

ESSAYS IN INDUSTRIAL ORGANIZATION

by

Edward Y. Zhong

A preliminary report submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy (Economics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2021

Date of final oral examination: January 21, 2021

The dissertation is approved by the following members of the Final Oral Committee:

Alan Sorensen (chair), Professor, Economics

Jean-Francois Houde, Associate Professor, Economics

Ken Hendricks, Professor, Economics

Menzie D. Chinn, Professor, Public Affairs and Economics

© Copyright by Edward Y. Zhong 2021

All Rights Reserved

To my parents, brother, and my partner.

ACKNOWLEDGMENTS

I would especially like to thank my advisors and my committee, including Alan Sorensen, Jean-Francois Houde, Ken Hendricks, and Menzie Chinn, for their invaluable advice and support.

ABSTRACT

Chapter 1. Estimating Bank Deposit Market Power through Pass-Through

I argue that one source of market power for banks' deposit franchise is the lack of alternative liquid savings instruments for many households (ie. most households do not hold stocks or bonds). I use micro-data from the Survey of Income and Program Participation to estimate a discrete-continuous demand system for bank deposits. With these demand estimates, I utilize pass-through data to estimate market power, as measured by a market conduct parameter. In addition to the pass-through framework, I also implement a framework to estimate a lower bound of conduct.

Chapter 2. Subsidies for Vaccine Development

I focus on the development of new vaccines, recognizing the important role (and constraint) of finance for small biotech companies that are a large and growing part of drug development. In particular, I adopt [?]'s framework for calibrating the vaccine free-rider externality in the product market, but I also focus on frictions in the development process, which lead to inefficient investment in vaccines. I find that even absent financing constraints, a monopolist does not invest at the socially optimal level. When financing constraints are added, government capital subsidies can dramatically increase welfare. My work highlights the role of government as cheap providers of capital, and I argue that a reasonable, and sustainable infectious-disease subsidy program should mimic private market joint venture arrangements.

1 Chapter 1. Estimating Bank Deposit Market Power through Pass-through

1.1 Introduction

How competitive is the deposit market for banks? In addition to competition policy, Drechsler et al. (2017) has argued that bank market power for deposits can explain the entire measured effect of monetary policy and real economic activity ¹.

My analysis is centered on using a novel pass-through framework to estimate bank competition, and to reassess the results of Drechsler et al. (2017) using that measure of competition in place of HHI's, which can mismeasure competitiveness. To examine the competitiveness of the deposit market, I focus on the pass-through of short-term market interest rates to banks' deposit rates. Unlike other markets, the price (offered deposit rate) and a substantial fraction of marginal costs (Fed Funds rate) are public and vary at relatively high frequency. Other analyses have argued that the incomplete (and asymmetric) pass-through is a function of market power. For example, Hannan & Berger (1997) shows that deposit rates are sticky, and they tend to fall faster than they rise when interest rates change, and both Neumark & Sharpe (1992) and Drechsler et al. (2017) have correlated this stickiness with measures of market concentration.

To measure competitive intensity and its relationship with pass-through in oligopoly, Weyl & Fabinger (2013) illustrate a framework that incorporates a competitive conduct parameter, which can be interpreted as a "degree" of competition (similar to Genesove &

¹Jeremy Stein has argued that the empirical relationship between the short rate and long-term market interest rates indicates that there are some mechanisms of monetary policy which affect the term premium (Stein & Sunderam (2018)). The bank lending channel is one such mechanism, but while the empirical relationship between interest rates and bank lending is quite strong (Kashyap & Stein (2000)), there hasn't, until recently, been a strong-enough theoretical rationale to explain that relationship. To complement the theoretical arguments, Drechsler et al. (2017) show that, when the Fed raises interest rates, bank branch deposit rates in concentrated markets rise slower and do not fully pass-through rates as much as branches in competitive markets *for branches belonging to the same bank*. Because deposits are the main source of bank lending, this deposit market friction induces changes in bank lending, thus affecting the larger economy. This imperfection is greater the more interest rate insensitive households, and they consider this parameter, the "bank beta", to be the relationship between bank deposit rates and market interest rates.

Mullin (1998)). However, in this framework, a key parameter is demand and the second order properties of demand and, in particular, when demand is highly concave (convex), then oligopolists may pass a little (a lot) of any cost decreases.

Household deposit demand is highly heterogeneous, and some households have all of their liquid assets in bank deposits, while some hold combinations of stocks and bonds (though household participation in these markets is limited, Vissing-Jørgensen (2002)). To correct for the endogeneity of participating in the stock or bond market, I utilize the discrete-continuous demand system framework highlighted by Dubin & McFadden (1984), where households make a discrete choice about the type of portfolio they have (bank deposits only, deposits and bonds, deposits and stocks, or all three), and then chose the share of assets conditional on that portfolio.

To estimate this demand model, I require sufficient price variation in interest rates, and stock returns. As such, I utilize the Survey of Income and Program Participation (SIPP), which is a repeated household level survey taken from the mid-1980s to the present. This period overlaps with substantial interest rate variation: until March of 1986, federal interest rate legislation capped the interest rate on savings accounts that was strongly binding (the cap was 5.25% in 1983, while the 1 year Treasury rate was 9%). Furthermore, the general level of interest rates was quite volatile throughout this period, which provides substantial variation in prices to estimate demand concavity.

Once the demand side² has been estimated, I subsequently use estimates of demand curvature, in conjunction with pass-through estimates, to estimate bank “conduct parameters”. In the context of solar subsidies, Pless et al. (2017) has used a similar test.

Unfortunately, there are some data limitations from my analysis, so that measured conduct suggests more collusion than is plausible. As such, in a separate analysis, I also utilize a lower bound on conduct developed by Sullivan (1985). The lower bound approach effectively rejects a high degree of collusion in markets where there are more than a few banks.

²While households “supply” deposits, to limit any semantic confusion, I consider households “demanding” liquidity, with concomitant opportunity cost of holding higher-yielding assets.

For a roadmap of the remainder of this paper, Section 1.2 describes related literature, and Section 1.3 describes the institutional background of banking and deposit markets. Section 1.4 presents the model of market power and the key relationship between pass-through and demand curvature. Section 1.5 describes the bank deposit demand estimates. Finally, section 1.7 presents the pass-through estimates, estimates of conduct, and relationship between conduct and measured HHIs.

1.2 Related Literature

Several strands of literature have attempted to measure the elasticity of demand for deposits. In the micro-banking literature, for example, Egan et al. (2017) find that both "insured and uninsured deposits are quite price insensitive: with a demand elasticity of 0.56 for insured deposits and 0.16 for uninsured deposits" using data in the 2000s and 2010s. Likewise, Dick (2008) finds an elasticity of approximately 0.5 for deposit rates (assuming a bank with 5 percent market share and a 1 percent interest rate) using data from 1993-1999.

Others have attempted to estimate deposit demand, highlighting different aspects of bank deposits. For example, there may be stickiness in the deposit market (Sharpe (1997)) because of switching costs, because of incomplete consideration of alternative deposit options (Abrams (2019)), or potentially because deposit information can inform lending decisions (Sharpe (1990)); second, deposits may be a differentiated good, as households may prefer characteristics like the breadth of their ATM network (Dick (2008)). The latter source of market power is often modeled using discrete choice methods, and while competition within the deposit market can be modeled well, granular data of county-level holdings of alternative asset holdings does not exist, which limits interpretation of price elasticities given mismeasurement of the outside good.

Several recent papers have attempted to isolate the forms of bank deposit market power. For example, Abrams (2019) is very similar in spirit to my paper, although he focuses on the market power that banks accrue because of household behavioral "inattention". In

particular, households may not shop for potential bank deposit deals, and as such, banks may hold more market power than measured according to a standard discrete choice analysis. Finally, Lin (2019) presents reduced-form evidence suggesting that, in counties where stock market participation is high and stock returns are high, bank deposits fall. His paper provides suggestive evidence that household asset market segmentation is an important driver of bank deposit rates.

In an older generation of literature, there have been methodologies to model a household's holdings of financial assets and holding of money, with the intent of estimating costs of inflation, following an inventory analysis of money holdings (Baumol (1952), Tobin (1956)). For example, Mulligan & Sala-i Martin (2000), using the 1989 Survey of Consumer Finances (1989) cross-section, argue that the extensive margin is more important than the intensive margin: many people hold zero account balances, and zero financial assets. Interestingly, they claim that they can identify the sensitivity of interest rates at very low interest rates, because the hurdle to hold assets is a fixed cost.

Finally, there is significant work on household portfolio choice, typically trying to explain equity market limited participation. For example, Vissing-Jørgensen (2002), using CEX data, tests whether asset holdings are important in estimating the elasticity of intertemporal substitution (EIS), which requires both consumption data and asset holdings for her hypothesis. She finds that the EIS is significantly higher for non-asset holders, suggesting that stock market participation is either by itself or as a proxy to something else, useful to smooth consumption. Bertaut (1998) uses the SCF panel survey (though she does not include 1986) to estimate the likelihood a household becomes a stock-holder, motivating her analysis using a consumption capital asset pricing model (CCAPM).

1.3 History of Bank Deposit Regulation

“After 1980, however, as the political winds shifted, many of the regulations on banks were lifted – and banking became exciting again.”

- Krugman (2009)

After the Great Depression, banking became a heavily regulated industry: ceilings were placed on interest rates for bank deposits, geographic bank entry into new markets was severely curtailed, and traditional banking activity was separated from other brokerage-type activities. The period from 1940 to 1980 was thus “boring banking”, characterized by the 3-6-3 rule: pay 3 percent on deposits, lend at 6 percent, go to the golf course by 3pm. Consistent with this characterization, banks were typically inefficient (Berger & Hannan (1998)), and managers were incentivized to maintain stable, but limited rents ³.

The post-1980 period deregulated all of these Depression-era barriers. During that era of financial liberalization, households were given access to more financial savings products and the structure of commercial banking changed significantly. This provides, as described in the first section, significant price variation in interest rates to estimate demand concavity. In the second subsection, I discuss bank branching deregulation, which motivates the discussion about estimating the conduct of banks, and finally, I conclude with some limited discussion about the Gramm-Leach-Bliley Act (GLBA).

1.3.1 De-Regulation of Reg Q

Regulation Q, as part of the Glass-Steagall Act of 1933, prohibited banks from paying interest on the majority of demand accounts (checking accounts) and heavily restricted the maximum rate of interest on non-demand accounts. These regulations were motivated by the bank failures of the Great Depression, where “ruinous” deposit competition increased costs for banks and was cited as a contributor to their failure. For most of the history of Reg Q, the interest rate ceilings were not binding, though that changed as a result of high inflation in the 1970s and 1980s. However, rather than limiting competition in the 1970s and

³For example, Philippon & Reshef (2012), for example, show that bankers’ relative wages and relative education declined by nearly 40 percent and 25 percent respectively from 1940 to 1980. Interestingly, these rents were shared with labor in the form of labor market discrimination: after bank branching was allowed, wages for male bankers fell 12 percent, while women’s wages fell only 3 percent, suggesting that “rents were shared mainly with men” ?.

80s, banks found that deposit ceilings were quite restrictive as new institutions, like money market funds, competed aggressively against banks and took market share.

Banks, in response, lobbied to offer new products, which differ in the degree of liquidity provision: some assets that are more money-like (can be used for transactions) and non-monetary assets (“interest-bearing assets”). For example, banks started to offer Negotiable Order of Withdrawal (NOW) accounts and some non-transaction accounts including MM-DAs⁴. Interestingly, and not inconsequentially for subsequent discussion, the Fed broadened the definition of M1 to include these financial innovations⁵.

Figure 1 shows selected interest rates. Interest rates on savings accounts (passbook savings) were capped at 5 percent throughout the late 70s, and were increased to 5.25 percent in 1979 (Gilbert et al. (1986)). Banks were allowed to increase their products to include ATS accounts (automatic transfer service accounts, which automatically transfers funds from a savings account to a checking account), and NOW accounts (negotiable order of withdrawal accounts, effectively a savings account with check writing privileges)⁶. Note

⁴For reference, a demand deposit account (e.g. a checking account) differs from a negotiable order of withdrawal account based on the amount of notice you need to give to the bank or credit union before making a withdrawal. “Most demand deposit accounts (DDAs) let you withdraw your money without advance notice, but the term also includes accounts that require six days or less of advance notice. NOW accounts are essentially checking accounts where you earn interest on the money you have deposited. With a NOW account, the bank or credit union has the right to require at least seven days written notice of a withdrawal, though this is rarely done. Not all accounts that give you checks are checking accounts. Other deposit products, such as money market accounts, may allow you to write checks, but they are not generally suited for day-to-day business, given the restrictions on their use. In addition, a lender may give you checks to access credit, such as a personal loan, home equity loan, or other lines of credit. These types of checks allow you to access your loan.”

⁵For reference, M1 currently consists of (1) currency outside the U.S. Treasury, Federal Reserve Banks, and the vaults of depository institutions; (2) traveler’s checks of nonbank issuers; (3) demand deposits at commercial banks (excluding those amounts held by depository institutions, the U.S. government, and foreign banks and official institutions) less cash items in the process of collection and Federal Reserve float; and (4) other checkable deposits (OCDs), consisting of negotiable order of withdrawal (NOW) and automatic transfer service (ATS) accounts at depository institutions, credit union share draft accounts, and demand deposits at thrift institutions. Seasonally adjusted M1 is constructed by summing currency, traveler’s checks, demand deposits, and OCDs, each seasonally adjusted separately.

In contrast, M2 consists of M1 plus (1) savings deposits (including money market deposit accounts); (2) small-denomination time deposits (time deposits in amounts of less than \$100,000), less individual retirement account (IRA) and Keogh balances at depository institutions; and (3) balances in retail money market mutual funds, less IRA and Keogh balances at money market mutual funds. Seasonally adjusted M2 is constructed by summing savings deposits, small-denomination time deposits, and retail money funds, each seasonally adjusted separately, and adding this result to seasonally adjusted M1.

⁶Not to be confused with Super-NOW accounts

that both accounts were capped at 5.25 percent, in line with passbook savings, and interest caps were eliminated on all three types of accounts by March 1986.

Note that after the interest rates caps were eliminated on savings (March of 1986, just before the SCF 1986 survey wave), the surveyed rates **did not change**, which immediately suggests some significant interest rate insensitivity. Regulation Q was gradually weakened by the Depository Institutions Deregulation and Monetary Control Act (DIDMCA) and in 1982 with Garn-St. Germain, so that by the 2000s, there were no interest rate caps on non-transaction deposits and there were some loopholes for interest rate payments on transaction deposits⁷.

1.3.2 Bank Branching

There are a few permutations of branching restrictions through the boring bank period: many states restricted de novo intra-state and inter-state branching, and some states restricted branching by M&A⁸. While many of these restrictions were lifted in bilateral negotiations between states, all banking restrictions were fully abrogated under the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994⁸. From 1980 to 1999, there were on average 400 bank mergers per year (increasing as states allowed mergers and then after the 1994 act); the total number of banks fell almost 50%. Some of the decline in banking firms is the result of bankruptcies. Interestingly, it is possible that entry restrictions may have incentivized banks to be safer, given the moral hazard associated with deposit insurance (Keeley (1990)).

The final piece of major deregulatory legislation was the Financial Services Modernization Act (also known as the Gramm-Leach-Bliley Act) of 1999, which formally repealed the last vestiges of Glass-Steagall. Through this period of deregulation and financial liberalization, bank deposit markets consolidated significantly, while also competing in new markets formerly outside of their domain.

⁷Post-script: business checking accounts were not deregulated until 2011, under Dodd-Frank. Interestingly, last vestige of small bank versus large bank divide (Capitol One).

⁸There is some interesting political economy (e.g. if firms use regulation to deter entry, why is there deregulation?), which Kroszner & Strahan (1999) tries to explain.

1.4 Pass-Through and Market Power

Bank deposit rates exhibit asymmetric interest rate pass-through (Hannan & Berger (1997)), and the degree of downward price rigidity and upward price flexibility has been associated with market concentration (Neumark & Sharpe (1992)). More contemporaneously, this association of limited pass-through has been verified by Drechsler et al. (2017), where an increase in the Herfindahl-Hirschman Index (HHI) is associated with a decrease in pass-through rates even controlling for the same bank⁹.

However, as Weyl & Fabinger (2013) and Ritz (2015) point out, limited pass-through (and even asymmetric pass-through) is not by itself sufficient to identify market power, and measures of market power through HHIs have been problematic. Alternatively, this strand of analysis provide a general model of competitive conduct that describes the relationship between pass-through, properties of demand, and market power.

Consider now the Monti-Klein oligopolistic model of banking, where banks confront an downward-sloping demand for loans $L(r_L)$ and an upward-sloping supply of deposits $D(r_D)$. In this homogeneous goods setting, the bank's profit function is given by:

$$\pi(l_i, d_i) = r_L(L)l_i + rm_i - r_D(D)d_i - C(d_i, l_i)$$

where m_i is the net position of the bank in the interbank market. If α is the share of deposits held in reserve, then $m_i = (1 - \alpha)d_i - l_i$ as an accounting definition¹⁰.

Within a local market, total loans and deposits are given by $L = \sum_n l_i$ and $D = \sum_n d_i$. Consider a conjecture $\tilde{\theta}_i = \frac{\sum_{j \neq i} dq_i}{dq_i}$. Substituting the net interbank position, the bank maximizes:

⁹In concentrated markets, banks are able to increase the spread between the risk free rate and their cost of deposits. They estimate a “spread beta” (ie. pass-through) to be 0.54, which they argue indicates “substantial market power”. Furthermore, they argue that “relative to keeping rates unchanged, a typical 400-bps Fed hiking cycle induces a 1,458 bps reduction in deposits and a 995 bps reduction in lending”.

¹⁰Of course, this static profit function ignores several considerations that banks have in practice, including loan loss provisions and liquidity requirements (which would require modeling dynamics).

$$\pi(l_i, d_i) = (r_L(L) - r)l_i + (r(1 - \alpha) - r_D(D))d_i - C(d_i, l_i)$$

where the first order condition for deposits is:

$$\frac{\partial \pi}{\partial D} = -r'_D(D)(1 + \tilde{\theta}_i)d_i + r(1 - \alpha) - r_D - C'_D(d_i, l_i) = 0$$

where θ represents the perceived conduct of other banks to bank i 's shift in deposits. I have to assume that changes in conduct are invariant to changes in quantity (ie $\epsilon_\theta = 0$). In some sense, by not taking a stand *a priori* on conduct, this is impossible to get around. However, in several contexts, this is a natural consequence, for example, in the setting of homogenous-products Nash-Cournot. This assumption, however, is not innocuous for differentiated-products discrete choice models with Nash-Bertrand competition.

In the Monti-Klein model, the interbank market, and the corresponding price—the Fed Funds rate—allow excess deposits to be spread nationally to equalize the marginal return on deposits. Assuming that the cost of servicing deposits and loans are separable, then the first order condition on deposits depends only on the local deposit market, and not on loans.

Hence, I rewrite the deposit condition as:

$$-r'_D(D)(1 + \tilde{\theta}_i)d_i + r(1 - \alpha) - r_D - c_d = 0 \tag{1}$$

This first order condition is very similar to Genesove & Mullin (1998), except the $r(1 - \alpha)$ represents the marginal return on deposits for a bank. Except for the sign differences, the Fed Funds rate acts similarly to a marginal cost shock. In symmetric equilibrium, where $d_i = \frac{D}{N}$,

$$\frac{r(1 - \alpha) - r_D^* - c_d}{r_D^*} \epsilon_D(r_D^*) = \frac{1 + \tilde{\theta}_i}{N}$$

and a higher conduct parameter, all else equal, results in a lower equilibrium deposit interest rate.

In the spirit of the pass-through literature, consider equilibrium price $r_D^*(r)$ and quantity $D^*(r)$ as a function of r , the Fed Funds rate. Define pass-through as $\rho(r) \equiv \frac{\partial r_D^*(r)}{\partial r}$, and note that $\rho(r) = r'_D \frac{dD}{dr}|_{D=D^*}$. Finally, define the curvature of inverse demand to be $\xi \equiv -\frac{r''_D D}{r'_D}$. Taking the envelope condition with respect to the Fed Funds rate in equation 1:

$$[-r''_D(D)\theta D - r'_D(D)\theta] \frac{dD}{dr} + (1 - \alpha) - r'_D \frac{dD}{dr} = 0$$

where $\theta = \frac{1+\tilde{\theta}}{N}$. Note that perfect competition corresponds to $\theta = 0$, collusion is $\theta = 1$, and Cournot corresponds to $\theta = \frac{1}{N}$. Note that marginal costs, similar to Egan et al. (2017), are the real costs of servicing deposits. In practice, based on Call Reports data, the interest expense on deposits is the largest fraction of total variable costs, while expenses on branches and the ATM network are fixed in the short-run. However, servicing incremental depositors, given a bank branch or ATM network, is not assumed to vary with the Fed Funds rate. Substituting pass-through and demand curvature, the relationship between pass-through and conduct is given by:

$$\theta \xi \rho - \theta \rho + (1 - \alpha) - \rho = 0 \Rightarrow \theta = \frac{1 - \alpha - \rho}{\rho(1 - \xi)} \quad (2)$$

For the remainder of my analysis, I begin with an empirical study of pass-through rates. How are deposit rates passed-through under various conditions? Next, I estimate conduct parameters, and finally, I compare my results with an alternative method of bounding conduct parameters, given by Sullivan (1985).

1.5 Estimating Bank Deposit Demand

Household participation in different asset markets is quite limited. For example, Vissing-Jørgensen (2002), using data from the PSID, finds that the majority of households at any given time do not hold stock, though households change from non-stockholders to stockholders and vice versa. These findings are consistent with Table 4, where the majority of

households do not change their portfolios. For example, 84% of households that hold only bank assets (“money”) in the prior year continue to hold only bank assets the following year.

If the markets for bank deposits, bonds, and stocks are segmented for some households, then households that are restricted to bank deposits are effectively “captured” by the banking sector. This can induce a non-linear aggregate demand function if the deposit rate sensitivity of households with different portfolios is substantially different. For the purposes of this paper, I don’t take a stand on the particular frictions that segment households—these could be psychological costs to holding stocks and bonds, or some form of risk aversion¹¹.

This initial portfolio choice introduces some endogeneity in demand, and in particular, suppose you ran the following linear demand equation: $y_i = \alpha\theta + \beta X_i' + \eta_i$ where i indexes individual households. Running this regression is likely to bias any results for deposits because the unobserved characteristics of households η_i are likely to be correlated with the decision to have multiple potential assets. The key insight of Perraudin & Sørensen (2000) is to think of the “discrete” choice as deciding whether to have a portfolio of mixed assets (ie. deposits and stocks), while the “continuous” choice is based on the share of assets allocated to each asset class.

For ease of exposition, I start by discussing the discrete choice of portfolios, assuming a household has picked “money and bonds”. I focus on households choosing between money (M), money and stocks (MS), and money, bonds and stocks (MSB), since based on Table 4, the population of households choosing other portfolios is small. I consider a linear model of demand for the share of a households’ liquid assets y_i^{MS} , with $y^{MS,*}$ as an indicator function for household i to choose portfolio MS .

¹¹Interestingly, while banks face this demand curve, they appear not to price discriminate to determine whether a household holds alternative assets

$$y_{MS} = x\beta_{MS} + u_{MS}$$

$$y_{MS}^* = z\gamma_{MS} + \eta_{MS}, j \in \{M, MS, MSB\}$$

Note that u_i^{MS} can have any form, though I assume that $E[u_{MS}|x, z] = 0$ and $V(u_{MS}|x, z) = \sigma^2$. While the vector x includes variables that affect the allocation of assets into money, the vector z includes the same decision variables but also an indicator for the prior portfolio choice, which acts as an exclusion restriction and helps identify the selection equation.

Note that, for households selecting portfolio MS , we only observe the demand for money conditional on selecting MS (consequently, households that selection money only do not have an outcome equation, since their allocation to money as a fraction of liquid assets is always one). Hence:

$$y_{MS}^* > \max_{j \neq MS} (y_j^*)$$

Define ϵ_j as the difference between utilities in the selection equation, such that

$$\epsilon_j = \max_{j \neq MS} (y_j^* - y_{MS}^*) = \max_{j \neq MS} (z\gamma_j + \eta_j - z\gamma_{MS} - \eta_{MS})$$

Assume that the η_j 's are independent and identically Gumbel distributed, which leads to the following multinomial logit portfolio selection model:

$$P(\epsilon_1 < 0|z) = \frac{\exp(z\gamma_1)}{\sum_j \exp(z\gamma_j)}$$

.

The issue is estimation of β_{MS} given the potential correlation of u_{MS} and η_j 's. Define

$$\Gamma = \{z\gamma_1, \dots, z\gamma_{MSB}\}$$

so that:

$$E[u_{MS}|\epsilon < 0, \Gamma] = \int \int_{-\infty}^0 \frac{u_{MS} f(u_{MS}, \epsilon_1 | \Gamma)}{P(\epsilon_1 < 0 | \Gamma)} d\epsilon_1 du_{MS} = \lambda(\Gamma)$$

where $f(u_{MS}, \epsilon_1 | \Gamma)$ is the conditional joint density of u_{MS} and ϵ_1 . Importantly, this conditional expectation can be substituted to the demand equation (where $E[u_1 | \epsilon_1 < 0, \Gamma] = \mu(P_1, \dots, P_M)$) to $y_1 = x_1 \beta_1 + \mu(P_1, \dots, P_M) + w_1$ where w_1 is uncorrelated to y_1 .

The key assumption for Dubin & McFadden (1984) is the following:

Assumption 1 $E(u_1 | \eta_1, \dots, \eta_M) = \sigma \sum_j r_j (\eta_j - E[\eta_j])$, with $\sum_j r_j = 0$

This is an assumption on the covariance structure of the selection equation and the demand equation. By substituting the constraint that all of the correlations add to 0 and taking advantage of the *iid* assumption on the η_j 's, we can re-write this as: $E(u_1 | \eta_1, \dots, \eta_M) = \sigma \sum_j r_j (\eta_j - \eta_1)$.

Given that $E(\eta_j - \eta_1 | y_1^* > \max(y_s^*), \Gamma) = \frac{P_j \ln(P_j)}{1 - P_j} + \ln(P_1) \cdot \forall j > 1$, we can write the demand equation as:

$$y_{MS} = x_{MS} \beta_{MS} + \sigma \sum_{j \in \{M, MS, MSB\}} r_j \left[\frac{P_j \ln(P_j)}{1 - P_j} + \ln(P_{MS}) \right] + w_{MS} \quad (3)$$

which can be estimated through least squares.

1.5.1 Data

The main data come from two sources: first, I use the Survey of Income and Program Participation (SIPP) to examine household asset holdings, and second, I get price data from RateWatch, a company that has surveyed bank branches weekly for over 20 years. To supplement these sources, I also look at stockholding (at the county level) from the IRS.

1.5.2 Household Asset Data: Survey of Income and Program Participation (SIPP)

The ideal household dataset should include repeated measurements of detailed household asset holdings, in addition to demographic and other characteristics, given the many idiosyncratic factors affecting household asset demand. Furthermore, the survey should be sufficiently long, so that one can use the variation across time in asset prices to estimate price elasticities. Several household surveys potentially have this information, though through an exhaustive search, the Survey of Income and Program Participation is the closest to meeting these requirements.

The SIPP data is a repeated survey of households—there are approximately 20,000 households interviewed per survey wave (see Table 1). These households are then followed for several years (the length of follow-up is variable since the survey has changed over time). There are at least two supplemental interviews (“topical modules”) where households provide detailed data on their liquid asset holdings (stocks, bonds, etc.), while income is reported during each interview.¹² Unfortunately, the data does not include consumption information, though income data is provided with granularity. Finally, the data aggregates liquid financial asset ownership into bank assets, bonds, and stocks¹³.

Because I use repeated cross-sections, I deflate both deposit variables and wealth/income variables by CPI. The relevant cost (the “spread”) s_t^d is defined as the 1 year end-of-month Treasury interest rate minus the deposit ceiling interest rate (pre-1986) or floating rate (post-1986) described in the previous section. I define net worth as the sum of all assets net of secured debt (e.g. mortgages, vehicle debt), which excludes credit card debt (which is largely unavailable across other surveys). I define liquid financial assets as the sum of interest-bearing bank assets, stocks, and bonds (US government securities, municipal or

¹²Unfortunately, modules from 1987 to 1993 do not ask about certain asset holdings. As such, I have imputed some asset holding information for those years.

¹³It is possible to infer some asset holdings based on the income provided during each interview. However, this tends to overstate the asset holdings, especially for bank deposits

corporate bonds; does not include US bonds)¹⁴. Income in the data is the sum of wage income, farm/self-employment income, interest and dividends, social security income, retirement/disability, public assistance, and other. Finally, primary residence equity and mortgage data is not available for certain years (1990-1993). For primary residence equity, I merge in state-level house prices. For mortgage data, I use monthly payment data and calculate the amortization schedule. Other non-liquid assets include retirement accounts, vehicle wealth, and other real estate, which are all observed for the relevant time periods.

There are some specific issues with the data. First, many households report having zero savings (and I restrict the analysis to households with a threshold level of net worth). Second, I do not observe the exact interest rate that individuals receive on their deposit accounts, the expected returns on equity for their portfolios, or the exact interest rate of bonds they hold (for example, they may hold a portfolio of distressed debt, which yields higher expected returns). Finally, there may be other unobserved characteristics that affect a household's level of savings that are not captured in the survey (e.g. consumption).

Table 2 describes a variety of filters to clean the data. I drop the 1987 data because it lacks too much data. I drop retirees because their asset holdings may differ based on their life-cycle, and I drop self-employed households because their motivation to hold alternative assets differs from regular wage-earners¹⁵. Business owners have substantial net worth outside of the home, thus biasing some of the liquid asset holdings data. Furthermore, because small business owners have a substantial share of their net worth in an illiquid asset, they may be much more sensitive to interest rates. Interestingly, in the data, as a share of their net worth, they hold relatively less deposits. Finally, I restrict financial net worth to be at least \$5,000 and less than \$5,000,000 adjusted for CPI.

Table 3 presents the full summary statistics for the dataset. In the sample, most households are homeowners (historically, home ownership is hovers near 60%). Finally, most

¹⁴Note that the data includes more granular information on some liquid bank assets, including regular (passbook) savings, money market accounts, CDs, and NOW/Super NOW Accounts. For my purposes, I aggregate these assets together into interest-bearing bank assets.

¹⁵Additionally, business equity and business secured debt are not reported for several years.

households in the dataset have a net worth of nearly \$200,000. The primary dependent variable is the share of bank assets to other liquid assets held by households. I use two different measures of household wealth: income and total net worth. Finally, for demographic controls, X_{it} , I (can) include household wage, number of earners, urban or rural, family size, age of head of household, and education of the head-of-household.

1.5.3 Price and Return Data

I use data from RateWatch for advertised deposit rates, which represents about a quarter of all bank branches in 2007. Bank branches are surveyed weekly, and I limit the analysis to a quarterly frequency. To exclude double-counting, I only include rate-setting branches. Finally, I use data for money market accounts with an account size of \$25,000, the most common savings account (similar to Drechsler et al. (2017)). The data spans 1997 to 2008.

For context, figure 1 shows the history of interest rates¹⁶. For the demand analysis, I look at the spread between the Fed Funds rate, a measure of the risk free rate, and offered rates. Note that the savings rate in the early 1980s was capped by Reg Q's interest rate ceilings. Interestingly, after interest rates caps were eliminated on savings (March of 1986), the surveyed rates **did not change** for several years.

In addition to the deposit rate spread, I also include data on S&P500 returns, with a two-year return history (see Figure 2). While this is similar to Lin (2019), I have also used expected returns from Campbell & Thompson (2008). Likewise, for bond returns, I use the spread between the Moody's AAA rated total return against Fed Funds (see Figure 3). Note that for both this spread and other interest rate variables, I use nominal (rather than real) interest rates.

¹⁶Historical data comes from the Federal Reserve's FR2042 series (the Federal Reserve's Monthly Survey of Selected Deposits and Other Accounts), which was the primary data for the studies by Hannan & Berger (1991) and Neumark & Sharpe (1992). The data was a monthly survey sent to a rotating panel of commercial banks to record a variety of deposit rates, including time deposits (CDs), IRA/Keogh plans, and most relevant, a variety of savings accounts from 1980 to 1994

1.6 Demand Estimation Results

Before getting to the estimation, I first discuss the qualitative evidence of limited participation, and differential elasticity by type of individual. First, even among a sample of households with sufficient liquid assets, about a third of households hold only bank deposits (see Table 4). Furthermore, households tend to keep their portfolio choices: 84% of households that held only bank deposits in the prior survey wave continue to hold only bank deposits. I restrict the analysis to portfolios of “money only” (M), “money and stocks” (MS), and “money, stocks, and bonds” (MSB), since the other portfolios have very few households. Second, wealthier households, on average, are more likely to have alternative assets (see Table 5), which is not very surprising.

Interestingly, households tend to report a constant level of prior savings. For example, nearly two thirds of the data report 0 savings accounts; of those that report positive accounts, more than half of the data show that households do not change their reported savings from one year to the next.

Before estimating the discrete-continuous model, I estimate a log-linear specification without the correction for selection. Table 6 shows the estimates using both a linear deposit spread and squared deposit spread. With a linear specification, both MS and MSB households decrease their share of deposits holdings equally given an increase in the deposit spread. However, after adding the squared term, MSB households are slightly more interest rate sensitive than MS households across empirically relevant levels of the spread. Interestingly, MSB households are also more sensitive to bond spreads, presumably because they already have access to bond market securities and may allocate some of their liquid wealth to those assets relative to MS households.

To estimate the the discrete-continuous demand model given by Equation 3, I implement the two-step estimator described by Bourguignon et al. (2007). Table 7 shows the results of the selection equation. Note that the spread¹⁷ is positively related to choosing a

¹⁷As a reminder, this is defined as the Fed Funds rate minus the deposit rate, which is usually positive.

portfolio of bank-alternative assets. Finally, the relative signs and magnitudes of choosing a portfolio with bonds is correlated to the relative attractiveness of bonds. Finally, financial sophistication, as proxied by education, increases households' willingness to hold a broader portfolio.

Table 8 shows the regression results of the outcome equation. Standard errors are calculated based on bootstrap estimates (10,000 samples). Note that, correcting for selection, households' demand for money (bank savings) declines: a 1 percent increase in the spread between the Fed Funds rate and deposit rates decreases the share of savings allocated to money by about 4%, and 3.5% for "MS", and "MSB" households, respectively. Similarly, when S&P returns are high, households allocate less of their assets to money. Finally, households with high net worth tend to have less of their savings in money.

Note that the ρ parameters detail the estimated correlation between the unobserved selection error and the unobserved characteristics in the outcome equation. Interestingly, ρ_1 (the selection into money only) is negatively correlated for all of these choices, suggesting that individuals who only have bank assets differ in some unobserved ways.

1.7 Demand Curvature, Pass-through and Conduct

For some intuition for the second order properties, I first calculate the elasticity for a hypothetical "average" household. Figure 4 shows simulated demand, probabilities of choosing various portfolios, aggregate demand, and the curvature of demand varying the interest rate spread. Panel A shows that households, as the interest rate spread increases, leave the deposit-only portfolio and move to deposits and stocks. Because of the logit assumption, the rate of departure from the money-only portfolio is proportional to the existing probabilities of the various portfolios. Panel B describes the conditional demand as a function of the interest rate spread. The decline for *MSB* households is particularly striking: they hold almost no deposits when interest rate spreads rise to 4%. Panel C is the expected demand for the aggregate household as the interest rate spread increases. Finally, the Panel D shows

demand curvature, which decreases when interest rate spreads increase¹⁸.

1.7.1 Pass-through Estimates

For both the second-order and Sullivan approaches, a key relationship is an estimate of equilibrium interest rates as a function of the Fed Funds rate at the county level. However, estimation is confounded by three different factors. First, counties are demographically heterogeneous: some counties may have wealthier, more financially literate (and interest rate sensitive) populations. To control for this potential sensitivity, I include demographic controls interacted with the change in Fed Funds.

Second, counties have different market structure, and in particular, counties with more banks may be more sensitive to interest rate changes than concentrated counties. While it may seem inappropriate to include a measure of competitive intensity when trying to derive competitive intensity, I argue that entry is fixed in the short-run. For example, both de novo entry and entry through acquisition requires a long review from banking regulators. As such, interacting the log number of banks with the Fed Funds rate conditions on the existing market structure to estimate short-run competitive intensity.

Finally, banks may have different lending opportunities. For example, banks that are in only one market are subject to the local economic conditions of that market, while banks that are in multiple counties can optimize lending opportunities across those counties. To alleviate this concern, I include bank fixed effects, which isolates Fed Funds variation to banks with branches in multiple counties. This identification assumes that bank branches optimize local deposit conditions, while allocating capital to equalize the return on lending opportunities across counties.

Before estimating the pass-through equation, I provide some summary statistics of the

¹⁸To calculate elasticities, I follow Dubin & McFadden (1984):

$$E[D] = \sum E[D|j]P_j \Rightarrow \frac{\partial E[D]}{\partial r^M} = \sum \left(\frac{\partial E[D|j]}{\partial r^M} P_j + E[D|j] \frac{\partial P_j}{\partial r^M} \right)$$

. Appendix 1.12 details the calculations for elasticities.

pass-through data by year (Table 9). For each year, I provide the average number of banks, the average spread, the variance of the *spread* of interest rates, the median HHI of the whole country, the average HHI of the banks in the sample, and the median asset size of all banks and banks in the sample. For the most part, interest rates on savings accounts tend not to move substantially, while the Fed Funds rate tends to move a lot.

To implement this estimate of the equilibrium pricing function $r_D^*(r)$, I run the following regression:

$$\Delta r_{ijmt}^D = \alpha + \beta \Delta FF_t + \beta \Delta FF_t * X_{mt} + \delta_i + \lambda_j + \xi_m + \epsilon_{ijmt} \quad (4)$$

where Δr_{ijmt}^D is the change of the branch-level advertised deposit rate from one quarter to the next for bank j , county m , and time t , and ΔFF_t is the contemporaneous change in the Fed Funds rate. The interaction terms X_{mt} vary by county (and potentially time). Finally, I include branch fixed effects (δ_i), bank fixed effects (λ_j), and county fixed effects (ξ_m).

The branch county fixed effects control for potential effects of management (e.g. quality of management) or location within a county that yields a local advantage. County fixed effects control for county-specific variation.

Table 10 describes the estimated pass-through rates for money market accounts. The average pass-through, without interactions, is about 20%. In column 2, similar to Drechsler et al. (2017), increasing the number of banks (and consequently, HHI) increases the pass-through. With more controls in column 3, the effect of increasing banks is consistently positive. As a proxy for financial sophistication, the effect of college education increases the sensitivity of the deposit rate to Fed Funds.

I also consider the pass-through of CD rates in Table 11. There is substantially greater pass-through of this product, potentially because households trade-off liquidity for higher interest rates. Consistent with money market accounts, increasing the number of banks increases the sensitivity of deposit rates to the Fed Funds rate. Column 1, for example, suggests twice the amount of pass-through for CDs than for money market accounts.

1.7.2 Estimates of Conduct

To compute conduct, consider now ξ as described in equation 2, which is a function of the inverse demand curve. To re-write ξ as a function of demand, I follow Weyl & Fabinger (2013), noting that the elasticity of marginal surplus is equal to the ξ term: $\frac{1}{\epsilon_{ms}} = 1 + \frac{p''q}{p'} = 1 - \xi$. Noting that $(\log D)' = -\frac{1}{ms}$ and $(\log D)'' = -\frac{1}{\epsilon_{ms}} \frac{1}{ms^2}$, then $1 - \xi = \frac{(\log D)''}{[(\log D)']^2}$. Note that if demand is concave, as is the case here, then $\frac{1}{\epsilon_{ms}} > 1$, suggesting more "cost-absorbing" demand.

I compute curvature by county using both the log-linear specification and the discrete-continuous demand model. However, for the discrete-continuous model, the SIPP does not provide geographic information, so I have to rely on average characteristics for the county to provide second order

To aggregate the log-linear specification, I simply use the estimates on *spread* and *spread*² from table 6. For households with stocks and households with stocks and bonds, the point estimates for curvature are 1.2445 and 3.97 when the spread is 3%, respectively. To estimate the share of households that are in each bucket, I use the probability estimates from the discrete-continuous demand model.

Panel A of figure 5 shows the estimated conduct parameters relative to HHI, using the demand estimates from the linear specification. Using money market pass-through, the estimated conduct is well above 1. While the linear model produces a curvature greater than 1 ("cost absorbing"), the pass-through of money market rates is too low to be justified by estimated demand. There are several potential reasons for this.

First, my measure of the demand for deposits is based on the substitution between deposits and stocks and bonds. As illustrated by Abrams (2019), there may be other sources of market power for banks, such as household inattention and "lock-in" when households have the potential to switch banks.

Second, my measure of demand considers typical households that have substantially less than the FDIC insurance limit, whereas a (relatively large) minority of deposits (especially

brokered deposits) are greater than the FDIC limit. This is unlikely to contribute too much error, since the money market product has a minimum balance of \$25K, which is not used by that section of the market.

Finally, one potential source of error may be that my demand estimates are based on all of the deposits (checking, money market/savings, and CDs) that a household holds. If households are much more sensitive to CD products rather than money market products, then that would reconcile the conduct estimates from Panel B. In other words, the theoretically correct deposit rate would be a price index of the different advertised products, which is outside the scope of this paper.

While the scale of these conduct estimates are hard to interpret, the correlation with HHI is clear. In general, counties that are more competitive also have lower estimated conduct parameters. This is also true of the estimates from the discrete-continuous demand.

1.7.3 Re-examining Drechsler, Savov and Schnabl

While some of my measures of conduct are implausible, it may still be instructive to examine my measures of conduct in place of HHI's to examine the robustness of Drechsler et al. (2017)'s results. All of my measures of competitive conduct suggest that market power is substantial in bank deposits.

Table 12 replicates the specification with state interacted with quarter, branch, county, and quarter fixed effects. The coefficient on HHI is slightly higher than the published number, though that may be because the time period is slightly different. Drechsler et al. (2017) finds that when the Fed Funds rate rises 100bps, spreads in high concentration branches increase 14bps more than in low concentration branches. Similarly, my measures suggests that the different spread between the 90th percentile branch and the 10th percentile branch, in terms of competitiveness, increases by 17, 17, 14, and 52 bps for using $\theta_{MM,u}$, $\theta_{CD,u}$, $\theta_{MM,dc}$, and $\theta_{CD,dc}$, respectively. These estimates are surprisingly consistent.

Table 13 describes the changes in quantity using different measurements of conduct. To

put these values in context, Drechsler et al. (2017) calculate an implied semi-elasticity of deposits with respect to deposit spreads of -5.3. My estimates are -4.4, -3.0, -2.4, and -0.14, for $\theta_{MM,u}$, $\theta_{CD,u}$, $\theta_{MM,dc}$, and $\theta_{CD,dc}$, respectively. Using their hypothetical scenario of a 400 bps Fed hiking cycle in 2014 (when deposits were \$9.3 trillion), they find that there would be \$1.2 trillion of deposit outflows. Using my estimates, that deposit outflow would be close to what they measure (on the high end), half of what they measure (using the middle estimate), and only two percent of what they measure using the lowest estimate.

1.8 Lower Bound on Conduct

While the previous analysis suggested a high degree of market power, I consider now an approach pioneered by Sullivan (1985) which describes a lower bound of conduct. In the same setting of homogeneous goods competition, this framework utilizes only the first order conditions of profit maximization and allows for differences in marginal costs. Indeed, I only assume that there is some lowest marginal cost c_d for all firms in the industry.

Define equilibrium levels of deposits as a function of the Fed Funds rate, $d_i^*(r)$, so that $D^*(r) = \sum d_i^*(r)$ and $r_D^*(r) = r_D(D^*(r))$ (and note that $\frac{dr_D^*}{dr} = r_D' \frac{dD^*(r)}{dr}$).

Substituting as a function of r , $-\frac{dr_D^*}{dr} / \frac{dD^*(r)}{dr} [(1 + \theta_i)] d_i^*(r) + r(1 - \alpha) - r_D^*(r) - c_d \geq 0$, where equality holds for the lowest marginal cost. Summing all of the first order conditions and re-arranging, we have:

$$\sum \frac{1}{1 + \theta_i} \geq n^*(c, r) \equiv \frac{\frac{dr_D^*}{dr} D^*(r)}{[r(1 - \alpha) - r_D^*(r) - c_d] \frac{dD^*(r)}{dr}} \quad (5)$$

The term $n^*(c_d, r)$ represents the "effective" number of competitors, and it represents the lower bound of conduct. In other words, with a simple assumption on a minimum cost of servicing deposits, it is possible to derive a condition that bounds the conduct parameter.

To implement this framework, I utilize the pass-through estimates from the previous section, which relate equilibrium pricing to the Fed Funds rate. For quantities, I assume

that $D^*(r)$ is log-linear, and I run a similar specification to the pass-through analysis:

$$\Delta \ln(D_{ijmt}^*) = \alpha + \beta \Delta F F_t + \beta \Delta F F_t * X_{mt} + \delta_i + \lambda_j + \xi_m + \epsilon_{ijmt} \quad (6)$$

Here, I use deposit quantities from the FDIC's Summary of Deposits dataset from 1997 to 2008. For each branch i , I take the log yearly difference and project that onto the yearly change in the Fed Funds rate. I include fixed effects for each branch i , each bank j , and each county m .

Table 14 shows the regression results. A one percentage point increase in the Fed Funds rate increases deposit growth by about one percent. In column 2, increasing the number of banks decreases the growth of deposits, consistent with Drechsler et al. (2017). This effect is consistent across specifications.

For each county, I compute $\frac{dr_D^*}{dr}$ from the full specification of the pass-through estimates by county, using the Fed Funds coefficient and interactions. Because I estimate $\log(D^*)$, I similarly only need the coefficients on Fed Funds and the predicted interaction terms to compute $(\frac{D^*(r)}{dD^*(r)})$. Finally, I use the predicted interest rate for $r_D^*(r)$, assuming the Fed Funds rate is 5% (the level in 2005), the reserve requirement (α) is 0.1, and the marginal cost of deposits is 1 bps ($c_d = 0.01$).

For figure 6, I group counties into 10 buckets of HHI (from 0 to 0.1, 0.1 to 0.2, etc.). The effective number of competitors declines with HHI, and when HHI is near 1, the lower bound cannot reject that the county behaves as a monopolist (which is to be expected). Similarly, when HHI is very low, the county behaves as if there were at least 9 competitors. There is a slight increase when HHI is between 0.6 and 0.8, but there are also fewer counties in those buckets, which may reflect measurement error.

In figure 7, I group counties into 10 equally sized buckets based on the average number of banks in those counties. For counties with more than 5 banks, I can reject that those counties behave as monopolists, contrary to the extreme measured levels of concentration from the second-order analysis.

1.9 Conclusion

In this paper, I attempt to measure conduct in bank deposits, and compare those conduct estimates with Drechsler et al. (2017)’s use of HHIs. Unfortunately, I find that the pass-through framework fails to consistently produce plausible conduct parameters. As discussed above, there may be several data-related reasons for this departure from plausibility, mostly focused on mis-measurement of the demand for deposits. For example, I am only measuring one potential margin of substitution—households substituting liquid assets from deposits to stocks—but households may also display substantial inattention. Relatedly, banks may be competing on a bundle of goods, including CD’s and non-price characteristics (ATM network).

However, taking the relative variation of conduct across counties, I do replicate many of the findings from Drechsler et al. (2017), including the sensitivity of deposit spreads to the Fed Funds rate across different levels of competition. Interestingly though, my measurements of semi-elasticities are lower than theirs. Taking these estimates seriously, it would suggest that the deposit channel does not explain the full effects of monetary policy on the real economy. Speculatively, this “missing” impact of monetary policy might require some other arguments, including the financial accelerator (Bernanke et al. (1999)) or liquidity motivations (Kashyap & Stein (2000)).

Finally, I find that the lower bound approach does produce credible estimates of “effective competition”. While this approach does not produce point estimates of conduct, it does reject a variety of collusive models. In particular, I find that counties with at least 10 banks operate as if there were at least 6 competitors in the market. Insofar as this measure is correlated to HHI’s, it does suggest that HHI’s are at least a reasonable first-order approximation to competition.

While this current empirical investigation could not fully reconcile the pass-through and demand properties, an investigation with a more complete dataset on household holdings may be able to provide credible estimates. In particular, this pass-through approach to

market power is especially attractive in financial services markets, since prices are observed frequently, and it does not require *a priori* assumptions about conduct.

1.10 Tables

Table 1: For each SIPP year, there are multiple waves (ie. repeated surveys of the same household). Most of the asset surveys are consistently asked on waves 4 and 7, though that has changed for recent survey years. Note that the 1996 survey year includes the longest panel dimension, as there were budget cuts to the survey.

Year	Num HHs					
	Wave 3	Wave 4	Wave 6	Wave 7	Wave 9	Wave 12
1984		19,055		14,645		
1985	13,453			10,028		
1986		10,848		9,996		
1987		10,950		10,270		
1990		19,762		18,733		
1991		12,780		12,222		
1992		17,790		16,816		
1993		17,913		16,661		
1996	32,817		29,119		27,009	25,886
2001	26,646		25,150		23,760	
2004	38,099		34,987			

Table 2: This includes various data cleaning steps and filtering.

Waterfall: Filtering out households

	#HHs	%Dropped
Total # Households across all years:	495,395	
After dropping 1987 wave:	474,175	-4%
After dropping retirees:	366,721	-23%
After dropping self-employed:	327,237	-11%
Keep balanced panel:	270,666	-17%
After HPI merge:	265,189	-2%
After Net Worth restriction*:	55,092	-79%
<i>Total Attrition</i>	11%	

* Financial wealth $\geq 5K\$$ and net worth $\geq 5M\$$

Table 3: Summary statistics of SIPP data. All amounts adjusted by CPI (index=2000).

<i>Balance Sheet & Demographic Information</i>		
	<i>Mean</i>	<i>%Ownership</i>
Home Equity	\$ 93,813	86%
Vehicle Equity	\$ 10,161	96%
Own-Business Equity	\$ -	0%
Interest-bearing Bank Assets	\$ 24,543	98%
Bonds	\$ 7,670	17%
Stocks	\$ 40,333	57%
Rental Properties	\$ 23,908	24%
Other Assets	\$ 7,115	60%
IRA/Keogh	\$ 21,569	57%
Total Wealth	\$ 288,672	100%
Mortgage Debt	\$ 59,560	66%
Secured Debt	\$ 59,560	66%
Unsecured Debt	\$ 4,924	52%
Total Debt	\$ 65,542	73%
Net Worth	\$ 229,112	100%
Total Income	\$ 5,972	100%
Wage Income	\$ 5,256	93%
Property Income	\$ 287	99%
Transfer Income	\$ 6	2%
Other Income	\$ 423	33%
%White	93%	
Age, Head of HH	46	
%Elementary	2%	
%HS	23%	
%Some College	24%	
%College	26%	
%Post-Grad	24%	
%Married	72%	
N	55,092	

Table 4: Households in prior years transitioning to a different portfolio in subsequent years.

Prior Period	Current Period							
	<i>B</i>	<i>M</i>	<i>MB</i>	<i>MS</i>	<i>MSB</i>	<i>S</i>	<i>SB</i>	<i>Total</i>
Bonds Only	57%	4%	17%	4%	10%	3%	4%	70
Money Only	0%	84%	2%	13%	1%	0%	0%	10,740
Money & Bonds	1%	11%	63%	5%	19%	0%	0%	1,499
Money & Stocks	0%	9%	1%	85%	4%	1%	0%	13,964
Money, Stocks, & Bonds	0%	3%	7%	13%	75%	0%	1%	3,236
Stocks	0%	4%	0%	42%	1%	53%	1%	497
Stocks & Bonds	4%	1%	1%	6%	18%	10%	61%	102

Table 5: Summary Stats by Portfolio Choice

	<i>Money</i>	<i>Money £</i> <i>Bonds</i>	<i>Money £</i> <i>Stocks</i>	<i>Money, £</i> <i>Stocks, £</i> <i>Bonds</i>
Home Equity	\$ 82,513	\$ 86,046	\$ 94,366	\$ 109,738
Vehicle Equity	\$ 9,058	\$ 10,406	\$ 10,508	\$ 12,461
Own-Business Equity	\$ -	\$ -	\$ -	\$ -
Interest-bearing Bank Assets	\$ 26,746	\$ 22,056	\$ 22,509	\$ 27,525
Bonds	\$ -	\$ 39,530	\$ -	\$ 46,405
Stocks	\$ -	\$ -	\$ 66,277	\$ 87,275
Rental Properties	\$ 17,833	\$ 26,147	\$ 23,120	\$ 39,036
Other Assets	\$ 5,683	\$ 5,948	\$ 6,220	\$ 13,554
IRA/Keogh	\$ 14,257	\$ 18,548	\$ 23,410	\$ 32,286
Total Wealth	\$ 207,778	\$ 250,315	\$ 312,890	\$ 422,241
Mortgage Debt	\$ 51,695	\$ 41,666	\$ 66,480	\$ 53,962
Secured Debt	\$ 51,695	\$ 41,666	\$ 66,480	\$ 53,962
Unsecured Debt	\$ 4,200	\$ 2,810	\$ 5,498	\$ 4,106
Total Debt	\$ 56,181	\$ 39,077	\$ 74,227	\$ 54,946
Net Worth	\$ 156,083	\$ 208,648	\$ 246,410	\$ 368,279
Total Income	\$ 5,181	\$ 5,167	\$ 6,393	\$ 6,636
Wage Income	\$ 4,595	\$ 4,265	\$ 5,731	\$ 5,395
Property Income	\$ 153	\$ 433	\$ 269	\$ 740
Transfer Income	\$ 9	\$ 7	\$ 5	\$ 2
Other Income	\$ 425	\$ 462	\$ 388	\$ 498
Age, Head of HH	46	47	46	48
%White	91%	95%	93%	96%
%Elementary	3%	1%	1%	1%
%HS	31%	25%	20%	14%
%Some College	25%	21%	25%	18%
%College	22%	22%	29%	29%
%Post-Grad	19%	31%	25%	39%
Household Size	2.7	2.7	2.8	2.7
N	19,437	2,902	23,440	5,974

VARIABLES	(1) Ln(Share Savings) MS	(2) Ln(Share Savings) MSB	(3) Ln(Share Savings) MS	(4) Ln(Share Savings) MSB
Dep. Spread	-0.126*** (0.0176)	-0.127*** (0.0311)	-0.188*** (0.0401)	-0.435*** (0.0745)
Dep Spread, Sq			0.0101* (0.00597)	0.0469*** (0.0103)
Equity Return	-0.636*** (0.0551)	-0.389*** (0.126)	-0.606*** (0.0579)	-0.315** (0.127)
Bond Spread	-0.149*** (0.0196)	-0.159*** (0.0379)	-0.162*** (0.0212)	-0.239*** (0.0417)
HH Net Worth	-0.0133*** (0.000369)	-0.0105*** (0.000575)	-0.0133*** (0.000369)	-0.0106*** (0.000574)
HH Income	0.307*** (0.0230)	0.334*** (0.0447)	0.307*** (0.0230)	0.335*** (0.0446)
Age	0.00420*** (0.00103)	0.00211 (0.00209)	0.00421*** (0.00103)	0.00208 (0.00209)
White	-0.103*** (0.0381)	0.0667 (0.0930)	-0.103*** (0.0381)	0.0670 (0.0928)
Married	-0.0134** (0.00670)	-0.0229* (0.0130)	-0.0134** (0.00670)	-0.0230* (0.0130)
Edu., Elementary	0.0328 (0.103)	-0.248 (0.238)	0.0325 (0.103)	-0.239 (0.238)
Edu., HS	-0.0267 (0.0287)	-0.00450 (0.0648)	-0.0268 (0.0287)	-0.00148 (0.0647)
Edu., Some College	-0.0748*** (0.0260)	-0.0909 (0.0581)	-0.0747*** (0.0260)	-0.0831 (0.0580)
Edu., College	-0.0250 (0.0259)	-0.0506 (0.0469)	-0.0242 (0.0259)	-0.0427 (0.0468)
HH Size	-0.0328*** (0.00869)	-0.0238 (0.0185)	-0.0327*** (0.00869)	-0.0235 (0.0184)
Year	-0.0349*** (0.00201)	-0.0209*** (0.00425)	-0.0353*** (0.00203)	-0.0226*** (0.00426)
Constant	69.15*** (4.057)	40.63*** (8.513)	70.09*** (4.094)	44.61*** (8.544)
Observations	23,440	5,974	23,440	5,974
R-squared	0.080	0.073	0.080	0.076

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Linear Demand Regression

VARIABLES	(M) 1	(MS) 2	(MSB) 3
Dep. Spread		0.369*** (0.0290)	0.443*** (0.0445)
Equity Return		1.965*** (0.0901)	1.798*** (0.146)
Bond Spread		0.626*** (0.0314)	0.771*** (0.0495)
HH Net Worth		0.0257*** (0.000896)	0.0387*** (0.00104)
HH Income		0.178*** (0.0388)	0.376*** (0.0534)
Age		-0.0139*** (0.00147)	0.00689*** (0.00234)
White		0.327*** (0.0502)	0.694*** (0.0972)
Married		-0.0227** (0.00939)	0.0219 (0.0148)
Edu., Elementary		-1.430*** (0.121)	-2.153*** (0.231)
Edu., HS		-0.532*** (0.0401)	-1.242*** (0.0697)
Edu., Some College		-0.156*** (0.0382)	-0.552*** (0.0637)
Edu., College		-0.0218 (0.0398)	0.268*** (0.0563)
HH Size		-0.0132 (0.0125)	-0.0795*** (0.0205)
Year		0.0294*** (0.00313)	-0.112*** (0.00466)
Constant		-61.35*** (6.327)	217.0*** (9.368)
Observations	27,898	27,898	27,898

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Portfolio Selection Regression

VARIABLES	(1) <i>M</i>	(2) Anciliary, <i>M</i>	(3) <i>MS</i>	(4) Anciliary, <i>MS</i>	(5) <i>MSB</i>	(6) Anciliary, <i>MSB</i>
Dep. Spread	-0.00112 (0.00108)		-0.0414*** (0.00562)		-0.0346*** (0.00969)	
Equity Return	-0.00620 (0.00600)		-0.164*** (0.0194)		-0.00835 (0.0352)	
Bond Spread	-0.00178 (0.00182)		-0.0374*** (0.00666)		-0.0332*** (0.0114)	
HH Net Worth	1.23e-05 (4.04e-05)		-0.000792*** (0.000244)		-0.00201*** (0.000490)	
HH Income	-0.00176 (0.00125)		0.0519*** (0.00608)		0.00498 (0.0124)	
Age	4.21e-05 (4.16e-05)		0.000719** (0.000351)		-0.00280** (0.00117)	
White	0.000143 (0.000478)		0.00208 (0.0126)		-0.0174 (0.0319)	
Married	-8.17e-05 (6.45e-05)		0.000256 (0.00206)		-0.00839* (0.00469)	
Edu., Elementary	0.00157 (0.00266)		-0.0812** (0.0397)		0.0478 (0.0771)	
Edu., HS	-0.000186 (0.000861)		-0.0278** (0.0134)		0.0583* (0.0337)	
Edu., Some College	-0.000697* (0.000424)		-0.0186** (0.00900)		0.0324 (0.0247)	
Edu., College	0.00203 (0.00190)		0.0113 (0.00716)		-0.0349* (0.0190)	
HH Size	-3.17e-05 (8.27e-05)		-0.00557** (0.00275)		0.00307 (0.00557)	
Year	-0.000236 (0.000209)		-0.0142*** (0.00132)		0.0122* (0.00723)	
Sigma2		0.000744 (0.000953)		0.132*** (0.0194)		0.190* (0.0974)
rho1				-0.567*** (0.0873)		-0.612*** (0.0540)
rho2		-0.557** (0.241)				0.951*** (0.145)
rho3		0.485** (0.218)		0.530*** (0.121)		
Constant	1.471*** (0.417)		28.89*** (2.619)		-23.25 (14.17)	
Observations	.	.	.	.	.	.

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Table 8: Outcome Regression

Table 9: Summary statistics of interest rate data. The average spread can range from very low (nearly 50 bps) to very high (around 4 percent).

<i>Year</i>	<i>Source</i>	<i>Num Banks</i>	<i>Num Branches</i>	<i>Avg Spread, Savings</i>	<i>StDev in Spread</i>	<i>Fed Funds</i>	<i>Savings Rate</i>	Δ <i>ings</i>	<i>Sav-</i>	Δ <i>Funds</i>	<i>Fed</i>
1998	RateWatch	32	31	2.3%	0.7%	5.3%	2.9%	0.0%			-0.3%
1999	RateWatch	12343	5988	2.7%	0.7%	5.1%	2.3%	0.0%			0.2%
2000	RateWatch	12629	6245	3.9%	0.7%	6.3%	2.4%	0.0%			0.3%
2001	RateWatch	12665	6323	1.6%	1.2%	3.5%	1.9%	-0.2%			-1.1%
2002	RateWatch	12531	6364	0.4%	0.6%	1.6%	1.2%	-0.1%			-0.1%
2003	RateWatch	11711	6375	0.4%	0.4%	1.1%	0.7%	-0.1%			-0.1%
2004	RateWatch	11624	6509	0.8%	0.6%	1.5%	0.6%	0.0%			0.3%
2005	RateWatch	11677	6682	2.6%	0.7%	3.4%	0.7%	0.0%			0.5%
2006	RateWatch	11243	6804	4.2%	0.6%	5.0%	0.9%	0.0%			0.3%
2007	RateWatch	11189	6842	4.0%	0.7%	4.9%	0.9%	0.0%			-0.2%
2008	RateWatch	11227	6978	1.0%	1.0%	1.6%	0.6%	-0.1%			-1.0%

Table 10: Using RateWatch data, estimating price as a function of Fed Funds.

VARIABLES	(1) Δ Dep. Rates	(2) Δ Dep. Rates	(3) Δ Dep. Rates
Δ Fed Funds	0.203*** (0.000944)	-0.120** (0.0585)	0.604*** (0.0806)
Δ Fed Funds * $\log(\#Banks)$		0.0399*** (0.00276)	0.0427*** (0.00277)
Δ Fed Funds * $\log(Income)$		0.0319*** (0.00578)	-0.0409*** (0.00852)
Δ Fed Funds * $\log(Dividend\ Inc)$		-0.0112*** (0.00123)	-0.0155*** (0.00130)
Δ Fed Funds * Age			-0.183*** (0.0306)
Δ Fed Funds * HS Edu			0.104*** (0.0208)
Δ Fed Funds * College			0.170*** (0.0186)
Constant	-0.0319*** (0.000630)	-0.0318*** (0.000630)	-0.0318*** (0.000629)
Observations	291,825	291,797	291,797
R-squared	0.161	0.162	0.163

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 11: Pass-through of Fed Funds using CD rates from RateWatch data.

VARIABLES	(1) Δ Dep. Rates	(2) Δ Dep. Rates	(3) Δ Dep. Rates
Δ Fed Funds	0.462*** (0.00101)	0.970*** (0.0633)	1.158*** (0.0866)
Δ Fed Funds * $\log(\#Banks)$		0.0484*** (0.00297)	0.0483*** (0.00299)
Δ Fed Funds * $\log(Income)$		-0.0468*** (0.00624)	-0.0462*** (0.00915)
Δ Fed Funds * $\log(Dividend\ Inc)$		-0.0118*** (0.00132)	-0.0182*** (0.00140)
Δ Fed Funds * Age			-0.216*** (0.0328)
Δ Fed Funds * HS Edu			-0.179*** (0.0223)
Δ Fed Funds * College			0.216*** (0.0198)
Constant	-0.00675*** (0.000675)	-0.00668*** (0.000675)	-0.00663*** (0.000674)
Observations	323,081	323,053	323,053
R-squared	0.412	0.413	0.413

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 12: Using other measures of conduct, impact of competition on spreads. $\theta_{MM, ll}$ is conduct measured using money market rates and log-linear demand. $\theta_{CD, ll}$ is conduct measured using CD rates and log-linear demand. $\theta_{MM, dc}$ is conduct measured using money market rates and discrete-continuous demand. $\theta_{CD, dc}$ is conduct measured using CD rates and discrete-continuous demand.

VARIABLES	(1) Δ spread	(2) Δ spread	(3) Δ spread	(4) Δ spread	(5) Δ spread
Δ FF*HHI	0.273*** (0.00699)				
Δ FF* $\theta_{MM, ll}$		0.156*** (0.00206)			
Δ FF* $\theta_{CD, ll}$			0.952*** (0.0103)		
Δ FF* $\theta_{MM, dc}$				0.0955*** (0.00142)	
Δ FF* $\theta_{CD, dc}$					0.345*** (0.00618)
Constant	-0.0657*** (0.000635)	-0.0316*** (0.000819)	-0.00319*** (0.000970)	-0.0296*** (0.000892)	-0.0400*** (0.000858)
Observations	287,422	287,422	287,422	287,378	287,378
R-squared	0.752	0.755	0.758	0.754	0.753

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 13: Using other measures of conduct, impact of competition on $\log(D)$. $\theta_{MM,ll}$ is conduct measured using money market rates and log-linear demand. $\theta_{CD,ll}$ is conduct measured using CD rates and log-linear demand. $\theta_{MM,dc}$ is conduct measured using money market rates and discrete-continuous demand. $\theta_{CD,dc}$ is conduct measured using CD rates and discrete-continuous demand.

VARIABLES	(1) $\Delta \ln(D)$	(2) $\Delta \ln(D)$	(3) $\Delta \ln(D)$	(4) $\Delta \ln(D)$	(5) $\Delta \ln(D)$
$\Delta FF*HHI$	-2.299*** (0.190)				
$\Delta FF*\theta_{MM,ll}$		-0.692*** (0.0445)			
$\Delta FF*\theta_{CD,ll}$			-2.811*** (0.245)		
$\Delta FF*\theta_{MM,dc}$				-0.232*** (0.0317)	
$\Delta FF*\theta_{CD,dc}$					-0.0498 (0.123)
Constant	8.191*** (0.0272)	8.026*** (0.0308)	7.997*** (0.0359)	8.128*** (0.0333)	8.272*** (0.0314)
Observations	795,531	795,531	795,531	795,477	795,477
R-squared	0.288	0.288	0.288	0.288	0.288

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 14: Sullivan regression on quantity using the FDIC Summary of Deposits.

VARIABLES	(1) $\Delta \ln(\text{Deposits})$	(2) $\Delta \ln(\text{Deposits})$	(3) $\Delta \ln(\text{Deposits})$
Δ Fed Funds	0.00951*** (0.000443)	0.100*** (0.0266)	0.165*** (0.0308)
Δ Fed Funds * $\log(\#\text{Banks})$		-0.00693*** (0.00128)	-0.00914*** (0.00129)
Δ Fed Funds * $\log(\text{Income})$		-0.0127*** (0.00269)	-0.0204*** (0.00323)
Δ Fed Funds * $\log(\text{Dividend Inc})$		0.00585*** (0.000630)	0.00739*** (0.000666)
Δ Fed Funds * Age			-0.117*** (0.0120)
Δ Fed Funds * HS Edu			0.0390*** (0.00703)
Δ Fed Funds * College			-0.0366*** (0.00722)
Constant	0.0839*** (8.20e-05)	0.0836*** (7.64e-05)	0.0835*** (7.13e-05)
Observations	887,139	886,924	886,924
R-squared	0.255	0.256	0.257

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

1.11 Figures

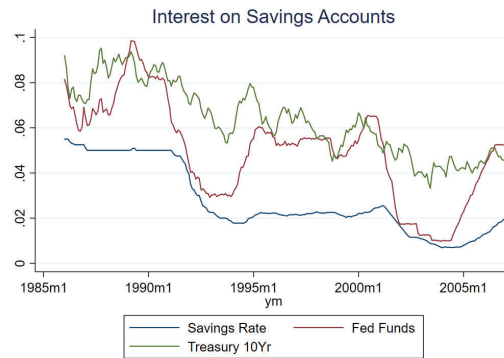


Figure 1: Share of Households with no liquid assets is quite high

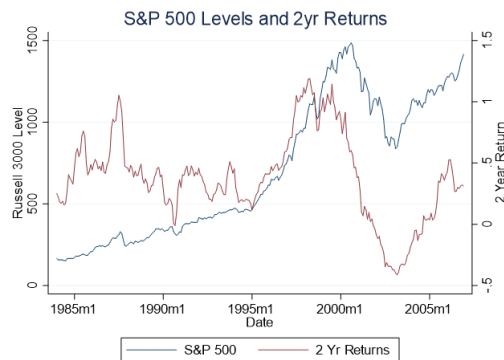


Figure 2: Share of Households with no liquid assets is quite high

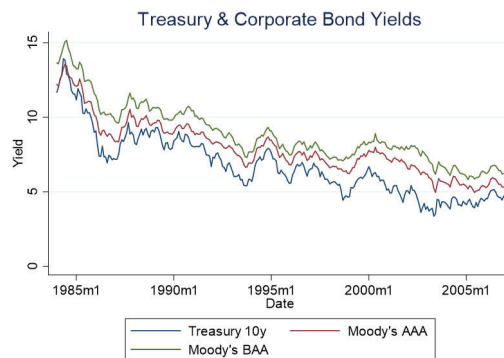


Figure 3: Share of Households with no liquid assets is quite high

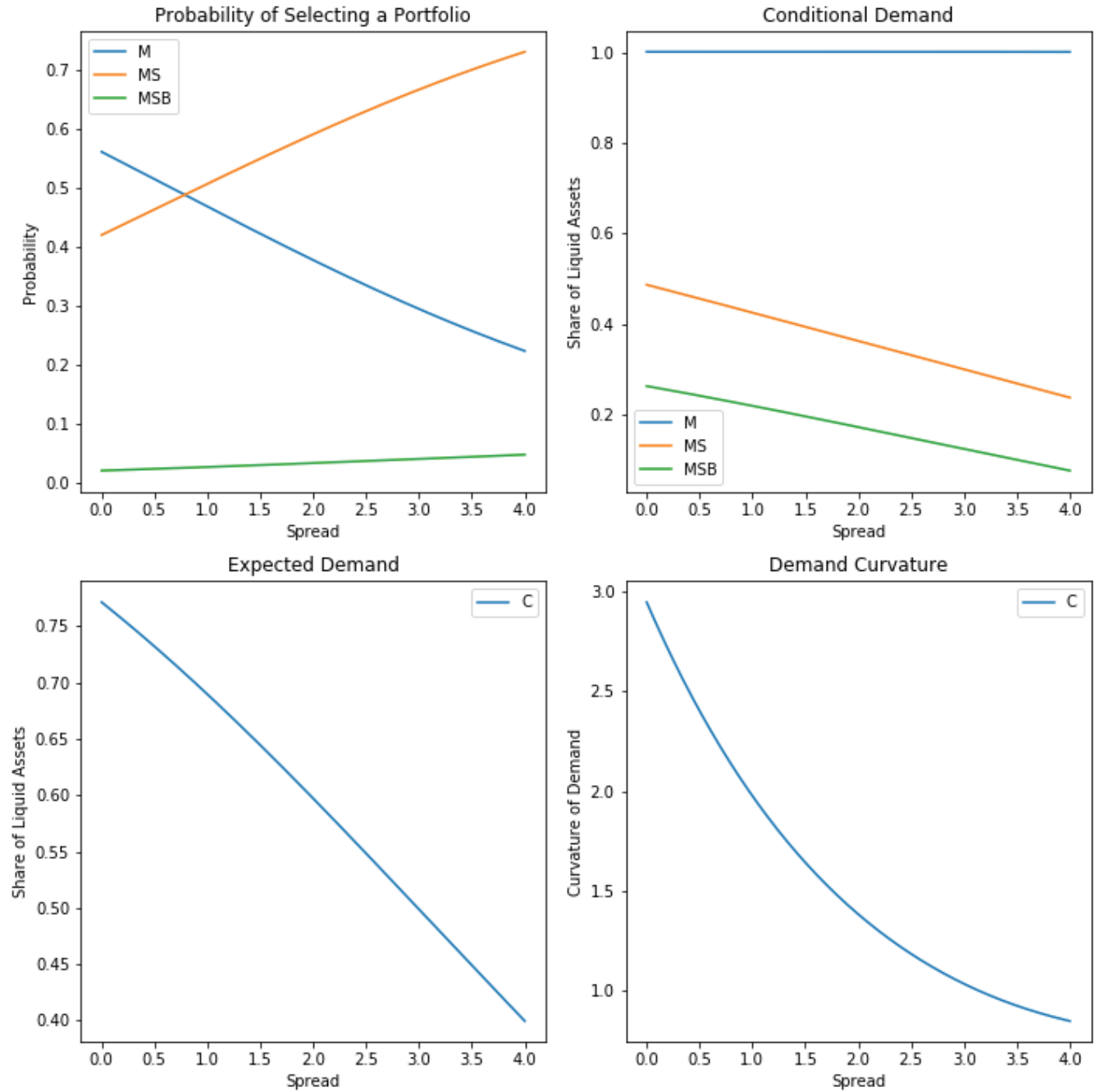


Figure 4: Estimated Demand and Curvature

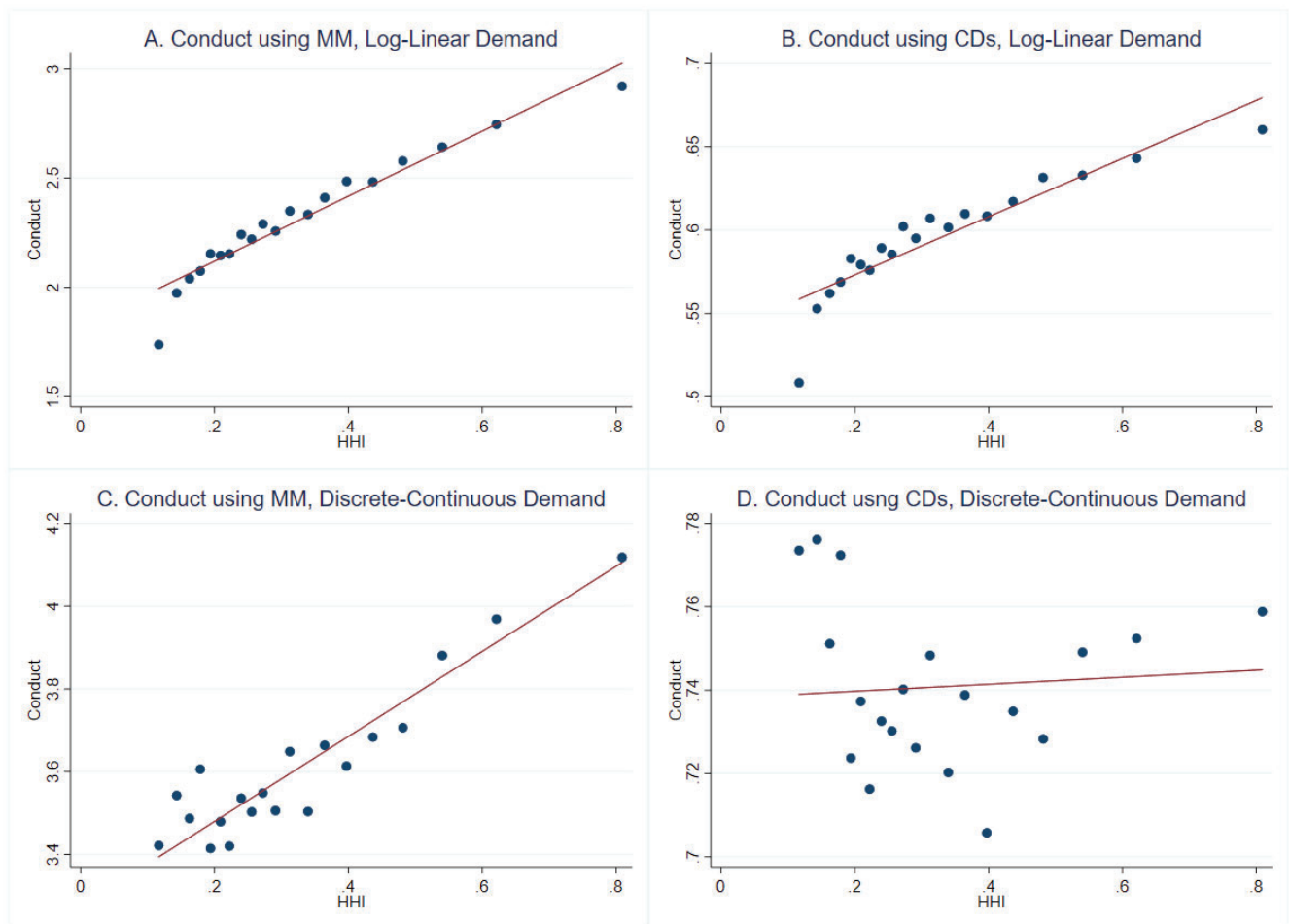


Figure 5: Estimated Conduct using Various Specifications.

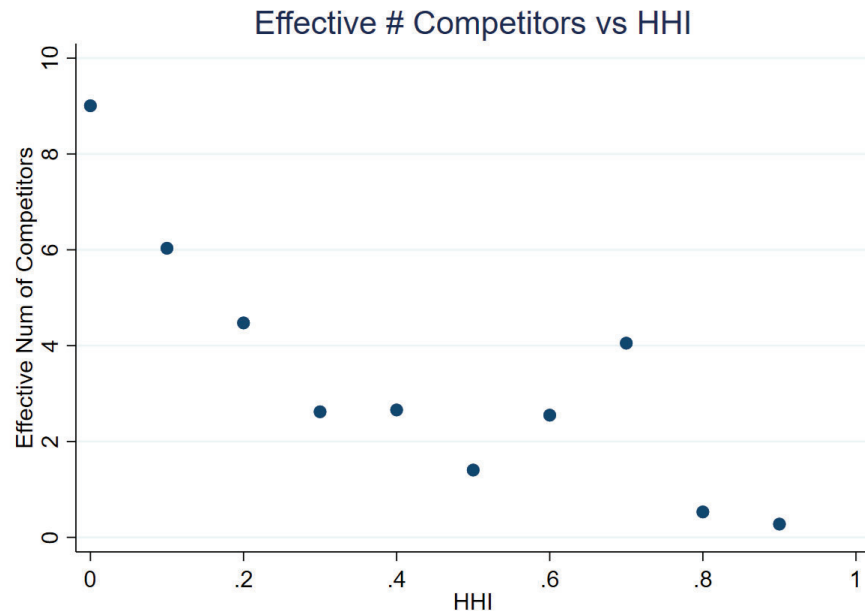


Figure 6: The correlation of HHI and the effective number of banks is -0.22 .

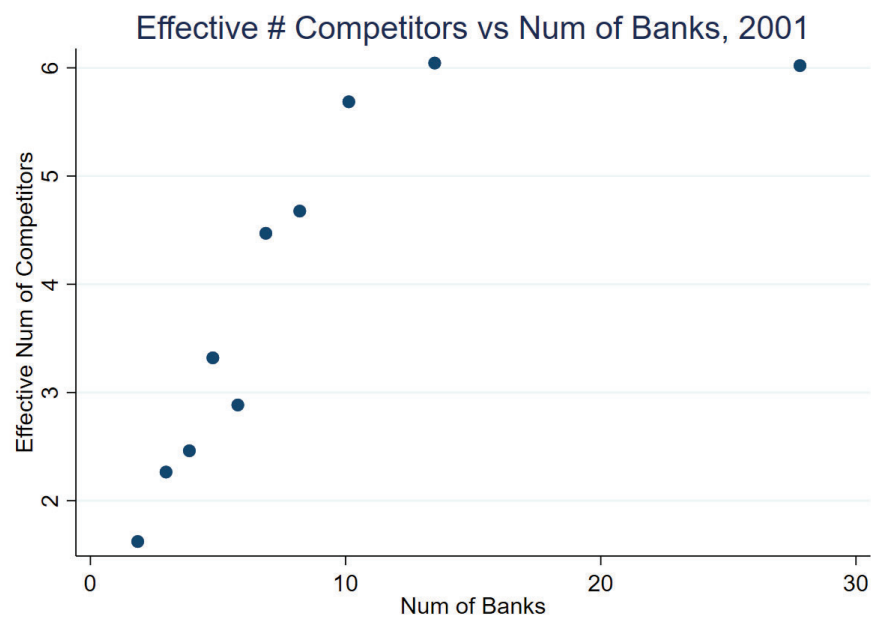


Figure 7: The correlation between the effective number of banks and the true number of banks by county is 0.25 .

1.12 Appendix A. Calculating Demand and Elasticity

Consider the demand for money assets for a household of type MB :

$$y_{i,MB} = x_i \beta_{MB} + \sigma \sum_{j \in \{M, MS, MSB\}} \rho_j \underbrace{\left(\frac{P_{ij} \ln(P_{ij})}{1 - P_{ij}} + \ln(P_{i,MB}) \right)}_{A_j} + w_{i,MB} \quad (7)$$

The expected demand is given by:

$$E[y_i] = E[y_i|M]P_{i,M} + E[y_i|MB]P_{i,MB} + E[y_i|MS]P_{i,MS} + E[y_i|MSB]P_{i,MSB} \quad (8)$$

Note that to calculate the aggregate demand:

$$E[y] = \int_i \sum_{j \in \{M, MB, MS, MSB\}} E[y_i|j] P_{i,j} di \quad (9)$$

1.12.1 Calculating First Derivative

To compute the derivative of conditional demand ($\frac{\partial y_{i,MB}}{\partial x_i}$), notice that the derivative of A_j with respect to the spread x_i is:

$$\frac{(1 - P_j) \left[\frac{\partial P_j}{\partial x_j} \ln(P_j) + \frac{\partial P_j}{\partial x_j} \right] + P_j \ln(P_j) \frac{\partial P_j}{\partial x_j}}{(1 - P_j)^2} = \frac{\frac{\partial P_j}{\partial x_j} [\ln(P_j) + 1 - P_j]}{(1 - P_j)^2} = \frac{\partial A_j}{\partial x_i}$$

Hence:

$$\frac{\partial y_i}{\partial x_i} = \beta_{MB} + \sigma \sum_{j \in \{M, MS, MSB\}} \rho_j (A'_j + \frac{1}{P_{MB}} P'_{MB}) \quad (10)$$

The first derivative of expected demand is given by:

$$\frac{\partial E[y]}{\partial x_i} = \sum_{j \in \{M, MB, MS, MSB\}} \left(\frac{\partial E[y|j]}{\partial x_i} P[j] + E[y|j] \frac{\partial P[j]}{\partial x_i} \right) \quad (11)$$

1.12.2 Calculating Second Derivative

Consider now the second derivative:

$$\frac{\partial^2 y_i}{\partial x_i^2} = \sigma \sum_{j \in \{M, MS, MSB\}} \rho_j \left(A''_j - \frac{1}{P_{MB}^2} (P'_{MB})^2 + \frac{1}{P_{MB}} P''_{MB} \right) \quad (12)$$

where

$$A_j'' = \frac{(1 - P_j)^2 \left[\frac{\partial^2 P_j}{\partial x_i^2} (\ln(P_j) + 1 - P_j) + \frac{\partial P_j}{\partial x_i} \left(\frac{1}{P_j} \frac{\partial P_j}{\partial x_i} - \frac{\partial P_j}{\partial x_i} \right) \right] - \frac{\partial P_j}{\partial x_j} [\ln(P_j) + 1 - P_j] (2)(1 - P_j) \frac{\partial P_j}{\partial x_j}}{(1 - P_j)^4}$$

and P_j' and P_j'' can be expressed more conveniently.¹⁹

To compute the second derivative of demand, re-write

$$\frac{\partial^2 E[y]}{\partial x_i^2} = \sum_{j \in \{M, MB, MS, MSB\}} \left(\frac{\partial^2 E[y|j]}{\partial x_i^2} P_j + 2 \frac{\partial E[y|j]}{\partial x_i} P_j' + E[y|j] P_j'' \right) \quad (13)$$

1.13 Appendix B. Pass-Through Framework

Define now the spread between the market rate and the deposit rate that is paid to households $s = f - r^D$. Using the terminology of Weyl & Fabinger (2013), $\rho = \frac{\partial s}{\partial f}$ is the pass-through of the market interest rate with respect to the spread s . In the model of Weyl & Fabinger (2013), the firms optimize according to a Lerner rule, where

$$\frac{p - mc - t}{p} \epsilon_D = \theta \quad (14)$$

In a symmetric equilibrium, quantity must be chosen such that the price minus the conduct parameter times the marginal surplus ($ms(q) \equiv p'q$)

$$p(q) - \theta ms(q) = mc(q) + t \quad (15)$$

Differentiating with respect to t yields (noting that $\rho = p' \frac{dq}{dt}$):

$$p' \frac{dq}{dt} - \left[\frac{d\theta}{dq} \frac{dq}{dt} ms - \theta ms' \frac{dq}{dt} \right] = mc' \frac{dq}{dt} + 1 \Rightarrow \rho - \rho \frac{d\theta}{dq} \frac{ms}{p'} - \theta \frac{ms'}{p'} \rho = \frac{mc'}{p'} \rho + 1$$

Hence, noting that $ms = -p'q$, we get

$$\rho = \frac{1}{1 - \frac{d\theta}{dq} \frac{ms}{p'} - \theta \frac{ms'}{p'} \frac{p \cdot q \cdot ms}{p \cdot q \cdot ms} - \frac{mc'}{p'} \frac{p \cdot q \cdot mc}{p \cdot q \cdot mc}} \Rightarrow \rho = \frac{1}{1 + \frac{d\theta}{dq} q + \theta \frac{\epsilon_D}{\epsilon_{ms}} \frac{ms}{p} + \frac{\epsilon_D}{\epsilon_S} \frac{mc}{p}}$$

¹⁹Note that: $P_j' = \frac{(\sum_k \exp(x_k \beta_k)) \beta_j \exp(x_j \beta_j) - \exp(x_j \beta_j) (\sum_k \beta_k \exp(x_k \beta_k))}{(\sum_k \exp(x_k \beta_k))^2} = P_j [\beta_j - \sum_k \beta_k P_k]$

Furthermore: $P_j'' = \frac{\partial P_j}{\partial x_i} [\beta_j - \sum_k \beta_k P_k] - P_j \sum_k \beta_k \frac{\partial P_k}{\partial x_i}$

Finally, $\frac{p-mc}{p} = \frac{\theta}{\epsilon_D}$, and writing $\epsilon_\theta = \frac{d\theta}{dq} \frac{q}{\theta}$, we finally get:

$$\rho = \frac{1}{1 + \frac{\theta}{\epsilon_\theta} + \frac{\epsilon_D - \theta}{\epsilon_S} + \frac{\theta}{\epsilon_{ms}}} \quad (16)$$

There are several important quantities here: first, note that

$$(\log D)' = \frac{D'}{D} = \frac{1}{p'q} = -\frac{1}{ms} \Rightarrow (\log D)'' = \frac{ms'}{ms^2} \frac{1}{p'} = -\frac{1}{\epsilon_{ms}} \frac{1}{ms} \left(-\frac{1}{p'q}\right) = -\frac{1}{\epsilon_{ms}} \frac{1}{ms^2}$$

Hence, the second derivative of the log-curvature of demand is related to the elasticity of marginal surplus. One can also simply re-write the elasticity in terms of the inverse demand function: $\frac{1}{\epsilon_{ms}} = \frac{ms'q}{ms} = \frac{(p''q+p')q}{p'q} = 1 + \frac{p''q}{p'}$. Note that a linear demand curve thus implies that $\frac{1}{\epsilon_{ms}} = 1$.

Second, the term $\frac{\theta}{\epsilon_\theta}$ represents the change in market power as quantity changes. Weyl & Fabinger (2013) notes that in many canonical models, this quantity is 0 (it is 0, for example, in the Dixit-Stiglitz model). Importantly, it is not 0 in the standard logit demand-differentiated goods with Nash-Bertrand firms model.²⁰

The elasticity of supply can reasonably be assumed to be 0 in this model, since the vast majority of marginal costs for deposits will be the interest rate a bank pays. Hence, we can write the simple relation:

$$\theta = \frac{1 - \rho}{\rho} \epsilon_{ms}$$

Finally, one interpretation of θ has been that $\theta = 1$ corresponds to monopoly, while $\theta = 0$ corresponds to perfect competition. Hence, if demand is linear, then market power is a simple function of pass-through: if pass-through is low (in fact, below $\frac{1}{2}$), then market power is quite high.

²⁰If demand is constant elasticity of substitution (e.g. $D(p) = kp^{-r}$), then $\epsilon_{ms} = -r$.

2 Chapter 2. Subsidies for Vaccine Development

2.1 Introduction

Despite the vast burden of infectious disease²¹, private drug development in this area lags other therapeutic areas such as cancer²². Vaccination provides a spillover benefit—if more people get vaccinated, there’s less impetus for an individual to value their own vaccination, so that a monopolist selling a vaccine cannot fully internalize all of the value (Goodkin-Gold et al. (2020))—which the academic literature suggests is the critical externality limiting investment. While Gardasil²³ and Prevnar²⁴ are rare exceptions, there are a number of commercial failures where companies with novel anti-infective agents have gone bankrupt *after* FDA approval²⁵.

To spur private investment, Kremer (2000) has argued that the best approach is for governments to pre-commit to bulk purchases of vaccines if they meet certain efficacy and safety specifications, especially for tropical diseases where there is virtually no private investment (the “pull” method). While there may be an incentive to renege upon a successful vaccine, Kremer argues that pre-committed bulk purchases is more efficient than directly subsidizing private drug development (the “push” method), given the political constraints of something akin to industrial policy (ie. regulatory capture) and the difficulty of monitoring. He fur-

²¹While the current COVID19 pandemic has been extremely costly, endemic infectious diseases (HIV/AIDS, pneumonia) continue to result in premature death and poor quality of life. As Fan et al. (2016) estimates, pandemic influenza is a persistent and regular risk, with expected costs estimated at 0.7% of global income.

²²Public investment in infectious disease primarily funds pre-clinical (58%) and public health (28%) research Head et al. (2020), while most funding for clinical trials comes from private investment.

²³Gardasil is the HPV-vaccine developed by Merck. While many forms of HPV are asymptomatic, HPV types 16 and 18 cause 70% of cervical cancers. Unlike most vaccines, Gardasil generated around \$3.7 billion in 2019 for Merck, though development of the vaccine took more than 20 years and cost an estimated \$250 million dollars.

²⁴The vaccine for pneumonia, recommended for elderly patients.

²⁵For example, Achaogen declared bankruptcy in 2019 after successfully creating a novel antibiotic, plazomicin, for multidrug-resistant Enterobacteriaceae infections, despite market sales projections of around half a billion dollars. While there are multiple proximate causes for Achaogen’s failure (one industry participant, the CEO of a different anti-infectives company, has argued that a major reason for Achaogen’s failure is in the perverse hospital reimbursement model), fundamentally, the demand for Achaogen’s drug was insufficient at a price that would compensate investors for the R&D investment made.

ther argues that private financial markets have developed so that the marginal benefit of subsidizing biotech and pharmaceutical research is limited.

In contrast, I argue that innovation in pharmaceuticals is now powered by small biotech firms, and these these biotech firms are highly sensitive to external financing conditions. For example, these small biotechnology firms have been responsible for a substantial share of therapeutic innovation: half of newly approved drugs in 2018 came from biotech companies rather than large pharma companies. Currently, there are several Coronavirus vaccines in development from small biotech companies (Novavax, Inovio, Moderna, Vir, BioNTech), and the development of some vaccines from large pharma companies are partnered with smaller biotech peers (e.g. while Pfizer is partnered with BioNTech, Pfizer is responsible for manufacturing BioNTech's vaccine candidate).

Further, small biotech firms are more likely to face external financing constraints. Half of the publicly listed biotech companies today IPO'd less than 3 years ago, so they generally do not have any approved products. Thus, they generate no internal cash flow, and they require enormous amounts of capital to get a drug candidate approved. Despite their importance in drug development, these companies have an extremely high cost of capital²⁶, which is neither surprising nor controversial. The probability of failure is high: while most development candidates fail the screening and pre-clinical phase, drugs that enter Phase 1 clinical trials have on average only a 10% rate of eventual FDA approval. A large corpus of empirical literature confirms this "external finance" wedge on investment (Brown et al. (2009) and Howell (2017)).

I argue that, in addition to subsidizing vaccine purchases, there is a strong rationale for the government to subsidize the cost of vaccine development, given the numerous, compounding market failures. Using Goodkin-Gold et al. (2020)'s framework, I examine the welfare value of incrementally better vaccines. Interestingly, welfare, without any subsidies,

²⁶For example, the XBI index (a value-weighted index of small- and mid-cap biotech stocks) has a market beta between 1.5 and 2. In contrast, an index of broader healthcare (insurance companies, large pharma), has a market beta closer to 1.

increases dramatically when a vaccine has low-to-moderate efficacy; however, when a vaccine is very efficacious, welfare rises slowly. A product subsidy can alleviate the externality from a vaccine, but at a substantial cost.

Next, I consider a model of financing constraints motivated by moral hazard, where a firm can choose between private consumption or true investment (Holmstrom & Tirole (1997)). First, I show that even without a financing constraint, a monopolist does not invest at social optimum. Second, if firms are constrained, then even a small amount of external financing from the government can increase investment to the unconstrained optimum²⁷. My argument for government subsidies does not rely on any special monitoring by the government: the subsidies just provide a cash-on-hand endowment, which credibly commits management to raise additional capital from the private market.

While those theoretical contributions provide some rationale for government R&D subsidies, I argue that a good policy framework is for the government to mimic existing non-dilutive capital sources, ie. to mimic joint ventures between biotechs and large pharma partners. We have a database of joint venture partnerships (many of these partnerships are unredacted because our database vendor was able to access them through FOIA), and their deal terms (an upfront payment, followed by milestone payments, and a subsequent division of revenue either as a profit split or a royalty stream to the licensee). I provide some brief summary statistics that describes reasonable deal terms, and we generate a framework for additional government capital.

This policy paper is organized as follows: section 2.2 describes the external financial frictions for biotech firms; section 2.3 describes the welfare returns to increasing vaccine efficacy; section 2.4 describes the financing constraints and calibrated levels of investment; and section 2.5 illustrates principles for designing joint private-public partnerships.

²⁷There are several strands of literature discussing innovation. For example, there are games based on contest success functions (Tullock (2001) and Skaperdas (1996)), where players win a prize based on a "contest success function". There are also market structure and R&D games that focus on the timing of experiments (Loury (1979) and Lee & Wilde (1980) and Taylor (1995)). These games have a structure similar to an all-pay auction (Baye & Hoppe (2003)). My model focuses on a simple monopolist conducting research.

2.2 Evidence for External Financing Costs

With perfect capital markets, the external financing of firms should be related to the marginal q , or the marginal market value of capital should be equal to the marginal cost of capital (Hayashi (1982)). However, substantial evidence²⁸ suggests that firms' financing decisions do not correspond to this perfect capital market benchmark. Indeed, several indices have been suggested to identify firms with financial constraints (Whited & Wu (2006), Kaplan & Zingales (1997), FAZZARI et al. (1988)), focusing on proxies like low cash flow and high debt loads for financial constraints.

There are two groups of stylized facts that characterize biotech financing. First, small biotech firms are primarily R&D organizations, with almost no cash flow and high capital needs. Second, public-market financing in the biotech sector is highly pro-cyclical, correlating with general equity market conditions.

2.2.1 Stylized Fact 1: Biotech Balance Sheets

Biotech firms are uniquely susceptible to external financing constraints, since they generally have no FDA-licensed products to sell, and they require an enormous amount of capital. For example, the median cost of developing a new drug, including accounting for failures, is estimated to be nearly a billion dollars (Wouters et al. (2020)). Table 15 displays the revenue from all publicly US-listed companies with market capitalizations between \$100M and \$10B on Jan 1, 2020²⁹. In this sample, the median biotech firm recorded about \$3 million revenue, and that small amount is typically the result of licensing agreements to larger pharma/biotech partners that are booked as revenue.

Most of the external capital raised is through follow-on equity issuance in equity capital markets, and not through debt. Strikingly, table 16 shows that the average debt on

²⁸See FAZZARI et al. (1988) for reduced form regressions, and Hennessy & Whited (2007) for evidence from a structural estimate.

²⁹I exclude very small firms because they are close to de-listing and represent idiosyncratic situations. I also exclude large firms, since these may be large biotech companies (ie. Amgen) that have substantial sales and are not expected to be cash-constrained.

the balance sheet is limited. Consistent with theories of debt based on collateral values, the assets of most biotech firms are intangible (and mostly human capital). Interestingly, most biotech companies, especially those without marketed products, do not go into formal bankruptcy procedures—instead, most companies become public shells that reverse merge into other private biotech companies trying to become public.

Finally, because biotech assets are binary (the drug is approved or not) and the development process is risky, the cost of capital for biotech firms is extremely high. One method of calculating the cost of equity capital is through the capital asset pricing model (CAPM), an equilibrium model of asset pricing which prices assets according to their relative riskiness. Table 17 shows that biotech stocks have higher betas than other sectors.

2.2.2 Stylized Fact 2: Biotech Financing is Dynamic and Pro-Cyclical

The biotech industry has grown substantially since the first biotech public listings in the 1980s. Generally, these companies are incubated through specialized biotech venture capital firms. Once the companies have demonstrated enough data through either clinical or pre-clinical evidence, they often have capital needs that private market investors, even through a syndicate, cannot fulfill, and so they IPO.

First, figure 8 shows IPO issuance of new biotech stocks³⁰. While there has been an increasing secular trend of new listings since 2010, not surprisingly, IPO activity tends to be flat when equity markets (as proxied by the S&P500 index) decline, but is very rapid when equity markets are elevated. Ideally, capital is injected into these companies in a staggered fashion throughout drug development. For example, an IPO financing is sufficient to cover a drug’s development until a large milestone (e.g. a successful clinical trial) is completed, upon which the company can raise additional equity capital through a follow-on financing at a higher valuation. This process raises sufficient capital without excessive dilution of existing shareholders (including the management team).

³⁰This does not include reverse mergers or SPACs.

However, in practice, clinical trials may fail, or development can get delayed for a variety of idiosyncratic events. As a result, biotech firms can also be very opportunistic when issuing follow-on financing (for example, when equity prices are high). Figure 9 shows follow-on financing all US-listed biotech companies. Like IPOs, these financing events tend to be limited when equity markets are low, but flood in when markets become buoyant. Alternatively, biotech companies can finance themselves through licensing deals that often involve ex-US commercial rights, in exchange for upfront cash to the biotech firm, and downstream royalties conditional on drug approval and sales in those territories.

While the industry trend is pro-cyclical, this fact cannot distinguish between capital-demand and capital-supply: for example, opportunistic managers may wait for high equity values in order to issue follow on capital to limit dilution. However, the evidence from the GFC natural experiment suggests that capital supply is an important constraint in drug development. I use a large financial shock, the Great Financial Crisis (GFC), to test the hypothesis of external financing, comparing large firms that generate internal cash versus small firms that do not generate cash-flow.

2.2.3 Financial Crisis as Evidence of External Frictions

To test for financial frictions, I use the Great Financial Crisis as a natural experiment to examine changes in real R&D investment. In particular, the GFC was an exogenously timed shock to credit markets, and also resulted in a substantial reduction in equity valuations. Figure 10a shows the issuance of additional equity offerings during the financial crisis, with the market performance of the S&P index and the XBI, a biotech exchange-traded fund (an index of biotech stocks). Panel 10b describes the median gross spread of follow-on issuances, which is a measure of the underwriting banks' discount, where a higher spread indicates a costlier issuance (to the company). While the standard spread is about 6%, the spread declines around the financial crisis, Panel 10c shows that the reason for the decline was a difference in the composition of companies accessing external capital. These results suggest

that, for small biotech companies, the window for external capital was effectively closed.

Because biotech financing is dynamic and biotech firms prefer to keep a buffer stock of cash (about 2 years of expected operating expenditures), the GFC shock is an experiment of financial constraints for firms that otherwise anticipated the ability to raise capital, but were not able to do so in late 2008 and early 2009. In particular, Panel 10d describes the relationship between cash and operating spending in 2008 and 2009. The vertical axis is the ratio of cash on the balance sheet as of the end of the fourth quarter of 2008 and the total operating expenditures in 2008. This ratio can be interpreted as the expected years of cash remaining, assuming a constant rate of expenditures and no influx of capital (in practice, this is often referred to as the "burn" rate). The horizontal axis captures the change in expenditures between cumulative 2009 and cumulative 2008. Notice that for companies that are less than a billion in market capitalization ("small"), cash remaining has a strong relationship with operating expenditures³¹.

I test this relationship using a difference-in-difference estimator, testing whether operating spending varies for those companies that happened to be constrained as of the fourth quarter of 2008. These constrained companies are tested against small companies, and large pharma companies that generate substantial cash flow and would not be expected to be constrained³². One might be concerned that any analysis of external financing constraints is actually identifying companies' assets where the perceived probability of drug approval has actually gone down. However, because the GFC was large and unexpected (as measured by stock prices), I argue that this shock was orthogonal to companies' normal investment process. Additionally, one might be concerned that the GFC shock was not just to the supply of capital, but also to the underlying demand for pharmaceuticals. However, the demand for pharmaceuticals is generally regarded as inelastic, and the comparison group includes large pharma companies, who did not decrease R&D expenditures (on average) and were actually

³¹Each of cash remaining corresponds to an increase in operating expenditures of 12% ($p < 0.001$). The regressions for other market capitalization groups is not significant.

³²In fact, there were three large acquisitions in 2009: Pfizer-Wyeth (\$68bn), Roche-Genentech (\$47bn), and Merck-Schering-Plough (\$41bn).

selling drugs at the time.

The data for this analysis is from all US-listed biotech and pharma companies from 2007 to 2010. Unfortunately, data for R&D specific expenses are limited prior to 2009; this is not a substantial concern for small biotech firms, since total expenses are almost entirely R&D expenses, though it may cloud the interpretation for large pharma companies whose expenses also include large sales forces.

For the difference-in-difference estimators, I employ the following regression model:

$$\ln(\text{spending}_{ft}) - \ln(\text{spending}_{2008}) = \tau_t 1_{\text{constrained}} + FE_f + FE_t + \epsilon_{it} \quad (17)$$

where t is measured in quarters, and f is a fixed effect at the firm level. Table 18 shows the difference-in-difference estimate. While the effect is not statistically significant, it shows that R&D spending declined by about 4%, on average, for firms that were constrained.

2.3 Product Market for Vaccines: Epidemiology Model

I begin with an epidemiological model and consumer valuations. In particular, vaccination is a public good, so that the marginal private valuation differs from the social valuation, motivating government intervention.

2.3.1 SIR Models

The susceptible-infectious-resistant (SIR) is a canonical, epidemiological model of infection for a non-fatal, chronic disease. In the model, individuals are born and die at a rate γ , regardless of infection status, so that life expectancy is $\frac{1}{\gamma}$. Individuals can be in three states—vaccinated, susceptible, and infected (or $V(t)$, $S(t)$, and $I(t)$, respectively). Once infected, the individual stays infected and can infect susceptible individuals, but cannot infect those that have been successfully vaccinated. Finally, the total mass of the population is normalized to one (and stays constant):

$$V(t) + I(t) + S(t) = 1 \quad (18)$$

Individuals at birth have one shot of being vaccinated. The probability of a successful vaccination is θ , so that the change in the vaccinated population is:

$$\dot{V}(t) = \theta Q(t) - \gamma V(t) \quad (19)$$

where $Q(t)$ as the mass of newborns who are vaccinated. Note that, necessarily, $Q(t) < \gamma$, the rate of new births. In words, the growth of the vaccinated population increases with the inflow of successfully vaccinated individuals, less the death rate of vaccinated individuals.

The infected population increases with the probability an infected individual interacts with a susceptible individual. For a given susceptible individual, the probability of contact is given by $\beta I(t)$, where β is a parameter that encompasses the conditional probability of infection given contact. Hence, the infectious population increases with $\beta I(t)S(t)$. Because this is a chronic and non-lethal disease, the infectious population decreases with the rate of death $\gamma I(t)$, so that the infectious population increases at the given rate:

$$\dot{I}(t) = \beta I(t)S(t) - \gamma I(t) \Rightarrow \dot{I}(t) = R_0 \gamma I(t)S(t) - \gamma I(t) \quad (20)$$

In the epidemiological literature, it is common to report the "infectiousness" of a disease as $R_0 = \frac{\beta}{\gamma}$, which is the ratio of how many susceptible individuals an infected individual can infect divided by how long the infected individual is infectious. Because this is a model of chronic infection, the denominator is γ (which is fixed), and subsequent comparative statics on R_0 are just changes in β .

Note that because there is no change net change in the population, $\dot{V}(t) + \dot{I}(t) + \dot{S}(t) = 0$, and one can calculate the change in $\dot{S}(t)$. I focus on steady-state values, so that the changes in vaccination, infection and susceptible populations must be zero ($\dot{I}(t) = \dot{V}(t) = 0$). Solving these yields the following unique equilibrium. In particular:

$$V = \frac{\theta Q}{\gamma} \quad (21)$$

Note that if R_0 is sufficiently small, then infections will necessary go to 0 without any vaccination. Hence, if the equilibrium $I = 0$, then $S = 1 - \frac{\theta Q}{\gamma}$. However, if $I > 0$, then the susceptible population is $\frac{1}{R_0}$ and the number infected is $I^* = 1 - \frac{\theta Q}{\gamma} - \frac{1}{R_0}$. Importantly, this equilibrium depends on the mass of newborns who choose vaccination Q (which we will determine later), and this equilibrium is summarized below:

$$I = \max\{0, 1 - \frac{\theta Q}{\gamma} - \frac{1}{R_0}\} \quad (22)$$

$$S = \min\{1 - \frac{\theta Q}{\gamma}, \frac{1}{R_0}\} \quad (23)$$

2.3.2 Consumer Valuation and Demand Under Monopoly

Given the steady-state values of vaccination and infection from above, I now discuss how individuals value health, with particular attention to the free-rider problem. I assume homogeneous, risk-neutral consumers, who have a flow value of health, $h > 0$, if not infected. The total health value of the population is thus $HV^* = h(1 - I^*)$.

The consumer's value of private benefit of vaccination is thus:

$$MPB = \frac{\theta h}{\gamma} \left(\frac{IR_0}{1 + IR_0} \right) \quad (24)$$

which is the discounted value of the flow value of health $\frac{h}{\gamma}$, times the probability of successful vaccination θ , times the probability of infection over the lifetime (in steady-state). Importantly, this relationship describes the perverse free-rider problem: as steady-state infections decline, the marginal private benefit (ie. willingness to pay) declines as well. In contrast, consider the marginal social health value, which is the marginal benefit of an incremental vaccination: $MHV^* \equiv \frac{dHV^*}{dQ^*} = \frac{\theta h}{\gamma}$. The marginal social health value is constant,

and is always greater than the marginal private benefit.

The demand relation for a monopolist charging price P is thus determined by the marginal private benefit relation (24) and the steady-state level of infection (22). Substituting these relations together yields the demand function

$$\tilde{D}(P) = \frac{\gamma}{\theta} \left[1 - \frac{\theta h}{(\theta h - \gamma P) R_0} \right] \text{ such that } \tilde{D}(P) \leq \gamma \quad (25)$$

Finally, a monopolist maximizes price to solve:

$$\max_P (P - c) D(P) \text{ s.t. } D(P) \leq \gamma \quad (26)$$

I now solve for equilibrium. For ease of exposition, define the parameter $\rho \equiv \frac{\theta h}{\theta h - \gamma c}$ ³³, and assume that $\frac{\theta h}{\gamma} > c$. For the remainder of the analysis, fix $R_0 > 1$, which is necessary for there to be some infections in steady-state absent any vaccination.

Before moving on to the algebraic derivation, I discuss the characteristics of equilibrium as θ varies. In particular, there are three different regimes for the monopolist:

Case 1. When the efficacy of the vaccine is very low, then no one takes the vaccine ($Q = 0$).

In Panel A of figure 11, this is represented by the flat portion near $\theta \sim 0$.

Case 2. When the efficacy of the vaccine is moderate, every newborn takes the vaccine ($Q = \gamma$).

However, because the efficacy of the vaccine is low, there are still infections in steady-state ($I = 1 - \theta - \frac{1}{R_0}$). In Panel A of figure 11, this is represented by the solid line and flat portion near $\theta \sim 0.2$.

Case 3. When the efficacy of the vaccine is high, the rate of vaccination decreases as efficacy increases. However, as long as costs are positive ($c > 0$), there are always some infections in steady-state³⁴. This is represented in Panel A of figure 11 after $\theta > 0.3$.

³³Similarly, note that the inverse demand curve is $P(Q) = \frac{\theta h}{\gamma} \left[1 - \frac{\gamma}{(\gamma - \theta Q) R_0} \right]$.

³⁴A simple way to show this is that if costs are positive (ie. no subsidy), then a monopolist will only produce ($Q^* > 0$) if it can charge a price $P^* > 0$. But by equation 24, $I^* > 0$.

In fact, in this regime, the number of infected patients is almost constant (see Panel A, Figure 12, solid blue line) despite pushing the efficacy of the vaccine to one hundred percent efficacy. In this region, the difference between the marginal health value and the marginal private benefit (ie. the marginal externality) is extremely high. This can be reflected in figure 13, where the profit to the firm increases dramatically when everyone gets vaccinated, but only modestly increases as vaccine efficacy is high. One implication of this, which I will explore in the next section, is the dulling of innovation incentives for highly efficacious vaccines.

Algebraically, the equilibrium quantity Q^* , when unconstrained, is given by (details are given in Appendix 2.8):

$$Q^* = \frac{\gamma}{\theta} \left(1 - \sqrt{\frac{\rho}{R_0}}\right)$$

At the constraint $Q^* = \gamma$, the threshold for vaccine efficacy is given by θ^* which solves $\theta = 1 - \sqrt{\frac{\rho}{R_0}}$ ³⁵.

Monopoly profits can be written as:

$$\Pi^M = \begin{cases} 0 & \text{if } \theta < \frac{R_0 \gamma c}{h(R_0 - 1)} \\ \frac{h}{\rho} \left(1 - \sqrt{\frac{\rho}{R_0}}\right)^2 & \text{if } \theta \in \left[\frac{R_0 \gamma c}{h(R_0 - 1)}, \theta^*\right) \\ \theta h \left[\frac{1}{\rho} - \frac{1}{(1 - \theta)R_0}\right] & \text{if } \theta \in (\theta^*, 1) \end{cases} \quad (27)$$

Table 19 summarizes most of the results. Note that the top condition stipulates that the efficacy has to be at least minimally effective given the cost and the private valuation of health. Second, the increase in profits ($\frac{\partial \Pi^M}{\partial \theta}$) is substantially higher in the second case (everyone vaccinates because efficacy is low), than in the third case ("free-rider" incomplete vaccination as efficacy increases).

³⁵After substitution of ρ , θ^* solves $\frac{\theta h}{\theta h - \gamma c} = (1 - \theta)^2 R_0$. Unfortunately, this results in an equation with a third-degree polynomial, but for empirically relevant parameter values, there is a single root.

2.3.3 Product Subsidies

I consider now optimal government subsidies paid to the firm, and I argue that the optimal product subsidy either has every newborn vaccinated or sets steady-state infections to 0. In particular, at an optimum, the government should subsidize costs as long as the marginal health value is greater than marginal social costs. Recall from the epidemiological equilibrium that the total health value is $HV^* = h(1 - I^*)$ (ie. health times those not infected), where the steady-state level of infection is $I^* = 1 - \frac{\theta Q}{\gamma} - \frac{1}{R_0}$.

Now, define welfare as $W^* = HV^* - cQ^*$, which is the total health value minus the true resource cost of the vaccine. Note that an incremental increase in vaccination $\frac{dW^*}{dQ^*} = \frac{\theta h}{\gamma} - c > 0$, where the last inequality follows from a necessary assumption on the parameter values (see the prior section). Because the incremental vaccination is always socially optimal, the optimal subsidy is to set either $Q^* = \gamma$ or $I^* = 0$. To calculate the threshold value when $I^* = 0$, set $Q = \frac{\gamma}{\theta}(1 - \frac{1}{R_0})$, so that $I^* = 0$ when $\frac{1}{\theta}(1 - \frac{1}{R_0}) \leq 1 \Rightarrow \theta \geq \frac{R_0-1}{R_0}$. However, if $\theta < \frac{R_0-1}{R_0}$, then at best, all households get the vaccination, but there is still some infection in steady-state.

To implement this, the government sets a marginal cost of $c - GS^*$, where GS^* is paid directly to the firm. The magnitude of the government subsidy, especially when the vaccine is very efficacious, can be quite large. For example, figure 14 visualizes the subsidy paid to the firm. Those subsidies correspond to substantially more individuals getting vaccinated (see figure 11, where the dashed line shows the impact of subsidies). Note that when the quantity of vaccinations drops with optimal subsidies (ie. dash line falls in figure 11), the steady-state level of infections is 0, which is shown figure 12.

Now, I consider changes to profit (see figure 13). Importantly, subsidies dramatically raise the profits of the firm. However, while the level of profit is higher, profit is very flat when equilibrium infections are zero with subsidies.

Table 19 (below) summarizes many of the salient endogenous variables. Note that *case 3* from above is now two different cases: when an optimal subsidy pushes all newborns to be

vaccinated, and when an optimal subsidy pushes equilibrium infections to 0.

2.3.4 Calibrating Health Variables

Under the framework discussed, there are several parameters of interest: First, γ is interpreted as the birth and death rate of the population to keep the population at steady-state. As such, $\frac{1}{\gamma}$ is the expected lifetime of an individual, so setting $\gamma = 0.015$ is approximately a life expectancy of 66 years. I examine steady-states, without regard to discounting or transition paths.

h can be interpreted as the flow benefit of health, which disappears when infected (alternatively, this is the flow harm from the disease). This measure can vary highly across different infectious diseases: for example, the economic burden of chronic HIV infection is estimated to be about ten thousand dollars without co-morbidities (Magoni et al. (2011)). In contrast, the total direct costs of chronic Hepatitis B is about one thousand five hundred dollars (Kavosi et al. (2014)). I assume a health value of 5,000.

The marginal cost of a vaccine is estimated to be pennies, especially for traditional attenuated viruses. However, the marginal cost of the mRNA COVID vaccines has not been published. I simply assume a marginal cost of 5.

Table 20 describes a range of estimated R_0 values for various viruses. In my simulations, I assume R_0 is 2 (though I provide some examples where $R_0 = 5$).

2.4 Innovation and Financing Constraints

My central contention is that investment in vaccine research is inefficiently low, partly because the monopoly profits of vaccine research are insufficient, and partly because there are frictions in financing capital and information intensive firms. In fact, independent of a moral hazard constraint, a monopolist does not invest enough capital relative to the social planner's solution. With a moral hazard constraint, the investment problem is exacerbated, consistent with recent work on small firm financial frictions (Howell (2017)). Here, I first discuss the

general process of biotech R&D, and then I discuss the potential for moral hazard.

Drug development is done in stages. For a standard small molecule that targets a receptor, a common approach to develop a drug candidate is to screen a variety of "backbone" candidates that hit the target in experimental conditions (e.g. assays, or live cells). After a series of backbones is produced, candidates are subsequently modified to maximize pharmacological parameters, including half life, metabolism, and target exposure. After enough iteration in *in vitro* models, a lead molecule is taken into development in *in vivo* (animal) models. Once in animal models, there are a variety of tests that can help characterize the candidate, by examining and testing a range of hypotheses or through surrogate biomarkers. Finally, once drug candidates are in clinic, there is substantial work finding appropriate patients for a drug trial (many drugs have failed because they have recruited patients that are not a true test of the efficacy of the drug).

The drug development process is hard, and there is rarely much visibility into each of these steps that is available to outside investors. As such, some firms may "skimp" on additional experiments to fully understand (and predictably detect) drug properties.

Given this motivation for potential moral hazard, I consider a simple model of moral hazard (adapted from Holmstrom & Tirole (1997)), where a biotech firm with initial endowment A requires outside investment to create a vaccine candidate, where total amount of capital is K (ie. outside investors add $K - A$). In the first stage of the model, the investors and the firm decide on financing terms, where the investors promise the firm's managers a share of expected monopoly profits (R_m) from the product model given above (II). However, the firm's managers can choose instead to consume the investment and "run away", to yield private consumption BK , resulting in no experiments.

Hence, the investors must choose a return of the profits to the managers (R_b), such that the managers behave. This corresponds to the incentive compatibility (*IC*) constraint:

$$\int_0^1 R_b(\theta) dF(\theta) \geq BK \quad (28)$$

Investors are risk-neutral, the market for investment is perfectly competitive, and the risk free rate is zero. Hence, the participation constraint (PC) for investors is:

$$\int_0^1 (\Pi(\theta) - R_b(\theta)) dF(\theta) \geq K - A \quad (29)$$

Because the market for investment is competitive, the PC binds with equality, and substituting this into the IC , yields the investment constraint as a function of the initial endowment:

$$\frac{\int_0^1 \Pi(\theta) dF(\theta) + A}{1 + B} \geq K \quad (30)$$

.

Now, the firm uses capital K and runs experiments, with cost c_τ , in parallel³⁶. Hence, the firm runs $\left\lfloor \frac{K}{c_\tau} \right\rfloor \equiv N$. Each experiment yields a vaccine with potential efficacy θ_i , so that the firm advances $\theta = \max\{\theta_1, \dots, \theta_N\}$ to the product market.

For simplicity, suppose that each potential vaccine candidate θ_i is drawn from a uniform distribution $U[0, 1]$. Then, the distribution of the maximum of N draws has pdf $f(\theta) = N\theta^{N-1}$.

Hence, the firm's profit function is given by:

$$U_f = \max_I \int_0^1 \Pi(\theta) \frac{K}{c_\tau} \theta^{\frac{K}{c_\tau}-1} d\theta - K \text{ s.t. } \frac{\int_0^1 \Pi(\theta) \frac{K}{c_\tau} \theta^{\frac{K}{c_\tau}-1} d\theta + A}{1 + B} \geq K \quad (31)$$

2.4.1 Unconstrained Solution

Define K^U as the unconstrained level of investment. The first order condition for any firm is thus:

$$\underbrace{\int_0^1 \Pi(\theta) \left[\frac{1}{c_\tau} \theta^{\frac{K}{c_\tau}-1} + \frac{K \ln(\theta)}{c_\tau} \theta^{\frac{K}{c_\tau}-1} \right] d\theta}_{MPK} = 1$$

³⁶This is important to make sure that the firm does not learn from sequential experiments, nor does it stop early

The left hand side of the equation represents the marginal product of capital (MPK), which is a function of the expected monopoly profit for a given vaccine efficacy times the derivative of the pdf of the distribution of θ for different numbers of draws. In particular, the effect of a subsidy increases the profit conditional on high levels of θ (which only occurs when $\theta > \theta^*$, defined in the prior section).

For some sense of the magnitude of subsidies on investment, consider Figure 15, where the x-axis across all of the panels varies the cost of experiments (c_τ). Panel A plots optimal investment, Panel B shows the number of experiments, Panel C describes the expected efficacy, and Panel D shows the expected profit of the firm.

2.4.2 Welfare without Moral Hazard

Absent a financing constraint, does the market choose an optimal level of investment? First, I compare the optimal number of experiments that a monopolist chooses, and the level of capital that would be preferred by the social planner. Next, I consider implementation of the social planner's preferred capital investment through product subsidies alone.

In the product market, there is an initial monopoly distortion in the pricing of the vaccine. However, the monopolist also invests less in the R&D phase of discovery. In particular, the social planner's criterion is given by:

$$W(k) = \max_k \int_0^1 (HV^*(\theta(k)) - cQ^* - K)dF(\theta) \quad (32)$$

Here as in the product market analysis from above, there is still a marginal externality associated with increasing the value of health relative to the true resource cost of vaccine manufacturing. However, the monopolist does not internalize that value, and so it sets a maximal amount of capital less than the social planners ($K \sim 6$). However, as can be seen in Panel D of figure 16, the social planner would like to set $K \sim 8$. This difference is the distortion resulting from the free-rider problem.

Interestingly, adding small product subsidies does not resolve this inefficient level of investment, and actually decreases investment. Figure 17 shows that expected efficacy actually declines, while expected profit, health value, and total welfare increases. In particular, an across-the-board subsidy raises profits most at medium levels of efficacy, while the effect of the free-rider problem attenuates the effect of the subsidy at higher levels of efficacy. Hence, the monopolist invests even less in R&D.

2.4.3 Binding Moral Hazard Constraint

Consider now when the constraint from equation 30 binds. In that case, using the envelope theorem:

$$\frac{dU_f}{dA} = \left\{ \int_0^1 \Pi(\theta) \left[\frac{1}{c_\tau} \theta^{\frac{K}{c_\tau}-1} + \frac{K}{c_\tau} \frac{\ln(\theta)}{c_\tau} \theta^{\frac{K}{c_\tau}-1} \right] d\theta - 1 \right\} \frac{dK}{dA}$$

Because the constraint binds with equality, substituting $\frac{dK}{dA}$ from the constraints leads to the “shadow value” of initial net worth:

$$\frac{MPK - 1}{B - (MPK - 1)} \equiv v$$

which is a function of the marginal product of additional investment and the degree of private benefit. This shadow cost of outside capital has a natural interpretation, mainly, as the cost of capital for biotech firms.

Figure 18 shows the level of investment with and without capital subsidies. In the left panel, when $B = 3$, the firm is constrained in investment (the horizontal orange line is the unconstrained level of investment), until the initial endowment is very high. Importantly, note that the level of endowment is actually higher than the unconstrained level, suggesting that the firm could invest with internal funds. While this may seem puzzling, one interpretation of the private benefit is that a manager may “run away” even with their own funds to consume, against the wishes of existing shareholders. In that case, the endowment acts as collateral for additional funds.

In the right panel, a private benefit of $B = 3$ for the same parameters does not make the firm constrained. This highlights the use of ex-ante product incentives in alleviating constraints. However, when $B = 5$, an initial endowment of $A \sim 4$ effectively eases the constraints, and allows the firm to invest at the private-market, unconstrained level.

2.4.4 Comparing product subsidies and financing subsidies

Are product subsidies or financing subsidies better? To a certain extent, product subsidies ease financing constraints, although they can be extremely costly. On the other hand, incremental experiments can dramatically increase the efficacy of a vaccine. However, increasing vaccine efficacy can push households to decrease their vaccination rate, which dulls the welfare impact of loosening financing constraints.

For a social planner, the total expected welfare is:

$$W(\theta) = \int_0^1 (HV^*(\theta) - cQ^* - K)dF(\theta) \text{ s.t. } \frac{\int_0^1 \Pi(\theta)dF(\theta) + A}{1 + B} \geq K \quad (33)$$

Here a product subsidy is again a subsidy on the marginal cost of production, while a capital subsidy is a grant for the firm's endowment A . Figure 19 describes the changes in total welfare as subsidies are allocated between capital and product market. In particular, Panel A shows that, at the far left side of the graph, when all subsidies go to development and alleviating capital constraints, welfare is still higher than no subsidies. However, as product subsidies increase (with concomitant decreases in capital subsidies), total welfare increases (including the expected value of health). However, there is an optimal level of product market subsidies, since once the subsidies are too small, the expected efficacy of the vaccine is limited, and expected health value declines dramatically.

2.5 Discussion and Conclusion

The free-rider problem for vaccines generates several pernicious effects. First, in the product market absent any subsidies, there is always some equilibrium level of infection, even with extremely efficacious vaccines³⁷. Second, even a monopolist without any financing constraints will not invest at the level desired by a social planner, because the social planner internalizes the free-rider problem, while the monopolist does not. Third, a minor subsidy, while increasing welfare, actually decreases the incentive for investment because the subsidy is most profitable at intermediate levels of vaccine efficacy. Finally, a firm that is capital constrained will invest substantially less than an unconstrained firm, and even less than the social optimum.

Hence, government subsidies of capital can be justified, by itself, because of the free-rider problem. With a capital constraint, the government can alleviate that constraint by giving grants that increase the cash-on-hand endowment of the firm. Despite these theoretical arguments, government funding of private companies focused on infectious disease is limited, and primarily channeled through the Biomedical Advanced Research and Development Authority (BARDA)³⁸. Traditionally, BARDA has used two policy levers: first, they stockpile a variety of medicines for both bio-defense and potential pandemics³⁹ (which can be interpreted as a product subsidy); second, they issue grants⁴⁰ (which can be interpreted as a capital subsidy).

While I argue that subsidizing biotech financing can be welfare enhancing, Kremer (2000) argues that the "push" of government grants, in contrast to "pull" programs of subsidies, is subject to several limitations: first, there may be asymmetric information about the quality of

³⁷This finding is similar to Geoffard & Philipson (1997).

³⁸BARDA was originally established in 2006 principally to manage threats from bioterrorism. BARDA has also had a prominent role in funding Coronavirus vaccine development. For example, they have distributed grants to Moderna, BioNTech/Pfizer, and others.

³⁹For example, Bavarian Nordic is contracted through BARDA to stockpile millions of smallpox and monkeypox vaccines.

⁴⁰Historically, these grants have been relatively small. Some of these grants have gone to anti-infectives companies like Spero Therapeutics, developing an oral antibiotic against UTIs.

vaccine candidates, and the government may run into a lemons problem; second, governments may require control terms, and they may be unwilling to shut down negative NPV projects; third, there may be monitoring issues, where firms may divert investments to "pet projects"; and fourth, the determination of grant recipients may be vulnerable to political lobbying and interests. Kremer (2000), in support of his argument, cites a fascinating story of USAID's efforts to develop a malaria vaccine ⁴¹.

I argue that the lemons problem is, in practice, less concerning than Kremer (2000) suggests. For example, private market license and joint venture agreements have very similar theoretical issues, though those issues have not (for the most part) derailed development programs. In particular, there is ample evidence suggesting that there is positive selection:

⁴¹Taken directly from Kremer (2000): Desowitz [1991] chronicles the sad story of the U.S. Agency for International Development's (USAID) 1980s push program to develop a malaria vaccine. USAID's efforts focused on three teams. A candidate vaccine was developed by the first team. Tests with 9 volunteers found that only 2 were protected from malaria, and suggested that the vaccine created side effects. These results, mixed at best, did not prevent USAID from issuing wildly overoptimistic statements. In 1984, the agency claimed that there had been a "major breakthrough in the development of a vaccine against the most deadly form of malaria in human beings. The vaccine should be ready for use around the world, especially in developing countries, within five years." ¹⁸ Fifteen years later, the world is still waiting for a malaria vaccine. Early work by the second team yielded disappointing results, but not surprisingly, the principal investigator argued that his approach was still worth pursuing and requested an additional \$2.38 million from USAID. The expert consultants assigned to review the project recommended that the research not be funded. However, USAID's malaria vaccine project director told the USAID Office of Procurement that the expert panel "had endorsed the scientific methodology and the exceptional qualifications and experience of the researchers." ¹⁹ Once the grant came through, the principal investigator transferred grant funds to his personal account. He was later indicted for theft.

The external evaluations of the third proposal called it mediocre and unrealistic. The USAID project director ignored the report and arranged for the project to be fully funded. The principal investigator and his administrative assistant were later indicted for theft and criminal conspiracy in diverting money from the grant to their personal accounts. Two months before his arrest, the Rockefeller Foundation had provided him with a \$750,000 research grant, and on the very day that he was arrested, USAID announced it was giving him an additional \$1.65 million for research.

By 1986, USAID had spent over \$60 million on its malaria vaccine efforts, with little progress. Since USAID believed that there would soon be many candidate malaria vaccines suitable for testing, it tried to obtain monkeys as test subjects for these vaccines. USAID's malaria vaccine project director, James Erickson, arranged for a contract to acquire monkeys to go to an associate who paid him a kickback. Erickson eventually pleaded guilty to accepting an illegal gratuity, filing false tax returns, and making false statements. What about outside oversight? USAID had arranged for independent oversight to be provided by the American Institute of Biological Science (AIBS). Erickson and the AIBS assigned project manager were lovers.

The USAID case is extreme, and many push programs are quite successful. But the general principle remains that researchers funded under push programs have incentives to be overoptimistic, and since they are paid before delivering a product, they may be tempted to divert resources away from the search for a vaccine. The USAID example may shed light on why the administrators of push programs and the researchers financed by such programs often believe that push programs are somehow cheaper than pull programs.

projects that are part of joint venture agreements are more likely to be FDA approved (Danzon et al. (2005)), consistent with the prominent role of business development roles in pharma companies.

To alleviate the other concerns, I argue that a government policy which focuses on replicating private market joint ventures—maybe by providing cheap capital while taking upside equity—can align incentives for both parties. While this joint-venture structure is not modeled here, it is a very common arrangement between cash-rich pharma and small biotech companies. For example, table 22 describes a dataset of several hundred public joint venture agreements between 1995 and 2014⁴². The types and distribution of funding tend to vary by development phase, where later stage (Phase 2 and 3 assets) command higher deal prices and greater upfront payments. The exception to this is the very earliest stage deals (ie. "Development" deals), which are typically a very large "hypothetical" amount with very limited funding upfront and poor licensing terms for the biotech firm. As the deals become de-risked, licensing deals become more favorable to the biotech company, as the deal size increases and the share of the deal that is paid upfront increases.

⁴²The agreements are made public because a small biotech company must disclose material changes to the SEC. Note that many of these contracts, in the public SEC database, are redacted. However, the data vendor has been able to FOIA most of these contracts to get unredacted terms, including upfront and milestone payments, and different control terms (Lerner & Merges (1998)).

2.6 Tables

Table 15: As of Jan 1, 2020, last twelve months of reported revenue (\$M) for all US-listed public companies between \$100M and \$10B market cap by sector.

Sector	Median	Mean	25th%tile	75th%tile	N
Biotech	\$ 4	\$ 57	\$ -	\$ 51	211
Consumer Discret	\$ 1,496	\$ 3,199	\$ 614	\$ 3,311	343
HC Services	\$ 1,065	\$ 2,261	\$ 310	\$ 3,110	57
Med Devices	\$ 120	\$ 326	\$ 42	\$ 388	108
Software	\$ 303	\$ 489	\$ 140	\$ 548	131
Utilities	\$ 1,131	\$ 1,936	\$ 418	\$ 1,855	46

Table 16: As of Jan 1, 2020, last twelve months of reported short term and long term debt (\$M) for all US-listed public companies between \$100M and \$10B market cap by sector.

Sector	Median	Mean	25th%tile	75th%tile	N
Biotech	\$ 16	\$ 67	\$ 3	\$ 48	211
Consumer Discret	\$ 730	\$ 1,646	\$ 194	\$ 2,047	343
HC Services	\$ 325	\$ 1,408	\$ 38	\$ 1,359	57
Med Devices	\$ 28	\$ 169	\$ 5	\$ 148	108
Software	\$ 85	\$ 341	\$ 15	\$ 404	131
Utilities	\$ 1,667	\$ 4,289	\$ 519	\$ 3,237	46

Table 17: As of Jan 1, 2020, beta to S&P500 index for all US-listed public companies between \$100M and \$10B market cap by sector.

Sector	Median	Mean	25th%tile	75th%tile	N
Biotech	1.2	1.4	0.9	1.5	211
Consumer Discret	0.9	0.6	0.8	1.1	343
HC Services	0.9	0.8	0.7	1.1	57
Med Devices	1.0	1.2	0.8	1.3	108
Software	1.0	0.8	0.8	1.2	131
Utilities	0.6	0.6	0.5	0.7	46

Table 18: Difference-in-difference estimate of R&D spending during the Great Financial Crisis.

VARIABLES	(1)
	log Opex Diff
t1	-0.0376 (0.0199)
Constant	-1.407*** (0.0343)
Firm Fixed Effects	Yes
Quarter Fixed Effects	Yes
Observations	1,088
R-squared	0.420

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Variable	$\theta \in [0, \frac{\gamma c}{h} \frac{R_0}{R_0-1}]$	$\theta \in [\frac{\gamma c}{h} \frac{R_0}{R_0-1}, \theta^*]$	$\theta \in (\theta^*, \frac{R_0-1}{R_0})$	$\theta \in (\frac{R_0-1}{R_0}, 1]$
Number Vaccinated (Q^*)	0	γ	$\frac{\gamma}{\theta}(1 - \sqrt{\frac{\rho}{R_0}})$	same as left
Monopoly Profit (Π)	0	$\theta h[\frac{1}{\rho} - \frac{1}{(1-\theta)R_0}]$	$\frac{h}{\rho}(1 - \sqrt{\frac{\rho}{R_0}})^2$	same as left
Welfare (W^*)	$\frac{h}{R_0}$	$\theta h(\frac{1}{\rho} + \frac{1}{\theta R_0})$	$\frac{h}{\rho}(1 - \sqrt{\frac{\rho}{R_0}} + \frac{\rho}{R_0})$	same as left
Subsidy GS^*	NA	NA	$c + \frac{\theta h}{\gamma}[\frac{1}{(1-\theta)^2 R_0} - 1]$	$c + \frac{\theta h(R_0-1)}{\gamma}$
Q_m under subsidies	NA	NA	γ	$\frac{\gamma}{\theta}(1 - \frac{1}{R_0})$
Π_m under subsidies	NA	NA	$\frac{\theta^2 h}{(1-\theta)^2 R_0}$	$h \frac{(R_0-1)^2}{R_0}$

Table 19: Solutions to the product market.

Infection	Time Period	R0
HIV (Type I)	1981-85	2-5
Rubella	1960-70	6-7
Chicken pox	1944-68	10-12
Mumps	1960-80	11-14
Measles	1950-68	16-18
Pertussis	1944-78	16-18

Table 20: Various estimated R_0 's from history.

Table 21: Data from database of joint venture agreements.

		Exclusive			Semi-Exclusive			Non-Exclusive		
		Deal Size	%Upfront	Sales*	Deal Size	%Upfront	Sales*	Deal Size	%Upfront	Sales*
Other	Mean	\$ 45	21.2	11.0	\$ 26	34.8	0.3	\$ 11	33.9	0.6
	25%tile	\$ -	-	-	\$ -	-	-	\$ -	-	-
	50%tile	\$ 3	-	-	\$ 1	-	-	\$ 1	-	-
	75%tile	\$ 25	30.3	-	\$ 18	88.2	-	\$ 9	100.0	-
	n	611	611	611	42	42	42	204	204	204
Development	Mean	\$ 146	8.8	26.6	\$ 39	18.1	5.6	\$ 16	25.7	0.9
	25%tile	\$ 8	-	-	\$ -	-	-	\$ 0	-	-
	50%tile	\$ 35	2.6	-	\$ 6	-	-	\$ 4	0.9	-
	75%tile	\$ 153	9.9	-	\$ 21	21.5	-	\$ 12	49.6	-
	n	658	658	658	47	47	47	159	159	159
Lead Molecule	Mean	\$ 61	13.0	4.0	\$ 17	2.6	-	\$ 14	24.7	5.2
	25%tile	\$ 3	-	-	\$ -	-	-	\$ 1	-	-
	50%tile	\$ 16	4.4	-	\$ -	-	-	\$ 4	3.7	-
	75%tile	\$ 56	14.0	-	\$ 52	7.7	-	\$ 16	37.5	-
	n	157	157	157	3	3	3	30	30	30
Pre Clinical	Mean	\$ 95	11.6	23.4	\$ 20	27.6	0.6	\$ 35	21.8	6.5
	25%tile	\$ 2	-	-	\$ 1	2.1	-	\$ 0	-	-
	50%tile	\$ 23	3.1	-	\$ 5	10.0	-	\$ 2	10.2	-
	75%tile	\$ 85	11.4	-	\$ 36	47.3	-	\$ 24	28.2	-
	n	489	489	489	9	9	9	16	16	16
Phase 1	Mean	\$ 156	12.1	25.9	\$ 202	4.8	55.4	\$ 17	2.8	-
	25%tile	\$ 8	-	-	\$ -	-	-	\$ 2	-	-
	50%tile	\$ 52	5.1	-	\$ -	-	-	\$ 17	2.8	-
	75%tile	\$ 173	13.0	5.0	\$ 112	7.2	32.5	\$ 32	5.7	-
	n	213	213	213	6	6	6	2	2	2
Phase 2	Mean	\$ 180	13.5	42.5	\$ 636	15.5	208.9	\$ 4	35.3	-
	25%tile	\$ 9	-	-	\$ 187	7.7	-	\$ 3	2.0	-
	50%tile	\$ 60	6.9	-	\$ 520	12.0	-	\$ 4	35.3	-
	75%tile	\$ 217	17.1	33.0	\$ 1,070	27.1	100.0	\$ 5	68.6	-
	n	335	335	335	9	9	9	2	2	2
Phase 3	Mean	\$ 181	18.7	48.4	\$ 373	26.6	48.9	\$ 3	27.2	-
	25%tile	\$ 11	4.0	-	\$ 100	7.5	-	\$ -	-	-
	50%tile	\$ 65	13.0	-	\$ 192	15.7	-	\$ 1	22.9	-
	75%tile	\$ 215	23.3	35.0	\$ 377	38.2	126.0	\$ 9	58.8	-
	n	273	273	273	10	10	10	3	3	3

Table 22: Summary statistics of various joint venture agreements between a small biotech and a larger pharma company. Sales* are sales milestones.

2.7 Figures

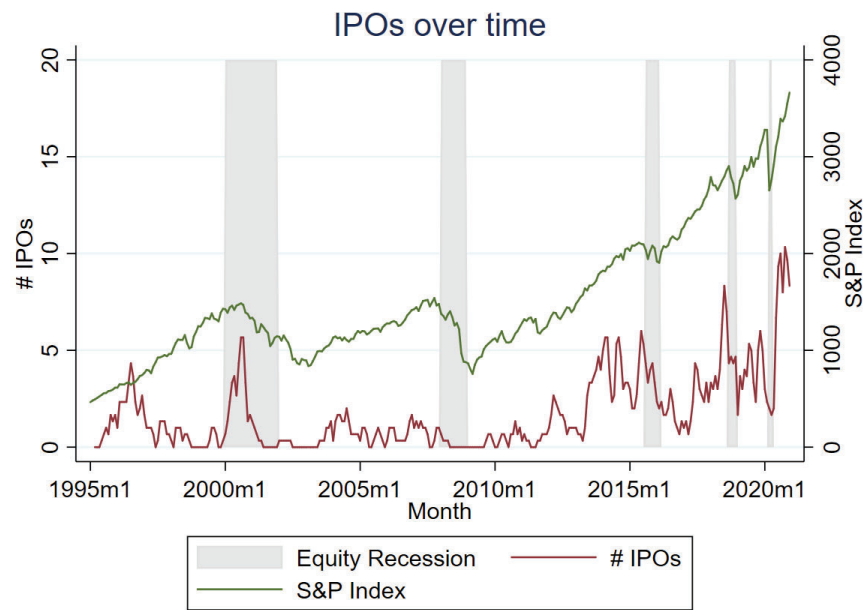


Figure 8: IPOs by month, 3 month moving average.

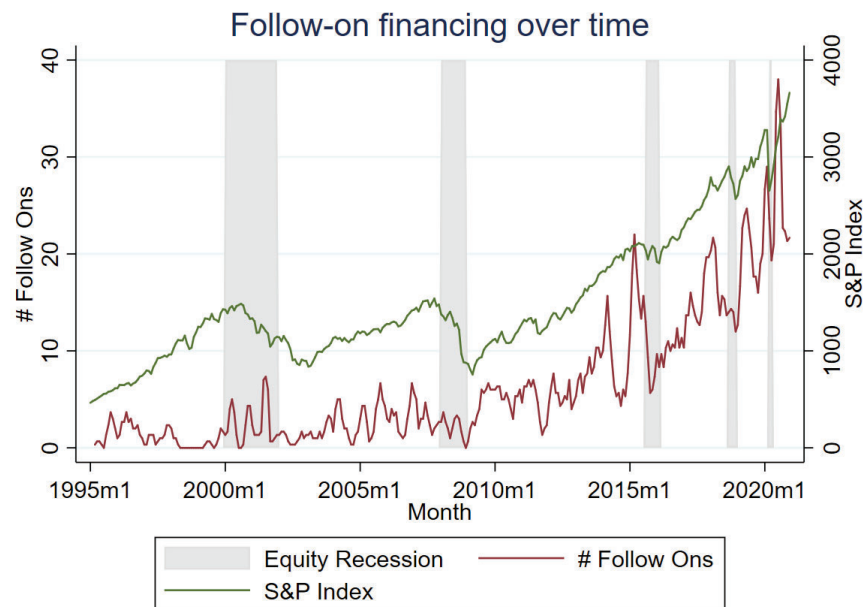
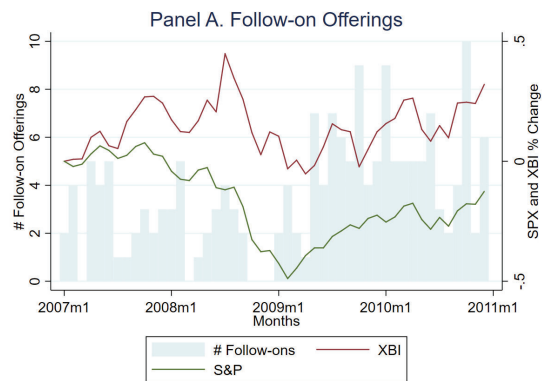
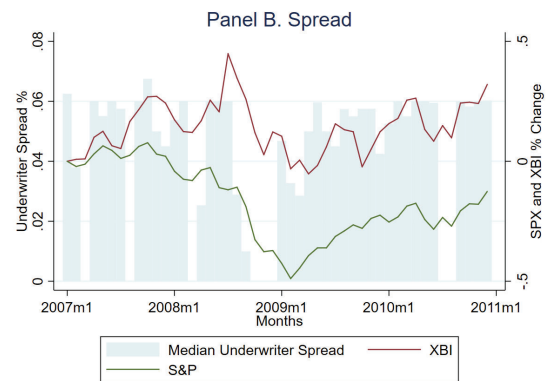


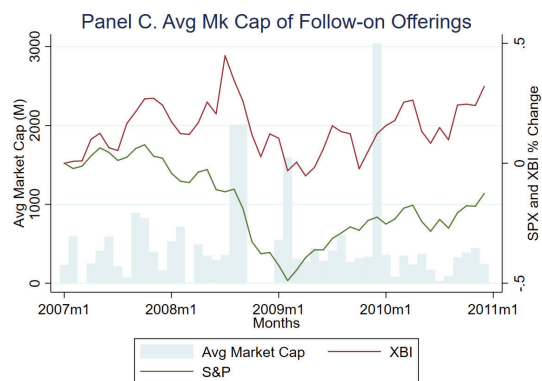
Figure 9: Additional equity financings, by month, 3 month moving average.



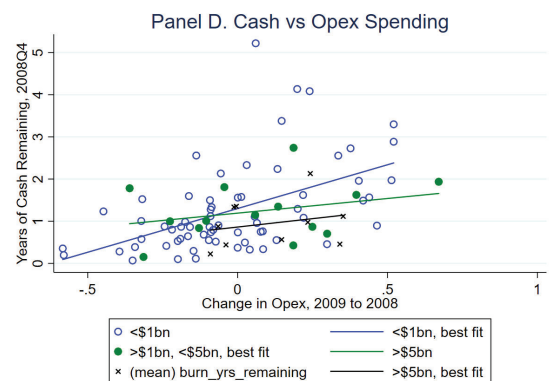
(a) Follow-on Offerings



(b) Underwriter Spread



(c) Composition of Follow-ons



(d) Change in R&D related to cash on hand

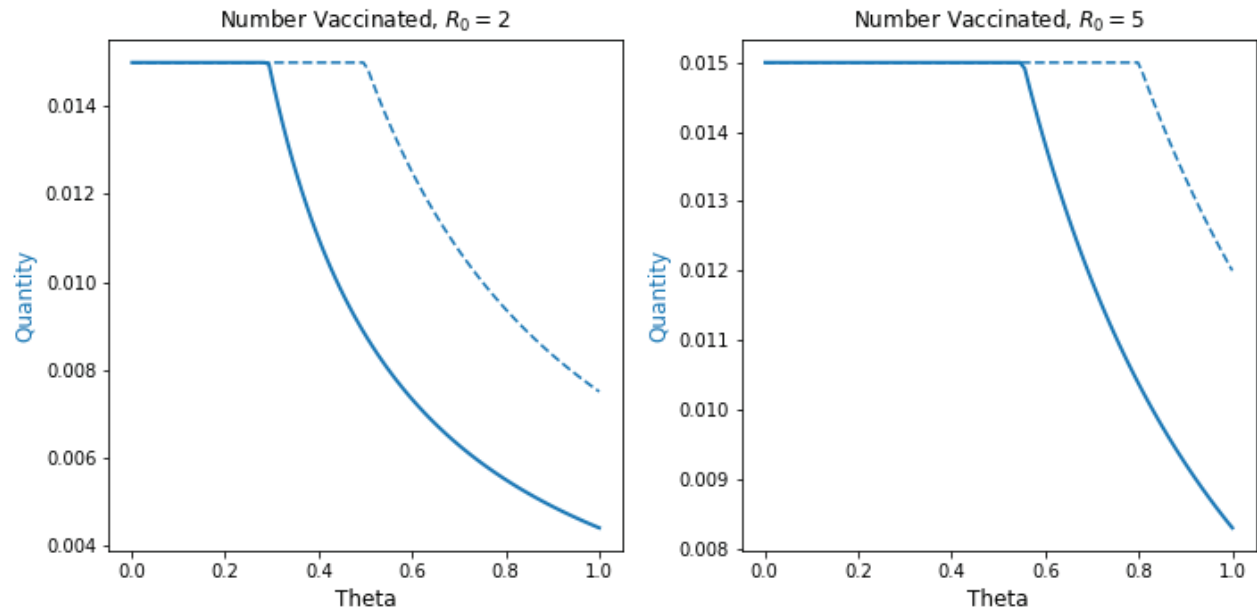


Figure 11: When the θ is at least minimally effective, everyone takes the vaccine. However, as the vaccine becomes increasingly effective, the externality effect pushes vaccination rates lower. A subsidy increases usage to either the maximum, or until equilibrium infections are 0.

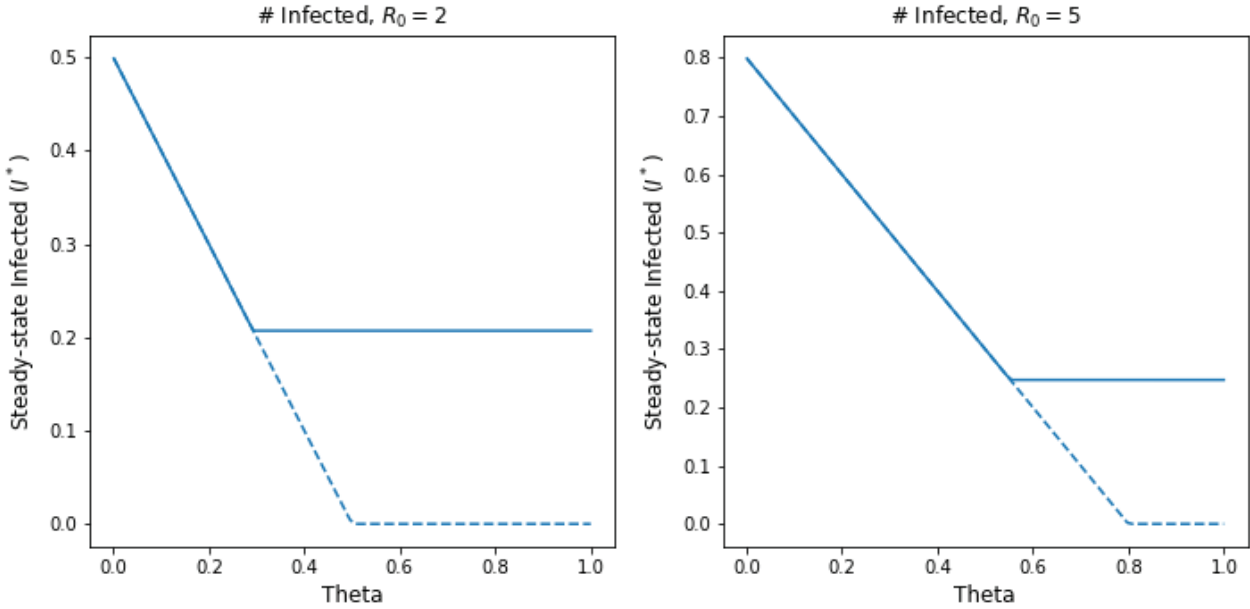


Figure 12: The solid blue line describes equilibrium without the optimal product subsidy, while the dash line includes the optimal subsidy. As the efficacy of the vaccine increases, the steady-state level of infection falls and then flat-lines without subsidies. The flat portion of the infection rate corresponds to the free-rider problem. With optimal subsidies, the steady-state infections can get to 0.

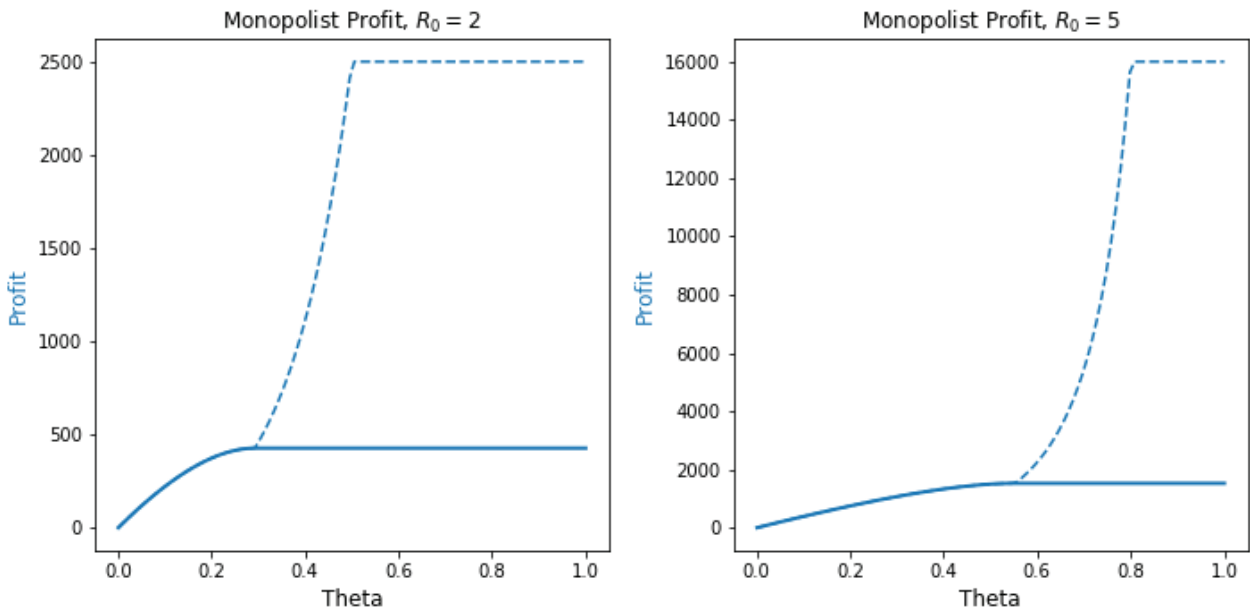


Figure 13: Profit for the monopolist without subsidies (solid line) increases until the free-rider problem lowers the incentive for individuals to get vaccinated.

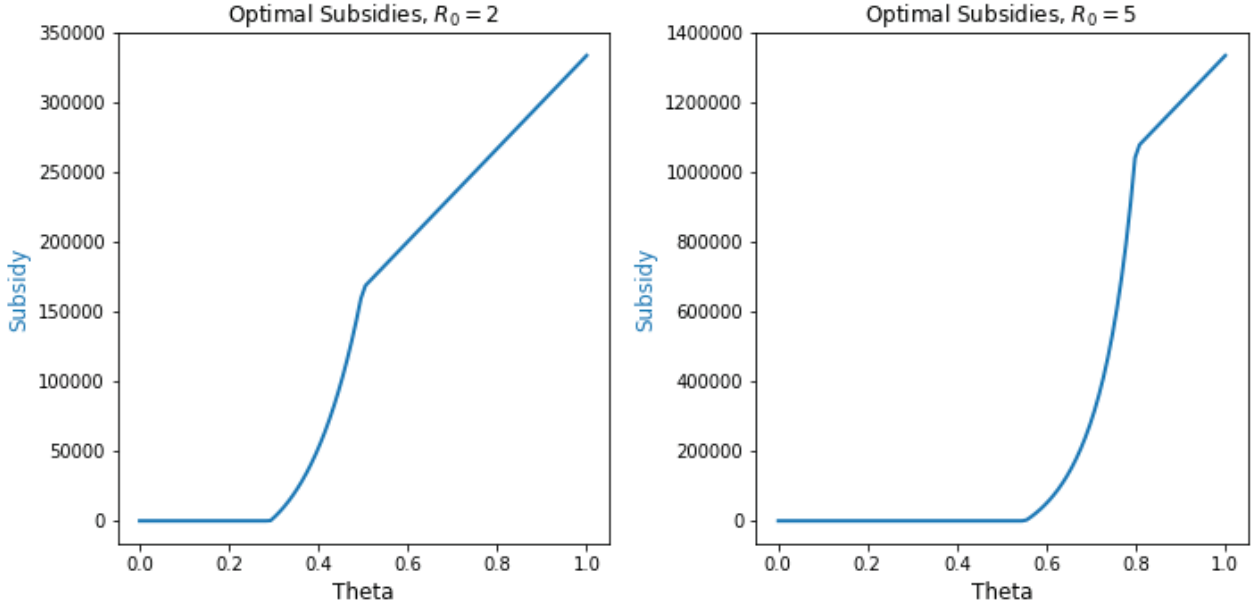


Figure 14: Notice that the subsidy increases as the efficacy of the drug increases. This is because the efficacy of the vaccine induces fewer people to purchase the vaccine (free-riding), which requires substantially more of a subsidy to increase usage.

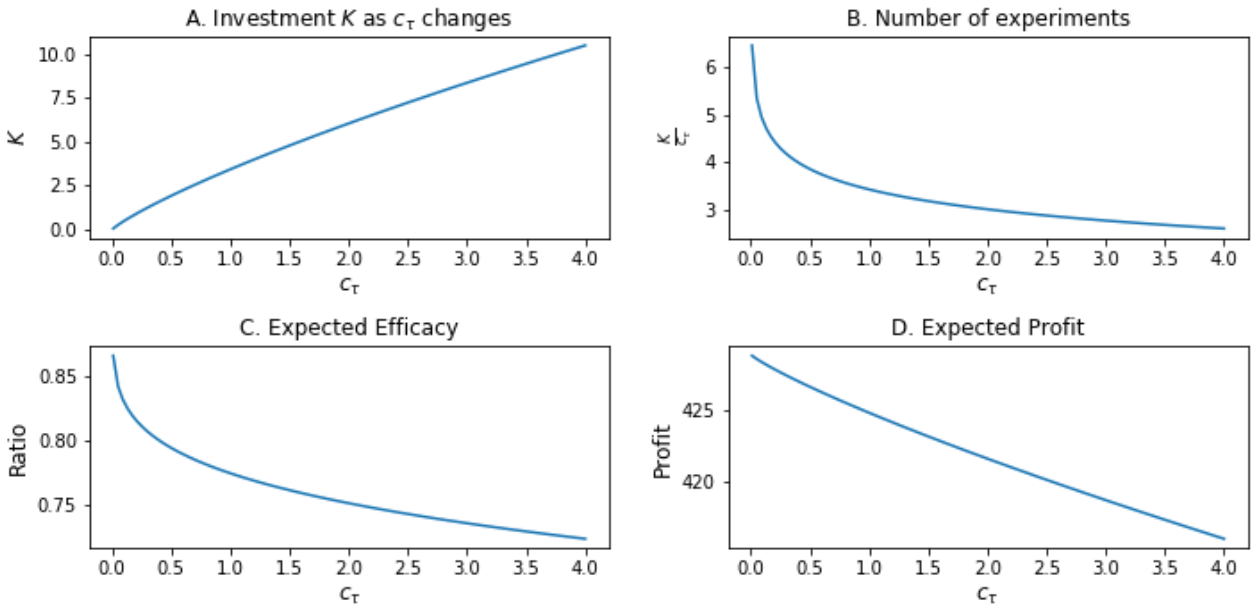


Figure 15: As the cost of running experiments increase, the required level of capital increases, while the number of experiments and expected profit decrease.

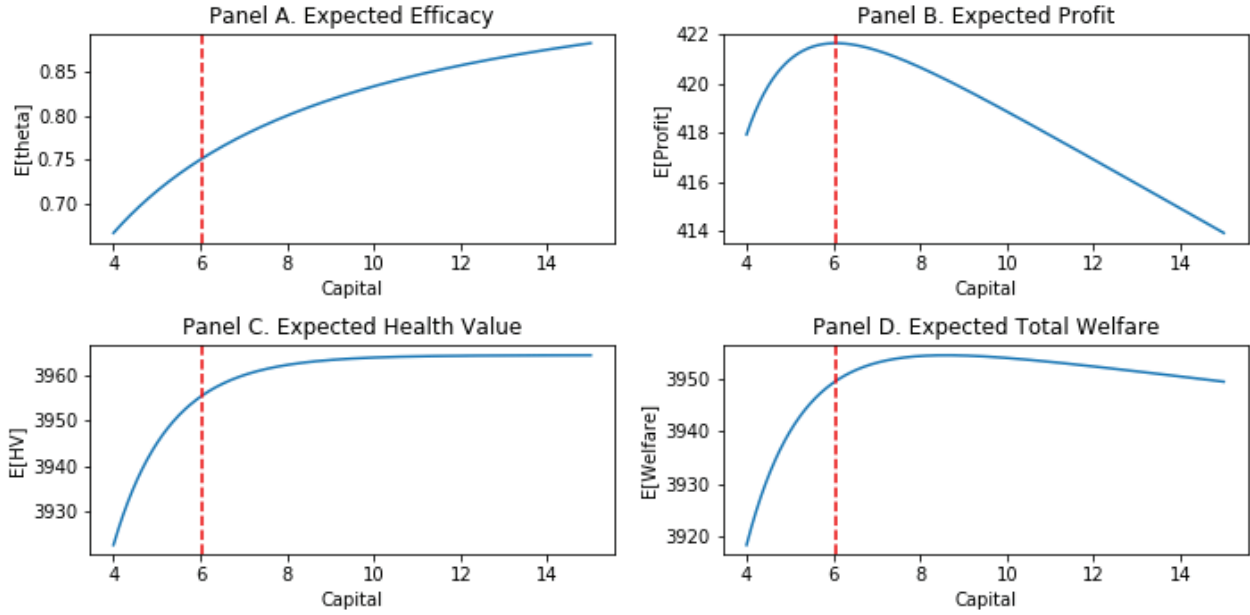


Figure 16: While the unconstrained, private-market level of capital is around 6, the total welfare enhancing level of capital is higher—at about 8.

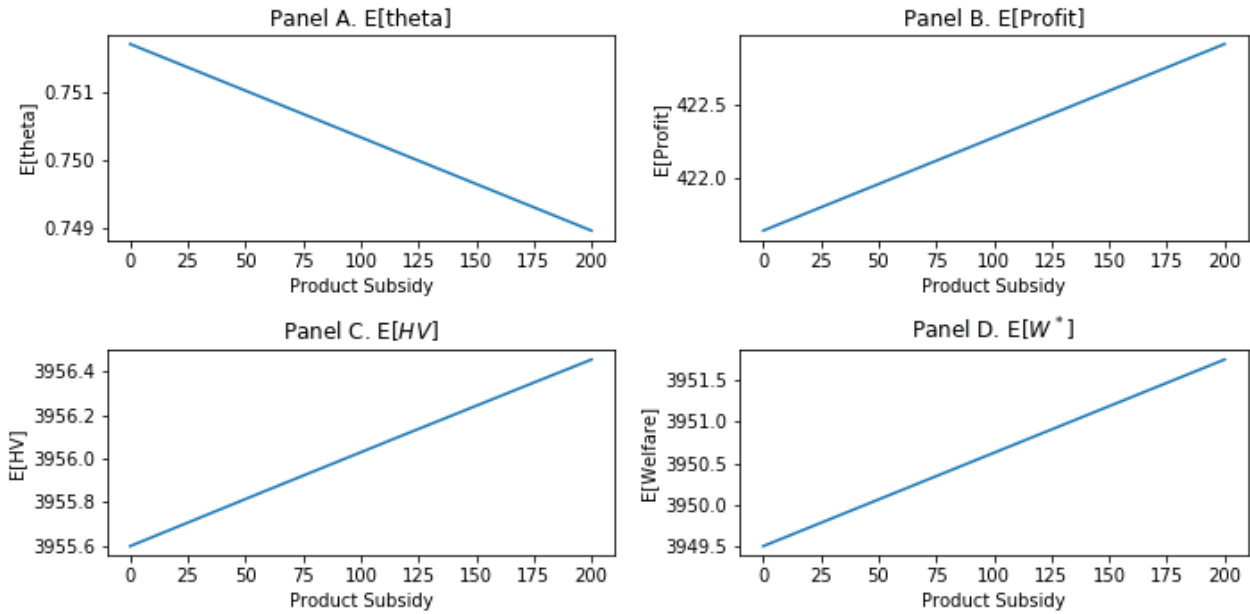


Figure 17: Small product subsidies actually decrease the level of investment, though the subsidies do increase welfare.

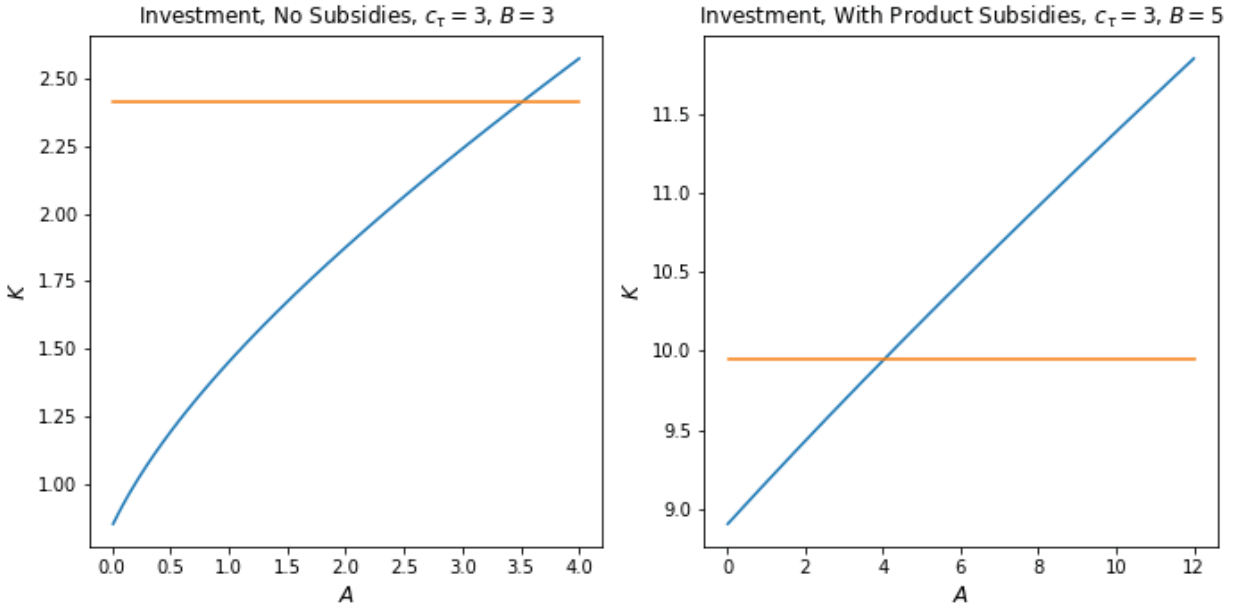


Figure 18: As a monopolist's cash endowment increases, the monopolist can invest more, up to the unconstrained level of investment (orange line).

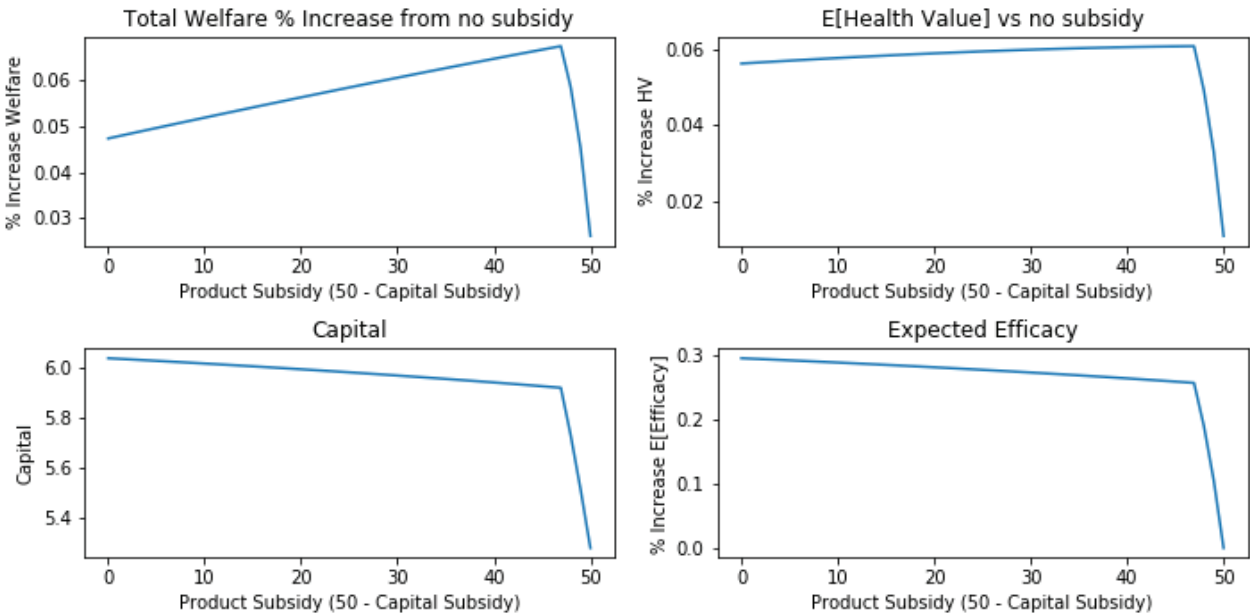


Figure 19: The x-axis varies the subsidy amount between capital and product market subsidies. The far left is all capital subsidies, while the far right is all product market subsidies.

2.8 Appendix A. Proofs

2.8.1 No Subsidies

Consider the inverse demand curve

$$P(Q) = \frac{\theta h}{\gamma} \left[1 - \frac{\gamma}{(\gamma - \theta Q) R_0} \right] \quad (34)$$

The monopolist maximizes:

$$\max_Q Q[P(Q) - c] \text{ s.t. } Q(P) \leq \gamma$$

Consider first when the constraint does not bind. The first order condition for Q is:

$$Q^* = \frac{\gamma}{\theta} \left(1 - \sqrt{\frac{\rho}{R_0}} \right)$$

and hence, the optimal price is given by:

$$P^* = \frac{\theta h}{\gamma} \left(1 - \frac{1}{\sqrt{\rho R_0}} \right)$$

There are three cases:

- 1 $Q^* \leq 0$. This occurs when $\theta \leq \frac{R_0}{R_0-1} \frac{\gamma c}{h}$
- 2 Constraint does not bind, which occurs when $Q^* \in (0, \gamma)$.
- 3 Constraint does bind, which occurs when $Q^* \geq \gamma$, and in that case, P^* is equal to the marginal private benefit $\frac{\theta h}{\gamma} \left(\frac{I R_0}{1 + I R_0} \right)$ where $I = 1 - \theta - \frac{1}{R_0}$.

The constraint binds when $Q^* = \gamma$, which occurs when $1 = \frac{1}{\theta} \left(1 - \sqrt{\frac{\rho}{R_0}} \right)$.

2.8.2 Subsidies

The optimal subsidy brings $Q^* = \min\{\gamma, \frac{\gamma}{\theta}(1 - \frac{1}{R_0})\}$. Hence, when θ is very high, the subsidy brings the number of vaccinated to the total population, but brings steady-state infection to 0.

Hence, when $\theta \in (\theta^*, \frac{R_0-1}{R_0})$, then the subsidy needed to get $Q^* = \gamma$ is given by:

$$\gamma = \frac{\gamma}{\theta}(1 - \sqrt{\frac{\theta h}{\theta h - \gamma \tilde{c}}}) \Rightarrow \tilde{c} = \frac{\theta h}{\gamma}[1 - \frac{1}{(1-\theta)^2 R_0}]$$

Note that the price the monopolist charges when $Q^* = \gamma$ is equal to the private marginal benefit, which in this case is $\frac{\theta h}{\gamma}[1 - \frac{1}{(1-\theta)R_0}]$, which is the same if you substitute the $\tilde{c}$ into the equilibrium pricing function ⁴³.

The monopolist's profit is thus

$$[\frac{\theta h}{\gamma}[1 - \frac{1}{(1-\theta)R_0}] - \frac{\theta h}{\gamma}[1 - \frac{1}{(1-\theta)^2 R_0}]]\gamma = \frac{\theta h}{R_0}[\frac{\theta}{(1-\theta)^2}]$$

When $\theta \in [\frac{R_0-1}{R_0}, 1]$, then $Q^* = \frac{\gamma}{\theta}(1 - \frac{1}{R_0})$, and solving for $\tilde{c}$ yields:

$$\frac{\gamma}{\theta}(1 - \frac{1}{R_0}) = \frac{\gamma}{\theta}(1 - \sqrt{\frac{\theta h}{\theta h - \gamma \tilde{c}}}) \Rightarrow \tilde{c} = \frac{\theta h}{\gamma}(1 - R_0)$$

.

Hence, the equilibrium price is thus 0, because the marginal private benefit is 0 when steady-state infections is 0. The monopolist's profit is thus:

$$\frac{\gamma}{\theta}(1 - \frac{1}{R_0})(-\frac{\theta h}{\gamma}(1 - R_0)) = h \frac{(R_0 - 1)^2}{R_0}$$

⁴³ $P^* = \frac{\theta h}{\gamma}(1 - \frac{1}{\sqrt{\rho R_0}})$, and note that $\rho R_0 = \frac{\theta h}{\theta h - \gamma \frac{\theta h}{\gamma}[1 - \frac{1}{(1-\theta)^2 R_0}]} R_0 = (1-\theta)^2 R_0^2$

References

- Abrams, E. (2019). Assessing bank market power given limited consumer consideration. *Available at SSRN 3431374*.
- Baumol, W. J. (1952). The transactions demand for cash: An inventory theoretic approach. *The Quarterly Journal of Economics*, (pp. 545–556).
- Baye, M. R. & Hoppe, H. C. (2003). The strategic equivalence of rent-seeking, innovation, and patent-race games. *Games and economic behavior*, 44(2), 217–226.
- Berger, A. N. & Hannan, T. H. (1998). The efficiency cost of market power in the banking industry: A test of the “quiet life” and related hypotheses. *Review of economics and statistics*, 80(3), 454–465.
- Bernanke, B. S., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, 1, 1341–1393.
- Bertaut, C. C. (1998). Stockholding behavior of us households: Evidence from the 1983–1989 survey of consumer finances. *Review of Economics and Statistics*, 80(2), 263–275.
- Bourguignon, F., Fournier, M., & Gurgand, M. (2007). Selection bias corrections based on the multinomial logit model: Monte carlo comparisons. *Journal of Economic Surveys*, 21(1), 174–205.
- Brown, J. R., Fazzari, S. M., & Petersen, B. C. (2009). Financing innovation and growth: Cash flow, external equity, and the 1990s r&d boom. *The Journal of Finance*, 64(1), 151–185.
- Campbell, J. Y. & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4), 1509–1531.
- Danzon, P. M., Nicholson, S., & Pereira, N. S. (2005). Productivity in pharmaceutical–biotechnology r&d: the role of experience and alliances. *Journal of health economics*, 24(2), 317–339.
- Dick, A. A. (2008). Demand estimation and consumer welfare in the banking industry.

-
- Journal of Banking & Finance*, 32(8), 1661–1676.
- Drechsler, I., Savov, A., & Schnabl, P. (2017). The deposits channel of monetary policy. *The Quarterly Journal of Economics*, 132(4), 1819–1876.
- Dubin, J. A. & McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, (pp. 345–362).
- Egan, M., Hortaçsu, A., & Matvos, G. (2017). Deposit competition and financial fragility: Evidence from the us banking sector. *The American Economic Review*, 107(1), 169–216.
- Fan, V. Y., Jamison, D. T., & Summers, L. H. (2016). *The inclusive cost of pandemic influenza risk*. Technical report, National Bureau of Economic Research.
- FAZZARI, S. M., HUBBARD, R. G., & PETERSEN, B. (1988). Financing constraints and corporate investment. *Brookings Papers on Economic Activity*, 1988(1), 141–195.
- Genesove, D. & Mullin, W. P. (1998). Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914. *The RAND Journal of Economics*, (pp. 355–377).
- Geoffard, P.-Y. & Philipson, T. (1997). Disease eradication: private versus public vaccination. *The American Economic Review*, 87(1), 222–230.
- Gilbert, R. A. et al. (1986). Requiem for regulation q: What it did and why it passed away. *Federal Reserve Bank of St. Louis Review*, (Feb), 22–37.
- Goodkin-Gold, M., Kremer, M., Snyder, C. M., & Williams, H. (2020). Optimal subsidies for prevention of infectious disease. *Working Paper*.
- Hannan, T. & Berger, A. (1991). The rigidity of prices: Evidence form the banking sector. *American Economic Review*, 81, 938–945.
- Hannan, T. H. & Berger, A. N. (1997). The rigidity of prices: Evidence from the banking industry. *J. Reprints Antitrust L. & Econ.*, 27, 245.
- Hayashi, F. (1982). Tobin’s marginal q and average q: A neoclassical interpretation. *Econometrica: Journal of the Econometric Society*, (pp. 213–224).
- Head, M. G., Brown, R. J., Newell, M.-L., Scott, J. A. G., Batchelor, J., & Atun, R. (2020).

- The allocation of usdollar; 105 billion in global funding from g20 countries for infectious disease research between 2000 and 2017: a content analysis of investments. *The Lancet Global Health*, 8(10), e1295–e1304.
- Hennessy, C. A. & Whited, T. M. (2007). How costly is external financing? evidence from a structural estimation. *The Journal of Finance*, 62(4), 1705–1745.
- Holmstrom, B. & Tirole, J. (1997). Financial intermediation, loanable funds, and the real sector. *the Quarterly Journal of economics*, 112(3), 663–691.
- Howell, S. T. (2017). Financing innovation: Evidence from r&d grants. *American Economic Review*, 107(4), 1136–64.
- Kaplan, S. N. & Zingales, L. (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The quarterly journal of economics*, 112(1), 169–215.
- Kashyap, A. K. & Stein, J. C. (2000). What do a million observations on banks say about the transmission of monetary policy? *American Economic Review*, 90(3), 407–428.
- Kavosi, Z., Zare, F., Jafari, A., & Fattahi, M. R. (2014). Economic burden of hepatitis b virus infection in different stages of disease; a report from southern iran. *Middle East journal of digestive diseases*, 6(3), 156.
- Keeley, M. C. (1990). Deposit insurance, risk, and market power in banking. *The American Economic Review*, (pp. 1183–1200).
- Kremer, M. (2000). Creating markets for new vaccines. part i: rationale. *Innovation policy and the economy*, 1, 35–72.
- Kroszner, R. S. & Strahan, P. E. (1999). What drives deregulation? economics and politics of the relaxation of bank branching restrictions. *The Quarterly Journal of Economics*, 114(4), 1437–1467.
- Krugman, P. (2009). Making banking boring. *New York Times*, 9, 2009.
- Lee, T. & Wilde, L. L. (1980). Market structure and innovation: A reformulation. *The Quarterly Journal of Economics*, 94(2), 429–436.
- Lerner, J. & Merges, R. P. (1998). The control of technology alliances: An empirical analysis

-
- of the biotechnology industry. *The Journal of Industrial Economics*, 46(2), 125–156.
- Lin, L. (2019). Bank deposits and the stock market. *Review of Financial Studies*, *Forthcoming*.
- Loury, G. C. (1979). Market structure and innovation. *The quarterly journal of economics*, (pp. 395–410).
- Magoni, M., Scarcella, C., Vassallo, F., Lonati, F., Carosi, G., Castelnovo, F., Quiros-Roldan, E., Albini, L., Gennaro, N., Bishai, D., et al. (2011). The evolving burden of hiv infection compared with other chronic diseases in northern italy. *HIV medicine*, 12(3), 129–137.
- Mulligan, C. B. & Sala-i Martin, X. (2000). Extensive margins and the demand for money at low interest rates. *Journal of political Economy*, 108(5), 961–991.
- Neumark, D. & Sharpe, S. A. (1992). Market structure and the nature of price rigidity: evidence from the market for consumer deposits. *The Quarterly Journal of Economics*, 107(2), 657–680.
- Perraudin, W. R. & Sørensen, B. E. (2000). The demand for risky assets: Sample selection and household portfolios. *Journal of Econometrics*, 97(1), 117–144.
- Philippon, T. & Reshef, A. (2012). Wages and human capital in the us finance industry: 1909–2006. *The Quarterly Journal of Economics*, 127(4), 1551–1609.
- Pless, J., Van Benthem, A., et al. (2017). *The surprising pass-through of solar subsidies*. Technical report, CEPR Discussion Papers.
- Ritz, R. (2015). The simple economics of asymmetric cost pass-through.
- Sharpe, S. A. (1990). Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *The journal of finance*, 45(4), 1069–1087.
- Sharpe, S. A. (1997). The effect of consumer switching costs on prices: A theory and its application to the bank deposit market. *Review of Industrial Organization*, 12(1), 79–94.
- Skaperdas, S. (1996). Contest success functions. *Economic theory*, 7(2), 283–290.
- Stein, J. C. & Sunderam, A. (2018). The fed, the bond market, and gradualism in monetary

-
- policy. *The Journal of Finance*, 73(3), 1015–1060.
- Sullivan, D. (1985). Testing hypotheses about firm behavior in the cigarette industry. *Journal of Political Economy*, 93(3), 586–598.
- Taylor, C. R. (1995). Digging for golden carrots: An analysis of research tournaments. *The American Economic Review*, (pp. 872–890).
- Tobin, J. (1956). The interest-elasticity of transactions demand for cash. *The review of Economics and Statistics*, (pp. 241–247).
- Tullock, G. (2001). Efficient rent seeking. In *Efficient Rent-Seeking* (pp. 3–16). Springer.
- Vissing-Jørgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4), 825–853.
- Weyl, E. G. & Fabinger, M. (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy*, 121(3), 528–583.
- Whited, T. M. & Wu, G. (2006). Financial constraints risk. *The Review of Financial Studies*, 19(2), 531–559.
- Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *Jama*, 323(9), 844–853.