Bank Incentives and the Effect of the Paycheck Protection Program [*]

Gustavo Joaquim [†]
FRB of Boston

Felipe Netto
Columbia University

January 8, 2021

Abstract

The Paycheck Protection Program (PPP) administered hundreds of billions of dollars of loans and grants to small business through private banks. In this paper, we explore the optimal allocation of funds across firms, the distortions caused by allocating these funds through the banking system, and how these distortions affect the empirical estimation of the program. First, we show that it can be optimal to allocate funds to the least or most affected firms depending on the nature of the shock, the firms' financial position and program design. Second, we show that banks distort the allocation towards firms with more pre-pandemic debt per-employee and a higher probability of survival without PPP funds, which is consistent with the empirical literature and significantly reduces the effectiveness of the program. Third, we show that bank incentives lead to a selection bias where the aggregate effect of the program is overestimated empirically. As bank incentives determine the set of compliers, we show that this selection bias cannot be fully addressed by an instrumental variable at the bank-level — as typically done in the empirical literature — even in the case of an exogenous bank shock. Our model thus provides a framework to understand the results found in the empirical literature, estimate the effect of the PPP, and perform counterfactuals.

[*]We thank Falk Bräuning, Joe Peek, Joanna Stavins, Adi Sunderam, Christina Wang, Paul Willen, Eric Zwick and participants of the Fama-French/Journal of Finance Conference on the Financial Consequences of the Covid-19 pandemic for valuable comments and discussions. We also thank Morgan Klaeser for outstanding research assistance. **The views expressed herein are those of the author and do not indicate concurrence by the Federal Reserve Bank of Boston, the principals of the Board of Governors, or the Federal Reserve System.** All errors are our own.

[†]Contact: gustavo.joaquim@bos.frb.org

1 Introduction

The COVID-19 pandemic has led to an unprecedented decrease in economic activity, affecting in particular small businesses (Bartik et al. (2020*a*)). In April of 2020, small business revenues decreased by more than 40% compared to January of the same year, and are still 20% down in August (Chetty et al. (2020)). As a response, Congress created the novel Paycheck Protection Program (PPP) as part of the larger CARES Act. The main goal of the program was to preserve jobs of small and medium business that were substantially affected by COVID-19. Over $650 billion dollars were allocated in the first two tranches of the program, almost two thirds of *all* interventions in the financial crisis (Bartik et al. (2020*b*)). To speed up the delivery of funds to businesses, the government used the private banking system to make decisions on applications. This gave banks the ability to target funds to preferred borrowers, particularly in the first tranche, where demand for funds overwhelmingly exceeded supply.[1]

In this paper, we develop a framework to understand the role of bank incentives in the allocation of PPP funds and how these incentives affect the estimation of the effects of the program. We focus on three main questions. First, we characterize the optimal allocation from the perspective of the government of PPP funds, that is, what is the target of the PPP. Second, we explore how banks distort the allocation of these funds or, alternatively, if the bank allocation hits the target. Finally, we explore how the distortion in the allocation of PPP through private banks impacts the estimation of the effect of the program at the firm and regional levels. Although we focus mostly on the PPP, we highlight the generalizable lessons from our paper to the design and evaluation of other lending programs.

The starting point of our theoretical framework are firms potentially in need of external finance to survive the pandemic shock. Firms have a fixed cash flow and debt level, and face a fixed cost shock to stay open as a result of the pandemic (as in Guerrieri et al. (2020)). As our focus is on the incentives of the banking system, we assume throughout most of the paper that firms, banks and government have the same information.[2] All agents know the parameters that characterize the distribution of the pandemic shock, but not the actual

[1] According to the Census Small Business Pulse Survey, by the last week of April, 2020, 75% of eligible firms requested financial assistance, but less than 40% received it.

[2] We relax this assumption in an extension of our model.

realization of each firm shock.[3] Firms must choose to apply and the amount they apply for in the program, limited to a multiple of their total payroll. Firms balance the trade-off between borrowing more — and thus increasing their debt burden in the future — with an increase in the probability of survival. Consistent with the empirical evidence, larger, more profitable, more affected by the pandemic and firms with a higher effect of receiving PPP are more likely to apply.

We characterize the optimal allocation of PPP funds in two steps. First, we focus on the optimal allocation of a government that maximizes the number of preserved jobs if it can choose how much to lend to each individual firm given the size of the program, which we denote as *constrained first best* allocation. The constrained first best allocation targets firms where each additional dollar's marginal effect on firm survival probability is higher. When the shock has a concave distribution, the optimal allocation of funds equalizes the marginal probability of survival across different firms. For firms equally likely to be affected by the pandemic, this implies that the government finds it optimal to allocate money toward firms with a lower cash-flow (e.g. with low cash-on-hand or high debt levels), as those are the ones most in need of funds. When firms are affected differently by the pandemic (due to different sector or regional exposure), the government finds it optimal to allocate funds in a inverted U-shaped pattern with respect to shock exposure, that is, firms intermediately affected will be targeted first. This implies that the government will not necessarily allocate funds to firms most affected by pandemic, as this will be the case only if the size of the program is sufficiently large.

In the second part of our analysis of the government's optimal allocation under PPP rules, we assume that the government must follow the same rules as the banks. Specifically, the government can only accept or reject applications for firms that apply instead of choosing the specific amount it lends to every firm in the economy. Contrary to the constrained first best, it is optimal to allocate money to firms with a worse financial position if, and only if, the shock is likely to be small. When the shock is relatively small, the government attempts to save firms with a low probability of survival (worse financial position), while for large shocks it focus on firms with a higher probability of survival (better financial position). Moreover, for any shock distribution, the government finds optimal to allocate funds to

[3] This is a reasonable assumption since the first tranche of PPP loans was introduced in March 27th, 2020, long before the full dimension of the pandemic was known. For example, a survey of small business in Bartik et al. (2020*a*) shows that there is substantial disagreement and uncertainty regarding the duration of crisis.

firms likely to be intermediately affected by the pandemic, as in the constrained first best case. As the government is constrained by the PPP rules, it must choose in the extensive margin firms with the highest *gain* in the probability of survival. Firms which are expected to be severely affected by the pandemic will likely shut down, while firms expected to be not affected by the shock have a higher probability of survival regardless of the allocation PPP funds.

Given our optimal allocation of PPP funds from the perspective of the government, we study the potential misallocation from deploying these funds through the private banking system. Banks have incentives that are different from the government for two reasons. First, banks have outstanding loans with firms that will default they do not survive the pandemic. As a result, the banks distort the allocation toward more indebted firms. Second, banks potentially lose clients if they reject their PPP applications and are concerned that some of the loans may not be forgiven.[4] As a result, banks distort the government allocation toward firms less affected by the pandemic than the government finds optimal. Both of these effects are consistent with the empirical findings in Bartik et al. (2020*b*), who show that approval rates are higher for less distressed firms and borrowers with larger existing debt. Together, these effects predict that the program will be systematically less effective than it would otherwise have been if funds were allocated through the government. We show that this misallocation problem is worse reduces the optimal size of the program and can affect the optimal maximum loan per firm.

In the final part of our paper, we study how bank incentives affect the empirical estimation of the effects of the PPP. We consider a general version of our model with bank and regional heterogeneity that mimics the data used in the empirical literature (e.g., Granja et al. (2020)). Within this general framework, we discuss what the estimation of the PPP effect recovers using OLS and an exogenous bank-level technological shock to instrument for the allocation of PPP loans. First, we show that conditioning on PPP application is not sufficient to control for selection bias of the PPP program, and both the firm and regional regressions using OLS conditional on PPP applications deliver biased results for causal parameters of interest. Furthermore, we show that the bias using firm level and regional level variation are different, and that is is possible that firm level results overestimate while re-

[4]This connects our theoretical results to empirical studies on the importance of bank relationships to the allocation decisions of banks, such as Li and Strahan (2020) and Amiram and Rabetti (2020). This idea is also discussed on this WSJ article.

gional level variation underestimates the average treatment effect of the program. Second, and more importantly, we show that when banks are exogenously different in their approval of PPP rates and there is no bank and regional heterogeneity beyond this technological factor, using bank approval rates as an instrument delivers the Local Average Treatment Effect (LATE) of the program. The LATE, however, is not the relevant parameter for evaluating the effectiveness of the program. Under some empirically founded conditions, we show that it will consistently overestimate the overall effect of the program. Finally, we show that a model that features a rich heterogeneity at the bank and regional level will generate additional channels through which bank intermediation of PPP funds affects the estimation of the effect of the program.

Literature Review. This paper joins the growing literature exploring the economic impact and policy response following the COVID-19 pandemic, in particular the impact of the PPP. Autor et al. (2020) and Chetty et al. (2020) use the 500 employee eligibility cut-off to run a difference-in-differences analysis at the firm level. Autor et al. (2020) finds that PPP increased employment at the firm level by 2 to 4.5%, which corresponds to an effect on aggregate employment between 1.4-3.2 millon jobs through the first week of June, 2020. Neilson, Humphries and Ulyssea (2020) focus on the informational differences among small and large firms in terms of terms of application and approval rates. Erel and Liebersohn (2020) shows that there is a significant level of substitutability between traditional banks and fintechs in the PPP. Chodorow-Reich et al. (2020) study differences in liquidity provision to small and larger firms, showing how the PPP ameliorated liquidity shortfalls experienced by SMEs, who have reduced access to credit lines relative to larger firms.

While several papers on the literature focus on firm level evidence from surveys or private data providers, Granja et al. (2020) takes a regional approach to understand the effect of the PPP. The authors show that the PPP funds did not flow to the areas most economically affected by the pandemic. Moreover, the authors show that there is a substantial heterogeneity at the bank level in terms of approval and disbursement, which gives rise to their identification strategy of using regional exposure to the PPP by weighting the relative disbursement of PPP banks at the national level with each bank's pre-pandemic market share in deposits by region. Using this strategy, Granja et al. (2020) find insignificant results in terms of employment and business shutdowns from the program.

Closest to our paper, Bartik et al. (2020*b*) empirically investigates the targeting of funds in the PPP through the banking system. Using an IV strategy that leverages previous bank-firm relationships and heterogeneity in PPP processing at the bank level, Bartik et al. (2020*b*) shows that applications from firms more distressed business (for instance, those with less cash on hand), were less likely to be approved, despite large effects of the PPP for this set of firms. Moreover, banks favored firms with closer relationships to the bank, rather than those in greater distress. A similar conclusion regarding lending relationships follows from Li and Strahan (2020) and Amiram and Rabetti (2020), who investigate the connection between bank-firm relationships and access to PPP funds.[5]

Our contribution to this literature is twofold. First, we propose a model of the PPP that allows us to characterize the optimal allocation of PPP funds from the government and understand the sources of misallocation of using the private banking system to disburse these funds. Our model is consistent with the empirical evidence of Bartik et al. (2020*b*), where banks distort the allocation towards less distressed firms and to those with previous relationships with banks. Importantly, our model shows that this distortion is not monotonic. For instance, in our framework banks prefer to allocate funds to firms more indebted when compared to the government allocation, but not monotonically to the firms with the highest levels of pre-pandemic debt. This suggests that the strategy of constructing dummies at the median of these variables adopted by Bartik et al. (2020*b*) can be underscoring the true magnitude of these channels.

Second, as Autor et al. (2020) stresses, understanding why the firm and regional level approaches yield different conclusions is an important line of research. Our model provides a unifying framework where bank incentives distort the allocation toward firms with a higher probability of survival *ex-ante*. If this probability is unobserved, we show analytically that OLS estimates will be biased due to these incentives. Furthermore, we also show that the use of an instrumental variables approach using market shares can, if the shares are plausibly exogenous, recover the local average treatment effects, but that these are not informative about the overall effect of the program.

The paper proceeds as follows. The next section describes the details and design of the PPP program and, more generally, the CARES act. In the next section we then focus on the model of the PPP, where we characterize the solution of the government and banking sector

[5]Other PPP focused studies include Hassan et al. (2020), Elenev, Landvoigt and Van Nieuwerburgh (2020), Faria-e Castro (2020), Cororaton and Rosen (2020) and Barrios et al. (2020)

optimal allocations. In the fourth section we discuss the implications of our model to the estimation of the PPP effect in firm and regional level data, as well as the bank IV identification strategy. Finally, the fifth section concludes the paper and points for directions to quantify the channels we discuss and the design of lending programs in a post-crisis world.

2 The Paycheck Protection Program

Created in early April 2020 as part of the CARES act, the Paycheck Protection Program (PPP) was designed to address liquidity shortages that could lead to employment losses from small businesses. The PPP is a loan which incorporates mechanisms to incentivize firms to keep their labor force, as it should be used to pay for employee wages and other fixed costs such as rent. Funds must be used to pay for these costs over the eight-week period following the provision of the loan. The allocation of funds for the program worked on a first come, first served basis. The initial provision of funds for the PPP was of $ 349 billion, and due to high demand, got depleted by April 16. Additional $300 billion were approved by late April to increase the support of the program.

The PPP is fully forgiven if funds are used for the specified purposes of employment maintenance. Originally, to obtain full loan forgiveness, businesses were required to use at least 75% of the amount on payroll expenses and to maintain employment headcount and wage levels. This percentage was reduced to 60% after the Flexibility Act was passed in June. Importantly, loan forgiveness would be reduced if wages decrease or full time headcount decreases.

Given its small business focus, only firms with less than 500 employees were eligible to apply,[6] and each firm could apply to at most one loan from the PPP program, with a maximum of $10 million for each loan or 2.5 times the firm's average monthly payroll costs. PPP loans have an interest rate of 1%, deferred payments for six months, and maturity of two years for loans issued in the first phase of the program and five years for loans issued after June 5. Moreover, PPP loans do not require collateral or personal guarantees.

Loans processing is performed by authorized banks, for example federally insured depository institutions and credit unions, which are responsible for checking documentation submitted by applicants, and are paid fees to cover these processing costs. Importantly,

[6]The exception were firms in restaurant and hospitality sectors (NAICS code 72), which were allowed to apply as long as they had at most 500 employees in each location.

loans from the PPP carry zero risk weight for the calculus of risk weighted assets, with the purpose of minimizing the impact on banks' capital requirements. Additionally, Federal Reserve Banks were authorized to provide liquidity to banks through the Paycheck Protection Program Lending Facility (PPPL Facility). This allows Federal Reserve Banks to extend loans for institutions which are eligible to make PPP loans using such loans as collateral. Overall, the program was designed to allow a large number of institutions to process loan requests while minimizing impacts on their balance sheet structure.

As a result of the facilitated access structure, the PPP attracted a substantial number of applications. As of the beginning of August, more than 5 million loans were granted, with a total amount of about $525 billion when the program closed and stopped receiving applications. In short, the PPP was an important liquidity tool for small firms with mechanisms that incentivize employment maintenance and broad access, facing substantial demand from borrowers since its creation.

More recently, the approval of the $900 Billion dollar stimulus package in late 2020 includes plans for another round of the PPP program.[7] Despite common characteristics with the first two rounds approved earlier in 2020, the next rounds is expected to require proof that businesses revenue was affected by the pandemic and has a smaller limit of $ 2 million per loan. Furthermore, the next round will have specific allocations to community lenders and first and second time borrowers that satisfy specific size requirements. Such features are designed to address the difficulties faced by small business that had no previous relationships with banks to access PPP funds. These challenges are captured our theoretical approach in the next section.

3 A Theoretical Model of the Lending Program

In this section we describe the setting of our model, discuss the optimal allocation of program's funds for the government and the potential misallocation of using the banking sector to allocate these funds.

[7] More information is available on this WSJ article

3.1 Firms

We consider a continuum of mass one of firms indexed by j. Each firm has N_j workers. We will define our model in terms of *per-worker* variables. Firm's j cash flow *per-worker* before the pandemic and the lending program is given by (1)

$$c_j \equiv \rho_j - b_j \tag{1}$$

where b_j are their debt payments per-worker to be made and ρ_j is the remainder of the cash flow per-worker (includes productivity, wages, cash-on-hand hand etc.). Without loss of generality, we normalize N_j such that $\int_j N_j dj = 1$. We assume that applying for the PPP has a fixed cost of F, and firms either choose to apply ($a_j = 1$) or not ($a_j = 0$) for the program. Each firm also chooses ω_j, the amount they apply for in the program per-worker, subject to a program limit based on their current employment level of φN_j.

We model the pandemic following Guerrieri et al. (2020). Each firm faces a reduction ν_j in cash flows (revenue shortfalls, extra costs to remain open). The per-worker magnitude of the shock is ν_j with c.d.f. denoted by Φ and p.d.f. ϕ parametrized by η_j (we define the specific functional form for the distribution below). A firm that borrows ω_j from the lending program *can* survive the pandemic if

$$\nu_j < c_j + \omega_j \equiv \Gamma_j(\omega_j)$$

where $\Gamma_j(\omega)$ corresponds to the available funds per employee to guarantee firm survival. We assume that $\Gamma_j(0) > 0$, $\forall j$, that is, all firms across all sectors and regions are profitable enough pre-pandemic to remain open.[8] A firm that borrows ω_j from the lending program *wants* to survive the pandemic if

$$\nu_j < c_j - r_G\omega_j + \pi_j^{LR} \equiv \Pi_j(\omega_j)$$

where π_j^{LR} is the perpetuity value of long-run profits of the firm (per-employee) and Π_j is the total profit of the firm (per-employee). We assume that all firms that *can* survive *want* to survive - that is, $\Gamma_j < \Pi_j$, $\forall j$. The motivation for this assumption is twofold. First, the

[8] As our focus is between the allocation of funds across firms, it is natural to assume that firms that are not profitable before the pandemic will shut-down and won't receive any funds from the program.

lending programs are designed as a short-term source of finance for these firms, such that it is expected that $\pi_j^{LR} > (1 + r_G)\varphi$ (which guarantees that all firms that can survive want to survive). Second, we are analyzing how to allocate funds among firms. Firms that do not want to survive won't be part of the program as is (that is, they don't want to survive even conditional on getting program funds) and thus won't participate in the optimal allocation regardless.

The problem of the firm is given by (2), where each firm chooses to apply or not for the program ($a_j \in \{0, 1\}$) and the amount to request from the program ($\omega \in [0, \varphi]$):

$$\max_{a \in \{0,1\}, \omega \in [0,\varphi]} \int_0^{\Gamma_j(a\omega)} N_j \left[\Pi_j(a\omega) - \nu\right] d\Phi(\nu\,;\,\eta_j) \tag{2}$$

In (2), we assume that the firm chooses ω_j before observing the realization of ν_j, which stands for the fact that the firm does not know the extent of the pandemic or of its own exposure to it ex-ante, but do know the distribution of shocks they can face. This is a reasonable assumption given the uncertainty in terms of depth and duration of pandemic. For instance, in a survey of over 5,800 small business, Bartik et al. (2020*a*) shows that there is substantial disagreement on the expected duration of the crisis across small business and the reported levels of confidence in their expected duration is low.

Moreover, as we want to highlight here the role of bank incentives, we assume in our benchmark model that banks and government also observe η_j (the parameter of the distribution), but not ν_j (the realization of the shock) in the main text. We solve the version where the government has less information and operational capacity than banks in Section 3.5.

For tractability, we follow Guerrieri et al. (2020) and assume that the c.d.f. of the fixed cost shock distribution is given by (3)

$$\Phi(\nu\,;\,\eta) = \begin{cases} 0, \ \text{if} \ \nu < 0 \\ \left(\frac{\nu}{c_0}\right)^{\eta}, \ \text{if} \ \nu \leq c_0 \\ 1, \ \text{if} \ \nu > c_0 \end{cases} \text{, with } \eta > 0 \tag{3}$$

The distribution in (3) has two characteristics that great simplify our analysis while still allowing us to focus on the difference between bank and government incentives. First, The shape parameter η controls the concavity of the c.d.f., and thus we have a monotonic p.d.f., which is increasing if $\eta > 1$ and decreasing if $\eta < 1$. Second, a distribution with higher η first order stochastically dominates a distribution with lower η, making it easier the comparison between more affected (higher η) and least affected (lower η) firms. It is worth noting that for $\eta < 1$, the distribution in (3) is a truncated Pareto distribution (and converges to Pareto as $c_0 \to \infty$).

For notation purposes, define a_j^*, ω_j^* as the solution to the problem in (4). Moreover, we simplify the notation by defining: $\Phi_j(\omega) \equiv \Phi\left(\Gamma_j(\omega)\,;\,\eta_j\right)$, which is the probability a firm survives the pandemic if it receives ω from the program.

The objective function of the firm can be rewritten as (4). The expected profit is given by the probability of survival (firm wants to and can survive) multiplied by the expected profit conditional of survival, subtracted of the application cost (if the firm chooses to apply):

$$\max_{\omega\in[0,\varphi],a\in\{0,1\}} \underbrace{\Phi_j(a\omega)}_{\text{Prob. Survival}} \cdot \underbrace{\left[\Pi_j(a\omega) - \mathbb{E}\left(\nu_j \mid \nu_j \leq \Gamma_j(a\omega)\right)\right]}_{\text{Expected Profit}} - aF \tag{4}$$

In (4), the problem of the firm is to balance borrowing to increase the probability of survival with reduced profitability in the future and the application cost. We can solve the problem in steps. First, consider a firm that has chosen to apply for the program. If the interest rate r_G is too high, then the firm does not want to borrow from the program, as Π_i^{LR} is decreasing and linear in r_G, and thus $\omega_j^* = 0$. On the other hand, if $r_G < 0$, as it is the case in the PPP given the implicit grants in the program, then borrowing increases the probability of survival *and* increases profits in the future, thus, conditional on applying, $\omega_i^* = \varphi$.

Second, a firm applies for the program if the benefits of applying are larger than the fixed cost F. Firms with more workers (N_j) apply more often given a smaller per-worker cost of applying. This is consistent with the survey evidence in Neilson, Humphries and Ulyssea (2020), who show that small business were less likely to be aware and apply for the PPP program. To analyze the firm benefits of the program, we introduce a key variable in our model. Let T_j be the treatment effect for firms of type j between receiving φ or 0 of loans from the PPP as in (5)

$$T_j \equiv \Phi_j(\varphi) - \Phi_j(0) \tag{5}$$

The variable T_j plays a special role as it measures the expected effectiveness of the PPP program for each type of firm j. All else equal, firm j is more likely to apply for the program if $(c_j + \pi_j^{LR})T_j$ is higher, that is, if the *increase* in expected profits is higher. However, this is not the only term that affects the benefits of applying for firms. Applying for the PPP increases the expected cost to be paid in terms of survival, that is, $\mathbb{E}\left(\nu_j \mid \nu_j \leq \Gamma_j(\varphi)\right) > \mathbb{E}\left(\nu_j \mid \nu_j \leq \Gamma_j(0)\right)$ and the loan will have to be repaid at r_G. Putting all together, firms that apply for the PPP are those that satisfy (6), which using our specific distribution can be written as (7). We define the set of all firms as $\mathcal{F}$ and the set of firms that apply as $\mathcal{A} \equiv \left\{j \mid a_j^* = 1\right\}$. First, we will analyze the optimal allocation of the government outside the rules of the program across all firm, $\mathcal{F}$. Then, we will focus on the differences between government and bank allocations under the rules of the program for firms that do apply, $\mathcal{A}$. In Section 4, we re-analyze the role of the set of applying firms for the estimation of the effect of the lending program empirically. All proofs and derivations are in the appendix.

Firm's Choice in the PPP. *If $r_G < 0$, then all firms apply for the maximum amount of PPP funds, that is, $\omega_j^* = \varphi$. Firms apply for the PPP ($a_j^* = 1$) if:*

$$T_j \Pi_j(0) - T_j \mathbb{E}\left[\nu_j \mid \nu_j \in [\Gamma_j(0), \Gamma_j(\varphi)]\right] - \Phi_j(\varphi) r_G \varphi > \frac{F}{N_j} \tag{6}$$

For the distribution in (7)

$$\left[\frac{1}{\eta_j + 1} c_j + \pi_j^{LR}\right] T_j - \Phi_j(\varphi)\left(\frac{\eta_j}{\eta_j + 1} + r_G\right)\varphi > \frac{F}{N_j} \tag{7}$$

3.2 Constrained First Best

Our first theoretical results is based on the problem of the government when it can choose the amount ω_j^G per-worker to lend to each firm. We assume that the objective of the government is to "provide a direct incentive for small businesses to keep their workers on the payroll", as stated in the SBA website.[9] Since there is no intensive margin adjustment at the firm level (such as downsizing), we model this objective as maximizing the number of surviving jobs. The government observes the types of firms j, but not their actual realization

[9]See here.

of the pandemic shock, so we denote the solution of this problem as the *Constrained First Best*, since the government is constrained by its information set. This assumption is realistic in the case of the PPP, where the government had to allocate funds quickly without fully understanding the depth and duration of the pandemic shock to small businesses.

We show that the government want to allocate funds to where their marginal effect is the highest, which does not necessarily corresponds to the places most/least affected by the shock. The marginal effect depends on the shock distribution, the size of the program, and the initial financial condition of the affected firms. The optimal target of the lending program thus is neither obvious nor invariant to the nature of the economic shock (as more information on the depth of the pandemic becomes available). For instance, if the shock for a firm (or region/sector) is large enough, the government does not always find optimal to save this firm as the opportunity cost of not allocating these funds for other firms is too high. Our analytical results below formalize this intuition and characterizes what is then the government optimal allocation.

We separately compare firms with the same shock exposure (η_j) but heterogeneous financial position first and then firms with the same financial position (c_j) and different shock exposures separately in this section and throughout the paper. The motivation for this is twofold. First, it highlights the key channels of the PPP allocation in our model for different sources of heterogeneity across firms. Second, it speaks more directly with the empirical literature which generally tries to control for either of these factors (with firms controls, fixed effects etc.) and focus solely on of them at a time (for instance, in Bartik et al. (2020*b*)).

We assume that the government observes for each firm j their cash-flow per-employee c_j and the number of workers N_j. The government does not observe the cost shock to each firm, ν_j, and thus can only choose the amount ω_j^G per-worker to lend to each firm based on the distribution of the shock.

For now we assume the amount of lending in the program, denoted by M, is fixed. footnoteWe return to the choice of M in Appendix C.3. The problem of the government is given by (8)

$$\max_{\{\omega_j^G\}} \int_{\mathcal{F}} N_j \cdot \Phi\left[\Gamma_j(\omega_j^G) \mid \eta_j\right] \; dj \;\; \text{s.t.} \quad \int_{\mathcal{F}} N_j \cdot \omega_j^G \, dj = M \tag{8}$$

We denote the solution to the maximization problem in (8) as the *Constrained First Best*

(CFB). In the constrained first best, we have that the solution is to allocate funds to where their marginal effect is higher, that is, where the *marginal* dollar will increase the probability of survival of firm j the most.

Our main analytical result of this section is Lemma 1, which considers the case where $\eta_j < 1, \ \ \forall j$, that is, all firms face a concave distribution of the pandemic shock. Define $\overline{M} \equiv M + \overline{\pi}$, which is the relative size of the lending program. The higher M is, more can be done within the program and the higher $\overline{\pi}$ is, the least needs to be done.

Lemma 1. ***Constrained First Best Allocation with*** $\eta_j < 1$**.** *The solution to (8) entails an equal gain in the probability of survival across firms, that is, for firms i, j*[10]

$$\phi\left[\Gamma\left(\omega_j^G\right)\right] = \phi\left[\Gamma\left(\omega_i^G\right)\right], \ \ \forall i, j \tag{9}$$

Using the distribution in (3), we have that

$$N_j \omega_j^{G,*} = N_j \tau(\eta_j, \overline{M}) + M - \left[N_j c_j - \overline{c}\right] \tag{10}$$

where $\overline{c} \equiv \int_j N_j c_j dj$ and $\tau(\eta, \overline{M})$ is a exposure-based per-worker transfer that sums to zero, that is, $\int_j N_j \tau(\eta_j, \overline{M}) = 0$. Furthermore, we have that:

- *$\overline{M}$ small: $\tau(\eta, \overline{M})$ is inverted U-shaped in $\eta \Rightarrow$ funds should flow to intermediately affected firms.*
- *$\overline{M}$ large: $\tau(\eta, \overline{M})$ is strictly increasing in $\eta \Rightarrow$ funds should flow to most affected firms.*

Lemma 1 implies that for two firms i, j with the same shock exposure $\eta_j = \eta_i$, we have that $N_j \omega_j^{G,*} - N_i \omega_i^{G,*} = N_i c_i - N_j c_j$, while for two firms with the same financial position $c_j = c_i$, we have that $N_j \omega_j^{G,*} - N_i \omega_i^{G,*} = N_j \tau(\eta_j, \overline{M}) - N_i \tau(\eta_i, \overline{M})$. Intuitively, Lemma 1 shows that (i) the optimal policy maximizes the marginal probability of survival across firms (eq. (9)) and (ii) this can be decomposed in the cash flow needs from firm j relative to the average cash flow needs in the economy and a transfer based on the size of the program and exposure to the shock (eq. (10)). Therefore, firms in a more fragile financial situation (as small firms) would receive more funds from the PPP. Moreover, if the relative size of the program is large, the government can allocate enough funds to the most affected firms to

[10] This is the interior solution to the problem of the government when c_0 is sufficiently large (e.g. $c_0 > M + \overline{\pi}$ is sufficient).

significantly increase their probability of survival. On the other hand, if the program is relatively small, the government must focus on firms that are intermediately affected by the pandemic. Firms that are strongly affected would cost too much to save, while the least affected firms can likely survive without PPP funds.

Sectoral/Regional Allocation. Our result in Lemma 1 is also useful to analyze the optimal allocation of funds across different sectors and regions in the country, as there is evidence that the funds did not flow to the most affected regions (Granja et al. (2020)). The result that optimal policy will equate the marginal probability of survival across firms in (9) is still true across sectors and regions. For instance, if different sectors have different initial levels of debt per-employee, sectors with relatively *more* debt per-employee should receive more of the funds since the probability that a firm of this sector survives the pandemic absent the government program is small, hence the marginal effect of funds on survival probability is large. However, if sectors or regions have shocks with a different distributions (that is, different exposures to the pandemic), the optimal transfers across sectors are those given by $\tau(\eta_j, \overline{M})$, and shouldn't necessarily go to the most affected sectors or regions.

Other cases. In Lemma 1, we focused in the case where $\eta_j < 1, \ \forall j$. In Appendix C.1, we show in a simple example that when $\eta_j > 1$, the problem of the government is convex and the solution will be to allocate funds to either the firms with the lowest or highest π or η, that is, the government is indifferent between allocating funds to the least or most affected firms, as long as all of the funds flow to either. More generally, take any distribution $\Upsilon\left(\pi_j + \omega_j \mid \theta_j\right)$ parametrized by θ_j. We show in Appendix C.2 that the government wants to allocate money to the highest θ_j (most affected) if Υ is supermodular in ω, θ (and to the least affected if it is submodular). Overall, this shows that the optimal target of PPP funds depends on the benefit of the marginal dollar rather than the funding needs of each individual firm.

3.3 Government's Optimal Allocation under the PPP Rules

In Section 3.2 we considered the government allocation when the government could choose how much to lend to every firm. To make a direct comparison with the bank allocation, we now focus on the optimal government allocation under the same rules as the banks in the PPP, conditional to the set of firms that in fact applied for the PPP, that is, $j \in \mathcal{A}$. In this case, governments can choose to accept or reject applications from firms, but cannot change the loan allocation in the intensive margin. This problem ensures that we are comparing the bank allocation with an equally constrained by the program government allocation (instead of the constrained first best), and thus that any difference comes from banks' incentives.

The problem of the government is to choose the probability $l_j^G \in [0,1]$ to accept the application from firm j, as in (11):

$$\max_{\{l_j^G \in [0,1]\}} \int_{\mathcal{A}} N_j \left[l_j^G \Phi_j^{\Gamma}(\varphi) + (1 - l_j^G) \Phi_j^{\Gamma}(0) \right] dj \ \text{ s.t. } \ \int_{\mathcal{A}} N_j l_j^G \, dj = \frac{M}{\varphi} \tag{11}$$

Distribution of firms/workers in the population. Note that so far we have used the short hand notation of dj to represent the integral over the distribution of firms, but we haven't defined how types of firms are present in the population. In what follows, it will be useful to spell out this integral. Let $G(\rho, b, \eta, N)$ be the joint distribution of ρ, b, η, N in the population of firms. For any variable at the firm level that is not a function on the number of employees, $x(\rho, b, \eta)$, we can write:

$$\int N x(\rho, b, \eta) dG(\rho, b, \eta, N) = \int x(\rho, b, \eta) \overline{N}(\rho, b, \eta) dG(\rho, b, \eta) \tag{12}$$

where $\overline{N}(\rho, b, \eta) \equiv \int_N N dG(N \mid \rho, b, \eta)$. The term $\overline{N}$ is the average number of employees of firms of a given type $\{\rho, b, \eta\}$, and it acts in our model as a shifter in the distribution of firms of type $\{\rho, b, \eta\}$. What matters in our model is not the marginal distribution of firms, but rather the marginal distribution of the firm variables at the *job* level. Thus, our model can encompass various other channels highlighted in the literature that focus on jobs and not firms. For instance, see the evidence on Bartlett and Morse (2020) of how firm survival varies by firm size.

As in our model the treatment effect T_j is not a function of N_j and the resource constraint is linear on it, we have that the optimal allocation l_j^G is also not a function of N_j where dj in this case represents the integration of all firms of types $\{b_j, \rho_j, \eta_j\}_j$ with the cummulative distribution of type j given by $G(\rho, b, \eta) \times \overline{N}(\rho, b, \eta)$. Following the argument in (12), we can write the problem of the government as (13)

$$\max_{\{l_j^G \in [0,1]\}} \int_{\mathcal{A}} l_j^G T_j \, dj \ \text{ s.t. } \ \int_{\mathcal{A}} l_j^G \, dj = \frac{M}{\varphi} \tag{13}$$

where the distribution of firms implicit in the integral dj is the job weighted distribution one, as in (12). The exact same argument can be made in everything that follows in this paper, so from now on, we leave the dependence on N_j implicit.

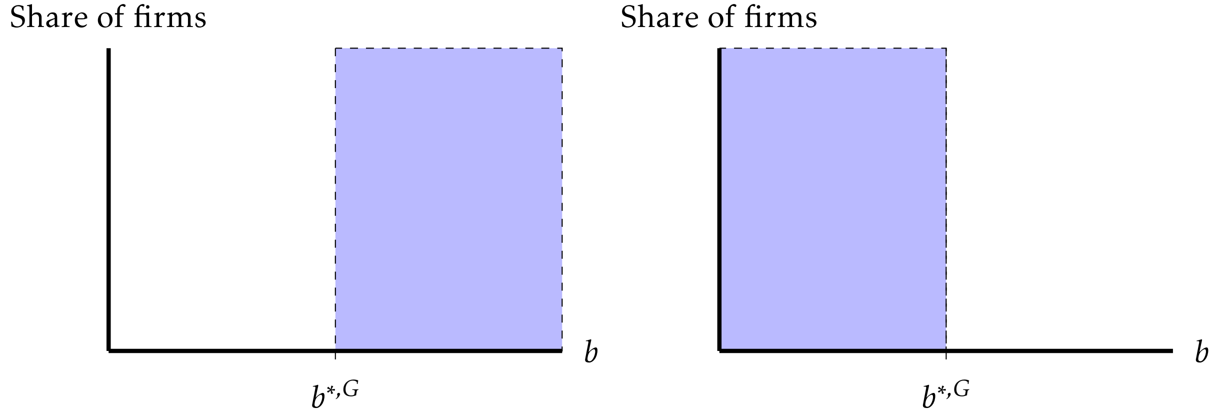
From (13), it is clear that the government wants to approve the application of firms with the *highest treatment effect*. The key question is which firms are those in the case of the PPP. In the first result of Lemma 2, we show that for firms with the same η, the government wants to allocate funds to firms with high levels of debt b_j (or low ρ_i) if $\eta < 1$, and to firms with low levels of debt b_i if $\eta > 1$. If the shock is likely to be relatively small ($\eta < 1$) the government can try to save the firms that have the lowest probability of survival, which are those with high levels of debt per-worker (*ceteris paribus*). On the other hand, for shocks that are most likely large, the government simply prefers to allocate the funds to firms with relatively low levels of debt, as those are the ones the government can still save in face of the pandemic. The insight here is that the treatment effects are a joint product of the firm financial position and the nature of the shock distribution, and thus do not have a distribution or model free ranking. This first result of Lemma 2 is illustrated in Figure 1.

Lemma 2. ***Government PPP allocation.***
Debt heterogeneity. *Consider that all firms in the economy are equal except for their level of debt b_j, that is, $c_j = c$ and $\eta_j = \eta$. The solution to (13) implies that $\exists! b^*$ such that: (i) for $\eta < 1$, $l_j^G = 1$ if $b_j > b^{*,G}$ and $l_j^G = 0$ otherwise, and (ii) the opposite for $\eta > 1$.*[11]
Shock Exposure Heterogeneity. *Consider that all firms are equal except for their shock exposure η_j, that is, $c_j = c$ and $b_j = b$. The solution to (13) implies that $\exists!\ \underline{\eta}_G, \overline{\eta}_G$ such that the government chooses $l_j^G = 1$ if $\eta_j \in [\underline{\eta}_G, \overline{\eta}_G]$ and $l_j^G = 0$ otherwise.*

In the second result of Lemma 2, we show that for firms with the same financial position c_j, the government wants to allocate funds to firms with intermediate exposure to the pandemic shock. This result has the same intuition as the case where M is small in the constrained first best. The most affected firms won't survive with the extra φ, while the least affected firms will likely survive regardless, such that φ is too much to allocate torwards them. Here, contrary to the constrained first best, this is not a function of the total size of the program, M, since the amount in the intensive margin the government can allocate to

[11] Note that $b^{*,G}$ is different if η is < or > than 1. Moreover, for $l_j^G = b^{*,G}$, the government is indifferent in terms of the allocation. As there is a continuum of firms, the allocation to this specific set of firms is irrelevant at the aggregate level.

Figure 1: Optimal Government Allocation under the PPP for Firms with Different Debt Per-worker b: $\eta < 1$ (left) vs $\eta > 1$ (right)



Note: This figure is the pictorial representation of Lemma 2. We denote b as the pre-pandemic debt per-worker of each firm.

each firm is fixed. Pictorially, this second result of Lemma 2 is in Figure 3, where we compare the optimal allocation of the government vs the private banking sector (Section 3.4). This result reinforces the idea that the optimal target of the PPP is not the most affected firms (or sectors and regions), since the return can be lower given the likely magnitude of the shock.

3.4 Banks' Allocation in the PPP program

We now focus on the private banking sector allocation in the PPP program. As in the government optimization problem (11), banks can choose to accept or reject applications from firms, but can't change the loan allocation in the intensive margin. The difference is that the banking sector is interested in maximizing their profits — and not firm and job survival. For now, we assume that there is a single representative bank. We relax this assumption in Section 4.

Banks receive positive profits from making more loans (and thus the constraint on the total amount available for the program will be binding). If the banks accepts the PPP application of a firm, there are two possible scenarios. If the firm survives, the bank recovers b_j of the current loan and a present value of ψ_F of potential future loans to this firm. If the firm does not survive, the bank receives a share $\delta \in (0,1)$ of their outstanding loans, that is, δb_j. If the bank rejects the PPP application, there are same two possible scenarios. However, we additionally assume that if the firm survives, the bank loses a share of ψ_C of future clients that switch bank providers. Additionally, to incorporate potential uncertainty regarding loan guarantees, we assume that with probability q the bank has to face the costs of the PPP loan, a concern for some banks in the pandemic.[12] Mathematically, the problem of the bank is given by (14):

$$\max_{\{l_j^B \in [0,1]\}} \int_{\mathcal{A}} \left\{ \Phi_j^{\Gamma}(\varphi) \left[1 + \frac{\psi_F}{b_j} \right] + \left[1 - \Phi_j^{\Gamma}(\varphi) \right] \left(\delta - q \frac{\varphi}{b_j} \right) \right\} b_j l_j^B \, dj \tag{14}$$
$$+ \int_{\mathcal{A}} \left\{ \Phi_j^{\Gamma}(0) \left[1 + (1-\psi_C) \frac{\psi_F}{b_j} \right] + \left[1 - \Phi_j^{\Gamma}(0) \right] \delta \right\} b_j (1 - l_j^B) \, dj$$
$$\text{s.t.} \int_{\mathcal{A}} l_j^B \, dj = \frac{M}{\varphi}$$

or, alternatively,

$$\max_{\{l_j^B \in [0,1]\}} \int_{\mathcal{A}} \Omega_j l_j^B \, dj \quad \text{s.t.} \quad \int_{\mathcal{A}} l_j^B \, dj = \frac{M}{\varphi} \tag{15}$$

where Ω_j is the profit of the bank for firms of type j (up to a constant), that is

$$\Omega_j \equiv T_j \left[b_j (1 - \delta) + \psi_F + q\varphi \right] + \Phi_j^{\Gamma}(0) \left[\psi_C \psi_F + q\varphi \right] \tag{16}$$

The misallocation in our setting comes exactly from $\Omega_j \neq T_j$, that is, the difference between the treatment effect and profits from allocating PPP funds to a given firm. In our setting, there are two channels through which profits of the banking sector deviate from the objective function of the government. We explore analytically these channels in Lemma 3. First, the banking sector already has an heterogeneous exposure to firms that they have outstanding loans and potential future loans to be made to this firm, which is captured by $b_j(1-\delta)$. Everything else equal, this implies that banks allocate funds with more pre-shock debt per employee than the government.[13] In comparison with the results in Lemma 2, this

[12] For instance, on March 31st of 2020, the Treasury and the SBA released guidelines for lenders, including one that said that administration said banks would need to verify some of the borrowers information for the loan to be elegible for forgiveness — see here. These guidelines were eventually reviewed several times, inducing even more uncertainty on lender.

[13] Note that under the conditions of the first part of Lemma 3, Ω_j is strictly increasing in b_j. However, if we take into account that b_j can also enter in the probability of survival without PPP funds, $\Phi_j(0)$, we can show that Ω_j is hump shaped in b_j. This means that banks want to allocate funds to firms with more debt compared to the government allocation, but not necessarily those firms with the highest levels of pre-pandemic debt, as the probability of survival for some of those is too small. We opt here for the simpler statement of Lemma 3 as it captures the channels we want to highlight and is consistent with the empirical evidence in Bartik et al.

bank incentive generates a misallocation whenever $\eta < 1$, since when $\eta > 1$ it is optimal for the government to allocate PPP funds to firms with the highest pre-pandemic debt levels.

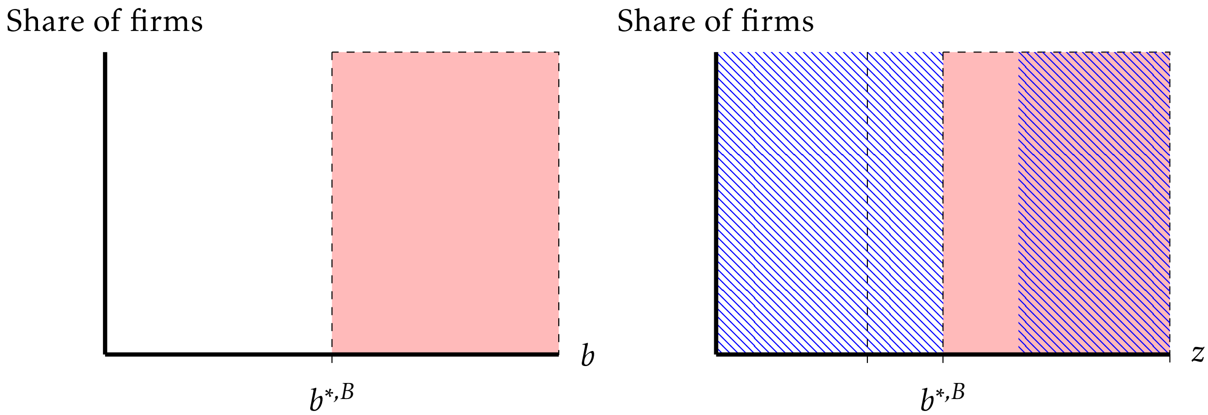
The pictorial representation of this result is in Figure 2. In the left panel ($\eta < 1$), there is no misallocation. In the right panel, the misallocation is given by the dotted area, where banks distort the allocation towards firms with more debt. The latter case is consistent with the empirical findings of Bartik et al. (2020*b*). In particular, Bartik et al. (2020*b*) shows that conditional on the set of firms with the relationship with a bank (in the extensive margin), banks approved more loans of firms with higher pre-existing debt at what the authors call a "striking magnitude".

Second, banks are also concerned about the probability of survival of the firm $\Phi_j^\Gamma(0)$, as those are potential clients to switch banks if they don't receive PPP loans or can have loans that are not forgiven to due to uncertain rules of the program. Everything else equal, this implies that banks allocate funds with a higher probability of survival without PPP funds. This incentive can be particularly perverse for the effectiveness of the program, since the firms that do receive funds are exactly those that could have survived regardless. In the second result of Lemma 3, we show that the banking sector distorts the optimal government allocation towards firms with lower η's and thus a higher probability of survival without PPP funds. Pictorially, this second result of Lemma 2 is in Figure 3. Intuitively, this result comes from the second term of Ω_j in (16), that is, the fact that banks also derive larger profits from firms that have a higher probability of survival ex-ante. This result is also consistent with the evidence in Bartik et al. (2020*b*), which finds that banks accept more applications from less distressed firms.

Lemma 3. ***Banks' PPP allocation.***
Debt heterogeneity. *Consider that all firms in the economy are the same except for their level of debt b_j, that is, $c_j = c$ and $\eta_j = \eta$. The solution to (15) is such that banks give preference to firms with higher b_j, that is, $l_j^{*,B} = 1$ if $b > b^{*,B}$.*
Shock Exposure Heterogeneity. *Consider that all firms are the same except for their exposure η_j. The solution to (15) implies that $\exists!\ \underline{\eta}_B, \overline{\eta}^B$ such that the bank chooses $l_j^{*,B} = 1$ if $\eta_j \in [\underline{\eta}, \overline{\eta}]$ and $l_j^{*,B} = 0$ otherwise. Additionally, $\underline{\eta}_B < \underline{\eta}_G$ and $\overline{\eta}_B < \overline{\eta}_G$, that is, bank's distort the allocation to firms with a higher probability of survival without PPP funds.*

(2020*b*).

Figure 2: Bank Allocation and Misallocation under the PPP rules: $\eta < 1$ (left) vs $\eta > 1$ (right)



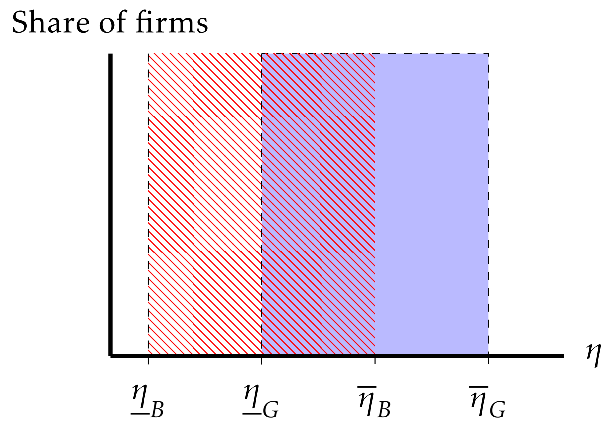
Note: The shaded area in the RHS figure denotes the misallocation, i.e., the difference from the government to the bank solution. There is no misallocation in the case where $\eta < 1$(Lemma 3).

Misallocation and Program Design. In Appendix C.3 we show how bank incentives affect the program design in terms of program size M and loan-size per firm φ. Given a cost function for the program, the lower effectiveness of the program reduces its optimal size when funds are allocated through the banking system. The optimal choice of φ is more nuanced. When firms are only heterogeneous in their debt levels and $\eta > 1$ (i.e., there is misallocation), banks can allocate even more funds to the highest indebted firms. However, misallocation is not a function of φ, since changes in φ do not affect the relative treatment effects of firms that receive funds in the bank and government allocations. Therefore, the optimal φ is the same in both cases. On the other hand, when firms are heterogeneous in their shock exposure, the effect is ambiguous. A higher φ changes the allocation at the extensive increasing misallocation (more funds are now channelled to a different set of firms), but it can be the case that the treatment effects of these firms gets closer, and the misallocation is reduced.

3.5 Differential Information and Operational Capacity

Although our main focus in on bank incentives, in this section we extend our model to include differential information and operational capacity between the government and the banking sector. In the benchmark version of our model (Section 3), as banks' and the government have the same information and ability to disburse funds, there is no reason for the government to intermediate the allocation through the private banking system. We thus augment our model in a way where banks' have more information and can disperse funds

Figure 3: Credit Allocation under the PPP for firms with Heterogeneous Shock Exposure (η_j): Government (blue, solid) vs Banks (red, dotted)



Note: This figure is the pictorial representation of Lemmas 2 and 3. The blue solid rectangle is the government allocation. The red, dotted allocation is the one in the banking sector. The η_G's are the lower and upper bound of the regions/sectors for the government, and η_B's for the banking sector.

more quickly than the government, the reasoning behind the use of the banking system in the program (Bartik et al. (2020*b*)).

First, we introduce the differential information between banks and the government. We assume that the government observes the true treatment effect for a share of firms, and a random one for the remaining firms. Importantly, we do not need to say which part of the terms of the treatment effect the government does not observe (i.e., if it is the financial position of the firm or the probability of survival), and simply that the government does not observe the true treatment effect, which is a sufficient statistic for the allocation in our setting. Mathematically, let $H(T)$ denote the distribution of treatment effects T_j in the population. Suppose that the government observes for every firm j a signal $\hat{T}_j$ of the true treatment effect T_j. The signal is given by $\hat{T}_j = T_j$ for a share μ of the population and $\tilde{T}_j$ for a share $1-\mu$, where $\tilde{T}_j$ is independent of T_j, but has the same distribution. Note that both $\hat{T}_j$ and T_j have the same distribution, which we denote by H.

Second, we introduce the operational capacity distortion. We assume that until the government can create a disbursement system, a share $1-\iota$ of firms will have faced their shocks ν_j (independently of their signals) such that their treatment effects will be zero, that is: either the firm has already survived with no PPP funds or the firm has not survived.

Let $H_{j|s}(j|s)$ the distribution of T_j conditional on signal T_s (i.e., $\hat{T}_j = T_s$). The government now chooses an allocation l_s^G for each signal s. Let the expected average treatment effect of firms with signal s, T_s, be given by:

$$ETE_s \equiv \int_j \iota T_j dH_{j|s}(j|s)$$

Our information structure implies that $ETE_s = (1-\mu)\iota\bar{T} + \mu\iota T_s$. Therefore, the problem of the government can be written as:

$$\max_{\{l_s^G\}} \int_s T_s l_s^G dH(s) \text{ s.t. } \int_s l_s^G dH(s) = \frac{M}{\varphi} \tag{17}$$

Note that the problem in (17) is the same as (13). Therefore, the allocation of firms with signal T_s is the same as the allocation for type j with treatment effect T_j in Lemma 3. Our main result of this section is Lemma 4. We show that that there is a decreasing curve in the μ — ι space such that if the government has sufficient information (high μ) or the delay cost is sufficiently low (high ι), then the government prefers to allocate PPP funds itself. In a future version of this paper, we will compute this region using a calibrated version of the model.

Lemma 4. ***Delegation vs Misallocation.*** *There exists a differentiable decreasing curve $\iota^*(\mu)$ in the μ — ι space such that for any combination of $\{\mu, \iota\}$ s.t. $\iota < \iota^*(\mu)$, the government prefers to allocate PPP funds through banks.*

4 Bank and Regional Heterogeneity

We have so far developed a model with a representative bank and region and focused on the different incentives between this bank and the government in the allocation of PPP funds. Based on the empirical evidence of Granja et al. (2020), we now consider the role of bank and regional heterogeneity and how it affects the PPP allocation and estimation of the effect of the PPP program. We present a generalized version of the bank problem that encompasses the representative bank framework we discussed before. Within this general framework, we discuss what the estimation of the PPP effect recovers using OLS and an exogenous bank-level technological shock to instrument for the allocation of PPP loans.

First, we show that conditioning on PPP application is not sufficient to control for selection bias of the PPP program, and both the firm and regional regressions using OLS conditional on PPP applications deliver biased results for causal parameters of interest. Furthermore, we show that the bias using firm level and regional level variation are different, and that is is possible that firm level results overestimate while regional level variation underestimates the average treatment effect of the program. Second, and more importantly, we show that when banks are exogenously different in their approval of PPP rates and there is no bank and regional heterogeneity beyond this technological factor, using bank approval rates as an instrument delivers the Local Average Treatment Effect (LATE) of the program. The LATE, however, is not the relevant parameter for evaluating the effectiveness of the program. Under some empirically founded conditions, we show that it will consistently overestimate the overall effect of the program. Finally, we show that a model that features a rich heterogeneity at the bank and regional level will generate additional channels through which bank intermediation of PPP funds affects the estimation of the effect of the program.

4.1 Generalized Bank Problem

Before proceeding with our results, we need to introduce additional notation in our model. Suppose that we have banks $b = 1, ..., B$. A bank b has a market share $\mu_{b,r}$ in region $r = 1, ..., R$. Each region has a weight W_r, $\sum_r W_r = 1$. Let the market share of bank b be given by $\mu_b \equiv \sum_r W_r \mu_{b,r}$. We model the bank level shock as a shifter $s_b > 0$ of the expected volume of small business loans this bank makes. That is, if the market share of bank b in small business loans is given by μ_b, we assume that bank b makes a share $s_b \mu_b$ of small business loans, and scale s_b to be such that $\sum_b s_b \mu_b = 1$.

We assume for simplicity that each firm has only one bank,[14] and thus that we can use the approval rate of this bank as an instrument for PPP approval at the firm level. Let b_j denote this bank for firm j, and r_j the region of this firm. The regional instrument we construct for PPP is given by the shift-share instrument s_r is the baseline market-share weighted average of s_b

$$s_r \equiv \sum_b \mu_{b,r} s_b \tag{18}$$

Moreover, let $p_{j,b,r}$ be the probability a firm of type j receives a loan from bank b in region r, the expected probability of survival of firms that use bank b in region r be given by $\theta_{b,r} \equiv \mathbb{E}\left[\theta_j | b, r\right]$ and the average treatment effect of these firms be given by $T_{b,r} \equiv \mathbb{E}[T_j | PPP_j =$

[14]Citation for US here

$1, b, r]$. For all variables $X_{b,r}$ denoted at the bank-regional level, let $X_r \equiv \sum_b \mu_{b,r} X_{b,r}$ and $X_b \equiv \sum_r W_r \frac{\mu_{b,r}}{\mu_b} X_{b,r}$ be these variables aggregated at, respectively, the regional and bank levels using the appropriate weights. Finally, we denote averages across all firms to be given by $\overline{x}$. For instance, $\overline{p}$ is the share of all firms that receive PPP funds in the economy.

Our true model of the effect of the PPP at either the firm or job levels can be written as (19)

$$y_j = \theta_j + T_j PPP_j + \varepsilon_j \tag{19}$$

where y_j is a dummy that is equal to one if the firm/job survives and PPP_j is a dummy if firm j receives PPP funds, θ_j is the probability of the firm surviving without PPP funds, that is, $\theta_j \equiv \Phi^j(0)$, T_j is the treatment effect $T_j \equiv \Phi^j(\varphi) - \Phi^j(0)$, and ε_j is a firm/job true idiosyncratic shock, with $\varepsilon_j \perp\!\!\!\perp \theta_j, PPP_j, T_j$. At the regional level, the true model can be written as

$$y_r = \theta_r + T_r PPP_r + \varepsilon_r \tag{20}$$

where y_r is the share of surviving business in region r, θ_r is the average probability of survival without intervention, T_r is the regional ATT and ε_r is the regional level shock.

When bank b makes a PPP loan to firm j in region r, it has an expected payoff of $\Omega(T_j, \theta_j, b_j, \zeta_b, \zeta_r)$. This framework encompasses the problem of the bank we consider in, but adds the possibility that all of the parameters (ψ_C, ψ_F, δ) vary both at the bank *and* regional levels. Therefore, this takes into account banks that have different abilities to recover collateral (different δ's), banks that can capitalize different of client's in the future (different ψ_F's) etc.. Let $G_{b,r}$ be the distribution of types of firms bank b faces in region r.

The general version of the problem of bank b we consider is given by maximizing (21) s.t. (22) by choosing a probability of accepting a PPP application $l_{j,r} \in [0,1]$ of firm j in region r.

$$\max_{\{l^b_{j,r}\}_{j,r}} \sum_r \mu_{b,r} W_r \int_j l^b_{j,r} \Omega(T_j, \theta_j, b_j, \zeta_b, \zeta_r)\, dG_{b,r}(j) \tag{21}$$

$$\sum_r \mu_{b,r} W_r \int_j l^b_{j,r} dG_{b,r}(j) = s_b \mu_b \frac{M}{\varphi} \tag{22}$$

The three differences between this problem and the one in Section 14 are: (i) the payoff per firm (which we leave as a general $\Omega(.)$ function for now), (ii) the constraint that is given by s_b times the expected market share of these banks in the small business lending market, and (iii) the bank-regional distribution $G_{b,r}$ of firms.

4.2 OLS Estimation of Causal Effect

Consider the estimation of the following firm-level regression by OLS

$$y_j = \beta_0 + \beta_F PPP_j + \epsilon_j \tag{23}$$

In our exercise, we don't need to include region-industry-week fixed effects, firm fixed effects and controls as we assume here we are comparing firms within the same sector-region-time and that are equal in everything other than their θ_j, T_j and PPP_j allocation.[15]

From (19), the OLS estimation of β_F in (23) delivers $\hat{\beta}_{F,OLS} \to ATT + B_{F,OLS}$, where the bias $B_{F,OLS}$ is given by (for the derivations of this section, see Appendix A.2)

$$B_{F,OLS} \equiv \frac{\text{Cov}(\theta_j, PPP_j)}{\mathbb{V}(PPP_j)} = \frac{\mathbb{E}(\theta_j | PPP_j = 1) - \overline{\theta}}{1 - \overline{p}} \tag{24}$$

The potential bias from firm level regression is a selection bias. In our theoretical framework, *even when conditioning of firms that apply*, firms that receive PPP loans are different from those that don't as banks target firms with higher treatment effects T_j and probability of survival ex-ante θ_j. In this case, firms that receive PPP will more likely survive because they receive PPP *and* because they would be more likely to survive regardless. The second representation makes this clear: the bias comes from a different probability of survival without PPP funds of firms that receive PPP funds versus the average of all firms in the population. This targeting channel is consistent with the empirical evidence in Bartik et al. (2020*b*), which shows that banks allocate PPP funds to firms less affected by the pandemic and with more cash on hand (even when those firms apply at lower rates).

Importantly, note that this effect is *not* specific to bank intermediation, but to any intermediation based on treatment effects. If the PPP was being allocated by the government, there would still be targeting of firms with higher treatment effects, which are not firms with the same probability of survival without the effect of the PPP.

[15] That is, we assume that all firms/jobs are alive at period zero and we are simply running a regression of period one variables where we observe differences across firms/regions — we don't have to run a difference-in-differences as our pre-period is equal across treatment and control groups.

Example 1: Firm-Level Selection Bias. We provide a simple example of the bias from targeting of PPP funds in (24). Consider a region where there are two types of firms, A and B, each consisting of half of the population. We assume that all firms have a treatment effect $T_A = T_B = T$. Firms of type A survive with probability $\theta_A = 1-T$ without PPP funds, while of firms of type B survive with $\theta_B = 0$ without PPP funds. In this setting, the $ATE = ATT = T$. Moreover, suppose that banks have enough funds to allocate to half of the firms. Since firms A have a higher θ_A, they will be targeted by banks to receive PPP funds. In this case, an OLS regression of the effects of PPP will deliver $\beta_{F,OLS} = 1$ since all firms that receive are of type A and will survive, and all firms that don't receive PPP of type B and won't survive. ∎

Similarly, the OLS estimation of β_R in (25)

$$y_r = \beta_0 + \beta_R PPP_r + \epsilon_r \tag{25}$$

using W_r as weights delivers $\hat{\beta}^{OLS,R} \to ATT + B_{R,OLS}$, where the bias $B_{R,OLS}$ is given by

$$B_{R,OLS} \equiv \frac{\text{Cov}_w(\theta_r, PPP_r)}{\mathbb{V}_w(PPP_r)} + \frac{\text{Cov}_w(T_r, PPP_r^2)}{\mathbb{V}_w(PPP_r)} \tag{26}$$

where the subscript w denotes a moment computed using W_r as weights. At the regional level, we have the same selection bias that we observe at the firm level given by the correlation of θ_r and PPP_r. In our framework, places with firms more likely to survive will be target by banks and thus will receive more PPP funds. There is, however, an additional term that relates the local average treatment effect T_r and the allocation of PPP funds to that region. When a region has more PPP, it will also endogenously change T_r as banks will target a different set of firms.

Example 2: Firm vs Regional Level bias. We provide a simple example of the bias from targeting of PPP funds at the firm level – (24) – and at the regional level – (26). Consider that there are multiple regions identical to our region of the firm-level example. However, now some regions exogenously have more PPP funds. In particular, half of the regions have PPP funds for 25% of firms and half for 75% of firms (for instance, each has a single market bank that has a different s_b). Since firms A have a higher θ_A, they will be targeted by banks to receive PPP funds. In regions with 25% of firms receiving, only type A firms will receive PPP funds. In regions with 75% of firms receiving, all type A and half of type B firms will receive PPP funds. In this case, $\beta_{R,OLS} = T$. To see that, note that the OLS estimator will compare the difference in surviving firms with the difference of PPP allocation, that is

$$\beta_{R,OLS} = \frac{1/2 + 1/4T - 1/4 - 1/4(1-T)}{3/4 - 1/4} = T = ATT$$

The regional regression works in this example since we are assuming that PPP is not correlated with either $\theta_{A/B}$ or the treatment effects, T, that are constant. However, consider now that the same bank is allocating funds across regions. The bank has enough funds to allocate to 1/2 of firms overall. In this case, the bank will allocate funds to type A firms only, and we will have the same selection bias we did from the firm-level example and get $\beta_{R,OLS} = 1$. This distinction is important in practice: how much banks are strategically choosing firms and thus moving funds around to different regions is related to the profitability of firms in each region, which can then induce the selection bias.

Consider now a second example. In this second example, we change the source of the bias from the selection bias (the covariance of PPP and θ's) to the endogeous effect of PPP on the ATT (second term in (26)). Suppose that there are multiple regions, each region has firms of type C and D, each consisting of half of the population. We assume that all firms will not survive without PPP funds, that is, $\theta_C = \theta_D = 0$. Firms of type C however have a treatment effect of larger than those of type D, $T_C > T_D$. Half of the regions have PPP funds for 50% of firms and half for 100% of firms. Due to bank targetting firms with higher treatment effects, in regions with 100% of firms receiving, all type C and D firms will receive PPP funds. In regions with 50% of firms receiving PPP funds, only firms of type C and none of the firms of type D receive funds. In this setting, we have that $ATE = \frac{1}{2}T_C + \frac{1}{2}T_D$ and $ATT = \frac{2}{3}T_C + \frac{1}{3}T_D$. The estimation at the regional level will imply

$$\beta_{R,OLS} = \frac{\frac{1}{2}T_C + \frac{1}{2}T_D - \frac{1}{2}T_C}{1 - 1/2} = T_D < ATE < ATT$$

The intuition here is that the treatment effect on the region with more PPP will endogenously be smaller (i.e., eventually funds will flow to firms with type D), since banks are targeting firms with higher treatment effects. In this case, there is a covariance between PPP allocation and the ATT at the regional level — the second term in (26). ■

In practice, both biases are likely true: firms that receive PPP are more likely to survive and there is regional targeting that endogenously affects the estimation of the the effect of the program. Our model shows that these two approaches — firm and regional — do not necessarily deliver the same estimation and one of them can superestimate while the other underestimates the true effect. To deal with this selection bias, Bartik et al. (2020*b*) and Granja et al. (2020) use the bank level shocks as instruments. We focus on the IV estimation using s_b as an instrument in the next section.

4.3 IV Estimation of Causal Effect

In section 4.2, we showed that even conditioning on application, there is a selection bias that affects the OLS estimation as banks target firms based on their treatment effect and *ex-ante* probability of survival. In this section, we explore the use a bank-level instrument, the shock s_b. The idea is that firms/regions that depend on banks with high s_b will have more access to PPP funds, but that this shock is potentially exogenous to other characteristics of these firms/regions. We consider first the case of firm-level regressions. The estimation of

$$y_j = \beta_0 + \beta_F PPP_j + \epsilon_j \tag{27}$$

using the approval rate of the bank firm j uses, s_{b_j}, as an instrument for PPP_j. Let $\mathbb{E}_\mu(.)$ denote the weighted expectation of any variable using bank market shares μ_b, and let $\mathbb{V}_\mu$ and $\text{Cov}_\mu(.,.)$ represent the analogous definition for variance and covariance, respectively. From (19), we have that the IV estimation of β^F in (23) delivers $\hat{\beta}^{IV,F} \to ATT + B_{F,IV}$, where the term $B_{F,IV}$ is given by

$$B_{F,IV} = \frac{\varphi}{M}\frac{\text{Cov}_\mu(\theta_b, s_b)}{\mathbb{V}_\mu(s_b)} + \frac{\text{Cov}_\mu(T_b, s_b^2)}{\mathbb{V}_\mu(s_b)} \tag{28}$$

If banks with different s_b's are consistently different in either their pool of applicants or their treatment effects on the firms each banks chooses to treat, the variation at the bank level does not yield the effect of the program. For instance, Li and Strahan (2020) document that smaller, relationship focused banks expanded credit more at the start of the program. Since the pool of applicants of these smaller banks are likely specific in certain characteristics, this might lead to biased estimates.

In our framework, there are several channels through which bank heterogeneity can induce both terms in (28) to be different from zero. We provide three examples. First, if large banks, which empirically have lower s_b's, also have a higher application cost F_b, these banks will face a consistently different pool of firms — those with higher treatment effects and lower probabilities of survival without PPP funds (Eq. 6). Second, if large banks also generally have larger clients, these banks will different pool of firms (since the application cost per worker is smaller). Additionally, if these larger firms are less exposed to the pandemic (have lower η's), the treatment effects at the bank level will also be correlated with s_b. Third, if larger banks also invest less in relationship banking and risk losing more clients (higher ψ_C), they will systematically choose firms with lower treatment effects, but with a higher probability of survival without PPP funds. In all of these examples, bank-heterogeneity that is correlated with s_b's (Which is empirically correlated with bank size) induces a systematic correlation between the instrument s_b and the outcome that is not through PPP allocation, that is, the instrument violates the exogeneity condition.

More importantly, if s_b is purely an exogenous technological shock, the IV strategy only partially solves the selection problem highlighted in the previous section, that is, we replace the covariance between the allocation of PPP funds with the probability of survival ex-ante with the covariance between the shock and this probability. This strategy introduces, however, an additional term that captures the effect that banks with higher s_b's, which will lend to more firms, will potentially lend to firms with different treatment effects. We explore this channel in detail in section 4.4, and show that using s_b as an instrument will lead to an underestimation of the effect of the program.

Similarly, using the regional exposure variation as an instrument, that is, s_r in (18) (and either running the firm-level or weighted regional regressions), we have that $\hat{\beta}_{R,IV} \to ATT + B_{R,IV}$, where the term $B_{R,IV}$ is given by

$$B_{R,IV} = \frac{\text{Cov}_w(\theta_r, s_r)}{\text{Cov}_w(PPP_r, s_r)} + \frac{\text{Cov}_w(T_r, PPP_r s_r)}{\text{Cov}_w(PPP_r, s_r)} \tag{29}$$

Equation (29) is a generalization of (28), where we don't assume necessarily a mechanic relationship for PPP_r and s_r, since multi-market banks can engage in regional targeting of PPP funds and thus endogenously affect the treatment effect on regions that receive more PPP loans.

Example 3: ATT estimation when s_b is a technological shock. Consider multiple regions where there are two types of firms, E and F, each consisting of half of the population. Firms of type E have a treatment effect of T_E, larger than the treatment effect of firms of type F, T_F, i.e., $T_E > T_F$. Each region has a different bank that serves only that region. For an exogenous reason, half of the regions have banks that can provide PPP funds for 50% of firms, while the other half has banks that can provide PPP funds for 100% of firms. Consider the case where banks target firms with higher treatment effects: we have that at regions with 50% of firms receiving PPP funds, all of E and none of F firms will receive PPP funds. Note that the effect of the PPP, the ATT, is given by:

$$ATT = \frac{1}{2}\left[\frac{1}{2}T_E + \frac{1}{2}T_F\right] + \frac{1}{2}T_E = \frac{3}{4}T_E + \frac{1}{4}T_F$$

while the $ATE = \frac{1}{2}T_E + \frac{1}{2}T_F$. Using s_b as an instrument for PPP_j in a regression with survival as the dependent variable yields:

$$\beta_{IV} = \frac{\frac{1}{2}T_E + \frac{1}{2}T_F - \frac{1}{2}T_E}{1 - \frac{1}{2}} = T_F < ATE < ATT$$

Alternatively, in the case where banks target firms with smaller treatment effects ($T_E < T_F$), we will have that $\beta_{IV} > ATT$, that is, we are potentially overestimating the effect of the program. The same is true in this case for the estimation at the regional level, since each region has a different bank. ■

4.4 Exogenous Bank PPP Capacity

We maintain our focus on the bank heterogeneous disbursement of PPP loans, s_b, we now assume that this shock is strictly exogenous, and is not correlated with any bank, region or firm observables and unobservables. Apart from this heterogeneity, all banks have similar market shares, face similar distribution of firms, and all regions are identical. Our goal is to prove our main result of this section: the bank-IV strategy can estimate a causal effect of the PPP, but will not estimate the effectiveness of the program or indicate how effective would the program had been had the government randomly allocated funds across firms or allocated the money itself. In particular, we show that the bank-IV strategy will estimate the Local Average Treatment Effect (LATE) of the PPP, but that this treatment effect is different from the effect of the program. Under some conditions on the support of firms, we argue that the effect of the PPP across all firms that received PPP funds is underestimated empirically.

The problem of the bank with heterogeneous disbursement can be written as maximizing (30) subject to (31)

$$\max_{\{l_j^b\}} \int_j l_j^b \, \Omega(T_j, \theta_j, b_j) dG(j) \tag{30}$$

subject to

$$\int_j l_j^b dG(j) = s_b \mu_b \frac{M}{\varphi} \tag{31}$$

where we simplify the Generalized problem in (21) to leave simply the s_b as a source of bank-heterogeneity. As in Lemma 3, the problem of the bank is linear in $\Omega(.)$ and thus the allocation will be given by $l_j^b = 1$ if, and only if, $\Omega(T_j, \theta_j, b_j) > \underline{\Omega}$, where $\underline{\Omega}$ solves

$$\int_j \mathbb{1}_{\Omega(T_j, \theta_j, b_j) > \underline{\Omega}_b} dG(j) = s_b \frac{M}{\varphi} \tag{32}$$

that is $\mathbb{P}[\Omega(\theta_j, T_j, b_j) > \underline{\Omega}_b] = s_b \frac{M}{\varphi}$. Therefore, a higher s_b implies a lower $\underline{\Omega}_b$, which expands the set of firms j such that $\Omega(\theta_j, T_j, b_j) > \underline{\Omega}_b$ and thus create a relation between the probability of survival without PPP funds θ_j and treatment effect T_j for firms that do receive PPP funds. A heterogeneous PPP capacity s_b will endogenously create a relationship between s_b and the treatment effect of firms that receive PPP from this bank, thus making the bank-IV coefficient drift away from the effect of the program on the firms that do receive PPP funds — the ATT. The same intuition applies for regional level regressions. We state this result formally in Lemma 5.

Lemma 5. ***Heterogenous PPP capacity and bank IV.*** *Even when the only source of heterogeneity across banks is the technological PPP capacity shock s_b, we have that estimating $y_j = \beta_0 + \beta_F PPP_j + \epsilon_j$ with s_{b_j} as an instrument for PPP_j will almost surely lead to an incorrect estimation of the effect of the program, that is, $B_{F,IV} \neq 0$, where $B_{F,V}$ is defined in (28). Similarly, for regions that are similar except for their exposure to each bank (heterogeneous $\mu_{b,r}$),[16] we have $B_{R,IV} \neq 0$ where $B_{R,IV}$ is defined in (29).*

If the shock s_b is in fact purely technological, as we assume in this section, why are we not able to estimate the effect of the program using a bank-IV strategy? We descriptively answer this question for the continuous IV case (using either s_b or s_r) now, but provide

[16] Which is what guarantees a strong first stage, as empirically observed in Granja et al. (2020).

a much simpler version of the result in the next section, where we focus on a two bank case. As shown in Angrist, Graddy and Imbens (2000) for the continuous IV case, the bank-IV strategy at the firm level can recover an average causal response when s_b is strictly exogenous. This causal response is a weighted average of the treatment effects of firms that would have a different potentially a different PPP allocation if they were clients of different banks. In our setting, as banks choose which firms to lend to based on their treatment effects, this set of compliers is systematically different than the set of always and never takers in terms of the treatment effect, such that the average treatment effect is always different than the average causal response.

4.4.1 Empirically Relevant Region

We can characterize even further the difference between the average causal response and ATT in our setting under some additional assumptions on the support of shock exposure η. We ground these assumptions in the empirical work of Bartik et al. (2020*b*). Specifically, we assume that the parameters of our model and the support of η are such that

1. Banks allocate funds to the firms less affected by the shock; and
2. Randomly assigning PPP funds instead of using the private banking system does not significantly change the effectiveness of the program.

In the proof of Lemmas 2 and 3 we show that both T_j and Ω_j are hump-shaped in η_j, with the max of Ω_j being at a lower η than the max of T_j, as in Figure 4. For conditions 1-2 to be valid, we have that the support of η is such that Ω_j is decreasing – banks want to allocate funds less affected by the pandemic – and up until the point where the average treatment effect of the firms that do not receive PPP funds are the same as those who do. Let $\mathcal{E} \equiv \left[\underline{\eta}_{\mathcal{E}}, \overline{\eta}_{\mathcal{E}}\right]$ be this empirically relevant region, where:

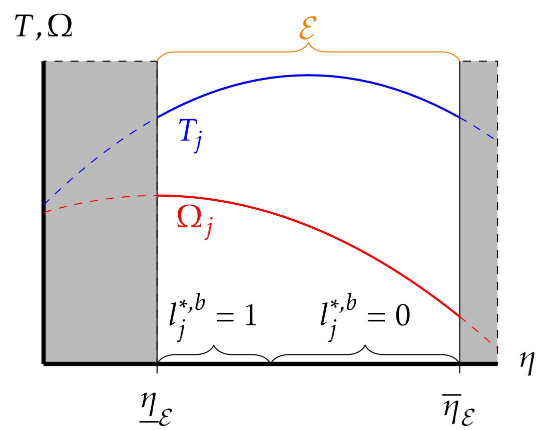
$$\arg\max_{\eta_j} T_j > \underline{\eta}_{\mathcal{E}} > \arg\max_{\eta_j} \Omega_j \tag{33}$$

and, given $\underline{\eta}_{\mathcal{E}}$, $\overline{\eta}_{\mathcal{E}}$ solves

$$\sum_b \mu_b \int_{\underline{\eta}_{\mathcal{E}}}^{\overline{\eta}_{\mathcal{E}}} l_j^{*,b} T_j dj = \sum_b \mu_b \int_{\underline{\eta}_{\mathcal{E}}}^{\overline{\eta}_{\mathcal{E}}} (1 - l_j^{*,b}) T_j dj \tag{34}$$

as depicted in Figure 4. Note that for $\mathcal{E}$ to be well defined (and able to satisfy condition 2), it must be the case that the share of firms receiving PPP is small enough for all banks such that the maximum treatment effect of firms receive PPP loan is still to the left of the max of the curve T_j, as in Figure 4.[17]

Figure 4: Empirically Relevant Region $\mathcal{E}$ of Shock exposure η



Note: This figure is the pictorial representation of the region $\mathcal{E}$, which is the support in η consistent with the evidence in Bartik et al. (2020*b*).

We show in Lemma 6 that if firms are heterogeneous in their shock exposure and $\eta_j \in \mathcal{E}$, $\forall j$, the bank-IV strategy will always overestimate the overall effect of the program, that is, $B_{F,IV} > 0$. Intuitively, the bank-IV strategy recovers an average treatment effect for firms that would have a different PPP allocation if they were clients of different banks. In region $\mathcal{E}$, these are firms with a larger treatment effect than firms that would receive PPP loans from all banks.

Lemma 6. ***Heterogeneous Shock Exposure and PPP Effect Estimation.*** *Suppose that firms are all equal except for their shock exposure η_j, and that all η_j's lie in the empirically relevant set defined in 34, that is $\eta_j \in \mathcal{E}$, $\forall j$. The effect of the program is overestimated using the bank-IV identification, that is, $B_{F,IV} > 0$.*

4.4.2 The two bank case

To provide a clear illustration of our channel, we present the simplified case when we have only two banks. This case provides all of the insight from the previous results, but it is

[17]If we only want to satisfy condition 1, then we can't bound the upper bound of $\mathcal{E}$, but can still provide analytical results similar to the ones presented here. We opt for imposing both conditions one and two to directly speak with the empirical literature.

simpler analytically as we can use a dummy for relationship with a bank as the instrument. Consider that there are two banks: A and B, that are equivalent in every measure except their PPP capacity s_b, $b \in \{A, B\}$. Note that we must have $s_B = 1 - s_A$, since both banks have the same market share. Without loss of generality, we suppose $s_A > s_B = 2 - s_A$. Let Z_j be equal to 1 if firm j is a client of bank A and zero otherwise. We consider now the use of Z_j as an instrument for PPP at the firm-level.

By assumption, the instrument Z_j is independent of the treatment effects in the population, the shocks in the second stage and the potential outcomes of the dependent variable conditional on PPP allocation. Moreover, note that if $s_A > s_B$, $\underline{\Omega}_A < \underline{\Omega}_B$ and, therefore, the monotonicity condition is satisfied. In this case, we have that the IV estimation yields the LATE, that is

$$\begin{aligned}\beta_{F,IV} &= \frac{\text{Cov}(y_j, Z_j)}{\text{Cov}(PPP_j, Z_j)} = \frac{\mathbb{E}[T_j PPP_j | Z_j = 1] - \mathbb{E}[T_j PPP_j | Z_j = 0]}{\mathbb{E}[PPP_j | Z_j = 1] - \mathbb{E}[PPP_j | Z_j = 0]} \\ &= \frac{T_A s_A - T_B s_B}{s_a - s_b} \equiv LATE\end{aligned}$$

Alternatively, we can follow the steps in the derivation of (28) to write (Appendix A.3):

$$LATE = ATT + \frac{\text{Cov}(T_j, PPP_j Z_j)}{\text{Cov}(PPP_j, Z_j)} = ATT + 2\frac{s_A}{s_A - s_B}[T_A - ATT] \tag{35}$$

If we further assume that each region only has either bank A or B, but not both, and use Z_r, we have that $\beta_{F,IV} = \beta_{R,IV}$, even if each region has a different distribution of firms and thus there is regional targetting in our sample (Appendix A.4).

Although the bank-IV strategy can recover the LATE, that is, the effect of the PPP on firms that would have received PPP with bank A, but that did not receive from bank B, the LATE is *systematically* different than the ATT in our model, since $T_A \neq ATT$. This means that we can recover the causal effect of the PPP from the bank-IV, but not the estimate the effect of the program. Under the same conditions as in Lemma 6, we show in Lemma 7 that the difference between the LATE and ATT is positive, but hump-shaped in s_A. Intuitively, if banks select firms for a given variable that is negatively correlated with their treatment effect, than firms that are selected by bank A, but not bank B, are exactly those that have the highest treatment effect. For a concrete example, note in *Example 3* that $\beta_{IV} = T_F$, that is, the treatment effect of the firms that receive funds when in a region where banks lend to

100% of firms and that do not receive funds in the region where banks lend to 50% of firms — the LATE.

Lemma 7. ***Heterogeneous Shock Exposure and PPP Effect Estimation with two banks.*** *In the same set-up as in Lemma 6, where $\eta_j \in \mathcal{E}$, $\forall j$, we have that the ATT is always smaller than the LATE, that is, $B_{F,IV} = B_{R,IV} = LATE - ATT > 0$.*

5 Concluding Remarks and Next Steps

This paper provides a framework to understand the role of the private banking sector in the allocation of the Paycheck Protection Program (PPP), a large novel crisis response adopted by the U.S. as a response to the Covid-19 crisis. We consider three main dimensions. First, what is the optimal target for the PPP? Second, how far are we from this optimal by using the private banking sector to allocate these funds? Third, what is the effect of bank incentives in the estimation of the effect of the PPP at the firm and regional levels?

We characterize the optimal PPP alloction in both a constrained first best and under the rules of the PPP under the assumption the government wants to maximize the number of firm-worker relationships maintained. We find that it is optimal to allocate PPP funds where their marginal effect is the highest, and thus PPP applications should be approved for firms with the highest treatment effects. These firms are not necessarily those most or least affected by the pandemic. Intuitively, firms that are the least affected by the pandemic shouldn't receive PPP funds, as they will likely survive regardless. On the other hand, it is likely that firms that are the most affected by the pandemic shouldn't receive PPP funds, as they likely won't survive regardless.

Banks have different incentives from the government, and thus deviate from the optimal allocation. First, banks already have outstanding loans to small business and thus have an incentive to distort the allocation towards firms with more debt. Second, banks potentially lose clients whose PPP applications are rejected and have uncertainty regarding the guarantee offered by the government. This implies that banks have an incentive to distort the allocation towards firms with a higher probability of survival without PPP assistance. These deviations are consistent with the empirical evidence in Bartik et al. (2020*b*) and provide a theoretical reasoning for why the effect of the PPP has been small in empirical literature so far.

Finally, we show that PPP loans being allocated to firms less affected by the pandemic due to bank incentives leads to a selection bias (even conditional on applications). Firms that receive PPP are less affected by the pandemic compared to those who do not receive funds, and thus more likely to survive regardless. To deal with this bias, the empirical literature has used a bank-level variation in PPP disbursement to instrument for PPP allocation. We show that even if this bank-level disbursement shock is exogenous, the bank-IV estimation does not recover the overall effect of the program. As banks distort their allocation towards certain firms with different probabilities of survival without PPP funds, they also distort the allocation towards firms with lower treatment effects. This implies that the local average treatment effect in our sample, which can arguably be estimated using the IV strategy, is biased, and likely overestimates, the treatment effect on the treated. In a more general setting with bank and regional heterogeneity, we show that there are other potential sources of bias when using bank-level shocks.

Our framework shows that, although extremely informative, the empirical estimation by itself cannot fully account for the total effect of the PPP or be used to estimate the effect of subsequent rounds of the program (where the pool of borrowers would be significantly different). Our ultimate goal is to use the same data as Granja et al. (2020) to calibrate our model, and use it to quantify the effects of the PPP and complement the reduced form results documented by the literature, as well as perform counterfactuals.

References

Amiram, Dan, and Daniel Rabetti. 2020. "The Relevance of Relationship Lending in Times of Crisis." SSRN Working Paper. 4, 6

Angrist, Joshua D., Kathryn Graddy, and Guido W. Imbens. 2000. "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish." *Review of Economic Studies*, 67(3): 499–527. 33

Autor, David, David Cho, Leland D Crane, Byron Lutz, Joshua Montes, William B Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz. 2020. "An Evaluation of the Paycheck Protection Program Using Administrative Payroll Microdata." 5, 6

Barrios, John Manuel, Michael Minnis, William C. Minnis, and Joost Sijthoff. 2020. "Assessing the Payroll Protection Program: A Framework and Preliminary Results." *SSRN Electronic Journal*. 6

Bartik, Alexander W., Marianne Bertrand, Zoe Cullen, Edward L. Glaeser, Michael Luca, and Christopher Stanton. 2020*a*. "The impact of COVID-19 on small business outcomes and expectations." 2, 3, 10

Bartik, Alexander W., Zoe E. Cullen, Edward L. Glaeser, Michael Luca, Christopher T. Stanton, and Adi Sunderam. 2020*b*. "The Targeting And Impact of Paycheck Protection Program Loans to Small Businesses." NBER Working Paper No. 27623. 2, 4, 5, 6, 13, 19, 20, 22, 26, 29, 33, 34, 36

Bartlett, Robert P, and Adair Morse. 2020. "Small Business Survival Capabilities and Policy Effectiveness: Evidence from Oakland." 16

Chetty, Raj, John N Friedman, Nathaniel Hendren, Michael Stepner, and Opportunity Insights Team. 2020. "The Economic Impacts of COVID-19: Evidence from a New Public Database Built from Private Sector Data." 2, 5

Chodorow-Reich, Gabriel, Olivier Darmouni, Stephan Luck, and Matthew Plosser. 2020. "Bank Liquidity Provision Across the Firm Size Distribution." National Bureau of Economic Research, Cambridge, MA. 5

Cororaton, Anna, and Samuel Rosen. 2020. "Public Firm Borrowers of the US Paycheck Protection Program." *SSRN Electronic Journal*. 6

Elenev, Vadim, Tim Landvoigt, and Stijn Van Nieuwerburgh. 2020. "Can the Covid Bailouts Save the Economy?" NBER Working Paper 27207, Cambridge, MA. 6

Erel, Isil, and Jack Liebersohn. 2020. "Does FinTech Substitute for Banks? Evidence from the Paycheck Protection Program." NBER Working Paper 27659, Cambridge, MA. 5

Faria-e Castro, Miguel. 2020. "Fiscal Policy during a Pandemic." NBER Working Paper. 6

Granja, João, Christos Makridis, Constantine Yannelis, and Eric Zwick. 2020. "Did the Paycheck Protection Program Hit the Target?" 4, 5, 15, 23, 29, 32, 37

Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Iván Werning. 2020. "Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?" National Bureau of Economic Research, Cambridge, MA. 2, 9, 10

Hassan, Tarek Alexander, Stephan Hollander, Laurence van Lent, Markus Schwedeler, and Ahmed Tahoun. 2020. "Firm-level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1." NBER Working Paper 26971, Cambridge, MA. 6

Li, Lei, and Philllip Strahan. 2020. "Who Supplies PPP Loans (And Does It Matter)? Banks, Relationships And The Covid Crisis." NBER Working Paper 28286. 4, 6, 29

Neilson, Christopher, John Eric Humphries, and Gabriel Ulyssea. 2020. "Information Frictions and Access to the Paycheck Protection Program." NBER Working Paper 27624, Cambridge, MA. 5, 11

Appendix

A Derivations

Auxiliary Result. For the distribution in (3), we have that $\mathbb{E}[\nu \mid \nu \leq X] = \frac{\eta}{\eta+1}X$

$$\mathbb{E}[\nu \mid \nu \leq X] = \left(\frac{X}{c_0}\right)^{-\eta} \int_0^X \eta t \frac{1}{c_0}\left(\frac{t}{c_0}\right)^{\eta-1} dt = (X)^{-\eta}\,\eta \int_0^X t^\eta dt = X^{-\eta}\eta\frac{X^{\eta+1}}{\eta+1} = \frac{\eta}{\eta+1}X \quad \blacksquare$$

A.1 Firm's Choice in the PPP

We proceed in two steps. First, we consider the case where $a = 1$ (firm applies), and then compute the amount of funds in the application, ω. Then, we focus on which firms choose to apply.

Step 1: Choice of ω given $a = 1$. From the problem of the firm in (2), we can take the FOC w.r.t. ω when $a = 1$ to obtain

$$\phi_j(\omega)\cdot\left[\Pi_j(\omega) - \frac{\eta_j}{\eta_j+1}\Gamma_j(\omega)\right] - \Phi_j(\omega)\cdot r_G > 0$$

from $\Phi_j(\omega) \geq 0$ and $\Pi_j(\omega) > \Gamma_j(\omega) > \frac{\eta_j}{\eta_j+1}\Gamma_j(\omega)$.

Step 2: Choice of a. From the firm objective function in (4), a firm chooses to apply if

$$\Phi_j(\varphi)\Big(\Pi_j(\varphi) - \mathbb{E}\Big[\nu_j \mid \nu_j \leq \Gamma_j(\varphi)\Big]\Big) - \Phi_j(0)\Big(\Pi_j(0) - \mathbb{E}\Big[\nu_j \mid \nu_j \leq \Gamma_j(0)\Big]\Big) > \frac{F}{N_j}$$

Therefore, $a_j^* = 1$ if :

$$T_j\Pi_j(0) - \Phi_j(\varphi)r_G\varphi - \int_{\Gamma_j(0)}^{\Gamma_j(\varphi)} \nu\, d\Phi(\nu \mid \eta_j) > \frac{F}{N_j}$$

which delivers (6). Using the distribution in (3):

$$T_j\Pi_j(0) - T_j\mathbb{E}\Big[\nu_j \mid \nu_j \in [\Gamma_j(0), \Gamma_j(\varphi)]\Big] = (c_j + \pi_j^{LR})T_j - \Phi_j(\varphi)\frac{\eta_j}{\eta_j+1}(c_j+\varphi) + \Phi_j(0)\frac{\eta_j}{\eta_j+1}c_j \qquad (36)$$

which delivers (7). $\blacksquare$

A.2 OLS and IV estimation and Treatment Effects

A.2.1 OLS Estimation

Using firm-level variation, we have that the OLS estimator in (23) will converge to β_F given by

$$\begin{aligned}\beta_F &= \mathbb{V}(PPP_j)^{-1}\left[\mathrm{Cov}(\theta_j, PPP_j) + \mathrm{Cov}(T_j PPP_j, PPP_j)\right] \\ &= \mathbb{V}(PPP_j)^{-1}\left[(\mathbb{E}(\theta_j|PPP_j = 1) - \overline{\theta})\overline{p} + ATT\overline{p}(1-\overline{p})\right] = ATT + \frac{\mathbb{E}(\theta_j|PPP_j = 1) - \overline{\theta}}{1-\overline{p}}\end{aligned} \tag{37}$$

where the second equality comes from

$$\mathrm{Cov}(T_j PPP_j, PPP_j) = \mathbb{E}\left[T_j PPP_j\right] - E\left[T_j PPP_j\right]E\left[PPP_j\right] = ATT\overline{p}(1-\overline{p})$$

Using regional-level variation, we have that the OLS estimator in (25) will converge to β_R given by

$$\begin{aligned}\beta_R &= \mathbb{V}_w(PPP_r)^{-1}\left[\mathrm{Cov}_w(\theta_r, PPP_r) + \mathrm{Cov}_w(T_r PPP_r, PPP_r)\right] \\ &= ATT + \mathbb{V}_w(PPP_r)^{-1}\left\{\mathrm{Cov}_w(\theta_r, PPP_r) + \mathrm{Cov}_w((T_r - ATT)PPP_r, PPP_r)\right\} \\ &= ATT + \mathbb{V}_w(PPP_r)^{-1}\left\{\mathrm{Cov}_w(\theta_r, PPP_r) + \mathrm{Cov}_w(T_r, PPP_r^2)\right\}\end{aligned}$$

The last equality comes from

$$\mathrm{Cov}_w((T_r - ATT)PPP_r, PPP_r) = \mathbb{E}_w\left[(T_r - ATT)PPP_r^2\right] - E_w\left[(T_r - ATT)PPP_r\right]E_w\left[PPP_r\right]$$

and $ATT = \frac{\mathbb{E}_w[T_r PPP_r]}{\mathbb{E}_w[PPP_r]}$ $\blacksquare$

A.2.2 Bank-IV Estimation

For the of the firm level estimation in (27), we have that:

$$\hat{\beta}^{IV,F} \xrightarrow[p]{} \frac{\mathrm{Cov}(y_j, s_{b_j})}{\mathrm{Cov}(PPP_j, s_{b_j})} = \frac{\mathrm{Cov}(\theta_j, s_{b_j})}{\mathrm{Cov}(PPP_j, s_{b_j})} + \frac{\mathrm{Cov}(T_j PPP_j, s_{b_j})}{\mathrm{Cov}(PPP_j, s_{b_j})}$$

Note that we can write:

$$\text{Cov}(PPP_j, s_{b_j}) = \mathbb{E}[\mathbb{E}[PPP_j|s_{b_j}]s_{b_j}] - \frac{M}{\varphi}\bar{s}^2 = \frac{M}{\varphi}\left\{\mathbb{E}\left[s_{b_j}^2\mu_{b_j}\right] - \bar{s}^2\right\} = \frac{M}{\varphi}\mathbb{V}_\mu(s_b)$$

where we use the fact that $\bar{p} = \frac{M}{\varphi}\sum_b \mu_b s_b$. Moreover,

$$\text{Cov}(\theta_j, s_{b_j}) = \mathbb{E}[\theta_j s_{b_j}] - \mathbb{E}[\theta_j]\mathbb{E}[s_{b_j}] = \mathbb{E}\left[\left(\mathbb{E}[\theta_j|s_{b_j}] - \bar{\theta}\right)s_{b_j}\right] = \text{Cov}_\mu(\theta_b, s_b)$$

Finally, we have that

$$\begin{aligned}
\text{Cov}(T_j PPP_j, s_{b_j}) &= \mathbb{E}[T_j PPP_j s_{b_j}] - \mathbb{E}[T_j PPP_j]\mathbb{E}[s_{b_j}] \\
&= \mathbb{E}[\mathbb{E}[T_j|PPP_j, s_{b_j}]\mathbb{P}[PPP_j = 1|s_{b_j}]s_{b_j}] - ATT\frac{M}{\varphi}\bar{s} \\
&= \frac{M}{\varphi}\left\{\mathbb{E}[ATT_b s_{b_j}^2] - ATT\bar{s}\right\} = \frac{M}{\varphi}\left\{\mathbb{E}_\mu[(ATT_b - ATT)s_{b_j}^2] + ATT\left[\mathbb{E}_\mu[s_{b_j}^2] - \bar{s}\right]\right\} \\
&= \frac{M}{\varphi}\left\{\mathbb{E}[(ATT_b - ATT)s_{b_j}^2] + ATT \times \mathbb{V}_\mu(s_b)\right\}
\end{aligned}$$

Therefore:

$$\beta_{F,IV} = ATT + \frac{\varphi}{M}\frac{\text{Cov}_\mu(\theta_b, s_b)}{\mathbb{V}_\mu(s_b)} + \frac{\text{Cov}_\mu(T_b, s_b^2)}{\mathbb{V}_\mu(s_b)}$$

Running the regression at the firm or regional level using the regional exposure from (18), we have that

$$\hat{\beta}_{IV,R} \xrightarrow[p]{} \frac{\text{Cov}(y_j, s_r)}{\text{Cov}(PPP_j, s_r)} = \frac{\text{Cov}_w(y_r, s_r)}{\text{Cov}_w(PPP_r, s_r)} = \frac{\text{Cov}_w(\theta_r, s_r)}{\text{Cov}_w(PPP_r, s_r)} + \frac{\text{Cov}_w(T_j PPP_r, s_r)}{\text{Cov}_w(PPP_r, s_r)}$$

Note that

$$\text{Cov}_w((T_r - ATT)PPP_r, s_r) = \mathbb{E}_w[(T_r - ATT)PPP_r s_r] - \mathbb{E}_w[(T_r - ATT)PPP_r]\mathbb{E}_w[s_r] = \text{Cov}_w(T_r, PPP_r s_r)$$

Therefore:

$$\beta_{IV,R}^R = ATT + \frac{\text{Cov}_w(\theta_r, s_r)}{\text{Cov}_w(PPP_r, s_r)} + \frac{\text{Cov}_w(T_r, PPP_r s_r)}{\text{Cov}_w(PPP_r, s_r)} \quad \blacksquare$$

A.3 Two Bank Example: LATE vs ATT

For the of the firm level estimation in (27), we have that:

$$\beta_{IV,F} = \frac{\text{Cov}(y_j, Z_j)}{\text{Cov}(PPP_j, Z_j)} = \frac{\text{Cov}(\theta_j, Z_j)}{\text{Cov}(PPP_j, Z_j)} + \frac{\text{Cov}(T_j PPP_j, Z_j)}{\text{Cov}(PPP_j, Z_j)}$$

Note that we can write:

$$\text{Cov}(PPP_j, Z_j) = \frac{1}{2}s_A - \frac{1}{2}\left(\frac{1}{2}s_A + \frac{1}{2}s_B\right) = \frac{1}{4}(s_A - s_B)$$

Moreover,

$$\begin{aligned}\text{Cov}((T_j - ATT)PPP_j, Z_j) &= \mathbb{E}[(T_j - ATT)PPP_j Z_j] - \left\{\mathbb{E}[T_j PPP_j] - ATT\,\mathbb{E}[PPP_j]\right\}\mathbb{E}[Z_j] \\ &= \mathbb{E}[(T_j - ATT)PPP_j Z_j]\end{aligned}$$

Furthermore

$$\begin{aligned}\text{Cov}(T_j PPP_j, Z_j) = \mathbb{E}[T_j PPP_j Z_j] - \mathbb{E}[T_j PPP_j]\mathbb{E}[Z_j] &= \frac{1}{2}s_A T_A - \frac{1}{2}\left(\frac{1}{2}s_A + \frac{1}{2}s_B\right)ATT \\ &= \frac{1}{4}(s_A - s_B)ATT + \frac{1}{2}s_A T_A - \left(\frac{1}{4}s_A + \frac{1}{4}s_B\right)ATT - \frac{1}{4}(s_A - s_B)ATT \\ &= \frac{1}{4}(s_A - s_B)ATT + \frac{1}{2}s_A(T_A - ATT)\end{aligned}$$

Therefore:

$$\beta_{IV,F} = ATT + \frac{\text{Cov}(T_j, PPP_j Z_j)}{\text{Cov}(PPP_j, Z_j)} = ATT + 2\frac{s_A}{s_A - s_B}[T_A - ATT] \quad \blacksquare$$

A.4 Two Bank Example: LATE vs ATT in the Regional Regression

In the two bank case with only one bank per region, we have that:

$$\beta_{IV,R} = \frac{\text{Cov}(T_r PPP_r, Z_r)}{\text{Cov}(PPP_r, Z_r)}$$

Note that we can write:

$$\text{Cov}(PPP_r, Z_r) = \mathbb{E}[PPP_r Zr] - \mathbb{E}[Z_r]\mathbb{E}[P_r] = \frac{1}{2}s_A - \frac{1}{2}\left(\frac{1}{2}s_A + \frac{1}{2}s_B\right) = \frac{1}{4}(s_A - s_B)$$

Furthermore,

$$\begin{aligned}\text{Cov}(T_j PPP_j, Z_j) &= \mathbb{E}[T_r PPP_r Z_r] - \mathbb{E}[T_r PPP_r]\mathbb{E}[Z_r] = \frac{1}{2}s_A T_A - \frac{1}{2}\left(\frac{1}{2}s_A + \frac{1}{2}s_B\right)ATT \\ &= \frac{1}{4}(s_A - s_B)ATT + \frac{1}{2}s_A(T_A - ATT)\end{aligned}$$

Therefore: $\beta_{IV,R} = \beta_{IV,F}$.

B Proofs

B.1 Lemma 1

Proof. Let λ be the Lagrange Multiplier on the constraint that $\int_{\mathcal{F}} N_j \omega_j^G di = M$. Taking the FOC of (8) w.r.t. ω_j^G

$$N_j \phi\left[\Gamma\left(\omega_i^G\right)\right] \cdot \frac{\partial \Gamma_i}{\partial \omega_i^G} - N_i \lambda = 0 \Rightarrow \phi\left[\Gamma\left(\omega_i^G\right)\right] = \phi\left[\Gamma\left(\omega_j^G\right)\right],\ \forall i, j$$

where we use that $\frac{\partial \Gamma_i}{\partial \omega_i^G} = 1$ of the last equation. Let $\tilde{\lambda} \equiv \lambda \cdot c_0^\eta$. Using the equation for the distribution $\phi(\nu)$ in (3):

$$\eta_j \left[\pi_j + \omega_j^{G,*}\right]^{\eta - 1} - \tilde{\lambda} = 0 \Rightarrow \omega_j^{G,*} = \left(\frac{\tilde{\lambda}}{\eta_j}\right)^{\frac{1}{\eta_j - 1}} - c_j \Rightarrow M + \bar{c} = \int_{\mathcal{F}} N_j \left(\frac{\tilde{\lambda}}{\eta_j}\right)^{\frac{1}{\eta_j - 1}} dj \tag{38}$$

where the last equality comes from integrating $N_j \omega_j^{G,*}$ across firms to solve for λ. This is the unique global maximum of the problem as the constraint is linear and the objective function is strictly concave.

Note that the RHS is strictly *decreasing* in $\tilde{\lambda}$ since $\eta_j < 1$, so there is always a unique solution for

$\tilde{\lambda}$ from (38). We can use (38) in the individual firm j equation to obtain:

$$N_j \omega_j^{G,*} = M - \left[N_j c_j - \overline{c}\right] + N_j \left(\frac{\tilde{\lambda}}{\eta_j}\right)^{\frac{1}{\eta_j - 1}} - \int_{\mathcal{F}} N_j \left(\frac{\tilde{\lambda}}{\eta_j}\right)^{\frac{1}{\eta_j - 1}} \, dj \tag{39}$$

Thus, we have that $N_j \tau(\eta_j, \overline{M}) \equiv N_j \left(\frac{\tilde{\lambda}}{\eta_j}\right)^{\frac{1}{\eta_j - 1}} - \int_{\mathcal{F}} N_j \left(\frac{\tilde{\lambda}}{\eta_j}\right)^{\frac{1}{\eta_j - 1}} \, dj$. Therefore:

$$\frac{\partial \omega_j^{G,*}}{\partial \eta_j} = \frac{\partial \tau(\eta_j, \overline{M})}{\partial \eta_j} = -\left(\frac{\tilde{\lambda}}{\eta_j}\right)^{\frac{1}{\eta_j - 1}} \frac{1}{(\eta_j - 1)^2} \left[\ln(\tilde{\lambda}) + 1 - \frac{1}{\eta_j} - \ln(\eta_j)\right] \tag{40}$$

Let $f(\eta) = 1 - \eta^{-1} - \ln(\eta)$. We know that $f(1) = 0$ and $f'(\eta) = \eta^{-2} - \eta^{-1}$. Thus, $f(\eta) > 0$ for $\eta < 1$ and $f(\eta) < 0$ for $\eta > 1$.

Case 1. If $\tilde{\lambda}$ small enough, we have that $\frac{\partial \omega_{i,j}^{G,*}}{\partial \eta_j} > 0$. Too see this, note that $f(\eta_j) < 0$ and $\ln(\tilde{\lambda}) < 0$, thus in (40) the RHS is positive. For $\tilde{\lambda}$ small, we must have a large $M + \overline{c}$ from (38).

Case 2. if $\tilde{\lambda} > 1$, we have that $\frac{\partial \tau(\eta_j, \overline{M})}{\partial \eta_j}$ is positive for $\eta_j < \bar{\eta} < 1$ and negative otherwise (that is, the transfer function τ is concave in η_j). To see this, note that $\lim_{\eta \to 0^+} f(\eta) = -\infty$. Therefore, any positive $\ln(\tilde{\lambda})$ is simply a shifter downward of this function. We still have that $\lim_{\eta \to 0^+} f(\eta) + \ln(\tilde{\lambda}) = -\infty$, $f(\eta) + \ln(\tilde{\lambda}) > 0$ and $f(\eta) + \ln(\tilde{\lambda})$ always decreasing. Therefore, by the intermediate value theorem, we have that $\exists! \ \bar{\eta} < 1$ s.t. $\frac{\partial \tau(\eta_j, \overline{M})}{\partial \eta_j} > 0 \Leftrightarrow \eta_j < \bar{\eta}$. For $\tilde{\lambda} > 1$, it must be the case where $\overline{M}$ large, in particular: $M + \overline{c} > \int_{\mathcal{F}} \left(\eta_j\right)^{\frac{1}{1 - \eta_j}} \, dj$. ∎

B.2 Lemma 2

Proof. Let $G(\{l_j^G\}_j)$ be the Lagrangian of the problem of the government in (11). The derivative of the Lagrangean $G(.)$ with respect to l_j^G, that is, the marginal allocation

$$G_l \equiv \frac{\partial G}{\partial l_j^G} = T_j - \varphi \lambda$$

where λ is the Lagrange multiplier in the resource constraint.

Case 1. Debt heterogeneity. Consider that all firms in the economy are the same except for their level of debt b_j. Then:

$$T_j = \left[\rho - b_j + \varphi\right]^{\eta} - \left[\rho - b_j\right]^{\eta} \Rightarrow \frac{\partial T_j}{\partial b_j} = -\eta\left(\left[\rho - b_j + \varphi\right]^{\eta-1} - \left[\rho - b_j\right]^{\eta-1}\right)$$

For $\eta < 1$, T_j is thus increasing in b_j. For $\eta > 1$, T_j is decreasing in b_j.

Case 2. Shock exposure Heterogeneity. Consider that all firms are the same except for their shock exposure η_j. Define $\tilde{c} \equiv \frac{c}{c_0}$ and $\tilde{\varphi} \equiv \frac{\varphi}{c_0}$. Then:

$$\frac{\partial T_j}{\partial \eta_j} = (\tilde{c} + \tilde{\varphi})^{\eta_j} \cdot \ln(\tilde{c} + \tilde{\varphi}) - \tilde{c}^{\eta_j} \ln(\tilde{c}) > 0 \Leftrightarrow (\tilde{c} + \tilde{\varphi})^{\eta_j} \cdot \ln(\tilde{c} + \tilde{\varphi}) > \tilde{c}^{\eta_j} \ln(\tilde{c})$$

Which implies:

$$\eta_j \ln\left(1 + \frac{\varphi}{c}\right) + \ln(-\ln(\tilde{c} + \tilde{\varphi})) < \ln(-\ln(\tilde{c})) \Leftrightarrow \eta_j < \eta_G^* \equiv \frac{\ln\left(\frac{\ln(\tilde{c})}{\ln(\tilde{c}+\tilde{\varphi})}\right)}{\ln(1 + \frac{\varphi}{c})} > 0$$

Therefore, T_j is strictly increasing up to $\eta_G^* > 0$ and strictly decreasing afterwards. The optimal allocation is thus $l_j^G = 1$ if $\eta_j \in [\underline{\eta}_G, \overline{\eta}^G]$, where $T_{\underline{\eta}} = T_{\overline{\eta}}$ and $\int_{\underline{\eta}_G}^{\overline{\eta}^G} \varphi dj = M$, which (i) exists, since the resource constraint is binding and (ii) is unique, since T_j is quasi-concave in η_j. ■

B.3 Lemma 3

Proof. We will proceed as in the proof of Lemma 2. Let $B(\{l_j^B\}_j)$ be the Lagrangian of the problem of the government in (11). The derivative of the Lagrangean of $B(.)$ with respect to l_j^B, that is, the marginal allocation

$$B_l \equiv \frac{\partial B}{\partial l_j^B} = \Omega_j - \varphi\lambda$$

where λ is the Lagrange multiplier in the resource constraint.

Case 1. Debt heterogeneity. When firms are only heterogeneous in b_j, we have $\frac{\partial \Omega_j}{\partial b_j} = (1-\delta)T > 0$, that is, banks want to allocate funds to the firms with the highest levels of pre-pandemic debt per employee.

Case 2. Shock Exposure Heterogeneity. Consider that all firms are the same except η_j. Then:

$$\frac{\partial \Omega_j}{\partial \eta_j} = \left[\kappa\,(c+\varphi)^{\eta_j} \cdot \ln(c+\varphi) - (\kappa - \tilde{\psi})c^{\eta_j} \ln(c)\right]$$

where $\kappa \equiv (1-\delta)b + \psi_F + q\varphi$ and $\tilde{\psi} \equiv \psi_F \psi_C + q\varphi$. Therefore

$$\frac{\partial \Omega_j}{\partial \eta_j} > 0 \Leftrightarrow (c+\varphi)^{\eta_j} \cdot \ln(c+\varphi) > \left[1 - \frac{\tilde{\psi}}{\kappa}\right] c^{\eta_j} \ln(c)$$

Which implies:

$$\eta_j \ln\left(1 + \frac{\varphi}{c}\right) + \ln(-\ln(c+\varphi)) < \ln\left(-\left[1 - \frac{\tilde{\psi}}{\kappa}\right]\ln(c)\right) \Leftrightarrow \eta_j < \eta_B^* \equiv \frac{\ln\left(\left[1 - \frac{\tilde{\psi}}{\kappa}\right]\frac{\ln(c)}{\ln(c+\varphi)}\right)}{\ln(1 + \frac{\varphi}{c})}$$

since $\tilde{\psi} < \kappa$.

Therefore, B_j is strictly increasing up to $\eta_B^* > 0$ and strictly decreasing afterwards. The optimal allocation is thus $l_j^B = 1$ if $\eta_j \in [\underline{\eta}^B, \overline{\eta}^B]$, where $B_{\underline{\eta}_j^B} = B_{\overline{\eta}_j^B}$ and $\int_{\underline{\eta}_B}^{\overline{\eta}_B} \varphi dj = M$, which (i) exists, since the resource constraint is binding and (ii) is unique, since B_j is quasi-concave.

Finally, we will show that: $\overline{\eta}_B \geq \overline{\eta}_G$ and $\underline{\eta}_B \geq \underline{\eta}_G$. By contradiction, assume that $\overline{\eta}_B \geq \overline{\eta}_G$. In this case, $\overline{\eta}_B \geq \overline{\eta}_G$ (from the resource constraint). The strategy of the proof is to take an η smaller, but sufficiently close to $\underline{\eta}_G$. The T_j at this point will be closer to a point at $\overline{\eta}^G$, but the probability of survival will be much higher, and thus this point will offer a much higher profit for the bank. Mathematically, given that T_η is a continuous function at $\eta > 0$, we have that $\forall\ \varepsilon > 0, \exists\ \zeta > 0$

$$|\eta - \underline{\eta}_G| < \zeta \Rightarrow |T_\eta - T_{\underline{\eta}_G}| < \varepsilon$$

Take $\varepsilon < \tilde{\psi}\left[\Phi_{\underline{\eta}_G}(0) - \Phi_{\overline{\eta}_G}(0)\right]$. Then, there $\exists\ \eta = \underline{\eta}_G - \zeta$, with $\zeta > 0$ such that:

$$\kappa T_{\underline{\eta}_G} + \tilde{\psi}\Phi_{\overline{\eta}_G}(0) = \kappa T_{\overline{\eta}_G} + \tilde{\psi}\Phi_{\overline{\eta}_G}(0) < \kappa T_\eta + \tilde{\psi}\Phi_{\underline{\eta}_G}(0) < \kappa T_\eta + \tilde{\psi}\Phi_\eta(0)$$

Therefore, $\overline{\eta}_B \geq \overline{\eta}_G$ cannot be optimal for the bank. $\blacksquare$

B.4 Lemma 4

Proof. Let $\mathcal{W}_G(\mu,\iota)$ be the social welfare under the government allocation (the same as in Lemma 2) and $\mathcal{W}_B$ the social welfare under the optimal bank allocation (the same as in Lemma 2). We know that $\mathcal{W}_G$ is differentiable and strictly increasing in both arguments, since:

$$\mathcal{W}_G(\mu,\iota) = (1-\mu)\iota\overline{T} + \mu\iota\int_{\underline{T}}^{\infty} T_j dj$$

where $\underline{T}$ is the minimum value of the treatment effect that receives PPP funds, that is $\int_{\underline{T}}^{\infty} dj = \frac{M}{\varphi}$. In Lemma 2, for instance, this corresponds to the treatment effect at the border of the interval $[\underline{\eta}_G, \overline{\eta}_G]$.

Moreover, we know that the curve $\iota^*(1)$ exists through the Intermediate Value Theorem, since $\mathcal{W}_G(1,1) > \mathcal{W}_B$ and $\mathcal{W}_G(0,1) < \mathcal{W}_B$. Finally, at a curve where $\mathcal{W}_G(\mu,\iota)$ is constant, we have from the implicit function theorem that $\frac{d\iota^*(\mu)}{d\mu} = -\frac{\frac{\partial \mathcal{W}_G}{\partial \mu}}{\frac{\partial \mathcal{W}_G}{\partial \iota}}$, which exists and is negative. Therefore, we have that $\iota^*(\mu) = \iota^*(1) - \int_{\mu}^{1} \frac{d\iota^*(t)}{d\mu} dt$. ■

B.5 Lemma 5

Proof. Given the linearity of (30) on l_j^b, the optimal allocation at the bank level from is given by choosing the firms j with maximum profits $\Omega(.)$, that is

$$l_j^b = 1 \Leftrightarrow \Omega(\theta_j, T_j, b_j) > \underline{\Omega}_b$$

where $\underline{\Omega}_b$ is defined from

$$\mathbb{P}\left[\Omega(\theta_j, T_j, b_j) > \underline{\Omega}_b\right] = s_b \frac{M}{\varphi}$$

Therefore, $\underline{\Omega}_b$ is strictly decreasing in s_b, and $T_b \equiv \mathbb{E}\left[T_j \mid \Omega(\theta_j, T_j, b_j) > \underline{\Omega}_b\right]$ is thus a function of s_b. Although it is possible that $\mathrm{Cov}(s_b, T_b) = 0$, it does not happen with positive probability given the relation between s_b and T_b. The same is true at the regional level, since T_r and s_r are weighted averages of T_b and s_b, respectively, and PPP_r is a function of s_r. ■

B.6 Lemma 6

Heterogeneous Shock Exposure and PPP Effect Estimation. Suppose that firms are all equal except for their shock exposure η_j, and that all η_j's lie in the empirically relevant set defined in 34, that is $\eta_j \in \mathcal{E}, \forall j$. The effect of the program is overestimated using the bank-IV identification, that is, $B_{F,IV} > 0$.

Proof. Given that $\eta_j \in \mathcal{E}, \forall j$ and $\mathcal{E}$ is well defined, we have that if $s_b > 1$, then $T_b > T$. To see that, we can replicate the proof of Lemma 7 replacing bank B as averages across all banks. Similarly, if $s_b < 1$, $T_b > T$. Jointly, this implies that $\text{Cov}(s_b, T_b) > 0$.

∎

B.7 Lemma 7

Proof. Suppose that all firms are identical except for their shock exposure, η_j. Let the optimal bank allocations be given by $[\underline{\eta}_{\mathcal{E}}, \overline{\eta}_A]$ for $b \in \{A, B\}$. Let $\overline{\eta}_1$ be given by

$$\int_{\underline{\eta}_{\mathcal{E}}}^{\overline{\eta}_1} dj = \frac{1}{2}\int_{\underline{\eta}_{\mathcal{E}}}^{\overline{\eta}_A} dj + \frac{1}{2}\int_{\underline{\eta}_{\mathcal{E}}}^{\overline{\eta}_B} dj$$

Given that $\eta_j \in \mathcal{E}, \forall j$ and $\mathcal{E}$ is well defined, we have that $\eta_1 < \eta_G^*$ (the argmax of treatment effects) such that

$$T_B = \mathbb{E}[T_j | \eta_j < \overline{\eta}_B] < \mathbb{E}[T_j | \eta_j < \overline{\eta}_1] \tag{41}$$

since T_j is increasing for $\eta < \overline{\eta}_1$ and $\overline{\eta}_1 < \overline{\eta}_B$ given that

$$\int_{\underline{\eta}_{\mathcal{E}}}^{\overline{\eta}_B} dj = (2 - s_A)\frac{M}{2\varphi} < \frac{M}{2\varphi} = \int_{\underline{\eta}_{\mathcal{E}}}^{\overline{\eta}_B} dj$$

Moreover, note that

$$T_A = \mathbb{E}[T_j | \eta_j < \overline{\eta}_A] \geq \mathbb{E}[T_j | \eta_j < \overline{\eta}_{\mathcal{E}}] = \mathbb{E}[T_j | \eta_j < \overline{\eta}_1] \tag{42}$$

To see that, we have to consider two cases. Let $\eta^{**} \equiv \arg\max_\eta \mathbb{E}[T_j | \eta_j < \eta]$, which is well defined in $\mathcal{E}$ given that T_j is hump-shaped in η and to satisfy condition 2 of $\mathcal{E}$ there is a positive probability of firms close to $\overline{\eta}_{\mathcal{E}}$ with $T_j < \mathbb{E}[T_j|\eta_j < \overline{\eta}_1]$ that is, that are average reducing. If $\overline{\eta}_A \leq \eta^{**}$, we have that T_j is increasing in this interval and our inequality is true given that $\mathbb{E}[T_j|\eta_j < \overline{\eta}_{\mathcal{E}}] = \mathbb{E}[T_j|\eta_j < \overline{\eta}_1]$. If $\overline{\eta}_A > \eta^{**}$

$$T_A = \mathbb{E}[T_j|\eta_j < \overline{\eta}_A] = \mathbb{E}[T_j|\eta_j \leq \eta^{**}]\mathbb{P}\left[\eta_j < \eta^{**}|\eta_j < \overline{\eta}_A\right] + \mathbb{E}[T_j|\eta^{**} < \eta_j < \overline{\eta}_A]\mathbb{P}\left[\eta^{**} < \eta_j|\eta_j < \overline{\eta}_A\right]$$

which is decreasing in $\overline{\eta}_A$, since $\mathbb{E}[T_j|\eta_j \leq \eta^{**}] \geq \mathbb{E}[T_j|\eta^{**} < \eta_j < \overline{\eta}_A]$ and $\mathbb{P}\left[\eta_j < \eta^{**}|\eta_j < \overline{\eta}_A\right]$ is decreasing in η^{**}. Putting together (41) and (42), $T_A > T_B \Rightarrow T_A > ATT$.

■

C Extensions

C.1 Example of Constrained First Best with $\eta_j > 1$.

Suppose that there are two equally present types of firms in the economy, H and L. All firms have zero pre-pandemic profits $\pi_j = 0$. However, each firms has its own η_F, $F \in \{L, H\}$, with $\eta_H > \eta_L$. Let the total amount of the program be $M = 1$. The CFB in this case:

$$\max_{d \in [0,1]} d^{\eta_H} + (1-d)^{\eta_L}$$

For $\eta_H > \eta_L > 1$, this function is maximized with $d = 0$ or $d = 1$. The government in this case is indifferent between allocating funds to the most or least affected firms. ■

C.2 Super and Submodular Distributions.

Let $\nu_j \sim \Upsilon(\pi_j + \omega_j, \theta_j)$, where $\theta_j \in \Theta$, a complete lattice, parametrizes the distribution Υ and can be different across firms and that, as in the text, a higher θ implies that a firm is more affected in a FOSD. Take $j, \hat{j}$ such that $\theta > \hat{\theta}$. Let $\omega^*, \hat{\omega}^*$ be candidates for an optimum for these two types.

Suppose by contradiction that for $\omega^* \leq \hat{\omega}^*$. For the strict inequality, since Υ is strictly supermodular

$$\Upsilon(\hat{\omega}^*, \theta) - \Upsilon(\omega^*, \theta) > \Upsilon(\hat{\omega}^*, \hat{\theta}) - \Upsilon(\omega^*, \hat{\theta}) \Leftrightarrow \Upsilon(\omega^*, \theta) + \Upsilon(\hat{\omega}^*, \hat{\theta}) < \Upsilon(\hat{\omega}^*, \theta) + \Upsilon(\omega^*, \hat{\theta})$$

For the case where $\omega^* = \hat{\omega}^*$, $\forall \varepsilon > 0$: $\Upsilon(\omega^* - \varepsilon, \theta) + \Upsilon(\omega^* + \varepsilon, \hat{\theta}) < \Upsilon(\omega^* + \varepsilon, \theta) + \Upsilon(\omega^* - \varepsilon, \hat{\theta})$. The same argument applies for the submodular case. ■

C.3 Misallocation and Program Design

A natural question from comparing the optimal allocation of the banking system in Lemma 3 with that of the government in Lemma 2 is: what is the optimal program design in terms of the size of the program M and maximum amount per firm φ?

To answer this question formally, we introduce some notation. First, denote $\tilde{M} \equiv M/\varphi$ as the share of workers in firms that receive PPP funds. We will write the problem of the government in terms of $\tilde{M}$ and φ. Let $\mathcal{W}^G_{\mathcal{A}}(\varphi, \tilde{M}) \equiv \int_{\mathcal{A}} l_j^{G,*} T_j dj$ be the welfare under the optimal government allocation and, analogously, $\mathcal{W}^B_{\mathcal{A}}(\varphi, \tilde{M}) \equiv \int_{\mathcal{A}} l_j^{B,*} T_j dj$ the welfare under the optimal bank allocation. Furthermore, suppose that there are cost functions $C_M(\tilde{M})$, $C_\varphi(\varphi)$ of changing the share of firms/loans per firm in the program (tax distortions, opportunity cost of funds etc.). We assume that both cost functions are strictly increasing, convex, continuously differentiable and such that $C'_M(0) = 0$ and $C'_M(1) = +\infty$ (and similarly for C_φ) to guarantee an interior solution. The optimization problem of the planner is

$$\max_{\varphi, \tilde{M}} \; \mathcal{W}^P_{\mathcal{A}}(\varphi, \tilde{M}) - C_M(\tilde{M}) - C_\varphi(\varphi) \tag{43}$$

for $P \in \{G, B\}$, that is, allocated through banks or the government. Let:

$$\varphi^{*,P}(\tilde{M}) \equiv \quad \arg\max_{\varphi} \mathcal{W}^P_{\mathcal{A}}(\varphi, \tilde{M}) - C_M(\tilde{M}) - C_\varphi(\varphi), \;\; P \in \{G, B\}$$

and

$$\tilde{M}^{*,P} \equiv \quad \arg\max_{\tilde{M}} \mathcal{W}^P_{\mathcal{A}}(\varphi^{*,P}(\tilde{M}), \tilde{M}) - C_M(\tilde{M}) - C_\varphi(\varphi^{*,P}(\tilde{M})), \;\; P \in \{G, B\}$$

We have three results in Lemma 8. First, the optimal size of the program is smaller when funds are allocated through banks. Intuitively, when funds are channelled through the banking system, their marginal gain is smaller (since banks select firms with a lower treatment effect). Therefore, for a given cost function, the optimal size of the program will also be smaller. Second, when firms are only heterogeneous in their debt levels and $\eta > 1$ (that is, there is misallocation), the optimal loan-size per firm is the same when funds are allocated through the banking system for any program size. Larger loans per firm do not affect the treatment effects differently, and thus have the same marginal gain regardless of how big the misallocation is. Finally, in the case of heterogeneous shock exposure, our model shows that the relationship between φ and misallocation is ambiguous, since the derivative of the treatment effect T_j with respect to φ is hump-shaped and depends on η, that is: banks lending more to firms in $[\underline{\eta}_B, \overline{\eta}_B]$ and the government further lending to firms in $[\underline{\eta}_G, \overline{\eta}_G]$ can have an ambiguous effect on the difference between the treatment effect of these two groups.

Lemma 8. *The optimal program size is smaller when funds are allocated through the banking system, that is, $\tilde{M}^{*,B} < \tilde{M}^{*,G}$. Moreover, for the case of firm debt heterogeneity with $\eta > 1$, the optimal loan per firm is smaller for any program size $\varphi^{*,B}(\tilde{M}) = \varphi^{*,G}(\tilde{M})$.For the case of shock exposure heterogeneity, $\varphi^{*,B}(\tilde{M})$ can be greater or smaller than $\varphi^{*,G}(\tilde{M})$.*

Proof. **Part 1: Program Size** We can write the solution to the problem of the government in (11) as a threshold strategy, where $l_j^{*,G} = 1$ for $T_j \geq \underline{T}(\tilde{M})$ and $\underline{T}(\tilde{M})$ solves $\int_{j:T_j > \underline{T}(\tilde{M})} dj = \tilde{M}$. Therefore, the optimal program size is determined by the following FOC (note that the LHS is decreasing and the RHS is stricly increasing in $\tilde{M}$ and satisfies the appropriate boundary conditions, so a solution to this equation exists and is unique):

$$\underline{T}(\tilde{M}^{*,G}) = C_M^{'}(\tilde{M}^{*,G})$$

Similarly, we can write the solution to the problem of the bank in (14) as a threshold strategy, where $l_j^{*,B} = 1$ for $\Omega_j \geq \underline{\Omega}(\tilde{M})$ and $\underline{\Omega}(\tilde{M})$ solves $\int_{j:\Omega_j > \underline{\Omega}(\tilde{M})} dj = \tilde{M}$. Let $\underline{\underline{T}}(\tilde{M})$ be the treatment effect at j such that $\Omega_j = \underline{\Omega}(\tilde{M})$. The optimal program size under bank disbursement is determined by the following FOC:

$$\underline{\underline{T}}(\tilde{M}^{*,B}) = C_M^{'}(M^{*,B})$$

since $\underline{\underline{T}}(\tilde{M}) < \underline{T}(\tilde{M})$ as the government chooses the maximum combination of treatment effects, $\tilde{M}^{*,B} < \tilde{M}^{*,G}$.

Part 2: Debt Heterogeneity. Let $\eta_j = \eta > 1$, $c_j = c$ such that firms are only heterogeneous in their debt level and there is misallocation. For a given $\tilde{M}$, the optimal choice of φ when the government allocates PPP funds is given by

$$\varphi^{*,P}(\tilde{M}) \equiv \arg\max_{\varphi} \int_{b_j \leq b^{*,P}(\tilde{M})} T\,dj - C_{\varphi}(\varphi)$$

Thus, $\varphi^{*,P}(\tilde{M})$ is determined by $\frac{\partial T(\varphi^{*,P}(\tilde{M}))}{\partial \varphi} = C_{\varphi}^{'}(\varphi^{*,P}(\tilde{M}))$ and is equal for $P = B$ or $P = G$.

Part 3: Shock Exposure Heterogeneity. We will show here particular cases where $\varphi^{*,B}(\tilde{M}) < \varphi^{*,G}(\tilde{M})$ and, alternatively, where $\varphi^{*,B}(\tilde{M}) > \varphi^{*,G}(\tilde{M})$. Let $\eta_j < 1$. The FOC that determines $\varphi^{*,P}(\tilde{M})$ is given by

$$\int_j l_j^{*,P} \frac{\partial T_j(\varphi^{*,P}(\tilde{M}))}{\partial \varphi} dj = C_{\varphi}^{'}(\varphi^{*,P}(\tilde{M})) \Rightarrow \int_j l_j^{*,P} \eta_j \left[\rho - b + \varphi\right]^{\eta_j - 1} dj = C_{\varphi}^{'}(\varphi^{*,P}(\tilde{M}))$$

As $\eta_j < 1$, the LHS is strictly decreasing and thus the FOC determines the unique optimum. Moreover, note that we can control the format of function C_{φ} to guarantee that $\varphi^{*,P}(\tilde{M}) < \overline{\varphi}$ (just assume that $\lim_{\varphi \to \overline{\varphi}} C_{\varphi}^{'}(\varphi) =$). Take any $\overline{\varphi}$ such that $\overline{\varphi} < \exp^{-1} -(\rho - b)$. Let $\underline{\tilde{\eta}} \equiv \frac{1}{-\ln(\rho - b)}$ and $\tilde{\eta} \equiv \frac{1}{-\ln(\rho - b + \overline{\varphi})}$. If $\eta_j < \underline{\tilde{\eta}}$, we have that $\frac{\partial T_j(\varphi^{*,P}(\tilde{M}))}{\partial \varphi} dj$ is decreasing in η and thus for larger values of φ the gains of allocating through the banking sector are smaller at the margin, and thus $\varphi^{*,B}(\tilde{M}) < \varphi^{*,G}(\tilde{M})$. On the other hand, if $\eta_j > \tilde{\eta}$, $\frac{\partial T_j(\varphi^{*,P}(\tilde{M}))}{\partial \varphi} dj$ is increasing in η and thus for larger values of φ the gains of allocating through the banking sector are larger at the margin, and thus $\varphi^{*,B}(\tilde{M}) > \varphi^{*,G}(\tilde{M})$. ■