynamics associated with occasionally binding constraints. The presence of constraints that always bind will likely make it very difficult to achieve realistic risk premiums, intermediary leverage, and volatility dynamics in HKM's appendix model.[5]

iv. In HKM's appendix model, the authors take as given some well-known equilibrium properties of state-dependent asset pricing moments established in the literature. HKM *assume* that return volatility is countercyclical (which has empirical support) and do not solve for equilibrium asset prices and volatility. In contrast, I solve for endogenously-determined asset prices, interest rate, portfolio weights, risk premium, volatility, and state variable dynamics. In particular, as I show in Fig. 2, return volatility is hump-shaped in state variables (x, y) leading to $\cup$-shaped A-type intermediary leverage in parts of the state space where the constraint

[5] As I discussed in the main paper, my model matches the ratio of broker-dealer to bank leverages in the data but is unable to match high levels of and variations in intermediary leverage.

binds.[6]

v. Guided by my model, I present two main empirical results. First, I show that the wealth share of broker-dealers in the financial sector, a measure of the composition of the intermediary sector, strongly forecasts future market excess returns with additional predictive power beyond that of many popular forecasting variables in the literature. Second, this measure of heterogeneity has strong explanatory power for the cross-section of assets: shocks to the relative wealth share of broker-dealers in the financial sector explain the cross-section of equity and bond returns about as well or better than existing intermediary asset pricing models.

vi. Finally, in addition to matching leverage dynamics within the financial sector, my model with heterogeneous agents and recursive preferences can also generate aggregate moments for equity premium, interest rate, and consumption consistent with the empirical evidence. With myopic agents and no hedging demand, the model in HKM's appendix will most likely struggle to produce asset pricing moment consistent with the data with realistic parameter values. However, as mentioned above, the purpose of their appendix model is to reconcile the puzzling intermediary leverage dynamics documented in AEM and HKM.

F Internal capital market regulation

Affiliate transactions (Regulation W) Section 23A of the Federal Reserve Act (12 USC 371c) is the primary statute governing transactions between a bank and its affiliates. Section 23A (1) designates the types of companies that are affiliates of a bank; (2) specifies the types of transactions covered by the statute; (3) sets the quantitative limitations on a bank's covered transactions with any single affiliate, and with all affiliates combined; and (4) sets forth collateral requirements for certain bank transactions with affiliates.

Overview of Section 23A:

> *Section 23A prohibits a bank from initiating a "covered transaction" with an affiliate if, after the transaction, (i) the aggregate amount of the bank's covered transactions with that particular affiliate would exceed 10 percent of the bank's capital stock and surplus, or (ii) the aggregate amount of the bank's covered transactions with all affiliates would exceed 20 percent of the bank's*

[6] When the economy is populated by only one type of agent (i.e., when $(x, y) = (1, 1)$, $(x, y) = (1, 0)$, or $x = 0, \forall y$) the volatility of the endowment claim coincides with the fundamental volatility, i.e., $\sigma = \sigma_D$.

capital stock and surplus.[7]

Section 23A requires all covered transactions between a bank and its affiliate to be on terms and conditions consistent with safe and sound banking practices ("Safety and Soundness Requirement").

Extensions of credit to an affiliate and guarantees, letters of credit, and acceptances issued on behalf of an affiliate ("credit transactions") must be secured by a statutorily defined amount of collateral, ranging from 100 to 130 percent of the covered transaction amount. Securities issued by an affiliate and low-quality assets are not acceptable collateral for any credit transaction with an affiliate. In addition, the attribution rule provides that any transaction by a bank with any person is deemed to be an affiliate transaction subject to section 23A to the extent that the proceeds of the transaction are used for the benefit of, or transferred to, an affiliate.

Overview of Section 23B:

Section 23B requires that certain transactions, including all covered transactions, be on market terms and conditions ("Market Terms Requirement"). In addition to covered transactions, the Market Terms Requirement applies to: (i) any sale of assets by the bank to an affiliate; (ii) any payment of money or furnishing of services by the bank to an affiliate; (iii) any transaction in which an affiliate acts as agent or broker for the bank or any other person if the bank is a participant in the transaction; and (iv) any transaction by the bank with a third party if an affiliate has a financial interest in the third party or an affiliate is a participant in the transaction. In the absence of comparable transactions for identifying market terms, the bank must use terms, including credit standards that are at least as favorable to the bank as those that would be offered in good faith to nonaffiliated companies.

Source: Federal Reserve Supervisory Policy and Guidance Topics, Affiliate Transactions (Regulation W).

G Data sources

Broker-dealer and holding company data

Balance sheet data for broker-dealers and bank holding companies are from Tables L.130 and L.131 of Financial Accounts of the United States (Flow of Funds) from the Federal Reserve, respectively.

Test assets

Test assets for time-series and cross-sectional asset pricing tests are taken from two sources: (i) equity portfolios (25 portfolios formed on size and book-to-market and 10 momentum portfolios) are from Ken French's Data Library, and (ii) non-equity assets are from HKM and obtained from Asaf

[7]Covered transactions include loans and other extensions of credit to an affiliate, investments in the securities of an affiliate, purchases of assets from an affiliate, and certain other transactions that expose the bank to the risks of its affiliates.

Manela's website and include 10 maturity-sorted US government and 10 corporate bond portfolios sorted on yield spreads, six sovereign bond portfolios based on a two-way sort on a bond's covariance with the US equity market and the bond's S&P rating, 54 portfolios of S&P 500 index options sorted on moneyness and maturity split by contract type (27 calls and 27 puts), 20 CDS portfolios sorted by spreads using single-name 5-year contracts, 23 commodity portfolios with at least 25 years of return data, and 12 foreign exchange currency portfolios, six sorted on interest rate differentials and six sorted on momentum. Except for Treasury bond portfolios, which are from CRSP, non-equity test assets in HKM are from previous studies.

Intermediary asset pricing factors

AEM and HKM factors are obtained from Tyler Muir's and Asaf Manela's websites, respectively. The AEM leverage factor is defined as the seasonally adjusted growth rate in broker-dealer book leverage from Table L.130 of the Flow of Funds, where leverage is defined as total financial assets divided by total financial assets minus total volatility. The intermediary capital ratio in HKM is the ratio of total market equity to total market assets (book debt plus market equity) of primary dealer holding companies of the New York Fed. Shocks to the capital ratio (HKM's capital factor) are defined as AR(1) innovations in the capital ratio, scaled by the lagged capital ratio. Data for publicly-traded holding companies of primary dealers are taken from CRSP/Compustat and Datastream. Primary dealers are large and sophisticated institutions and serve as trading counterparties of the NY Fed in its implementation of monetary policy. For the current and historical list of primary dealers see this link.

H Robustness checks for empirical results

H.1 Predictive regressions

H.1.1 Exclude the Great Recession from the sample

As a robustness check, I remove the Great Recession (years 2007 to 2009) from the sample and rerun the predictive regressions in Eq. (31) with the market excess return as the dependent variable. Table H.1 presents the results. Consistent with the first two columns of Table 5 with the full sample, we also see a negative and significant coefficient γ_y with additional predictive power over control variables in the sample that excludes the 2008 financial crisis. Importantly, the predictive regression results are robust to excluding the Great Recession from the sample: It is not just the financial crisis that drives my predictability results.

Table H.1.
Predictive regressions: Excluding the Great Recession.
This table provides results for one-year-ahead predictive regressions according to $R^i_{t+1\to t+4} - r^f_t = \gamma^i_0 + \gamma^i_y\, y_t + \gamma^i_{\text{Ctrl}}\, \text{Ctrl}_t + \varepsilon^i_{t+1\to t+4}$, using the lagged equity share of broker-dealers in the financial sector, y^{data}_t, as defined in Eq. (30). Ctrl represents the vector of control variables that are known in the literature to forecast returns. I use the following control variables: the wealth share of the aggregate financial sector (x from the model defined in Eq. 29), the *cay* variable from Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings ratios (CAPE) from Robert Shiller's website, and the variance risk premium (VRP) from Bollerslev et al. (2009). The dependent variables are excess holding period returns from quarter $t+1$ to quarter $t+4$ on the CRSP value-weighted portfolio (Mkt_{t+1}). Broker-Dealer leverage is calculated using data from Table L.130 of Flow of Funds and is defined as Total Financial Assets/(Total Financial Assets − Total Liabilities). The sample is quarterly from 1974Q1 to 2018Q4. The Great Recession (years 2007–2009) is excluded from the sample. Hodrick (1992) standard errors are reported in parentheses to adjust for the fact that overlapping quarterly observations are used to forecast annual returns.

	Dependent variable:	
	Mkt	
	(1)	(2)
y_t		-1.06^{***}
		(0.34)
Constant	-0.12	-0.13
	(0.14)	(0.10)
Controls	Yes	Yes
Observations	97	97
R^2	0.48	0.54
Adjusted R^2	0.45	0.51
Note:	*p<0.1; **p<0.05; ***p<0.01	

H.1.2 Using the cyclical component of y as a predictor

One might be concerned that time-series trends in state variable y could impact the predictive regression results. Fig. H.1 presents the trend and cyclical components of y^{data}, defined in Eq. (30). The Hodrick-Prescott filter with a smoothing parameter of 1600 is used to separate the time series into trend and cyclical components. We observe that the cyclical component shows procyclical behavior, particularly in the later part of the sample.

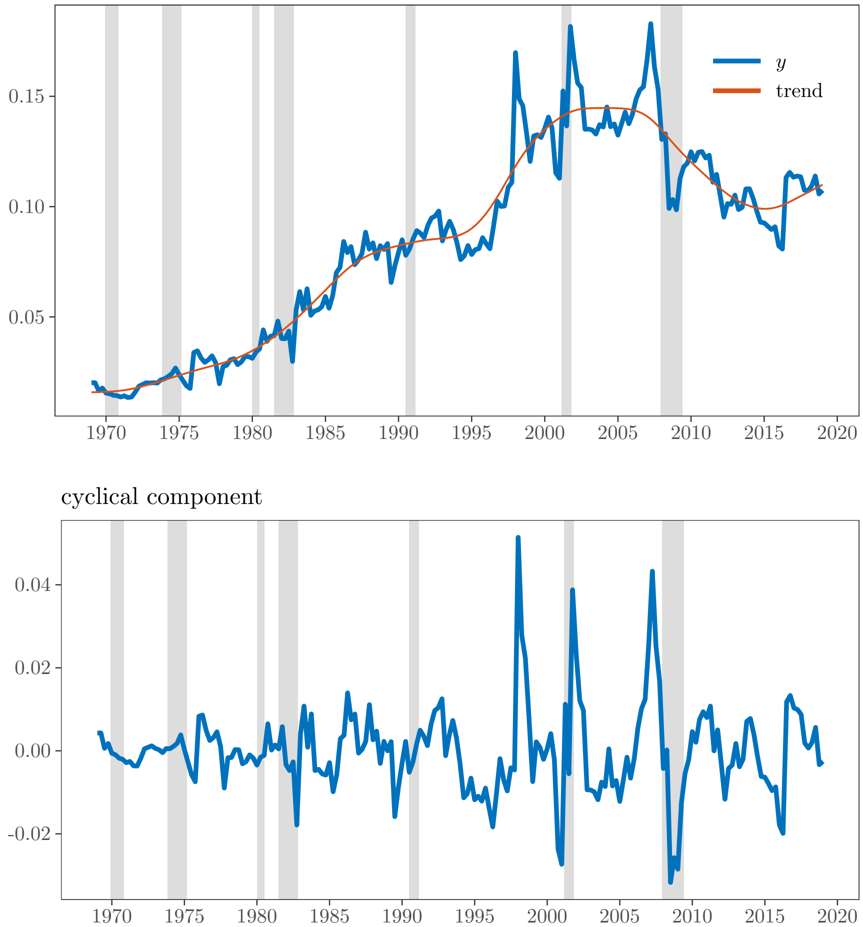


Fig. H.1. Trend and cyclical components of state variable y in the data.
This figure presents the trend and cyclical components of quarterly book equity share of broker-dealers in the financial sector, y^{data}, defined in Eq. (30). The Hodrick-Prescott filter with a smoothing parameter of 1600 is used to separate the time series into trend and cyclical components. Book equity for BDs and commercial banks are computed from the Flow of Funds Tables L.130 and L.110, respectively. The sample period is from 1970Q1 to 2018Q4. The vertical shaded bars indicate NBER recessions.

Table H.2 presents the results. Consistent with the results reported in the first two columns of Table 5, we also see a negative and significant coefficient γ_y for the lagged cyclical component of state variable y (y^{cyc}) with additional predictive power over the control variables. I also considered other ways to detrend the time series. The predictive regression results are robust to excluding the time trends in state variable y and using only its cyclical components as the main forecasting variable. In unreported regression, the forecasting regression is also robust to using 1-year and 5-year growth rates as well as the AR(1) residual of y.

Table H.2.
Predictive regressions: Using cyclical component of y as predictor.
This table provides results for one-year-ahead predictive regressions according to $R^i_{t+1\to t+4} - r^f_t = \gamma^i_0 + \gamma^i_y\, y^{cyc}_t + \gamma^i_{\text{Ctrl}}\,\text{Ctrl}_t + \varepsilon^i_{t+1\to t+4}$, using the lagged *cyclical component* of the equity share of broker-dealers in the financial sector, y^{cyc}_t, as the main predictor. The Hodrick-Prescott filter with a smoothing parameter of 1600 is used to separate the time series of y into trend and cyclical components. Ctrl represents the vector of control variables that are known in the literature to forecast returns. I use the following control variables: the wealth share of the aggregate financial sector (x from the model defined in 29), the *cay* variable from Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings ratios (CAPE) from Robert Shiller's website, and the variance risk premium (VRP) from Bollerslev et al. (2009). The dependent variables are excess holding period returns from quarter $t+1$ to quarter $t+4$ on the CRSP value-weighted portfolio (Mkt_{t+1}). Broker-Dealer leverage is calculated using data from Table L.130 of Flow of Funds and is defined as Total Financial Assets/(Total Financial Assets − Total Liabilities). The sample is quarterly from 1970Q1 to 2018Q4. Hodrick (1992) standard errors are reported in parentheses to adjust for the fact that overlapping quarterly observations are used to forecast annual returns.

	Dependent variable:	
	Mkt	
	(1)	(2)
y^{cyc}_t		-4.47^{***}
		(1.06)
Constant	0.04	-0.06
	(0.19)	(0.16)
Controls	Yes	Yes
Observations	103	103
R^2	0.33	0.42
Adjusted R^2	0.29	0.39
Note:	*p<0.1; **p<0.05; ***p<0.01	

H.1.3 Include factors from representative intermediary-based models

To further test robustness, in Table H.3, I examine the predictive power of the composition of financial intermediaries (captured by state variable y) in the presence of factors from *representative* intermediary asset pricing models studied in AEM and HKM. That is, I include AEM and/or HKM factors in predictive regressions in Eq. (31):

$$R^i_{t+1\to t+4} - r^f_t = \gamma^i_0 + \gamma^i_y\, y_t + \gamma^i_{\text{Rep}}\, \text{Rep}_t + \gamma^i_{\text{Ctrl}}\, \text{Ctrl}_t + \varepsilon_{t+1\to t+4},$$

where Rep represents the vectors of representative intermediary factors: broker-dealer leverage from AEM and BHC capital ratio from HKM. Ctrl represents the vector of control variables that are known in the literature to forecast returns. I use the following control variables: the wealth share of the aggregate financial sector (x from the model defined in 29), fluctuations in the aggregate consumption-wealth ratio (the *cay* variable) defined in Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings (CAPE) ratios from Robert Shiller's website, and the variance risk premium (VRP) from Bollerslev et al. (2009). For reference, the first column repeats the regression in Column (1) of Table 5. In Column (2), HKM's intermediary capital ratio (CapRatio) is added as an additional predictor. We observe that the coefficient on CapRatio is not statistically significant and the R^2 is only slightly increased (from 0.43 to 0.45). Removing y from Column (3) substantially reduces the R^2, by 13%, emphasizing the predictive power of my measure of intermediary heterogeneity beyond CapRatio. In Column (4), AEM's broker-dealer leverage (BDLev) is added as an additional predictor. Similarly, the coefficient on BDLev is not statistically significant and the R^2 increases only slightly increased (from 0.43 to 0.47). Removing y from Column (5) however, does not substantially reduce R^2 (only by 2%). Finally, in Column (6), all three predictors are included simultaneously: The coefficient on y remains negative and highly significant with a large R^2 of 0.47. The coefficient is also economically significant: a 1% decrease in the wealth share of dealers in the financial sector predicts a 1.2% (quarterly, 4.8% annualized) increase in the market risk premium over the next four quarters. As before, I include several control variables that are known in the literature to forecast returns: Wealth share of the aggregate financial sector (x from the model defined in Eq. 29), fluctuations in the aggregate consumption-wealth ratio (*cay* variable) defined in Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings (CAPE) ratios from Robert Shiller's website, and the variance risk premium (VRP) from Bollerslev et al. (2009).

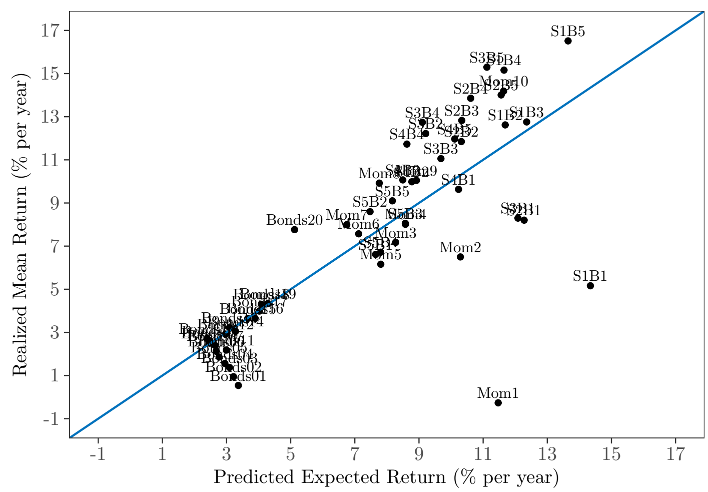
Table H.3.
Predictive regressions for the MKT excess return: Robustness.
This table provides results for one-year-ahead predictive regressions using the lagged equity share of broker-dealers in the financial sector, the empirical proxy for the composition of the financial sector as defined in Eq. (30), as well as the intermediary equity capital ratio (from HKM) and the leverage of broker-dealers (from AEM) as predictors of interest. I use the same control variables as in Table 5: the wealth share of the aggregate financial sector (x from the model defined in 29), the *cay* variable from Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings (CAPE) ratios from Robert Shiller's website, and the variance risk premium (VRP) from Bollerslev et al. (2009). The dependent variable is excess holding period returns from quarter $t+1$ to quarter $t+4$ on the CRSP value-weighted portfolio (Mkt_{t+1}). The sample quarterly from 1974Q1 to 2017Q3. Broker-Dealer leverage is calculated using data from Table L.130 of Flow of Funds and is defined: Total Financial Assets/(Total Financial Assets − Total Liabilities). The capital ratio for the New York Fed's primary dealer holding companies are downloaded from Asaf Manela's website. Hodrick (1992) standard errors are reported in parentheses to adjust for the fact that overlapping quarterly observations are used to forecast annual returns.

	Dependent variable: Mkt_{t+1}					
	(1)	(2)	(3)	(4)	(5)	(6)
y_t	-1.80^{***}	-1.98^{***}		-0.97^{*}		-1.19^{***}
	(0.43)	(0.48)		(0.51)		(0.43)
CapRatio_t		3.31	1.22			1.68
		(3.58)	(4.58)			(2.25)
BDLev_t				-1.35	-1.98^{***}	-1.13^{*}
				(0.85)	(0.56)	(0.59)
Constant	0.01	0.18	0.10	0.05	0.09	0.13
	(0.13)	(0.26)	(0.38)	(0.15)	(0.17)	(0.22)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	109	109	109	109	109	109
R^2	0.46	0.49	0.32	0.51	0.48	0.51
Adjusted R^2	0.43	0.45	0.28	0.47	0.45	0.47
Note:				*p<0.1; **p<0.05; ***p<0.01		

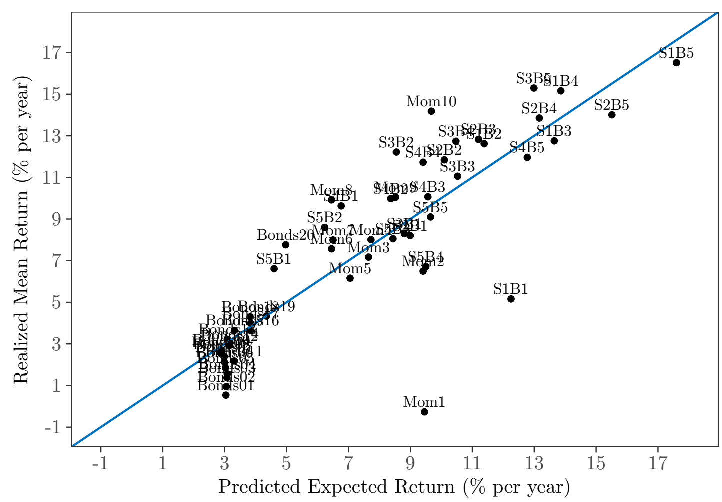
H.2 Cross-sectional asset pricing tests

H.2.1 HIFac's pricing performance

Fig. H.2 shows the HIFac's pricing performance visually: the top panel plots the annualized realized against the predicted excess returns for the 55 equity and bond portfolios when HIFac is the only pricing factor (Column (1) in Table 6). Most of the portfolios line up closely with the 45-degree line. The bottom panel is similar to the top panel when HIFac and AEM are used as pricing factors, corresponding to Column (4) in Table 6. The model slightly outperforms the one used for panel (a) as shown above.



A. HIFac only.



B. HIFac + AEM.

Fig. H.2. Realized versus predicted mean returns: intermediary heterogeneity factor.
This figure presents the realized mean excess returns of 35 equity portfolios (25 size and book-to-market-sorted portfolios and 10 momentum-sorted portfolios), 10 Treasury bond portfolios (sorted by maturity), and 10 US corporate bond portfolios (sorted by yield spread) against the mean excess returns predicted by the single heterogeneous intermediary risk factor when only the heterogeneous intermediary factor (HIFac) (panel a) and HiFac and AEM factors (panel b) are used as pricing factors, respectively. The sample is quarterly from 1970Q1 to 2017Q4. Returns are reported in percentages per year (quarterly percentages multiplied by four).

Sorted CRSP portfolios

To empirically verify the positive price of risk for innovations in the wealth share of dealers in the financial sector, I sort stocks based on their exposures to these shocks and form portfolios by quintiles on a 10-year trailing window. I consider all common stocks (share codes 10 and 11) in the CRSP universe from Amex, NASDAQ, and the NYSE (exchange codes 1, 2, and 3). For every stock i in quarter t, I regress its quarterly excess return on a constant and the HIFac, defined in Eq. (33):

$$R^e_{i,t} = \alpha_i + \beta_{i,\text{HIFac}}\,\text{HIFac}_t + \xi_{i,t}$$

The coefficient $\beta_{i,\text{HIFac}}$ measures the exposure of firm i's stock to the factor's innovations. I then sort stocks into quintiles every quarter according to their $\beta_{i,\text{HIFac}}$.

The average returns on the beta-sorted portfolios are reported in Table H.4, along with return volatilities, average book-to-market ratios, average market caps, and alphas from CAPM and the Fama-French three-factor model. Consistent with the model's implications, when sorted on β_{HIFac}, the average risk premiums are increasing from the portfolio of low-beta stocks to the high-beta quintile. Excess returns are monotonically increasing from quintile one to quintile five and the top portfolio earns an approximately 5% premium over the lowest quintile.

Table H.4.
Sorted CRSP portfolios on exposures to the heterogeneous intermediary factor.
This table reports average excess returns, alphas, volatilities, average book-to-market ratios, and average market capitalizations for portfolios formed on their exposure to shocks to the dealer wealth share in the financial sector. Shocks to the dealer wealth share (HIFac) are defined as AR(1) innovations in the wealth share, scaled by the lagged wealth share as shown in Eq. (33). Data is quarterly from 1970Q1 to 2017Q3. Returns, volatilities, and alphas are annualized.

	L (1)	(2)	(3)	(4)	H (5)	HML (6)
Average Excess Return (%)	11.66	11.53	12.81	14.34	16.65	4.98
Volatility (%)	19.69	19.08	21.76	26.41	35.53	26.72
β_{HIFac}	−0.20	0.33	0.69	1.19	2.21	2.41
t-stat	−0.89	1.50	2.77	4.07	5.89	10.67
α_{CAPM}	4.77	4.33	4.56	4.75	4.44	−0.32
t-stat	3.40	3.90	3.63	2.84	1.57	−0.10
α_{FF3}	3.89	2.88	3.36	3.98	4.02	0.14
t-stat	3.14	4.21	5.45	5.04	2.24	0.05
Average Market Cap ($bn)	5.28	3.66	2.40	1.97	0.89	–

H.2.2 The Heterogeneous intermediary factor-mimicking portfolio

As emphasized above, the main argument of the paper is that heterogeneity in the intermediary sector has important implications for asset prices. To conduct additional robustness tests, in this section, I project the heterogeneous intermediary factor (HIFac) onto the space of traded returns to form a factor-mimicking portfolio that mimics the HIFac. To further verify that this heterogeneity is an important source of risk, I evaluate the heterogeneous intermediary factor-mimicking portfolio (HIMP) relative to the mimicking portfolios for representative intermediary factors in AEM and HKM. I show that the mimicking portfolios for these representative intermediary factors cannot fully span the HIMP and there is more to be captured by considering heterogeneity within the financial sector.

This approach also allows me to run tests using higher frequency data and longer time series. Moreover, because mimicking portfolios are traded excess returns, I can evaluate the model by testing alphas in time-series regressions without the need to estimate the cross-section of risk prices.

Construction of the HIMP To construct the mimicking portfolio of HIFac, I follow AEM and project this factor onto the space of excess returns by running the following regression:

$$\text{HIFac}_t = a_{\text{HI}} + b'_{\text{HI}} \left[BL, BM, BH, SL, SM, SH, Mom, Bond\right]_t + \varrho_t, \tag{H.1}$$

where HIFac is the heterogeneous intermediary factor defined in Eq. (33), and BL, BM, BH, SL, SM, SH are, respectively, the excess returns of the six Fama-French portfolios on size (Small and Big) and book-to-market (Low, Medium, and High), and Mom is the momentum factor, obtained from Ken French's data library. $Bond$ is the first principal component (PC) of excess returns on six Treasury bond portfolios sorted by maturity from CRSP. The heterogeneous intermediary mimicking portfolio (HIMP) is then given by

$$\text{HIMP}_t = \widetilde{b}'_{\text{HI}} \left[BL, BM, BH, SL, SM, SH, Mom, Bond\right]_t, \tag{H.2}$$

where $\widetilde{b}_{\text{HI}} = \frac{b'_{\text{HI}}}{\sum b_{\text{HI}}} = [-0.34, 0.20, -1.04, -0.09, 0.41, 1.64, 1.04, -0.83]$ positively loading on the momentum factor.

HIMP vs. mimicking portfolios for AEM and HKM factors To further verify that my heterogeneous intermediary factor captures sources of risk beyond the factors from representative intermediary asset pricing models, in this section I evaluate the performance of HIMP with mimicking portfolios for AEM and HKM factors. I similarly construct mimicking portfolios for AEM's broker-dealer leverage and HKM's holding company capital factors using quarterly data for the two factors from Tyler Muir's and Asaf Manela's websites, respectively.[8] The mimicking portfolio for the heterogeneous intermediary factor has a Sharpe ratio of 0.45 over the sample period (1970Q1 to 2017Q3), which is much higher than the Sharpe ratios for AEM and HKM factor-mimicking portfolios (0.21 and 0.27, respectively).

To evaluate the importance of heterogeneity in the financial sector above an beyond representative intermediary factors, I regress HIFac on mimicking portfolios for AEM and HKM factors in the following regression:

$$\text{HIMP}_t = \alpha_{\text{MP}} + \beta'_{\text{FMP}} \, \text{FMP}_t + \epsilon_t, \tag{H.3}$$

where FMP is either the mimicking portfolio for broker-dealer leverage factor from AEM (AEM_MP), or the mimicking portfolio for capital factor for primary dealers' holding companies from HKM (HKM_MP), or both AEM_MP and HKM_MP. Notice that the mimicking portfolios are traded excess returns, enabling me to evaluate the model by testing alphas in the time-series regression without the need to estimate the cross-section risk prices. If HIMP is fully "explained" by AEM_MP, HKM_MP, or both, I expect to see small and insignificant α_{MP} in the regression above. I find the opposite to be true, however.

Table H.5 presents the results. In Columns (1) and (2), I run univariate regressions where the dependent variables are AEM_MP and HKM_MP, respectively. In both cases the intercept, α_{MP}, is statistically significant at the 1% level and the R^2 of the regressions are relatively low, at 0.14 and 0.32, respectively. In Column (3), I add value-weighted returns from CRSP (MktRF) to HKM_MP as independent variables, which generate results that are very similar to those reported in Column

[8] The loadings for AEM and HKM factor-mimicking portfolios are $\widetilde{b}_{\text{AEM}} = [-0.98, 0.50, -0.03, -0.26, 0.96, 0.05, 0.16, 0.59]$ and $\widetilde{b}_{\text{HKM}} = [0.30, 0.03, 0.58, -0.06, -0.16, 0.25, 0.09, -0.03]$.

(3). In Column (4), I add both AEM and HKM factor-mimicking portfolios as right-hand-side variables in Eq. (H.3). We observe a large and significant α_{MP} and relatively small R^2. Adding MktRF in Column (5) to the regression in Column (4) further strengthens these results.

As an additional (unreported) robustness check, I build factor-mimicking portfolios by projecting HIFac, AEM, and HKM factors instead onto the Fama-French three factors , the momentum factor, and the first PC of bond portfolios, and repeat the regressions in Table H.5. I arrive at very similar results: time-series alphas are large and significantly different from zero with low R^2 in all regressions.

Table H.5.
The heterogeneous intermediary mimicking portfolio (HIMP): Comparing models.
This table presents time-series regression results of the HIMP on mimicking portfolios for the representative intermediary factors in AEM and HKM according to: $\text{HIMP}_t = \alpha_{\text{MP}} + \beta'_{\text{FMP}}\,\text{FMP}_t + \epsilon_t$, where FMP is either the mimicking portfolio for the broker-dealer leverage factor from AEM (AEM_MP), or the mimicking portfolio for the capital factor for primary dealers' holding companies from HKM (HKM_MP), or both AEM_MP and HKM_MP. The factor-mimicking portfolios are constructed by projecting the HIFac, AEM's leverage factor, and HKM's capital factor, unto the space of equity and bond returns according to Eq. (H.1) and (H.2). The sample is quarterly from 1970Q1 to 2017Q3. Standard errors are in parentheses.

	Dependent variable: HIMP				
	(1)	(2)	(3)	(4)	(5)
α_{MP}	5.03***	4.03***	4.14***	3.74***	3.91***
	(0.94)	(0.85)	(0.85)	(0.83)	(0.83)
AEM MP	0.72***			0.37***	0.49***
	(0.13)			(0.12)	(0.13)
HKM MP		0.94***	0.68**	0.82***	0.15
		(0.10)	(0.27)	(0.10)	(0.30)
MktRF			0.27		0.65**
			(0.26)		(0.27)
Observations	191	191	191	191	191
R^2	0.14	0.32	0.33	0.35	0.37
Adjusted R^2	0.14	0.32	0.32	0.35	0.36
Note:			*p<0.1; **p<0.05; ***p<0.01		

This exercise confirms my earlier results: the heterogeneity in the financial sector is an important risk factor and has pricing information above and beyond representative intermediary asset pricing factors in AEM and HKM.

References for the Internet Appendix

References

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. The Review of Financial Studies 22, 4463–4492.

Coimbra, N., Rey, H., 2017. Financial cycles with heterogeneous intermediaries. Unpublished working paper 23245. National Bureau of Economic Research.

Duffie, D., Lions, P.-L., 1992. PDE solutions of stochastic differential utility. Journal of Mathematical Economics 21, 577 – 606.

He, Z., Krishnamurthy, A., 2019. A macroeconomic framework for quantifying systemic risk. American Economic Journal: Macroeconomics 11, 1–37.

Hodrick, R. J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. The Review of Financial Studies 5, 357–386.

Judd, K. L., 1992. Projection methods for solving aggregate growth models. Journal of Economic Theory 58, 410 – 452.

Judd, K. L., 1998. Numerical Methods in Economics. MIT Press, Cambridge.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. The Journal of Finance 56, 815–849.

Ma, S., 2017. Heterogeneous intermediaries and asset prices. Unpublished working paper. Federal Reserve Board.