FEDERAL RESERVE BANK OF ST. LOUIS

REVIEW

THIRD QUARTER 2019
VOLUME 101 | NUMBER 3

REVIEW

Volume 101 • Number 3

Review
Review is published four times per year by the Research Division of the Federal Reserve Bank of St. Louis. Complimentary print subscriptions are available to U.S. addresses only. Full online access is available to all, free of charge.

Online Access to Current and Past Issues
The current issue and past issues dating back to 1967 may be accessed through our Research Division website: **http://research.stlouisfed.org/publications/review**. All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this website.

Review articles published before 1967 may be accessed through our digital archive, FRASER: **http://fraser.stlouisfed.org/publication/?pid=820**.

Review is indexed in Fed in Print, the catalog of Federal Reserve publications (**http://www.fedinprint.org/**), and in IDEAS/RePEc, the free online bibliography hosted by the Research Division (**http://ideas.repec.org/**).

Authorship and Disclaimer
The majority of research published in *Review* is authored by economists on staff at the Federal Reserve Bank of St. Louis. Visiting scholars and others affiliated with the St. Louis Fed or the Federal Reserve System occasionally provide content as well. *Review* does not accept unsolicited manuscripts for publication.

The views expressed in *Review* are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Subscriptions and Alerts
Single-copy subscriptions (U.S. addresses only) are available free of charge. Subscribe here: **https://research.stlouisfed.org/publications/review/subscribe/**.

Our monthly email newsletter keeps you informed when new issues of *Review*, *Economic Synopses*, *Regional Economist*, and other publications become available; it also alerts you to new or enhanced data and information services provided by the St. Louis Fed. Subscribe to the newsletter here: **http://research.stlouisfed.org/newsletter-subscribe.html**.

Economic Data
General economic data can be obtained through FRED® (Federal Reserve Economic Data), our free database with over 500,000 national, international, and regional data series, including data for our own Eighth Federal Reserve District. You may access FRED through our website: **https://fred.stlouisfed.org**.

ISSN 0014-9187

What Determines Debt Maturity?

Rodolfo E. Manuelli

What determines the maturity structure of debt? In this article, I develop a simple model to explore how the optimal maturity of debt issued by a firm (or a country) depends both on the firm's cyclical state and other features of the economic environment in which it operates.

I find that firms with better current earnings and better growth prospects issue debt with longer maturity, while firms operating in more-volatile environments issue debt with shorter maturity. Yield to maturity is a poor indicator of the risk of debt issued by a firm. The reason is simple: Yield to maturity captures both default risk and a component that is a pseudo term premium. In the model, the market does require a term premium and one appears only because of the risk of default. It is not possible to separate the impact of maturity and risk. (JEL G12, G30)

Federal Reserve Bank of St. Louis *Review*, Third Quarter 2019, *101*(3), pp. 155-76.
https://doi.org/10.20955/r.101.155-76

1 INTRODUCTION

It is a standard view that borrowing short term exposes debtors—both firms and countries—to potential refinancing shocks. In some cases, the existence of short-term debt has been blamed for the subsequent poor performance of the borrower. In the case of the domestic financial system, the mismatch between (short) liabilities and (long) assets is sometimes viewed as a factor that explains the deep Financial Crisis of 2007-09 (see, for example, Brunnermeier and Oemkhe, 2013). Kacperczyk and Schnabl (2010) review the evidence of the role that commercial paper (which is short-term debt) played in the Finacial Crisis. In the case of sovereign debt, Rodrik and Velasco (1999) argue that short-term debt increases the likelihood of a crisis. Benmelech and Dvir (2011) document how East Asian countries shortened the maturity of their debt before the financial crisis in the 1990s. Broner, Lorenzoni, and Schmukler (2013) find that during crisis times it is optimal for countries to issue short-term debt.

Why do some firms (and countries) borrow short and others long? There is some evidence that shows that firms (or countries) that find themselves in a "weak" position at the time they

Rodolfo E. Manuelli is a professor of economics at Washington University in St. Louis and a research fellow at the Federal Reserve Bank of St. Louis.
The author thanks Sid Sanghi for excellent research assistance and the Federal Reserve Bank of St. Louis for financial support.

need to borrow—for example, when their current income is relatively low—tend to issue relatively short-term debt. In some cases, this has been interpreted as the cause of a subsequent default. This simplistic interpretation hinges on the observation that, since default is caused by the inability of a firm or a country to repay its debt and had the firm or country issued long-term debt, it would have had "more time" to recover and default could have been avoided.

If we are going to go beyond simplistic interpretations of observed correlations, a theory of the factors that determine the choice of maturity is needed. In addition to such a theory's ability to rationalize the evidence, having it is important from the point of view of a policymaker to the extent that it describes the supply of debt for different maturities as a function of economic variables.

The literature on firms' choice of maturity includes the early work by Diamond (1991) and Leland and Toft (1996). Recent work on default risk over the cycle in the corporate sector includes Chen et al. (2017). In the sovereign debt literature, Arellano and Ramanarayanan (2013) determine maturity as a function of the costs of different types of debt. Aguiar et al. (2019), following Arellano and Ramanarayanan (2013), show that a country that has issued long and short bonds and needs to reduce its debt should not intervene in the market for long bonds and make the adjustment using exclusively short-term debt.

In this article, I develop a simple model of the choice of the structure of debt by a risk neutral firm that needs to raise a certain level of funds. I restrict the analysis to the case of a pure discount bond.[1] I assume that the borrower and the financial market share exactly the same discount factor, that they are both risk neutral, and that information is complete. Thus, the choice of debt structure does not depend on differences in the valuation of risks between borrowers and lenders. Finally, I consider the case in which the borrower issues one bond. This eliminates the incentive that borrowers might have to issue short-term debt to dilute the value of long-term liabilities. I do not believe that strategic factors and differences in valuation criteria between borrowers and lenders are unimportant. Rather, I want to emphasize that factors other than those related to valuation and strategic considerations are also relevant to determine the maturity of the debt.

Even though the analysis could be applied to the study of sovereign debt (in this interpretation the borrower in the model is a country), I will assume that the borrower is a firm. This is mostly for convenience and to keep the presentation simple.

I find that there are several factors that influence the optimal (from the point of view of the borrower) maturity of the debt:

(i) *The cyclical component*. Firms that have better prospects (as measured by their current earnings) issue longer-term debt, as do firms with better growth prospects.

(ii) *The market environment*. Firms that operate in more-volatile environments choose to issue shorter-term debt.

(iii) *Asset saleability*. Firms that have high collateral—measured as the market value of their assets in the case of default—also issue longer-term debt.

The intuition underlying these results is somewhat complex. Consider first the impact on the market value of a debt instrument of a given maturity and face value of an increase in current income of the issuer. To the extent that income is positively serially correlated, the increase in current income increases expected future income, lowering the risk of default and, consequently, increasing the market value of the debt. However, a borrower that needs to raise a given amount has the opportunity to redesign the debt instrument to lower the debt's market value to match the financing needed. To this end, the borrower can (and does) increase the maturity (and the face value of the bond), reducing the market value of the debt to the desired level. In the context of the economy that I study, causality runs from the state of the borrower (in this case, the level of current earnings) to maturity: Lower current income induces the borrower to issue shorter-term debt, even when it was possible to issue longer-term debt. This choice is not driven by risk aversion on the part of lenders[2] but rather by the changes in the default risk associated with lower current income.

The impact of higher uncertainty about the growth rate of the firm's income has exactly the opposite effect as an increase in earnings. Higher uncertainty increases the risk of default, lowering the market value of the debt. For a given financing need, the borrower has to change the structure of the debt to restore the debt's market value, which can be accomplished by lowering the expected maturity.

Finally, a higher level of collateral increases the market value of the debt (it lowers the cost of default from the point of view of the debt holders and, hence, the price of risk). In order to reduce this value to match the financing needed, the firm finds it optimal to increase the price of risk by lengthening the expected maturity of the debt it issues.

2 REVIEW OF EMPIRICAL FINDINGS

There is a large empirical literature on the factors that influence debt maturity of business firms. In a recent paper, Manuelli and Sanchez (2019) provide a short review of this literature and perform an up-to-date empirical analysis using data on nonfinancial and nonregulated firms in Standard & Poor's Compustat dataset.[3] They define maturity as the ratio of debt with maturity of less than five years to total debt. Table 1 reports some of their results.

Manuelli and Sanchez's (2019) results parallel many others in the literature. They find that firms that operate in more-uncertain environments, as measured by the different volatility measures that they compute, decrease the maturity of the debt (that is, they increase the share of short-term debt). In their sample, asset maturity—measured as a weighted average of short- and long-term assets—does not have a significant impact on maturity. In this sample, there is very little evidence of "matching" the maturity of the asset with the liabilities of the firm. Asset salability—measured as the ratio of total net property, plant, and equipment to total assets—has a significant impact on maturity. I interpret this as indicating that firms that have assets that can be more easily sold in the market (as opposed to firms whose assets have a large intangible component) choose to issue longer-term debt. Finally, size—measured here by sales—has a positive impact on maturity: Firms with higher sales tend to issue, on average, longer-term debt.

The results reported here are consistent with the findings of Dinlersoz et al. (2018), who use a large data set that includes both traded and private firms. They find, in a cross section, that larger firms choose a higher ratio of long-term to short-term debt. Moreover, maturity—

Table 1

Dependent Variable: Ratio of Short-Term Debt

R&D investment	–0.0102*	–0.0139*	–0.1250*
Volatility (sales)	0.0945***		
Volatility (EBITDA)		0.1249***	
Volatility (returns)			0.0097***
Asset maturity	–0.0009	–0.0010	–0.0009
Asset salability	–0.1596***	–0.1601***	–0.1631***
Size	–0.0404***	–0.0405***	–0.0394***
Year fixed effect	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes
No. observations	17,169	17,169	17,169
R^2	0.19	0.19	0.19

NOTE: ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively. EBITDA, earnings before interest and taxes.

SOURCE: Manuelli and Sanchez (2019).

proxied by the ratio of long-term to short-term debt—is higher for firms that have higher collateral, measured as the ratio of tangible assets to fixed assets.

The take away from the empirical evidence can be summarized as follows:

- Firms that have higher profits, income, and are larger on average issue longer-term debt.
- Firms that operate in more-volatile environments—for example, as measured by the standard deviation of sales or net income—tend to issue shorter-term debt.
- Firms that can provide better collateral tend to issue longer-term debt.

There is also a fairly significant literature that studies the factors that influence the maturity structure of sovereign debt (see, for example, the discussion in Dvorkin et al., 2018). Some–times in the literature—especially in the case of sovereign debt—the correlation between maturity and the state of the country is interpreted as reflecting a specific direction of causality. More precisely, in the presence of possible interruptions in funding, countries are advised to "borrow long." Moreover, it is not uncommon for countries whose income is low to refinance existing debt using shorter maturities. In this article, I will argue that it is possible that causality goes in the opposite direction: When income is low—and in this model this implies that the probability of default is high—firms *choose* to issue shorter debt even though the market was willing to lend longer. The reason is simple: The price that the market charges for high-maturity bonds makes them less attractive than short-maturity bonds.[4]

In what follows, I present a model that captures the stylized facts listed above and that is useful to disentangle the role played by different economic factors in determining the optimal structure of debt.

3 THE MODEL

I study the case of a risk-neutral firm that has to incur an irreversible cost to implement an investment project. I assume that the returns of the project are completely described by a few parameters that determine the associated stochastic processes for net output (or earnings). I denote net earnings by x_t, and I assume that they evolve according to the following stochastic differential equation:

$$dx_t = \mu x_t dt + \sigma x_t dW_t,$$

where W_t is a Wiener process or Brownian motion and μ (the expected growth rate) and σ (the volatility of the growth rate) are constants.

I assume that x_t is known at all times by all market participants.

Given that the stochastic process for net income is nonnegative, in the absence of debt obligations, it is efficient for this firm to operate forever. In this case, the unlevered value of the firm is

$$V^*(x) = E\left[\int_0^\infty e^{-rt} x_t dt \middle| x_0 = x\right] = \frac{x}{r-\mu}.$$

3.1 The Value of the Firm

I assume that the firm has to finance either part or all of the cost of the project by issuing *noncontingent* debt. The firm has access to a risk-neutral credit market that will price any debt issued by the firm, using the same discount rate, r. Thus, none of the results are driven by either differences in the discount factor or differences in the curvature of direct payoffs.

I consider debt in the form of pure discount bonds with stochastic maturity. Thus, a debt instrument is completely described by two parameters: T, the expected maturity, and K, the face value that must be paid at maturity. I assume that maturity arrives as a Poisson process with rate $\eta = 1/T$. I also assume that if the firm defaults on its debt, bondholders receive a fraction $1 - \delta$ of the unlevered value of the firm.

The value of a firm that has issued a (K,T) bond is

$$(1) \qquad V(x;T,K) = \left\{ E\int_0^{N_T} e^{-rt} x_t dt + e^{-rN_T} \max\left(\frac{x_{N_T}}{r-\mu} - (K+F), 0\right) \middle\| x_t = x \right\}.$$

Here N_T is the first time that the Poisson process with parameter $1/T$ jumps. I assume that, at maturity, there is a cost of repaying the debt, which is captured by F. In this simple model that ignores refinancing, this feature is necessary to prevent the firm from issuing an extremely short bond.[5]

The first term in equation (1) is the value of earnings until the debt matures, while the second term is the residual value of the firm. The residual value of the firm is either the unlevered value minus the face value of the debt or zero in the case of default. The optimal default rule at maturity is very simple: Default if and only if

$$\frac{x_{N_T}}{r-\mu} < K,$$

that is, if the value of the assets falls short of the face value of the outstanding debt. Let the value of income that triggers a default be denoted as $\bar{x}(K)$. Thus,

$$\bar{x}(K) = (r-\mu)K.$$

Since the face value of the debt is a choice variable of the firm, the default decision is endogenous.

The appendix shows that the value of the firm, $V(x;T,K)$ can be split into two branches. The upper branch, $V_H(x,T,K)$, which corresponds to $x \geq \bar{x}(K)$, gives the value of the firm when the debt that it has issued is "in the money," while the lower branch, $V_L(x,T,K)$, reports the equity value when the debt is "out of the money"; that is, if the debt matures when earnings are in this (low) region, the firm's optimal choice is to default.

The expressions show that the value of the firm depends on the cyclical component, as captured by the value of net earnings, x. The appropriate expressions are

$$\text{(2)} \quad \begin{aligned} V_H(x,T,K) &= \frac{x}{r-\mu} - \frac{(K+F)}{1+Tr} + \bar{V}_H^1 \left(\frac{x}{\bar{x}(K)} \right)^{\lambda_1}, \text{if } x \geq \bar{x}(K), \\ V_L(x,T,K) &= \frac{Tx}{1+T(r-\mu)} + \bar{V}_L^2 \left(\frac{x}{\bar{x}(K)} \right)^{\lambda_2}, \text{if } x \leq \bar{x}(K), \end{aligned}$$

where the constants $\bar{V}_H^1$ and $\bar{V}_L^2$ depend in a very nonlinear way on the parameters of the model. The appendix provides details.

The expressions illustrate the role of default. Consider, for example, the value of the equity in this firm in "good times." The value can be split into two components. The first term,

$$\frac{x}{r-\mu} - \frac{(K+F)}{1+Tr},$$

gives the expected present discounted value of earnings minus the default-free value of the debt. The second term,

$$\bar{V}_H^1 \left(\frac{x}{\bar{x}(K)} \right)^{\lambda_1}$$

—the appendix shows that $\bar{V}_H^1$ is positive—implies that the value of the firm is higher than what the computation that ignores default would suggest: This term captures the option that the firm has to default and avoid paying the principal on the debt.

3.2 The Valuation of the Debt

The value of the bond issued by the firm is

$$\text{(3)} \quad B(x;T,K) = E\left[e^{-rN_T} R\left(x_{N_T}\right) \middle| x_t = x \right],$$

where

$$R_B\left(x_{N_T},K,\delta\right)=\begin{cases}K & \text{if } x_{N_T}\geq \bar{x}(K)\\ (1-\delta)\dfrac{x_{N_T}}{r-\mu} & \text{if } x_{N_T}<\bar{x}(K)\end{cases}.$$

Thus, if net earnings of the firm are high, bond holders receive payoff K if $x_{N_T} \geq \bar{x}(K)$, while if net earnings are low, bond holders receive the fraction $1 - \delta$ of the unlevered value of the firm.

For future reference, the value of a (T,K) bond in the *absence of default risk* is

$$B^*(T,K)=\frac{K}{1+rT}.$$

In this model, the firm and the market discount cash flows at exactly the same rate. Thus, the choice of maturity cannot be driven by different perceptions of risk as reflected in the stochastic discount factor. Moreover, the market does not charge a "term premium." The discount rate, r, is constant and independent of the maturity of the bond. Of course, in the presence of default, there will be a pseudo term premium, but this is endogenous and determined by the choices made by the firm.

3.3 Choosing the Optimal Debt Structure

The problem faced by a firm that has to raise $\bar{K}$ is

$$\max_{(T,K)} V(x,T,K)$$

subject to the constraint that the debt issued raises at least $\bar{K}$; that is,

$$B(x,T,K)\geq \bar{K}.$$

The appendix describes the functions $V(x,T,K)$ and $B(x,T,K)$.

The problem faced by the firm is fairly nonlinear, and I present some results about the optimal debt structure in the quantitative section. However, by exploring how different factors affect the price of risk, it is possible to gain intuition on how they influence the optimal choice of debt structure.

To illustrate the mechanism, I concentrate on the case in which the firm issues a bond that is in the money, that is, a bond that if matured instantaneously would be repaid.

The value of a bond in this "high" region can be written as

$$B_H(x,T,K)=B^*(T,K)\left[1-P_H(x,T,K)\right].$$

In this setting $P_H(x,T,K)$ can be interpreted as the (proportional) price of risk of a (T,K) bond issued by a firm with current earnings given by x. In the appendix, I show how $P_H(x,T,K)$ depends on specific factors. In particular, this price of risk can be decomposed as

$$P_H(x,T,K)=\Phi^H(T,K)E\left[e^{-rT_{\bar{x}(K)}}|x\right], \tag{4}$$

where $T_{\bar{x}(K)}$ is the first time that the process for net earnings hits the value $\bar{x}(K)$ given that $x > \bar{x}(K)$. Formally, this "stopping time" is given by

$$T_{\bar{x}(K)} \equiv \inf\{t : x_t = \bar{x}(K)\}.$$

Recall that if the bond matures and $x < \bar{x}(K)$, the firm will default, while if $x \geq \bar{x}(K)$, the bond will be repaid. Thus, I interpret $E\left[e^{-rT_{\bar{x}(K)}}|x\right]$ as the expected present discounted value of a dollar payable when the borrower's net earnings reach the default threshold. The higher this value, the higher the risk, as the payoff depends on the borrower reaching the default state.

Equation (4) describes the price of risk for a bond that is in the money. This price of risk, in turn, depends on the following:

- **The state of the firm, as measured by the level of current net income.** Thus, a higher value of x corresponds to a firm that needs to issue debt in "good" times, while a lower value of x captures the impact of negative cyclical conditions. In the context of this simple model, the level of current net earnings, x, captures the impact of cyclical conditions on risk.
- **The features of the economic environment.** In this model, these features are completely summarized by the interest rate, r; the mean growth rate of the firm's earnings, μ; and the volatility of the growth rate, σ. Increases in μ correspond to better growth prospects, while increases in σ are interpreted as capturing higher uncertainty in the earnings process.[6]
- **The structure of the debt as captured by the face value, K, and the expected maturity, T.**

The formulas in the appendix show that all those elements interact in a fairly nonlinear manner to determine the price of risk.

Before I describe how those factors influence the riskiness of the debt, it is useful to compute the yield to maturity, y, of a bond. It is defined as the discount rate, y, that makes the value of a (T,K) bond with a *zero default rate* equal to the market value of the debt of that same (T,K) bond when the issuer optimally defaults. Formally, when the bond is in the money (when $x \geq \bar{x}(K)$, which I view as the normal case), y satisfies

$$\frac{K}{1+yT} = B_H(x,T,K).$$

This equation is equivalent to

$$\frac{K}{1+yT} = \frac{K}{1+rT}\left[1 - P_H(x,T,K)\right].$$

Then, the excess of the yield to maturity over the risk-free interest rate is

$$y - r = \left(r + \frac{1}{T}\right)\left(\frac{P_H(x,T,K)}{1 - P_H(x,T,K)}\right).$$

Thus, there is a direct connection between the price of risk, $P_H(x,T,K)$, and the excess yield that the debt commands. Similar formulas apply to bonds that are out of the money (i.e., when $x < \bar{x}(K)$); the details can be found in the appendix.

3.4 Cyclical Effects

What is the impact of the current level of earnings on the price of risk of a given (T,K) bond? Since $P_H(x,T,K) = \Phi^H(T,K)E\left[e^{-rT_{\bar{x}(K)}}|x\right]$, the only term that depends on x is $E\left[e^{-rT_{\bar{x}(K)}}|x\right]$. Higher values of x increase the time until x reaches $\bar{x}(K)$ (recall that $x \geq \bar{x}(K)$), which is the level of net earnings at which the firm defaults. Thus, higher current earnings lower the risk of default by making the expected value of the random discount factor lower.

3.5 Economic Environment

Simple but tedious algebra show that increases in the growth prospects, μ, decrease the price of risk. There are two forces at work. First, higher growth increases the unlevered value of the firm, $V^*(x)$, for any given level of earnings, x. Since default occurs when the unlevered value of the firm is lower, a higher μ lowers the cost of default. A factor that lowers the price of risk is that net earnings "grow away" from $\bar{x}(K)$ at a faster rate the higher the value of μ. Thus, the expected value of the random discount factor increases.

Increases in σ correspond to a lower market value of the assets in case of default. The formulas in the appendix show that increases in δ are associated with higher risk prices. The intuitive argument is that a lower value of the marketable assets in case of default (higher δ) implies that the cost of default (from the point of view of the debt holder) is higher and, hence, the firm is assessed a higher price by the market.

Consider next the effect of uncertainty. It is possible to show that the price of risk is increasing in the level of uncertainty. There are two components. An increase in σ decreases the time to reach the default level (and has the opposite effect of an increase in μ). Thus, $E\left[e^{-rT_{\bar{x}(K)}}|x\right]$ increases and converges to 1 as σ increases without bound. The second component, $\Phi_H(T,K)$, increases as well, and its limit, when $\sigma \rightarrow \infty$, is also 1. Thus, higher uncertainty increases the price of risk (and the excess yield of any bond) and the price of risk converges to 1 as σ grows (the excess yield, $y - r$, goes to infinity). To sum up, the market value of debt decreases when uncertainty increases because the price of risk increases.

3.6 The Structure of the Debt

The impact of a high face value on the price of risk is pretty intuitive: Higher K increases the default threshold, $\bar{x}(K)$. This, in turn, increases the present value of the stochastic discount factor, $E\left[e^{-rT_{\bar{x}(K)}}|x\right]$, and the price of risk. The impact of longer maturities is less straightforward. Consider first a bond of very short maturity. Formally, take the limit as T goes to zero. In this case, the bond matures instantly and, given that the firm is in the no-default zone, there is no risk; the price $P_H(x,T,K)$ is zero. This bond also has $y = r$, that is, no premium. As T increases, the effects are more complicated. A longer maturity has an impact similar to a lower discount factor and pushes the term $E\left[e^{-rT_{\bar{x}(K)}}|x\right]$ up. It also increases the first component. It is possible to show that as T grows without bound, the price of risk converges to a positive number less than 1.

3.7 The Excess Yield

In the finance literature, it is standard to interpret the excess yield of a bond in terms of a "term premium"—the higher return that a longer bond has to offer—and a "risk premium"—the higher return associated with the risk of default. In this model, attempting to decompose the excess yield, $y - r$, into a term premium and a risk premium would be somewhat misleading. To see this, assume that there is no default. One way of capturing no default is to drive the current level of income to infinity. The discussion of the impact of x on the price of risk shows that the price of risk goes to zero and hence y converges to r. In this model, the market does not charge a term premium. There is a premium that responds to the maturity of the bond, but it is not a pure term premium: It critically depends on risk.

How does the excess yield of a given (T,K) bond depend on cyclical factors? Simple algebra shows that the elasticity of the excess yield with respect to x is given by

$$\frac{\partial \ln(y-r)}{\partial \ln(x)} = \frac{\lambda_1(T,\sigma^2)}{1-P_H(x,T,K)},$$

where $\lambda_1(T,\sigma^2)$ is defined in the appendix. The term $\lambda_1(T,\sigma^2)$ is negative, and it increases (becomes less negative) when σ increases and decreases when expected maturity, T, increases.

Since $P_H(x,T,K)$ is decreasing in x, the model implies that the excess yield when income is low is not only higher but also much more responsive to "news" about income. In the case of firms with low current income, small increases in income have a potentially large impact on the excess yield. At the other end, when the firm issuing the debt has a high level of current income, small changes in that level have a smaller impact on the excess yield.[7]

Uncertainty also impacts the excess yield. In this model, the (instantaneous) standard deviation of the growth rate of income is given by σ. Since higher σ increases $P_H(x,T,K)$, it, in turn, increases the excess yield. The model implies that firms that operate in more-uncertain environments will have to offer a higher return for similar bonds.[8] Note that this implication is restricted to a given bond, that is, for a fixed (T,K). It does not necessarily imply that higher volatility and observed yields are positively correlated, since a firm that faces a more uncertain environment can choose a different structure of its debt, that is, a different (T,K) bond.

3.8 Environment Changes and the Structure of Debt

In this section, I offer a heuristic discussion that suggests how changes in the environment affect the structure of debt. The next section uses a quantitative version of the model to display the impact of changes in the economic environment faced by the firm.

It is useful to consider the situation of a borrower that has chosen the debt structure optimally—that is, it has chosen a bond (T,K) that maximizes the equity value and raises the necessary amount of resources—and that, unexpectedly, finds itself in a different environment. In this case, I ask, what is the likely response of the firm to the change in conditions—that is, what is the "new" (T,K) bond the firm would like to issue.

To get some intuition about the forces at work, it is useful to describe the connection between the structure of the debt and the debt's market value in elasticity form. The appropriate expressions are (with the arguments omitted to keep the expressions simple)

$$\frac{\partial B_H}{\partial K}\frac{K}{B_H}=1+\frac{\lambda_1 P_H}{1-P_H}$$

and

$$\frac{\partial B_H}{\partial T}\frac{T}{B_H}=\frac{T}{1+rT}\left[-1-\frac{\partial P_H/\partial T}{1-P_H}\right].$$

Consider first an improvement in current income. It lowers the price of risk, P_H, and increases the value of the "existing" debt. Thus, the firm that has a fixed demand for funds needed to restructure the debt to lower the debt's value to the level it was before the increase in income. This can be done by lengthening the maturity of T or increasing the face value of K. The previous expressions show that longer maturity always lowers the value of the debt, while a higher face value lowers the value of the debt only when net income is low since, as x goes to infinity, the impact of higher K is positive. Thus, we tentatively expect higher x to result in longer maturity and a higher face value.

Consider next the impact of an unexpected increase in uncertainty. It lowers the market value of the existing debt, and, in this case, the firm needs to adjust the structure of the debt, (T,K), so as to increase the debt's value. This can be accomplished by shortening the maturity and decreasing the face value.

As it is clear from the equations, the impact of features of the environment on the valuation of the firm and the bond is fairly nonlinear. Instead of a somewhat boring derivation of the impact, I report the results from a quantitative exercise.

4 QUANTITATIVE MODEL

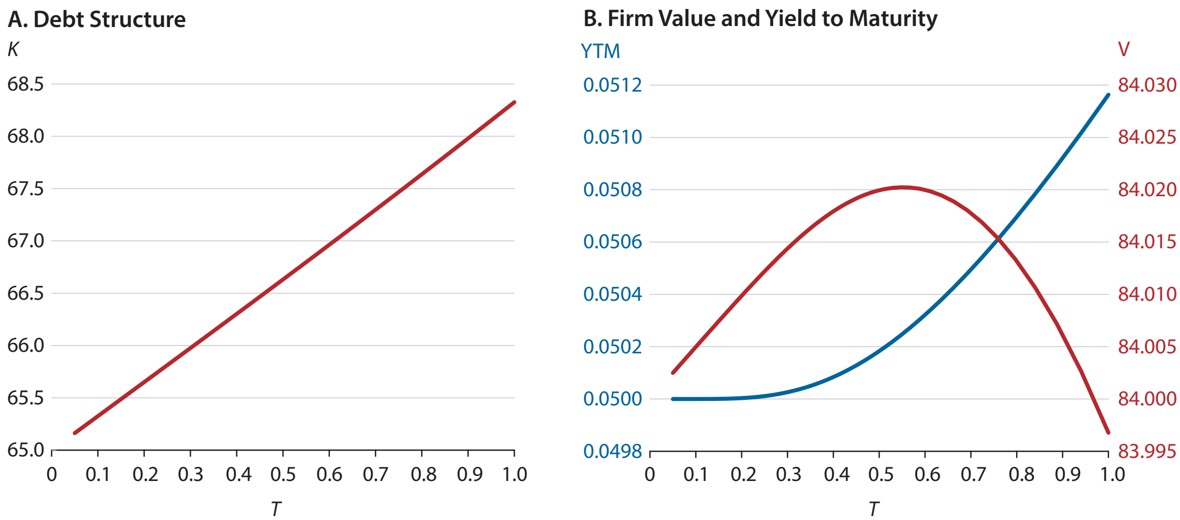
In this section, I report the implications of the model for the expected maturity of the debt, as measured by T as a function of the cyclical state, as captured by x, and other features of the environment that do not change with the cyclical component.

4.1 Calibration

The purpose of the model is to illustrate how the different features of the economic environment interact to determine optimal (from the point of view of the borrower) debt maturity. To this end, I chose a reasonable calibration that is not meant to mimic the U.S. situation but that produces reasonable values of the variables of interest. I assume that the interest rate is 5 percent and that the expected growth rate of the firm's income is 1 percent. The standard deviation of the growth rate is 20 percent, which is close to the standard deviation of a broad index of the stock market. In the base case, bondholders recover about 60 percent of the value of the firm in the case of default ($\delta = 0.4$) and the fixed cost, F, is set to 5. I consider a firm that has a financing need of $\bar{K} = 65$.

Figure 1

Debt Structure, Firm Value, and Yield to Maturity (Low *x*)



NOTE: YTM, yield to maturity; V, firm value.

SOURCE: Author's calculations.

4.2 A First Look at the Options

Before I discuss the optimal debt structure, it is useful to see the set of options that the firm had available when choosing a particular bond. Panel A of Figure 1 contains the set of pairs (T,K) that raise a given amount of revenue (in this case, $\bar{K} = 65$) when x is relatively low $(x = 6)$. As expected, if the firm chooses to issue longer debt, it has to offer a higher face value. Panel B shows the consequences of those possibilities for the value of the firm and the yield to maturity.

For this particular (low) value of net income, the optimal maturity is about 5.5 months. It is clear that the firm had the option of issuing longer (or shorter) debt but chooses not to. Moreover, by choosing shorter debt, the firm could have lowered the excess yield of the debt issued. However, value maximization does not imply excess yield minimization.

Figure 2 has similar information but for a higher level of earnings.

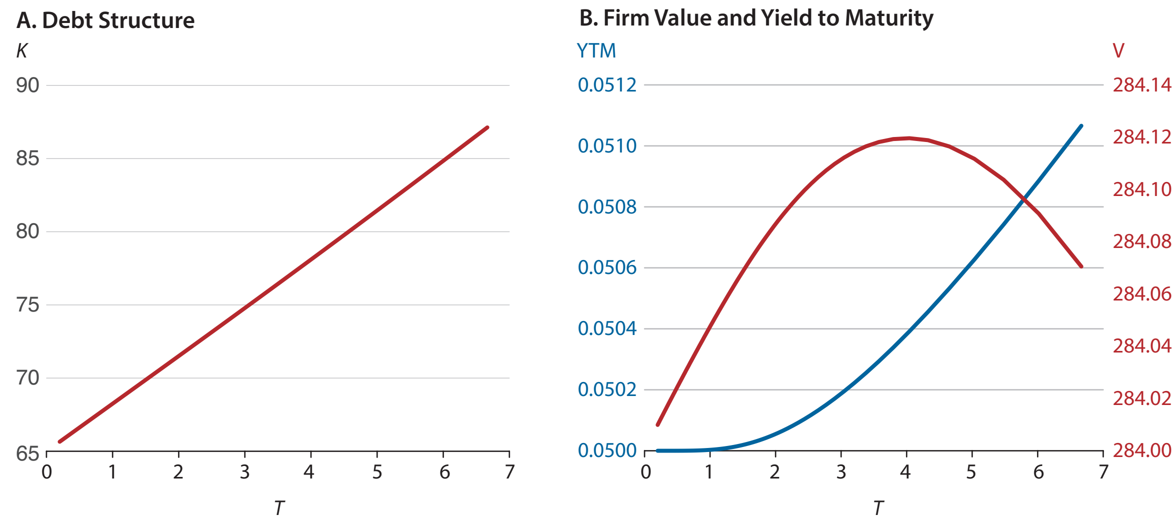
In this case, the optimal (in the sense of value maximization) maturity is close to four years. As in the previous case, the market was willing to purchase debt of different maturities provided the firm was willing to offer the necessary excess return.

What does an increasing yield as a function of maturity say about how face value changes with maturity? The yield to maturity of a bond that has a market value equal to $\bar{K}$ is

$$y(T) = \left(\frac{K(T)}{\bar{K}} - 1\right)\frac{1}{T},$$

Figure 2

Debt Structure, Firm Value, and Yield to Maturity (High *x*)



NOTE: YTM, yield to maturity; V, firm value.

SOURCE: Author's calculations.

where $K(T)$ depends on $(\bar{K},x)$ as well as all other variables that have an effect on the value of the bond. If the function $y(T)$ is differentiable, then it is increasing in maturity if and only if

$$K(T) < \bar{K} + K'(T)T.$$

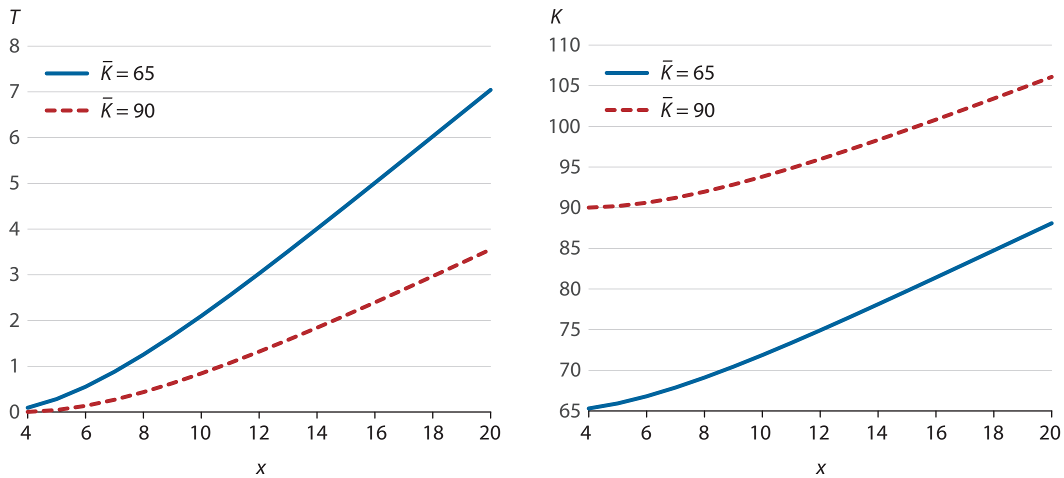
Since the function $K(T)$ must have the property that $\lim_{T\to 0} K(T) = \bar{K}$, the previous condition shows that the face value must be a concave function of maturity. Thus, the ratio of the face value to maturity is decreasing in the average maturity.

The model also illustrates the difficulties of using observed values of yields to estimate the term premium. As mentioned before, there is no "pure" term premium in this model. However, suppose that one had data on just the two bonds that are optimally chosen by two firms that are identical except for the level of current earnings. Both bonds raise exactly the same amount, $\bar{K}$, but have different maturities. Since the model is not calibrated, it implies very small excess yields. To make the argument transparent, I use the difference between the excess yield of a long-term bond with a 4-year maturity and a short-term bond with a half-year maturity.

I define the "observed" term premium as the difference between the excess yields for the bonds that are actually issued. In this example, this difference is (after adjusting units) 13. I can now ask, what would have been the excess yield if the low-net-income firm (i.e., when $x = 6$) had chosen to issue a 4-year bond instead of a half-year bond? In this example, the equivalent value is 70. The same calculation for the high-x firm yields 38. Thus, I conclude that, for these examples, what could be called the true "term structure" is steeper the lower the current level

Figure 3

Leverage and Optimal Debt Structure



SOURCE: Author's calculations.

of income and that the *observed* term structure—which reflects optimal maturity choice—is much flatter.

4.3 Optimal Debt Structure: Cyclical Effects

What is the optimal debt structure as defined by face value, K, and expected maturity, T? What are the factors that influence the choice of debt structure by a firm? I first look at what I view as a cyclical component, that is, how the optimal debt structure changes as net earnings change. The intuitive interpretation is that higher earnings are associated with good times (and in this model larger firms).

Figure 3 reports for average maturity (Panel A) and face value (Panel B) the optimal debt structure (T,K) as a function of net income, x, for two levels of financing requirements (i.e., $\bar{K}$): 65 and 90. For reference, when x is less than 10, the resulting leverage is close to 100 percent, while when x is equal to 10, the implied leverage decreases to 50 percent. Overall, I view the results of when x is less than 10 as capturing the optimal debt structure for highly indebted firms (or countries) and when x is greater than 10 as being more representative of the case in which financing needs are low.

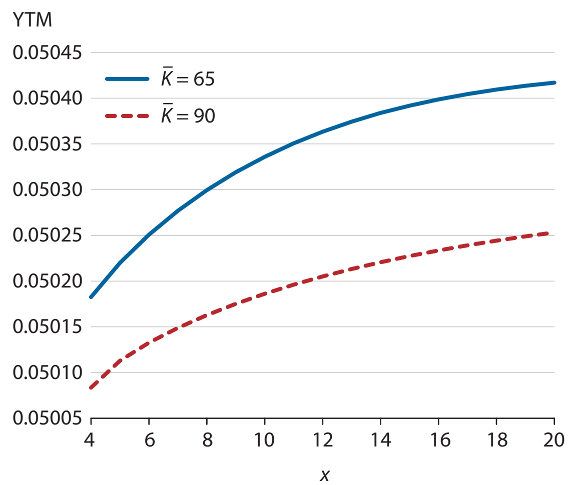
As expected, firms with better prospects—as measured by their current earnings x—issue longer debt. This finding is consistent with the previous discussion that showed that higher current earnings—which in this model also correspond to higher future expected earnings—reduce the cost of default and allow a firm to postpone paying the cost of repaying the debt. The results show that a higher level of leverage—captured here by a higher value of funds that the firm has to raise—is associated with shorter maturity. The difference in leverage—as captured by the difference between 65 and 90—also has a large impact on optimal maturity when net earning, x, are high, but the impact tends to disappear as net earning become low. The results suggest that firms with low levels of earnings choose short maturities even when the market value of the debt they issue varies significantly.

Since higher x is associated with longer maturities, the face value of the bond must increase as well. This is illustrated in Panel B of Figure 3.

Figure 4 shows the yield to maturity of debt issued by a firm. Note that the yield is increasing in x. As I discussed in the previous section, at each point x, the firm has multiple options in terms of debt structure—combinations of (T,K)—that raise the required funds but differ in yield to maturity. The results show that the firm does not choose the debt structure that minimizes the excess yield. At any level of net earnings, the market is willing to lend to the firm at a lower premium if the firm chooses debt with a shorter maturity. The optimal choice—the choice that maximizes the value of the firm—does not minimize the excess yield. As the situation of the firm improves (x increases), the firm chooses to issue longer debt and to pay a higher return on that debt.

Figure 4

Leverage and Yield to Maturity



NOTE: YTM, yield to maturity.

SOURCE: Author's calculations.

One could easily merge the data from Figures 3 and 4 to create a market-based "term structure" but, as discussed in the previous section, this cannot be used to estimate the costs of debt of different maturities issued by the same firm. In general, I expect that the term-structure curve faced by a firm—at, say, a given x—to be steeper than the one that emerges from looking at the optimally chosen debt structure.

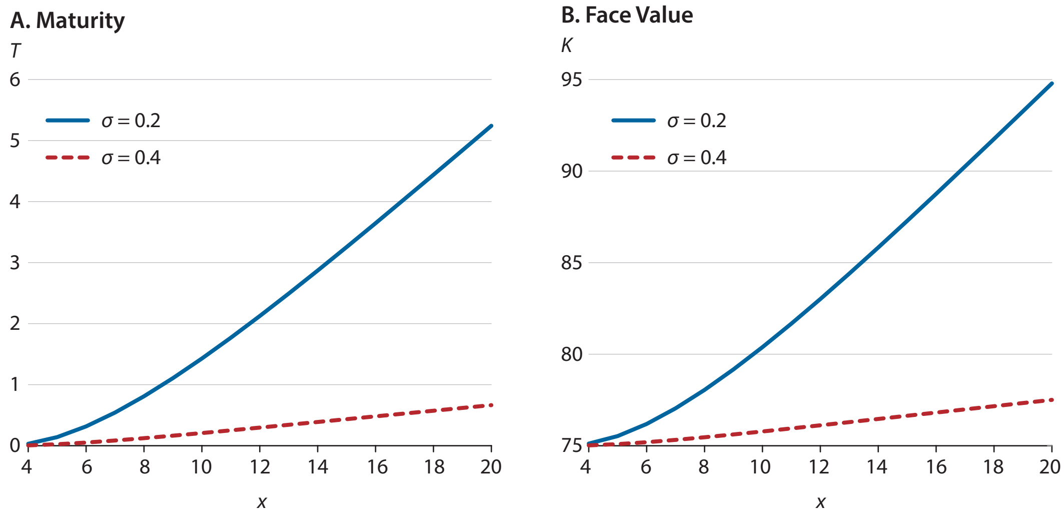
4.4 Optimal Debt Structure: The Impact of Uncertainty

Figure 5 presents the results on the impact of uncertainty, σ, on average maturity, T (Panel A), and face value, K (Panel B).

As in the case of differences in financing needs, the responses are fairly nonlinear. Borrowers that operate in more-volatile environments find it optimal to issue shorter debt and, somewhat mechanically, debt with lower face value. For example, at low levels of net earnings (say for $x < 6$), firms that operate in a low-uncertainty environment (i.e., $\sigma = 0.2$, the blue line) and firms that operate in a high-uncertainty environment (i.e., $\sigma = 0.4$, the red line) both issue very short-term debt (maturity less than a year). However, for high levels of net earnings (which also have low levels of leverage), while firms in the low-uncertainty environment issue debt with an average maturity of over five years, firms in the high-uncertainty environment

Figure 5

Uncertainty and Optimal Debt Structure



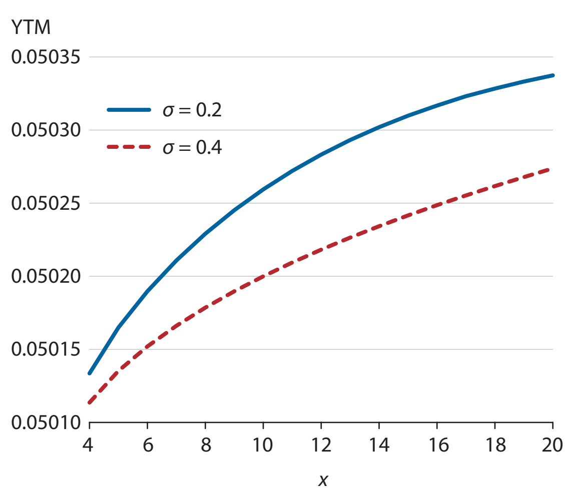
SOURCE: Author's calculations.

still issue debt with an average maturity of less than a year. Put differently, maturity responds more elastically to changes in net income at low levels of uncertainty. When volatility is high, these results suggest that the cyclical component (the value of x) has a small impact on maturity.[9]

Figure 6 shows that, somewhat surprisingly, debt issued by the borrower in the low-uncertainty environment has a higher yield. Thus, uncertainty and excess yield move in the opposite directions. To understand what drives this result, consider the forces at work when volatility increases. On the one hand, higher uncertainty increases risk and tends to increase yields. However, the firm with more-uncertain revenue issues shorter debt and this second effect—which moves yields in the opposite direction—dominates. Thus, it is the endogenous choice of the borrower that determines the yield.

Figure 6

Uncertainty and Yield to Maturity

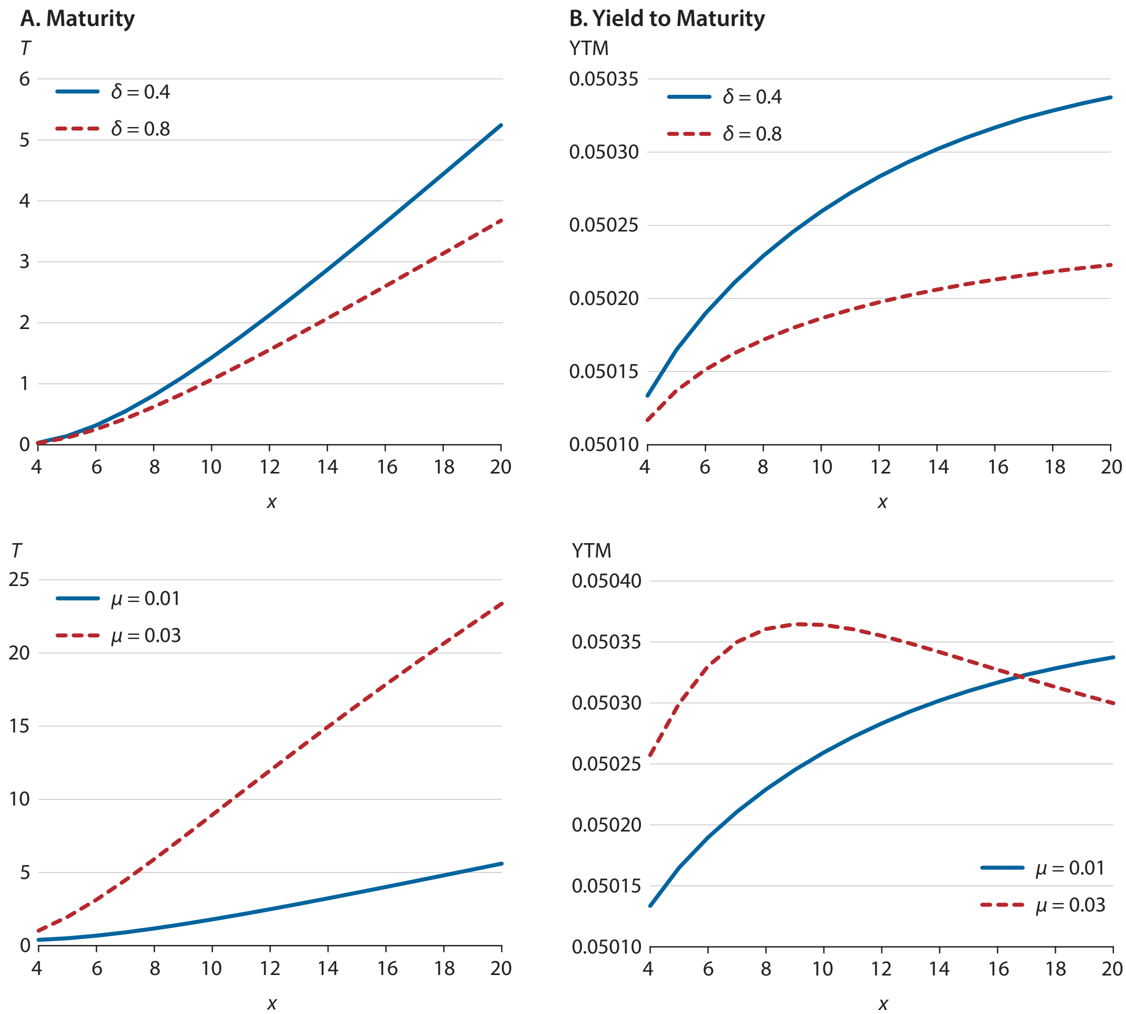


NOTE: YTM, yield to maturity.

SOURCE: Author's calculations.

Figure 7

Asset Saleability, Growth, and the Optimal Debt Structure



NOTE: YTM, yield to maturity.
SOURCE: Author's calculations.

4.5 Optimal Debt Structure: Collateral and Growth Prospects

The impact of the value of collateral and growth prospects on the average maturity and yield are summarized in the four panels of Figure 7. Higher collateral results is longer debt due to the lower cost of a default. As in the previous case, maturity is the dominating force because the yield is also higher (even though the risk is lower).

The impact of higher growth is as expected: Firms with better growth prospects issue longer debt. The implication for yield to maturity is not monotone. There are two forces at work: Higher growth lowers the risk of default, which works in the direction of lowering the yield demanded by the market. However, high-growth firms issue longer-term debt, which works in the opposite direction. For relatively low values of net earnings, the second effect dominates, but for high values of x, the lower probability of default is the dominant force because both types of firms (high and low growth) optimally choose to issue bonds with the same maturity.

5 CONCLUSION

In this article, I presented a very simple model that sheds light on some of the factors that influence the choices of debt structure, maturity, and face value, by borrowers that have a stochastic earnings stream. The model is designed so that financial markets and borrowers share exactly the same discount factor and, to prevent strategic forces that pitch holders of different maturities against each other, I look at the case of a single bond. Because of the simplicity of the model, the results are suggestive, but more work needs to be done before taking the model to the data.

Several findings are broadly consistent with the evidence:

(i) The higher the level of income, the longer the maturity of the debt. In the model, this captures the tradeoff between longer debt (that increases the price of risk) and short debt (that incurs a payment cost that can be delayed).

(ii) Firms in more-uncertain environments issue shorter-term debt. However, higher uncertainty does not translate into a higher price of risk, as measured by the excess yield to maturity over the (risk free) rate. The reason is that, faced with higher uncertainty, firms issue shorter-term debt, which lowers the return demanded by the market. Thus, it would be erroneous to use yields (say, relative to the risk-free rate) when estimating default risk since in this simple model there is an endogenous "term premium like" component of the yield as well.

(iii) Better collateral and higher growth prospects result in firms choosing longer maturities. The impact on the yield to maturity is ambiguous. In general, better collateral is associated with a higher yield to maturity because of the longer debt. Higher growth is associated with higher yields for firms with low income, but lower yields for firms with high income.

This simple model suggests that simplistic policy recommendations of the form "it is always better to borrow long" fail to understand that, even in situations in which it is possible to borrow long, the firm may choose to borrow short because longer maturities have associated prices of risk that are endogenous. More relevant to interpreting both the evidence on firm borrowing and sovereign borrowing is that when income is low, firms/countries choose shorter debt—and not the other way around. ■

APPENDIX

In this appendix, we derive the values of the firm and the debt as given by equations (2) and (3).

A.1 The Value of the Firm

The problem of the firm has an associated Hamilton-Jacobi-Bellman equation given by

$$rV(x;T,K)=x+\frac{\partial V}{\partial x}\mu(x)+\frac{\partial^2 V}{\partial x^2}\frac{\sigma^2(x)}{2}+$$
$$\frac{1}{T}\left[\max\left(\frac{x}{r-\mu}-(K+F),0\right)-V(x,\theta)\right].$$

It is standard to show that the solution to the previous equation is of the form

$$V(x,T,K)=\begin{cases}V_H(x,T,K) & x\geq\bar{x}(K)\\ V_L(x,T,K) & x\leq\bar{x}(K)\end{cases}.$$

The value of the firm is

$$V_H(x,T,K)=\frac{x}{r-\mu}-\frac{(K+F)}{1+rT}+\bar{V}_H^1\left(\frac{x}{\bar{x}(K)}\right)^{\lambda_1},$$
$$V_L(x,T,K)=\frac{Tx}{1+(r-\mu)T}+\bar{V}_L^2\left(\frac{x}{\bar{x}(K)}\right)^{\lambda_2},$$

where $\lambda_1(\lambda_2)$ is the negative (positive) root of

$$r+\frac{1}{T}-\lambda\mu=\frac{\sigma^2}{2}\lambda(\lambda-1)$$

and the constants are given by

$$\bar{V}_H^1=\frac{1}{\lambda_2-\lambda_1}\left[(\lambda_2-1)\frac{T(r-\mu)K}{1+T(r-\mu)}+\frac{TK}{1+Tr}\left(r+\frac{1}{T}-\lambda_2 r\right)+\frac{\lambda_2 F}{1+Tr}\right],$$
$$\bar{V}_H^1=\frac{1}{\lambda_2-\lambda_1}\left[(\lambda_1-1)\frac{T(r-\mu)K}{1+T(r-\mu)}+\frac{TK}{1+Tr}\left(r+\frac{1}{T}-\lambda_1 r\right)+\frac{\lambda_1 F}{1+Tr}\right].$$

A.2 The Market Value of the Debt

The market value of a (T,K) bond depends not only on its maturity but also on its default probability. The relevant Hamilton-Jacobi-Bellman equation whose solution is the market value of the bond is

$$rB(x;T,K)=b+\frac{\partial B}{\partial x}\mu(x)+\frac{\partial^2 B}{\partial x^2}\frac{\sigma^2(x)}{2}+T\left[R_B(x,K,\delta)-B(x,\theta)\right],$$

where

$$R_B(x,K,\delta)=\begin{cases} K & \text{if } x \geq \bar{x}(K) \\ \dfrac{(1-\delta)x}{r-\mu} & \text{if } x < \bar{x}(K) \end{cases}.$$

One can show that the solution to this equation can be described as simple functions that vary depending on which "branch" the value of net earnings is. Formally, the solution is of the form

$$B(x,T,K)=\begin{cases} B_H(x,T,K) & x \geq \bar{x}(K) \\ B_L(x,T,K) & x \leq \bar{x}(K) \end{cases}.$$

The functions $B_H(x,T,K)$ and $B_L(x,T,K)$ are given by

$$B_H(x,T,K)=\frac{K}{1+Tr}+\bar{B}_H^1\left(\frac{x}{\bar{x}(K)}\right)^{\lambda_1}, \text{ for } x \geq \bar{x}(K),$$

and

$$B_L(x,T,K)=\frac{(1-\delta)}{1+T(r-\mu)}\frac{x}{r-\mu}+\bar{B}_L^2\left(\frac{x}{\bar{x}(K)}\right)^{\lambda_2}, \text{ for } x \leq \bar{x}(K).$$

The constants are

$$\bar{B}_H^1=\frac{K}{1+Tr}\frac{1}{\lambda_2-\lambda_1}\left[\frac{\lambda_2(\mu T-\delta(rT+1))-(rT+1)(1-\delta)}{1+T(r-\mu)}\right]$$

$$\bar{B}_L^2=\frac{K}{1+Tr}\frac{1}{\lambda_2-\lambda_1}\left[\frac{\lambda_1(\mu T-\delta(rT+1))-(rT+1)(1-\delta)}{1+T(r-\mu)}\right].$$

A.3 The Price of Risk

The definition of the price of risk associated with a bond is the value of P that accounts for the difference between the market price of a bond, B, and the price in the absence of risk, B^*. Formally,

$$\frac{B(x,T,K)}{B^*(T,K)}=(1-P(x,T,K)).$$

Since the market value of a bond depends on whether it is in the money (i.e., $x \geq \bar{x}(K)$) or out of the money (i.e., $x < \bar{x}(K)$), there are two corresponding prices of risk: P_H and P_L, respectively. Using the previous results on bond prices, it follows that

$$P_H(x,T,K)=\Phi^H(T,K)\left(\frac{x}{\bar{x}(K)}\right)^{\lambda_1(T,\sigma^2)},$$

where

$$\Phi^H(x,T,K)=\frac{\phi_N\left(\lambda_2;T,\sigma^2\right)}{2\phi_D\left(T,\sigma^2\right)}$$

and

$$\phi_D\left(T,\sigma^2\right)=\left[\left(\frac{1}{2}-\frac{\mu}{\sigma^2}\right)^2+2\frac{(1+rT)}{T\sigma^2}\right]^{1/2},$$

$$\phi_N\left(\lambda_2;T,\sigma^2\right)=\frac{(1+rT)\left(1-\delta+\lambda_2\left(T,\sigma^2\right)\delta\right)-\lambda_2\left(T,\sigma^2\right)\mu T}{1+(r-\mu)T}$$

and

$$\lambda_2\left(T,\sigma^2\right)=\left(\frac{1}{2}-\frac{\mu}{\sigma^2}\right)+\phi\left(T,\sigma^2\right)>1$$

and

$$\lambda_1\left(T,\sigma^2\right)=\left(\frac{1}{2}-\frac{\mu}{\sigma^2}\right)-\phi\left(T,\sigma^2\right)<0.$$

It is possible to show that

$$\left(\frac{x}{\bar{x}(K)}\right)^{\lambda_1\left(T,\sigma^2\right)}=E\left[e^{-rT_{\bar{x}(K)}}|x\right].$$

This implies that

$$P_H(x,T,K)=\Phi^H(T,K)E\left[e^{-rT_{\bar{x}(K)}}|x\right],$$

which is equation (4) in the text.

In the case of a bond that is out of the money, $x\leq\bar{x}(K)$, the price of risk is given by

$$P_L(x,T,K)=1-\underbrace{\frac{(1-\delta)x}{(r-\mu)K}\frac{1+rT}{1+(r-\mu)T}}_{+}+\frac{\phi_N\left(\lambda_1;T,\sigma^2\right)}{2\phi_D\left(T,\sigma^2\right)}\left(\frac{x}{\bar{x}(K)}\right)^{\lambda_2\left(T,\sigma^2\right)}.$$

NOTES

[1] It turns out that allowing for more general debt that includes coupon payments does not expand the set of options relative to the pure discount bond.

[2] This is the mechanism in Broner, Lorenzoni, and Schmukler (2013).

[3] Most of the empirical literature on determinants of debt maturity uses data on publicly traded companies. Ideally, one would like to analyze data on all firms, but available balance sheet data on non-traded firms are very limited.

[4] This is quite different from the risk-aversion mechanism in Broner, Lorenzoni, and Schmukler (2013).

[5] Manuelli and Sanchez (2019) study a more general version of the model with different regimes, more general debt instruments, and refinancing.

[6] Even though in this simple version uncertainty is exogenous, it is possible to interpret a higher level of σ as driven in part by uncertainty about government policies.

[7] Formally, the elasticity converges to $\lambda_1(T,\sigma^2)$ as x goes to infinity.

[8] The uncertainty can be either the result of the type of industry that the firm operates in or the consequence of government policies that result in higher variability in the growth rate of a firm's earnings.

[9] To the extent that volatility characterizes a given market, the results imply that a cross section of the expected maturity of the debt issued by firms, controlling for the amount raised, does not vary much in highly volatile markets and the opposite is true in more-stable markets.

REFERENCES

Aguiar, M.; Amador, M.; Hopenhayn, H. and Werning, I. "Take the Short Route: Equilibrium Default and Debt Maturity." *Econometrica*, March 2019, *87*(2), pp. 423-62; https://doi.org/10.3982/ECTA14806.

Arellano, C. and Ramanarayanan, A. "Default and the Maturity Structure in Sovereign Bonds." Staff Report 410, Federal Reserve Bank of Minneapolis, 2013.

Benmelech, E. and Dvir, E. "Does Short-Term Debt Increase Vulnerability to Crisis? Evidence from the East Asian Financial Crisis." NBER Working Paper No. 17468, National Bureau of Economic Research, 2011; https://doi.org/10.3386/w17468.

Broner, F.; Lorenzoni, G. and Schmukler, S. "Why Do Emerging Economies Borrow Short Term?" *Journal of the European Economic Association*, January 2013, *11*(Issue Suppl. 1), pp. 67-100; https://doi.org/10.1111/j.1542-4774.2012.01094.x.

Brunnermeier, M. and Oehmke, M. "The Maturity Rat Race." *Journal of Finance*, 2013, *68*(2), pp. 483-521; https://doi.org/10.1111/jofi.12005.

Chen, H.; Cui, R.; He, Z. and Milbradt, K. "Quantifying Liquidity and Default Risks of Corporate Bonds over the Business Cycle." Unpublished manuscript, 2017; https://doi.org/10.1093/rfs/hhx107.

Diamond, D. "Debt Maturity and Liquidity Risk." *Quarterly Journal of Economics*, 1991, *106*(3), pp. 709-37; https://doi.org/10.2307/2937924.

Dvorkin, M.; Sanchez, J.; Sapriza, H. and Yurdagul, E. "News, Sovereign Debt Maturity, and Default Risk." Working Paper No. 2018-033, Federal Reserve Bank of St. Louis, 2018; https://doi.org/10.20955/wp.2018.013.

Kacperczyk, M. and Schnabl, P. "When Safe Proved Risky: Commercial Paper During the Financial Crisis of 2007-2009." *Journal of Economic Perspectives*, 2010, *24*(1), pp. 29-50; https://doi.org/10.1257/jep.24.1.29.

Leland, H. and Toft, K. "Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads." *Journal of Finance*, 1996, *51*(3), pp. 987-1019; https://doi.org/10.1111/j.1540-6261.1996.tb02714.x.

Manuelli, R. and Sanchez, J. "Endogenous Debt Maturity: Liquidity Risk vs Default Risk." Unpublished manuscript, 2019; https://doi.org/10.20955/wp.2018.034.

Rodrik, D. and Velasco, A. "Short-Term Capital Flows." NBER Working Paper No. 7364, National Bureau of Economic Research, 1999; https://doi.org/10.3386/w7364.

How Have Banks Been Managing the Composition of High-Quality Liquid Assets?

Jane Ihrig, Edward Kim, Cindy M. Vojtech, and Gretchen C. Weinbach

Banks' liquidity management practices are fundamental to understanding the implementation and transmission of monetary policy. Since the Global Financial Crisis of 2007-09, these practices have been shaped importantly by the liquidity coverage ratio requirement. Given the lack of public data on how banks have been meeting this requirement, we construct estimates of U.S. banks' high-quality liquid assets (HQLA) and examine how banks have managed these assets since the crisis. We find that banks have adopted a wide range of HQLA compositions and show that this empirical finding is consistent with a risk-return framework that hinges on banks' aversion to liquidity and interest rate risks. We discuss how various regulations and business model choices can drive HQLA compositions in general, and connect many of the specific compositions we see to banks' own public statements regarding their liquidity strategies. Finally, we highlight how banks' preferences for the share of HQLA met with reserves affect the Fed's monetary policy implementation framework. (JEL E51, E58, G21, G28)

Federal Reserve Bank of St. Louis *Review*, Third Quarter 2019, *101*(3), pp. 177-201.
https://doi.org/10.20955/r.101.177-201

1 INTRODUCTION

Liquidity management—ensuring access to sufficient quantities of assets that can be converted easily and quickly into cash with little or no loss of value—has always been a key component of banks' balance sheet management. However, liquidity management has become an even more important consideration in banks' operations in the wake of the Global Financial Crisis of 2007-09 with the introduction of new regulations aimed at ensuring banks' ability to meet their cash and collateral obligations during times of financial stress. In particular, beginning in 2015, large banks in the United States have needed to comply with the liquidity coverage ratio (LCR) by holding sufficient "high-quality liquid assets" (HQLA), a requirement

Jane Ihrig is an economist and senior advisor, Edward Kim was a research assistant, Cindy M. Vojtech is a principal economist, and Gretchen C. Weinbach is an economist and senior associate director at the Board of Governors of the Federal Reserve System. The authors thank Roop Ambardekar for his initial help with the data, Akber Khan for his excellent research assistance, and Ashish Kumbhat for his input on an earlier version of this article. They also thank seminar participants at the Board of Governors of the Federal Reserve System, the Federal Reserve Bank of New York, and the MBAA International 2018 conference.

that has induced significant changes to banks' balance sheet management. In this article, we examine how U.S. banks have managed the *composition* of their HQLA to meet the LCR and other liquidity considerations over the past several years.[1] In particular, we address the following questions: Which particular liquid assets have banks chosen to hold and in what shares? Have those liquid shares changed over time? Do banks' preferences for these liquid shares vary? If so, what factors may be driving banks' preferences in this regard?

Understanding banks' liquidity management is central to both the implementation and the transmission of monetary policy. Banks' preferences regarding the compositions of their liquid pools—and in particular their demand for excess reserve balances—interact with various short-term market interest rates and thus the Federal Reserve's setting of its administered rates. In turn, the constellation of these interest rates affects banks' choices regarding the composition of their balance sheets—that is, the trade-off between lending and holding liquid assets—and hence the transmission of monetary policy (Bianchi and Bigio, 2017). In addition, we and others have noted that banks' preferences regarding the compositions of their liquid assets have influenced the Federal Open Market Committee's determination of "ample" reserves needed to effectively and efficiently implement monetary policy in the longer run; ultimately, these preferences affect the long-run size of the Federal Reserve's balance sheet.[2]

Because no historical time series of HQLA are available, we begin our analysis by constructing bank-level, quarterly estimates of HQLA from 1997 to the present. We focus on the three largest components of HQLA—banks' reserve balances held at the Federal Reserve and their holdings of both Treasury securities and certain mortgage-backed securities (MBS). We use our bank-level estimates to document how U.S. banks have managed the compositions of their HQLA pools over time. Not surprisingly given the Federal Reserve's large-scale asset purchases (LSAPs) that injected reserves into the banking system at the time, we find that during the run-up to becoming LCR compliant, banks in aggregate took on a significant quantity of excess reserves. However, after becoming compliant, many such banks adjusted their liquid asset holdings, reducing their stocks of reserve balances and raising their holdings of other HQLA components, presumably to achieve a more optimal configuration.

To explain this subsequent compositional adjustment, we use a risk-return framework that captures the relative return of the different HQLA components, the covariance of these returns, and the sensitivity of banks' preferences for these assets to a metric of their risk aversion. One can think of this model as capturing the decisions of a bank's treasury department. The bank treasurer oversees investments in various securities and cash instruments while providing for the bank's daily liquidity needs and meeting regulatory constraints such as the LCR. In particular, in managing the bank's liquid assets, the treasurer considers liquidity risk—the risk that cash is not immediately available when needed—and interest rate risk—in this case, the risk that the value of a liquid asset will change due to a change in interest rates. Interest rate risk is a particular concern for holders of fixed-income securities such as Treasury securities. The more risk averse the bank is in this context, the more its treasury department tilts its bank's HQLA composition toward cash—that is, the more the bank prefers holding a relatively high share of reserve balances in its HQLA pool, helping to insulate it from both liquidity and interest rate risks.

With this model in hand, we then look at individual, bank-level data for the eight U.S global systemically important banks (GSIBs) based on publicly available sources. We find that the HQLA compositions of these large institutions differ widely, with some institutions relying on reserves and others more so on longer-term assets. This result holds even for some institutions with similar business models. Digging deeper with daily, bank-reported confidential HQLA data, we find that the volatilities of these banks' HQLA shares differ. In particular, we find that banks that exhibit a relatively higher level of daily volatility in their cash balances tend to rely more heavily on reserves to meet their liquidity needs.

We conclude that banks have different tolerances for exposure to liquidity and interest rate risks, and, based on research into individual banks' own descriptions of their liquidity management strategies, we identify a number of factors influencing their strategies. These factors include their individual evaluations of how they choose to meet post-crisis financial regulations as well as a range of other influences including differing business models, products and services, and unobservable, internal operating procedures. Where possible, we back up our empirical findings with specific public statements. For example, large banks have cited resolution planning requirements—which rely on bank-dependent stress-test models—as a key driver of their reserves share of HQLA.

The remainder of this article proceeds as follows. Section 2 positions our work in the related literature. Section 3 provides background information about the LCR and HQLA and describes the two different sets of time-series data that we construct. Using those data, Section 4 describes how banks initially adjusted, and then subsequently managed, the compositions of their HQLA. Section 5 introduces our model, and Section 6 explores our bank-level estimates for the eight GSIBs, including the ranges of both the sizes and volatilities of these banks' HQLA shares. Section 7 highlights key factors that drive banks' decisionmaking regarding their HQLA shares and summarizes the differences evident across the GSIBs. Section 8 concludes.

2 RELATED LITERATURE

Given that international-based regulatory liquidity requirements are relatively new for the banking industry, research regarding how banks are adjusting to these post-crisis regulations is nascent and growing. For example, Allen (2014) and Diamond and Kashyap (2016) survey the existing literature and generally conclude that more research is needed to understand the effects of liquidity regulation on banks' behavior.

That said, our work is complementary to a few studies. Banerjee and Mio (2015) describe their work as being the first study to estimate the causal effect of liquidity regulation on banks' balance sheets. They examine the impact on banks' balance sheets of the implementation of a new LCR-like requirement in the United Kingdom and find that U.K. banks adjusted the compositions of both their assets and liabilities at the onset of the new requirement, including increasing HQLA. Our findings are similar in that we document an increase in U.S. banks' holdings of HQLA in response to new U.S. liquidity regulation. Our study differs in that we also consider how banks subsequently managed the compositions of their liquid assets once initially becoming regulatory compliant.

Cetina and Gleason (2015) use examples to show how compositional asset and liability caps in the U.S. LCR rule introduce nonlinearities in the calculation of the ratio and conclude that banks' LCRs can vary in complex ways not necessarily related to underlying liquidity risk. In our analysis, we also account for the nonlinearity of the LCR's asset caps in our model estimates. However, our work differs from that study in that we focus on the compositional shares of HQLA and abstract from issues associated with the computation of the LCR itself.

Balasubramanyan and VanHoose (2013) use a theoretical model to examine how the LCR is likely to affect banks' balance sheet dynamics and conclude that the LCR may generate bank responses that are not necessarily fully consistent with a policy objective of greater stability of bank deposits and loans. Our study differs in that we focus on a different slice of banks' balance sheets—HQLA portfolios—and explain how such holdings have evolved in light of the LCR and why that may be the case.

Other LCR-related research includes theoretical work that explores the interaction between liquidity regulation and monetary policy (Bech and Keister, 2017, and Duffie and Krishnamurthy, 2016) and dynamic general equilibrium models that explore the interactions between banks' responses to liquidity and capital regulations (Adrian and Boyarchenko, 2013, and Covas and Driscoll, 2014). Our article adds to the literature on theoretical approaches to studying the effects of liquidity regulation on bank behavior by using a risk-return framework to help motivate banks' preferences for the compositions of their HQLA portfolios.

Next we review key aspects of the U.S. LCR requirement, including why it was implemented, and then describe how we compute our time-series estimates of bank-level HQLA.

3 BACKGROUND AND DATA CONSTRUCTION

During the Global Financial Crisis of 2007-09, substantial stress in U.S. funding markets—illiquidity—caused solvency issues for several large financial institutions. With financial markets quite fragile, funding shocks easily spread across the financial system. In the wake of the crisis, international financial regulators sought to improve the resiliency of the financial system by incorporating liquidity requirements into the international Basel III framework for enhanced regulation of banking institutions. Basel III is a comprehensive set of reform measures, developed by the Basel Committee on Banking Supervision (BCBS), to strengthen the regulation, supervision, and risk management of the banking sector.[3] One key liquidity measure, and the focus of this article, is the LCR.

The LCR aims to strengthen the liquidity positions of financial institutions by creating a standardized minimum daily liquidity requirement for large and internationally active banking organizations.[4] In particular, relative to the pre-crisis period, the LCR requires that bank holding companies (BHCs) maintain ready access to a pre-determined minimum level of highly liquid assets to meet demand for cash over the short term, a rolling 30-day period. (In this article, we will use the terms "BHCs" and "banks" interchangeably.) The formula for calculating the LCR is generally represented by equation (1):

$$(1) \qquad LCR = \frac{HQLA}{Estimated\ net\ cash\ outflows} \geq 100\%$$

The numerator of the ratio, HQLA, is made up of a range of liquid assets grouped into categories according to their approximate "level" of liquidity, or ease of conversion to cash. "Level 1" assets, the first column in Table 1, comprise the most liquid forms of HQLA, free of haircuts and limiting compositional caps. This category includes excess reserves and securities issued or guaranteed by the U.S. government. Excess reserves are balances held at Federal Reserve Banks in addition to any that banks must hold to meet reserve requirements against their deposit liabilities. The allowable securities in the Level 1 asset category include U.S. Treasury securities, Government National Mortgage Association (GNMA) MBS, and obligations issued by U.S. government agencies (or "non-GSE [government-sponsored enterprise] agency debt"). U.S. government agencies (non-GSEs) include the GNMA, the FDIC, and the Small Business Administration.

Table 1

Selected Components of HQLA

Level 1 assets (no haircut)	Level 2A assets (15% haircut)
Excess reserves	GSE debt**
Treasury securities	GSE MBS**
GNMA MBS	GSE CMBS**
Non-GSE agency debt*	

NOTE: The highlights indicate the three largest components. *Non-GSE agency debt includes U.S. government agency securities such as the debt of the GNMA, the FDIC, and the Small Business Administration. **GSEs include Fannie Mae, Freddie Mac, the Federal Home Loan Bank System, and the Farm Credit System.

"Level 2" assets, which cannot account for more than 40 percent of total HQLA, comprise two subcategories:

- Level 2A assets, the second column in Table 1, are subject to a 15 percent haircut and include securities issued or guaranteed by a U.S. GSE, such as GSE debt securities as well as these institutions' residential MBS and commercial MBS (CMBS). (U.S. GSEs are defined in the Table 1 note.)
- Level 2B assets (not shown), which include corporate debt securities and tend to comprise a much smaller portion of banks' balance sheets, are subject to a substantial haircut (50 percent) and can be no more than 15 percent of total HQLA. Without loss of generality, we abstract from these assets in our analysis.[5]

Of the various HQLA assets, banks in aggregate hold the largest amounts of excess reserves, Treasury securities, and GSE MBS, the highlighted cells in Table 1 and the predominant focus of our analysis.

A bank's size and the degree of its international exposure determine which of two required LCR stringency levels it must meet. The largest banking organizations are subject to the "standard" LCR rule (hereafter, standard LCR banks or standard BHCs). At year-end 2017, 15 U.S. banks were in this category. Smaller banking organizations are subject to a less stringent, "modified" LCR rule (hereafter, modified LCR banks or modified BHCs).[6] At the end of 2017, 22 banks were in this category, but we excluded 5 such institutions from our sample because of their insufficient time spent in the category or because they underwent a significant structural change during the sample period. Thus, our sample includes 15 standard LCR banks and 17 modified LCR banks. For comparison, we also consider banks not subject to the LCR requirement over our period of study ("non-LCR banks"). Note that for all analysis in this article, we keep each bank's LCR membership—standard, modified, or neither—fixed across time based on its year-end 2017 status.

Although the largest U.S. banks began publicly disclosing a limited set of metrics related to the LCR requirement in the second quarter of 2017, data for periods prior to that time were not required to be reported. In addition, while some large banks began publically disclosing their LCRs as early as early 2014 during events such as quarterly earnings conference calls and other investor presentations, they did so without releasing the detailed data that underlie the ratio. Here, using a range of data sources, we construct two different time series of banks' HQLA holdings. The first is a quarterly series based on publicly available data sources. The second is a daily series based on confidential supervisory data reported to the Federal Reserve. Our proxy HQLA series allow us to analyze which assets these institutions chose to hold to satisfy their LCRs over time. The data appendix provides a full description of our data sources, both public and confidential, as well as our calculation methods and assumptions made for simplicity or due to data limitations; we describe in subsequent sections how we take asset haircuts and caps into account.

As we show in the next section, although the deadline for banks' full LCR compliance was phased in between January 2015 and January 2017, banks' balance sheet management was affected well before this period.[7]

4 HOW DID BANKS INITIALLY INCREASE AND THEN SUBSEQUENTLY MANAGE THEIR HQLA?

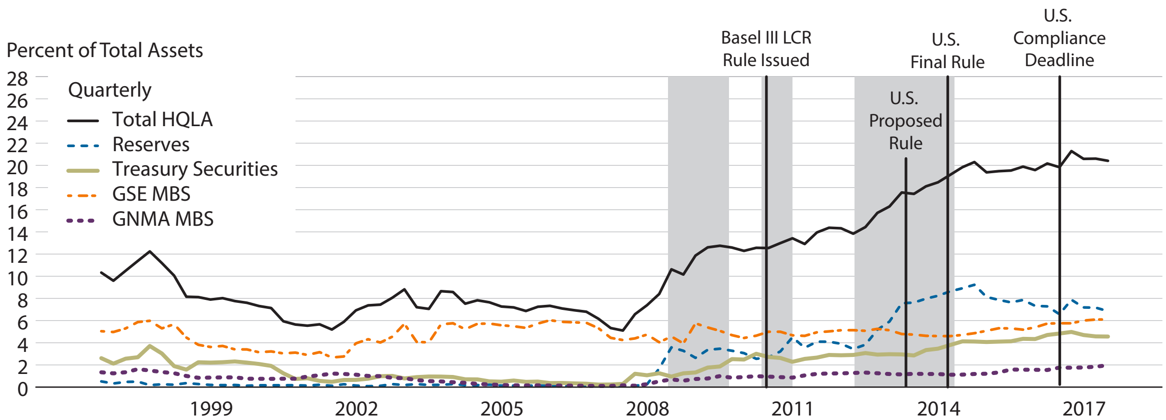
Figure 1 shows our quarterly time-series measure of standard LCR banks' aggregate HQLA (total holdings of HQLA, abstracting from the LCR's haircuts and compositional caps) and each of the three key components of HQLA highlighted in Table 1—reserve balances, Treasury securities, and GSE MBS.[8] For comparison, we also show GNMA MBS, a much less widely held Level 1 component of HQLA. Each of the five data series is plotted as a share of these banks' *total assets*. Looking at the key LCR-related announcement dates (vertical bars), one sees that in the years leading up to the initial deadline for LCR compliance (January 2015), these institutions substantially increased the shares of liquid assets on their balance sheets, particularly in 2013 and 2014, and did so primarily by taking on substantial reserve balances (blue-dashed line).

It is important to remember that accumulating reserve balances over this period was easy for banks; in fact, it was not a choice for the banking sector as a whole. Prior to the Financial Crisis, the amount of reserve balances in the banking system was consistently quite small—that is, prior to 2008, the blue-dashed line is very close to zero. However, as a result of the Federal Reserve's LSAPs, which were conducted between 2009 and 2014 (gray-shaded regions), reserve balances grew over many quarters at varying rates.[9] In total, the amount of reserves in the banking system increased by over $2 trillion as a result of the Federal Reserve's LSAPs.

The figure also illustrates that standard LCR banks actively acquired other HQLA-eligible securities in the lead-up to the full-compliance deadline (January 2017). In particular, the share of these institutions' holdings of Treasury securities (green line) rose over 2014 and

Figure 1

Major Components of High-Quality Liquid Assets at Standard LCR Banks

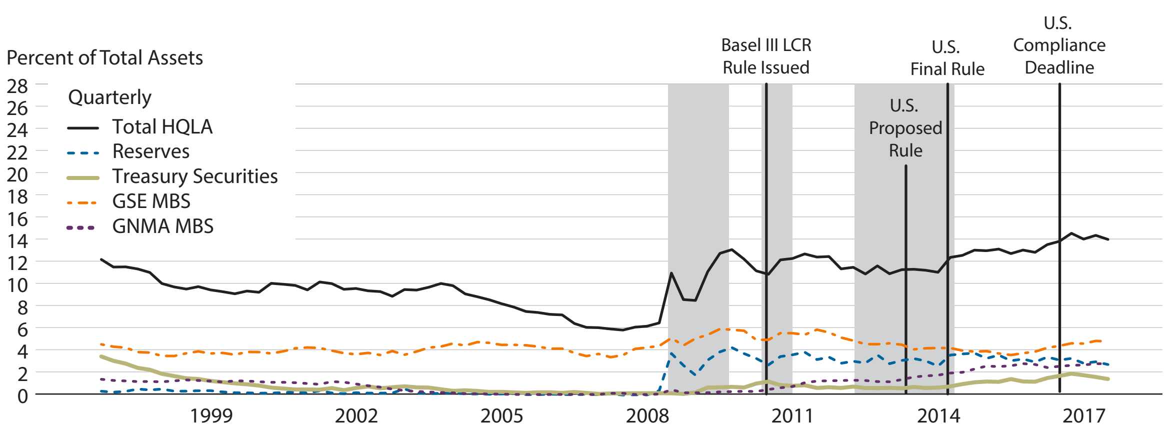


NOTE: Standard LCR banks are defined as BHCs with $250 billion or more in total assets or $10 billion or more in on-balance sheet foreign exposures. Shaded bars indicate periods of the Federal Reserve's LSAPs. See the data appendix for explicit line items used to construct the categories.

SOURCE: FR Y-9C, FR 2900 (for reserve balances).

Figure 2

Major Components of High-Quality Liquid Assets at Modified LCR Banks

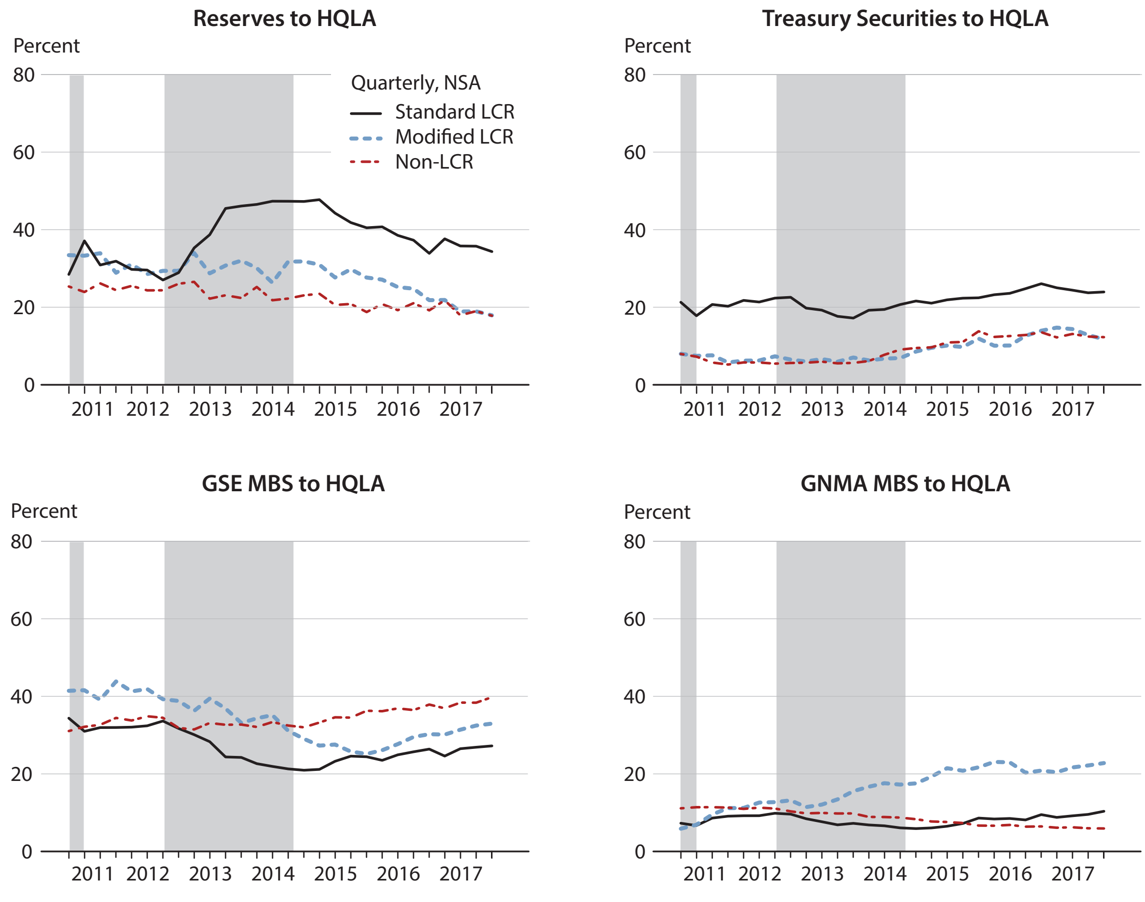


NOTE: Modified LCR banks are defined as BHCs with $50 billion or more in total assets. Shaded bars indicate periods of the Federal Reserve's LSAPs. See the data appendix for explicit line items used to construct the categories.

SOURCE: FR Y-9C, FR 2900 (for reserve balances).

Figure 3

Comparing Shares of Selected HQLA-Eligible Assets by Bank Group (units as a percent of HQLA)



NOTE: NSA, not seasonally adjusted. Reserves defined as balances held at Federal Reserve Banks as reported in Federal Financial Institutions Examination Council (FFIEC) Call Reports. Only reserve balances held in excess of required amounts are HQLA eligible. All other data are from the FR Y-9C. Security balances are based on fair values for available-for-sale (AFS) and held-to-maturity (HTM) securities, as well as securities held in trading accounts. Shaded bars indicate periods of the Federal Reserve's LSAPs. See the data appendix for explicit line items used to construct the categories.

SOURCE: FR Y-9C and Call Reports (FFIEC 031/041).

continued to grow, on balance, thereafter. In addition, the share of these banks' holdings of GSE MBS (orange-dashed line) also increased over most of that same period. Meanwhile, these banks' holdings of GNMA MBS (purple-dashed line) stayed relatively low and flat as a share of total assets.

Figure 2 reports similar data for the set of modified LCR (smaller) banks. Here we see a very different pattern. These banks' reserve holdings as a share of total assets (blue-dashed line) were roughly steady from the time of the issuance of the Basel III rule through the end of 2016. In contrast to the standard LCR banks, over that same period, these firms mostly increased their holdings of GNMA MBS (purple-dashed line), a Level 1 asset, and also increased their holdings of Treasury securities.

Of course, another important difference between the two sets of banks is that the modified LCR banks needed to undertake a much smaller overall buildup of HQLA: At the end of 2017, the HQLA of modified LCR banks were about 14 percent of total assets (black line), while those for standard LCR banks were about 20 percent. This differential is approximately accounted for by the differential treatment of the denominator of the modified LCR rule (see endnote 6).

Next we compare the behavior of the two LCR bank groups just discussed side by side with similar data for the set of non-LCR banks. In particular, Figure 3 shows each of the same four HQLA components—reserves, Treasury securities, GSE MBS, and GNMA MBS—but now as a share of *total HQLA* (instead of total assets) for standard LCR banks (black lines), modified LCR banks (blue-dashed lines), and non-LCR banks (red-dashed lines). The same component contours described above for the LCR banks are apparent. Meanwhile, the liquid-asset holdings of non-LCR banks remained relatively flat during the period of transition to meet the liquidity requirement, suggesting that the LCR was in fact an important driver of banks' balance sheet management over that period.

Banks likely faced somewhat different incentives in the run-up to becoming LCR compliant than they subsequently faced. In the run-up, Wall Street analysts were regularly asking about banks' progress and some institutions began publicly reporting updates; as a result, banks may have felt some pressure to build a sufficient *stock* of HQLA to become compliant. If so, managing the *composition* of HQLA over that period may not have been banks' top priority.

But in 2015 and 2016, after standard LCR banks first met initial LCR compliance, *total* HQLA generally leveled off—HQLA comprised about 20 percent of the total assets of standard LCR banks at the end of 2014 and remained in that neighborhood thereafter (see Figure 1). During this period, banks' primarily made compositional adjustments to their HQLA. In particular, reserves comprised 47 percent of standard LCR banks' HQLA at the end of 2014 and then declined by about 13 percentage points (see Figure 3).[10] Meanwhile, the GSE MBS and Treasury shares at these same institutions each increased. Similarly, as noted above, modified LCR banks also adjusted the compositions of their HQLA after meeting initial LCR compliance, increasing their shares of both GNMA MBS and Treasury securities. Moreover, all of these institutions continued to adjust their HQLA compositions once the deadline for full LCR compliance passed, generally reducing their shares of Treasury securities in favor of holding additional MBS.

This reshuffling of banks' HQLA raises the question, What drives banks' decisionmaking regarding their desired shares of HQLA? The next section addresses this question using a risk-return framework.

5 TO WHAT EXTENT DOES A RISK-RETURN TRADE-OFF EXPLAIN BANKS' HQLA SHARES?

We begin by thinking about an active bank treasurer who needs to invest cash to maximize the bank's risk-adjusted return on its asset holdings and at the same time ensure sufficient funds are available for the institution's liquidity needs. A bank's liquidity needs can stem both from external regulatory constraints—such as the LCR and resolution-planning requirements—

and from self-imposed internal controls and preferences regarding liquidity and capital or other aspects of the bank's balance sheet management. In managing its liquidity needs, the bank treasurer accounts for both liquidity risk and interest rate risk.[11]

Our risk-return framing contrasts with an approach in which a bank's choice of HQLA is modeled in the context of its portfolio value at risk, such as in Cuoco and Liu (2006). Value at risk is a statistical technique used to estimate the quantity of potential losses on an entity's investment portfolio over a specific timeframe. Such an approach is useful when the focus of study is banks' potential for realizing losses and the extent of those potential losses, particularly on relatively risky assets, such as those whose market pricing tends to be highly volatile, or on various other exposures such as foreign currency or derivatives positions. Instead, we focus on banks' holdings of highly liquid assets and on explaining the risk-return and other trade-offs that banks face in choosing the optimal allocations of such assets. Our modeling is similar to Halaj (2013, 2016), who focuses on a bank's optimal risk-return problem; there, however, the focus is on banks' entire balance sheets in context of stress testing.

In our model, we assume a bank allocates HQLA across three assets—reserve balances, Treasury securities, and GSE MBS. Reserves are guaranteed to have the same principal value returned the next day, and these balances receive a fixed overnight return. Therefore, this asset carries no liquidity, interest rate, or credit risk, and we consider it the risk-free asset. We also assume that each individual bank can choose the level of reserve balances it wishes to hold. Of course, the Federal Reserve determines the aggregate quantity of reserve balances available to the banking system. Stated another way, while an individual bank may adjust its holdings of reserves, the banking system as a whole cannot.

Unlike reserve balances, the day-to-day value of, and returns on, Treasuries and MBS vary as their market prices adjust with current and expected future financial market conditions, including the likely future paths of interest rates. A bank treasurer may wish to reduce the liquidity and interest rate risks associated with its liquid portfolio for a number of reasons. For example, a bank may want more certainty regarding the value of its liquid portfolio, more immediate accessibility to its liquid assets on hand, or to limit unwanted volatility in its regulatory capital stock through marked-to-market changes in its liquid securities holdings.[12] Keep in mind that transactions to obtain cash by selling a security or by lending a security through a repurchase agreement (repo) do not settle until the following day, and the ultimate proceeds raised from such transactions depend on conditions in financial markets on a given day. To reduce interest rate risk from Treasury holdings, securities of shorter tenors can be purchased. But such a step will not completely eliminate the market price volatility of these holdings. Regarding MBS, while shorter tenors can also be used to manage interest rate risk, the duration of MBS generally lengthens when interest rates rise because mortgage prepayments tend to decline in that situation. As a result, interest rate risk can be higher for MBS than Treasury securities, even if the securities pay the same coupon rate. And both Treasuries and MBS carry some liquidity risk, as a repo or sales transaction is needed to monetize these assets.[13] Overall, banks have to determine the share of HQLA they want to hold in the form of reserves—the risk- free asset with a lower average return—versus these longer-term assets that yield higher average returns but carry varying degrees of liquidity and interest rate risks.

This model is consistent with bank treasurers' own descriptions of their HQLA management.[14]

To determine a bank's optimal liquid portfolio, we first solve for the optimal "risky" HQLA portfolio (denoted with subscript *R*)—that is, the share of Treasury securities and MBS—that maximizes the risk-return trade-off captured by the Sharpe ratio shown in equation (2):

$$\text{(2)} \qquad Sharpe\ ratio = \frac{E\left(r_R\right) - r_{RB}}{\sigma_R},$$

where $E(r_R)$ is the expected return of the risky portfolio, r_{RB} is the risk-free rate on reserve balances, and σ_R is the standard deviation of the risky portfolio. Daily returns are used to make these calculations.

After determining the optimal risky HQLA portfolio, a specific bank's complete HQLA portfolio (denoted with subscript *P*) will depend on its aversion to liquidity and interest rate risks, which is captured in parameter *A* in equation (3):

$$\text{(3)} \qquad U\left(r_P\right) = E\left(r_P\right) - \frac{1}{2} A \sigma_P^2,$$

where

$$\text{(4)} \qquad E\left(r_P\right) = (1 - w) * r_{RB} + w * E\left(r_R\right),$$

$$\text{(5)} \qquad \sigma_P^2 = w^2 * \sigma_R^2,$$

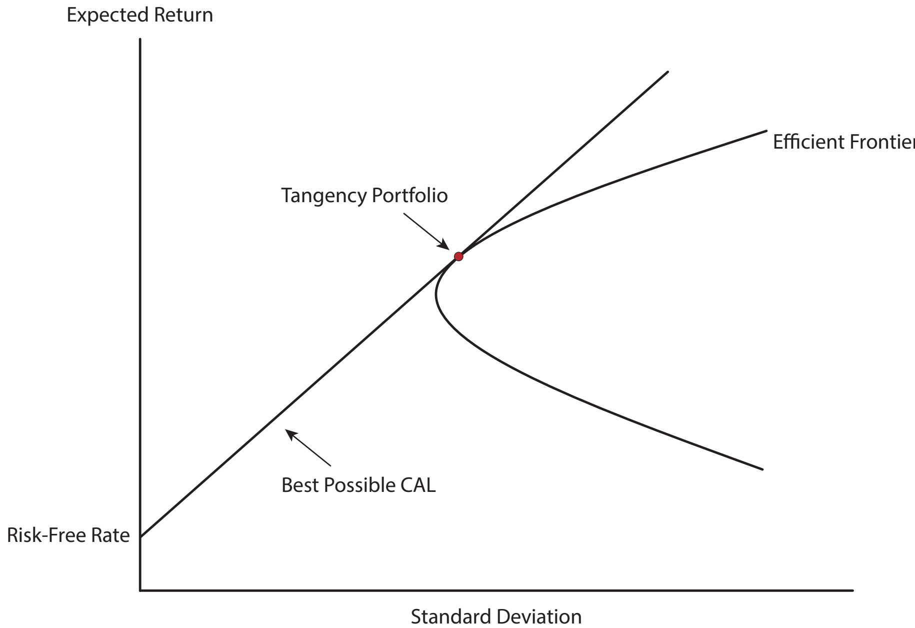
and *w* is the share of risky assets. The higher a bank's value of *A*, the more risk averse it is—that is, the more it prefers a high share of reserves (cash) in its holdings of HQLA.

The curve in Figure 4 provides a stylized illustration of the efficient frontier that represents the optimal share of Treasury securities and MBS holdings—the risky portfolio—for given risk-return trade-offs. When a risk-free asset is introduced, the straight line shown in the figure, known as the capital allocation line (CAL), is the new efficient frontier for the complete portfolio. The line's vertical intercept represents a portfolio that contains only the risk-free asset. The red dot designates the optimal risky portfolio. All points on the line between these two corner solutions represent portfolios that allocate some funds to both the risk-free asset and the optimal risky portfolio. Where a bank ends up on that line depends on its aversion to liquidity and interest rate risks—its value of *A*.

In solving for a bank's optimal portfolio shares, we use one representative Treasury security and one representative GSE MBS security in the analysis.[15] Information on the maturity of banks' securities holdings is not available through publicly available sources, so we use the Federal Reserve's confidential *Complex Institution Liquidity Monitoring Report* (FR 2052a) data as described in the data appendix to compute the average maturity of securities held by large banks. The maturity buckets collected are not sufficiently granular to distinguish among specific maturity holdings of Treasury securities or MBS, but we do see that the largest banks tend to hold securities of these types that mature in five or more years. Therefore, we assume that the market yields on 5-year constant maturity Treasury securities and 5- and 10-year

Figure 4

The Standard Investment "Efficient Frontier"



NOTE: This figure illustrates the construction of an optimal portfolio of assets given the efficient frontier of risky assets and the inclusion of a risk-free asset. The CAL is constructed by connecting the risk-free rate (y-axis) to the tangency of the efficient frontier, which is the optimal risky portfolio.

constant maturity MBS (Fannie Mae) are the best available proxies for the returns on the two risky assets in our model portfolio.

We let the level of the risk aversion parameter, A, vary from 1 (relatively risk tolerant) to 10 (relatively risk averse).[16] In our setting, a relatively risk averse bank prefers little liquidity or interest rate risk. Aversion to interest rate risk corresponds with a desire for protection from volatility in interest rates—or, in terms of fixed-income securities, volatility of the market pricing—associated with the bank's HQLA portfolio. A bank's aversion to interest rate risk in this context will depend in part on the extent to which it wants to reduce the effects of any mark-to-market volatility in its securities holdings on its regulatory capital measures. For example, relative to a bank with a large regulatory cushion, a bank that is operating close to its regulatory minimum leverage ratio may not want much market price volatility associated with its assets and, therefore, will have a relatively higher value of A. Separately, a bank that protects itself from liquidity risk by assuming higher haircuts than its competitors do for internal stress-testing purposes, such as in the context of its resolution planning, will also have a relatively higher value of A.

Regarding our sample period, to start, we consider a recent period, 2012 through 2016, in which the size of the Federal Reserve's balance sheet was large and held constant. In this case, our model suggests that a bank treasurer should not want to hold any amount of the

Table 2

Optimal Portfolio Composition Using Data from 2001 to 2016

	Maturity of the two "risky" assets		Portfolio compositions (optimal portfolio shares, %)			
Bank's risk aversion (*A*) (1)	Treasury securities (2)	GSE MBS (3)	Reserves (4)	Treasury securities (5)	GSE MBS (6)	Total (by definition) (7)
High risk aversion (*A* = 10)	5 year	5 year	79	0	21	100
		10 year	40	20	40	100
Middle of range (*A* = 5)		5 year	59	1	40	100
		10 year	15	45	40	100
Low risk aversion (*A* = 1)		5 year	0	60	40	100
		10 year	0	60	40	100

risk-free asset—reserve balances. This holds for all considered values of the risk aversion parameter, *A*. In fact, given the relatively high return on GSE MBS and these securities' limited volatility, each institution in this case prefers to hold all of its liquid assets in MBS. But, as noted above, there is a limit to the amount of Level 2 assets that may be used to satisfy the LCR; because no more than 40 percent of a bank's HQLA holdings may consist of such assets, the optimal composition of HQLA in this case is 60 percent Treasury securities and 40 percent GSE MBS.[17]

However, one may think that during this period the volatility in market interest rates was significantly constrained by their proximity to the zero lower bound and, given that our model relies on the relationship between asset returns and covariances, that this circumstance could bias our model results away from holding the risk-free asset toward holding a riskier portfolio. Therefore, we also consider the longer period from 2001 to 2016, which embodies much more variability in the returns on liquid assets. Because the Federal Reserve began paying interest on reserve balances during this longer period—in late 2008—we assume the risk-free asset is remunerated at the effective federal funds rate prior to the onset of interest on reserves. This assumption can be interpreted as one in which the Federal Reserve's pre-crisis monetary policy implementation framework was a corridor system that paid interest on excess reserves.[18]

For this longer time period, we do find demand for the risk-free asset. As shown in Table 2, the share of reserves in a bank's optimal HQLA portfolio (column 4) varies widely depending on the bank's assumed degree of risk aversion (column 1). For example, when a bank's risk aversion is high (that is, $A = 10$; first pair of rows), the optimal share of reserves ranges from 40 percent to about 80 percent of the bank's total HQLA, depending on the securities considered. In contrast, when a bank's risk aversion is assumed to be low ($A = 1$; bottom pair of rows), the optimal share of reserves is zero regardless of the other securities considered. We conclude that our model supports a fairly wide range of plausible liquid portfolio compositions.

These results have implications for the Federal Reserve's longer-run supply of reserves in the banking system. The LCR requirement has effectively caused a structural change by boosting banks' demand for highly liquid assets relative to the pre-crisis period. This change means that, relative to the pre-crisis period, for any given level of the federal funds rate, more excess reserves may be needed in the banking system to meet the LCR requirement. This finding is consistent with the Board of Governors of the Federal Reserve System's September 2018 Senior Financial Officer Survey where large banks responded that regulatory factors are important determinants of their demand for reserves.[19] That said, asset substitutability will also importantly be driven by interest rate differentials. If the yield on Treasury securities or GSE MBS is sufficiently higher than the administered rate of interest that the Federal Reserve sets on excess reserves, then any LCR-driven demand for reserves could be relatively dampened.

So far, we have very little direct information regarding the behavior of banks' underlying demand for excess reserve balances. For example, the aggregate level of reserves in the banking system declined significantly through 2018, from its peak of $2.8 trillion in late 2014 to about $1.7 trillion in December 2018, with no significant pressure evident on the various interest rates that one might expect if banks had a strong preference to hold on to those reserve balances.

We next consider bank-level data to see how individual banks' compositions of HQLA have evolved in recent years.

6 WHAT DO INDIVIDUAL BANK-LEVEL DATA SHOW?

To get an understanding of how individual large banks have been managing their liquidity needs post crisis, we look at bank-level data for the eight U.S. GSIBs. First we examine the longer-run trends in their holdings based on our quarterly estimates of HQLA. Then we consider high-frequency daily movements in these holdings by utilizing a relatively new, confidential supervisory data set collected via the Federal Reserve's FR 2052a described in our data appendix.

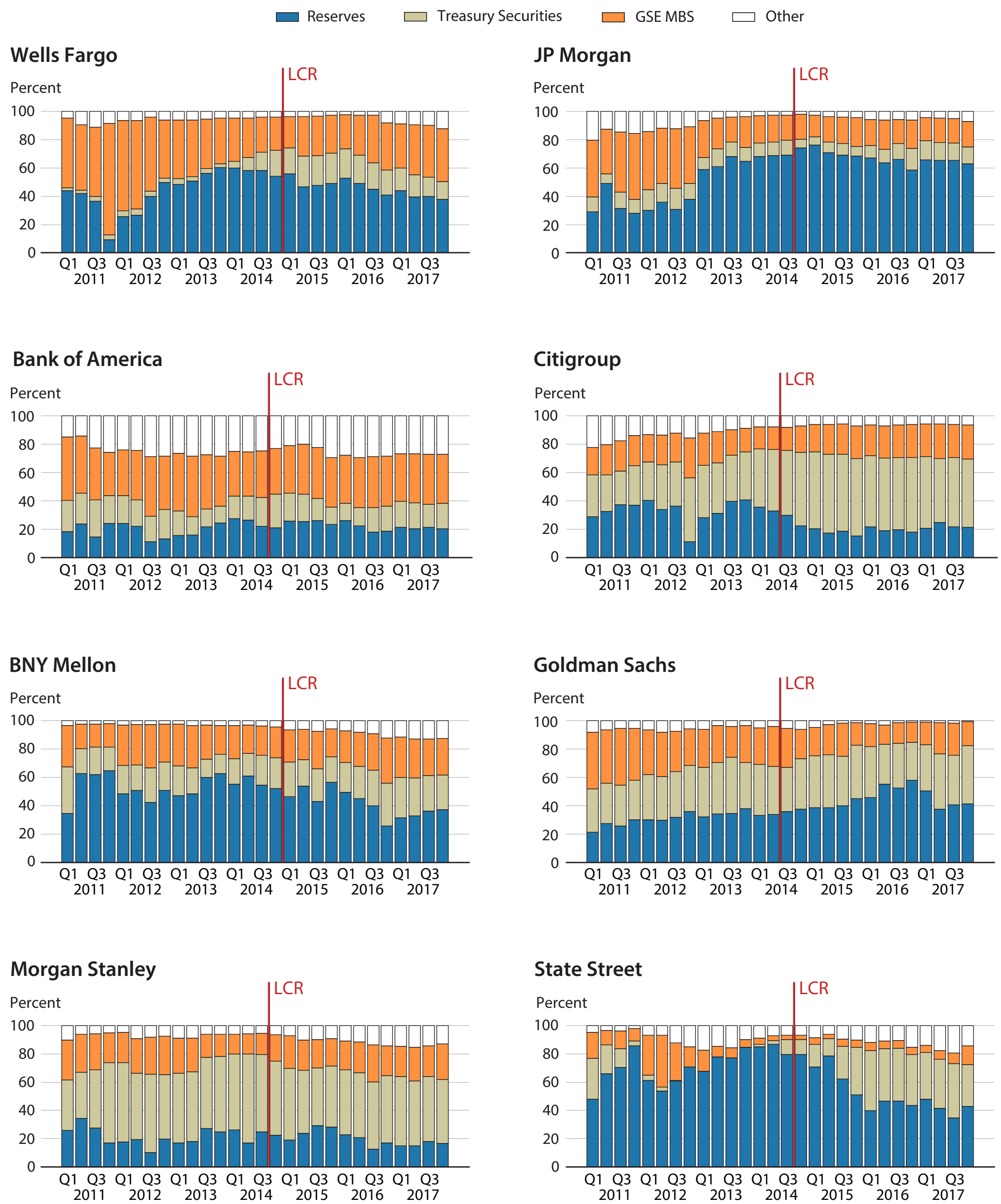
At the end of 2017, the eight GSIBs together held more than $2 trillion of aggregate HQLA and represented more than 80 percent of the total assets of standard LCR banks (those shown in Figure 1). Figure 5 displays the composition of each of these eight institution's HQLA, with the institutions ordered by dollar amount of HQLA holdings, largest to smallest from left to right. The red vertical lines indicate when each institution publically announced it had fully met its LCR requirement (that is, had a ratio of at least 100 percent). As shown, in the run-up to meeting the LCR, about half of these banks relied heavily on reserves (blue bar portions), while others relied on a wider range of assets.

However, after becoming compliant, most banks that had relied heavily on reserves exhibit a decline in such reliance, albeit to varying degrees. Such adjustments are consistent with banks subsequently aiming to improve the risk-return trade-off associated with their HQLA portfolios and determining that fewer reserves were needed to meet their liquidity needs.

Data through 2017:Q4 show that each of these large banks appears to have since settled into relatively stable shares of HQLA, and, across institutions, they display a wide range of reserve (risk-free asset) shares of HQLA—from as little as around 20 percent or less for about half of the institutions to above 50 percent. Even the four largest holders of HQLA, often cat-

Figure 5

Individual Banks' HQLA Shares

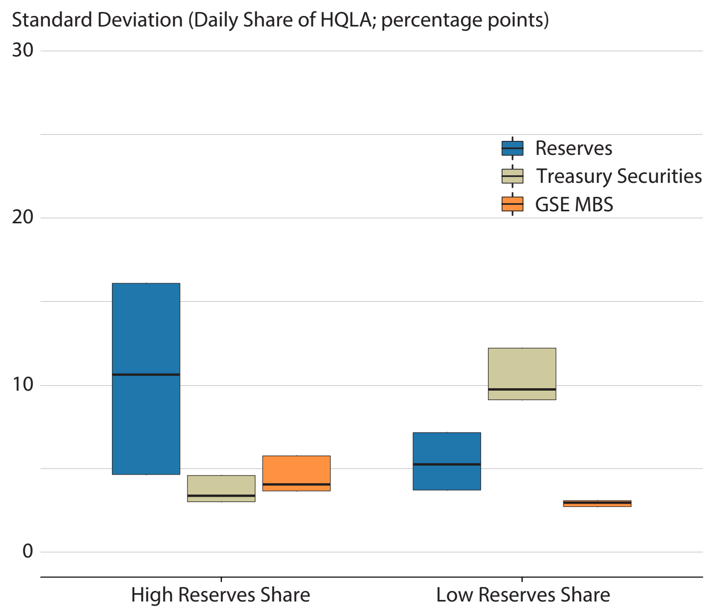


NOTE: Security balances based on fair value of AFS, HTM, and trading account securities. Other HQLA include GNMA MBS, agency CMBS, and agency debt. Red lines indicate quarter in which bank publicly stated it met the fully phased-in final U.S. LCR rule based on its own calculations.

SOURCE: Call reports (for reserves, or balances held at Federal Reserve Banks; FFIEC 031/041), FR Y-9C (all other data), and transcripts from quarterly earnings calls and financial updates (red lines).

Figure 6

Daily Variance of Banks' HQLA Shares Post LCR (January 2015–March 2017)



NOTE: "High Reserves Share" banks comprise the four GSIBs with the highest average shares of reserves to HQLA, and "Low Reserves Share" banks comprise the remaining four banks. The horizontal line through each bar is the median bank's standard deviation; the high and low ends of each bar signify the 75th and 25th percentiles, respectively. (With four observations in each bank group, medians and percentiles are the midpoint [average] of the appropriate neighboring observations.)

SOURCE: FR 2052a.

egorized as universal banks, are noticeably dissimilar in their compositional choices: Wells Fargo relies mainly on a mix of reserves and GSE MBS; JP Morgan mainly reserves; Bank of America mainly GSE MBS; and Citigroup mainly Treasury securities. The two investment banks, Goldman Sachs and Morgan Stanley, also exhibit very different compositions.

Turning to the *daily* movement in the GSIBs' individual HQLA shares, we compute the average daily standard deviation of each institution's reserves, Treasury securities, and GSE MBS shares. As shown in Figure 6, we divide the sample into two equal-sized bank groups—those with relatively high reliance on reserve balances (left-hand bars) and those with relatively low reliance on reserve balances (right-hand bars).[20] Overall, we find that banks that experience higher daily variance in their cash balances tend to hold more cash. As shown by the left-most blue bar, the median share of reserves that comprises reserve-reliant institutions' HQLA varies day-to-day by about 10 percentage points and the interquartile dispersion runs from 5 to 15 percentage points on any given day. Meanwhile, the shares of these banks' Treasury securities (left-most green bar) and MBS (left-most orange bar) are significantly more stable. In contrast, as shown by the right-hand set of bars, banks with much less variation in their reserves share exhibit relatively low reliance on reserve balances; for these institutions, the highest average daily variation in HQLA shares is in the Treasury share.[21] Next we explore what can be driving these and other differences across the GSIBs.

7 WHAT EXPLAINS THE DISPARATE BEHAVIOR OF BANKS' HQLA SHARES?

In terms of our model, the question is, what in *A* is driving the large differences in banks' HQLA shares and in their reserves shares in particular? A number of considerations appear to be in play and, for the largest banks at least, we conclude that the answer largely varies by institution. We attribute this result to three factors.

First, as was their intent, post-crisis financial regulations are affecting banks' liquidity management. The LCR is a predominant example, but it is important to remember that banks' resolution planning and capital management also affect their liquidity management. Respondents to the Senior Financial Officer Survey ranked internal liquidity stress tests, which are used in resolution planning, as the most significant factor affecting their demand for reserves. In an internal stress test, a bank that deems it difficult to raise large-scale funds quickly by engaging in sufficiently sizable repos or by conducting sufficiently sizable outright asset sales, may lean toward holding a relatively abundant level of reserves for precautionary reasons.[22] Indeed, in 2017 Citigroup pointed to its resolution planning as an important cause for growth in its cash balance.[23] Furthermore, capital constraints can also influence banks' choice of HQLA holdings. A bank with an effectively binding risk-based capital ratio would likely lean, on the margin, toward onboarding additional reserve balances and Treasury securities to meet its HQLA requirement because those assets carry a zero risk weight.

Second, observable factors, including a bank's business model and the particular products and services it offers, importantly help determine a bank's liquidity needs and drive its internal preferences for certain quantities and types of liquid assets. For example, a bank that engages in significant payment, settlement, and clearing activities may desire a relatively high share of reserve balances to meet its intraday and next-day liquidity needs, especially since the timing of many payments has recently shifted to earlier in the day, when settlements for repo transactions would not be available.[24] These factors would likely hold for large processing banks such as BNY Mellon and State Street. Other business factors that may boost demand for reserves include commercial lending activities and retail client credit lines such as credit cards and home equity lines of credit. These customer accounts can be unpredictably tapped, suggesting that banks providing such services may prefer certain quantities and types of liquid assets. And those that manage significant amounts of volatile deposits from institutional clients may prefer holding a relatively high share of cash balances to accommodate such variation.[25] Moreover, relative to pre-crisis times, reserve balances now earn interest, reducing the opportunity cost of holding such balances.

Finally, a third factor is unobservable, internal business preferences. Given that the two GSIB investment banks face similar business models and regulations but differ significantly in their preferences for cash holdings, we conclude that these banks' compositional HQLA choices are driven by unobservable, internal business preferences. In particular, banks' self-imposed operational cash targets may be established by internal liquidity stress tests; for example, JP Morgan disclosed that an internal duration target is important in driving its HQLA management.[26] Limiting the duration of a securities portfolio is an example of how banks actively manage interest rate risk, consistent with our portfolio model.

Overall, the LCR is only one constraint that banks face when choosing their optimal holdings of HQLA. Several other factors are importantly driving banks' liquidity management. How the various factors interact in determining banks' demand for reserves—and the relative importance of each factor in so doing—is complicated by the fact that banks prioritize these factors differently, resulting in a range of HQLA portfolios.

8 CONCLUSION

The LCR is a post-crisis liquidity requirement that has importantly affected banks' balance sheet management, research about which is only recently emerging. Our work contributes to the discussion by constructing time-series data to document how U.S. banks have managed the compositions of their HQLA over recent years and by providing a theoretical framework to model banks' preferences for the composition of their HQLA. We empirically documented how banks' HQLA holdings have evolved over time. We explained banks' decisions in the context of our model, which highlights that a bank's demand for reserves is sensitive to the opportunity cost of holding such balances as well as to its aversion to liquidity and interest rate risks. Overall, the more risk averse (tolerant) a bank is, the higher (lower) its demand for reserves.

Looking at the data we created to study both the compositions of banks' HQLA shares and the patterns of daily volatilities of those shares, we documented that banks' individual preferences for HQLA shares vary widely, even for banks with seemingly similar business models. We conclude that banks' liquidity management practices—and thus their underlying sensitivities to liquidity and interest rate risks—are likely driven by a range of considerations that may be grouped into three broad categories: considerations pertaining to liquidity and capital regulations, including the LCR and resolution planning; observable factors such as product lines and business models; and unobservable considerations related to banks' individual business practices and preferences.

We also explained why understanding banks' post-crisis liquidity management practices is important for both the implementation and the transmission of monetary policy. Going forward, the Federal Reserve will need to monitor total reserves in the banking system and decide if the level is "ample" for effective and efficient monetary policy implementation. Careful monitoring of large banks' management of HQLA, and in particular their demand for reserves, will be important in this process. ■

DATA APPENDIX

Here we detail the data sources, methods, and assumptions used to construct each of the bank-level HQLA estimates used in our article.

Quarterly Bank-Level HQLA Estimates

Our quarterly HQLA time series is based on publicly available data. Our data sources include quarterly regulatory filings, some of which are reported at the commercial bank (depository institution) level; when this was the case, we subsequently mapped individual commercial banks to their affiliated (parent) BHC. These sources included the following:

- *Consolidated Financial Statements for Holding Companies* (FR Y-9C) from which we obtained individual BHCs' securities holdings, including investment securities and securities in the banks' trading accounts. The FR Y-9C reporting form and instructions are available on the Board of Governors of the Federal Reserve System's website: https://www.federalreserve.gov/apps/reportforms/reportdetail.aspx?sOoYJ+5BzDal-8cbqnRxZRg==.
- *Consolidated Reports of Condition and Income for a Bank with Domestic Offices* (FFIEC 041) and *Domestic and Foreign Offices* (FFIEC 031) for data on reserve balances. The FFIEC 031 and 041 forms and instructions are available on the FFIEC's website: https://www.ffiec.gov/ffiec_report_forms.htm.

For simplicity and because of some data limitations, we made the following methodological choices in constructing our quarterly bank-level HQLA estimates:

- We ignored LCR-eligible HQLA assets outside of the seven asset categories listed in Table 1, including banks' holdings of Level 2B assets such as corporate debt. Banks generally hold relatively small shares of such assets, making our conclusions robust to their exclusion. For example, for the eight GSIBs we examined in detail, less than 2 percent of their HQLA consisted of Level 2B assets in the second quarter of 2017, as reported in their first required public LCR disclosures.
- We used data on banks' total reserve balances instead of excess reserves, but these measures differ little in aggregate over the period we primarily study. According to the Federal Reserve's H.3 statistical release, *Aggregate Reserves of Depository Institutions and the Monetary Base*, required reserves in the banking system totaled about $160 billion on average in February 2017, while total reserves averaged about $2,300 billion.
- For Figures 1 and 2, we ignored Level 2 asset caps (described in the main text); however, most BHCs are not near those caps. The share allocations in our risk-return model analysis did account for the caps.
- Due to limitations of the granularity of the publicly available data, some non-GSE agency securities (Level 1 assets) are included in our estimates of the GSE asset (Level 2A) categories. However, for very large banks, we think our HQLA asset category estimates are reasonable. For example, for the eight GSIBs examined, we estimate that use of the

Table A1

Construction of Quarterly Bank-Level HQLA Estimates from Public Sources

HQLA component	Calculation
Reserves	FFIEC 031/041 RC-A item 4
Treasury securities	FR Y-9C HC-B item 1 (column B) + Item 1 (D) + HC-D item 1 (A)
GNMA MBS	FR Y-9C HC-B item 4.a.(1) (B) + Item 4.a.(1) (D)
Non-GSE agency debt	FR Y-9C HC-B item 2.a (B) + Item 2.a (D)
GSE debt	FR Y-9C HC-B item 2.b (B) + Item 2.b (D) + HC-D item 2 (A)†
GSE MBS	FR Y-9C HC-B item 4.a.(2) (B) + Item 4.a.(2) (D) + Item 4.b.(1) (B)† + Item 4.b.(1) (D)† + HC-D item 4.a (A)† + Item 4.b (A)†
Agency CMBS	FR Y-9C HC-B item 4.c.(1)(a) (B)† + Item 4.c.(1)(a) (D)† + Item 4.c.(2)(a) (B)† + Item 4.c.(2)(a) (D)† + HC-D item 4.d (A)†

NOTE: † Includes obligations of both U.S. government agencies and U.S. GSEs.

publicly available data results in about 80 percent of banks' Level 1 assets sorted accurately, with the remaining 20 percent inaccurately included in our GSE (Level 2A) asset categories.

- Finally, also due to data limitations, we ignored whether the securities included in our HQLA estimates were deemed by each bank to be "unencumbered" or not (the LCR requires securities be unencumbered, that is, free of legal, regulatory, contractual or other restrictions on the ability of the bank to liquidate, sell, transfer, or assign the assets). We do not believe that making such a generality has biased our findings or conclusions.

In Table A1, we list each of the specific line items that we used to construct quarterly HQLA using these publicly available data sources. Given that the LCR requirement is based on assets valued at market values, the fair-value measures of banks' securities holdings are used (as reported on schedule HC-B for both available-for-sale [AFS] and held-to-maturity [HTM] securities).

Daily Bank-Level HQLA Estimates

To construct our daily bank-level measures, we used two sources of confidential micro data collected by the Federal Reserve:

- For historical data on individual banks' reserve balances, we relied on the confidential flows associated with the *Report of Transaction Accounts, Other Deposits and Vault Cash* (FR 2900). FR 2900 reporting forms and instructions are available on the Board's website: https://www.federalreserve.gov/apps/reportforms/reportdetail.aspx?sOoYJ+5BzDblI7g2+r203S0gg6NcUIj6.

- For daily data on HQLA reported directly by BHCs, we relied on the confidential supervisory data collected from the *Complex Institution Liquidity Monitoring Report* (FR 2052a). This form is collected from U.S. GSIBs—those with $700 billion or more in consolidated assets or with $10 trillion or more in assets under custody. These BHCs submit a report each business day and have been doing so since December 14, 2015; prior to that time, beginning in 2012, these BHCs submitted more-limited daily data. The FR 2052a data comprise the detailed balance sheet inputs necessary to calculate the LCR, but do not include the actual LCR ratios or the numerators and denominators of the ratios. FR 2052a reporting forms and instructions are available on the Board's website: https://www.federalreserve.gov/apps/reportforms/reportdetail.aspx?sOoYJ+5BzDbpqbklRe3/1zdGfyNn/SeV.

These source data for the daily measures of HQLA are quite limited prior to December 2015—they consist of reserves, Level 1 HQLA (which include reserves), and (total) Level 2 HQLA. Because we are interested in tracking banks' behavior over the past several years, we created separate proxy measures for banks' Treasury securities and GSE MBS holdings, as follows: For the period in which insufficiently detailed data are available, we assumed that each of the following two relationships held:

$$\text{(A.1)} \qquad \textit{Treasury securities} = \textit{Level 1 assets} - \textit{Total reserves}$$

$$\text{(A.2)} \qquad \textit{GSE MBS} = \textit{Level 2 assets}.$$

For robustness checks, we first compared each proxy share to the corresponding actual measure using data for 2016, when both measures are available. The performance of our Treasury proxy is somewhat mixed, while that of our MBS proxy seems quite accurate. Two banks largely account for the miss in the Treasury proxy, one with large GSE debt holdings and one with a large portion of encumbered (thus ineligible for the LCR) Treasury holdings. Second, we considered whether the proxies and actual measures behaved similarly in terms of the daily standard deviations. Our proxy shares exhibit volatilities that are similar to those of the actual ratios.

NOTES

[1] Our analysis focuses on banks' management of the numerator of the LCR—banks' adjustments to the compositions of their holdings of HQLA. We recognize that to implement and manage the LCR requirement, banks also make adjustments to balance sheet liabilities, which feed back into the calculation of necessary HQLA. Abstracting from the denominator does not affect our analysis.

[2] For an overview of how liquidity regulations are affecting demand for reserves, see Ihrig (2019). This factor is influencing the FOMC's determination of ample reserves as seen in the Federal Open Market Committee's March Press Release "2019 Balance Sheet Normalization Principles and Plans" available here: https://www.federalreserve.gov/newsevents/pressreleases/monetary20190320c.htm. Also, see Quarles (2018) for an overview of how large banks' reserve demand interacts with the size of the Federal Reserve's balance sheet.

[3] The BCBS is a committee of banking supervisory authorities that was established by the central bank governors of the G10 countries in 1975. The Office of the Comptroller of the Currency, Board of Governors of the Federal Reserve System, and Federal Deposit Insurance Corporation (FDIC) actively participate in the BCBS and its international efforts. Documents issued by the BCBS are available through the Bank of International Settlements website: http://www.bis.org/. The BCBS's description of the LCR may be found here: http://www.bis.org/publ/bcbs238.pdf. Information about the Basel III framework may be found here: http://www.bis.org/bcbs/basel3.htm.

[4] The text of the final U.S. LCR rule, issued in September 2014, may be found here: https://www.federalreserve.gov/newsevents/pressreleases/bcreg20140903a.htm. For a descriptive overview of the LCR rule, see House, Sablik, and Walter (2016). Of course, the LCR is not the first incidence of liquidity regulation of U.S. financial institutions. Reserve requirements of depository institutions, administered by the Federal Reserve, were originally implemented as a prudential requirement to promote banks' liquidity positions. However, the prevalence of banks' retail deposit sweep programs, begun in the mid-1990s, meant that reserve requirements were not commonly a binding consideration for most large banks prior to the Financial Crisis. See Bouwman (2015) for a synthesis of the theoretical and empirical literature on the economics of how banks create liquidity and of related issues regarding liquidity requirements. A historical overview of liquidity regulation, ending with Basel III and the Dodd-Frank Act, is also provided.

[5] When the largest banks began publicly disclosing their LCR-related data for the quarter ending June 30, 2017, Bank of America, Citigroup, JP Morgan, and Wells Fargo each reported that Level 2B assets made up less than 1 percent of their total HQLA. LCR-related disclosure requirements for U.S. banks are described here: https://www.federalreserve.gov/newsevents/pressreleases/bcreg20161219a.htm.

[6] The original requirements for meeting the standard versus the modified LCR rule are as follows: Banks subject to the standard LCR requirement have $250 billion or more in total consolidated assets or $10 billion or more in on-balance sheet foreign exposure, or are these banking organizations' subsidiary depository institutions with assets of $10 billion or more. Banks that do not meet these thresholds but have $50 billion or more in consolidated assets are subject to the modified LCR requirement; the denominator of the modified LCR is multiplied by 70 percent. After the time period we analyze, the $50 billion threshold was raised to $100 billion by the Economic Growth, Regulatory Relief, and Consumer Protection Act in May 2018. More information is available in the related press release found here: https://www.federalreserve.gov/newsevents/pressreleases/bcreg20180706b.htm.

[7] The U.S. LCR was phased in as follows: Standard LCR banks were required to maintain an LCR of 80 percent beginning in January 2015, all LCR banks were required to maintain a 90 percent ratio beginning in January 2016, and full compliance—a ratio of 100 percent or more—was required beginning in January 2017.

[8] As we describe in our data appendix, for simplicity we use data on banks' total reserve balances instead of excess reserves because these measures differ little in aggregate over the period we primarily study.

[9] For an explanation of the mechanism by which increases in the Federal Reserve's security holdings, such as via LSAPs, resulted in a commensurate increase in the amount of reserve balances in the banking system, see the data appendix to Ihrig, Meade, and Weinbach (2015a or 2015b).

[10] While the Federal Reserve was holding the size of its securities holdings constant over this period, the aggregate stock of reserves in the banking system was declining, reflecting the ongoing growth in the outstanding stock of currency in circulation. That said, these banks' reserve holdings did not decline in proportion to the aggregate drop in reserve supply. In fact, Ihrig, Milchanowski, and Detering (2019) show that the vast majority of the decline in aggregate reserves can be accounted for by declines in the reserve balances of regional banks without LCR

restrictions, implying that large, LCR-restricted banks intentionally chose to retain reserve balances in their HQLA holdings.

11 The bank treasurer we model must also account for credit risk, which we assume to be small in this context.

12 Under Basel III, some BHCs must include net unrealized gains and losses from available-for-sale (AFS) securities in their calculation of the common equity tier 1 capital to risk-weighted assets ratio. For such institutions, changes in the valuation of AFS securities add volatility to their capital ratios. While both AFS and held-to-maturity (HTM) securities qualify as HQLA, only AFS securities can be immediately sold without accounting rule penalties. There has been a notable increase in the share of HTM securities at large banks since 2011, thus reducing the impact of interest rate risk on the regulatory capital of such banks.

13 Treasuries and MBS also carry some credit risk, which as noted we abstract from. For example, in August 2011, Standard & Poor's downgraded the credit rating of U.S. Treasury debt from AAA to AA+ shortly after Congress voted to raise the debt ceiling of the federal government by means of the Budget Control Act of 2011.

14 For example, in a 2017:Q4 earnings call (S&P Global, 2018), Citigroup's treasurer noted, "we're constantly looking at the optimization within HQLA, which is cash and HQLA securities, but then also looking at are there other investment opportunities that makes sense? … As obviously, [how] the interest rate environment changes or the spread environment changes [is] something we're constantly focused on and evaluating in terms of how to optimize."

15 For simplicity, our analysis ignores the trade-offs banks face in choosing among various maturities of Treasury securities and MBS. Given that rates of return are not perfectly correlated, the choice of the maturities of the assets considered does affect the estimated shares. For example, if in the last row of Table 2 a 2-year Treasury security is considered instead of a 5-year Treasury security (column 2), the optimal reserves share (column 4) rises from 0 to 16 percent (with the shares of Treasury securities and MBS falling to 50 and 34 percent, respectively). In addition, as already noted, for simplicity our analysis ignores other HQLA-eligible assets that are held in much smaller shares.

16 It is difficult to benchmark our coefficient of relative risk aversion, *A*, to values, or a range of values, that might be representative in our context. The coefficient is mostly used in the risk aversion literature for households; see Grandelman and Hernández-Murillo (2014) for a review. While the most commonly accepted estimates in that context lie between 1 and 3, a wide range of estimates have been presented—from as low as 0.2 to 10 and higher. Thus, we examine a wide range of parameter values.

17 We do not account here for the 15 percent haircut that is applied to banks' holdings of GSE MBS; doing so in this scenario would result in banks holding even more MBS to achieve the 60-40 portfolio. We exclude the haircut in our portfolio model because it is not straightforward to adjust the relative yields we consider to reflect such a constraint.

18 Setting the return on reserves equal to the effective federal funds rate over the entire sample period—including the early portion of this sample period—may constitute a return that was higher at times than in practice, which would upwardly bias the resultant reserve balance shares derived from our model. More generally, standard models of the federal funds market show that an increase in the rate at which federal funds are remunerated decreases the opportunity cost of holding reserves, suggesting that banks would demand more reserve balances as the remuneration rate rises (see Figure 5 of Ihrig et al., 2015a).

19 A summary of the survey findings can be found here: https://www.federalreserve.gov/data/sfos/files/senior-financial-officer-survey-201809.pdf.

20 Because the source data are confidential, we cannot show our individual, bank-level estimates of HQLA at the daily frequency.

21 We do not find it surprising that the GSIBs' MBS shares, the orange bars in Figure 6, exhibit relatively little daily volatility for each bank group. Because MBS are less liquid than Level 1 assets and subject to duration risk and eligible MBS holdings are capped under the LCR, banks likely largely position their MBS holdings at a desirable level and do not seek to adjust this component on the margin, at a high frequency. As evidence of this, GSIBs have a much higher share of their MBS holdings booked as HTM (that is, not AFS) than they do Treasuries—about 30 percent versus 20 percent, respectively.

22 See Andolfatto and Ihrig (2019) for a summary of how resolution planning in particular may be causing banks to hold sizable reserve balances in the current economic environment.

[23] Citigroup in its 2017:Q4 fixed income call (S&P Global, 2018) responded to a question about growth in its liquid assets by saying that its resolution planning "caused… growth in [its] liquidity balance, especially cash… in the first half of [2017]."

[24] See Bech, Martin, and McAndrews (2012) for evidence that banks' payment flows have moved to earlier in the day.

[25] Indeed, a majority of the respondents to the Senior Financial Officer Survey indicated that meeting potential deposit outflows was an important or very important determinant of their reserve demand.

[26] In an earnings call with market analysts, JP Morgan responded to a question about why it held so many excess reserves rather than higher-yielding assets such as MBS by pointing out that it already has high mortgage exposure taking both its mortgage loans and MBS holdings into account, and that further increasing its MBS holdings to augment returns would cause it to exceed its internal duration target (Thomson Reuters StreetEvents, 2017).

REFERENCES

Andolfatto, David and Ihrig, Jane. "Why the Fed Should Create a Standing Repo Facility." Federal Reserve Bank of St. Louis On the Economy Blog, March 6, 2019; https://www.stlouisfed.org/on-the-economy/2019/march/why-fed-create-standing-repo-facility.

Adrian, Tobias and Boyarchenko, Nina. "Liquidity Policies and Systemic Risk." Staff Report No. 661, Federal Reserve Bank of New York, 2013; https://doi.org/10.2139/ssrn.2377451.

Allen, Franklin. "How Should Bank Liquidity Be Regulated?" Speech at the Federal Reserve Bank of Atlanta, March 27, 2014; https://www.frbatlanta.org/-/media/Documents/news/conferences/2014/fmc/Allen.pdf.

Balasubramanyan, Lakshmi and VanHoose, David D. "Bank Balance Sheet Dynamics under a Regulatory Liquidity-Coverage-Ratio Constraint." *Journal of Macroeconomics*, 2013, *37*, pp. 53-67; https://doi.org/10.1016/j.jmacro.2013.03.003.

Banerjee, Ryan N. and Mio, Hitoshi. "The Impact of Liquidity Regulation on Banks." Staff Working Paper No. 536, Bank of England, 2015; https://doi.org/10.2139/ssrn.2635522.

Bech, Morten and Keister, Todd. "Liquidity Regulation and the Implementation of Monetary Policy." *Journal of Monetary Economics*, 2017, *92*, pp. 64-77; https://doi.org/10.1016/j.jmoneco.2017.09.002.

Bech, Morten; Martin, Antoine and McAndrews, Jamie. "How the High Level of Reserves Benefits the Payment System." Federal Reserve Bank of New York *Liberty Street Economics*, February 27, 2012; https://libertystreeteconomics.newyorkfed.org/2012/02/how-the-high-level-of-reserves-benefits-the-payment-system.html.

Bianchi, Javier and Bigio, Saki. "Banks, Liquidity Management, and Monetary Policy." Staff Report 503, Federal Reserve Bank of Minneapolis, 2017; https://www.minneapolisfed.org/research/sr/sr503.pdf.

Bouwman, Christa H.S. "Liquidity: How Banks Create It and How It Should Be Regulated," in Allen N. Berger, Philip Molyneux, and John O.S. Wilson, eds., *The Oxford Handbook of Banking*. Oxford University Press, 2015, 2nd Ed., pp. 184-218.

Cetina, Jill and Gleason, Katherine I. "The Difficult Business of Measuring Banks' Liquidity: Understanding the Liquidity Coverage Ratio." Working Paper No. 15-20, Office of Financial Research, 2015; https://doi.org/10.2139/ssrn.2681372.

Cuoco, Domenico and Liu, Hong. "An Analysis of VaR-Based Capital Requirements." *Journal of Financial Intermediation*, 2006, *15*, pp. 362-94; https://doi.org/10.1016/j.jfi.2005.07.001.

Covas, Francisco and Driscoll, John C. "Bank Liquidity and Capital Regulation in General Equilibrium." Finance and Economics Discussion Series 2014-85, Board of Governors of the Federal Reserve System, 2014; https://www.federalreserve.gov/econresdata/feds/2014/files/201485pap.pdf.

Diamond, Douglas W. and Kashyap, Anil K. "Chapter 29 - Liquidity Requirements, Liquidity Choice and Financial Stability," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*. Volume 2. Elsevier, 2016, pp. 2263-303; https://doi.org/10.1016/bs.hesmac.2016.03.011.

Duffie, Darrell and Krishnamurthy, Arvind. "Passthrough Efficiency in the Fed's New Monetary Policy Setting." Working paper presented at the 2016 Jackson Hole Symposium of the Federal Reserve Bank of Kansas City, 2016; https://www.kansascityfed.org/~/media/files/publicat/sympos/2016/econsymposium-duffie-krishnamurthy-paper.pdf?la=en.

Grandelman, Néstor and Hernández-Murillo, Rubén. "Risk Aversion at the Country Level." Working Paper 014-005B, Federal Reserve Bank of St. Louis, October 2014; https://s3.amazonaws.com/real.stlouisfed.org/wp/2014/2014-005.pdf.

Halaj, Grzegorz. "Optimal Asset Structure of a Bank: Bank Reactions to Stressful Market Conditions." Working Paper Series No. 1533, European Central Bank, April 2013.

Halaj, Grzegorz. "Dynamic Balance Sheet Model with Liquidity Risk." Working Paper Series No. 1896, European Central Bank, April 2016; https://doi.org/10.2139/ssrn.2625983.

House, Mark; Sablik, Tim and Walter, John. "Understanding the New Liquidity Coverage Ratio Requirements." Economic Brief EB16-01, Federal Reserve Bank of Richmond, 2016.

Ihrig, Jane E. "Banks' Demand for Reserves in the Face of Liquidity Regulations." Federal Reserve Bank of St. Louis On the Economy Blog, March 5, 2019.

Ihrig, Jane E.; Milchanowski, Mike and Detering, Jennifer. "Which Banks Have Been Shedding Reserves?" Federal Reserve Bank of St. Louis On the Economy Blog, April 18, 2019; https://www.stlouisfed.org/on-the-economy/2019/april/bank-shedding-reserves.

Ihrig, Jane E.; Meade, Ellen E. and Weinbach, Gretchen C. "Monetary Policy 101: A Primer on the Fed's Changing Approach to Policy Implementation." Finance and Economics Discussion Series 2015-047, Board of Governors of the Federal Reserve System, 2015a; https://doi.org/10.17016/FEDS.2015.047.

Ihrig, Jane E.; Meade, Ellen E. and Weinbach, Gretchen C. "Rewriting Monetary Policy 101: What's the Fed's Preferred Post-Crisis Approach to Raising Interest Rates?" *Journal of Economic Perspectives*, 2015b, *29*(4), pp. 177-98; https://doi.org/10.1257/jep.29.4.177.

Quarles, Randal K. "Liquidity Regulation and the Size of the Fed's Balance Sheet." Presented at "Currencies, Capital, and Central Bank Balances: A Policy Conference," Stanford University, May 4, 2018.

S&P Global Market Intelligence. "Citigroup Inc. Fixed Income Call." January 24, 2018, accessed through SNL Financial.

Thomson Reuters StreetEvents. "Edited Transcript: JPM – Q3 2017 JPMorgan Chase & Co Earnings Call." October 12, 2017, accessed through SNL Financial. (Presentation available at https://www.jpmorganchase.com/corporate/investor-relations/quarterly-earnings.htm.)

An Empirical Economic Assessment of the Costs and Benefits of Bank Capital in the United States

Simon Firestone, Amy Lorenc, and Ben Ranish

We evaluate the economic costs and benefits of bank capital in the United States. The analysis is similar to that found in previous studies, though we tailor it to the specific features and experience of the U.S. financial system. We also make adjustments to account for the impact of liquidity- and resolution-related regulations on the probability of a financial crisis. We find that the level of capital that maximizes the difference between total benefits and total costs ranges from just over 13 percent to 26 percent. This range reflects a high degree of uncertainty and latitude in specifying important study parameters that have a significant influence on the resulting optimal capital level, such as the output costs of a financial crisis or the effect of increased bank capital on economic output. Finally, the article discusses a range of considerations and factors that are not included in the cost-benefit framework that could have a substantial impact on estimated optimal capital levels. (JEL G18, G21, G28)

Federal Reserve Bank of St. Louis *Review*, Third Quarter 2019, *101*(3), pp. 203-30.
https://doi.org/10.20955/r.101.203-30

1 INTRODUCTION

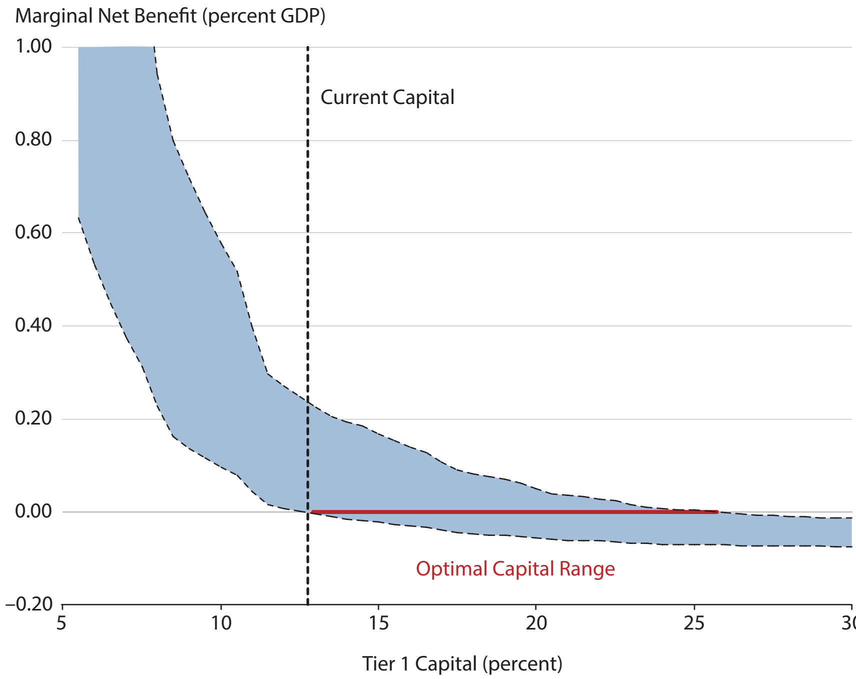
We perform an economic analysis of the long-term costs and benefits of different levels of system-wide bank capital and estimate optimal Tier 1 capital levels under a range of modeling assumptions.[1] Within our framework, the benefit of bank capital is to reduce the probability of costly future financial crises. The marginal benefit of bank capital generally decreases as capital levels rise; the potential improvement from reducing the frequency of crises becomes more limited as the frequency of crises approaches zero. The aggregate economic costs of bank capital stem from an increase in banks' cost of capital. This increase is passed on to borrowers in the form of higher credit costs and lowers gross domestic product (GDP). In our framework, the marginal cost of bank capital is constant.

The shaded region in Figure 1 shows a range for our estimated marginal net benefits of capital. At levels of capital up to about 13 percent, the shaded region lies above the horizontal

Simon Firestone is a principal economist, Amy Lorenc was a senior financial institution policy analyst, and Ben Ranish is a senior economist at the Board of Governors of the Federal Reserve System.

Figure 1

Estimates of the Net Benefits of Additional Tier 1 Capital (Marginal Net Benefits of Capital)



axis. This implies that our estimated range for the benefits of additional capital remains positive until Tier 1 capital ratios reach 13 percent. For levels of capital between 13 percent and 26 percent, the shaded region overlaps the horizontal axis. This overlap implies that our estimated benefits of additional capital for this range may be positive or negative, depending on the modeling assumptions used.

The width of the shaded region in the plot represents uncertainty around our estimates of the costs and benefits associated with the level of bank capital. The upper bound of this region represents our low estimate of the costs and high estimate of the benefits of bank capital. Specifically, it assumes that the effects of financial crises are permanent and that banks pass only 50 percent of capital-related cost increases on to borrowers. The lower bound of this region represents our high estimate of the costs and low estimate of the benefits of bank capital. Specifically, it assumes that the effects of crises diminish gradually over time and that banks pass all capital-related cost increases on to borrowers.[2] We note that this plot shows levels of actual capital rather than minimum required capital. Because banks generally hold a buffer above and beyond the minimum required capital, we expect that optimal minimum capital requirements may be a bit below the range shown in Figure 1.

There is an extensive literature on financial crises in advanced economies and the relationship between bank capital and macroeconomic risk. Our methodology builds on studies by the Basel Committee on Banking Supervision (BCBS) (2010), Brooke et al. (2015), and the

Federal Reserve Bank of Minneapolis (2016). All studies use the same basic framework; the effects of increased bank capital on the probability and severity of a crisis are compared with the increase in the cost of credit and associated reduction in GDP level. The BCBS (2010) study uses a meta-analysis of the academic literature, combined with results from various country-specific supervisory models, to quantify these effects. Brooke et al. (2015) and the Minneapolis Fed (2016) use both the existing literature and substantial original data analysis for calibration to the U.K. and U.S. economies.

Our approach differs from these studies in some significant ways. For example, we use adjustments and controls to account for the effects of new liquidity requirements and resolution requirements for failing firms. We also use Romer and Romer's (2015) generalized least squares (GLS) estimates of the severity of financial crises to reduce the result's dependence on data from inherently more volatile and smaller economies that are arguably less relevant to the United States.[3] We provide estimates of the severity of financial crises that assume permanent effects of financial crises on GDP, as well as alternative estimates that assume persistent but decaying effects. Finally, unlike the BCBS (2010) and Brooke et al. (2015) studies, we design the research to ensure, where possible, that the analysis is tailored to the specific features of the U.S. financial system so that the results are relevant for considering capital regulatory policy in the United States. Our results imply larger optimal capital levels than those in Brooke et al. (2015) and similar levels to those in the BCBS (2010) and Minneapolis Fed (2016) studies. Like past studies, our framework addresses broad changes in capital rather than targeted requirements that apply to specific banks, such as the global systemically important banks (GSIBs) surcharge and the Comprehensive Capital Analysis and Review (CCAR). Our framework assumes that all banks choose the same capital ratios, and we do not account for the heterogeneity of the U.S. capital framework resulting from targeted regulations.

2 INSTITUTIONAL ENVIRONMENT

Two new U.S. enhancements to large-bank safety and soundness affect the relationship between bank capital levels and the macroeconomy: increased resolvability of failing firms and liquidity requirements. In our analysis, we consider the effects of capital in the presence of both.

A number of resolution-planning requirements have been adopted to ensure rapid and orderly resolution in the event of financial distress or failure of a company. Of particular importance are the long-term debt and the total loss-absorbing capacity (TLAC) requirements. These requirements are designed to create a source of funds for recapitalization.[4] They apply to top-tier U.S. bank holding companies that have been designated as GSIBs and to U.S. intermediate holding companies of foreign GSIBs. These institutions must maintain a certain amount of eligible long-term unsecured debt that can be converted to equity for the purpose of absorbing losses or recapitalization in the event of failure. Requirements became effective January 1, 2019.

In addition to complying with the long-term debt and TLAC requirements, large U.S. bank holding companies must submit annual orderly resolution plans, commonly known as living wills, to the Federal Reserve System. These plans should facilitate the rapid and orderly resolution of a firm in the event of failure and prevent contagion among broader financial markets. Beyond resolution-planning requirements, the orderly liquidation authority provided to the Federal Deposit Insurance Corporation (FDIC) by Title II of the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 increases the resolvability of financial firms. The orderly liquidation authority provides an alternative to standard bankruptcy that allows the FDIC to carry out the liquidation of a failing financial company, aiming to minimize systemic risk and moral hazard.

Shocks to the value of financial system assets are amplified by liquidity shortfalls. Banks with short-term funding requirements exceeding their liquid assets may find themselves vulnerable to "fire sales" of illiquid assets when such shocks occur. Through their price impact, fire sales can trigger the loss of funding liquidity and asset sales at other banks, potentially threatening the solvency of the entire system.[5] Consequently, liquidity requirements are also likely to reduce the probability of a financial crisis. We include adjustments for the effect of liquidity regulations in our estimated financial crisis probabilities.

The liquidity coverage ratio (LCR)[6] was adopted in the United States in 2014, with all phase-in completed by January 2017. It applies in full to all U.S. institutions with at least $250 billion in total assets or at least $10 billion in on-balance sheet foreign assets. Less-stringent requirements apply to institutions with between $50 billion and $250 billion in total assets and less than $10 billion in on-balance sheet foreign assets. The LCR requires banks to maintain an amount of high-quality liquid assets sufficient to cover expected net cash outflows during a 30-day stress period. This amount is a weighted sum of specific high-quality liquid assets, with higher weights on the most-liquid assets. Expected net cash outflows are fixed by regulation and based on the type of liabilities. Shorter-term and more-runnable liabilities are associated with the highest outflow rates.

Both resolution and liquidity reforms are likely to reduce the expected frequency and costs of future financial crises, and thus reduce the marginal benefits of bank capital. As these reforms are new, their effect is subject to considerable uncertainty. In addition, the regulatory landscape continues to evolve as regulators consider ways to simplify and tailor post-crisis reforms. Future changes to resolution and liquidity reforms could require a reassessment of their potential impact on future financial crises. With that said, we discuss how we incorporate the current expected effects of these enhancements to large-bank safety and soundness within our framework in the next section.

3 THE ECONOMIC BENEFITS OF BANK CAPITAL

To measure the benefits of bank capital, we first estimate bank capital's role in reducing the probability of a financial crisis (Section 3.1) and then multiply this by the costs of financial crises (Section 3.2). We rely on definitions of "financial crisis" developed in previous surveys, as defining crisis is a non-trivial exercise. When estimating the probability of a financial crisis, we rely on criteria and crises identified by Laeven and Valencia (2012). When estimating the costs of financial crises, we use results from Romer and Romer (2015), who developed a continuous narrative-based measure of financial distress.[7] We discuss additional, potentially significant, benefits of bank capital not captured in this framework (Section 6).

3.1 Probability of a Financial Crisis

As with past studies, we estimate the probability of a financial crisis starting via the following two approaches with distinct strengths and weaknesses: a "bottom-up" bank-level simulation and a "top-down" country-level regression. In our reported results, we use the average of probabilities generated by the two approaches.[8] For use in both approaches, we collect data spanning 24 developed countries from the Organisation for Economic Co-operation and Development (OECD) for 1988-2014.[9] The two sections that follow describe each of these approaches and the results they deliver in greater detail.

Bottom-Up Approach. In this approach, we simulate net income for U.S. banks and then determine whether the aggregate capital shortfall that results is large enough to classify as a financial crisis, using the definition of financial crisis described in previous surveys. In the process, we adjust the simulated income to account for large U.S. banks now being subject to liquidity requirements as well. This adjustment is discussed in more detail below. To implement this approach, we use bank-level data on total assets, risk-weighted assets (RWA), net income, Tier 1 capital, and asset liquidity ratios from Bankscope. These data have good coverage for 1988-2014.[10] We use these historical data to simulate net income for U.S. banks. For each simulation, we first randomly pick a single country-year scenario; for example, "France 2008." Then, for each of the 5,935 U.S. banks at the end of 2015, we randomly draw (with replacement) a value for two-year net income/RWA from within the set of large banks present in that country-year scenario.[11,12] We use two-year net income to roughly match the length of CCAR stress periods. This captures the fact that bank losses are predictable over multiple years in stress periods, possibly because of the smoothing of asset value shocks on balance sheets.[13]

Next, we compute the aggregate capital shortfall that results from the simulated net income draws and a range of initial Tier 1 capital levels. We follow Brooke et al.'s (2015) approach by identifying crises as simulations where the aggregate shortfall exceeds 3 percent of GDP ($540 billion), the level Laeven and Valencia (2012) define as a significant recapitalization. This shortfall is measured relative to a threshold of 8.24 percent. This threshold is chosen so that the average probability of a crisis generated by the model when applied to the United States (1989-2014) equals the average probability of a crisis in an advanced economy (4.0 percent per Laeven and Valencia, 2012).[14] With our shortfall threshold and crisis definition, if all U.S. banks had the same Tier 1 capital ratios, a crisis occurs whenever this ratio falls below about 4.7 percent. We run our simulation 10,000 times, each time using a random set of net income draws from a randomly selected country-year scenario as described above.

For most of the historical data period, major liquidity regulations were not in effect. If they were, banks may have had more liquid assets to sell off during stress periods, reducing the extent of loss-magnifying fire sales. To account for the effect of liquidity regulations on loss magnitudes, we first estimate the relationship between asset liquidity and loss magnitudes. We use this to adjust the draws of negative bank income upward by an approximation of how

Table 1

Results of the Bottom-Up Approach: Estimated Probability of a Financial Crisis

Tier 1 capital ratio (%)	No adjustment for liquidity regulations or increased resolvability [1]	Adjustment for liquidity regulations, but not increased resolvability [2]	Adjustment for liquidity regulations, and 30% reduction for increased resolvability [3]
8.0	3.8	3.2	2.6
11.0	1.9	1.8	1.3
14.0	1.4	1.4	1.0
17.0	1.0	1.0	0.7
21.0	0.8	0.7	0.5
25.0	0.7	0.7	0.5

SOURCE: Bankscope and authors' calculations.

much higher income would have been if the draw had been from an LCR-compliant bank. This higher income translates into a lower aggregate net shortfall, and thus a lower probability of a financial crisis. Further details of this adjustment are available in the appendix.

Finally, other impact studies have suggested that loss-absorption requirements, such as TLAC, may reduce the probability of a financial crisis. In theory, greater reliance on debt that is priced to reflect risk may restrain bank risk-taking. TLAC may reduce the risk of individual or even systematic bank failures. Based on this argument, the Financial Stability Board (2015) relies on work by Afonso, Santos, and Traina (2014) and Marques, Correa, and Sapriza (2013) to estimate that TLAC reduces the probability of a financial crisis by 30 percent.[15]

Table 1 shows our bottom-up estimates of the probability of a financial crisis across a wide range of Tier 1 capital levels. Column [1] presents our estimates without the liquidity requirement-based adjustment to loss magnitudes described above. Column [2] includes this adjustment, showing that it reduces the estimated probability of a financial crisis by about 3 to 14 percent, depending on the level of capital. Column [3] presents a 30 percent reduction applied to all probabilities to account for increased resolvability of failing firms. Given the multiple uncertainties in this approach, we treat the magnitude of this adjustment with caution.

The probability of a financial crisis declines sharply as capital levels rise. However, the rate of this change in probability, in absolute and relative terms, is more gradual as the level of capital rises. This observation is due to the fat-tailed historical distribution of stress losses, upon which all empirical bottom-up approaches rely. Once the relatively common milder stress periods are avoided (at Tier 1 ratios around 15 percent), there are relatively few crises remaining that can be avoided without substantially greater amounts of capital. Liquidity requirements have a modest effect on the probability of a crisis, with the exception of cases when capital is very low.

Top-Down Approach. In our top-down approach, we use a cross-country logit model to predict the probability of a financial crisis starting given the financial system's capital and asset liquidity, and control for other sources of economic fragility. Our annual data covers the same set of countries and years as used in our bottom-up approach, although we omit country-years where there is an ongoing crisis.[16] We adopt the definition and dates of systemic banking crises developed in Laeven and Valencia (2012). This definition requires the following two conditions be met:

(i) significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations); and

(ii) significant banking policy intervention measures in response to significant losses in the banking system.

This definition is more flexible than the capital shortfall-based definition of a crisis used in the bottom-up approach, allowing for a more nuanced definition of what constitutes a crisis.

We derive country-level Tier 1 capital and asset liquidity ratios for each year using the bank-level data from Bankscope described previously. We collect implied volatility from the Chicago Board Options Exchange's Volatility Index® (VIX®) and the ratio of total private-sector credit to GDP from the World Bank.[17] To further control for fragility related to trade imbalances and asset valuation, we also collect data on the current account balance (as a percent of GDP) from the World Bank and the price-to-income ratio from the OECD. Note that VIX® is the only explanatory variable in our model that is not country specific.

We run two separate logit regressions. These regressions are indicated by the two specifications below, where the function *f* is the logistic function, and *c* and *t* denote country and year, respectively. Specification (1) uses only VIX® and credit to GDP as macroeconomic controls. Specification (2) adds current account balance and home price-to-income ratios. We opt to use two specifications because the additional control variables in specification (2) are missing for around one-third of our observations. The first specification thus uses a larger sample, while the second specification uses a more complete set of variables.

$$(1)\quad Probability\left(Crisis_{ct}\right)=f\left(Tier1\_Capital_{c,t-1},LiquidAsset\%_{c,t-1},VIX_{t-1},Credit\_GDP_{c,t-1}\right)$$

$$(2)\quad Probability\left(Crisis_{ct}\right)=f\begin{pmatrix}Tier1\_Capital_{c,t-1},LiquidAsset\%_{c,t-1},VIX_{t-1},\\Credit\_GDP_{c,t-1},CurrAcct\_GDP_{c,t-1},\ HomePrice\_Inc_{c,t-1}\end{pmatrix}$$

Results of both specifications are shown in Table 2. Specification (1) is a more appropriate fit for the data based on log likelihoods and the Akaike Information Criterion, but it is important to note that the differences in fit and common coefficients between the specifications are largely driven by the completeness of the data in the samples rather than the covariates chosen. The VIX® is highly statistically significant in both specifications. Liquid asset share is not statistically significant in either specification, which will impact how we consider the effect of liquidity regulations, as discussed shortly.

We first use estimates from these two regressions to predict the probability of a crisis under counterfactual levels of Tier 1 capital without any adjustments for regulations put in

Table 2

Regression Results of Both Specifications of the Top-Down Approach

	Dependent variable	
	Probability of a financial crisis	
	Specification (1)	Specification (2)
Tier 1 ratio	–7.251 (10.089)	–24.733* (15.691)
Liquid asset share	1.744 (1.768)	4.034 (2.603)
Credit to GDP ratio	1.243*** (0.454)	0.733 (0.702)
VIX®	0.090*** (0.031)	0.095*** (0.035)
Current account balance		–3.388 (5.231)
Price-to-income ratio		2.707** (1.134)
Constant	–5.879*** (1.230)	–6.893*** (1.797)
Observations	434	328
Log likelihood	–82.144	–66.512
Akaike Information Criterion	174.288	147.024

NOTE: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. GDP, gross domestic product; VIX®, Chicago Board Options Exchange's Volatility Index®.

Table 3

Results of the Top-Down Approach

	Annual probability of a financial crisis (%)			
	No adjustment		Adjustment for increased resolvability	
Tier 1 ratio	Specification (1)	Specification (2)	Specification (1)	Specification (2)
8.0	6.3	6.8	4.4	4.7
11.0	5.2	3.3	3.6	2.3
14.0	4.3	1.5	3.0	1.1
17.0	3.5	0.7	2.5	0.5
21.0	2.7	0.2	1.9	0.2
25.0	2.0	0.1	1.4	0.1

place after the 2007-09 Financial Crisis. For each Tier 1 capital level, we take the average of probabilities across all country-years as our estimate of the average probability of a crisis over the business cycle (i.e., over the distribution of non-crisis macroeconomic states) for that capital level. These baseline results are in columns 2 and 3 of Table 3. These results may overestimate the probability of a crisis because recent regulations have greatly increased the resiliency of the U.S. financial system and are not reflected in the data used to estimate the regressions.

To account for this overestimation of the probability of a financial crisis, we consider the effects of compliance with new liquidity regulations and the increased resolvability of failing financial firms. Liquidity regulations, particularly the LCR requirement, would be best captured by our share-of-liquid-asset variable. As shown in Table 2, the estimated coefficient for this variable is positive but insignificant in both specifications of our model. For this reason, accounting for liquidity regulation compliance would not significantly alter the results of the top-down approach, and we do not perform an adjustment for liquidity regulation.

We next consider the effect of increased resolvability of failing financial firms. We follow the adjustment method of the bottom-up approach and assume a 30 percent reduction in the probability of a crisis as a result of the increased resolvability of firms.[18] Results reflecting this reduction are shown in the final two columns of Table 3. These results are most representative of the probability of a crisis in the current regulatory environment.

The differences between the two specifications show that the relatively limited number of financial crises mean that results can be sensitive to the sample selection and choice of macroeconomic control variables. As in the bottom-up approach, both specifications of the model attribute higher Tier 1 capital ratios to lower probabilities of crises. However, specification (2) shows far greater sensitivity of probabilities to capital ratios and produces estimates closer to those of the bottom-up approach. In this case, most of this difference is driven by differences between the data sample used in the two specifications.[19]

3.2 Costs of Financial Crises

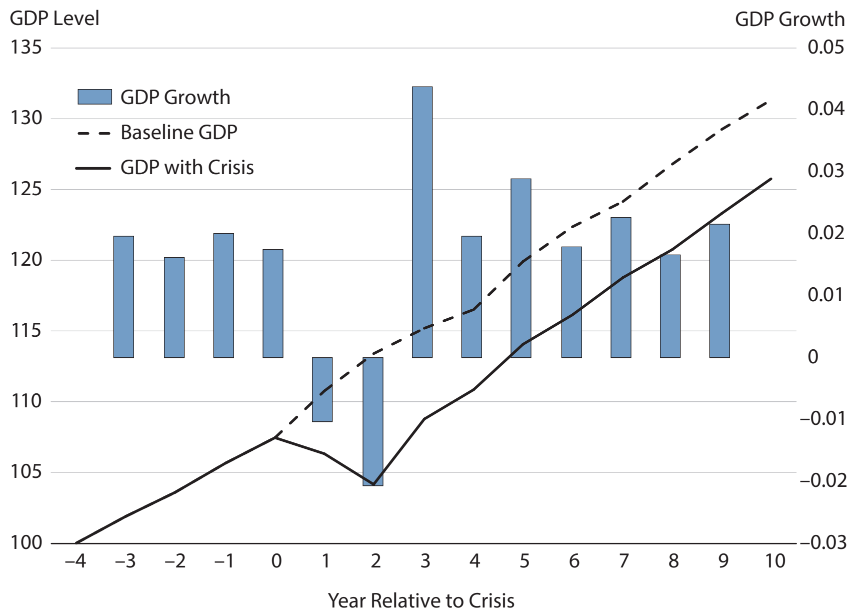
There are several reasons why financial crises reduce economic growth. A negative shock to lending reduces demand for output, leading to less investment in capital and research. The increased uncertainty of an economy in crisis can further reduce investment. We quantify the size of this effect in this section.

There are three steps to measuring the effects of a financial crisis on GDP: deciding if a crisis's effects are permanent, measuring the time-varying effects of the crisis throughout its duration, and calculating the present value of the sum of these effects over time. If there are no permanent effects, the cost is simply the present value of the short- and medium-term effects. We discuss each of these steps below.

Does a Crisis Have Permanent or Persistent Effects? Some past studies have assumed that crises have only temporary effects, while others assume a permanent decrease in output. The duration of the effects is especially important because it can change the cost estimate by an order of magnitude. To illustrate the difference between the two approaches that measure the cost of a crisis, consider Figure 2. The dashed line represents growth without a crisis, while the solid line represents the baseline until a financial crisis reduces growth for two years. The blue bars are annual GDP growth rates. Exiting the crisis leads to one year of above-trend growth, followed by a return to trend growth. The transitory costs of a crisis would be given by the area of the difference between the solid and dashed lines for the first two years of the crisis, adjusted for discounting.[20] Measuring the permanent effects would include the area of the discounted difference between the two lines from the start of the crisis. Of course, if

Figure 2

GDP Level (left axis) and Growth (right axis), Hypothetical Economy



the effects are not permanent, then at some point the solid and dashed lines coincide and the effects of a crisis are reduced to zero, as is the area between the two lines beyond that point.

Furceri and Mourougane (2012) analyze OECD countries and compare actual output after a crisis with a measure of potential input. They estimate autoregressive equations and the implied impulse response functions, finding an average permanent reduction in GDP of 2 percent per year. Cerra and Saxena (2008) analyze data from over 120 countries and find evidence that effects of a financial crisis on GDP are barely reduced by 1 percentage point after 10 years, remaining at a level of 6 percent. These studies provide evidence for robust long-lasting effects. Given the empirical evidence, we think it is sensible to include long-lasting effects. However, as it is very difficult to statistically distinguish between permanent and highly persistent effects, we present results using an alternative assumption of a persistent-but-temporary effect that fades gradually over time.

We use the Gordon Growth Model to compute the present value of the long-term costs of a financial crisis as

$$PV = \frac{Initial\ GDP\ Impact}{Discount\ Rate + Rate\ of\ Decay\ of\ Impact}.$$

This value is as of the beginning of the perpetuity flow, so we further discount it back to the beginning of the crisis. Cerra and Saxena (2008) find that the annual loss in GDP from a banking crisis shrinks by only 25 percent after 10 years, equivalent to an annual decay rate of

Table 4

Effect of Financial Crises on GDP (%), by Horizon

	Time since start of crisis			
Specification	0	2.5 years	5 years	Largest effect
OLS	–2.08	–4.72	–4.61	–5.96
GLS	–1.24	–3.26	–2.25	–4.07

NOTE: OLS, ordinary least squares; GLS, generalized least squares.

SOURCE: Romer and Romer (2015) and Table 2.

2.25 percent. There is considerable uncertainty around their point estimate, so we choose an annual decay rate of 5 percent to establish a larger range for our cost-of-crisis estimate.

Notice that in the growth formula, the present value is unaffected by whether an increase of a set amount impacts the discount rate or the decay rate. Therefore, in quantifying the costs of crises, we could think of this decay as equivalent to a 5 percent increase in the discount rate applied to the years following a crisis.

Short- and Medium-Term Severity of a Crisis. We use results from Romer and Romer (2015) of financial crises in advanced economies to estimate the short- and medium-term costs of a financial crisis in the United States.[21] The data are from 1967-2012 and include the 24 OECD members as of 1973. The authors use narrative data from semiannual OECD reports to generate a continuous measure of the intensity of a crisis and run a regression with GDP growth, the state of a country's financial system, lags of both variables, and other controls.[22]

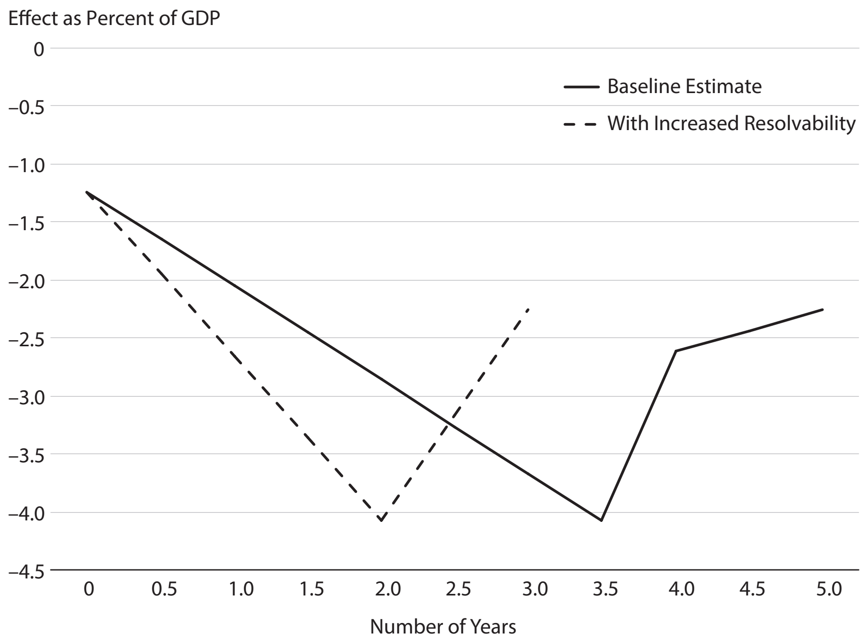
Romer and Romer (2015) present estimates of the effect on GDP using a standard, ordinary least squares (OLS) regression, which weights each observation equally. They also present estimates using a statistically superior GLS approach, which provides less weight to observations from countries with a high variance of GDP growth. Most of these high-variance countries are smaller or less economically advanced. For example, the variances of GDP in Turkey, Iceland, and Ireland are much larger than that in the United States or other advanced countries. Such countries' economies are more influenced by international capital markets and/or sovereign risk. Nonetheless, data on financial crises are limited, and it is useful to include these countries in the analysis. As a compromise, we focus on Romer and Romer's GLS estimates, as they provide less weight to these high-variance countries that are arguably less comparable to the United States.

The authors give estimates of the reduction in GDP due to financial crises in advanced economies on a semiannual basis over five years. Their results with both GLS and OLS are shown in Table 4.[23] Both specifications show that a relatively small effect, which is contemporaneous with a crisis, becomes larger in magnitude and reaches a peak after two-and-a-half years, slowly becoming smaller afterward.

Total Cost of a Crisis. We use a linear interpolation of Romer and Romer's (2015) GLS results for the effects of a financial crisis up to five years and assume a permanent reduction

Figure 3

Short- and Medium-Term Effects of a Financial Crisis on GDP



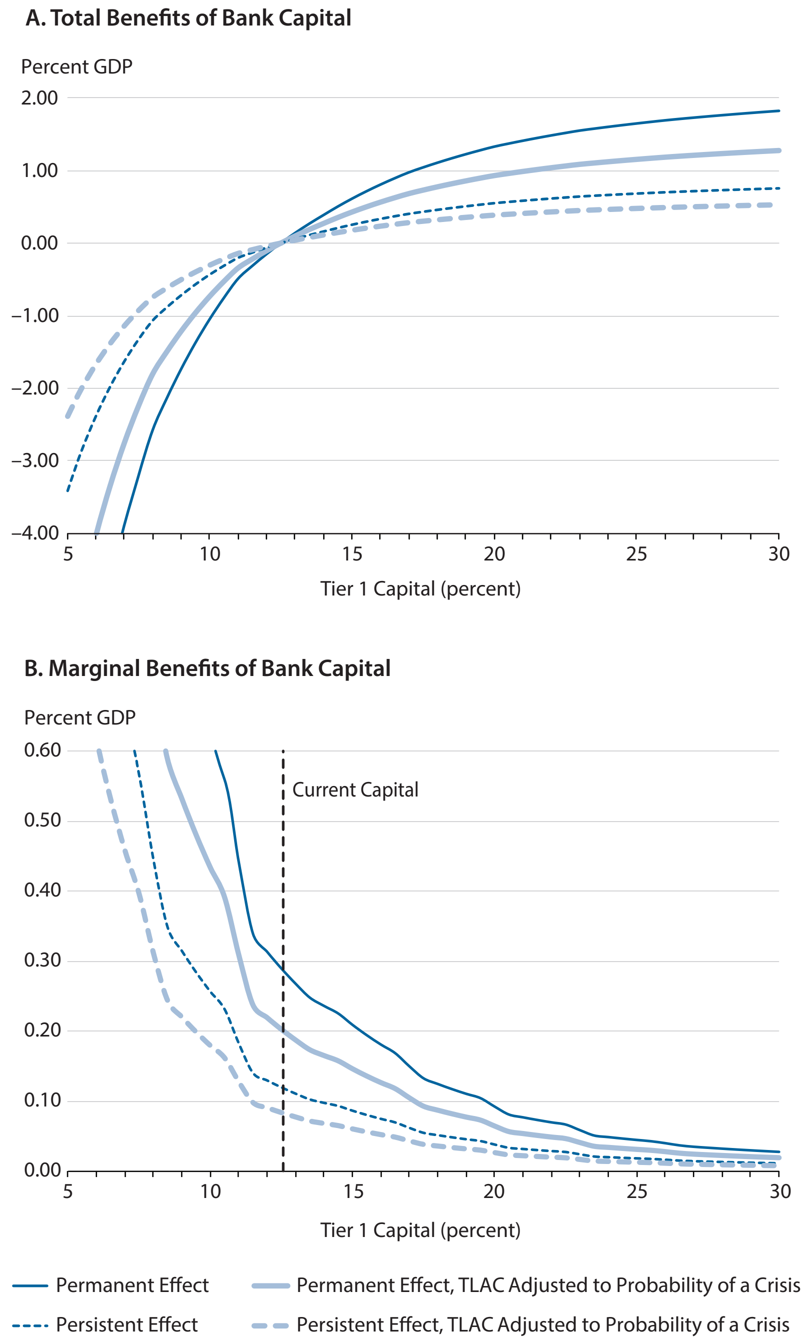
SOURCE: Baseline is a linear interpolation of Romer and Romer's (2015) Table 2 results with GLS estimation. "With increased resolvability" assumes that, per Homar and van Wijnbergen (2016), the duration of a crisis's effects is reduced. We assume the beginning, maximum, and final effects are as in the baseline, and again use linear interpolation to obtain results with prompt recapitalization requirements.

in GDP, as found by Furceri and Mourougane (2012), of 2 percent afterward. We also calculate the impact of an effect whose magnitude decays 5 percent a year.

As discussed below, we adjust these results for the effects of prompt recapitalization requirements and the increased resolvability of failing financial firms. To turn these estimates into a net present value, we need to apply a discount rate to future reductions in GDP. Standard asset pricing models imply that the real risk-free rate reflects the intertemporal preferences for the economy. We use the average real yield on 10-year Treasury bonds, or 2.7 percent.[24] The magnitude of the effect is quite sensitive to assumptions about decay (discount) rates, falling by more than half if we assume a 5 percent rate of decay. Improved resolvability requirements, discussed above, likely will cause future recapitalizations to happen more rapidly than was typical for advanced countries in the past. Research suggests that prompt recapitalization reduces the duration of financial crises. Homar and van Wijnbergen (2016) conclude that with prompt recapitalizations, the time to a GDP trough is two years rather than three-and-a-half years, and the duration of effects is three years rather than five years. We maintain the Romer and Romer (2015) estimates of the peak and final semiannual costs of a financial crisis and calculate the revised costs given the shorter duration. Figure 3 shows the estimated semi-

Figure 4

Total and Marginal Benefits of Bank Capital, Relative to Current U.S. Average Tier 1 Capital Ratios (12.5 percent)



annual costs under this assumption. With the effects of increased resolvability, the estimated costs of financial crises fall by 3 to 4 percentage points of GDP.[25]

Overall, assuming a permanent effect, we calculate the cost of a crisis as 99 percent of GDP. With a 5 percent rate of decay, the cost is 41 percent of GDP, a reduction of almost 60 percent. This wide range shows the importance of assumptions regarding the long-range effects of a crisis.

3.3 Estimated Range of Benefits of Capital

Within our framework, the benefits of bank capital equal the reduction in the probability that a financial crisis starts (estimated in Section 3.1), multiplied by the costs of the avoided financial crises (estimated in Section 3.2). Our estimates of each of these components vary because of different modeling assumptions. We either assume that increased resolvability has no effect or apply a 30 percent reduction in the probability of a financial crisis for it. We compute the cost of a crisis assuming that crises have either a permanent effect or a transitory, though highly persistent, effect on the level of GDP.

Figure 4 plots the range in benefits estimates due to this modeling variation.[26] Benefits are plotted relative to the benefits received at the current average level of Tier 1 capital ratios in the United States. Figure 4B, which presents marginal benefits, shows that at current capital levels, the marginal benefit of increasing capital ratios by 1 percentage point likely falls between 8 and 27 basis points of GDP per year. However, at lower capital levels, this marginal benefit is far higher.[27]

4 ECONOMIC COSTS OF BANK CAPITAL

Increased bank capital reduces GDP through its impact on lending costs. Banks' weighted average cost of capital rises in response to higher capital, and some of this cost increase may be passed on to borrowers. A higher cost of credit discourages investment, thus decreasing the equilibrium level of GDP. In the macroeconomic model we use, long-term growth rates are unaffected by bank capital.

The first step in estimating the impact of bank capital on the level of GDP is to estimate how changes in capital affect the cost of loans. To do this, we estimate the impact of capital on a representative bank's weighted average cost of capital. Next, we calculate how much loan rates would have to change to offset these changes to the cost of capital. We assume that banks are able to pass either 50 or 100 percent of these changes on in the form of increased lending spreads.

Once we have estimated the impact of capital on lending costs, we use the Federal Reserve Board (FRB)/US macroeconomic model to assess the effect of these changes in lending costs on the equilibrium level of GDP.[28] The FRB/US macroeconomic model is a large-scale macro model used by Federal Reserve staff to track the evolution of the U.S. economy over time. The model is well documented and versions of it are freely available to the public.[29] It is used here as a baseline economic model to translate increased lending spreads into reduced GDP levels and follows previous work conducted by the Minneapolis Fed (2016).

Table 5

Studies Showing Evidence of Modigliani-Miller (MM) Offsets

Study	Geography of sample	MM offset
Kashyap, Stein, and Hanson (2010)	U.S.	Partial
Clark, Jones, and Malmquist (2015)	U.S.	65-100%
Junge and Kugler (2013)	Switzerland	37%
Miles et al. (2013)	U.K.	45%
Toader (2015)	Europe	42%
European Central Bank (2011)	International	41-73%

NOTE: Clark, Jones, and Malmquist's (2015) results depend on the size of the bank and the time period. They find full offsets for large banks since the 2007-09 Financial Crisis, and a 65 percent offset for banks with assets between $100 and $200 billion.

4.1 Bank Capital Structure and the Modigliani-Miller Theorem

A key driver of the cost of increased capital requirements is applicability of the Modigliani-Miller theorem (Modigliani and Miller, 1958) to U.S. banks. To the extent that the theorem applies, banks' funding costs are insensitive to their capital ratios, and the resulting economic costs of bank capital are reduced. Notable articles arguing that the theorem applies to banks include Miller (1995) and Admati et al. (2014). However, there are reasons, even in the presence of full Modigliani-Miller effects, that higher capital may result in higher funding costs. These are discussed in Elliott (2013), among other places. For example, the Modigliani-Miller theorem assumes that capital structure has no effect on taxes. Interest payments on debt are tax-deductible, serving as a "debt shield," while dividends are not, so higher equity means that banks pay more in taxes.

Ultimately, whether the theorem applies is an empirical question. A common approach is to consider the effects of changes in bank capital structure on the risk of bank equity, and hence its required return. If the risk of bank equity falls as the equity share of bank capital structure increases, then the required return on bank equity also falls, supporting the applicability of the Modigliani-Miller theorem. Kashyap, Stein, and Hanson (2010) find this predicted relationship for a panel of publicly traded U.S. banks for 1976-2008. Clark, Jones, and Malmquist (2015) suggest that the relevance of the Modigliani-Miller theorem for U.S. banks is more nuanced, depending on bank size and time. They find it holds most strongly for the largest banks, mainly because their debt is closer to riskless. Since the 2007-09 Financial Crisis, Clark, Jones, and Malmquist (2015) find that the Modigliani-Miller theorem holds for banks with assets of at least $100 billion.

Other authors have studied European banks and find evidence of at least partial Modigliani-Miller offsets. We summarize them in Table 5. The body of empirical evidence strongly supports a partial Modigliani-Miller offset. We therefore assume a 50 percent Modigliani-Miller offset to rates of return throughout our estimates.

4.2 Increased Loan Rates

We estimate the changes in banks' weighted average cost of capital resulting from changes in the amount of bank equity. We then assume that these changes in the cost of capital are offset by changes in banks' return on lending. The cost of increasing equity capital ratios by 1 percentage point equals the amount of capital raised, multiplied by the greater cost of equity capital—the return premium on equity over debt capital—adjusted for the Modigliani-Miller offset. It also includes the value of the tax shield provided by debt. The following equation represents this cost as a percentage of total assets:

$$0.01 * \frac{RWA}{Assets} * \left[(1 - MM) * \left(R^E - R^D \right) + R^D * t \right],$$

where R^E is the return on equity, R^D is the return on debt, MM is the degree of the Modigliani-Miller offset, and t is the corporate tax rate.

For the average risk weight, RWA/assets, we use the asset-weighted average for large U.S. banks as of the end of 2015, which is 66 percent.[30] Substituting, this becomes

$$0.01 * 66\% * \left[(1 - MM) * \left(R^E - R^D \right) + R^D * t \right].$$

To estimate the return on bank debt, R^D, we look at the past 15 years. We consider annual data on all U.S. top-tier bank holding companies with assets of at least $50 billion in 2015 dollars. During the past 15 years, median interest costs were 1.61 percent on average. However, we consider recent, very-low interest rates to be a historic abnormality, so we use the pre-crisis (2001-06) period, for which median interest costs averaged 2.78 percent.[31]

To estimate the return on bank equity, R^E, we apply the capital asset pricing model. We estimate the risk, or market betas, of bank equity using 15 years of data. We calibrate the model to an average risk-free rate of 2.78 percent and average excess market returns of 5.24 percent. We estimate a current daily market beta of 1.33 for domestic U.S. top-tier holding companies, based on data from 2010-15, implying an expected return for bank equity of 9.75 percent.[32] To complete the estimate, we use a corporate tax rate, t, of 43.7 percent.[33]

We assume that half the Modigliani-Miller offset occurs:

$$0.01 * 66\% * \left[0.5 * (9.75\% - 2.78\%) + 2.78\% * 43.7\% \right].$$

We thus project an increase in asset return rates of 3.1 basis points per 1-percentage-point increase in capital ratios, of which 0.8 basis points is attributed to a reduced debt shield.

We assume this increase in cost is passed on entirely in the form of increased rates on loans. As loans currently comprise about 40 percent of bank assets, we divide the increase in the cost of capital by 40 percent for our estimate of the increase in loan costs, arriving at 0.00031/0.4 or 7.8 basis points.

There are also reasons to think that banks may not be able to pass along 100 percent of any increased capital cost to borrowers. Competition with non-bank lenders and the public debt markets may make it difficult for banks to pass along increases in the cost of capital. Even without such competition, banks might increase revenue by reducing the cost of operations, decreasing interest expense, or finding revenue increases from business lines other than lending.[34] We consequently assume that, alternatively, banks can pass only half of their increased cost on to borrowers, which implies an increase in loan costs of 3.9 basis points per percentage-point increase in equity capital.[35]

4.3 Translating Loan Rates into GDP

We use the FRB/US model to translate increases in lending rates into a long-term effect on the level of GDP. In this model, lending rates have no effect on long-term growth rates. FRB/US is a large-scale model of the U.S. economy featuring optimizing behavior by households and firms, as well as detailed descriptions of monetary policy and the fiscal sector. FRB/US has a neoclassical core that combines a production function with endogenous and exogenous supplies of production factors and key aspects of household preferences such as impatience. To account for cyclical fluctuations, the model features rigidities that apply to many decisions made by households and firms. The model does not explicitly include corporate loan spreads. Therefore, we use spreads on corporate bonds to model a shock to corporate lending rates and use spreads on home mortgages and auto loans to model a shock to consumer lending rates.[36]

The model predicts that the effects on the equilibrium level of GDP from an increase in lending rates are approximately linear. A 1-basis-point increase in lending rates implies a decrease in the level of GDP of approximately 1.07 basis points.[37] Combining this with our estimated impact of bank capital on lending rates, we estimate that the effect of a 1-percentage-point increase in capital ratios is a reduction in the level of long-term GDP of 8.3 basis points. If only half the cost is passed on in the form of higher rates on loans, the effect on long-term GDP is 4.2 basis points.

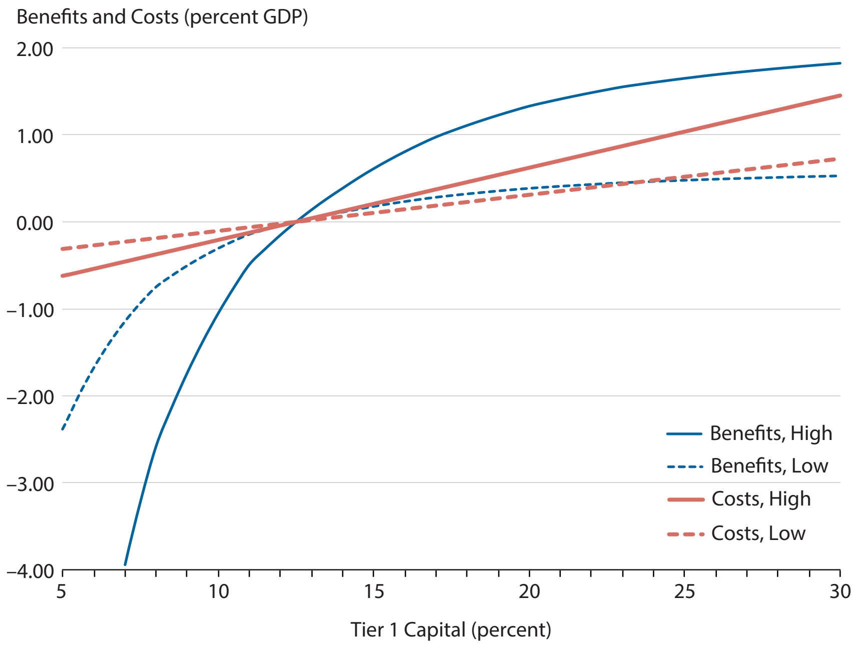
5 ASSESSMENT OF NET BENEFITS

Our framework for assessing the net benefits of bank capital requires that we estimate (i) the long-term effect of increased bank capital on the probability of a financial crisis in the United States, (ii) the net present cost of U.S. financial crises, (iii) the impact of capital on banks' lending spreads, and (iv) the effect of consequent higher lending rates on GDP. Because of uncertainty around the choice of modeling approach for several of these components, we report a range of estimates and refer the reader to Section 6 for a discussion of costs and benefits left out of our framework.

In Figure 5, we plot the range of economic costs and benefits that we model, relative to the costs and benefits of the current Tier 1 risk-based capital ratio of 12.5 percent for the average U.S. bank. Our low estimate of the benefits assumes that financial crises cause only gradually decaying reductions in GDP and that increased resolvability reduces the probability of crises by 30 percent. Our high estimate of the benefits assumes that financial crises result in permanent reductions in output and that increased resolvability affects only the costs, not the probability, of financial crises. Our low and high estimates of the costs vary only because

Figure 5

High and Low Estimates of the Costs and Benefits of Tier 1 Capital Relative to the Current Average U.S. Tier 1 Capital Ratio (12.5 percent)



of the extent that increases in banks' cost of capital are passed on to borrowers. Both low and high estimates of the costs assume that 50 percent of the Modigliani-Miller effect applies and that lending costs affect only long-term GDP levels, not growth.

Figure 6 plots net benefits based on the cost and benefit curves from Figure 5. Figure 7 plots marginal net benefits or the net benefit of additional capital. Optimal capital levels are found where marginal net benefits equal zero.[38] These points are indicated by large diamonds. As mentioned above, banks generally hold significant capital above the minimum capital requirements, so optimal capital requirements would be lower than these levels.

The net benefit of additional capital beyond current levels is very sensitive to assumptions about both the costs and benefits of financial crises. For example, the assumption of low, instead of high, cost of capital levels increases the optimal level of capital by about 4 or 5 percentage points. The impact of using our high estimate of the benefits of capital is even larger at about 8 percentage points. However, in all cases the economic benefits of moderate increases in capital levels above current levels exceed the economic costs. At current levels of capital, this net benefit amounts to between approximately 0 and 25 basis points of annual GDP per percentage-point increase in additional required capital. Depending on the modeling approach, net benefits are maximized at capital ratios that vary from a low of 13 percent (high costs and low benefits) to 26 percent (low costs and high benefits).

Figure 6

Estimates of the Net Benefits of Tier 1 Capital Relative to the Current Average U.S. Tier 1 Capital Ratio (12.5 percent)

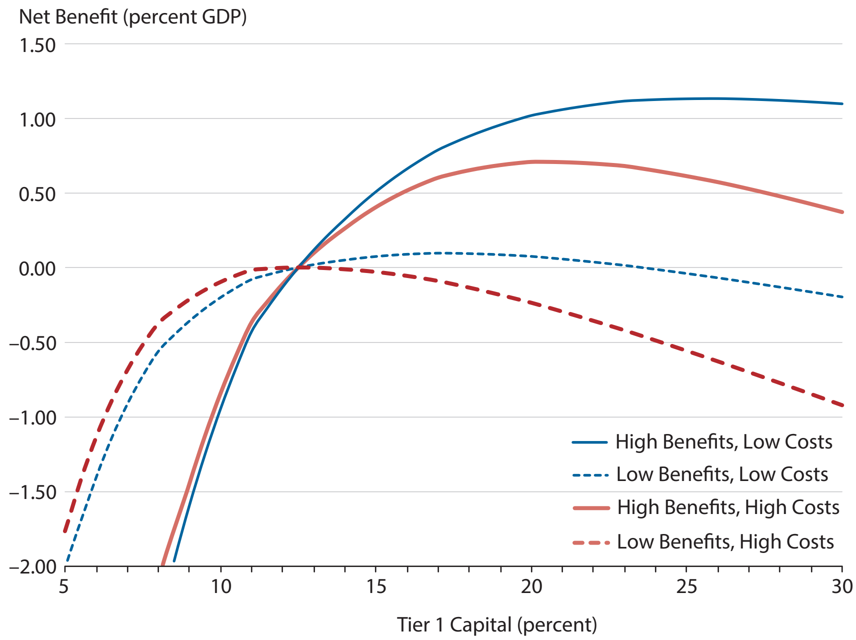


Figure 7

Estimates of the Marginal Net Benefits of Tier 1 Risk-Based Capital

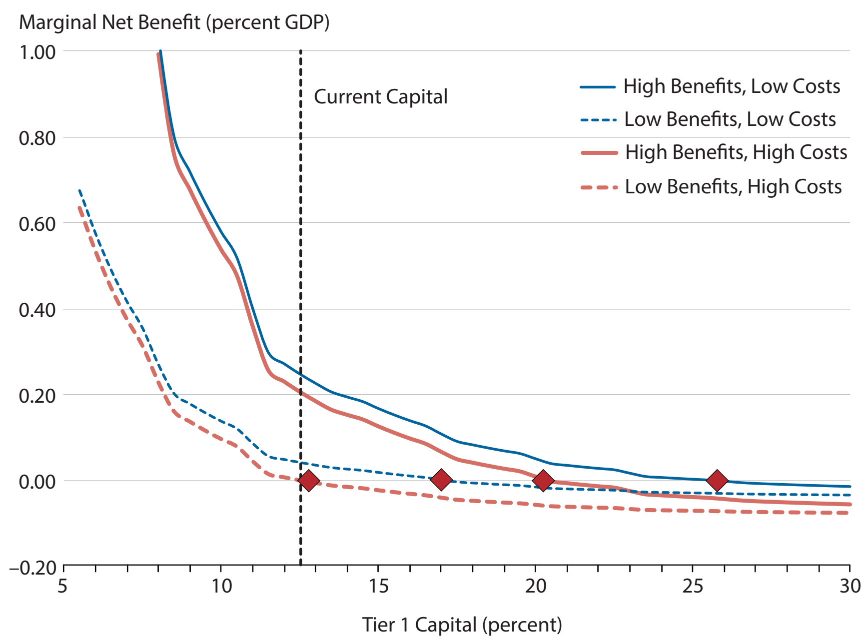


Table 6

Optimal Tier 1 Capital from Other Impact Analyses

Study	Optimal Tier 1 Capital
This study	13-26%
Federal Reserve Bank of Minneapolis (2016)	~22%
Dagher et al. (2016)	15-23%
Brooke et al. (2015)	10-14%
Miles et al. (2013)	~20%
BCBS (2010)	9-15+%

NOTE: The BCBS (2010) optimal Tier 1 capital represents a Basel II TCE/RWA ratio. We calculate the equivalent Basel III Tier I/RWA ratio as 9.3 percent to 15.5+ percent.

Table 6 compares these estimates of optimal capital levels with those estimated in other studies. Only the Brooke et al. (2015) analysis yields optimal capital requirement levels that are significantly lower. This is primarily due to the large size of the reduction to both probability and costs of financial crises that the authors apply for the presence of TLAC regulations.

6 OTHER ECONOMIC COSTS AND BENEFITS OF BANK CAPITAL

There are a number of economic costs and benefits of bank capital not explicitly captured in our results for a variety of reasons, including availability of data, required assumptions, and general feasibility. We will now discuss these costs and benefits and the expected impact they could have on capital.

6.1 Implications for a Risk-Averse World

A better capitalized banking system reduces the volatility of economic cycles, particularly where capital requirements include a countercyclical buffer. Using a DSGE model, BCBS (2010) suggests that output volatility decreases by a bit over 1 percent for each percentage-point increase in Tier 1 capital.[39] Risk-averse people prefer reduced consumption volatility and especially dislike large negative shocks. For example, suppose a country produces one unit of output per year, except during crises when output is instead 0.9 units. A hypothetical consumer of this output with a typical utility function would be willing to forgo up to about 0.4 percent of the expected output in return for reducing the probability of crises from 4 percent to 1 percent.[40] This consumer might be willing to give up significantly more to avoid crises if, for example, crises were associated with greater idiosyncratic volatility (e.g., likelihood of job loss) or the consumer has large inflexible expenses (e.g., housing payments).

In addition, the macrofinance literature suggests that "rare disasters," such as financial crises, may explain much of the risk premia that risk-averse investors require (e.g., Rietz, 1988, and Barro, 2006). A reduction in the probability of a financial crisis could thereby reduce the cost of capital across the economy and increase output.

6.2 Cross-Border Externalities

We have attempted to quantify only those costs and benefits of higher capital for U.S. banks that accrue within the United States. Improved stability of the U.S. financial system may also improve the stability of foreign financial systems. Kaminsky, Reinhart, and Vegh (2003) describe financial crises whose effects were felt across international borders even where the linkages between countries are weak, a phenomenon known as "contagion." They also provide a review of the theoretical literature on why contagion may occur. Most of their analysis focuses on less-developed economies. Kalemli-Ozcan, Papaioannou, and Perri (2013) describe how the 2007-09 Financial Crisis spread across U.S. borders. By decreasing the probability of domestic financial crises, higher U.S. bank capital also decreases the probability of overseas crises.

6.3 Migration of Financial Intermediation

To the extent that higher bank capital increases lending costs for banks, financial intermediation could migrate out of the regulated financial sector. This has an ambiguous effect on optimal capital levels. Benefits are reduced as systemic risks become less effectively controlled, but costs also decline as competition provided by non-regulated intermediation decreases lending costs. Pozsar et al. (2013) provide an overview of the institutional framework of the shadow banking system. Plantin (2014) presents a general equilibrium model that demonstrates the risk of banks shifting assets to less-regulated money market funds. Harris, Opp, and Opp's (2014) analysis of a general equilibrium model with bank and non-bank lenders shows that, depending on the size of the increase in capital, the presence of competing shadow banks can be helpful or harmful because of changes in the nature of projects that receive funding.

6.4 Interaction of Costs of Higher Capital with TLAC and Liquidity Requirements

The greater reliance on long-term unsecured debt due to TLAC likely increases banks' cost of capital. The liquidity rules are likely to also increase banks' reliance on longer-term funding, which would also raise banks' cost of capital. As we use a cost of capital that neglects the impact of future effects of liquidity and TLAC requirements to calculate the cost of increasing equity financing, we may overstate the effect of higher equity capital on banks' cost of capital. Such adjustments would likely decrease the costs of higher capital.

6.5 Transitional Costs

We exclusively analyze the long-term economic costs of bank capital. Kiley and Sim (2010) analyze the cost of transition to higher capital levels. They assume that banks achieve these levels through raising new equity, reducing dividends, and reducing lending. Their estimates, like ours, are characterized by considerable uncertainty and imply a reduction in GDP between 0.5 percent and 3 percent of GDP for a 2-percentage-point increase in capital ratios. The magnitude of their results is partially due to their assumption that when banks reduce lending, there is no other substitute for financing, leading to a credit crunch. This assumption may be unrealistic where there is a bond market and non-bank lenders, and thus their results likely represent an upper bound on transitional costs. Their analysis also shows that the magnitude of the effect on GDP is reduced when banks are given a long period to comply with new requirements. Including such transitional costs would decrease the net benefits of higher capital.

7 CONCLUSION

This article assesses the long-term net benefits of the level of bank capital to provide an estimated range of optimal Tier 1 risk-based capital levels. We provide estimates of optimal capital levels under a variety of assumptions about the costs of financial crises and the extent to which lending rates could rise in response to higher levels of bank capital. When using a high estimate of the costs and low estimate of the benefits of capital, we find an optimal capital ratio around 13 percent, which is close to the current average capital ratio of 12.5 percent for U.S. banks. When using higher estimates of the benefits of capital, we find optimal Tier 1 risk-based capital ratios are approximately 26 percent. As banks generally hold substantial buffers above the minimum requirements, optimal capital requirements would be lower.[41]

This range of estimates is limited by a framework that includes only two channels by which bank capital affects the real economy. Optimal capital levels may be even higher in a framework that accounts for the preferences of risk-averse consumers or the international benefits of financial stability, and they are likely lower when accounting for the cost of a transition to higher capital levels or allowing lending costs to affect long-term growth rates. ■

APPENDIX

Liquidity Requirements Adjustment to Simulation-Based Estimates of the Probability of a Financial Crisis

First, we construct a set of proxies for compliance with the LCR, as these figures are generally not available historically.[42] We identify a cross section of 12 large U.S. banks in 2012 for which both LCR and liquid asset share (from Bankscope) are available. The correlation of these two variables is significant (about 0.52), and a least squares regression of LCR on liquid asset share (without an intercept) yields: LCR = 2.55 ∗ *liquid_asset_share*. For each of these 12 observations, we then estimate the liquid asset share required to set LCR equal to one (the required level) as *required_liquid_asset_share* = (1 – LCR)/2.55. These 12 values are proxies for the liquid asset share equivalent to LCR compliance.

Next, we estimate the historical relationship between the (lagged) liquid asset share of a bank and the magnitude of banks' losses. We run least squares regressions of the log magnitude of losses (relative to RWA) on log liquid assets, alternately controlling for (lagged) log assets and bank type in addition to log assets. The estimated coefficient on log liquid assets varies from –0.095 to –0.107, so we use –0.10 as our estimate of the impact of log liquid assets on log loss magnitude.[43]

Table A1

Example of Liquidity Requirement Adjustments to Net Income Draws

	U.S. Bank A	U.S. Bank B	U.S. Bank C
Randomly drawn net income/RWA	–0.12	–0.11	0.09
Liquid asset share (from same random bank)	20%	40%	10%
Randomly assigned LCR proxy	40%	20%	35%
Net income multiplier	$0.93 = (40\%/20\%)^{-0.1}$	1 (bank already LCR compliant)	1 (net income is positive)
Net income/RWA to use	–0.1116 = 0.93 * –0.12	–0.11	0.09

To integrate these findings within the simulations described above, we draw the liquid asset share in addition to net income (relative to RWA) from the banks in each country-year scenario. We also randomly assign one of the 12 proxies for LCR compliance based on liquid asset share to each bank. Then, if the net income drawn is negative and the liquid asset share drawn is less than the assigned LCR proxy, net income is adjusted toward zero based on the relationship of log liquid assets to log loss magnitude estimated above. Table A1 provides three examples of this procedure. Bank A receives a draw of negative net income from some bank X with relatively low liquid assets. Based on the randomly assigned LCR proxy of 40 percent, bank X's log liquid asset share would need to be higher by ln(40 percent) – ln(20 percent) = 0.693 to have been LCR compliant. The regression of log net losses (relative to RWA) above implies, in turn, that log losses would have been lower by –0.1 ∗ 0.693 = –0.0693, or that losses would have been 93.3 percent as large. In this case, we apply this modifier to the losses before computing the size of any bank-specific or aggregate shortfall. The table shows that no such modifier is applied when the draw comes from either a bank deemed already LCR compliant (as in the case of Bank B) or the net income drawn is positive (as in the case of Bank C).

REFERENCES

Admati, A.; DeMarzo, R.P.; Hellwig, M. and Pfleiderer, P. "Fallacies and Irrelevant Facts in the Debate on Capital Regulation." *Central Banks at a Crossroads: Europe and Beyond*. London: Anthem Press, 2014.

Afonso, G.; Santos, J. and Traina, J. "Do 'Too-Big-to-Fail' Banks Take on More Risk?" Federal Reserve Bank of New York, *Economic Policy Review*, 2014; https://www.newyorkfed.org/medialibrary/media/research/epr/2014/1412afon.pdf.

Barro, R. "Rare Disasters and Asset Markets in the Twentieth Century." *Quarterly Journal of Economics*, 2006, *121*(3), pp. 823-66; https://doi.org/10.1162/qjec.121.3.823.

Basel Committee on Banking Supervision. "An Assessment of the Long-Term Economic Impact of Stronger Capital and Liquidity Requirements." 2010; https://www.bis.org/publ/bcbs173.pdf.

Brooke, M.; Bush, O.; Edwards, R.; Ellis, J.; Francis, B.; Harimohan, R.; Neiss, K. and Siegert, C. "Measuring the Macroeconomic Costs and Benefits of Higher U.K. Bank Capital Requirements." Financial Stability Paper No. 35, Bank of England, 2015; https://www.bankofengland.co.uk/financial-stability-paper/2015/measuring-the-macroeconomic-costs-and-benefits-of-higher-uk-bank-capital-requirements.

Brunnermeier, M. and Pedersen, L. "Market liquidity and Funding Liquidity." *Review of Financial Studies*, 2009, *22*(6), pp. 2201-38; https://doi.org/10.1093/rfs/hhn098.

Cerra, V. and Saxena, S. "Growth Dynamics: The Myth of Economic Recovery." *American Economic Review*, 2008, *98*(1), pp. 439-57; https://doi.org/10.1257/aer.98.1.439.

Cifuentes, R.; Ferrucci, G. and Shin, H.S. "Liquidity Risk and Contagion." Journal of the European Economic Association, 2005, 3(2-3), pp. 556-66; https://doi.org/10.1162/jeea.2005.3.2-3.556.

Clark, B.; Jones, J. and Malmquist, D. "Leverage and the Weighted-Average Cost of Capital for U.S. Banks." Working Paper, SSRN, 2015; https://doi.org/10.2139/ssrn.2491278.

Dagher, J.; Dell'Ariccia, G.; Laeven, L.; et al. "Benefits and Costs of Bank Capital." Staff Discussion Note, International Monetary Fund, 2016; https://www.imf.org/external/pubs/ft/sdn/2016/sdn1604.pdf.

Demirgüç-Kunt, A.; Detragiache, E. and Gupta, P. "Inside the Crisis: An Empirical Analysis of Banking Systems in Distress." *Journal of International Money and Finance*, 2006, *25*, pp. 702-18; https://doi.org/10.1016/j.jimonfin.2006.04.004.

Elliott, D. "Higher Bank Capital Requirements Would Come at a Price." *Brookings Report*, 2013; https://www.brookings.edu/research/higher-bank-capital-requirements-would-come-at-a-price/.

European Central Bank. "Common Equity Capital, Banks' Riskiness and Required Return on Equity." *Financial Stability Review*, December 2011, pp. 125-31.

Federal Reserve Bank of Minneapolis. "The Minneapolis Plan to End Too Big to Fail." 2016; https://www.minneapolisfed.org/~/media/files/publications/studies/endingtbtf/the-minneapolis-plan/the-minneapolis-plan-to-end-too-big-to-fail-2016.pdf.

Financial Stability Board. "Assessing the Economic Costs and Benefits of TLAC Implementation." Bank for International Settlements, 2015; https://www.bis.org/publ/othp24.pdf.

Furceri, D. and Mourougane, A. "The Effect of Financial Crises on Potential Output: New Empirical Evidence from OECD Countries." *Journal of Macroeconomics*, 2012, *34*(3), pp. 822-32; https://doi.org/10.1016/j.jmacro.2012.05.010.

Harris, M.; Opp, C. and Opp, M. "Higher Capital Requirements, Safer Banks? Macroprudential Regulation in a Competitive Financial System." Working Paper, SSRN, 2014; http://dx.doi.org/10.2139/ssrn.2181436; https://doi.org/10.2139/ssrn.2181436.

Hoggarth, G.; Reis, R. and Saporta, V. "Costs of Banking System Instability: Some Empirical Evidence." *Journal of Banking & Finance*, 2002, *26*, pp. 825-55; https://doi.org/10.1016/S0378-4266(01)00268-0.

Homar, T. and van Wijnbergen, S.J.G. "Bank Recapitalization and Economic Recovery after Financial Crises." *Journal of Financial Intermediation*, 2016, *32*, pp. 16-28; https://doi.org/10.1016/j.jfi.2016.11.004.

Hutchison, M. and Noy, L. "How Bad Are Twins? Output Costs of Currency and Banking Crises." *Journal of Money, Credit, and Banking*, 2005, *37*(4), pp. 725-51; https://doi.org/10.1353/mcb.2005.0043.

Junge, G. and Kugler, P. "Quantifying the Impact of Higher Capital Requirements on the Swiss Economy." *Swiss Journal of Economics and Statistics*, 2013, *149*(3), pp. 313-356; https://doi.org/10.1007/BF03399394.

Kalemli-Ozcan, S.; Papaioannou, E. and Perri, F. "Global Banks and Crisis Transmission." *Journal of International Economics*, 2013, *89*(2), pp. 495-510; https://doi.org/10.1016/j.jinteco.2012.07.001.

Kaminsky, Graciela L.; Reinhart, C.M. and Vegh, C.A. "The Unholy Trinity of Financial Contagion." *Journal of Economic Perspectives*, 2003, *17*(4), pp. 51-74; https://doi.org/10.1257/089533003772034899.

Kashyap, A.K.; Stein, J.C. and Hanson, S. "An Analysis of the Impact of 'Substantially Heightened' Capital Requirements on Large Financial Institutions." Booth School of Business, University of Chicago, Mimeo, 2010; https://faculty.chicagobooth.edu/anil.kashyap/research/papers/an_analysis_of_the_impact_of_substantially_heightened-Capital-Requirements-on-Financial-Institutions.pdf.

Kiley, M. and Sim, J. "Technical Background Report: An Assessment of the Macroeconomic Transition Costs Associated with Higher Levels of Capital at Financial Institutions in the United States." Unpublished Manuscript, Board of Governors of the Federal Reserve System, 2010.

Laeven L. and Valencia, F. "Systemic Banking Crises Database: An Update." Working Paper No. 12/163, International Monetary Fund, 2012; https://doi.org/10.5089/9781475505054.001.

Malherbe, F. "Self-Fulfilling Liquidity Dry-Ups." *Journal of Finance*, 2014, *69*(2), pp. 947-70; https://doi.org/10.1111/jofi.12063.

Marques, L.B.; Correa, R. and Sapriza, H. "International Evidence on Government Support and Risk Taking in the Banking Sector." Working Paper No. 13/94, International Monetary Fund, 2013; https://doi.org/10.5089/9781484309346.001.

Miles, D.; Yang, J. and Marcheggiano, G. "Optimal Bank Capital." *Economic Journal*, 2013, *123*(657), pp. 1-37; https://doi.org/10.1111/j.1468-0297.2012.02521.x.

Miller, M.H. "Do the M&M Propositions Apply to Banks?" *Journal of Banking & Finance*, 1995, *19*, pp. 483-89; https://doi.org/10.1016/0378-4266(94)00134-O.

Modigliani, F. and Miller, M.H. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review*, 1958, *48*(3), pp. 261-97.

Philippon, T. and Reshef, A. "Wages and Human Capital in the U.S. Finance Industry: 1909–2006." *Quarterly Journal of Economics*, 2012, *127*(4); https://doi.org/10.1093/qje/qjs030.

Plantin, G. "Shadow Banking and Bank Capital Regulation." *Review of Financial Studies*, 2014, *28*(1), pp. 146-75; https://doi.org/10.1093/rfs/hhu055.

Pozsar, Z.; Adrian, T.; Ashcraft, A. and Boesky, H. "Shadow Banking," Federal Reserve Bank of New York, *Economic Policy Review*, 2013, *458*, pp. 3-9.

Rietz, T. "The Equity Risk Premium: A Solution." *Journal of Monetary Economics*, 1988, *22*, pp. 91-115; https://doi.org/10.1016/0304-3932(88)90172-9.

Romer, C. and Romer, D. "New Evidence on the Impact of Financial Crises in Advanced Countries." Working Paper No. 21021, National Bureau of Economic Research, 2015; https://doi.org/10.3386/w21021.

Shleifer, A. and Vishny, R. "Fire Sales in Finance and Macroeconomics." *Journal of Economic Perspectives*, 2011, *25*(1), pp. 29-48; https://doi.org/10.1257/jep.25.1.29.

Toader, O. "Estimating the Impact of Higher Capital Requirements on the Cost of Equity: an Empirical Study of European Banks." *International Economics and Economic Policy*, 2015, *12*(3), pp. 411-436; https://doi.org/10.1007/s10368-014-0303-x.

NOTES

1 Throughout the article, unless otherwise noted, we refer to capital as Tier 1 risk-based capital.

2 The low estimate of benefits also incorporates an assumption that total loss-absorbing capacity (TLAC) regulations independently reduce the probability of a financial crisis by 30 percent. Our high estimate of benefits assumes that TLAC regulations reduce only the costs of financial crises, not their probability.

3 Brooke et al. (2015) use other methods to adjust for the presence of such countries in the sample.

4 See www.federalreserve.gov/newsevents/press/bcreg/20151030a.htm for more information on long-term debt and TLAC requirements. See www.federalreserve.gov/bankinforeg/resolution-plans.htm for more information on living wills. See www.fdic.gov/regulations/laws/federal/2011/11finaljuly15.pdf for more information on the orderly liquidation authority.

5 See Shleifer and Vishny (2011) for a review of the literature on asset fire sales. See Brunnermeier and Pedersen (2009) for a discussion of the destabilizing link between market and funding liquidity.

6 See www.federalreserve.gov/newsevents/press/bcreg/20140903a.htm for more information on the LCR.

7 Romer and Romer (2015) include an extensive comparison of their definition with that in Laeven and Valencia (2012). They are almost equivalent in identifying if crises occur but have a few differences with regard to crises start dates.

8 We apply a 50 percent weight to our bottom-up approach and a 25 percent weight to each of the two models in our top-down approach.

9 The sample includes Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

10 We exclude 30 country-years where coverage appears particularly incomplete and/or includes banks representing less than $10 billion in aggregate total assets.

11 In contrast to our approach, the Brooke et al. (2015) impact analysis pools country-years into just the following two scenarios: one where credit/GDP and equity volatility are high and one where they are not.

12 To account for the average difference between Basel III risk weights and Basel I and II risk weights, we increase RWA for all years prior to 2011 by the increase in U.S. average risk weights for 2011-15 (about 11 percent). For 2012-14, we increase RWA by the increase in U.S. average risk weights between that year and 2015.

13 We exclude small banks from the distribution of net income we draw from, as this distribution may be different for small banks. We define the cutoff between small and large banks so that 90 percent of each country's financial system assets in each year are within "large" banks. Across countries, we have an average of about 915 large banks per year.

14 When using a different threshold, as is done by Brooke et al. (2015), additional adjustments are required to match the historical frequency of financial crises.

15 The Financial Stability Board analysis uses 30 percent as a baseline estimate, but also provides a range of estimates for the impact of TLAC ranging from under 10 percent to over 40 percent.

16 We omit country-years where there is an ongoing crisis because we are interested in predicting the probability of a financial crisis starting. This probability is implicitly conditional on a financial crisis not yet having occurred.

17 We used the Chicago Board Options Exchange VXO® volatility index prices until the VIX® index became available in 1990. Historical VIX® data is available at www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data.

[18] We follow Financial Stability Board (2015), which relies on work by Afonso, Santos, and Traina (2014) and Marques, Correa, and Sapriza (2013) to estimate that TLAC reduces the probability of a financial crisis by 30 percent.

[19] Note that Tier 1 capital is not statistically significant in specification (1). We retain it because it incorporates much more data. If we were to use just specification (2), it would slightly increase the lower bound of our estimate of the optimal capital range to 14 percent and lower the upper bound of the range to 22 percent.

[20] Defining the end of a crisis's effects as when the economy exits a recession would give the same results. If we define the temporary effects as occurring when the economy reaches its pre-crisis growth path, the crisis would end when the two lines intersect.

[21] There are other studies measuring the overall short- and medium-term costs of a financial crisis. Examples include Hoggarth, Reis, and Saporta (2002), Hutchison and Noy (2005), and Demirgüç-Kunt, Detragiache, and Gupta (2006). We use results from Romer and Romer (2015) because they focus on advanced countries and are thus most relevant to the United States. Romer and Romer (2015) are also unique in that they provide estimates of the semiannual dynamics of the effect on GDP, rather than a single aggregated number. A further advantage is that they incorporate results from the 2007-09 Financial Crisis. The dynamics allow a more credible analysis of the effects of shortening crises through prompt recapitalization.

[22] Note that our analysis of the probability of a crisis uses the definition in Laeven and Valencia (2012). See section I.D of Romer and Romer (2015) for a comparison of financial crises identified by their method and an updated version of the data used by Laeven and Valencia (2012). The two are almost identical. Furceri and Mourougane (2012) use the same definition as Laeven and Valencia (2012).

[23] These numbers are calibrated to represent a "moderate" crisis. The last time they designated the United States as having been in such a crisis state was the second half of 2009.

[24] We use all monthly data for 1962-2016 and calculate inflation from the CPI for all urban consumers. Data were obtained from the Federal Reserve Bank of St. Louis Economic Research website at https://research.stlouisfed.org/.

[25] While this may seem small, given that U.S. GDP in 2015 was $18.2 trillion, this translates into a reduction in cost of $546 to $728 billion.

[26] To remove minor non-monotonicities in the plot of marginal benefits, we approximate the probability of a financial crisis generated by the bottom-up approach by a piecewise log function with segments that are 2 percentage points wide. This adjustment has little effect on the level of curves or the estimated optimal level of capital.

[27] The marginal benefit of capital is "lumpy" due to lumpiness in distribution of the net income implicit in the bottom-up approach. For example, the marginal benefit is slightly higher at a 21 percent capital ratio than at a 20 percent capital ratio, as there are more scenarios where a roughly 21 percent capital ratio is just barely required to avoid crisis. Miles, Yang, and Marcheggiano (2013) also look at a model where the presence of rare, but very large, shocks to capital means that the marginal benefit of capital is not decreasing everywhere.

[28] We thank Michael Siemer for using FRB/US to calculate the effect of an increase in lending costs on the level of GDP.

[29] For more information on the FRB/US model, see the Board of Governors of the Federal Reserve System's website at www.federalreserve.gov/econresdata/notes/feds-notes/2014/a-tool-for-macroeconomic-policy-analysis.html.

[30] Many recent changes have affected banks' asset composition and RWA calculations in recent years, including the implementation of Basel III, the Volcker rule, and liquidity and TLAC regulations. Therefore we use data on both RWA/assets, assets/loans, and interest expense from December 2015, the most recent full year for which data are available. Neither the interest rate environment nor regulations governing RWA calculations have changed since then in a way that would significantly affect our conclusions.

[31] Kayshap, Stein, and Hanson (2010) assume that additional equity will displace either long-term or short-term debt. TLAC and liquidity requirements mean that banks are now less free to adjust their liability composition. Thus, we use the average liability cost.

[32] We use data starting in 2010 for calculating beta because the large number of mergers and acquisitions following a crisis would muddy the analysis prior to that date.

[33] We follow Clark, Jones, and Malmquist (2015) in assuming a Delaware residence. Delaware has a tax rate of 8.37 percent, and the federal corporate tax rate is 35 percent. Only South Dakota and Wyoming have no corporate income or gross receipts tax. State corporate income taxes range from 4 percent to 12 percent for 2016. See http://taxfoundation.org/article/state-corporate-income-tax-rates-and-brackets-2016.

[34] Philippon and Reshef (2012) show that wages in the financial sector from 1990 onward earned a premium over other sectors. This premium grew over time and cannot be fully explained by risk or technological change. This research suggests that cost reductions are feasible.

[35] If only half the increased cost is passed on in the form of higher loan rates, banks might respond to increased costs in other ways, such as focusing more on higher-yield, and thus risky, lending. Banks might also raise fees for other services in a manner that could create distortions. To our knowledge, there is no adequate empirical basis for modeling such effects.

[36] Most of the effect on GDP is from an increase in corporate lending rates. To the extent that rate increases are limited to other loans, we may be overstating the effects.

[37] This is slightly different than the results calculated by the Federal Reserve Bank of Minneapolis (2016). The Minneapolis Fed (2016) calculates changes from a long-term baseline, while we use the present-day scenario.

[38] The irregularities in the marginal benefit curve occur because the density of loss severities in the historical data is not steadily decreasing as loss severity increases.

[39] The dynamic stochastic general equilibrium models used in BCBS (2010) also suggest that the use of a countercyclical buffer of plus-or-minus 2 percent reduces output volatility by around 15 percent.

[40] This calculation assumes the consumer has constant relative risk aversion utility with a coefficient of relative risk aversion of five.

[41] See the Federal Reserve Board working paper version of this analysis for an extensive comparison of the differences between our study's conclusions and that of similar work done by the BCBS (2010), Brooke et al. (2015), and the Minneapolis Fed (2016). It is available at https://www.federalreserve.gov/econres/feds/files/2017034pap.pdf.

[42] We focus on LCR compliance as we lack decent proxies for the NSFR in our Bankscope data.

[43] Depending on the specification, this coefficient is statistically significant at the 5 or 10 percent level with standard errors clustered at the country level. We treat this result with caution as we have not controlled for liquidity being an endogenous decision by banks. There is to our knowledge very little empirical work on the relationship between balance sheet liquidity and loss severity. Cifuentes, Ferrucci, and Shin (2005) use simulations and find that liquidity ratio requirements can limit "contagion" of bank failures because of illiquid asset fire sales. Malherbe (2014) finds in a theoretical model that liquidity ratio requirements can increase the probability of a financial crisis.

Occupational Mobility and Lifetime Earnings

Yongseok Shin and C.Y. Kelvin Yuen

People's occupations have a significant amount of information about their wages. However, because people—especially young workers—go through multiple occupations and employment statuses during their working lives, we find that their occupations at a young age do not predict their lifetime earnings well. When educational attainment and gender are considered, we find that across education-gender groups the differences in lifetime earnings are even larger than the differences in average occupational wages: Workers in high-wage education-gender groups (men with college degrees, for example) work more (at the extensive margin) and are more likely to have higher-paying occupations. (JEL E24, J24, J31, J62)

Federal Reserve Bank of St. Louis *Review*, Third Quarter 2019, *101*(3), pp. 231-44.
https://doi.org/10.20955/r.101.231-44

1 INTRODUCTION

Occupations signify people's economic standing: One can safely assume that doctors and lawyers are better off than cooks and janitors. In this article, we ask how much information on people's labor market outcomes is encapsulated in their occupations. We first consider how much of overall wage inequality can be accounted for by the differences in average wages of occupations (between-occupation inequality). We then examine how informative a person's occupation at a young age (i.e., 25 years of age) is about that person's lifetime earnings. Because the calculation of lifetime earnings takes into account the fact that people go through multiple occupations and employment statuses, we will look for patterns of mobility across occupations. Finally, we divide the data by education and gender groups to see how average wages and lifetime earnings vary across occupations within and across the groups.

Our first finding is that occupations provide a significant amount of information on workers' wages in the cross section—average wages of occupations vary significantly: Even when we divide occupations into only 22 occupation categories, the highest average wage

Yongseok Shin is a professor of economics at Washington University in St. Louis and a research fellow at the Federal Reserve Bank of St. Louis. C.Y. Kelvin Yuen is a Ph.D. candidate in economics at Washington University in St. Louis.

(legal) is nearly three times the lowest average wage (food preparation and serving). However, because workers go through multiple occupations and employment statuses during their working lives, especially when young, we find that their occupations at a young age do not predict their lifetime earnings well.

Our second finding is about the earnings gaps between education-gender groups. Even within the same occupation category, some groups (for example, men with college degrees) earn more than others. When we compute between-group differences in lifetime earnings, conditioning on occupation category at age 25, the fact that people go through multiple occupation categories and employment statuses *amplifies* the between-group wage differences within an occupation category. This amplification occurs because workers in high-wage education-gender groups work more (at the extensive margin) and are more likely to remain in higher-paying occupation categories.

In recent years, many macro-labor models of tasks and occupations have equated workers' occupations with their skill levels (see the review of Acemoglu and Autor, 2011). Once we factor in occupational mobility, however, our analysis suggests that the link between occupation and skill may be weaker than has often been assumed.

Our finding on inequality in wage and lifetime earnings between education-gender groups relates to studies showing that occupation-specific human capital and tenure explains wage growth and inequality (see the review of Sanders and Taber, 2012).

2 DATA

To compute wages and occupational mobility, we use the Current Population Survey (CPS). To be more specific, we use *monthly* CPS Outgoing Rotation Groups (ORG) data, which have a rotating panel structure: Each household participates in four consecutive monthly interviews, is dropped from the sample for the following eight months, and finally participates in another four consecutive monthly interviews. This structure allows us to match individuals from one month to another using their race, sex, and age (Nekarda, 2009). We restrict our sample to those 22 to 55 years of age and look at the 15-year period 2003-17.[1] In total, we have 7,696,838 observations in our ORG sample. In the computations, individuals are weighted by their composited final weight for the CPS.

The CPS records the occupation of an employed individual and the most-recent occupation of an unemployed individual. For those out of the labor force, there is no occupation information. To have enough observations for each occupation, we divide the 571 occupations in our sample into 22 occupation categories using the Detailed Occupation Recodes in the CPS, excluding the Armed Forces.

To compute the average wage of each of the 22 occupation categories, we use the microdata from the CPS Annual Social and Economic Supplement, which is conducted every March and asks more-detailed questions about each individual's labor income. We again restrict our sample to those who are 22 to 55 years of age and the period 2003-17. We have 1,026,605 observations. Each average hourly wage is calculated as the total annual wage and salary earnings divided by the product of the number of weeks worked and the usual hours worked per week for the given occupation category.[2] All wages are deflated to 2003 dollars using the consumer price index from the Bureau of Labor Statistics. For each year, we trim wages at the 0.1 and 99.9 percentiles to exclude outliers.

3 BETWEEN-OCCUPATION INEQUALITY IN WAGES AND LIFETIME EARNINGS

To give an overview of our analysis, in Figure 1 we start with three Lorenz curves. A Lorenz curve is a visual representation of inequality. On the horizontal axis, we order people in the economy by their wages, from 0 to 100 percentiles. On the vertical axis, we plot the fraction of total wages that belongs to the people to the left of a given percentile. If everyone has the same wage, then the Lorenz curve will be the 45-degree line: *x* percent of the population will account for exactly *x* percent of total wages—perfect equality. The farther away from the 45-degree line a Lorenz curve is, the more unequal the wage distribution is.

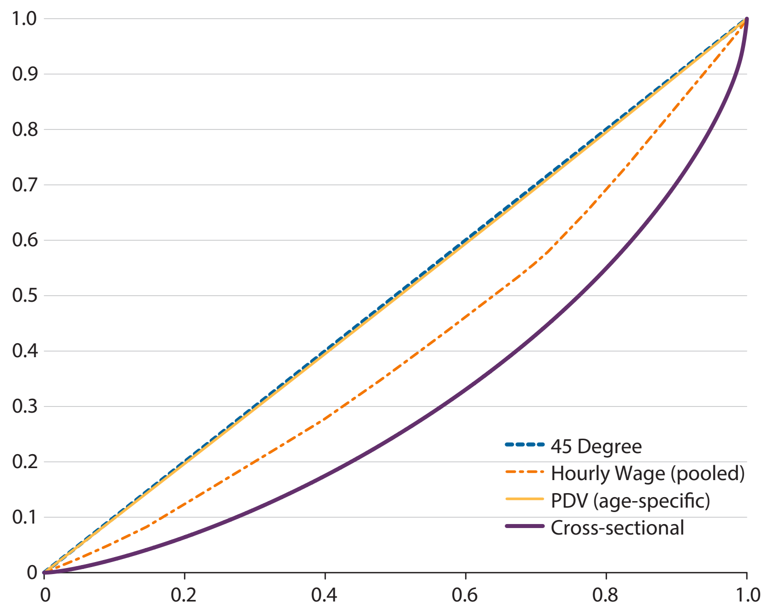
We first construct a Lorenz curve using the wage distribution of the entire sample, which is the purple curve labeled "Cross-sectional" in Figure 1. This curve is far from the 45-degree line, showing substantial wage inequality in the economy. One measure of the degree of inequality is the Gini coefficient, which is the area between a Lorenz curve and the 45-degree line divided by the area of the triangle of which the hypotenuse is the 45-degree line. Thus, a Gini coefficient ranges from 0 (no inequality) to 1 (the extreme case where all wages of the entire economy go to one person). The Gini coefficient of the wage distribution is 0.38.[3]

To show how much of this inequality is attributable to wage differences across occupation categories, we make the (very strong) assumption that everyone in the same occupation category has the same wage. Now all remaining wage inequality is solely due to wage differences across occupation categories. The orange curve labeled "Hourly Wage (pooled)" is the resulting Lorenz curve. It is closer to the 45-degree line, as it should be, because we assumed no wage inequality within any given occupation category. What is remarkable is that this curve sits almost halfway between the cross-sectional Lorenz curve and the 45-degree line. In other words, between-occupation-category wage differences account for as much as half of the Gini coefficient of the actual wage distribution (0.19 of 0.38)—a very large fraction considering we have only 22 occupation categories and do not control for conventional explanatory variables of wages, such as education, age, gender, and race. This exercise confirms that occupations indeed encapsulate a significant amount of information on people's economic standing.

However, people's wages at a given point in time are a very partial measure of their lifetime labor income, which is what matters the most for their lifetime economic well-being. We construct the present discounted value of lifetime earnings at 25 years of age (PDV at 25) for each occupation category, using an annual interest rate of 2 percent. We now assume that everyone of the same age and in the same occupation category earns the same wage.[4] Instead of assuming that people stay in the same occupation category throughout their working lives, we calculate from the data the probability of changing occupation categories or even becoming unemployed. (See the appendix for the details.) The yellow curve labeled "PDV (age specific)" is the resulting Lorenz curve and is nearly indistinguishable from the 45-degree line. Thus, we find that PDV at 25 varies very little across occupation categories.[5]

Figure 1

Lorenz Curves for Wages and PDV at 25



NOTE: This figure plots Lorenz curves for the hourly-wage distribution ("Cross-sectional"), the wage distribution assuming everyone with the same occupation earns the same wage ("Hourly Wage (pooled)"), and the distribution of the PDV at 25 by occupation group ("PDV (age-specific)"). The horizontal axis is the cumulative percentage of workers in the economy, ordered by their wages or PDV at 25, and the vertical axis is the cumulative percentage of their wages. See the data section for details about the data processing and the appendix for the derivation of PDV at 25.

SOURCE: Authors' calculations using the CPS Annual Social and Economic Supplement microdata for the period 2003-17.

There are two reasons why PDV at 25 is much more equal across occupation categories than wages are. First, differences in wages across occupations "fan out" over the life cycle: That is, the difference between high-paying occupations and low-paying occupations is larger for older workers than for young workers. Accounting for age and also discounting future earnings make PDV at 25 more equal across occupation categories. However, what is much more important is occupational mobility. People switch occupations over the course of their working lives and especially frequently when they are young: A 25 year old on average stays with one occupation for a mere eight months (before either changing occupations or dropping out of labor force); and even for a 45 year old, the expected duration is only slightly more than one year. As a result, people's occupation categories at age 25 have very limited information on their lifetime earnings, as illustrated by the Lorenz curve that collapses onto the 45-degree line. For the very same reasons, people's occupations at later ages do have more information on their remaining lifetime earnings; but even for 45 year olds, the PDV at 45 Lorenz curve is closer to the 45-degree line than to the hourly-wage Lorenz curve.

The analysis above looked at wages and labor supply at only the extensive margin (employment status). We also conducted a similar exercise using actual earnings (wage times hours

Table 1

Comparison of Occupations

Occupation category	PDV at 25 (norm.)	Wage rate (2003 $)	EU Separation (%)	EU+EN Separation (%)	Job finding (%)	UE exp. wage (2003 $)	Job switching (%)	Switching exp. wage (2003 $)	Gross flows (%)
Architecture and Engineering	1.041	28.50	0.55	1.31	21.44	19.53	1.39	22.23	10.89
Computer and Math	1.036	29.05	0.58	1.50	19.63	19.76	1.65	22.53	11.51
Healthcare Practitioner	1.023	26.87	0.41	1.68	27.97	19.78	1.58	21.48	9.14
Legal	1.022	36.10	0.43	1.62	23.60	20.45	1.70	24.02	9.21
Business and Financial	1.022	24.73	0.66	1.82	21.67	17.48	1.68	19.26	13.78
Life, physical, and Social Science	1.020	24.83	0.68	1.88	23.26	19.33	1.64	20.90	12.84
Management	1.013	28.53	0.58	1.67	21.17	17.04	1.45	19.64	11.68
Arts, Design, and Entertainment	1.009	20.83	1.60	4.06	27.21	17.17	2.43	17.47	17.73
Community and Social Service	1.006	16.47	0.58	2.11	21.83	16.20	1.83	16.97	12.63
Protective Service	1.004	19.26	0.74	2.09	21.05	14.72	1.46	15.71	10.47
Installation, Maintenance, and Repair	1.001	17.33	1.13	2.40	22.43	14.28	1.64	16.37	13.27
Education, Training, and Library	1.000	17.65	0.90	3.11	34.54	16.23	1.72	16.39	11.55
Sales	0.997	18.07	1.08	3.15	20.00	13.88	2.05	14.69	15.33
Office and Administrative Support	0.996	14.36	0.99	2.94	19.04	14.38	1.72	15.14	14.80
Production	0.993	14.53	1.65	3.41	21.65	12.34	1.68	12.67	15.54
Construction and Extraction	0.992	16.54	3.42	5.56	29.58	13.79	2.46	13.98	18.84
Healthcare Support	0.990	11.52	1.21	3.75	22.97	12.85	1.95	13.11	17.69
Transportation and Material Moving	0.989	14.27	1.94	4.07	23.53	12.08	2.20	12.66	17.69
Personal Care and Service	0.985	11.42	1.47	5.30	23.16	12.26	2.29	12.60	19.92
Building and Grounds Cleaning	0.983	10.60	2.24	5.63	22.92	10.66	2.35	10.71	20.87
Food Preparation and Serving	0.983	9.68	1.86	4.86	23.13	11.57	2.73	11.95	18.25
Farming, Fishing, and Forestry	0.981	10.20	3.79	7.19	30.05	10.98	2.79	11.70	25.00
Correlation with PDV	1.00	0.91	–0.70	–0.83	–0.21	0.95	–0.71	0.96	–0.78

SOURCE: Authors' calculations using the CPS Outgoing Rotation Groups and Annual Social and Economic Supplement microdata for the period 2003-17.

worked) data. There is more inequality in earnings in the cross section (a Gini of 0.43, compared with 0.38 for wages), but the analysis yields fairly similar results, suggesting that the intensive margin of the labor supply plays only a minor role in these calculations.

4 HOW DO OCCUPATIONS DIFFER?

The exercises so far show how much information people's occupation categories provide about their wages and lifetime earnings, but are silent about how occupations differ from one another, for example in terms of wages or mobility. We now make such comparisons.

As shown in the first column of Table 1, we have 22 occupation categories. They are sorted by PDV at 25 normalized by their averages, from highest to lowest (column "PDV at

25 (norm.)”). As discussed above, there is very little difference across occupation categories: The difference between the highest (architecture and engineering) and the lowest PDV at 25 (farming, fishing and forestry) is only 6 percent.

In the next column (“Wage rate”), we show the average wage of each occupation, which we used to construct the “Hourly Wage” Lorenz curve in Figure 1. Two facts stand out. First, the average wages of the occupation categories vary greatly, unlike PDV at 25: The highest average wage (legal) is nearly four times as high as the lowest (food preparation and serving). Second, in spite of this stark difference, average wages are very strongly correlated with PDV at 25 (correlation coefficient of 0.91). In other words, ranking occupations either by wages or by PDV at 25—the idea being that people want higher average wages and lifetime earnings—provides very similar rankings.

In the next five columns, we consider the notions of job security and occupational mobility. We first look at job separation rates by occupation category, defined as the rate at which a worker in a given occupation category becomes unemployed *in the next month* (column “EU Separation”; EU stands for employment to unemployment). For the majority of occupations, it is less than 1 percent, but there are outliers. It is nearly 4 percent for the farming, fishing, and forestry occupation category, likely reflecting the seasonal nature of employment in this category. The EU separation rate is also high for the construction and extraction occupation category (3.42 percent), partly because of the massive destruction of such jobs during the Great Recession. The EU separation rates are negatively correlated with PDV at 25 (correlation coefficient of –0.70), implying that occupation categories with high average lifetime earnings (and high average wages) lose fewer workers to unemployment. Since people want higher earnings and not to be unemployed, the rankings of occupation categories by PDV at 25 and by the inverse of the EU separation rates broadly coincide.

We also consider a broader notion of job separation, by calculating the rate at which a worker in a given occupation category becomes unemployed or leaves the labor force (nonparticipation) in the next month (column “EU+EN Separation”; EN stands for employment to nonparticipation). While the majority of EU transitions are involuntary, it is difficult to figure out which EN transitions are voluntary (for example, retirement) or involuntary. The broadly defined EU+EN separation rates vary across occupation categories more than the more conventionally defined (EU) job separation rates do, and they are even more negatively correlated with PDV at 25 (correlation coefficient of –0.83).[6]

The next column (“Job finding”) shows the job finding rates of unemployed workers by their *previous* occupation category. The job finding rate is the rate at which unemployed workers become employed in any occupation category in the next month. The job finding rates by previous occupation category vary quite a bit—for example, those who become unemployed from the education, training, and library occupation category have about a 35 percent chance of finding another job in a month, while those in the office and administrative support occupation category have about a 19 percent chance. However, the job finding rates do not show a clear pattern, as they are not strongly correlated with PDV at 25 (correlation coefficient of –0.21). What is interesting is that the correlation is negative: If a worker in a higher-paying occupation category becomes unemployed, that worker is expected to stay unemployed a little longer than those in lower-paying occupation categories.

Getting out of unemployment is a good thing, but it also matters what kind of job one lands in out of unemployment. We compute the wage an unemployed worker could expect when finding a new job, by the worker's *previous* occupation category ("UE exp. wage" column). Each expected wage is computed by first multiplying the chance of getting employed in a particular occupation out of unemployment by the average wage of that occupation (considering the worker's age), and then adding these products across all 22 occupation categories. Expected wages by previous occupation category are very strongly correlated with PDV at 25 (correlation coefficient of 0.95) and with average wages across occupation categories. The reason is fairly simple: Unemployed workers tend to return to their previous occupation category: 48 percent of those moving out of unemployment return to their previous occupation category, even though there are 22 possible categories.

We now consider occupational mobility that does not involve a spell of unemployment or nonparticipation. The first variable we construct is a job-switching rate (column "Job switching"). This is the rate at which employed workers in a given occupation category switch employers in the next month (known as job-to-job transition), regardless of whether they switch occupation categories. The job-switching rate is strongly negatively correlated with PDV at 25 (correlation coefficient of –0.71) and with average wages across occupation categories: Workers in high-paying occupation categories are less likely to switch jobs.

We also look at the expected wages of those who make job-to-job transitions (column "Switching exp. wage"). As for the expected wages out of unemployment above (UE exp. wage), this is computed by first multiplying the probability of getting employed in a particular occupation category by the average wage of the new occupation category (considering the worker's age), and then adding these products across all 22 occupation categories. The expected wages for job switchers are closely correlated with PDV at 25 (correlation coefficient of 0.96) and with average wages across occupation categories, again primarily because even job switchers tend to remain in the same occupation category: 54 percent of job switchers do not change occupation categories.

Finally, we construct another measure of occupational mobility that captures the flow of workers into and out of each occupation category (column "Gross flows"). Specifically, the monthly gross flow rate of each occupation category o is the sum of the outflow (the number of workers who are employed in occupation category o in month m but not in month $m + 1$) and the inflow (the number of workers who are employed in occupation category o in month $m + 1$ but not in m) divided by the average number of workers employed in occupation category o in months m and $m + 1$. A high gross flow rate implies a higher rate of turnover, or "churning," for the occupation category.[7] We note that, first, there are large variations in the gross flow rates across occupation categories: from lows of about 9 percent each for the health-care practitioner and legal occupation categories to a high of 25 percent for the farming, fishing, and forestry occupation category. By comparison, the net flow rates across occupation categories have a standard deviation of 4.2 percent over the sample period: In other words, even the smallest *gross* flow rate among occupation categories is more than twice the standard deviation of *net* changes in occupational employment. Second, the gross flow rates are strongly negatively correlated with PDV at 25 (correlation coefficient of –0.78) and with average wages across occupation categories: High-paying occupations have lower turnover rates.

The results in Table 1 are as follow. First, starting one's career (at age 25) in an occupation category with a high average wage promises a high PDV at 25, but the differences in PDV at 25 across occupation categories is tiny relative to the differences in average wages across them. Second, high-paying occupation categories (either in terms of PDV at 25 or average wages) have less turnover (churning), as measured by the separation rates (the fraction becoming unemployed and/or out of the labor force in a month) and the gross flow rates. The job finding rates for those who become unemployed from high-paying occupation categories are also somewhat lower as well, although this pattern is not very clear (correlation coefficient of –0.21).

What can we make out of these findings? For one, we can safely reject any notion that high wages of an occupation compensate for the risk of becoming unemployed. If that were true, one should find a positive relationship between wages and separation rates across occupation categories. For another, the negative relationship between wages and gross flows suggests that high wages of an occupation reflect special skills and occupation-specific human capital that take time to develop. The fact that the healthcare practitioner, legal, and architecture and engineering occupation categories have the lowest gross flow rates supports this idea. We explore this subject in our ongoing research projects.

5 WAGES AND LIFETIME EARNINGS ACROSS OCCUPATION CATEGORIES: THE ROLE OF EDUCATION AND GENDER

In our analysis so far, we differentiated people by only their age and occupation category. The more traditional approach in labor economics is to group people by their educational attainment and gender. One natural question is whether a person's occupation has more or less information than his or her demographic characteristics such as education and gender. To answer this question, we break down our sample into two educational groups (bachelor's degree or more vs. less than a bachelor's degree, including those with some college education) for each gender.

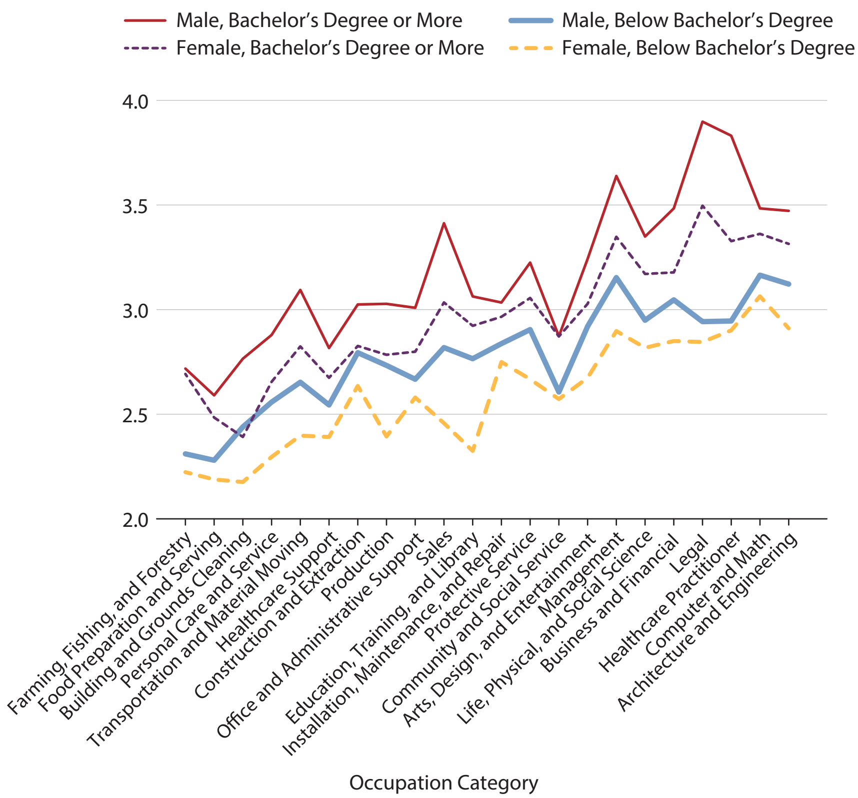
Figure 2 plots in log scales average wages of the 22 occupation categories (sorted by the PDV at 25 variable in Table 1, from lowest to highest) for each of the four education-gender groups. We can readily confirm two well-known facts in labor economics holding across all occupations. First, all occupations have a college premium: The average wages of those with a bachelor's degree or more are higher than the average wages of those with less education, *regardless of gender*.[8] Second, once educational attainment is controlled for, all occupation categories have a gender gap: Men have higher average wages than women of the same educational group in all occupations.[9]

Closer to the theme of this article, we note the sizable differences in average wages across occupation categories within each education-gender group. We also note that the rankings of the occupation categories by average wages are roughly the same across all four education-gender groups. We confirm that occupations contain a significant amount of information on people's economic standing, even after accounting for the roles of education and gender.

We also compute the PDV at 25 by education and gender. We use age-specific average wages and transition probabilities of each education-gender group to construct the PDV at 25.

Figure 2

Average Wages of Occupations, by Education and Gender



NOTE: This figure plots in logs scales the average hourly wages of occupations, from lowest to highest, by education and gender.

SOURCE: Authors' calculations using the CPS Annual Social and Economic Supplement microdata for the period 2003-17.

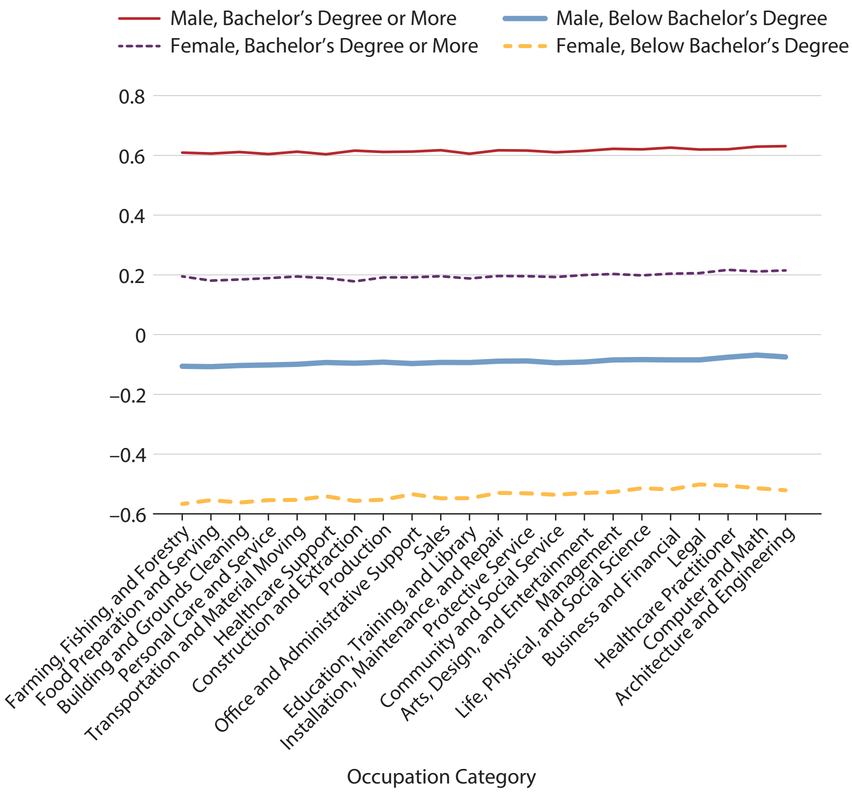
As already noted, there is very little difference in PDV at 25 across occupations in the overall sample (see column "PDV at 25 (norm.)" in Table 1). As shown in Figure 3, this result holds within each education-gender group: All four lines are essentially flat. For all education-gender groups, the ranking of the occupation categories (horizontal axis, from lowest to highest) more or less coincides with the ranking of PDV at 25 for the overall sample .

What is striking is the large gaps among the four education-gender groups. The average wage differences across occupation categories within an education-gender group do not translate into commensurate differences in PDV at 25, because people keep changing occupation categories. The wage differences across education-gender groups, however, persist even in the PDV at 25 calculations because the workers in our sample rarely change their education status, let alone gender.[10]

Furthermore, the differences (in logs) in PDV at 25 across education-gender groups (see Figure 3) are bigger than the differences in average wages across the groups (see Figure 2).

Figure 3

Average PDV at 25 of Occupations, by Education and Gender



NOTE: This figure plots in log scales the PDV at 25 of occupations by education and gender. All PDV at 25 are normalized by their averages.

SOURCE: Authors' calculations using the CPS Outgoing Rotation Groups and Annual Social and Economic Supplement microdata for the period 2003-17.

There are two reasons. First, some groups work more than others. Between 25 and 55 years of age, men with a bachelor's degree or more on average are employed 91.6 percent of all months. The figures are 80.5 percent for women with a bachelor's degree or more, 80.2 percent for men with less education, and 66.2 percent for women with less education. In other words, the groups with higher wages work more (to be precise, are employed for a larger fraction of their prime working years), amplifying the differences in wages across the four education-gender groups. Second, those in the education-gender groups with higher wages across occupation categories are more likely to work in high-wage occupation categories than low-wage occupation categories over the course of their working lives. The composition of occupations within the education-gender groups also contributes to the differences in PDV at 25 across groups being bigger than the between-group differences in average occupational wages.

6 CONCLUDING REMARKS

Occupations do encapsulate a significant amount of information about people's economic standing. However, because people change their occupations and even become unemployed over the course of their working lives, people's occupations at a young age are not very informative about their lifetime earnings. Calculation of the PDV at 25 necessitates estimation of the transition probabilities of moving from one employment status to another and also from one occupation category to another. Our analysis shows that high-paying occupation categories have lower turnover rates: Workers move into and out of high-paying occupation categories at a slower pace, be it employment-to-unemployment or job-to-job transitions.

The lifetime earnings results suggest that one should be careful about readily equating workers' occupations with their skill levels, a common practice in a large macro-labor literature (see Lee and Shin, 2017, and the references therein).

Conditioning on education and gender, the rankings of occupation categories by average wages and by PDV at 25 remain the same. The lifetime earnings differences across education-gender groups are larger than the differences in average wages across occupation categories because high-wage groups work more (at the extensive margin) and also are more likely to work in high-wage occupation categories.

This article leaves out some of our related findings. One is the patterns in worker flows across employment statuses and occupation categories separately computed by gender and education, which we used in our construction of Figure 3. Another is the pattern of worker flows during booms and recessions.[11] These deserve more thorough analyses and are left for future research. ■

APPENDIX

Computing the Present Discounted Value of Lifetime Earnings

There are 22 occupations. Although we know the previous occupation categories of those unemployed, we do not know the previous occupation categories of those out of the labor force. In a given month, there are 45 possible states: employed in occupation categories 1 through 22 ($E1$ to $E22$), unemployed from previous occupation categories 1 through 22 ($O1$ to $O22$), and out of the labor force (N for nonparticipation). For each occupation category o and age a, we denote the value of being employed by V_a^{Eo}, the value of being unemployed by V_a^{Uo}, and the value of being out of the labor force by V_a^N. We can construct the 45-by-45 transition matrix $P_a = \left\{p_a^{i,j}\right\}$, where $P_a^{i,j}$ is the monthly transition probability from state i to state j at age a. The value of any state i can be defined recursively as

$$V_a^i = w_a^i + \beta \sum_{j=1}^{45} p_a^{i,j} V_{a+1}^j,$$

where β is a discount factor and w_a^i is the real hourly wage of the current occupation category (in 2003 dollars) at age a if employed, 40 percent of the real hourly wage of the previous occupation category at age a if unemployed, and zero if out of the labor force.

In matrix form, this system of equations can be written as

$$V_a = w_a + \beta P_a V_{a+1}$$

where $V_a = \left(V_a^{E1}, V_a^{E2}, \ldots, V_a^N\right)^T$ and $w_a = \left(w_a^{E1}, w_a^{E2}, \ldots, w_a^N\right)^T$. The terminal condition at $a = A + 1$ is set to $V_{A+1}^i = 0$ for all i. Forward iteration gives us

$$\begin{aligned} V_a &= w_a + \beta P_a w_{a+1} + \beta^2 P_a P_{a+1} V_{a+2} \\ &= w_a + \beta P_a w_{a+1} + \beta^2 P_a P_{a+1} w_{a+2} + \ldots + \beta^{A-a}\left(P_a P_{a+1} \cdots P_{A-1}\right) w_A . \end{aligned}$$

The expected value of occupation category o in the first month of age 25 is V_{25m1}^{Eo} with $A = 55 \times 12$ months. The monthly discount factor β is assumed to be 0.99835, which is equivalent to a discount rate of 2 percent per year.

In computing the transition probabilities $P_a^{i,j}$, we follow Elsby, Hobijn, and Şahin (2015) to remove possible spurious transitions between the unemployment and out-of-the-labor-force states. For example, if a worker is out of labor force in the first and third months but unemployed in the second month of the survey, we treat the worker as out of the labor force for all three months. The case of a worker who is unemployed in the first and the third months but out of the labor force in the second month is similar: We consider the worker unemployed for all three months.

NOTES

[1] We drop those who usually work for less than 8 hours per week while employed.

[2] We again drop those who usually work for less than 8 hours per week.

[3] By comparison, the more commonly cited Gini coefficient is for household income, which was roughly 0.47 in the United States over the sample period. The numbers are different for two reasons. First, our unit of observation is individuals, not households. With assortative mating, the Gini coefficient would be higher for households. Second, and more important, capital income is much more unequally distributed than labor (wage) income.

[4] In the previous exercise, we assumed that everyone with the same occupation category earned the same wage, regardless of their age.

[5] The PDV at 25 of the cross-sectional Lorenz curve is difficult to construct because the panel dimension of our data is too short for an estimation of lifetime earnings.

[6] Occupation categories that have high EN tend to have more younger workers and also more women, although the correlation is fairly weak.

[7] Worker churning means those hires that replace separations from an employer, or the difference between worker flows and job flows at the employer level (Burgess, Lane, and Stevens, 2000). In the firm dynamics literature, churning is the sum of the number of jobs created and the number of jobs destroyed across firms or establishments in a sector or the whole economy (see, for example, Davis, Haltiwanger, and Schuh, 1996). Our notion of gross flows is defined at the occupation level to measure the turnover rate of an occupation.

[8] The lone exception is the building and grounds cleaning occupation category, in which women with a bachelor's degree have a slightly lower average wage than men with less education.

[9] In one occupation category, community and social service, the gender gap is virtually nonexistent.

[10] Lee, Shin, and Lee (2015) find that for more than 95 percent of the population in the National Longitudinal Survey of Youth, their educational attainment is finalized by 25 years of age.

[11] Aum, Lee, and Shin (2017) use a different dataset without panel dimensions to study the changes in employment across occupations during and after the Great Recession.

REFERENCES

Acemoglu, Daron and Autor, David. "Skills, Tasks and Technologies: Implications for Employment and Earnings," in D. Card and O. Ashenfelter, eds., *Handbook of Labor Economics*. Volume 4, Part B. Elselvier, 2011, pp. 1043-71; https://doi.org/10.1016/S0169-7218(11)02410-5.

Aum, Sangmin; Lee, Sang Yoon (Tim) and Shin, Yongseok. "Industrial and Occupational Employment Changes During the Great Recession." Federal Reserve Bank of St. Louis *Review*, First Quarter 2017, *99*(4), pp. 307-17; https://doi.org/10.20955/r.2017.307-317.

Burgess, Simon; Lane, Julia and Stevens, David. "Job Flows, Worker Flows, and Churning." *Journal of Labor Economics*, 2000, *18*(3), pp. 473-502; https://doi.org/10.1086/209967.

Davis, Steven J.; Haltiwanger, John C. and Schuh, Scott. *Job Creation and Destruction*. MIT Press, 1996.

Elsby, Michael W.L.; Hobijn, Bart and Şahin, Ayşegül. "On the Importance of the Participation Margin for Labor Market Fluctuations." *Journal of Monetary Economics*, 2015, *72*, pp. 64-82; https://doi.org/10.1016/j.jmoneco.2015.01.004.

Lee, Sang Yoon (Tim) and Shin, Yongseok. "Horizontal and Vertical Polarization: Task-Specific Technological Change in a Multi-Sector Economy." NBER Working Paper No. 23283, National Bureau of Economic Research, 2017; https://doi.org/10.3386/w23283.

Lee, Sang Yoon (Tim); Shin, Yongseok and Lee, Donghoon. "The Option Value of Human Capital: Higher Education and Wage Inequality." NBER Working Paper No. 21724, National Bureau of Economic Research, 2015; https://doi.org/10.3386/w21724.

Nekarda, Christopher J. "A Longitudinal Analysis of the Current Population Survey: Assessing the Cyclical Bias of Geographic Mobility." Working paper, Board of Governors of the Federal Reserve System, 2009; https://chrisnekarda.com/papers/lpd-geomob_20090527.pdf.

Sanders, Carl and Taber, Christopher. "Life-Cycle Wage Growth and Heterogeneous Human Capital." *Annual Review of Economics*, 2012, *4*, pp. 399-425; https://doi.org/10.1146/annurev-economics-080511-111011.