Banking with Risky Assets

Patrick C. Kiefer*† UCLA Anderson

April 26, 2018

Abstract

We show rebalancing risks in incomplete markets are mitigated by a bank. Ex-ante, the bank exchanges risky endowments for demandable liabilities. An ex-post withdrawal corresponds to exercising a put option on the market, used to resolve an unexpected portfolio choice problem. Portfolio choice opens a risk aversion channel that distinguishes our predictions from Diamond and Dybvig (1983) and related models. In these models, deposits resolve consumption-timing tensions by accommodating the investor's intertemporal elasticity of substitution (IES). The inclusion of risk-based incentives allow us to characterize the endogenous link between the intermediary balance sheet and the preference-based pricing kernel. Moreover, ex-post rebalancing incentives relax enforcement problems for ex-ante optimal policies in incomplete markets. This provides a justification for the coexistence of intermediation and market institutions.

*I thank Mark Grinblatt and Doug Diamond for detailed comments. I thank my committee for their patience and encouragement. This document is taken from my dissertation work on financial intermediation.

†Comments welcome: pkiefer@anderson.ucla.edu; https://sites.google.com/site/patrickckiefer

Introduction

A significant fraction of the typical individual's net worth is risky but not insurable. This comes with direct costs, from uninsurable losses, but also indirect costs. Losses in the uninsurable component of wealth change the composition of an investor's portfolio. This in turn requires rebalancing through the remaining, *tradeable* component of wealth, at prevailing prices. Direct losses are amplified if market prices are low and selling risky claims is costly.

We model a financial intermediary designed to mitigate these costs. The intermediary holds the tradeable component of investors' wealth on it's balance sheet and issues demandable claims as remuneration. Demandability allows bank investors to withdraw up to the cash value of their deposited endowment in exchange for a pro-rata share reduction in the risky asset. This arrangement improves welfare when markets are incomplete. The intermediary partially insures shocks directly by facilitating a transfer proportional to one minus the net capital gain on traded assets. The risk-sharing mechanism links the intermediary's balance sheet and the equilibrium pricing kernel. This link provides novel testable implications.

Investors can rebalance implicitly through the bank rather than directly in securities markets when bank assets are *risky*. To illustrate how this works, note that shocks to the uninsurable component of wealth are independent of the market, and that ex-ante, each investor holds her optimal portfolio. Now, ex-post low private income investors have an unexpectedly high fraction of their total wealth in the risky claim. In response, they liquidate risky holdings by exercising the cash option. Ex-post high private income investors are conversely under-exposed to the aggregate claim - but, they acquire a leveraged position in the risky asset *passively*. In the simplest case, the net result is that each type attains their optimal portfolio exactly.

Because the withdrawals are made ex-post based on portfolio demand, state dependence of the bank's capital structure reflects a wealth-weighted average of investor's effective risk aversion. The ex-post rollover rates capture the representative risk prices in each aggregate state. Moreover, the ex-ante asset levels capture the representative inter-temporal marginal rate of substitution (IES). As a result, our theory ties the bank balance sheet to a complete description of the equilibrium pricing kernel.

A challenge for theories linking intermediation to asset prices through individual preferences is that intermediary and market institutions should be jointly accessible. In the context of explaining an institution's role in resolving incentive conflicts, a common assumption is to prohibit investors from accessing markets directly. Institutional preferences that reflect the modified interests of a collection of individuals can then stand-in for a marginal investor. However, the market restriction is empirically implausible.[1] Resulting explanations for equilibrium asset prices are tenuous. A theory linking banks to asset prices through individual preferences cannot rely on restricting access of individuals to markets.[2]

Alternatively, by abstracting from preferences in the population, the institution can be endowed directly with preferences or interests and used as the representative agent.[3] However, this abstraction becomes costly in the face of welfare, benchmarking and policy analyses.[4] In contrast, a theory linking intermediation to the preferences of an investing population can be productively integrated in the normative space.

The theory presented in this paper links the intermediary balance sheet to investor preferences, and allows investors access to exchange and banking institutions simultaneously. This trade-off in each period is what keeps a positive measure of investors positioned in the bank. Moreover, the rebalancing motive is risk-based, while the deposit motive is IES-based. Together, they address the problem of time-consistent policies. Investors' ex-post policies are optimal solutions to the contemporaneous portfolio-choice problem. Thus, the mechanism addresses the enforcement problem in incomplete markets in addition to the verification problem.

[1]For exchange-traded equities and indices, the assumption is implausible. The importance of the incremental effort required to access options markets, bond markets, REITs etc, is debatable. In the latter case, lower observed participation rates are not because of inability to access, but rather a choice not to access.

[2]In full nuance, the theory cannot produce a representative agent that is restricted. Subtler forms of heterogeneity in place of blanket restrictions, e.g., the Lucas family device (Lucas 1990), can produce a plausible theory. Lucas (1990) models a representative "family" by restricting family members to certain tasks within each period, but then aggregating decisions at the family level. We do not require this device.

[3]Krishnamurthy (2014) models the intermediary's marginal value of reputation.

[4]Welfare ideally incorporates the effects on individual utility and efficiency including endogenous equilibrium effects, ruled out by this abstraction. Benchmarking against competing or complementary theories based on individual optimization is also limited.

Haubrich and King (1990) critique the view that banks uniquely produce liquidation options the are credible because of fragility. They argue that the bank services can be broken into a liquidity component, that can be provided in ex-post securities markets, and an insurance component, that can be replicated by a mutual fund with the right configuration of coupon payments and share purchases. However, our setting relies on aggregate risk for asset pricing. The securities markets do not provide liquidity when market price depend on the aggregate state. Although their original analysis is not done with aggregate risk, in principle a mutual could announce any coupon and share purchase policy made contingent on the aggregate state and thus may be able to reproduce the bank system allocations. We show that replication by a mutual is in fact not possible.

The key mechanism in our theory precluding replication by a mutual or other market institution is the *synthetic* leverage generated by the bank to accommodate the various claimants to its assets. When a bank financier makes a withdrawal, the bank debits the residual capital account. Bank capital becomes leveraged and the corresponding changes in bank capital risk are *synthetic*, because the bank does not need to clear its shares and liabilities in the market contemporaneously. In contrast, the mutual marks-to-market, in that changes in its liabilities must correspond to changes in its assets. Both the mutual and the economy are unlevered in equilibrium. The bank precludes contemporaneous unwinding of synthesized leverage, and generates both the concentrated risk and the negative cash position high types need to be indifferent to rolling over the bank position.

0.1 Related Work

This paper is motivated by at least two areas of research. The first is the theoretical literature on bank liability design based on Diamond and Dybvig (1983). This literature constitutes the basis for understanding endogenous intermediation liquidity creation. The second is the literature on asset pricing and financial intermediation, beginning with empirical work by Adrian, Etula and Muir (2015). Drawing on methodologies and outstanding questions from each of these areas, we show portfolio choice motives are sufficient to microfound a financial intermediary. Because the intermediary microfounded in this way is financed dynamically based on preferences for risk, the intermediary balance sheet is linked to the incomplete markets stochastic discount factor (SDF).

Diamond and Dybvig (1983) model a bank deposit contract that solves the ex-ante allocation problem for a large economy of individuals who are uncertain about the timing of their consumption needs, and where capital is only productive in the long term. Ex-post population frequencies of near and long term consumption are known and there is no aggregate risk. Deposits allow investors to delay their commitment to an allocation between short and long term investments until their consumption timing preference is revealed. Knowledge of the ex-post population frequencies permits the bank to allocate the deposits between short and long term investments more efficiently ex-ante.

Haubrich and King (1990) argue that ex-post securities markets can provide the same liquidity as deposits when there is no aggregate risk. Moreover, they argue the bank per-se is not preferred over a mutual unless transactions in securities markets between individuals are restricted. We show that the ex-post securities market does not provide this liquidity when there is aggregate risk and prices vary across states ex-post. As a result, preference for the intermediary does not rely on restricted access to markets for aggregate claims. In fact, there is no need for the bank to price discriminate based on timing, because the trade-off between ex-post market prices and the common initial price of the bank claim ensures investors adjust ex-post funding predictably.

Using constant elasticity of substitution (CES) utility, Haubrich and King (1990) argue deposit contracts driven by consumption timing, such Diamond and Dybvig (1983), are driven by the inter-temporal elasticity of substitution (IES). Our theory extends the work of earlier theories to portfolio allocation motives. We show our ex-post withdrawal policies are set from a portfolio rebalancing motive rather than a consumption-savings motive. Ex-ante, the IES drives savings policy and impacts bank deposit levels, but ex-post, the withdrawal policy is a function of risk aversion only. Through this channel, the bank's capital structure is connected to risk-based asset pricing, and hence the equilibrium SDF.

The Diamond and Dybvig (1983) model and its progeny use a sequential timing protocol for deposit withdraws to show coexistence of inefficient bank-run equilibria. Multiple equilibria led to an insight about the fragility of a bank funding structure as the source of its strength. Depositors *en-masse* credibly threaten runs simply by owning the demandability option, thus providing a source of discipline for banks. If a bank makes a promise ex-ante to honor ex-post withdrawals, the threat of runs prevents the banks from reneging on their promise. Common knowledge of this device ex-ante makes formation of the bank possible.

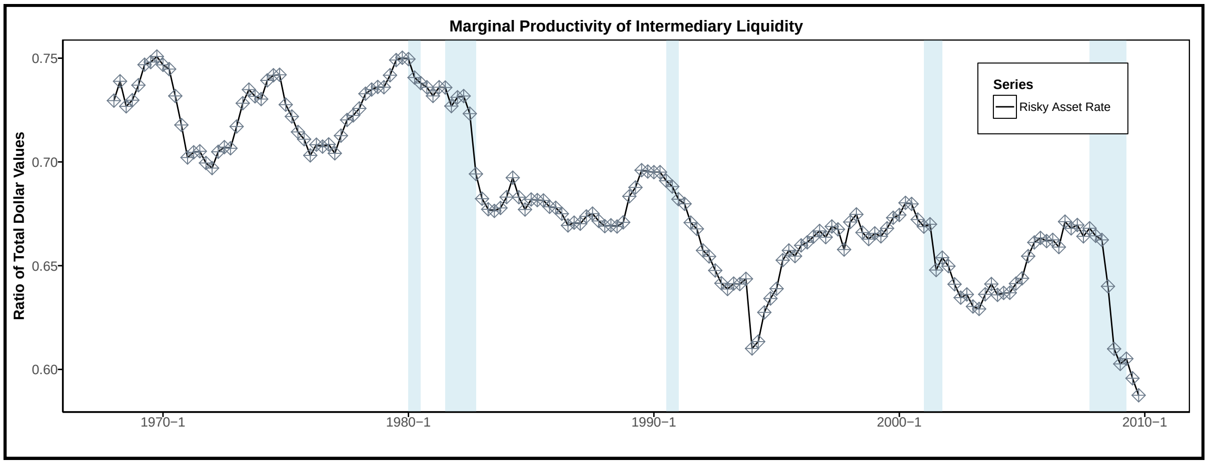
We do not use a sequential service constraint. However, the rebalancing motive gives novel insight into bank funding trade-offs. Risky assets that have near-zero cash flows with positive probability preclude bank formation. Disaster risk impacts the ability of the bank to credibly produce liquidity, requiring the bank hold a cash buffer. In contrast, risky assets in the absence of disaster risk can improve the efficiency of liquidity production. While runs as sunspot equilibria are not a feature of our model, we show bank funding is contingent on the competitiveness of its expected return.

Our study has implications for the coexistence of intermediation and securities markets with unrestricted access. Allen and Gale (2010) study an economy with aggregate risk where investors have restricted access to securities markets, and derive the constrained-optimal asset holdings for a financial intermediary. We find that when intermediation appeals to portfolio motives, investors trade-off bank financing with direct trade in securities markets based on the direction of their trade and market prices. Expected utility is maximized when both institutions are accessible.

Our investigation is influenced by the recent literature connecting intermediary balance sheet dynamics and asset prices. A key empirical contribution is Adrian, Etula and Muir (2015), who find that asset return exposure to shocks to dealer leverage can explain cross-sectional variation in average returns. He, Kelly and Manilla (2017) find that a measure of bank capital can be used to explain the cross-sections of asset classes outside of equity and bond markets. Adrian et al (2012) shows a related measure of leverage has predictive power for market returns.

The empirical literature also studies the balance sheet dynamics of financial institutions.

Figure 1: Ratio of High-Risk Assets to Liquid Liabilities



(a) The rate at which the highest risk assets contribute to liability-side liquidity production drops sharply in recessions. Separately, the rate exhibits a secular trend downward. Quarterly balance sheet data from 1967 $Q1$ to 2012 $Q4$ are from the Flow of Funds, Board of Governors of the Federal Reserve. We use private depository institutions, issuers of asset-backed securities, and securities brokers and dealers to measure liquidity production. The ratio of high risk assets to liquid liabilities is calculated by classifying liquid liabilities as large time deposits, uninsured checkable and savings deposits, asset backed commercial paper (ABCP) and repurchase agreements. Risky assets are corporate equities, mutual fund shares, and private residential and commercial mortgage-backed securities (MBS). NBER recessions are in blue.

Adrian et al (2010), and Boyarchenko et al (2011), document leverage dynamics of broker dealers and commercial banks respectively, each of which are liquidity producers. Krishnamurhty et al (2014) document that commercial banks take on debt to acquire risky assets in bad times, which are being sold off by non-liquidity producing investment institutions like hedge funds, mutual funds, pensions and others. In Fig.1, we plot the rate of high-risk assets to liquid liabilities. We see a sharp downturn in every recession, as well as a low frequency trend downwards.

A theoretical literature in this area includes and Krishnamurthy (2013), Brunnermeier and Sannikov (2015) and Adrian and Boyarchenko (2015). Our theory differs in emphasis and methodology, and as such the two approaches are complementary. IAP theories do not aim to justify the intermediary in equilibrium. These theories extend dynamic asset pricing models designed to produce quantitative statements about the dynamics of asset returns in a variety of experimental settings to cases where a financial intermediary is marginal in securities markets. They also provide economic insight into how financial intermediation impacts risk and return.

In our model economy the financial intermediary is endogenous. The asset pricing implications in our model arise because of the connection between a preference for holding risky assets in the cross-section and the aggregate bank capital structure. This connection is an equilibrium outcome. The cost of this insight is that quantitative exercises in our stylized model are difficult to justify. However, qualitative predictions from our model generate testable implications that are valid in dynamic settings. We provide reduced-form empirical evidence that supports our theory.

0.2 Example

To highlight the economic mechanism we present an example. The rigorous description of the model begins in section 1.

The Setting Consider two investors, I_1 and I_2, who each own equal claims $\frac{1}{2}V_0$ to a project valued V_0. The project pays an uncertain amount Y two periods from today. Each investor i is also entitled to an uncertain cash payment n_i that cannot be insured. The investors have log utility over final wealth. For simplicity we assume $n_i \in \{-\Delta, \Delta\}$ and $n_1 + n_2 = 0$. Each of the two configurations occurs with equal probability. The investors begin with equal stores of "cash" $\frac{1}{2}C > 0$ and have access to a storage technology with per-period gross return normalized to one. Total initial wealth is $W_0 = V_0 + C$.

Timing In the first period $t = 1$, the payments n_i are revealed. Prospects for the project payout Y are also revealed through a signal $m \in \{L, H\}$ corresponding to low and high productivity. In $t = 2$, Y_m is either above A or below B its conditional forecast $Y_{m,A} > E[Y_{m,k}|m]$, $Y_{m,B} < E[Y_{m,k}|m]$. Claims to the project output are traded competitively in periods $t = 0, 1$. Time $t = 1$ prices are V_m.

Implications Because the investors are identical ex-ante, and the income n_i is uninsurable, there is no incentive to modify holdings at time-zero. Each investor's initial portfolio Π_0 can be written $\Pi_0 = (\alpha_0 W_0, (1 - \alpha_0) W_0)$ for initial wealth W_0. When normalized by wealth, the first entry corresponds to the rate of wealth invested in equity at $t = 0$ and the second entry is the rate of cash investment. With no trade, $\alpha_0 = V_0/W_0$.

At $t = 1$, idiosyncratic income n_j is realized, along with the news about productivity m. We let $i = 1$ correspond to the unexpectedly wealthy investor: $n_1 = \Delta > 0$. The two investors have identical shares of the risky asset at the *beginning* of the first period $a_0 = \frac{1}{2}$. However, idiosyncratic shocks produce heterogeneous wealth levels $W_j = a_0 V_m + \frac{1}{2}C + n_j$. As a result, the unlucky investor has an oversized rate of investment in the risky project $a_0 V_m/W_2 > a_0 V_m/W_1$.

Securities Trading The investors adjust positions until investment rates are equalized at the *end* of the first period, i.e., $\Pi_{1,1}/W_{1,1} = (\alpha_1, (1 - \alpha_1)) = \Pi_{1,2}/W_{1,2}$. Subscripts (t, i) in $W_{t,i}, \Pi_{t,i}$ indicate the period and the investor, respectively. The market clearing share is $\alpha_1 = V_m [C + V_m]^{-1}$. Changes in individual wealth levels do not impact individual portfolio weights α. With log utility, each investor splits their idiosyncratic income pro-rata $n_i = [\alpha_1 n_i]_{\text{Equity}} + [(1 - \alpha_1)n_i]_{\text{Cash}}$. Transactions are made at the ex-post market prices.

Written with terms gathered by position, the final wealth shares are

$$\begin{aligned} W_{1,m,1} &= V_m \left[\frac{1}{2} + \frac{\Delta}{W_{1,m}}\right] + C\left[\frac{1}{2} + \frac{\Delta}{W_{1,m}}\right] \\ W_{1,m,2} &= V_m \left[\frac{1}{2} - \frac{\Delta}{W_{1,m}}\right] + C\left[\frac{1}{2} - \frac{\Delta}{W_{1,m}}\right] \end{aligned} \tag{WS.0}$$

The first term, scaled by the share price V_m, is the market value of each investors equity position. The second term is the risk-free position. Anticipation of the ex-post distribution of net-worth in WS.0 is reflected in initial prices.

The Bank The investors instead create the following arrangement. The investors deposit their claims in a bank, and allow the bank to hold the risky claim as an asset. The investors now hold the liabilities of the bank instead of the claim to the project. In turn the bank includes provisions in the liabilities that allow the investors to withdraw any amount of the cash value of their deposit before the project matures, at which point the proceeds are paid to the residual claimants of the bank assets, and the bank is dissolved.

Implications with the Bank For simplicity, investors deposit the cash value $b_0 = \alpha_0\Delta$ of their risky endowment, or equivalently, $\alpha_0\frac{\Delta}{V_0}$ shares of their risky endowment, in the bank. This leaves a direct equity position with cash value $[\frac{1}{2} - \alpha_0\frac{\Delta}{V_0}]V_0$, and the cash position $\frac{1}{2}C$. Write $\mathbf{b} = 2\Delta\frac{\alpha_0}{V_0}$ for the fraction of the risky claim intermediated at time-zero.

Now, at $t = 1$, in a recession $m = R$, the investor with $n_2 = -\Delta < 0$ withdraws cash from the bank in the amount of b_0. This transaction liquidates $\frac{\alpha_0}{V_0}\Delta$ shares with market value $\frac{\alpha_0}{V_0}\Delta V_m < \frac{\alpha_0}{V_0}\Delta$. The low type has no remaining exposure to the bank. We can calculate her shares explicitly: $\frac{b_0-\alpha_0\Delta}{2b_0-\alpha_0\Delta} = 0$. The high type passively acquires the residual bank position: $\frac{b_0}{2b_0-\alpha_0\Delta} = 1$. With no further action, the resulting portfolios are

$$\frac{\Pi_{1,1}}{W^b_{1,1}} = \underbrace{\left[\frac{1}{2} - \frac{\alpha_0\Delta}{V_0}\right]V_m}_{\text{Direct Equity}}, \quad \underbrace{\mathbf{b}\,V_m}_{\text{Bank Liabilities}}, \quad \underbrace{\frac{1}{2}C + \Delta(1-\alpha_0)}_{\text{Cash}}$$

$$\frac{\Pi_{1,2}}{W^b_{1,2}} = \underbrace{\left[\frac{1}{2} - \frac{\alpha_0\Delta}{V_0}\right]V_m}_{\text{Direct Equity}}, \quad \underbrace{0}_{\text{Bank Liabilities}}, \quad \underbrace{\frac{1}{2}C - \Delta(1-\alpha_0)}_{\text{Cash}}$$

By combining the bank and direct equity exposures into a single equity position, we can write the portfolios

$$\frac{\Pi_{1,j}}{W^b_{1,j}} = \alpha_1, (1-\alpha_1) \qquad\qquad j = 1, 2$$

for wealth shares $W^b_{1,j} = W_1\left(\frac{1}{2} + (-1)^{j-1}\Delta\right)$. Through the bank's balance sheet, the withdrawal policy of the unlucky investor successfully implements her required portfolio adjustment, *as well as the portfolio adjustments of the lucky investor.*

Remark Relative to the incomplete markets wealth share $W_{1,2} = \frac{1}{2}W_1 - \Delta$, the low-type saves $\Delta[V_0 - V_m] > 0$. Gains arise because the low income investor liquidates risky holdings at cost $\alpha_0\Delta\frac{V_{1,R}}{V_0} < \alpha_0\Delta$, thereby implementing an implicit share transfer from the ex-post high type. The high income investor is indifferent to this transfer at prevailing prices.

Discussion In this example, the converted payment is in the form of an I.O.U. in the amount of the withdrawal $\alpha_0\Delta$, to be paid when the output is realized. Residual claimants are entitled to the output *net* of the I.O.U., and the residual claimant's position is commensurately *more concentrated*. The change in exposure for investor I_1 is $0.5b_0[b_0]^{-1} \mapsto b_0[2b_0 - \alpha_0\Delta]^{-1} > 0.5b_0[b_0]^{-1}$. Passive rebalancing works when *bank assets are risky* by allowing exercised cash options to leverage residual exposures.

Policies in the $m = G$ Case In this example, the low-type can liquidate the risky claim at a better rate on the market directly when net capital gains are positive. The high type cannot acquire assets through the intermediary balance sheet unless the low-types exercise cash options. As a result, bank exposures remain symmetric, and rebalancing is carried out in securities markets.[5]

Put Option on Bank Assets The ex-post transfer $\Delta[V_0 - V_m]$ can be written ex-ante as a put option on the bank's assets

$$\tau = [K^* - S_1]^+ 1_{\{n_1 = -\Delta\}}\Delta$$

The liabilities embed an option that will only be exercised by investors with negative idiosyncratic shocks in bad times, when net capital gains are negative. Liquidity is created by allowing bank financiers to lock-in the ex-ante share price, through the strike $K = V_0$, as a contingency for the event that an ex-post liquidation is needed when market prices are low.

Remark For an ϵ-fee on deposits, investors will finance the bank by exactly the amount they will need to withdraw in the bad aggregate state. The reason is that, in the good state, the low type would prefer to liquidate risk claims on the market because the implied share price of her bank deposit is lower.

0.3 Organization

Complete markets, incomplete markets and intermediated incomplete markets versions of the model are developed for contrast. Section 1 details the resources, participants and market

[5] In a more general setting, in good times, the ex-post high type can exercise a call option on bank assets by supplying new cash funding for the bank. This policy leaves the low-type indifferent given their exposure to the risky assets is diluted proportionally.

arrangements that serve as a benchmark to each of the versions we develop. Section 1.4 specializes to the economy with financial intermediation and derives equilibrium allocations and prices. The complete and incomplete markets environments are specialized and solved in A.0. Section 2 details comparative implications. Proofs based on standard arguments are relegated to Appendix A. The final section 4 discusses a handful of applications, including policy implications.

1 The Model

1.1 Environment

There is a continuum $\mathcal{I} := [0,1]$ of ex-ante identical investors. Investors have log utility over terminal wealth and are endowed with equal claims e_0 to terminal output Y. Y can take four possible values $Y \in \{Y_{GA}, Y_{GB}, Y_{RA}, Y_{RB}\} =: \mathcal{Y}$, where $Y_{GA} > Y_{RA}$ and $Y_{GB} > Y_{RB}$. Uncertainty is resolved over two periods. In the first period, a public signal indicates growth G or recession R. In the final period, realized output Y will will be above or below market expectations. For example, Y_{GB} corresponds to below expected output in the growth regime. Aggregate states are denoted s_m for $t = 1$ and $s_{m,k} \in S$ for $t = T$ with indices $m = R, G$; $k = A, B$. We write $Y(s_{m,k}) =: Y_{m,k} = Y$. Probabilities are mutually independent, $\Pr(s_{m,k}) = \pi_{m,k} = \pi_m \pi_k$, with marginals $\Pr(s_m) = \pi_m$ and $\Pr(s_{m,k}|m = R) + \Pr(s_{m,k}|m = G) = \pi_k$.

Each investor is also endowed with a nontradeable claim to income n_j distributed according to

$$n_j = \begin{cases} \Delta_0 + \Delta & \text{with } \Pr(n_j = \Delta_0 + \Delta) = \frac{1}{2} \\ \Delta_0 - \Delta & \text{otherwise} \end{cases}$$

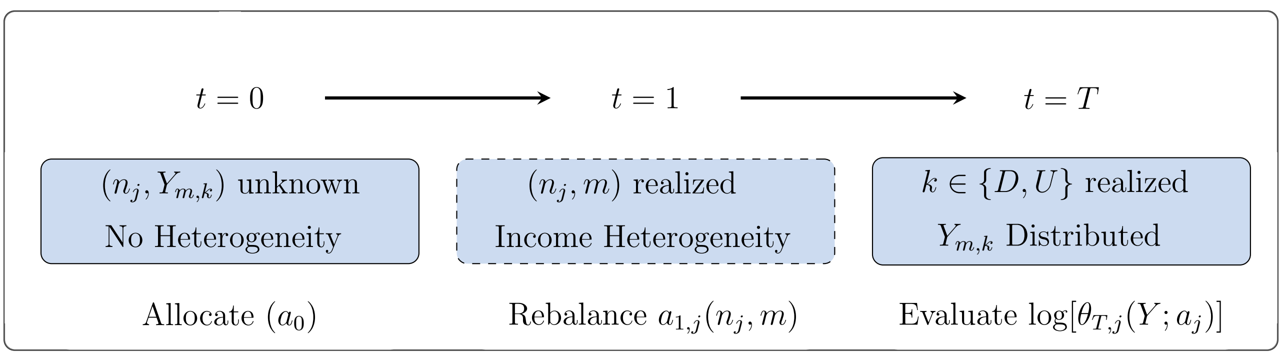
with $|\Delta| < \Delta_0$. The random variable n_j is revealed in the intermediate period. n_j is independent of the aggregate signal $m \in \{R, G\}$ and i.i.d. in the cross-section. Joint probabilities for $(s_{m,k}, n_j)$ are $\pi_{m,k,n_j} = \frac{1}{2}\pi_{m,k}$. There is no aggregate income risk. Total income is $Y_{m,k,0} := \Delta_0 + Y_{m,k}$ in every state. The resolution of uncertainty is depicted by a binomial tree in Figure 3.

An equity claim on output Y is traded at $t = 0$ and $t = 1$. Equity shares are fixed at one. Equity prices V_0 and $V_{1,m}$ are determined equilibrium. Time-zero share purchases in excess of the endowment are written a_0. After time-zero, share adjustments are denoted a_j. We write *gross* positions as a proportion of individual wealth α_0 and α_j. Allocations a, α are functions of time, individual wealth and the aggregate state.[6].

Finally, investors are endowed with equal deterministic amounts of a durable numeraire, "cash," written $\omega_0 > 0$. A riskless storage technology in infinitely elastic supply yields gross return normalized to $R = 1$. The securities markets open in response to news, as depicted in Figure 1.

Expenditures at time-zero are constrained by *tradeable* wealth $W_0 = \omega_0 + e_0$ where $e_0 = \mathbb{E}^Q[Y]$. The risk-neutral measure Q is determined in equilibrium. Investors can adjust their initial positions e_0 in the risky claim through choice of a_0 subject to $a_0 V_0 - e_0 \leq \omega_0$. Investors will forego trade if it is optimal.

Figure 2: Timing for the Economy without Intermediation



(a) For each t, uncertainty is described in the large middle rectangle, with available actions listed directly beneath. Investors choose initial market positions a_0. Economic conditions and the income distribution are realized in $t = 1$. Investors rebalance market positions $a_{1,j}(n_j, m)$ by type j at the corresponding market prices. At $t = T$, utility $u[\theta_{T,j}(Y_{m,k}; a_{1,j}(n_j, m))]$ is evaluated based on the realized output $Y_{m,k}$ and prior policies $a_{1,j}(n_j, m)$.

[6]The shorthand $a_0 = a(0, \cdot, \cdot)$ emphasizes initially identical policies, while $a_j = a(1, n_j, m)$ emphasizes the type-j dependence of allocations made at $t = 1$ given signal m. The same applies to α_0, α_j.

1.2 Equilibrium

1.2.1 Sequences

Write the investor's initial endowments $\boldsymbol{q}_0 = (V_0, \omega_0)'$ and define $q_0 = \boldsymbol{q}_0' \cdot \mathbf{1}$. The condition $e_0 = V_0$ gives $q_0 = W_0$, the initial level of tradeable wealth. We summarize each investor's final cash position $\omega_{j,m} := \omega_0 - a_0 V_0 + n_{1,j} - a_j V_{1,m}$, and the corresponding equity position $A_j := (1 + a_0 + a_j)$. Write $\theta_{T,j}$ for the period-T gross return on one dollar invested at time-zero for ex-post type j. $\theta_{T,j}$ is cum-income. It includes the income in $t = 1$, n_j, and the associated gain or loss on that investment between $t = 1$ and $t = T$.

Individual optimization problems are

$$\begin{aligned} J(q_0; V_0) \;&=\; \max_{\boldsymbol{a}} \mathbb{E}\left[\log(\theta_{T,j})\right] \qquad (1.A) \\ \text{s.t.} \quad & a_0 V_0 - V_0 \leq \omega_0 \\ & a_j V_{1,m} - (1 + a_0) V_{1,m} \leq (\omega_0 - a_0 V_0) + n_{1,j} \\ & \theta_{T,j} = A_j Y_{m,k} + \omega_{j,m} \end{aligned}$$

where $\boldsymbol{a} := (a_0, \{a_j\}_j)$. The second inequality is reproduced for every $(n_{1,j}, m) \in \{n_{1,1}, n_{1,2}\} \times \{R, G\}$ and the final equality for each $k \in \{D, U\} | (n_{i,j}, m)$. Expectations are with respect to the joint distribution of productivity, prices and income $(\mathcal{Y}, V, \{n_j\}_j)$.

The recursive analogue 1.B to the objective 1.A is provided in section 6.0.1.

Normalization We normalize $W_0 \equiv 1$ without loss of generality. Investors with log utility over final wealth care only about single-period gross returns. Aggregate wealth is not a state variable, although market incompleteness requires that individual wealth $q_{t,j} = q_{t,j} W_0^{-1}$ is a state variable for every individual. Standard arguments based on homothetic preferences, given in section 6.0.1 of Appendix A, justify this choice of state vector.

1.2.2 Resources

Put $\overline{Y}_{m,k} := Y_{m,k,0} + \omega_0$. In what follows, we distinguish between expectations over idiosyncratic income states and ex-post aggregation across population types by writing $\mathbb{E}_{n_j}$ and

$\sum_j \pi_j$, respectively.

Equilibrium The equilibrium is determined when every trader optimizes 1.A or 1.B and the following markets clear:

$$\begin{aligned} &\sum_j a_{0,j} = 0 && && \text{(1.C)} \\ &\frac{1}{2}\sum_j \alpha_j \tilde{W}_{1,j} = V_{1,m} && m \in \{R, G\} \\ &\frac{1}{2}\sum_j \theta_{2,j} = \overline{Y}_{m,k} && k \in \{U, D\}|m \end{aligned}$$

where we have applied $\pi_j = \frac{1}{2}$. The first two lines ensure securities markets clear at times $t = 0, 1$. The first line is simply $a_0 = 0$, or equivalently, $e_0 = V_0$, implying $W_0 = \omega_0 + V_0$. The second line states total shares are fixed at $1 = \frac{1}{2}\sum_j \overline{a}_j = \frac{1}{2}\sum_j \frac{\alpha_j W_{1,j}}{V_{1,m}}$. The third line is the final accounting for resources in terms of goods supply $Y_{m,k}$ and cash ω_0. When $k \in \{A, B\}$ is realized, output $Y_{m,k}$ is distributed according to the equity holdings A_j. Dependence on m is suppressed in some notation, e.g., $W_{1,j} = W_{1,j,m}$.

1.3 Discussion

Heterogeneous Wealth To highlight the rebalancing policies $a_j = a_j(n_{1,j}, m)$, we write the beginning of period wealth $W^-_{1,j}$ for each ex-post type j,

$$\begin{aligned} W^-_{1,j} &= \underbrace{W_0\,\alpha_0\,\frac{V_{1,m}}{V_0}}_{\text{Equity Capital Gain}} + \underbrace{W_0\,(1-\alpha_0)}_{\text{Storage}} + \underbrace{n_{1,j}}_{\text{Non-tradeable Income}} \\ &= W_1 + n_{1,j} \end{aligned}$$

In the first line, the first two terms are identical for every investor, suggesting we take $R_0 := R_{0,j} - n_{1,j}$ and write $W_1 = W_0 R_0$, giving the second line.[7] Now, write the end of

[7] $\alpha_0 W_0 = (1 + a_0)V_0$ relates the fraction of equity a_0 to the fraction of wealth invested in equity α_0. Therefore $W_0(1 - \alpha_0) = \omega_0 - a_0 V_0$ is the total cash position. Policies with subscript 0 are identical across investors.

period wealth $W_{1,j}^{+}$, for each ex-post type j

$$W_{1,j}^{+} = \underbrace{a_j V_{1,m} + (1 + a_0)\, V_{1,m}}_{\text{Equity Position}} \; + \; \underbrace{\omega_0 - a_0 V_0 + n_{1,j} - a_j V_{1,m}}_{\text{Cash Position}}$$

$W_{1,j}^{+}$ captures the composition of the investor's *outgoing* portfolio, expressed in terms of share policies a_j.[8] The first term is the value of their equity position after rebalancing at market prices $V_{1,m}$, and the second term is the value of the cash position. [9]

Securities Markets Take $j : n_j = \Delta > 0$ and $i \neq j : n_i = -\Delta < 0$. Clearly $a_j^* > 0 > a_i^*$. From the incoming portfolios $W_{1,j}^{-} > W_{1,i}^{-}$, together with the identical initial positions a_0, we see that $a_0 V_{1,m}/W_{1,j}^{-} < a_0 V_{1,m}/W_{1,i}^{-}$. Type j has a relatively lower *incoming* equity investment *rate* than type i. Since $W_{1,h}^{+} = W_{1,h}^{-}$ for any h, rebalancing operates entirely through a_h. With log preferences, the policies $a_j^* > 0 > a_i^*$ are chosen to equalize the equity investment rates $\alpha_j = \alpha_i$ across types.

1.3.1 Benchmark Implications

In Appendix A.0, we solve the model with log utility the for complete and incomplete markets cases and detail the implications. Complete markets naturally produce a degenerate wealth distribution and a representative agent pricing kernel. Incomplete markets imply the ex-post wealth distribution is bimodal with support that varies with the aggregate state. The corresponding pricing kernel takes the form of the complete markets kernel scaled multiplicatively by a term accounting for the wealth distribution. Details of the incomplete markets asset prices are reproduced alongside the intermediated-economy asset prices, in section 2.2. The intermediary economy is formalized in section 1.4.

[8]Policies $a_0 = \alpha_0 W_0 V_0^{-1} - 1 > 0$ correspond to an increase in the investor's risky position at time zero, $a_0 < 0$ represents a decrease, and $a_0 = 0$ gives the time-zero no-trade allocation.

[9]Recall that policies $a_j > 0$ represent an increase in equity levels while policies $a_j < 0$ indicate a decrease.

1.4 Intermediated Markets

1.4.1 Technology

The bank allows investors to deposit a fraction of their tradeable endowment e_0, denoted in levels by $b_0 \leq e_0$, in exchange for a claim to bank assets. The claim embeds the option to convert any amount $k_j = k(n_j, b_0)$, up to the cash value of the deposit $k_j \leq b_0$, into a certain payment. Total bank financing aggregates b_0 over investors and is written $\mathbf{b}$. The financing level $\mathbf{b}$ acquires $\frac{\mathbf{b}}{V_0}$ shares of the risky claim for intermediation. The remaining shares $\frac{1}{V_0}(V_0 - \mathbf{b})$ are held directly by investors. In each period, investors can access both the bank and the market directly. At time-zero, bank investors bear identical exposures to the bank asset risk.

Ex-post, an individual financier i has the option to respond to n_i via the rollover policy $k(n_i, b_0)$. However, the rollover policies *en-masse* $\{k_j\}_{j \in \mathcal{I}}$ control the risk composition of the residual claim to bank assets. Thus, an optimal funding policy can only be evaluated given an assessment of the aggregate effect of all funding policies on the residual capital.

Bank Capital For total withdrawals $\boldsymbol{\kappa} := \sum_j k_j \pi_j$, an infinitesimal investor j_0 choosing $k_{j_0} \in [0, b_0]$ acquires the residual exposure

$$\mathbf{k}^+(k_{j_0}, \boldsymbol{\kappa}) = \frac{b_0 - k_{j_0}}{\mathbf{b} - \boldsymbol{\kappa}} \tag{K.1}$$

The quantity $\mathbf{k}^+$ reports the fraction of total claims to bank assets. An investment of one dollar at time zero corresponds to a $\mathbf{b}^{-1}$ ownership stake in the bank. Ex-post, with no withdrawal by the dollar investor and an economy-wide withdrawal of $\boldsymbol{\kappa}$, the dollar investment corresponds to a $[\mathbf{b} - \boldsymbol{\kappa}]^{-1} > \mathbf{b}^{-1}$ ownership stake. For initial investment b_0 and ex-post policy k_{j_0}, we obtain K.1. From the stake $\mathbf{k}^+$, we can calculate its market value at time $t = 1$, $\mathbf{k}^+(k_{j_0}, \boldsymbol{\kappa})\, V_{1,m} \frac{\mathbf{b}}{V_0}$ and the corresponding ownership stake in equity $\mathbf{k}^+(k_{j_0}, \boldsymbol{\kappa}) \frac{\mathbf{b}}{V_0}$.

Payments to residual claimants are net of the bank's obligatory payments $\boldsymbol{\kappa}$, which are also deducted pro-rata. We define the dividend paid to a marginal unit of bank capital $D_k := [\frac{\mathbf{b}}{V_0} Y_{m,k} - \boldsymbol{\kappa}]$. Naturally, the residual position $\mathbf{k}^+$ entitles an owner to the cash flows

$\mathbf{k}^+ D_k$ in each final state $k \in \{B, A\}$.[10]

An individual financier i's rollover policy k_i therefore depends on the ex-post bank capital structure in addition to the income realization n_i. We occasionally write $k_j = k(n_j, b_0, \boldsymbol{\kappa})$ to emphasize this dependence. Optimal policies and the ex-post population frequencies $\pi_j\, j = 1, 2$ are common knowledge. An individual investor has no market power. Her optimal rollover policy takes the ex-post capital structure $(\boldsymbol{\kappa}, \mathbf{b})$ as given, where $\boldsymbol{\kappa} = \boldsymbol{\kappa}(m)$. The resolution of uncertainty is illustrated in Figure 2, along with the timing of allocation decisions.

1.4.2 Preferences

In the appendix A.7.1 we discuss a nonseparable preference specification for the banking model. A simple modification to the endowment to include a proportional dividend allows us to define preferences over consumption streams, which is a more natural specification for recursive utility. Below, we continue our analysis with logarithmic preferences.

1.4.3 Equilibrium

We invoke the result that $a_0 \equiv 0$. After incorporating the bank technology, every investor's time-zero objective can be written

$$
\begin{aligned}
J(q_0; V_0) \;&=\; \max_{\boldsymbol{a}, b} \mathbb{E}\left[\log(\theta_{T,j})\right] \qquad \text{(B.1)}\\
\text{s.t.} \quad & b_0 \le V_0 \\
& a_j V_{1,m} - \left[1 - \frac{b_0}{V_0} + \mathbf{k}^+(k_j, \boldsymbol{\kappa})\,\frac{\mathbf{b}}{V_0}\right] V_{1,m} \le \omega_0 + n_j + k(n_j, b_0, \boldsymbol{\kappa}) \\
& k(n_j, b_0, \boldsymbol{\kappa}) \;\le\; b_0 \\
& \theta_{T,j} = A_{j,m}(\mathbf{b})\, Y_{m,k} + \mathbf{k}^+(k_j, \boldsymbol{\kappa})\, D_k + \omega_{j,m} + k_j
\end{aligned}
$$

where $A_{j,m}(\mathbf{b}) := [1 - \frac{\mathbf{b}}{V_0} + a_j]$ is the fraction of equity held directly by investor j.

[10] A k without index or argument always refers to final stage uncertainty $k \in \{B, A\}$, while indexed or functional $k_j = k(n_j, b_0, \boldsymbol{\kappa})$ are always contingent rollover policies.

Resources In the banking economy, market clearing conditions 1.C interact with the definitions of b_0, $\mathbf{b}$ and k_j because ownership of the equity claim is partially intermediated.[11] We restate the market clearing conditions below and discuss the role of bank variables in equilibration. Having already imposed $a_0 \equiv 0$[12], we can write

$$\begin{aligned} \sum_j a_j &= 0 \\ \frac{1}{2}\sum_j \theta_{2,j} &= \overline{Y}_{m,k} \qquad k|m \end{aligned} \tag{2.C}$$

The first term in 2.C says net time $t = 1$ modifications through the market a_j are zero. A key definition is worth restating

$$\frac{1}{[\mathbf{b} - \boldsymbol{\kappa}]}\frac{1}{2}\sum_j [b_0 - k_j] = 1 \tag{Re.1}$$

The resource constraint $\sum_j a_j = 0$, together with Re.1, is equivalent to enforcing that equity is in fixed supply with shares normalized to one. Re.1 also ensures the total output paid to the bank is $\frac{\mathbf{b}}{V_0} Y_{m,k}$, which, together with the share accounting for direct holdings $(1 - \frac{\mathbf{b}}{V_0} + \frac{1}{2}\sum_j a_j)$, ensures total output distributed is $Y_{m,k}$. Finally, from Re.1 and the definition of D_k, the total level of precedent payments owed by bank capital owners is $\boldsymbol{\kappa}$.

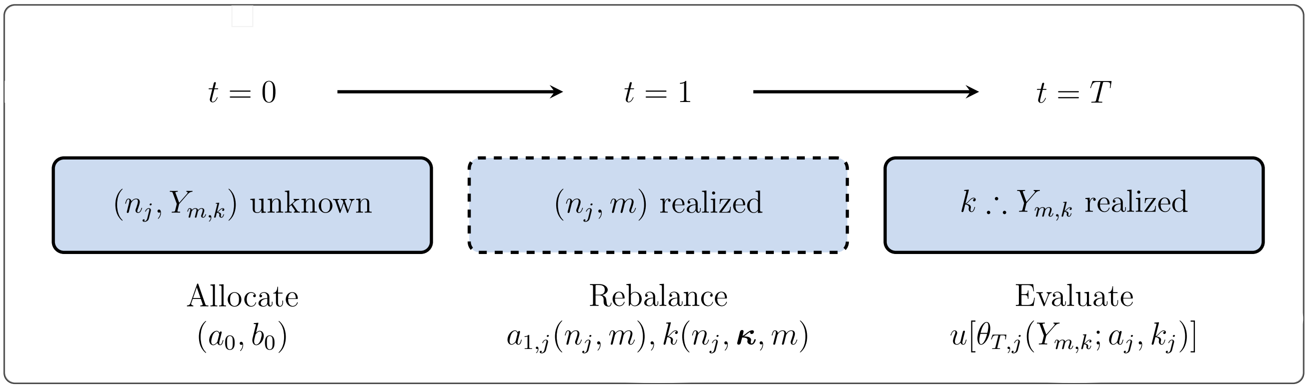
The second term in 2.C is an accounting of final payouts made to individuals. It is identical to the third term in 1.C with the exception that $\theta_{2,j}$, given in B.1, is a function of bank policies $k_j, \{k_{-j}\}_{-j \in \mathcal{J}}$. The first term in the original clearing list1.C is subsumed by the fact that $a_0 \equiv 0$ and the definition $\mathbf{b} = \frac{1}{2}\sum_j b_0$.

Equilibrium with Intermediation *An equilibrium with intermediary financing is a set of allocation policies a_j, b_0, k_j and prices $V_0, V_{t,m}$ such that every investor optimizes B.1, markets clear according to 2.C, and the policies $k(n_j, b_0)$ and ex-post population frequencies π_1, π_2 are common knowledge.*

[11] The definitions of $\{b_0, \mathbf{b}, k_j, \boldsymbol{\kappa}\}$ along with market clearing criteria from the incomplete markets model, 1.C, are alone sufficient for a well-defined equilibrium.

[12] The equivalent conditions for wealth rates α_t attain via the obvious substitution $a_j = \alpha_t W_{t,j}/V_{t,m} - 1$.

Figure 3: Timeline for the Economy with Bank Financing



(a) For each period t the state of the economy is given in the dashed rectangle and the corresponding actions available are listed underneath. Initially, identical investors choose market positions and bank financing levels (a_0, b_0). In $t = 1$, economic conditions $m \in \{R, G\}$ and idiosyncratic incomes $\{n_j\}_{j \in \mathcal{I}}$ are realized. Investors rebalance market positions and bank financing allocations $a_j, k(n_j, \boldsymbol{\kappa}, m)$ according to their realized type j and the aggregate state. In the final period, utility $\log[\theta_{T,j}(Y_{m,k}, a_j, k_j(n_j, \boldsymbol{\kappa}, m))]$ is evaluated for the realized $Y_{m,k} = Y$.

2 Implications

We develop the portfolio and bank funding policy implications of our theory, along with the corresponding asset pricing implications. Implications for industrial organization in the financial sector are postponed to section 2.4. We first state some key results.

Proposition 2.1 (Bank Financing Equilibrium) *An equilibrium in the economy with bank liability production exists and exhibits the following properties*

1. *Trade in the initial period organizes the bank, leading to welfare gains*
2. *Ex-post policies implement rebalancing through the bank's balance sheet when market prices are low* $k(n_1, \boldsymbol{\kappa}(s_R)) = 0, k(n_2, \boldsymbol{\kappa}(s_R)) = \alpha_0 \Delta$*, and through securities markets when market prices are high* $k(n_1, \boldsymbol{\kappa}(s_G)) = k(n_2, \boldsymbol{\kappa}(s_G)) = 0$
3. *Prices in the intermediated economy can be written in terms of the Lucas kernel*

$$\mathbb{M}[(s_{m,k})]_{IAP} = \left[M[\overline{Y}_{m,k}]_{Lucas}\right] e^{-s_{m,k}\eta_s - s_0 \overline{s}_{m,k}\zeta_s}$$

4. *Relative to the incomplete markets benchmark, the distribution of subjective valuations is more dispersed but with lower mean, and the distribution of wealth is less dispersed*

2.1 Policies

Heterogeneous Wealth We revisit the wealth expressions from section 1.3 in the context of the banking economy. With incomplete markets, incoming positions are symmetric up to the shocks n_j. In the banking economy, the positions $b_0 = b_0(V_0)$ nest a put option on risky assets that separates ex-post investors by type in bad times. When exercised, the options induce portfolio heterogeneity by implementing the ex-post swap of cash for risky claims at prices set ex-ante.

We can write the incoming and outgoing expressions from 1.3 in terms of intermediary positions b_0 and rollover policies k_j

$$W^{-}_{1,j} = \underbrace{\overbrace{[V_0 - b_0]\frac{V_{1,m}}{V_0}}^{\text{Market Position}} + \overbrace{b_0\frac{V_{1,m}}{V_0}}^{\text{Bank Deposit}}}_{\text{Combined Capital Gain}} + \underbrace{\omega_0}_{\text{Cash}} + \underbrace{n_{1,j}}_{\text{Non-tradeable Income}} \tag{W.B.1}$$

$$W^{+}_{1,j} = \underbrace{[a_j + 1]\, V_{1,m}}_{\text{Market Position}} + \underbrace{\frac{\mathbf{b}}{V_0}\,\mathbf{k}^{+}(k_j,\boldsymbol{\kappa})\, V_{1,m}}_{\text{Bank Capital}} + \underbrace{\omega_0 + n_{1,j} - a_j V_{1,m} + k_j - \boldsymbol{\kappa}\,\mathbf{k}^{+}}_{\text{Cash Position}}$$

where the residual claim is $\mathbf{k}^{+}(k_j,\boldsymbol{\kappa}) = \frac{b_0 - k_j}{\mathbf{b}-\boldsymbol{\kappa}}$ defined in K.1.

Rollover Policies For a withdrawal k_j, the investor rolls-over a fraction $b_0^{-1}[b_0 - k_j]$ of her initial stake in the bank. The level k_j augments the *Cash Position* of the investor's portfolio, illustrated in $W^{+}_{1,j}$ of W.B.1. The corresponding increase in exposure to asset risk is consolidated in the *Bank Capital* ledger from the same expression, and is written $\frac{b_0 - k_j}{\mathbf{b}-\boldsymbol{\kappa}}\frac{\mathbf{b}}{V_0}$, or equivalently, $\mathbf{k}^{+}\frac{\mathbf{b}}{V_0}$. The residual claims to bank assets $\mathbf{k}^{+}(k_j,\boldsymbol{\kappa})$ are by definition the bank capital.[13] For any withdrawal policy short of full divestment $k_j < b_0$, a third portfolio implication accounts for the precedent payments $\boldsymbol{\kappa}$ nested in the bank dividend D_k. These payments are recorded as the final entry in the *Cash Position* ledger, written $-\boldsymbol{\kappa}\,\mathbf{k}^{+}$, as a liability corresponding to the rollover policy k_j. Each of the three portfolio components corresponding to k_j are illustrated in W.B.1 via $W^{+}_{1,j}$.

[13] We use the term *bank capital* to refer to the market value of equity for the bank, consistent with the nomenclature in banking. In this model, equity is only defined implicitly ex-ante, but following the aggregate withdrawals, we unambiguously refer to the residual claims as bank capital.

Remark Investors implement the optimal allocation policy $\alpha_j = \alpha_i$ through a combination of intermediary leverage and direct trade in securities markets. In equilibrium, $\alpha_{1,h} = \alpha_1 = V_{1,m}\,[V_{1,m} + C]^{-1}$ for each $h \in \{1, 2\}$ as in the incomplete markets economy. However, the corresponding price levels and wealth shares differ. The implementation of $\alpha_i = \alpha_j$ is ex-post less costly for high marginal utility types, and therefore ex-ante expected to be less costly for every investor.

2.2 Asset Prices

We present and discuss asset pricing implications from the incomplete markets economy and the intermediated economy. Proofs and background details are given in Appendix A.6.0.

Time-zero We can express the time-zero incomplete markets NC pricing kernel in terms of the expected wealth distribution. One-period asset prices in the complete and incomplete markets economies are described by

$$M[\overline{Y}_m]_{\text{Lucas}} = [\nu_0]^{-1}\, W_{1,m}^{-1}\pi_m \tag{L.0}$$

$$\begin{aligned} M[\theta_{1,h}(s_m)]_{\text{NC}} &= [\nu_0]^{-1}\,[\theta_{1,j}^{-1} + \theta_{1,-j}^{-1}]\frac{1}{2}\,\pi_m \\ &= [\nu_0]^{-1}\, W_{1,m}[(W_{1,m} + \Delta)(W_{1,m} - \Delta)]^{-1}\,\pi_m \end{aligned} \tag{NC}$$

where $W_{1,m} = V_{1,m} + \omega_0$. The proof, given in section 6.2.4, uses the $t = 1$ marginal value of wealth $[\partial J_{1,j}]^{-1} = \theta_{1,j}$ for each j, that are consistent with backward induction from the final-period shares I.θ. The shares simultaneously obey the time-zero Euler equations.[14]

Oversaving The NC kernel includes a component correcting for the distribution of wealth that strictly raises state prices relative to the complete markets benchmark. The additional component vanishes as income transfers become negligible

$$[(W_{1,m} + \Delta)(W_{1,m} - \Delta)]^{-1} - W_{1,m}^{-2} \;>\; 0 \qquad \Delta > 0$$

$$\lim_{\Delta \searrow 0} W_{1,m}[(W_{1,m} + \Delta)(W_{1,m} - \Delta)]^{-1} = W_{1,m}^{-1}$$

[14] Time-consistency for incomplete markets, in the sense of Marcet and Marimon (1997),(2012), requires ex-post optimal policies agree with the policies that made ex-ante allocations optimal.

NC reduces to L.0 with $\Delta \searrow 0$. NC accommodates higher demand for savings when some states of the world are uninsurable. [15,16]

Time-one Market prices at time $t = 1$ are complete markets prices, but with heterogeneous investors. With log utility, the representative and heterogeneous-investor pricing kernels are equivalent

$$M[\overline{Y}_{m,k}|m]_{\text{Lucas}} = [\partial J_1]^{-1}\,\overline{Y}^{-1}_{m,k}\pi_{m,k} \tag{L.1}$$

$$\begin{aligned} M[\theta_{2,j}(s_{m,k})]_{\text{Hetero}} &= \frac{\theta_{2,j}^{-1}}{\partial J_{1,j}}\pi_{m,k} = \frac{\theta_{2,-j}^{-1}}{\partial J_{1,-j}}\pi_{m,k} \\ &= \frac{1}{2}[\partial J_{1,j}^{-1} + \partial J_{1,-j}^{-1}]\overline{Y}^{-1}_{m,k}\pi_{m,k} \\ &= [\partial J_1]^{-1}\overline{Y}^{-1}_{m,k}\pi_{m,k} \end{aligned} \tag{H}$$

at time $t = 1$, for each $s_{m,k}$.

Log utility gives $[\partial J_{1,j}]^{-1} + [\partial J_{1,-j}]^{-1} = 2[\partial J_1]^{-1}$ with $[\partial J_{1,j}]^{-1} = W_{1,m} + n_{1,j}$ and $[\partial J_1]^{-1} = W_{1,m}$. The first line for M_{Hetero} equates valuations by type. The second line uses the wealth shares $\theta_{2,j}$ and represents the kernel by averaging $\partial J_{1,j}$ and $\theta_{2,j}^{-1}$ over types and then normalizing. The third line is true for any aggregation rule, e.g., averaging after normalization.[17]

Lucas Exchange Model The incomplete markets pricing kernel can be written in terms of the Lucas kernel and a multiplicative term reflecting imperfect risk sharing in the cross-section of the investing population. Write $\sqrt{\sigma_\Delta}(s_m) := \Delta/W_{1,m}(s_m)$. We express the incomplete kernel NC in terms of L.0:

$$\begin{aligned} M[\theta_{1,h}(s_m)]_{\text{NC}} &= \left[M[\overline{Y}_m]_{\text{Lucas}}\right]\; e^{-\log[(1+\Delta/W_{1,m})(1-\Delta/W_{1,m})](s_m)} \\ &= \left[M[\overline{Y}_m]_{\text{Lucas}}\right]\; e^{\sigma_\Delta(s_m)(1+\frac{1}{2}\sigma_\Delta(s_m))-o(\frac{1}{\Delta})} \end{aligned}$$

[15] Over-saving in incomplete markets with convex marginal utility is well studied, see e.g., Weil (1989) and Mankiw (1986).

[16] Demand for savings grows proportionally with the fraction of net-worth that is nontradeable. State prices rise in equilibrium because the market cost of postponing consumption must offset its increased demand.

[17] The two aggregation rules produce identical pricing kernels because $[\partial J_{1,j}]^{-1} + [\partial J_{1,-j}]^{-1} = [\partial J_{1,j} + \partial J_{1,-j}]\,(\partial J_{1,j}\partial J_{1,-j})^{-1}$.

The rate $\sigma_\Delta(s_m)(1+\frac{1}{2}\sigma_\Delta(s_m))$ summarizes the distributional risks for $\Delta > 0$ up to a fourth-order expansion of $\log(1+\sqrt{\sigma_\Delta}(s_m)) + \log(1-\sqrt{\sigma_\Delta}(s_m))$. Exposure to higher moments of the wealth distribution is priced in the incomplete economy, even with myopic investors. Distributional risk becomes negligible as $\Delta W_{1,m}^{-1} \searrow 0$.

2.2.1 Remarks

I. $V_{1,m}$ can be reconstructed $V_{1,m} = \sum_{k\in\{D,U\}} M[\overline{Y}_{m,k}]_{\text{Lucas}}\, Y_{m,k}\, \pi_k$ for each $m \in \{R, G\}$. Because $W_{1,m} = V_{1,m} + \omega_0$, the time-zero prices for payouts at maturity obtain by plugging $V_{1,m}$ into NC, or L.0 for the complete-markets case. The state price kernels $M[\,\cdot\,]$ are a more flexible description of the economy than $V_{1,m}$ in part because, with log utility, the discount rate on a claim to aggregate consumption reduces to the subjective rate of time preference.[18,19]

II. In incomplete markets, investors with a negative shock are compelled to "take the hit," by renormalizing their marginal utility growth rates to be in line with market prices, but at higher individual marginal utility levels. Investors with low private income become poor relative to expectations. Proportionally, the loss of net worth is larger when aggregate productivity is low, holding the level $|\Delta|$ fixed. In the incomplete markets equilibrium, the cross-sectional dispersion of subjective valuations is countercyclical.

2.3 Asset Prices with Intermediation

Every investor's time-zero allocations are identical. Ex-post, each type $j = 1, 2$ trades-off the market and the intermediary differently depending on the state of the economy. Consider

[18] In the finite horizon model without intermediate consumption, we set $1 + \beta = 1$.

[19] In an endowment economy with log utility defined over a perishable numeraire c, risk premia on the aggregate claim collapse with a representative agent because $c(s_t)d\log(c(s_t)) = dc(s_t)$ state-by-state. The same would be true in an economy with a durable numeraire and utility defined over wealth, such as ours, if cash was in zero net supply. Notably, in either case, the incomplete markets pricing kernel retains the distributional term when pricing the aggregate claim.

the Euler equations for the low-type $j = 2$, i.e., $n_2 - \Delta_0 = -\Delta$. During a recession $m = R$,

$$\partial J_{1,j} \frac{V_R}{V_0} - \mathbb{E}\left[\theta_2(\mathbf{k}^+, k_j, s_{R,k}))^{-1} Y(s_{R,k})\right] = 0$$
$$\partial J_{1,j} V_R - \mathbb{E}\left[\theta_2(\mathbf{k}^+, k_j, s_{R,k}))^{-1} Y(s_{R,k})\right] \leq 0$$

where now $\partial J_{1,j} = [W_{1,j}(\mathbf{k}^+, k_j)]^{-1}$. Consider the high-type $j = 1$ in a recession,

$$-\partial J_{1,j} \frac{V_R}{V_0} + \mathbb{E}\left[\theta_1(\mathbf{k}^+, k_j, s_{R,k}))^{-1} Y(s_{R,k})\right] \geq 0$$
$$-\partial J_{1,j} V_R + \mathbb{E}\left[\theta_1(\mathbf{k}^+, k_j, s_{R,k}))^{-1} Y(s_{R,k})\right] = 0$$

whose incentives for trade in each institution are the complement of the low type. The high income investor needs to acquire more risky claims and in a recession they are cheaper on the market. In contrast, the low type must liquidate the claims, and can turn each share into more cash by pulling bank funds. The caveat is that the high type must also not withdraw, but she will never withdraw given her portfolio needs, unless limited liability is jeopardized.

Recall that $\mathbf{k}^+$ is the equilibrating variable for aggregating bank policies. Equating the marginal conditions from the two types gives

$$\frac{\mathbb{E}\left[\theta_1(\mathbf{k}^+, k_1, s_{R,k}))^{-1} Y(s_{R,k})\right]}{\partial J_{1,1}} = V_0 \frac{\mathbb{E}\left[\theta_2(\mathbf{k}^+, k_2, s_{R,k}))^{-1} Y(s_{R,k})\right]}{\partial J_{1,2}} \qquad \text{(Eq.1)}$$

Equation Eq.1 says that market forces relax the imposition on low types that their marginal utility growth equal that of the high type, given their unexpectedly high marginal utility level today. The Euler equation ensures the low type is not raising marginal utility today to make this happen, so the scale factor corresponds to lower marginal utility of wealth tomorrow, i.e., higher ex-post low-type wealth. The relaxation factor is the scale $V_0 < 1$.

The single-period pricing kernel can be written in terms of the Lucas kernel as in the incomplete markets economy. The kernels are

$$\mathbb{M}[(s_m)]_{\text{IAP}} = [M[(s_m)]_{\text{Lucas}}]\, e^{\sigma_\Delta \left[1 - o(\frac{1}{\Delta})\right](s_m, \mathbf{k}^+)}$$

The exponential terms parametrize an equivalent change of measure that distinctly characterizes the asset pricing implications of our financial intermediary.

2.3.1 The Price of Risk

At time-zero, the economy optimally reorganizes to include intermediation. Relative to complete markets, price levels are lower in equilibrium because of a decrease in the over-saving propensity. In addition to the level effect, a covariance effect activates when nontradeable income *levels* are conditionally independent of aggregate output in the time-series.[20] We model nontradeable income additively which, in conjunction with independence from the aggregate state, satisfies this criterion. Other possibly dependent specifications can also be tailored to violate Krueger and Lustig (2010).

To illustrate, consider $n_L < \mathbb{E}[n] < n_H$ and evaluate the excess Euler equations prior to trade. These are equalities at the allocations in expectation. The thought experiment is to consider the effect of news about private income only

$$
\begin{aligned}
-[\partial J_{L,t}]^{-1}\mathbb{E}[\partial J_{L,t+1}]r_f + [\partial J_{L,t}]^{-1}\mathbb{E}\left[\partial J_{L,t+1}\frac{Y_{t+1}}{P_t}\right] &< 0 \\
-[\partial J_{H,t}]^{-1}\mathbb{E}[\partial J_{H,t+1}]r_f + [\partial J_{H,t}]^{-1}\mathbb{E}\left[\partial J_{H,t+1}\frac{Y_{t+1}}{P_t}\right] &> 0
\end{aligned}
$$

With the additive private income specification, convex marginal utility induces not simply rebalancing incentives from changes in wealth, but also relative over-and-under valued types of assets delineated by risk. In contrast, consider a private income process N_j that scales output Y to determine the income level $n_j = N_j Y$, subject to the appropriate goods clearing protocol. The same thought experiment generates the above Euler equations that are still equalized after realizing n. The reason is that with homogeneous preferences, multiplicative income factors out just like aggregate wealth, so that marginal effects from income shocks are constant across asset types.

[20] This violates the criteria for risk-indifference in Krueger and Lustig (2010).

2.3.2 The Wealth Distribution

We analyse the ex-post wealth distribution in the banking economy against benchmark complete and incomplete markets wealth distributions. Naturally, the complete markets ex-post wealth distribution is degenerate with all mass located at the level of realized output $Y(s_{m,k})$ in each state $s_{m,k}$. A contrasting limit is given by the incomplete markets model, where the wealth distribution is bimodal in each state of aggregate cash flows. The distribution has two atoms of equal mass separated by length 2Δ for each $s_{m,k}$, but the absolute horizontal position of the atoms moves with $Y(s_{m,k})$.

2.4 Organizational Implications

We state a handful of implications for industrial organization in the financial sector implied by the benchmark case of our theory.

Corollary 2.2 (Banks are always Capitalized) $b_0 > 0$ *and* $\mathbb{P}(\omega(j) = b_0,\ j = 1, 2) < 1$ *Investors allocate strictly positive wealth levels to bank financing, and claim the bank's residual assets with strictly positive probability.*

Corollary 3.1 follows from proposition 2.3. Investors always allocate strictly positive wealth to bank formation initially, $b_0 > 0$. Ex-ante, each investor places positive probability on the event: "retain exposure" to bank assets. Ex-post, the population mass $\pi_1 > 0$ holds bank capital optimally, in bad times. The bank is always formed ex-ante, and always capitalized ex-post.

Remark This result fails for some changes of model assumptions. If the distribution of private income is not deterministic and $\mathbb{P}(\pi_1 = 0) > 0$, the bank may not be capitalized ex-post. If the distribution of aggregate output incorporates disasters, i.e., realized output of $Y_{\min} = \epsilon > 0$ has $\mathbb{P}(Y_{m,k} = Y_{\min}) > 0$, the bank may not be formed ex-ante.

Corollary 2.3 (No Market Segmentation) *Banks coexist with markets: the demand for risky asset intermediation does not rely on restricted access to markets.*

2.3 follows from corollary 2.2. Investors demand intermediation because risky assets are a necessary input for liquidity production.

Corollary 2.4 (Variation in the Investment Opportunity Set) *The demand for financial intermediation is distinct from hedging demand. Myopic investment policies that are indifferent to shocks that impact future productivity nonetheless finance the bank.*

Corollary 3.3 follows immediately from log preferences. Myopic investors prefer one-step ahead contingency plans to mitigate liquidation expenses.

Random Capital Structure The bank has a *random capital structure* in the sense that at any time t assets A_t are financed by a combination of debt and capital that is not known until $t+1$ when population rollover policies are observed and exposures net of liabilities can be computed. The residual claims are risky not only because the cash flows generated by assets are risky but also because the amount of debt financing that survives until residual payments are made is uncertain.

3 Conclusion

The formation of a bank improves welfare over an incomplete markets economy when a large risk-averse population faces uninsurable shocks to net worth. Narrow banking precludes this mechanism in its strictest form. Relatively risky assets facilitate the creation of liquidity and hence the impetus for the risk sharing through banks. Bank financing does not in general achieve the first-best allocation because bank liabilities do not allow digital contingencies to be assigned ex-ante. Several interesting implications emerge. We show banks are always capitalized under the assumptions of our economy, even for myopic investors. A corollary to this is that banks and financial markets always coexist in this economy, even when investors have unrestricted access to both. This prediction has evaded a long literature in corporate finance and is worthy of further scrutiny.

The principal thrust of the preceding investigation was to link dynamics of institutional capital structure to the preferences of investors making decisions on the margin who ultimately possess the wealth in the economy. The mechanism can be understood through a key intuitive ingredient that distinguishes a narrow class of institutions including banks from others: the *random capital structure*. Why is bank capital structure risky and why is this unique to banks? Within our stylized economy the answer is: bank capital structure is risky because of the production of liquid liabilities, and only institutions that produce liquidity in the same way exhibit the same patterns in equilibrium. Production of liquid liabilities is a function unique to banks and dealers, each of which contribute uniquely to our understanding of asset prices.

On any day, the assets on the bank's balance sheet are funded by a mix of equity and liabilities. But the mix is uncertain ex-ante. Before any obligatory or residual payments are made, liquidity holders can convert their investment to the numeraire corresponding to the time of their investment. When they do this, the component of the risky assets that their investment effectively financed is transferred pro-rata to the residual owners, who then become liable for the numeraire payment. Based on the investments today, residual claimants face ex-post leverage restrictions on their cash flow - the same asset has an effective leverage ratio for each state tomorrow in general. Bank and dealer balance sheets are unique in that, in addition to the cash flow risk from the assets, owners bear the risk embedded in the demand options of other liability holders.

The allocations and prices are characterized in several stylized economies for comparison. Asset prices reflect the quality of risk-sharing in the cross-section of the investment population. In each ex-post contingency, the wealth-weighted marginal valuations of investors determine the realized leverage ratio. As a result, changes in bank capitalization rates measure innovations to the average marginal value of wealth in the cross-section of investors. The risk sharing is improved over incomplete markets, which means a reduction in over-savings reflected in the stochastic discount factor.

The implications contribute to understanding the empirical successes of intermediary asset pricing tests. The model provides specific additional implications that can help reject this or other theoretical proposals. A discussion of the implications and evidence are discussed in the appendix. It is useful that the candidate explanation presented in this paper is an outcome in a general equilibrium with minimal primitive assumptions. Indispensable assumptions are few: investors are risk averse and experience uncertainty about the present value of their lifetime productivity, and the markets for the idiosyncratic shocks to valuations are not operational. We assume the income shocks are purely distributional, but the results do not depend on this assumption. The theoretical implications in this paper are transparently prone to rejection by new empirical tests, while exiting models rely on stark assumptions to replicate existing empirical evidence at the expense of generating productive testable implications.

References

[1] Tobias Adrian and Nina Boyarchenko, *Intermediary Leverage Cycles and Financial Stability*, Federal Reserve Bank of N.Y., Staff Report No. 567, August 2013

[2] Tobias Adrian, Erkko Etula and Tyler Muir *Financial Intermediaries and the Cross-Section of Asset Returns*, The Journal of Finance, Vol. LXIX, No. 6, Dec 2014

[3] Douglas W. Diamond and Philip H. Dybvig, *Bank Runs, Deposit Insurance, and Liquidity*, Journal of Political Economy, Vol. 91 No.3 1983

[4] He, Kelley and Manilla, *Intermediary Asset Pricing: New Evidence from Many Asset Classes*, Journal of Financial Economics, vol. 126, issue 1, 2017

[5] Zhiguo He, In Gu Khang and Arvind Krishnamurthy, *Balance Sheet Adjustments during the 2008 Crisis*, IMF Economic Review Vol. 58, No. 1, 2010

[6] Charles J. Jacklin, *Demand Equity and Deposit Insurance*, Stanford GSB Research Paper No.1062, 1987

[7] Joseph G. Haubrich and Robert G. King, *Banking and insurance*, Journal of Monetary Economics 26 (1990) 361-386

[8] Gary Gorton and Andrew Metrick, *Securitized Banking and the Run on Repo*, NBER and Yale SOM Working Paper, 2009

[9] Gary Gorton and Andrew Metrick, *Who Ran Repo?*, NBER and Yale SOM Working Paper, 2012

[10] Anil K. Kayshap, Raghuram Rajan and Jeremy C. Stein, *Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking*, Journal of Finance, Vol. LVII, No. 1, Feb. 2002

[11] Robert E. Lucas, *Liquidity and Interest Rates*, Journal of Economic Theory, No.50, 1990

Appendices

Contents

5 Benchmark Results

Here we state and discuss implications from the complete and incomplete markets benchmark models. We present the results necessary for the comparisons referenced in the body of the paper, including results on welfare, asset pricing and securities positions, as well as the ex-post wealth distributions. Some additional results not central to the comparisons above are listed along with the proofs in section A.1.

5.1 Complete Markets

To complete the markets, we include a perfectly enforced Arrow-Debreu contingent claim $a(s_{m,k}, j)$ for every distinct state of the economy $(s_{m,k}, j)$, $m = R, G$, $k = A, B$, $j = 1, 2$. Details of the trading technology are provided in section 6.0.1.

Proposition 5.1 (Complete Markets Benchmark) *In the complete markets economy with ex-ante identical investors and security menu S that spans consumption paths*

1. *The log-utility representative agent pricing kernel $M[(s_{m,k})]$ is constructed state-by-state*
$$\nu_0 \, M[(s_{m,k})] = \overline{Y}_{m,k}^{-1} \, \pi_{m,k}$$
for each $s_{m,k} \in S$ and time-zero marginal value of wealth ν_0.
2. *Efficient allocations can be implemented with assets from the benchmark economy 1*
3. *The ex-post distribution of wealth is degenerate.*

The complete markets implications are well known. In section 6.1 of Appendix A, we give a constructive proof that provides the equilibrium contracts $a^*(s_{mk,j})$ in terms of the assets $(a_0, a_j(n_{1,j}, m))$ from section 1.1. We state the key constructions here:

Supporting Positions A standard argument, reproduced in 6.1.2, shows the equilibrium complete markets wealth shares are identical $\theta_{T,j}(s_{m,k}) = \theta_T(s_{m,k}) = \overline{Y}_{m,k}$ for all individuals $j \in \mathcal{I}$. Securities positions a_j supporting the risk sharing rule are

$$a(s_{m,k}, n_j)_{\text{Complete}} =: a_j^0 = -n_{1,j}[Y_{m,k} - V_{1,m}]^{-1} \tag{S.0.1}$$

for every $s_{m,k} \in S$. Supporting positions are derived in section 6.1.3.

Asset Prices From the equilibrium wealth shares $\theta_{T,j}(s_{m,k})$, any investor's marginal value of wealth recovers the representative Lucas pricing kernel state-by-state

$$\frac{d}{d\theta}\log[\theta_{T,j}(s_{m,k})]\pi_{m,k} = [\theta_T(s_{m,k})]^{-1}\pi_{m,k} = \nu_0\, M[\overline{Y}_{m,k}]_{\text{Lucas}}$$

for each $s_{m,k}$. The full system of state prices for $s_{m,k} \in S$ are obtained in this way. In particular, $q(s_{m,k}, n_j) = q(s_{m,k})\pi_j = \nu_0^{-1}[\theta_{T,j}(s_{m,k})]^{-1}\pi_{m,k}\frac{1}{2} = \nu_0^{-1}\frac{1}{2}[\overline{Y}_{m,k}]^{-1}\pi_{m,k} = \frac{1}{2}M[\overline{Y}_{m,k}]_{\text{Lucas}}$.

5.2 Incomplete Markets

The set-up is identical to the baseline case presented in section 1.[21] The objective is restated with the proofs in section 6.2 of Appendix A.

Proposition 5.2 (Incomplete Markets Benchmark) *For the incomplete markets model 5.2, an equilibrium exhibits the following properties*

1. *There is a no trade equilibrium at time $t = 0$.*
2. *Trade at $t = 1$ induces a distribution of wealth $\theta_{T,j}(s_{m,k})$ with shares for each investor j proportional to her realized $n_{1,j} \in \{-\Delta, \Delta\}$.*
3. *The incomplete markets pricing kernel $M[\theta_{t,j}(s_{m,k})]_{NC}$ can be written in terms of the complete markets kernel and the wealth distribution state-by-state*

$$M[\theta_{T,j}(s_{m,k})]_{NC} = \left[M[\overline{Y}_{m,k}]_{Lucas}\right]\, e^{\sigma_\Delta(s_m)(1+\frac{1}{2}\sigma_\Delta(s_m)) - o(\frac{1}{\Delta})}$$

for $s_{m,k} \in S$, and where $\sigma_\Delta(s_m) = \left(\frac{\Delta}{W_{1,m}(s_m)}\right)^2$ and

$$M[\overline{Y}_{m,k}]_{Lucas} = \left[\frac{\partial J_1}{\nu_0}\pi_m\right]\frac{\overline{Y}_{m,k}^{-1}}{\partial J_1}\pi_k$$

   (a) *Prices are strictly higher than complete markets prices. The difference is proportional to the welfare loss.*

[21] Program 1.A is optimized and markets 1.C clear.

Remark The cross-sectional transfers $\{n_{t,j}\}_{j\in\mathcal{I}}$ stimulate rebalancing activities. At time-zero, expected marginal values are distorted by uncertainty about $n_{t,j}$. At time 1, the distribution of wealth is bimodal, with relatively poor and rich populations corresponding to realizations $-\Delta$ and Δ, respectively.

Remark Complete markets S.0.1 represent $(s_{m,k}, n_{1,j})$ - contingent payments. For example, policy a_j^0 for type j: $n_{1,j} = \Delta > 0$ requires payment to type $-j : n_{1,-j} = -\Delta < 0$ in the *low* productivity state, $Y_{m,k} - V_{1,m} < 0$. In contrast, for *any* aggregate state, incomplete markets policies a_j respond to $n_{1,j} = \Delta > 0$ with acquisitions, while policies a_{-j} respond to $n_{1,-j} = -\Delta < 0$ with liquidations.

Proof We first show that at time-zero, no-trade is an equilibrium. We then open markets in response to the realizations $n_{1,j}$ to derive wealth shares, securities positions and prices. Details omitted from the main text are found in section 6.2 in Appendix A

No Trade Given that there is no ex-ante heterogeneity, no trade at time-zero can be seen by assuming every investor consumes her endowment, then using the corresponding (IMRS) as a price system. The proof is given in section 6.2.1.

We now state the wealth shares and describe the supporting securities positions. Asset prices are developed in the following section, 2.2.1.

Risk Sharing Using $\partial J_{1,j} = \frac{\partial}{\partial a_j} J_{1,j}$ and $\overline{Y}_{m,k} = Y_{m,k} + \omega_0$, wealth shares are written

$$\theta_{2,j} = \frac{\partial J_{1,-j}}{\partial J_{1,j} + \partial J_{1,-j}} \overline{Y}_{m,k} \tag{I.θ}$$

The derivation of equilibrium wealth shares is in section 6.2.2 of Appendix A.

Securities Market clearing gives $a_j = -a_{-j}$. Write the wealth shares $\theta_{2,j} = a_j(Y_{m,k} - V_{1,m}) + n_{1,j} + \theta_0$ with common term $\theta_0 := \overline{Y}_{m,k} + a_0(Y_{m,k} - V_0)$. Put $W_{1,m} := V_{1,m} + \omega_0$. For log utility, $\partial J_{1,j} = [W_{1,m} + n_{1,j}]^{-1}$. Using I.$\theta$, these imply $[a_j - n_{1,j} W_{1,m}^{-1}][Y_{m,k} - V_{1,m}] = 0$.

We express the reallocation policies a_j as a partition of $n_{1,j}$ into two components. One component corresponds to adjustments in the equity position via $a_j V_{1,m}$, while the other component maps to "cash." An implementation of the policy a_j can be written, in units of wealth,

$$\begin{aligned} \textbf{Equity:} \quad & a_j V_{1,m} = n_{1,j} \frac{V_{1,m}}{W_{1,m}} = n_{1,j}\, \alpha_1 \\ \textbf{Cash:} \quad & n_{1,j} - a_j V_{1,m} = n_{1,j}[1 - \alpha_1] \end{aligned} \tag{S.1.2}$$

for $n_{1,1} = \Delta$, $n_{1,2} = -\Delta$ and every $s_{m,k} \in S$, and where α_1 is the fraction of aggregate wealth held in the risky asset in equilibrium.

□

Remark The partition in S.1.2 captures the myopic policy formation characteristic of log-utility populations. The response to $n_{1,j}$ simply splits the gain or loss into risky and risk-free components at the same rate that portfolios hold equity and cash in equilibrium.

6 Appendix: Empirical Implications and Evidence

Rather than understanding intermediaries as economic actors themselves our theory suggests the bank's balance sheet uniquely captures demand for liquidity in the cross-section of investors. The endogenous structure of the banking arrangement suggests a particular function of the wealth distribution is captured by a particular function of bank balance sheet. This generates testable implications for a "representative" agent asset pricing theory. The theory is doubly productive because it provides tests that exploit data on financial institutions rather than individual-level data on income, net worth, human capital, real estate etc.

The empirical implication of the representative agent prediction is that shocks to the marginal rate of liquidity production, measured by the rate of risky assets to uninsured liquid liabilities, will have cross-sectional pricing power in markets where assets are accessible broadly to both institutions and individuals. Assets relevant to our predictions include exchange traded stocks and indices. Arguably tests are relevant in fixed income and options markets, which exhibit lower participation rates because of decisions not to enter rather than prohibitions.

The theory also predicts that the component of the pricing kernel that bank balance sheet statistics capture uniquely arises from inter-temporal hedging motives of the reference investor.[22] In large incomplete markets economies with institutional liquidity production, myopic investors induce inter-temporal hedging motives in the SDF through changes in an aggregate measure of their propensity to over-save reflected in bank financing flows.[23,24] Moreover, the wealth distribution channel can operate holding aggregate cash flows fixed, which highlights the propensity for intermediary balance sheets to generate pure discount rate effects in the time series of asset returns. An interesting implication is that a decomposition of the log pricing kernel in this model using the present-value identity (Campbell and Shiller, 1989) should reflect that the liquidity factor yield does not predict cash flows.

6.1 Data

We use data from the Flow of Funds reports by the Federal Reserve for balance sheet information. The data are quarterly, and aggregated sector-wide. We collect data from sectors that finance a significant portion of their assets with demandable liabilities: commercial banks and broker-dealers. We also collect data for off balance sheet asset-backed commercial paper (ABCP) activities reported to the Federal Reserve, a significant fraction of which have bank holding companies or subsidiaries of bank holding companies as their conduits.

We connect repo financing and ABCP-type liabilities to the demandability of deposits. First, repo are provided by many institutional depositors (Gorton) and almost all of them are expected to rollover their financing. Commercial paper also tends to have a large number of buyers, although CP is less often used as a permanent financing policy. The expected perpetuity property of repo is the same as deposits. Empirically, a key difference is that deposits are countercyclical, while repo are pro-cyclical.

Hence, the liabilities used for construction of our series are repurchase agreements, large

[22]The term reference is used in place of representative when the marginal investor cannot be recreated from linear combinations of individual investors in the model.

[23]Similarly, Chien, Cole and Lustig (2012) and Chien and Lustig (2010) find i.i.d. dynamics can still produce persistent risk prices in large incomplete markets economies.

[24]This point is elaborated in a dynamic version of this model, available upon request.

time deposits, uninsured savings and checkable deposits and ABCP. The exclusion of ABCP appears to have little effect. The inclusion of insured deposits has significant effects on the time-series properties and the cross-sectional exposure patterns of the series. Similarly, repurchase agreements and uninsurable large time deposits are necessary for the series to produce a viable distribution of exposure in the cross-section of equities. The test-asset cross-sections data are from Ken French. Risky assets are measured by corporate equities, mutual fund shares, and private residential and commercial mortgage-backed securities (MBS). The productivity series is defined as changes in the ratio of risky assets to liquid liabilities. We report the β distributions and the liquidity production time series in the charts below.

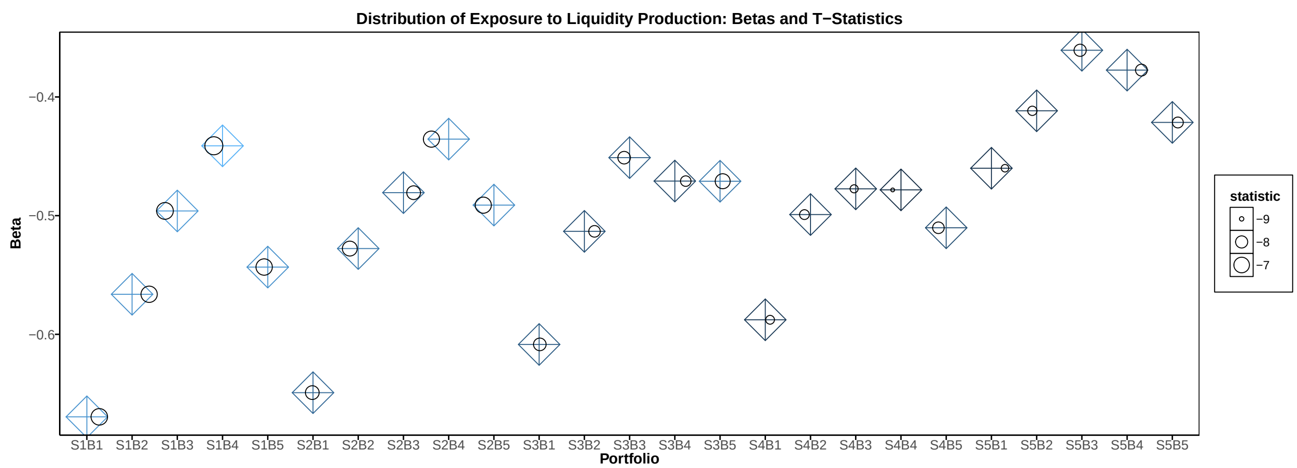
Table 1: Cross Sectional Exposure to Changes in Liability-Side Productivity

	Portfolio	term	estimate	statistic	std.error
1	S1B1	Δ Liquidity	-0.670	-6.678	0.100
5	S1B5	Δ Liquidity	-0.543	-6.731	0.081
6	S2B1	Δ Liquidity	-0.649	-7.577	0.086
10	S2B5	Δ Liquidity	-0.491	-6.808	0.072
11	S3B1	Δ Liquidity	-0.609	-7.881	0.077
15	S3B5	Δ Liquidity	-0.471	-7.206	0.065
16	S4B1	Δ Liquidity	-0.588	-8.641	0.068
20	S4B5	Δ Liquidity	-0.510	-8.134	0.063
21	S5B1	Δ Liquidity	-0.460	-8.806	0.052
25	S5B5	Δ Liquidity	-0.421	-8.287	0.051

(a) The bank balance sheet productivity measures the ratio of high risk assets to liquid liabilities. Comparison of large-small spread and high-low spread (high-low book to market (BTM) ratios). Quarterly balance sheet data for commercial banks and broker-dealers from the Flow of Funds, Board of Governors of the Federal Reserve. We use private depository institutions, issuers of asset-backed securities, and securities brokers and dealers to measure liquidity production. The ratio of high risk assets to liquid liabilities is calculated by classifying liquid liabilities as large time deposits, uninsured checkable and savings deposits, ABCP and repurchase agreements. Risky assets are corporate equities, mutual fund shares, and private residential and commercial mortgage-backed securities (MBS). Monthly Fama -French 3-factor and Carhart model returns data are from 1967 $Q1$ to 2016 $Q4$.

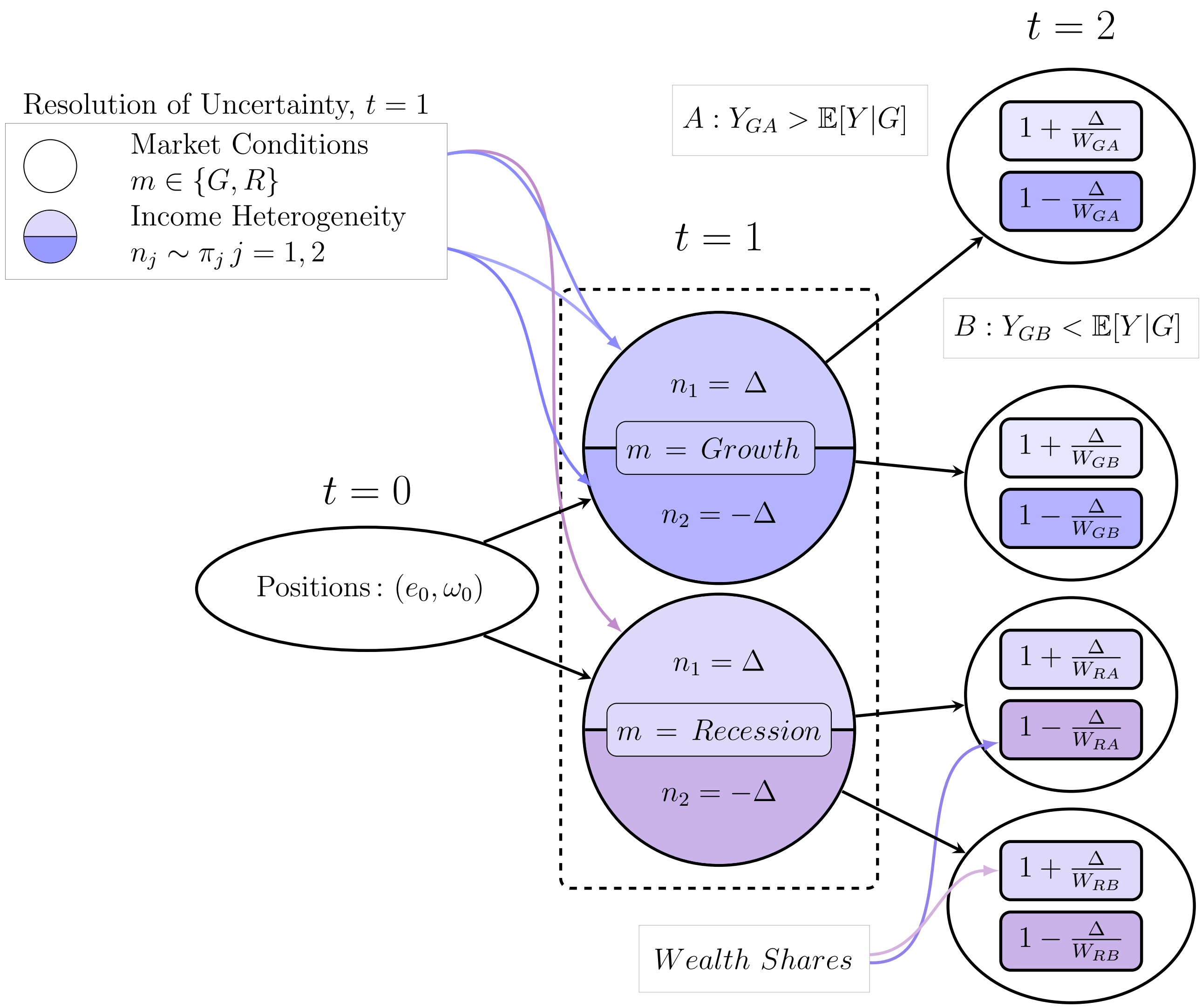
7 Charts and Figures

Figure 4: Exposures to Changes in Liability-Side Productivity



(a) The distribution of exposure spreads value and size. Both size and value spreads are large. The bank factor should not be identical to value, because the motive to save with large banks appears even for logarithmic investors, and the value premium captures intertemporal hedging demands (i.e., the HML factor is in zero-net supply). Quarterly balance sheet data for commercial banks and broker-dealers from the Flow of Funds, Board of Governors of the Federal Reserve. We use private depository institutions, issuers of asset-backed Securities, and securities brokers and dealers to measure liquidity production.The ratio of high risk assets to liquid liabilities is calculated by classifying liquid liabilities as large time deposits, uninsured checkable and savings deposits, ABCP and repurchase agreements. Inclusion of insured deposits significantly alters the time series. Risky assets are corporate equities, mutual fund shares, and private residential and commercial mortgage-backed securities (MBS). Monthly Fama -French 3-factor and Carhart model returns data are from 1967 $Q1$ to 2016 $Q4$.

Figure 5: Binomial Information Structure and Timing



$X_0 = (q_{0,0}, V_0, T) \quad | \quad | \; X_{1,j} = (q_{1,j}(n_j), V_{1,m}, 1) \; | \quad | \; X_{T,j} = (q_{T,j}, Y_{m,k}, 0)$

(a) The state of the system for each individual j is $X_{t,j} = (q_{t,j}, V_{1,m}, T - t)$. Each circular node is a distinct aggregate state. Contiguous nodes are indicated by straight black arrows. In $t = 1$, each ex-ante identical investor j learns her income $n_j \in \{-\Delta, \Delta\}$ and thus wealth $q_{t,j}(n_j)$. Population heterogeneity is captured by split colouration in aggregate nodes. Simultaneously, a public signal m reveals the aggregate productivity path. Signals $m = G$ and $m = R$ correspond to growth and recession. Investors rebalance in response to news. In the final node, $t = T$, output is either above A or below B expectations. Ratios $1 + \Delta W_{m,k}^{-1}$ are wealth shares by type j, and $W_{m,k} = Y_{m,k} + \omega_0$.

8 Proofs

Recursion

We show indifference to initial aggregate wealth $W_0 > 0$ and future levels of aggregate wealth $W_t > 0$ $t = 1, T$. Standard arguments are used, based on homothetic preferences and properties of log. The additional state variables are discussed.

Proof Recall gross positions as fractions of net worth, α_j, and share adjustments a_j, are connected via $\alpha_j W_{1,j} = (1+a_0+a_j)V_{1,m}$, for $W_{1,j} = (1+a_0)V_{1,m} + \omega_0 + n_{1,j} - a_0 V_0$. Similarly for $t = 0$, $\alpha_0 W_0 = (1+a_0)V_0$. Using wealth shares α define the returns to wealth over periods $0 \to 1$ and $1 \to 2$, respectively,

$$R_{0,j} := R(a_0, n_j; V_{1,m}) = \alpha_0 \frac{V_{1,m}}{V_0} + (1 - \alpha_0) + \frac{n_{1,j}}{W_0}$$
$$R_{1,j} := R(a_j; Y_{m,k}) = \alpha_j \frac{Y_{m,k}}{V_{1,m}} + 1 - \alpha_j$$

Now, write the wealth process $(W_{1,j}, W_{2,j})$ in terms of W_0 in the natural way. Set $W_{1,j} = W_0 R_{0,j}$ and then $W_{2,j} = W_{1,j} R_{1,j} = W_0 R_{0,j} R_{1,j}$. For convenience, denote the gross return on initial wealth $\theta_{T,j} := W_{2,j}$.

Log utility decouples today's allocation policies from cumulative effects of future policies. Together with the tower property of conditional expectations, and using $n_{1,j}$ *i.i.d.*, it is clear the objective from program 1.A can be written

$$\max_{\boldsymbol{a},\boldsymbol{\theta}} \mathbb{E}_0[\log(\theta_{T,j})] = \log(W_0) + \max_{a_0} \mathbb{E}_0[\log(R_{0,j})] + \mathbb{E}_j \left[\max_{a_j} \mathbb{E}_1[\log(R_{1,j})] \right] \tag{1.B.1}$$

The level $\log(W_0)$ is irrelevant for allocation decisions and therefore irrelevant for asset pricing, from 1.B.1. Moreover, expected utility is unique only up to order preserving transformations, so we can remove the scale factor $\log(W_0)$. Equivalently, without loss of generality, set $W_0 = 1$.

From 1.B.1 (and 1.B below) we can also disregard future levels of aggregate wealth W_t, $t \leq T$ because log investors require only single-period gross returns $R_{t,j}$ for allocation decisions. Moreover, there is no intermediate consumption. State prices are constructed from shadow values of one-period gross returns.

□

Write the state vector for every individual

$$X_{j,t} := \begin{bmatrix} q_{t,j} \\ V_t \\ T-t \end{bmatrix} \in \mathbb{R}^2_{++} \times \{2,1,0\} \tag{X.1}$$

for strictly positive prices $(q_{t,j}, V_t)$. At $t = 0$, the normalization $W_0 \equiv 1$ implies $q_0 = 1$ for every investor.

Define $J(q_{t,j}, V_t, T-t) := \mathbb{E}_t[\log(\Pi_{s=1}^{T-t} R_s(\boldsymbol{a}^*))]$ along the optimal policy path $\boldsymbol{a}^*$. The additional index in $J(q_t; V_t, T-t)$ monitors the number of periods prior to termination, $T-t$, although we adopt the conventional shorthand $J_0 = J(X_0) = J(1; V_0, 2)$ and $J_{1,j} = J(X_{j,1}) = J(q_j; V_{1,m}, 1)$. Indirect utility separates recursively

$$\begin{aligned} J(X_0) &= \max_{a_0} \mathbb{E}_0 \left[\log(R_{0,j})\right] + \mathbb{E}_j \left[J(X_{j,1})\right] \\ J(X_{j,1}) &= \max_{a_j} \mathbb{E}_1 \left[\log\left(R_{1,j}\right)\right)] \end{aligned} \tag{1.B}$$

Heterogeneity is tracked by treating the ratio of individual wealth to initial wealth $q_{t,j} W_0^{-1} = q_{t,j}$ as a state variable for each individual. This is equivalent to treating private income $n_{1,j}$ as the individual state following from the fact that, contemporaneously, $q_{1,j} = W_{1,m,j} = W_{1,m} + n_{1,j}$. The uninsurable shock $n_{1,j}$ is necessary conditioning information. Policies satisfying 1.A or 1.B / 1.B.1 are made contingent on type $j \in \{1, 2\}$ for $t \neq 0$.

Finally, policies are made contingent on aggregate prices. The deterministic probabilities $\pi_{m,k,j} = \pi_m \pi_k \frac{1}{2}$ are common knowledge. We conclude the vector $X_{j,t}$ in X.1 is a sufficient statistic for the state of the economy.

Note that prices are $V_t = V([s_{m,k}]_t, T-t, (q_{t,j}, q_{t,-j}))$ for $[s_{m,k}]_1 = s_m$, $[s_{m,k}]_2 = s_{m,k}$ and $[s_{m,k}]_0 = \text{null}$. The endogenous state can be altered in several ways, through a change of variables, and still produce a valid description of the economy.

8.1 Proposition 2.1

Back to Section 2.1

Market Arrangements We complete securities markets by including an Arrow-Debreu contingent claim $a(s_{mk,j})$ for each $s_{mk,j} \in S := \mathcal{Y} \times (n_{1,1}, n_{1,2})$, the set of all pairs $(Y_{m,k}, n_{1,j})$. For example $s_{GB,2} = (Y_{G,B}, n_{1,2})$. Each claim is traded at time $t = 0$. Contracts are fully enforceable. Arrow-Debreu prices are $q(s_{m,k}, n_j)$. We nest the positive supply endowments e_0 into the contingent claims menu S. Market clearing for every $s_{m,k}$ is $\sum_j a(s_{mk,j})\pi_j = \frac{1}{2}a(s_{m,k}, n_{1,1}) + \frac{1}{2}a(s_{m,k}, n_{1,2}) = \overline{Y}_{m,k}$.

8.1.1 The Economy

The present value of all expenditures net of endowments must equal zero. Write the objective

$$\begin{aligned} J_0(q_0; V) &= \max_{\boldsymbol{a},\theta} \mathbb{E}\left[\log(W_{1,j})\right] \qquad (2.0) \\ \text{s.t.} \qquad & \sum_k \sum_m q(s_{m,k}) \sum_{j=1,2} \frac{1}{2} a(s_{m,k}, n_j)\pi_{m,k} \leq \omega_0 + V_0 \\ & \theta_{1,j}(s_{m,k}) = W_{1,j} \end{aligned}$$

where now a single time zero budget constraint includes the complete set of marketable securities spanning aggregate and individual j-shocks. Shares of equity are fixed by the endowments giving $\mathbb{E}^Q[e_0] = V_0$. $\theta_{1,j}(s_{m,k}) = \theta_{t=1}(s_{m,k}, n_j)$ is the gross return to initial wealth $W_0 \equiv 1$ in state $(s_{m,k}, n_{1,j})$.[25] For details of the securitization of $n_{1,j}$ see the Securitization section below.

We can read off the first order conditions by inspection

$$-\nu_0\, q(s_{m,k}) + [\theta_{1,j}(s_{m,k})]^{-1}\pi_{m,k} = 0 \qquad (5.a)$$

for every $s_{mk,j}$, where $\nu_0 = J_0'(q_0, V)$ is the initial marginal value of wealth, and $q(s_{m,k})$ is the price of a claim to one dollar in state $s_{m,k}$. ν_0 is necessarily identical across investors.

[25] θ is a control dummy for final wealth, used for convenience. If a reader prefers to think of utility defined over consumption of a non-perishable numeraire good at the terminal date, written say, $c_1(s_{m,k}, n_j)$, then $\theta_{1,j}(s_{m,k}, n_j) = c_1(s_{m,k}, n_j)$.

Write $\lambda_{0,j}$ for the Lagrange multiplier on the initial budget constraint for investor $j \in \mathcal{I}$. The envelope is $\lambda_{0,j} = J'_{0,j}(q_0, V)$. By ex-ante symmetry, $\lambda_{0,j} = \lambda_0$ for $j \in \mathcal{I}$, so $J'_{0,j} = J'_0 = \nu_0$.

8.1.2 Wealth Shares

Proof From equation 5.a, for each $s_{m,k} \in S$ and any two $i, j \in \{1, 2\}$, market prices enforce

$$\theta_{T,i}(s_{m,k}) = \theta_{T,j}(s_{m,k})$$

Investors all have identical final wealth shares $\theta_{T,j}$, which must in turn equal the average share and the total level

$$\theta_{T,i}(s_{mk,i}) = \frac{1}{2}\sum_j \theta_{T,j}(s_{mk,j}) = \omega_0 + Y_{m,k}$$

where the second equality follows by market clearing for terminal wealth. In particular, for every infinitesimal investor and for all $s_{mk,j} \in S$, $\theta_{T,j}(s_{mk,j}) = \omega_0 + Y_{m,k} = \overline{Y}_{m,k}$ is the complete markets (perfect) risk sharing rule.[26] By populations j, for $\pi_j = 1/2$, the rule is $\widehat{\theta}_1(s_{k,h}, n_j) = \frac{1}{2}\left[\omega_0 + Y_{m,k}\right]$.

□

Back to Section 2.1

8.1.3 Supporting Positions

Proof We implement the complete markets risk sharing rule using assets from the benchmark economy. We use a completed menu (containing j-contingencies) of the two-step assets described in the binomial model.[27] Individual portfolio realizations for each $s_{mk,j}$, can be written

$$\theta_{1,i}(s_{mk,i}) = \omega_0 + A_j(s_{mk,j}) + n_{1,j}$$

[26] It is straightforward that when net private income distributions $\sum_j \pi_j \Delta_j =: \overline{\Delta} \neq 0$, shares by type j are $\pi_j \left[\omega_0 + \overline{\Delta} + Y_{m,k}\right]$.

[27] These assets are more useful for analyzing the different welfare implications across the three economies we consider. Unsurprisingly, allocations using (a_0, a_j) are equivalent for asset pricing and welfare analyses to allocations using the Arrow-Debreu menu $\mathbf{a}_0 := a(s_{m,k}, n_j)_{\{m,k,j\}}$. Formalities are addressed in the Equivalence discussion of this Appendix.

We recover the allocations by unpacking

$$A_j(s_{mk,j}) = (1 + a_0 + a_{j,m})Y_{m,k} - a_0 V_0 - a_{j,m} V_{1,m}$$

Using $\theta_{1,i}(s_{mk,i}) = \theta_{1,j}(s_{mk,j})$ and $a_0 = 0$, simple algebra reveals

$$[a_{j,m} - a_{-j,m}][Y_{m,k} - V_{1,m}] = n_{1,-j} - n_{1,j}$$

Finally, we appeal to scarce resources $\frac{1}{2}\sum_j(1 + a_{j,m}) = 1$, having used $a_0 = 0$. Recalling that $n_{1,j} + n_{1,-j} = 0$, the remaining allocations can be expressed, in terms of equilibrium objects,[28]

$$a(s_{m,k}, n_j)_{\text{Complete}} =: a_j^0 = -n_{1,j}[Y_{m,k} - V_{1,m}]^{-1} \tag{S.0.1}$$

for every $s_{m,k} \in S$.

□

Asset Prices Given the symmetric wealth shares $\theta_{2,j}(s_{m,k})$, any investor's marginal value of wealth can be written in terms of aggregates (identically) and used to price assets. See the Asset Pricing discussion in Section 2.1.

Securitization We can write $\mathbb{E}^Q[e_0] = V_0$ for the unit price of market equity. Note that $\mathbb{E}^P[n_{t,j}] = 0$ for $dQ = e^{-\eta(s)}dP$ but $\mathbb{E}^P[e^{-\eta(s)}n_{t,j}]$ is an equilibrium object.[29] There are several ways to allow $n_{t,j}$ to be marketable. We adopt the simplest case for the present-value representation of our economy by securitizing claims to $n_{t,j}$ at time zero. The equilibrium value for a claim to n_j is

$$\mathbb{E}^Q[n_{t,j}] = \sum_{k=U,D}\sum_{m=R,G} q(s_{m,k}) \sum_{j=1,2} \frac{n_j}{2}\pi_{m,k} = 0 \tag{1n}$$

[28]In terms of model primitives

$$a(s_{m,k}, n_j)_{\text{Complete}} = -n_{1,j}\left[Y_{m,k} - \mathbb{E}_1\left[Y_{m,k}[Y_{m,k} + \omega_0]^{-1}\right]\right]^{-1}$$

for every $s_{m,k} \in S$ and where $\mathbb{E}_1\left[Y_{m,k}[Y_{m,k} + \omega_0]^{-1}\right] = V_{1,m}\nu_0 = \mathbb{E}_1^Q[Y_{m,k}]\nu_0$ with $\nu_0 = 1$ for $u(W_2) = \log(W_2)$ and $W_0 \equiv 1$.

[29]When $n_{t,j}$ is orthogonal to the pricing kernel, $\mathbb{E}^Q[n_{t,j}] = 0$. In complete markets with private shocks that are aggregate-neutral this condition is satisfied. The process $\eta(s)$ has $\mathbb{E}^P[e^{-\eta(s)}] = 1$.

Securitizaion of n_j has no impact on the level of tradeable wealth at time-zero W_0.[30]

Budget Constraint We use the resource restriction

$$\sum_j \frac{1}{2} a(s_{m,k}, n_j) = Y(s_{m,k}) \qquad s_{m,k} \in S$$

together with 1n to write the investor's complete markets budget constraint

$$\sum_k \sum_m q(s_{m,k}) \sum_j \frac{1}{2}[a(s_{m,k}, n_j) - 2Y_{m,k}]\pi_{m,k} \leq \omega_0$$

which states the present value of all financed positions net of endowments is zero.

□

Back to Section 2.1

Equivalence Unsurprisingly, the allocations $\mathbf{a}_0$ in the time-zero economy are equivalent to the allocations (a_0, a_j) in the original two-step economy. By implementing the complete markets risk-sharing rule with a feasible allocation of assets consistent with the trading protocol from the two-step economy, we have shown that *an* allocation in the two-step economy $(\hat{a_0}, \hat{a}_j)$ is weakly preferred to $\mathbf{a}_0$.[31] Because $\mathbf{a}_0$ is Pareto efficient in a frictionless economy with resources and time-separable preferences that are identical to those in the two-step economy, it must also be that $\mathbf{a}_0$ is weakly preferred to *any* $(\tilde{a}_0, \tilde{a}_j)$. Out of these we pick $(\hat{a_0}, \hat{a}_j)$ and set $(a_0, a_j) = (\hat{a_0}, \hat{a}_j)$.

□

8.2 Incomplete Markets: Proposition 2.2

Back to Section 2.2

[30] Of course, the tradeability of $n_{t,j}$ shows up as an additional lever in the allocation policies $a = a(s_{m,k}, n_j)$.

[31] That is, you would never do worse by optimizing in the two-step economy directly.

8.2.1 The Economy

Each investor faces the objective

$$
\begin{aligned}
J(q_0; V_0) \ = \ & \max_{\boldsymbol{a},\theta} \mathbb{E}\left[\log(\theta_{2,j})\right] \\
\text{s.t.} \quad & a_0 V_0 - V_0 \leq \omega_0 \\
& a_j V_{1,m} - (1 + a_0) V_{1,m} \leq (\omega_0 - a_0 V_0) + n_{1,j} \\
& \theta_{2,j} = \overline{a}_j Y_{m,k} + (\omega_0 + n_{1,j} - a_0 V_0 - a_j V_{1,m})
\end{aligned}
\tag{IC.1}
$$

where $\overline{a}_j = (1 + a_0 + a_j)$ and we distinguish the final portfolio value $\theta_{2,j} = W_{2,j}$ from the wealth process $W_0, W_{1,j}, W_{2,j}$. Note that while $n_{t,j}$ cannot be securitized, after $n_{1,j}$ is realized all wealth is tradeable.

Definition: Incomplete Markets Equilibrium In equilibrium, every investor optimizes IC.1 and markets clear according to 1.C.

8.2.2 No Trade Equilibrium

Proof Endowments and preferences are identical. Suppose a price system at time-zero for aggregate states $s_{m,k} \in S$ is given by

$$
\begin{aligned}
q(s_{m,k})\nu_0 = & [\frac{1}{2}[\hat{e_0} + n_{1,1}]^{-1} + \frac{1}{2}[\hat{e_0} + n_{1,2}]^{-1}]\pi_{m,k} \\
= & \left[[\hat{e_0} + \Delta]^{-1} + [\hat{e_0} - \Delta]^{-1}\right] \frac{1}{2}\pi_{m,k}
\end{aligned}
$$

where $\hat{e_0} = \hat{e}_0(s_{m,k}, t)$ is the realization claimed by an investor owning e_0 in state $s_{m,k}$ and period t. In the terminal period, $\hat{e_0} = Y_{m,k}$, while in the interim period $t = 1$, $\hat{e}_0$ is the capital value $V_{1,m}$. The gross rate of time discount is $1 + \beta = 1$.

When we propose a no-trade allocation, feasibility is automatic. Every investor holds her endowment. Moreover, the two-period horizon circumvents the need to verify transversality conditions. We are left to verify optimality.

The investors face the same number of contingencies as in the complete markets case, but they can only access half the number of primitive assets, corresponding to the cardinality of

$\{s_{m,k}\}_{m,k}$. Consider an economy with trade and write the wealth shares $\theta_{t,j}(s_{m,k})$ for $t = 1, 2$. For each tradeable contingency $s_{m,k}$, first order conditions are

$$-\nu_0\, q_\theta(s_{m,k}) \; + [\frac{1}{2}[\theta_{t,j}(s_{m,k})]^{-1} + \frac{1}{2}[\theta_{t,j}(s_{m,k})]^{-1}]\pi_{m,k} = 0$$

where $\frac{1}{2} = \pi_j$ is used, and $\nu_0 = J_0'(q_0, V)$ is the initial marginal value of wealth, necessarily identical across investors.

In contrast, we have proposed prices that correspond to the intertemporal tradeoff

$$-\nu_0\, q(s_{m,k}) + \left[[\hat{e_0} + \Delta]^{-1} + [\hat{e_0} - \Delta]^{-1}\right] \frac{1}{2}\pi_{m,k} = 0$$

for every time-zero investor and any state $s_{m,k} \in S$. The only hope for improving this margin is to pick a $\theta_{1,j}$ to reduce the Jensen cost over states j conditioning on $s_{m,k}$. By assumption, there are no securities to trade on the realizations $n_j \in \{-\Delta, \Delta\}$ and hence, $\theta_{1,j}$ is contingent on n_j only through $\theta_{1,j} = \theta_1^* + n_{1,j}$ where θ_1^* is a control variable at time zero. Investors still must average over $n_{1,j}$ realizations for each $s_{m,k}$. Thus, we can take $\theta_{1,j} = \hat{e}_0 + n_{1,j}$ and $q_\theta = q$. The proposed price system is an optimum for every investor, is feasible, and clears markets.

□

Back to Section 2.2

8.2.3 Incomplete Markets Wealth Shares

Every investor has an identical portfolio *coming in* to the first period, prior to realization of shocks $n_{1,j}$. In response to $n_{1,j}$ and the signal m investors enter securities markets to arrange their final portfolios. Outgoing positions take the form

$$W_{1,j} = \underbrace{V_{1,m}a_j + V_{1,m}}_{\text{Equity Claim}} + \underbrace{\omega_0 + n_{1,j} - a_j V_{1,m}}_{\text{Risk Free Holdings}}$$

and differ for each m only through the pairs $(n_{1,j}, a_j)$.

Proof We derive policies a_j by first extracting the risk sharing rules $\theta_{2,j}$. Market prices enforce

$$-V_{1,m}\frac{\partial}{\partial a_j}J_{1,j} + [\theta_{2,j}]^{-1}Y_{m,k}\pi_{m,k} = 0$$

giving the rule

$$\theta_{2,j}\frac{\partial}{\partial a_j}J_{1,j} = \theta_{2,-j}\frac{\partial}{\partial a_{-j}}J_{1,-j}$$

Risk sharing is full conditioning on today's uninsured shock, so variation in the wealth shares $\theta_{2,j}$ across $j = 1,2$ is driven by today's marginal value of wealth. We adopt shorthand $\partial J_{1,j} = \frac{\partial}{\partial a_j}J_{1,j}$ and $\overline{Y}_{m,k} = Y_{m,k} + \omega_0$. Wealth shares

$$\theta_{2,j} = \frac{\partial J_{1,-j}}{\partial J_{1,j} + \partial J_{1,-j}}\overline{Y}_{m,k}$$

follow from final-period goods market clearing by state $s_{m,k}$, written $\theta_{2,j} + \theta_{2,-j} = 2\overline{Y}_{m,k}$.

□

8.2.4 Time-zero Shares and State Prices

First-order conditions for an asset that pays 1 in state s_m are

$$\nu_0\, q(s_m) - \frac{1}{2}\theta_{1,j}^{-1} - \frac{1}{2}\theta_{1,-j}^{-1} = 0$$

Market clearing gives $\sum_j \alpha_j\, W_{1,m,j}\,\pi_j = V_{1,m}$. Then $W_{1,m} = V_{1,m} + \omega_0$ and $W_{1,m,j} = W_{1,m} + n_{1,j}$. Put $\theta_{1,j} = \partial J_{1,j}^{-1} = W_{1,m,j} = W_{1,m}^{*} + n_{1,j}$ where the * indicates the component can be controlled from time-zero. Plugging $\theta_{1,j} = \partial J_{1,j}^{-1}$ into the time-one shares and using $\frac{1}{2}\sum_j \theta_{2,j} = \overline{Y}_{1,m}$ gives $V_{1,m}$. Plugging $V_{1,m}$ into time-zero FOCs using $\theta_{1,j} = W_{1,m} + n_{1,j}$ gives

$$\nu_0\, q(s_m) = W_{1,m}\left([W_{1,m} + \Delta][W_{1,m} - \Delta]\right)^{-1}\pi_m$$

in agreement with NC.

□

Nonseparable Preferences To preserve the comparison in the previous sections, we define perishable consumption in $t = 1$ to be a small dividend paid by the productive asset that is a constant proportion of the expected payout conditioning on that path

$$c(m; \epsilon) = \epsilon \, \mathbb{E}[Y(s_{m,k})|m]$$

Then, individual consumption policies are written $c = c(j, m)$. Define

$$u(c, \theta_{T,j}) := \left[c^{1-1/\psi} + \theta_{T,j}^{1-1/\psi}\right]^{\psi/(\psi-1)}$$

and

$$U = \frac{1}{1-\gamma} u(q_1, \theta_{T,j})^{1-\gamma}$$

In addition, resources are now constrained in $t = 1$ by

$$\frac{1}{2} \sum_j c(j, m) = c(m; \epsilon)$$

9 Organizational Implications: Internal Diversification

In this appendix we present some of the ancillary implications of the theory in more detail.

Liquidity Production and Asset Diversification

Banks and other intermediaries hold diversified assets on their balance sheets, but typical non-financial public firms do not. Conventional wisdom holds that investors are weakly better off when individual firms concentrate risk in their area of expertise. Does the expertise of financial firms require they hold diversified assets, or are these allocations inefficient?

Financial operations carried out by market makers, broker-dealers, prime brokerages, at IB trading desks, etc, require holding a variety of assets on behalf of clients, or available for trade, or in some cases for risk-management. Expertise in these businesses entails more internal diversification than, for example, a bio-tech start-up. However, notably commercial banks are omitted from this list, yet tend to diversify assets. Moreover, IB balance sheets are often concentrated via emphasis on a small number of issuances.

A separate function of financial institutions provides an alternative explanation for asset diversification. Broker-dealers, dealer banks, commercial banks and bank holding companies effectively use risky assets as inputs for the production of liquidity on their liability-side. This function is carried out optimally when the balance sheet is both diversified and risky.

Corollary 9.1 (Liquidity Production and Asset Diversification) *Efficiency of liquidity production increases with asset diversification. In particular, equilibrium liquidity producers hold the market.*

Value is created when a liquidity producer can consistently peel off *average returns* from risky investments - say the market returns - and direct them to the subset of investors with the *highest marginal valuation*. Over time this requires calibrating the distribution of asset returns through portfolio choice. It is costly, on average, to concentrate asset risk: competitive markets for liquidity production will drive out otherwise equivalent institutions with higher overall asset volatility. The most efficient liquidity producers will hold the most diversified asset portfolio, all else equal.