Revenue-Based Financing*

Dominic Russel† Claire Shi‡ Rowan Clarke§

October 20, 2023

Abstract

We use data from a major South African payment processor to study how digital payments mitigate asymmetric information challenges in small business "revenue-based financing" contracts, which tie repayment schedules to future revenue. Eight months post-financing, digital payments through the processor are 15% lower for takers than observably similar non-takers. We show this "gap" can be decomposed into three components: moral hazard from revenue hiding, adverse selection, and the causal effect of financing for takers. Two natural experiments suggest that takers shift more revenue off the platform when competition increases (moral hazard), and that financiers can increase repayment by waiting longer before extending offers (adverse selection). With estimates from both experiments, we bound the gap components, finding substantial adverse selection, but also positive short-run causal effects. Our results suggest digital payment platforms with "sticky" features can alleviate classic risk-sharing frictions by imposing hiding costs and limiting hidden information.

JEL codes: D21, D22, G20, G23

Keywords: small-business lending, moral hazard, adverse selection, digital payments, FinTech

*We are especially grateful to John Campbell, Edward Glaeser, Sam Hanson, Nathan Hendren, Rem Koning, David Scharfstein, Andrei Shleifer, Jeremy Stein, Adi Sunderam, Boris Vallée, and Yifei Wu for helpful comments and feedback. We gratefully acknowledge the Tech for All Lab within the (D^3) Digital, Data, and Design Institute at Harvard for data access and assistance. Russel and Shi contributed equally to this work.

†Harvard Business School: drussel@hbs.edu

‡Harvard Business School: cshi@hbs.edu

§Harvard Business School: rclarke@hbs.edu

Small business owners often rely on their firms for income, exposing them to revenue volatility and potentially discouraging them from pursuing risky, high expected-return investments. In perfect financial markets, risk-neutral financiers would fund these investments with state-contingent contracts; however, informational and incentive frictions hinder their existence (Arrow, 1974; Townsend, 1979).[1,2] These frictions are especially acute for small businesses as state-verifying audits can be impractical and costly. Nevertheless, in the past decade, a number of providers have introduced "revenue-based financing" contracts for small firms that tie repayment schedules to future revenues (see Rush, 2021, for one overview). Revenue-based financing is often implemented using digital payments, suggesting these technologies are important for overcoming classical challenges.[3]

Do digital payments mitigate asymmetric information challenges in state-contingent contracting? And if so, how? These questions are important for evaluating the long-term potential of such contracts, and for discerning the circumstances under which they will be more or less successful. In this paper, we use theory and transaction-level data from a major South African financial technology platform to explore these issues.

We organize our analysis with a stylized model in which small firms can raise capital for a risky investment. In the model, state-contingent contracts increase investment compared to debt but are hindered by moral hazard if borrowers can hide revenue from the financier after accepting the contract. Unlike in models with costly state-verifying audits (e.g., Townsend, 1979), digital payments provide *incentives* for the borrower to truthfully reveal their revenue to the financier, limiting moral hazard. This captures a key force: hiding payments from a digital platform is costly because the borrower loses access to the platform itself. For example, if repayment is automatically deducted from sales on an e-commerce website, the business must find customers off the website to avoid repaying. If repayment is through a payment processor, the firm must switch to cash or buy an alternative processor. The model also shows that state-contingent contracts suffer from an adverse selection problem, as they attract firms with hidden information about low future revenue.

[1] Arrow (1974) writes that "Such contingent markets are not entirely unknown...but they are relatively rare. Why this should be so follows again from the general problem of information costs and dispersal." Townsend (1979) notes that "there are few contingent dealings among agents relative to those suggested by theory" and explains this with a model in which agents are asymmetrically informed about the actual state of nature.

[2] For an overview of the evidence that risk aversion prevents small and medium enterprises from making risky, high-expected return investments in developing countries see Woodruff (2018) and De Mel *et al.* (2019).

[3] One provider highlights the importance of digital payments, stating "Revenue Based Financing is the best way to finance software vendors SaaS or e-commerce, subscription-based digital business models, D2C brands, service companies, and more generally for all digital companies. These activities benefit from good visibility on cash flows...".

Even with perfect state verification, this can cause the contract to unravel.

This conceptual framework makes two concrete predictions about digital payments and the performance of state-contingent contracts. First, due to a reduction in moral hazard, the contracts should perform better when there are higher costs associated with hiding revenue from a digital payment platform. Second, due to challenges with adverse selection, repayment should be higher in settings where the digital platform can limit hidden information about future revenue.

We test these predictions using data on over 100 million transactions from a South African financial technology platform. The platform processes payments and provides small businesses "capital advances" with repayment schedules based on future revenue. When a business takes an advance, the platform deducts a constant share of their daily transactions until the principal plus a fixed fee is repaid. An increase in processed revenue results in faster repayment, while a drop leads to slower or no repayment. This contract moves risk from the merchant to the financier, as there are no additional interest charges if revenue falls and no required repayment if the business fails.[4]

Our data includes processed transactions for advance takers and *non-takers* allowing us to compare these businesses. We show that the ex-post revenue "gap" between the takers and observably similar non-takers can be decomposed into three components: adverse selection, moral hazard in revenue hiding, and the causal effect of the contract on takers. Eight months after an advance, the revenue of takers is 15.4% lower than observably similar non-takers, consistent with the existence of asymmetric information. This gap is primarily driven by an intensive margin decrease in the revenue of takers relative to non-takers, rather than by takers being more likely to leave the platform. Even conditional on no default, the average revenue for takers is 14.4% lower than the matched control.

To test our first prediction, that digital payments mitigate moral hazard through hiding costs, we use a shock to a rival processor's pricing induced by funding from a World Bank Group member to "make digital payments systems more affordable." We first present evidence that capital takers shift large transactions to other processors, extending repayment time. We then use a difference-in-differences design that compares takers in areas where the rival does and does not operate, looking at post-advance revenue before versus after the shock. The funding shock for the competitor led to a post-advance decrease in takers' transactions of 10–15%, providing evidence that hiding responds to the cost of alternative processors. We also show the usage of two platform features, opening a

[4]We discuss the contract in greater detail in Section 1.

"manage" tab and exporting sales, are predictive of advance performance. This suggests platforms can improve repayment with "sticky" features which make it more costly to hide transactions.

Next, we explore our second prediction, which implies digital platforms must find ways to limit hidden information challenges associated with state-contingent contracts. We first note that features directly related to digital payments and platform usage—time on platform, pre-advance transaction volume and volatility—are all strongly predictive of advance performance, conditional on firm and advance characteristics. The predictive power of digital payment information is consistent with existing evidence on cashless payments and lending (Ghosh *et al.*, 2021). We then highlight two ways a financier can mitigate adverse selection: waiting longer before extending offers and repeat financing. Due to a temporary system error, businesses that joined the platform after March 20, 2022 and met minimum activity requirements were not offered an advance for six months, instead of the usual three. We show that controlling for pre-advance revenue, the quarter post-advance revenue of businesses that take an advance in their first year on the platform is 8% higher after the change compared to similar businesses before the change. We also show that repeat advances are 4.2 percentage points less likely to default than first-time advances conditional on other factors, supporting "learning-by-lending" to limit adverse selection (e.g., Botsch and Vanasco, 2019).

To quantify the importance of the forces highlighted by our predictions, we use evidence from both our natural experiments and bound the role of the three gap components—moral hazard in revenue hiding, adverse selection, and the causal effect of the advance on takers. We show variation from the temporary delay in offers allows us to estimate short-run adverse selection under an assumption of similarity between businesses who joined the platform before and after the error. Our point estimates imply adverse selection is roughly 60% the size of the gap, but is noisily estimated. If we additionally assume businesses shift 10% of their post-advance revenue (the estimated magnitude of response to the rival price drop),[5] our results imply a three-month positive causal effect equivalent to a 8% increase relative to the average transactions amount pre-advance.

Finally, we discuss what our estimates imply about the cost of revenue hiding, a revealed preference measure of how valuable businesses find the platform. Our back-of-the-envelope calculations suggest that if takers hide 10% of revenue, the total cost of hiding all additional transactions for the

[5]As discussed in Section 6 the size of shifting is inversely related to the size of the causal effect. Thus a smaller estimate of shifting provides a more conservative estimate of the causal effect. The magnitude of the post-period response bounds the *level* of post-period hiding from below (because businesses cannot negatively hide).

average taker is bounded below by 40,000 ZAR, slightly higher than their monthly revenue. This relatively large cost suggests customers find different processors to be imperfect substitutes, or that other factors—such as fear of being "caught" or moral considerations—play an important role.

Our work speaks to literatures on costly state verification (e.g., Townsend, 1979; Webb, 1992; Winton, 1995; Bond and Crocker, 1997) and equity-like financing in variety of settings (e.g., Friedman, 1955; Leland and Pyle, 1977; Alfaro and Kanczuk, 2005; Herbst and Hendren, 2021). We tie these ideas to a recent literature on the role of technology in business lending (e.g., Ghosh *et al.*, 2021; Howell *et al.*, Forthcoming) and a broader discussion on the increasing importance of non-bank and FinTech lending (e.g., Chen *et al.*, 2017; Schneider *et al.*, 2020; Gopal and Schnabl, 2022; Teima *et al.*, 2022). Rishabh and Schäublin (2021) study sales-linked loans in India and find evidence of revenue hiding immediately after disbursement for certain repeat borrowers. Relative to this work, we provide clarity on the asymmetric information challenges faced by state-contingent financing and the conditions that mitigate them. We also estimate adverse selection using a natural experiment, adding to recent methods that use price variation (Einav *et al.*, 2010; Einav and Finkelstein, 2011), surveys (Hendren, 2013; Herbst and Hendren, 2021), or experiments (Karlan and Zinman, 2009).

This paper is further motivated by an emerging literature on flexible repayment and micro-equity in development economics (e.g., Battaglia *et al.*, Forthcoming; Cordaro *et al.*, 2022; De Mel *et al.*, 2019; Woodruff, 2018). As in our model, this literature suggests risk aversion prevents small businesses from making positive NPV investments, motivating risk-sharing contracts. In the field, many of these contracts are also implemented using financial technology.[6] Our work provides insight into how and when these technologies can successfully overcome classical frictions.

The rest of the paper is structured as follows. Section 1 describes the setting and platform's revenue-based financing contract. Section 2 presents the model. Section 3 introduces the data and provides evidence of asymmetric information. We then test our model's predictions about moral hazard in Section 4 and adverse selection in Section 5. Section 6 quantifies these forces and the causal effect by decomposing the revenue gap between takers and non-takers. Section 7 discusses what our estimates imply about the cost of hiding transactions. Section 8 concludes.

[6] See, for example, Cordaro *et al.* (2022) who notes that they "leverage innovations in technology and digital finance that improve the observability of microenterprise performance."

1 Setting

Our data comes from a major South African financial technology platform ("the Platform"). The Platform offers payment processing machines and online interfaces, processing over \$1 billion USD in transactions per year for 250,000 active small business users. These users are estimated to represent 10% of all South African small and medium-sized enterprises (SMEs).[7] The Platform takes a small percentage (2-3%) of each transaction before depositing the rest into business bank accounts.

In addition to processing payments, the Platform offers a "capital advance" product in which a borrower pays back an advance with their daily processed payments. Advance offers are extended to small businesses that have processed transactions for at least three months and meet minimum activity requirements.[8] Since 2018, the Platform has issued over 40,000 advances and re-advances. In general, a large majority of South African SMEs are self-funded, and many face financing challenges, as there is little small business lending from banks.[9,10]

The Platform's advance offers consist of a principal, a factor rate, and a charge rate. If the business takes the offer, the Platform deposits the principal amount within one business day. The Platform then automatically takes a share of each future transaction (the charge rate) that goes through their platform until the principal $\times$ factor rate is paid off. There is no fixed term and repayment is quicker if revenue increases and slower if revenue falls. There are no additional fees or interest for slower repayment and the Platform only has a claim on the revenues they process (as opposed to revenue in cash or on other processors). Two partner companies, who observe past transactions and time on platform only, make underwriting decisions and provide the capital.

For illustration, consider an advance with a \$2000 principal, a factor rate is 1.3, and a charge rate of 20%. When the offer is accepted, \$2000 will be directly deposited into the business' account. To repay the advance, the Platform then deducts 20% off each transaction they process until these

[7] Estimates of the total number of SMEs in South Africa vary widely. While tax and registration data from 2016 pointed to only around 250,000 small businesses *total*, this excludes a large informal sector in which firms remain unregistered and bypass taxation (Small Business Institute, 2019). By other estimates, the number of total SMEs is over 2 million (OECD, 2022; United Nations, 2023).

[8] As of July 2023, these requirements were a minimum of 18 card transactions over the last 90 days and a monthly turnover of more than 3,000 ZAR per month.

[9] In one survey, South African SMEs reported obtaining financing as their second most important obstacle, behind only unreliable electricity supply (World Bank, 2021). Among exiting SMEs, 22% said they had challenges due to "problems getting financing" (United Nations, 2023). Another survey reports that 6% of SMEs receive government funding, 9% receive non-government funding, and only 2% are reliant on banks (SME South Africa, 2018).

[10] Access to finance has also become important for businesses to invest in backup power solutions (e.g., generators and solar systems), given South Africa's recent surge in rolling blackouts (Clarke *et al.*, Forthcoming; Kozak, 2023).

deductions sum to $\$2000 \times 1.3 = \$2600$. If the business processes \$1500 in transactions per month, the \$2600 is paid over 8.7 months. If, instead, the business processes \$1000 in transactions per month, the same \$2600 is paid over 13 months, lowering the implicit APR. We provide summary statistics on the terms of the contract in Section 3.

This contractual structure provides risk sharing in two ways. First, the flexible duration lowers the present value of repayments when revenues are low. Second, the lack of default consequences (as the Platform only has a claim on the revenue they process) insures businesses against failure. The contract is also motivated by the fact that the processor earns a small percentage of all transactions independent of advances, so aims to minimize exits. An additional late payment penalty could drive businesses away if they fall behind.

Globally, while other payment processors use similar technology to make advances and collect repayment, the structure of these contracts varies widely, changing the amount of risk sharing. For example, several U.S. payment processors require a personal guarantee and minimum monthly payments if transaction volume falls below a threshold.[11] These contracts then function more similarly to traditional debt contracts. Conversely, some non-processors use payments data to offer revenue-based financing.[12] We believe the Platform provides an ideal setting due to its size and its contract design, which offers substantial repayment flexibility compared to debt.

2 Conceptual Framework

In this section, we present a stylized model that helps explain the connection between digital payments and revenue-based financing. We use the model's framework and predictions to structure our empirical analysis. All proofs are in Appendix A.

2.1 Setup

Businesses can undertake a risky investment project of cost L. Conditional on undertaking the project, they earn a stochastic revenue payoff $\tilde{y} \sim \mathcal{N}(\mu, \sigma^2)$. If they decide to not undertake the investment they have a fixed revenue y.[13] Businesses are risk averse with CARA utility over revenue

[11] Square and Stripe can directly debit from business bank accounts to collect this minimum payment.
[12] See, for example, Ritmo, Uncapped, Pipe, and Vitt.
[13] We assume that $\mu \gg 0$ so the probability that $\tilde{y} < 0$ is negligible.

so that $E[u(\tilde{y})] = \mu - \gamma\sigma^2$. Assume $\mu > L \geq u(y)$ so there is a risk-neutral benefit to investing. However, if the borrower's risk-aversion γ is high enough, this can prevent some positive NPV investments from being made. Lenders are risk-neutral and make zero profits.

2.2 Risk Sharing

There are two types of contracts available to businesses to finance their investment: debt, in which borrowers receive and repay L, and a revenue-sharing contract, in which borrowers repay a share, η, of their revenue after investment.[14] Lenders make zero profits which pins down η. If there exists no $\eta \in [0, 1]$ such that the lender breaks-even, revenue-based financing will "unravel."

Proposition 1. *With only debt contracts available, there exist thresholds $\overline{y}_d$ and $\overline{y}_r$ such that for all $y < \overline{y}_d$ the debt contract is taken and all $y < \overline{y}_r$ revenue-based financing is taken. Then, $\overline{y}_d < \overline{y}_r$.*

Intuitively, revenue-sharing contracts move risk to the lender, decreasing the variance of the investment payoff for the borrower. Less is paid back when revenue is low, more is paid back when revenue is high. This can attract new risk-averse investors to accept revenue-based financing.

2.3 Moral Hazard

The classic challenge associated with state-contingent contracts is state verification. Townsend (1979) introduces "costly" state verification, whereby an agent learns the true state only after incurring an audit cost. In contrast, digital providers obtain audit-like information not by paying a cost, but by providing *incentives* for borrowers to reveal their true revenue. For instance, to continue using a specific payment processor or e-commerce platform, a borrower allows the provider to monitor and automatically deduct their revenue for repayment. Increasing the cost of shifting revenue off the platform—e.g., by adding more valuable features for the businesses or lending in an environment with few alternative forms of payment available—improves state-verifiability.

To illustrate, we let businesses hide revenue from the lender so that only a fraction $v(c)$ of revenues are observable. Here, c is the cost of revenue hiding determined by the types of factors

[14] This abstracts from the Platform's contract, in which the financier captures a share of payments until a fixed repayment threshold is met. Our stylized framework allows us to capture the two central forces of the contract. First, as the Platform does not hold a claim on the business if it never reaches the repayment threshold, the amount owed is in practice often lower than the threshold. Second, when revenue falls the *present value* of repayments falls even if the payment threshold does not change, lowering the time-discounted cost of the contract. Indeed, our model can capture this force by letting $\tilde{y} = \sum_{t=0}^{\infty} R^{-t}\tilde{y}_t$, the present value of all future revenue with some discount rate R.

described above. $v'(c) > 0$ as borrowers' returns to hiding decrease in the cost of cost of hiding. The cost c determines the feasibility of revenue-based financing, as shown in Proposition 2.[15]

Proposition 2. *With moral hazard, there exists some $\bar{c}$ such that for all $c < \bar{c}$, revenue-based financing is impossible.*

Intuitively, when $v = 0$, the lender would not offer revenue-based financing due to the guaranteed loss of L on every contract. The simple debt contract doesn't experience this issue, as repayment isn't tied to performance. In our analysis, we'll provide evidence that reduced payment processing competition and the provision of other "sticky" add-on features — which can increase the costs of hiding — each improves repayment.

2.4 Adverse Selection

Even in the absence of moral hazard, revenue-based financing faces adverse selection challenges. In particular, businesses who know they will have low future revenue find revenue-based financing attractive. In the extreme, they consume L and repay nothing. No monitoring technology would stop these "bad types" from selecting in. To illustrate, suppose there are two types of businesses:

- "Good types" face investment opportunities with expected payoff μ and variance σ^2.
- "Bad types" have no investment opportunities and instead consume L.

The lender cannot distinguish between good and bad types, conditional on characteristics X, but knows that $P(G|X) = p$ and $P(B|X) = 1-p$. The lender does not like lending to bad types because they make $-L$ on each contract. Assume that $\mu > 2\gamma\sigma^2$ so good types do not prefer to give up a share of the investment altogether.[16]

Proposition 3. *With adverse selection, there exists a $\bar{p}$ such that for all $p < \bar{p}$, revenue-based financing is impossible.*

[15] An alternative form of moral hazard associated with equity contracts arises in a principal-agent setting where managers reduce effort, lowering the investment returns. Here, we hold the contract fixed and model the manager's incentives only. Empirically, shirking would decrease $Y(1)$ in Definition 1 which appears in both the causal effect and moral hazard terms.

[16] For more details on the need for this technical assumption see the proof of Proposition 3.

Improvements in the screening technology enable the lender to more effectively screen out bad types, increasing p and making revenue-based financing more viable. Intuitively, good types have to cross-subsidize the bad types for the lender to break even, increasing η. If η becomes too high, the cost of financing for good types becomes too expensive and they select out, unraveling the market. Debt contracts do not face the same adverse selection issues as the repayment amount is independent of future revenue. In our analysis, we will provide evidence that digital payments have predictive content and that a longer history of payment data better predicts performance.

3 Facts & Evidence of Asymmetric Information

In this section, we first summarize our data, which includes processed transactions for both advance takers and non-takers. We then show the difference in revenue between takers and observably similar non-takers can be decomposed into three components: moral hazard in revenue hiding, adverse selection, and the causal effect of the advance on takers. We estimate this difference empirically.

3.1 Data & Summary Statistics

As described in Section 1, our data comes from a South African financial technology company that processes payments and offers capital advances. We have all transaction-level payments for both advance takers and others, including the size, amount, and time of transactions. We also observe all information on advances, including the principal amount, pricing components, and repayments. Lastly, we have business-level information including location, industry, and owner demographics. We focus on advances made from June 2020 onwards (to exclude the highest level of COVID-19 lockdown) and for which we can observe for 12 months post-advance as of August 2023.

Table 1 provides summary statistics on the 10,136 first advances and 13,014 repeat/re-advances.[17] Advance takers are generally small, consumer-facing businesses (e.g., food trucks, hair salons). First advance takers have an average of 110,000 in South African Rand ($\sim\$6,000$ USD) in sales over the prior three months.[18] The average principal of first advances is around one month's worth of revenue. Combined with the average charge rate of 20% and factor rate of 1.3, this implies an estimated

[17] The processor considers a repeat advance an additional advance taken out after fully paying off previous advances. By contrast, a re-advance is taken out *before* fully paying off previous advances, adding to the outstanding amount due. We will generally refer to both as repeat advances.

[18] By purchasing power parity, 110,000 ZAR is roughly \$16,000 USD.

repayment period of 7-8 months.[19,20] On average, first-time capital takers paid 1.06 times their original principal, or 1.04 and 1.02 when each payment is discounted at annualized rates of 5% and 15% respectively.[21] Panel B shows that repeat advances have higher 1-year repayment: 1.22 times the original principal on average.

Figure 1 shows the outcomes of advance takers one year later. 21% of first-time takers have an open advance but no transactions in the last 30 or more days, nearly twice the rate of repeat takers (12%). The figure also shows that for both groups, despite the 7-8 month average estimated repayment period, a sizeable majority have some open advance one year later.

3.2 Predictors of Performance

We test the relationship between ex-ante observables and two outcome measures of advance performance. First, we say a business has *defaulted* if they have no transactions (and therefore no payments) in the eighth month after taking an advance.[22] Second, for businesses that did not default, *conditional revenue* is the sum of revenue over the eight months after taking the advance.[23] We begin exploring predictability by regressing each measure on ex-ante observables with:

$$Y_{i,t} = \mathbf{T}'_i \boldsymbol{\beta_1} + \beta_2 \text{First}_i + \mathbf{X}'_i \boldsymbol{\beta_3} + \delta_t + \epsilon_{i,t} \tag{1}$$

For advance i given in quarter t, $Y_{i,t}$ is either default or log conditional revenue. $\mathbf{T}_i$ is a vector of characteristics related to platform use and transactions in the three months before disbursement. First_i is an indicator for whether the advance is a first or repeat advance. $\mathbf{X}_i$ is a vector of comprehensive firm and loan characteristics.[24] δ_t are quarter by year fixed effects.

Table 2 shows that both time on platform and transaction volume before the advance are strong predictors of default and conditional revenue.[25] Our preferred specifications in Columns (3) and

[19] As described in Section 1, the charge rate is the fixed percentage of daily sales that is automatically deducted to pay back the advance. The factor rate times the principal is the total amount due (regardless of repayment timing).

[20] The implicit APR, if sales from the last three months stay constant, is 81%. To understand this, note first that a factor rate of 1.3 over 8 months implies a lump sum repayment APR of around 40%. However, because repayment is *daily*, early payments are made at a much higher effective APR. This roughly doubles the APR again (for related discussion, see Stango and Zinman, 2009).

[21] Repayment is daily and begins immediately, resulting in many payments not being substantially discounted.

[22] Appendix Figure C.1 provides hazard plots of this measure over the life of an advance.

[23] Eight months is the median estimated repayment period in Table 1.

[24] Firm characteristics: Fixed effects for industry segmentation, business type, owner citizenship, location classification, and province. Loan characteristics: linear controls for the principal, charge rate, and factor rate.

[25] The sample in this analysis is slightly larger than in Section 3.1 because we use all advances for which we can

(4) (which control for owner demographics, quarter, and advance characteristics) show that for first advances, an additional year on the platform corresponds to a 3.6 percentage point decrease in default and a 3.3% increase in conditional revenue. A doubling of the transaction volume in the three months prior corresponds to a 1.4 percentage point decrease in default. Table 2 also shows that a measure of ex-ante "stability", the weekly volatility of transactions in the months prior, positively predicts default.[26] Columns (5) and (6) suggest that these results hold when looking across all advances, but first advances are more likely to default than repeat advances. We will return to adverse selection and further discussion of repeat lending in Section 5.

3.3 Revenue Gap Between Advance Takers and Non-Takers

A natural empirical object to test for the existence of asymmetric information is the revenue "gap" between takers of the advance and observably similar non-takers. To understand the determinants of this gap, let $Y(1)$ and $Y(0)$ be potential revenue for a business with and without an advance. Takers can be seen as "compliers" who will take the advance if eligible and non-takers as "never takers." This framework provides a straightforward tie between the model forces described above and the ex-post revenue gap between observably identical businesses that do and don't take advances.

Definition 1. *Suppose X is the set of characteristics observed by the financier. The conditional Gap is:*

$$Gap|(X = x) \equiv \mathbb{E}[Y(0)|X = x, \textit{Non-Taker}] - \mathbb{E}[vY(1)|X = x, Taker] \tag{2}$$

*The **Gap** is:*

$$Gap \equiv \int (Gap|X = x) \cdot f(x|Taker)dx \tag{3}$$

When advances are randomly assigned and $v = 1$, the Gap provides an estimate of the (opposite signed) average treatment effect of advances among the takers (ATT). However, asymmetric information invalidates this interpretation.

Definition 2. *Define adverse selection, the causal effect of the loan on takers, and moral hazard in*

observe 8 months post-advance—rather than 12 months post-advance—as of August 2023.

[26]Note that this measure predicts default, but not conditional revenue. This is consistent with ex-ante volatility predicting ex-post volatility (i.e., "stability" or the frequency of negative shocks). Ex-post volatility would make a business more likely to default, but conditional on staying alive, not predict performance.

the following way:

$$AS \equiv \mathbb{E}[Y(0)|\textit{Non-Taker}] - \mathbb{E}[Y(0)|\textit{Taker}] \tag{4}$$

$$CE \equiv \mathbb{E}[Y(1) - Y(0)|\textit{Taker}] \tag{5}$$

$$MH \equiv \mathbb{E}[(1-v)Y(1)|Taker] \tag{6}$$

Then, by linearity of expectation, given $X = x$:

$$Gap|(X = x) = (AS - CE + MH)|(X = x) \tag{7}$$

A large Gap provides evidence for the existence of moral hazard, adverse selection, and/or negative causal effects. If business owners are risk averse and rational, they will only accept the advance if the expected value of taking the advance is greater than not, generally leading to non-negative causal effects.[27] With non-negative causal effects, the Gap bounds adverse selection and moral hazard from below. To study these forces, we next estimate the size of the Gap empirically.

3.4 Gap Estimation

To estimate the Gap we match each first-time advance-taking business to their nearest "control" non-taker.[28] In particular, for each taker, we find a match non-taker in the same month and industry who met the minimum advance eligibility requirements and is closest in terms of time on platform and transaction amount in the month before the advance (according to normalized Euclidean distance). The average difference in post-advance revenue between takers and their matched control businesses then provides an estimate of the Gap defined in Equation 3. Figure 2 shows that the revenues of the takers and matched control diverge in the year after the advance, consistent with the existence of adverse selection and moral hazard.[29] Average revenue of the advance takers is 15.4% lower than the matched control eight months after taking the advance. In Appendix B, we show that panel regression and machine learning approaches provide similar estimates.

[27]One force that would lead the causal effect to be negative is a reduction effort as described in footnote 15. However, unlike a pure equity contract, the owner retains full control of business revenues after the advance is paid off, reducing the potential for a negative impact on effort.

[28]In our baseline results we use $K = 1$, but our results are nearly identical when averaging across larger sets of K matched neighbors.

[29]Here we use individual monthly outcomes, analogous to letting $Y(1)$ be a vector of post-period outcomes.

Average outcomes in Figure 2 can be driven downward by intensive margin revenue decreases or exits from the platform. Panel (A) of Figure 3 separates out the former by limiting to matched pairs in which both the taker and control business transacted in the eighth month. The advance takers' average revenue is 14.4% lower than the matched control in month eight, only slightly smaller than the unconditional gap in Figure 2. Accordingly, Panel (B) shows that advance takers are only 6% more likely to disappear. These results suggest that "running away"—taking the advance with the intent to close or no longer use the processor—is relatively less important than the drop in intensive margin revenue. This is consistent with a story in which removing marginal transactions off the platform is less costly than switching away entirely (e.g., because some, but not all, transactions can be shifted to cash), supporting the intuition of Proposition 2. We explore this proposition next.

4 Moral Hazard from Revenue Hiding

Do advance takers reduce the cost of financing by "hiding" revenue using alternative processors or cash? In this section, we explore this question and test Proposition 2's prediction that the feasibility of revenue-based financing depends on the hiding costs associated with a digital payment platform. To do so, we use a natural experiment that led the Platform's rival to reduce prices. We also explore the relationship between advance performance firms' usage of add-on platform features.

4.1 Competition

Unlike its competitors, the Platform charges no fixed or "daily settlement" fees which makes their product relatively cheaper for smaller transactions. This, combined with the fact that moving the largest transactions off the platform is the most effective way (per transaction) to extend the advance duration, suggests larger transactions could be shifted to other processors. Figure 4 shows that indeed, large transactions more sharply decline after an advance.[30]

While this post-advance decline in large transactions is suggestive of hiding, this finding could also be driven by takers having private information about large future sales. To address this issue, we use a natural experiment. Note that if advance takers are incentivized to shift transactions to another processor, they should be more inclined to do so when a rival reduces its prices. We focus on

[30]This result is unchanged when residualizing on business revenues, suggesting that within-business variation, rather than across-business variation, drives our result.

a primary rival ("the Competitor") of the Platform that offers similar payment processing products. The Platform operates out of Cape Town, whereas the Competitor started in Durban and primarily operated in the surrounding region before expanding recently. In March 2021, the Competitor's parent company received $15 million USD from a World Bank Group member to "make digital payments systems more affordable."[31] Accordingly, they cut the price of their flagship product by more than 50% over the next six months, as shown in Appendix Figure C.2.

As the Competitor reduced its prices, it became cheaper for the Platform's advance takers in areas around Durban (where the Competitor operated) to purchase another processor and hide transactions. If advance takers shift revenue to rival processors, one would expect a greater post-advance decline in the number of transactions after price cuts among the Platform's users in these areas relative to their counterparts around Cape Town. To test this empirically, we employ a difference-in-differences regression framework with the following specification:

$$Y_{i,t} = \alpha_1 D_i + \sum_{t \neq 3} \beta_t \left(D_i \times \text{Quarter}_t\right) + \mathbf{X}_i' \boldsymbol{\alpha}_3 + \text{Quarter}_t + \epsilon_{i,t}. \tag{8}$$

For business i who took a first advance in quarter t, Y_{it} is the ratio of the average monthly number of transactions in the eight months following the advance over three months before the advance. D_i is an indicator for whether borrower i is in the Kwazulu-Natal and Eastern Cape provinces around Durban or the Western and Northern Cape provinces around Cape Town. $\mathbf{X}_i$ are controls for pre-advance transaction amount and industry.

Figure 5 shows that β_t decreases after the price cut, consistent with revenue hiding caused by the Competitor's price drop. Borrowers exposed to the price drop experienced a 10–15 percentage point decrease in transactions on the Platform.[32] β_t increases as the Competitor expands throughout South Africa in 2022. In line with the mechanism suggested in Figure 4, Appendix Figure C.4 shows that the same difference-in-differences specification appears somewhat stronger for large transactions. Our results suggest that advance takers hide revenue, at least in part, by moving transactions to rival processors. When there are fewer incentives to keep transactions on the Platform, moral hazard makes revenue-based financing less viable financier.

[31] See the archived press release here.

[32] Appendix Figure C.3 shows the corresponding raw averages of R_{it}. The counterfactual mean is close to one in the post period, which implies that a 10 percentage point decrease is approximately a 10% decrease in transactions on the Platform. Our estimate of α_1 is 0.049.

4.2 "Sticky" Features

In addition to price, the cost of using an alternative processor is determined by any other factors that limit substitutability. "Sticky" features (e.g., add-ons, a higher quality machine, better customer support) that provide value to businesses *outside* the advance, then, may have the potential to reduce moral hazard. Consider, for example, a business that uses a processor's accounting platform which is automatically tied to payments. Shifting transactions would disrupt their accounting, as only a fraction of transactions would be recorded. Moving transactions entirely off the platform would require an entirely new accounting system.[33]

We test this intuition by focusing on two interactions with the platform: (1) opening a "manage" tab on the platform app that allows businesses to track their staff, customers, and inventory; and (2) clicking a button to export sales history to a CSV file. The latter feature is commonly used by businesses with third-party management tools. Accordingly, we predict businesses to have a *higher* cost of shifting when they use the manage feature, and a *lower* cost of shifting if they export sales.

We add usage of these features to Equation 1 in Section 3.2 to test their relationship with advance performance. Columns (1) and (2) in Table 3 show that among takers, the "manage" measure is positively predictive of both default and conditional revenue.[34] In particular, pre-advance usage of the feature predicts a 1.3% percentage point decrease in default probability and a 5.4% increase in transaction amount. It is possible, however, that those who use the feature are better performing in general. We use the second measure, exporting sales, to address this concern. If businesses that use external accounting are better performers, exporting sales should *positively* predict performance. But if this feature decreases the cost of shifting, exporting sales should *negatively* predict performance for takers compared to non-takers. To test this, in Columns (3) and (4) we include non-takers and interact usage of the export feature with being a taker.[35] The interaction on default is significantly positive, a *negative* impact on performance, consistent with decreased shifting costs.

In sum, our results suggest that revenue hiding on digital platforms is mitigated by "sticky

[33]Similar sticky features exist in a variety of "embedded finance" settings where platforms offer capital to small businesses. For one overview of embedded finance, see Dresner *et al.* (2022).

[34]The features we analyze were introduced in February 2022 and June 2021, respectively, reducing the sample sizes for these analyses.

[35]To include non-takers, we use the eight months of outcomes starting one year after they joined the platform (as if they counter-factually took an advance at this time). One year is approximately the median time on the platform before advance for takers. Our results do not significantly change when using other times on the platform.

features" which provide users value and make it more costly to hide transactions. They also suggest platforms can improve the performance of revenue-based financing by providing these features and screening on businesses' usage of them.

5 Adverse Selection

Even without the moral hazard discussed in the prior section, our stylized framework's Proposition 3 shows that hidden information can cause revenue-based financing contracts to unravel. In this section, we explore how digital platforms might address this adverse selection. We first briefly revisit our results on predictors of performance, then explore evidence on two additional interventions.

As discussed in Section 3.2, Table 2 shows that three measures directly related to digital payments and platform usage—time on platform, pre-advance transaction volume and volatility— predict default. Columns (1) and (2) include only these predictors, while Columns (3) and (4) add many controls for business characteristics, loan characteristics, and quarter-by-year fixed effects. This large number of additional controls increases the adjusted R^2 by no more than 50%, consistent with measures linked to digital payments and platform usage having meaningful predictive power. In turn, this suggests verifiable digital information can potentially reduce adverse selection, a notion supported by Ghosh *et al.* (2021). To provide insight into two additional features of this context that might limit hidden information, we next use both quasi-experimental and observational evidence.

5.1 Waiting Longer Before Financing

Given advance performance increases with time on the platform, should platforms wait longer before offering advances? A digital platform may benefit from observing a longer history of payments before offering financing, as this could reveal information about business quality and reduce adverse selection. Appendix Figure C.5 supports this, showing that time on platform not only predicts future revenue unconditionally, but also reduces the Gap between capital takers and observably similar non-takers. Waiting, however, might not be advantageous if the same "bad types" still demand financing as soon as they are eligible or if it significantly reduces demand from "good types."

To test this, we use a natural experiment that delayed advance offers to businesses on the Platform. In particular, due to a temporary system error, businesses who joined the platform between March 20, 2022 and March 1, 2023 and met the minimum activity requirements were not offered an advance until six months after joining, instead of the usual three.[36] Figure 6 shows the variation introduced by the initial offer change.[37] The top panel shows around 10% of businesses who joined before the change and were eligible to take an advance in month three did so before month six. Due to the system error, there are no early takers after the change. But the second panel shows that the number of advances made between months six and twelve increased, suggesting demand was pushed to later months. The final panel shows this increase in demand did not fully catch up by month twelve: the share of takers drops from 16% to 11%.

We next ask "does the drop in advances come from better screening or decreased demand from 'good types'?" We test this using advances made within 12 months of joining and the regression:

$$Y_{i,t} = \beta_1 Y_{i,t-1} + \beta_2 \text{AfterCut}_i + \mathbf{X}_i' \boldsymbol{\beta}_3 + \text{Month}_t + \epsilon_{i,t}. \tag{9}$$

Here, $Y_{i,t}$ is revenue in the quarter post advance, $Y_{i,t-1}$ is revenue in the prior quarter, and $\mathbf{X}_i$ are industry fixed effects. We control for seasonality with Month_t fixed effects. The coefficient of interest is β_2 on AfterCut_i which is an indicator for whether the business joined the platform after the March 20, 2022 cutoff date. Columns (1) and (2) of Table 4, show that controlling for pre-advance revenue and seasonal variation, the quarter post-advance revenue of takers is around 7,000 ZAR higher after the cutoff than before, an 8% increase in revenue relative to the overall average. To rule out that this variation is driven by other trends in advance performance over time, in Columns (3) and (4) we focus only on businesses that joined the platform in the six weeks immediately before and after the cutoff. The effect size remains similar, around 7,500 ZAR. Our results provide evidence that waiting six months instead of three improved screening.

5.2 Repeat Financing

A long literature underscores the importance of repeated interactions and firm-bank relationships in lending, especially for small businesses (Petersen and Rajan, 1994; Berlin and Mester, 1999; Berger *et al.*, 2005). A potential implication is that digital platforms could limit adverse selection

[36] A small number of new very high-revenue businesses continued to receive advance offers throughout this period. We exclude these businesses from these analyses.

[37] We focus on the initial change because, at most, we can only observe the first five months on the platform for businesses who joined after the latter change.

by following a "learning by lending" approach in which they start small and offer more credit over time (Botsch and Vanasco, 2019). Indeed, Table 1 shows that repeat advances have higher average returns than first-time advances, as discussed in Section 3. However, it is also possible that platforms with a long history of digital transactions learn nothing additional from repeat loan performance.

We test this possibility by using a sample of first-time and repeat advances and modifying Equation 1 to include an indicator for initial advance. Column (5) of Table 2 shows that even conditional on observables, first plans are 4.2 percentage points more likely to default than repeat advances. In Figure 7 we regress default on weeks on platform separately for first, second, and third or later advances. The downward slopes across all three groups show default decreases steeply with time on platform; however, the level of default decreases monotonically with plan number suggesting repeat advances provide additional information beyond time on the platform.

6 Separating Moral Hazard from Adverse Selection

To quantify the importance of the forces described in the prior sections, in this section we use the natural experiments described in Sections 4.1 and 5.1 to decompose the Gap into adverse selection, moral hazard from revenue hiding, and the causal effect (as each is defined in Equations 4–6). First, to build intuition, we describe a hypothetical experiment that identifies the combined size of moral hazard and the causal effect. Subtracting this quantity from an estimate of the Gap identifies adverse selection.[38] Second, we discuss the identifying assumptions necessary to use our natural experiment in Section 5.1 instead of the hypothetical experiment. Third, we produce a short-run (three months post-advance) estimate of adverse selection under these assumptions. Fourth, we use a series of back-of-the-envelope estimates on the overall level of revenue hiding—informed by our natural experiment in Section 4.1—to fully decompose the short-run Gap into its three components.

6.1 Natural Experiment

Ideal Experiment

Consider an experiment in which one randomly assigns eligible businesses into two groups. If neither group receives advance offers, their expected revenues will be the same. Instead, suppose one group

[38] To see this, simply subtract $(MH - CE)|(X = x)$ from both sides of equation 7.

receives offers and the other does not. While the observed revenue of non-takers in each group will not change, the observed revenue of takers in the offered group will change due to revenue hiding and the causal effect on true revenue. Adverse selection plays no role in the difference between groups, as they are randomly assigned. Proposition 4 formalizes this intuition.

Proposition 4. *If businesses are randomly assigned offers, the expected difference in reported revenue between those without offers and with offers is* $\mathbb{P}(\textit{Taker}) \times (MH - CE)$.

As noted above, subtracting this quantity from a Gap estimate identifies adverse selection.

Natural Experiment

Following the intuition of the ideal experiment, we use variation from the natural experiment described in Section 5.1—in which eligible businesses were given offers six months after joining the Platform instead of three months—to identify adverse selection in the three months post-advance. To map the natural experiment to the ideal one, observe that businesses who joined the Platform just before March 20, 2022, are in an "offer-receiving group" in their third to sixth months on the Platform, while those joining just after the cutoff form a "non-receiving group." The arbitrary cutoff randomly assigns businesses to either group. Thus, comparing the revenues of the groups can provide an estimate of moral hazard and causal effects, as Proposition 4 describes.

The necessary identifying assumption is that businesses on either side of the cutoff have the same counterfactual distributions of $Y|X$ and $\mathbb{P}(\text{Taker}|X)$. That is, if businesses who joined in the months after March 20 were instead offered financing at three months, their rate of uptake and expected revenue would be the same as observably similar businesses who joined before March 20. This assumption would be violated by seasonality in revenue. To address this, we use two full years of data—with businesses in the 18 months before the cutoff and businesses in the six months after the cutoff—and always include controls for month-of-year fixed effects.[39] To address the fact that businesses after the cutoff may have taken advances after month six, we look only at "short-run" three-month outcomes: for businesses that took an advance before month six, we use the next three months' revenue; for others, we use revenue from months four through six on the platform.

[39] We limit to six months because our transaction data end before we are able to analyze outcomes for businesses more than six months after the cutoff.

6.2 Estimation

Estimating Moral Hazard Minus the Causal Effect

We first estimate the difference in revenue between those offered and not with the regression:

$$Y_{i,t} = \alpha_1 Y_{i,t-1} + \alpha_2 \text{Offered}_i + \mathbf{X}_i' \boldsymbol{\alpha}_3 + \text{Month}_t + \epsilon_{i,t}. \tag{10}$$

$Y_{i,t}$ is observed three-month revenue, $Y_{i,t-1}$ is revenue lagged by one quarter, $\mathbf{X}_i$ are industry fixed effects, and Month_t are month fixed effects. The coefficient of interest is α_2 on the indicator Offered_i. Column (3) of Table 5 shows that with industry fixed effects α_2 is around -300. It is worth emphasizing, however, that the standard error on this estimate is relatively large and that our analysis should be seen as providing directional guidance rather than a sharp point estimate. Following Proposition 4, to recover $MH - CE$ we divide $-\alpha_2$ by the probability of taking an advance.[40] Dividing by the share of those offered that are takers, 11.1%, gives $ME - CE \approx$ 2,600.[41]

Estimating Adverse Selection

To estimate adverse selection, we next estimate the short-run Gap conditional on observables with:[42]

$$Y_{i,t} = \gamma_1 Y_{i,t-1} + \gamma_2 \text{Taker}_i + \mathbf{X}_i' \boldsymbol{\gamma}_3 + \text{Month}_t + \epsilon_{i,t}. \tag{11}$$

Variables are defined similarly to Equation 10. Table 5, Columns (2) and (4) show γ_2 (Gap) estimates of around 6,400 ZAR. Subtracting $ME - CE$ implies adverse selection of around 3,800 ZAR, accounting for roughly sixty percent of the Gap.

Estimating the Full Decomposition

Our ability to separate moral hazard from the causal effect is limited by the fact that the Platform cannot directly observe revenue hiding. However, Section 4.1 shows that in response to increased

[40] Under our identifying assumption $-\alpha_2 = \mathbb{P}(\text{Taker}) \times (MH - CE)$, then divide by $\mathbb{P}(\text{Taker})$.

[41] Alternatively, we can make in-sample predictions using the logistic regression $\text{Taker}_i = \text{Month}_t + \sigma_0 Y_{i,t-1} + \sigma_1 X_i$. Then the average of α_2 divided by $Taker_i$ yields an estimate for $ME - CE$. This method gives nearly identical results, consistent with the observables having little predictive power on whether an individual is a taker.

[42] This Gap differs from the Gap estimated in Section 3.4 and Appendix B because we only include businesses that (1) were eligible for an advance at the end of their third month on the platform, (2) took advances in their fourth through sixth month on platform, and (3) took in advance in the 18 months before the March 20, 2022 cutoff.

competition transactions fall by 10–15%, consistent with revenue hiding. We use this range as a back-of-the-envelope estimate of $1 - v$.[43] Assuming an identical v to all businesses allows us to identify $MH|(X = x) = (1 - v)\mathbb{E}[Y(1)|\text{Taker}, X = x]$. We can then estimate $CE|(X = x)$ as the Gap less adverse selection and moral hazard. Figure 8 illustrates what different assumptions about the overall level of revenue hiding v imply about the magnitude of the causal effect. For any $v < 0.975$, the causal effect of the advance for takers is positive. As businesses hide a greater share of revenue, the size of the causal effect increases.[44] For example, if advance takers hide 10% of their transactions, the short-term causal effect is a revenue increase of around 7,700 ZAR, an 8% increase in quarterly revenue relative to pre-advance for the average taker.

This section provides suggestive evidence of both substantial revenue hiding and adverse selection, but also positive short-run causal effects. These estimates highlight both the promise of revenue-based financing (Proposition 1), and the challenges (Propositions 2 and 3).

7 Value of the Platform

In this section, we discuss what our estimates imply about the cost of hiding transactions, a revealed preference measure of how valuable businesses find the Platform. Our analysis in Section 4.1 suggests that businesses hide transactions to slow repayment and that this hiding responds to the cost of outside options. Yet, Figure 3 shows there is also very little intentional extensive margin default, suggesting businesses place high value, at least implicitly, on the Platform. We now use our previous estimates to provide back-of-the-envelope guidance on the cost of shifting.

Assuming a discount rate for the value of payments made in the future, we can observe the gains from hiding an additional percent of revenue directly from the repayment schedule. We cannot directly observe the costs of shifting; however, if businesses are optimizing, the shape of the marginal benefit curve will place bounds on the marginal cost curve.

Define $h \equiv 100 \cdot (1 - v)$ as the percentage of revenue hidden and $g(h)$ as the present-value "gain" from hiding $h\%$ of transactions each day. In particular, $g(h^*)$ is the difference in the net present

[43] Smaller estimates of hiding result in smaller estimates of the causal effect. Our use of the hiding estimates from Section 4.1 is conservative, as it assuming no hiding in the Competitor's region after the price drop and no hiding in either region before the price drop. It is also conservative in that it assumes transactions of all sizes are equally likely to be hidden, whereas Appendix Figure C.4 provides evidence that larger transactions are more likely to be hidden.

[44] Note that this is the positive effect of the advance on *true* revenue, even if the financier cannot see it.

value of repayments between $h = 0$ and $h = h^*$. $g'(h)$ is the marginal benefit of hiding. Let $c(h)$ be the "cost" of hiding, and $c'(h)$ is the costliness of moving an additional percentage point of revenue off the platform. If businesses are optimizing and there is an interior solution, $g'(h^*) = c'(h^*)$.

For our back-of-the-envelope analysis, we let $h^* = 10$ as in Section 4.1, which implies $c'(10) = g'(10)$. We assume small businesses have an annual discount rate of $r = 30\%$ and constant daily revenues for simplicity.[45] Appendix Figure C.8 shows the corresponding $g(h)$ and $g'(h)$ using the characteristics of the average advance.[46] Under these assumptions, the marginal cost of moving an additional 1% of transactions each day is $c'(10) = g'(10) = 40$ ZAR.

Intuitively, Figure C.8 shows that $g'(h)$ is steep because the discount rate has a non-linear effect on payments far into the future. The marginal benefit of hiding is increasing in the percentage hidden (duration effect).[47] If the optimal amount of hiding is 10%, then $c'(h)$ must be below $g'(h)$ for $h < 10$. For $h > 10$, we need $\int_{10}^{100} c'(h)dh > \int_{10}^{100} g'(h)dh \Rightarrow c(100) - c(10) > g(100) - g(10)$, otherwise businesses would hide all of their revenue.[48] For example, if marginal costs rose faster than marginal benefits for $h > 10$, then $h^* = 10$ could be possible. Consequently, the total cost of additional shifting to justify $h^* = 10$ is bounded below by $g(100) - g(10) \approx 43{,}000$ ZAR.

This cost of shifting is relatively large, slightly higher than the average monthly revenue of advance takers. For comparison, the cost of switching to a competitor processor after the price drop described in Section 4.1 is 2,000 ZAR. This suggests customers may find different platforms to be imperfect substitutes (e.g., because of machine quality, customer support, switching costs, or "sticky" add-on features as discussed in Section 4.2). Other factors, such as fear of being "caught" or moral considerations, may also play an important role.

[45] We chose 30% based on several articles that value small businesses in practice (e.g., Mercer Capital) and academic work showing that the discount rate for small businesses should be fairly high, due to idiosyncratic risk. See, for example, Trevino (1997) and Jagannathan *et al.* (2016).

[46] Factor rate of 1.3, charge rate of 20%, and principal of $36,224$ (Table 1). We also assume that daily revenue is constant, given by $103,570/90$.

[47] This differs from a traditional equity contract in which the marginal benefit of hiding would be constant.

[48] In Figure C.8 we assume two functional forms for $c(h)$. If $c(h)$ was cubic and $c'(h)$ was in the form ah^2 starting at $(0,0)$ and passing through $(10, g'(h))$, then $h = 10$ could be optimal. However, quadratic costs and $c'(h) = bh$ (passing through $(0,0)$ and $(10, g'(h))$ would not work because $c(100) - c(10) \ll g(100) - g(10)$.

8 Conclusion

While revenue-based financing contracts offer risk-sharing benefits that can encourage small business investment, they also amplify asymmetric information challenges. We argue that digital payments technology has played a key role in mitigating these challenges and explore these forces using data on over 100 million transactions from a major payments processor in South Africa.

A key insight from our work is that revenue-based financing can be viable if a financier is able to incentivize borrowers to reveal their true revenue, limiting moral hazard. This intuition differs from costly state-verification models requiring audits to sustain state-contingent contracts (Townsend, 1979). Borrowers have incentives to reveal revenue to digital platforms because concealing payments can be costly. For example, if a small business's financing is automatically repaid from sales on an e-commerce website, the business must find customers off the website to avoid repaying. If repayment is through a payment processor, the firm must switch to cash or use alternative processors. With a natural experiment, we show that capital takers in our setting shift large transactions to other processors and that this hiding responds to the cost of alternatives. We also provide evidence that platforms can improve repayment with "sticky" features that make it costly to revenue hide.

Revenue-based financing also generates adverse selection challenges that digital platforms must overcome. Intuitively, businesses with hidden information find state-contingent repayment attractive which can cause the contract to unravel. Using a second natural experiment, we estimate that adverse selection is the size of roughly sixty percent of the gap in revenue between advance takers and observably similar non-takers. We also provide evidence that digital payment data improves screening and offer two suggestions for increasing repayment.

Finally, our decomposition of the "gap" in revenue between takers and non-takers provides evidence of revenue-based financing having a positive causal effect on small business performance. While our work highlights that the increased use of digital payments technology in recent decades has helped sustain these contracts, it remains unclear how these forces will affect equilibrium credit market outcomes in the years to come. For example, increased competition and open banking regulations might reduce the costs of hiding revenue. Further understanding how financial technology and asymmetric information interact in equilibrium remains an important avenue for future work.

References

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, **74** (1), 235–267.

Alfaro, L. and Kanczuk, F. (2005). Sovereign debt as a contingent claim: a quantitative approach. *Journal of International Economics*, **65** (2), 297–314.

Arrow, K. J. (1974). Limited knowledge and economic analysis. *American Economic Review*, **64** (1), 1–10.

Battaglia, M., Gulesci, S. and Madestam, A. (Forthcoming). Repayment flexibility and risk taking: Experimental evidence from credit contracts. *Review of Economic Studies*.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G. and Stein, J. C. (2005). Does function follow organizational form? evidence from the lending practices of large and small banks. *Journal of Financial Economics*, **76** (2), 237–269.

Berlin, M. and Mester, L. J. (1999). Deposits and relationship lending. *The Review of Financial Studies*, **12** (3), 579–607.

Bond, E. W. and Crocker, K. J. (1997). Hardball and the soft touch: the economics of optimal insurance contracts with costly state verification and endogenous monitoring costs. *Journal of Public Economics*, **63** (2), 239–264.

Botsch, M. and Vanasco, V. (2019). Learning by lending. *Journal of Financial Intermediation*, **37**, 1–14.

Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5–32.

Chen, B. S., Hanson, S. G. and Stein, J. C. (2017). The decline of big-bank lending to small business: Dynamic impacts on local credit and labor markets. *NBER Working Paper No. w23843*.

Clarke, R., Kuloszeswki, Z. and Wong, J. (Forthcoming). Small firm impacts of electricity reliability and planned outages.

Cordaro, F., Fafchamps, M., Mayer, C., Meki, M., Quinn, S. and Roll, K. (2022). Microequity and mutuality: Experimental evidence on credit with performance-contingent repayment. *NBER Working Paper No. w30411.*

De Mel, S., McKenzie, D. J. and Woodruff, C. (2019). Micro-equity for microenterprises. *CEPR Discussion Paper No. DP13698.*

Dresner, A., Murati, A., Pike, B. and Zell, J. (2022). Embedded finance: Who will lead the next payments revolution? `https://www.mckinsey.com/industries/financial-services/our-insights/embedded-finance-who-will-lead-the-next-payments-revolution`.

Einav, L. and Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, **25** (1), 115–138.

—, — and Cullen, M. R. (2010). Estimating welfare in insurance markets using variation in prices. *The Quarterly Journal of Economics*, **125** (3), 877–921.

Friedman, M. (1955). The role of government in education. *Economics and the Public Interest*, **2** (2), 85–107.

Ghosh, P., Vallee, B. and Zeng, Y. (2021). Fintech lending and cashless payments. *HBS Working Paper.*

Gopal, M. and Schnabl, P. (2022). The rise of finance companies and fintech lenders in small business lending. *The Review of Financial Studies*, **35** (11), 4859–4901.

Hendren, N. (2013). Private information and insurance rejections. *Econometrica*, **81** (5), 1713–1762.

Herbst, D. and Hendren, N. (2021). Opportunity unraveled: Private information and the missing markets for financing human capital. *NBER Working Paper No. w29214.*

Howell, S., Kuchler, T., Snitkof, D., Stroebel, J. and Wong, J. (Forthcoming). Lender automation and racial disparities in credit access. *Journal of Finance.*

Jagannathan, R., Matsa, D. A., Meier, I. and Tarhan, V. (2016). Why do firms use high discount rates? *Journal of Financial Economics*, **120** (3), 445–463.

Karlan, D. and Zinman, J. (2009). Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, **77** (6), 1993–2008.

Kozak, A. (2023). Shedding the load: Power shortages widen divides in south africa. https://kleinmanenergy.upenn.edu/news-insights/shedding-the-load-power-shortages-widen-divides-in-south-africa/.

Leland, H. E. and Pyle, D. H. (1977). Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance*, **32** (2), 371–387.

OECD (2022). South africa. In *Financing SMEs and Entrepreneurs 2022*.

Petersen, M. A. and Rajan, R. G. (1994). The benefits of lending relationships: Evidence from small business data. *The Journal of Finance*, **49** (1), 3–37.

Rishabh, K. and Schäublin, J. (2021). Payment fintechs and debt enforcement. *Center of Business and Economics, University of Basel, Working Paper*.

Rush, T. (2021). Revenue-based financing: The next step for private equity and early-stage investment. https://techcrunch.com/2021/01/06/revenue-based-financing-the-next-step-for-private-equity-and-early-stage-investment/.

Schneider, K., Stocks, C. and Dietrich, J. (2020). Small business lending and the great recession. *Consumer Financial Protection Bureau Office of Research Reports Series*.

Small Business Institute (2019). The number of formal micro, small & medium businesses in south africa. https://www.smallbusinessinstitute.co.za/wp-content/uploads/2019/01/SBIbaselineStudyAlertfinal.pdf.

SME South Africa (2018). An Assessment of South Africa's SME Landscape: Challenges, Opportunities, Risks & Next Steps' Report 2018/2019. https://smesouthafrica.co.za/sme-resources/an-assessment-of-south-africas-sme-landscape-challenges-opportunities-risks-next-steps-report-2018-2019/.

Stango, V. and Zinman, J. (2009). Exponential growth bias and household finance. *The Journal of Finance*, **64** (6), 2807–2849.

Teima, G., Istuk, I., Maldonado, L., Soriano, M. and Wilson, J. (2022). Fintech and sme finance: Expanding responsible access. *World Bank Group*.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, **21** (2), 265–293.

Trevino, G. A. (1997). A note on formulating and corroborating discount rates for small firms. *J. Legal Econ.*, **7**, 45.

United Nations (2023). South africa: Entrepeneurship strategy review. https://unctad.org/system/files/official-document/tcsdiaeinf2023d3_en.pdf.

Webb, D. C. (1992). Two-period financial contracts with private information and costly state verification. *The Quarterly Journal of Economics*, **107** (3), 1113–1123.

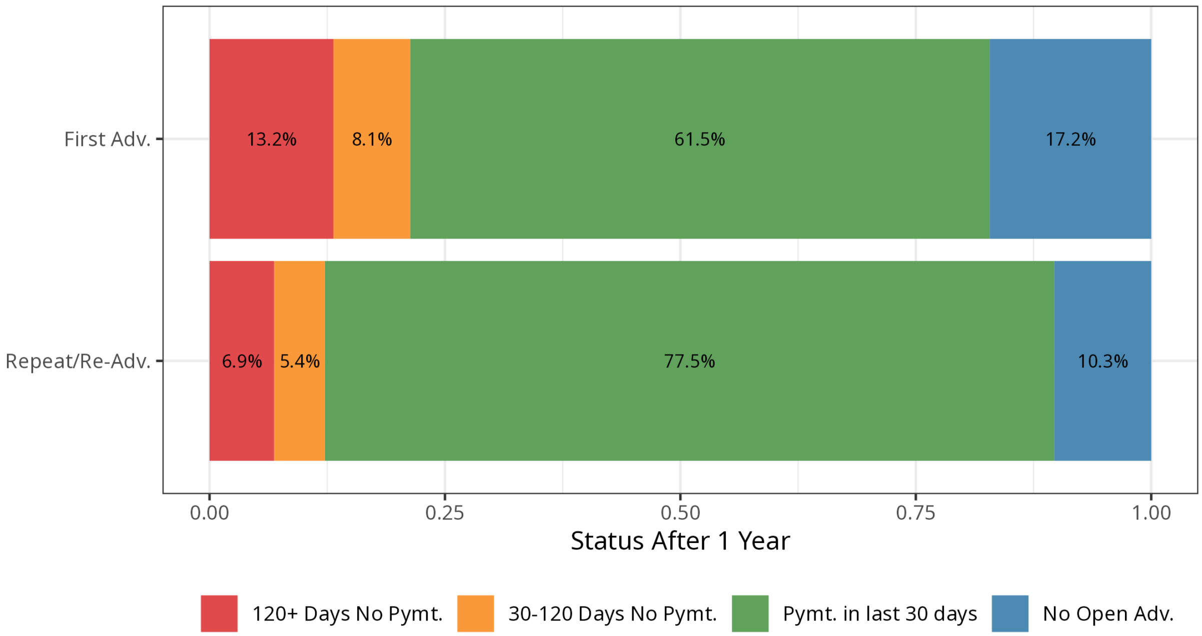
Winton, A. (1995). Costly state verification and multiple investors: the role of seniority. *The Review of Financial Studies*, **8** (1), 91–123.

Woodruff, C. (2018). Addressing constraints to small and growing businesses. *International Growth Centre, London*, **20**.

World Bank (2021). Enterprise surveys: South africa. www.enterprisesurveys.org.

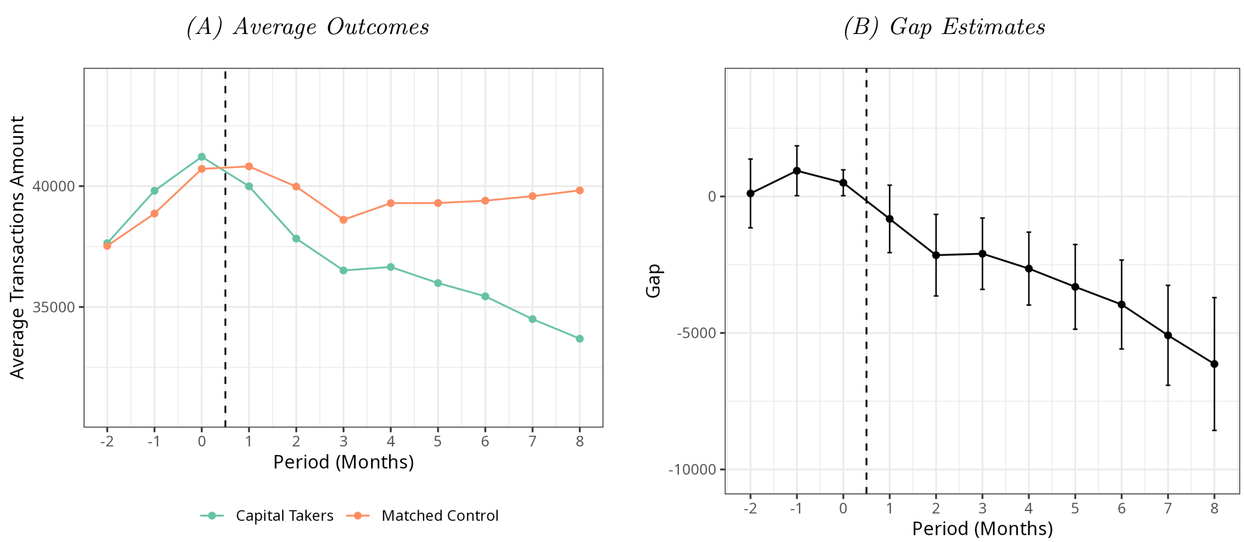
Figures

Figure 1: *Advance Taker One-Year Outcomes*



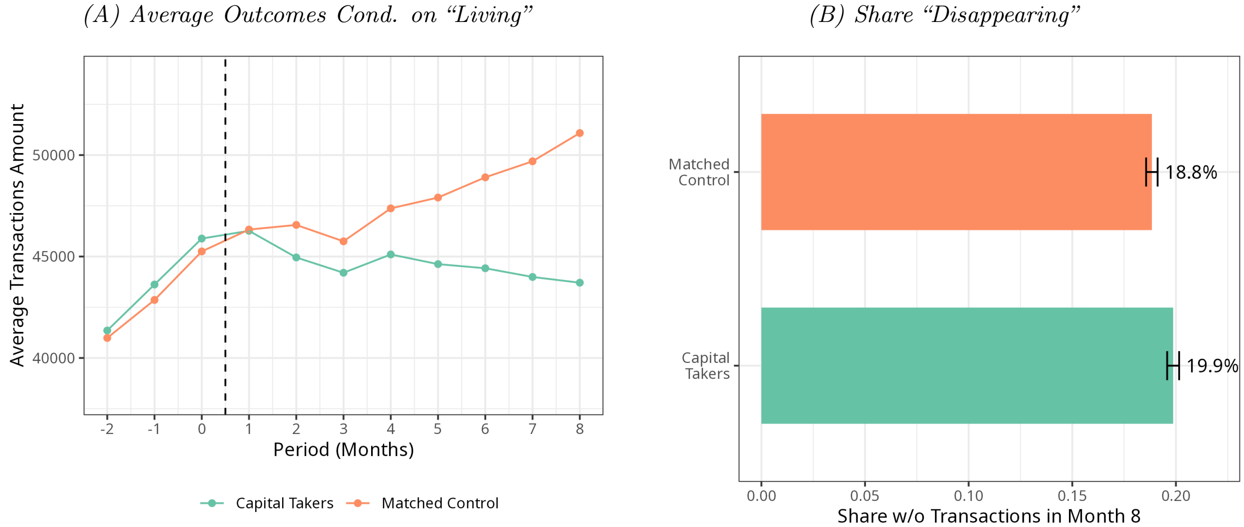
Note: Figure shows outcomes for first-time and repeat advance takers, respectively, one year after taking an advance. Outcomes shown are for *any* advance the business has one year later.

Figure 2: *Gap Between Capital Takers and Non-Takers*



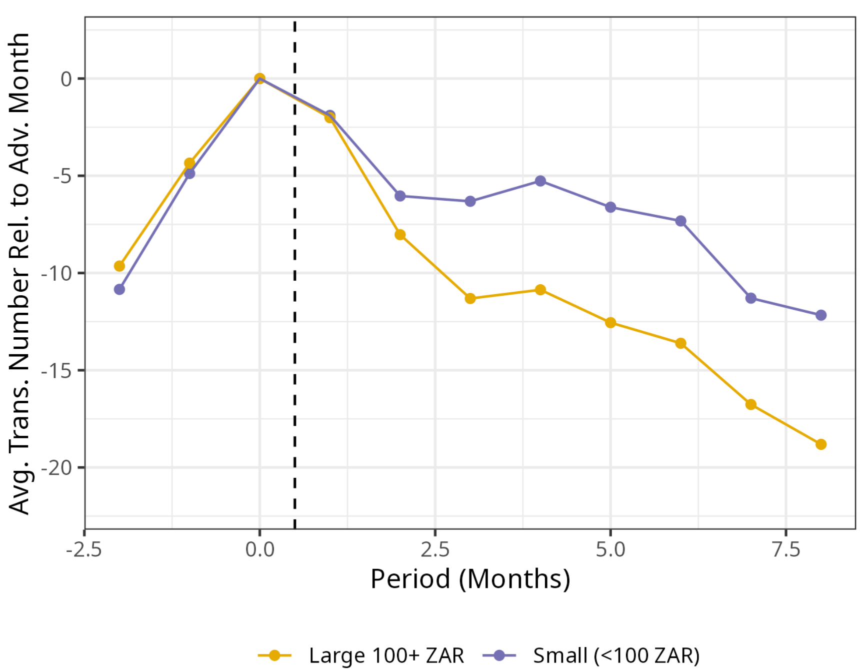
Note: Figure shows the average monthly transactions of capital takers and matched control businesses. Advances were taken in month 0. Each taker is matched to a control business in the same month and industry with the smallest Euclidean distance according to (normalized) time on platform and transaction amount in month 0. Panel A shows the average transactions amount capital takers and the matched control group. Panel B displays the difference between groups. Bars display the 95% confidence intervals with Abadie and Imbens (2006) adjusted standard errors.

Figure 3: *Gap Primarily Driven by "Intensive Margin"*



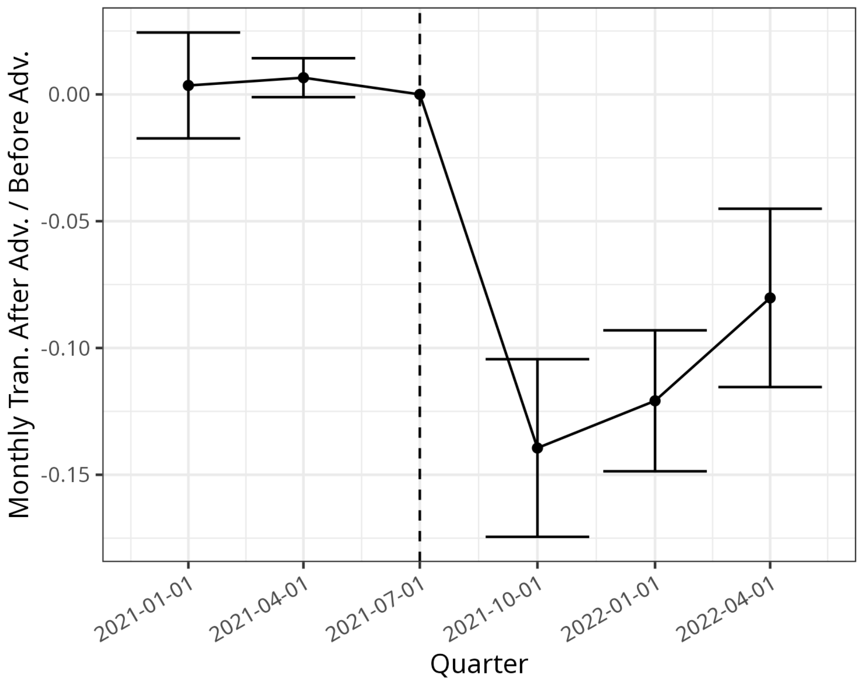
Note: Figure shows the intensive and extensive margin contributions to the gap shown in Figure 2. Panel A shows the average monthly transactions of capital takers and matched control businesses as in Figure 2, but selected only from those who are transact on the platform in the eighth month after the advance. Panel B shows the share of each set of business in Figure 3 that transact on the platform in the eighth month after the advance.

Figure 4: *Capital Taker Revenue by Transaction Size*



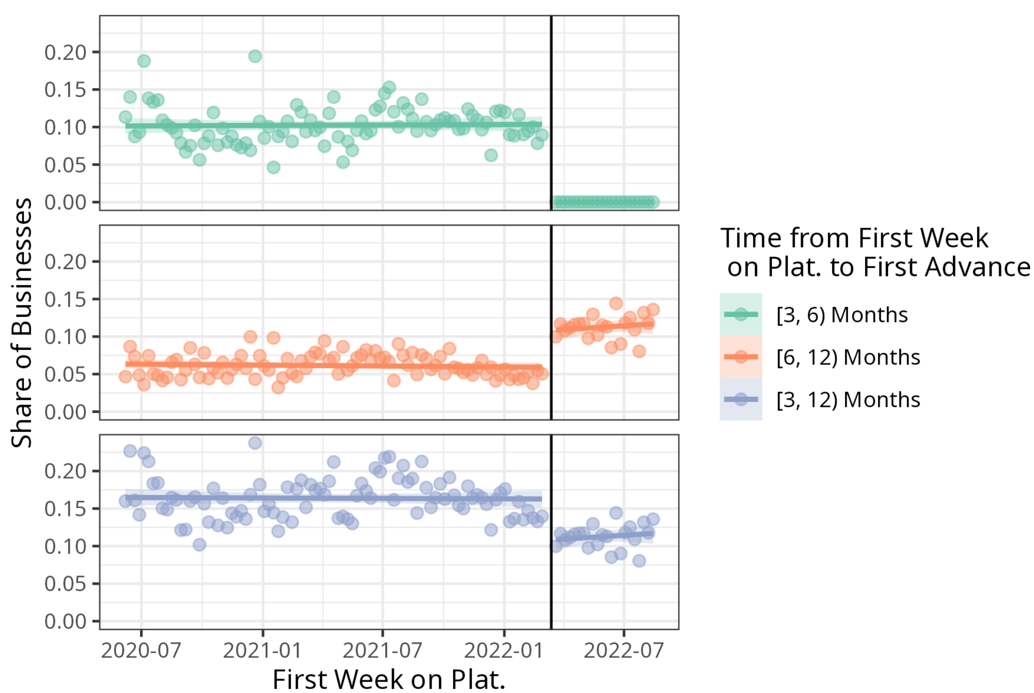
Note: Figure shows the average monthly number of transactions of capital takers by transaction size. Advances were taken in month 0. Both series are centered at 0. 100 ZAR is roughly $15 USD by purchasing power parity.

Figure 5: *Post-Capital Transactions When Rival Lowers Price*



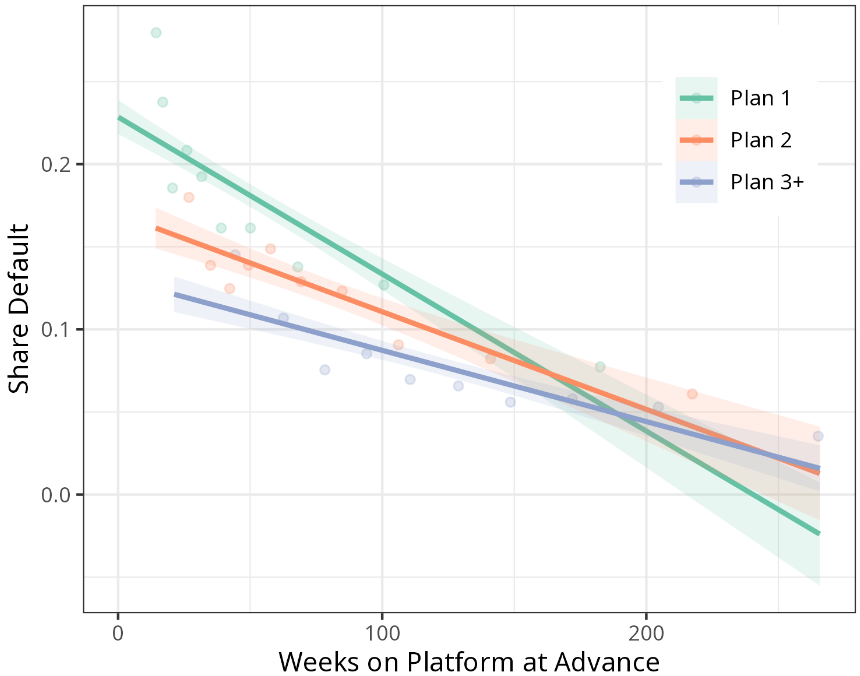
Note: Figure shows estimates of β_t from a difference-in-differences specification given in Equation 8. The quarter of the Competitor's price drop is July 2021, as shown in Figure C.2. Bars display 95% confidence intervals.

Figure 6: *Time to First Advance, Around Policy Change*



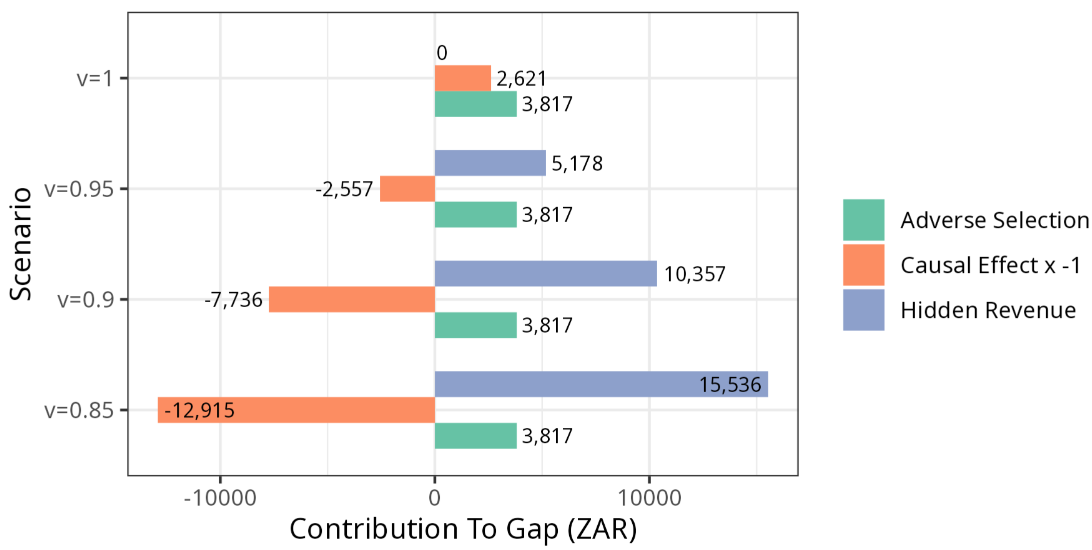
Note: Figure shows, by first week on the platform, the share of businesses who met the minimum transaction eligibility criteria in month three that took an advance within different time frames. From top to bottom, the panels show the share of businesses who took an advance in months 3-5, 6-11, and 3-11, respectively. The week of the policy change described in Section 6 is excluded.

Figure 7: *Share "Disappearing" by Advance Number and Time on Platform at Advance*



Note: Figure shows the share of businesses that default against the number of weeks the business had been on the platform, split by advance number. Default is whether the business (on the advance) has an open advance and no transactions 8 months after the start of the advance. Weeks on platform is calculated from the date of first transaction on the platform to the advance being given.

Figure 8: *Decomposition Scenarios*



Note: Figure shows the contributions of adverse selection, causal effects, and moral hazard to the overall Gap—as each term is defined in Equations 4–6—under various revenue hiding scenarios. The estimates are for the three-month post-advance revenue of businesses who took an advance between March 2021 and February 2022. Our methodology for constructing these estimates is described in Section 6. As defined in Section 2, v is the share of revenue the financier can observe. Each scenario applies a given v to all advance takers.

Tables

Table 1: *Summary of Primary Sample*

Panel A: 1st Advances					
Var	Mean	SD	p25	p50	p75
Prior Weeks on Platform	51.71	50	19.07	32.25	63.12
Sales Amount in Prior 3mo (ZAR)	110,396	227,342	26,599	53,782	115,437
Sales N in Prior 3mo	451.86	875.28	88	199	478
Principal Amt. (ZAR)	36,224	62,597	7,500	16,000	39,000
Charge Rate (%)	19.19	6.51	16	22	23
Factor Rate	1.28	0.04	1.27	1.3	1.3
Est. Repayment Period (Months)	7.62	2.16	6	8	9
1 yr. Amt. Paid / Princip.	1.06	1.14	0.97	1.26	1.3
Discounted (5%) Amt. Paid 1 yr. / Princip.	1.04	1.13	0.96	1.24	1.28
Discounted (15%) Amt. Paid 1 yr. / Princip.	1.02	1.09	0.92	1.22	1.25

Panel B: Repeat / Re-Advances					
Var	Mean	SD	p25	p50	p75
Prior Weeks on Platform	108.23	64.57	56.2	93.2	146.91
Sales Amount in Prior 3mo (ZAR)	152,117	222,162	44,663	85,481	171,839
Sales N in Prior 3mo	601.02	984.83	115	272	680
Principal Amt. (ZAR)	46,802	74,331	12,000	22,500	50,000
Charge Rate (%)	20.02	5.37	18	22	23
Factor Rate	1.35	0.14	1.3	1.3	1.39
Est. Repayment Period (Months)	8.43	2.12	8	8	9
1 yr. Amt. Paid / Princip.	1.22	0.82	1.22	1.3	1.37
Discounted (5%) Amt. Paid 1 yr. / Princip.	1.21	0.82	1.21	1.28	1.36
Discounted (15%) Amt. Paid 1 yr. / Princip.	1.18	0.81	1.18	1.25	1.33
Nth Advance	3.4	1.85	2	3	4

Note: Table presents summary statistics describing characteristics of first and repeat advances. The sample is cut to include advances made from June 2020 until June 2022 (so we can observe outcomes for at least 12 months as of June 2023). When discounting repayments, we assume an annual rate of return of x_a (5% or 15%) and that interest payments are compounded daily giving a daily discount rate of $x_d = (1 + x_a)^{\frac{1}{365}} - 1$. Then, daily repayments are discounted by $\frac{1}{(1+x_d)^t}$ where t is the number of days since the advance was opened. We use a daily discount rate because repayments are collected at the end of each day. Section 1 provides more details on how the advance works.

Table 2: *Predictors of Advance Performance*

	(1) Default	(2) Log Total Amt. 8 Months	(3) Default	(4) Log Total Amt. 8 Months	(5) Default	(6) Log Total Amt. 8 Months
Years on Platform	-0.045***	0.038**	-0.036***	0.033***	-0.026***	0.021***
	(0.0037)	(0.0076)	(0.0037)	(0.0057)	(0.0015)	(0.0023)
Log Amt. -3 Months	-0.021***	0.92***	-0.014**	0.86***	-0.013**	0.88***
	(0.0025)	(0.0098)	(0.0040)	(0.020)	(0.0028)	(0.0045)
Relative Sd.	0.065***	0.031	0.077***	0.035	0.071***	-0.028
	(0.0085)	(0.048)	(0.0087)	(0.046)	(0.0041)	(0.019)
First Plan					0.042***	-0.0087
					(0.0043)	(0.0098)
Sample	First Plans	First Plans, No Default	First Plans	First Plans, No Default	All Plans	All Plans, No Default
Demographic FE	No	No	Yes	Yes	Yes	Yes
Quarter X Year FE	No	No	Yes	Yes	Yes	Yes
Loan Controls	No	No	Yes	Yes	Yes	Yes
Observations	11903	9798	11903	9798	29157	25423
Adjusted R^2	0.024	0.67	0.035	0.68	0.045	0.72

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table shows regressions of various measures on advance performance. The unit of observation is an advance. The dependent variable "default" in columns (1), (3) and (5) is whether the advance taker has an open advance and no transactions in the 8th month post-advance. The dependent variable in columns (2), (4) and (6) is log of total transaction amounts within 8 months post advance, conditional on no default. Years on platform is calculated from the date of first transaction on the platform to the advance being given. Log amount 3 months prior is the log of total transactions three months before the advance was given. Relative standard deviation is the standard deviation of weekly transactions amounts, divided by the mean, in the three months before the advance was given. First plan is an indicator for whether the advance was the first advance taken by the business. Columns (3)–(5) include demographic fixed effects (industry segmentation, business type, citizenship, location classification, province), quarter by year fixed effects, and loan controls (principal, charge rate, factor rate). Standard errors are clustered at the province level.

Table 3: *Stickiness on Advance Performance*

	(1) Default	(2) Log Total Amt. 8 Months	(3) Default†	(4) Log Total Amt. 8 Months†
Manage Button	-0.013^{+}	0.054***		
	(0.0063)	(0.0045)		
Exported Sales	0.024	-0.0024	-0.020*	0.045
	(0.015)	(0.024)	(0.0067)	(0.024)
Years on Platform	-0.023***	0.014***	-0.035***	0.036***
	(0.0015)	(0.0024)	(0.0053)	(0.0045)
Log Amt. -3 Months	-0.016*	0.90***	-0.018***	0.91***
	(0.0049)	(0.011)	(0.0024)	(0.0048)
Relative Sd.	0.056**	-0.080**	0.17***	-0.14***
	(0.014)	(0.019)	(0.0069)	(0.013)
First Plan	0.031***	0.00031		
	(0.0051)	(0.022)		
Taker			0.0059*	-0.076***
			(0.0025)	(0.0085)
Taker × Exported Sales			0.052*	0.028
			(0.017)	(0.039)
Sample	Takers, All Plans	Takers, All Plans	All, First Plans	All, First Plans
Demographic FE	Yes	Yes	Yes	Yes
Quarter X Year FE	Yes	Yes	Yes	Yes
Loan Controls	Yes	Yes	No	No
Observations	10751	9527	22725	18179
Adjusted R^2	0.041	0.74	0.076	0.73

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table shows regressions of various measures on advance performance. The unit of observation is an advance. The dependent variable "default" in column (1) is whether the advance taker has an open advance and no transactions in the 8th month post-advance. The dependent variable in column (2) is log of total transaction amounts within 8 months post advance, conditional on no default. In columns (3) and (4) we include non-takers, using their outcomes starting one year after they joined the platform, as if they counterfactually took an advance. Default† in column (3) also includes businesses that have no open advance. Log Total Amt. 8 Months† in column (4) is conditional on no default according to Default†. The measure "manage button" is whether the business opened the manage tab to track staff, customers, and inventory. The measure "exported sales" is whether the business exported its sales history to a csv. All other independent variables are defined in Table 2. Standard errors are clustered at the province level.

Table 4: *Revenue Quarter Post-Advance Relative to Revenue Quarter Pre-Advance, Around Policy Change*

	Dependent Variable: Amt. 3 Months Post-Advance			
	(1)	(2)	(3)	(4)
Amt. -3 Months	0.80***	0.80***	0.94***	0.93***
	(0.05)	(0.05)	(0.02)	(0.02)
After Cutoff	7075.27^{+}	7072.23^{+}	7520.85^{+}	7361.25^{+}
	(3618.75)	(3573.00)	(4369.58)	(4395.69)
Sample	Full	Full	Near Cutoff	Near Cutoff
Month of Year FE	Yes	Yes	No	No
Industry FE	No	Yes	No	Yes
Observations	6975	6975	1043	1043
Adjusted R^2	0.633	0.633	0.685	0.684

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table shows results from regression 9. The unit of observation is a business. The dependent variable is their transaction amount in the quarter after taking a first advance. After cutoff is an indicator for whether the business joined the platform after the March 20, 2022 cutoff described in Section 5.1. Columns 1 and 2 include all businesses who joined the platform between the June 2020 and the start of May 2022 and took an advance within 12 months. The near cutoff sample in columns 3 and 4 additionally filters to businesses who joined the platform in the six weeks on either side of the cutoff. In columns 1 and 2 standard errors are clustered at the month of year level.

Table 5: *Decomposition Regressions*

	Dependent Variable: Amt. 3 Months			
	(1)	(2)	(3)	(4)
Amt. -3 Months	0.92***	0.87***	0.92***	0.87***
	(0.06)	(0.05)	(0.06)	(0.05)
Offered	−115.13		−293.61	
	(2196.25)		(2301.86)	
Taker		−6424.92*		−6439.42*
		(2484.48)		(2575.43)
Sample	Full	Offered Only	Full	Offered Only
Month of Year FE	Yes	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Observations	42346	30454	42346	30454
Adjusted R^2	0.682	0.672	0.682	0.672

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table shows results from regressions to decompose the short-term Gap, as described in Section 6. Columns 1 and 3 show results from equation 10 over all businesses who joined the platform between September 2020 and September 2022. Columns 2 and 4 show results from equation 11, limiting to businesses who joined the platform before the March 20, 2022 cutoff. Standard errors are clustered at the month of year level.

Appendix

A Proofs

Proof of Proposition 1

Proof. A firm will invest under a debt contract iff:

$$\mu - \gamma\sigma^2 - L \geq u(y)$$

A firm will invest under a revenue-based financing contract iff:

$$(1-\eta) \cdot \mu - (1-\eta)^2\gamma\sigma^2 \geq u(y)$$

As lenders are perfectly competitive, the η offered will be given by $\eta\mu = L \Rightarrow \eta = \frac{L}{\mu} < 1$. Thus, the above expression can be written as:

$$\mu - L - (1 - L/\mu)^2\gamma\sigma^2 \geq u(y)$$

As $\eta < 1$, this implies that the threshold y for taking revenue-based financing is lower. □

Proof of Proposition 2

Proof. If lenders cannot make non-positive profits, no revenue-based financing contracts will be offered. The lender's zero-profit condition is now:

$$v(c)\eta\mu = L \Rightarrow v(c)\eta = \frac{L}{\mu}$$

Let $\overline{c} = v^{-1}(L/\mu)$. Then,

$$\overline{c} < c \Rightarrow v^{-1}(L/\mu) < c \Rightarrow \frac{L}{\mu} < v(c)$$

as v is an increasing function. Consequently, for any $x \leq 1$:

$$x \cdot v(c)\mu - L < v(c)\mu - L < L - L = 0$$

which implies that the lender must make negative profits. $\square$

Proof of Proposition 3

Proof. As $L \geq u(y)$ this implies that all bad types will take revenue-based financing if offered. However, this increases η because the good types will need to cross-subsidize the bad types. If η becomes too high, it is possible that the good types will no longer take revenue-based financing. In this case, the lender will be unwilling to lend because they would incur negative profits of $-(1-p)L$.

The lender's zero-profit condition is:

$$(1-p)(-L) + p(\eta\mu - L) \Rightarrow \eta = \frac{L}{p\mu}$$

All the bad types take revenue-based financing, but the good types will only take revenue-based financing if:

$$(1-\eta)\mu - (1-\eta)^2\gamma\sigma^2 \geq u(y)$$

The LHS is expected utility from investment with revenue-based financing. Here it's clear that we need the $\mu > 2\gamma\sigma^2$ assumption, otherwise it's possible that borrowers would prefer a higher η (more expensive financing) in exchange for lower variance. If $\mu > 2\gamma\sigma^2$, then the derivative of the LHS with respect to η is:

$$-\mu + 2(1-\eta)\gamma\sigma^2 < -2\gamma\sigma^2 + 2(1-\eta)\gamma\sigma^2 \leq 0$$

as $\gamma \leq 1$. So the LHS is *decreasing* with respect to η. Consequently, as η increases, fewer good types take the contract and invest. Set $\overline{p}$ to be such that:

$$\left(1 - \frac{L}{\overline{p}\mu}\right)\mu - \left(1 - \frac{L}{\overline{p}\mu}\right)^2 \gamma\sigma^2 = u(y)$$

Then, for any $p < \overline{p}$, the LHS will be lower than $u(y)$ which means that no good types will take revenue-based financing, which makes revenue-based financing impossible as lenders must make negative profits. If $p \geq \overline{p}$, the good types will select in and make positive NPV investments. $\square$

Proof of Proposition 4

Consider two groups that are randomly assigned, but Group 1 receives offers and Group 2 does not. The difference in expected reported revenue Y_{obs} between the two groups is $\mathbb{E}[Y_{obs}|\text{Group 2}] - \mathbb{E}[Y_{obs}|\text{Group 1}]$. Notice that:

$$\mathbb{E}[Y_{obs}|\text{Group 2}] = \mathbb{E}[Y(0)|\text{Taker}] \cdot \mathbb{P}(\text{Taker}) + \mathbb{E}[Y(0)|\text{Non-Taker}] \cdot \mathbb{P}(\text{Non-Taker})$$
$$\mathbb{E}[Y_{obs}|\text{Group 1}] = \mathbb{E}[vY(1)|\text{Taker}] \cdot \mathbb{P}(\text{Taker}) + \mathbb{E}[Y(0)|\text{Non-Taker}] \cdot \mathbb{P}(\text{Non-Taker})$$

where we can remove the conditioning on group because the groups are randomly assigned. The first equation comes from the fact that "Taker" refers to "would take if offered." Thus,

$$\begin{aligned}\mathbb{E}[Y_{obs}|\text{Group 2}] - \mathbb{E}[Y_{obs}|\text{Group 1}] =& \mathbb{E}[Y(0)|\text{Taker}] \cdot \mathbb{P}(\text{Taker}) - \mathbb{E}[vY(1)|\text{Taker}] \cdot \mathbb{P}(\text{Taker}) \\ =& \mathbb{P}(\text{Taker}) \cdot (-CE + MH)\end{aligned}$$

where we have used the definitions of CE and MH from Equations 4–6.

B Alternative Approaches to Estimating the Gap

In this Appendix, we present two alternative methods for estimating the Gap between capital takers and non-takers, in addition to the matching approach detailed in Section 3.4.

Panel Regression Approach

We use a panel of business-by-quarter observations for every business that ever met the minimum advance eligibility requirements. To estimate the Gap we run a regression of the form:

$$Y_{i,t} = \delta_{c(i),t} + X_i + \beta_0 \text{Taker}_{i,t} + \beta_1 \text{Taker}_{i,t-1} + ... + \beta_8 \text{Taker}_{i,t-8}. \tag{B.1}$$

Here, $Y_{i,t}$ is the revenue of business i in month t; $\delta_{c(i),t}$ are cohort (first month on platform) by time fixed effects; and X_i are industry fixed effects. The indicators $\text{Taker}_{i,k}$ equal one when a business took a first advance in month k.

Intuitively, β_0, the coefficient on taker $\text{Taker}_{i,t}$, will be positive. This is because, while our sample includes only businesses that were eligible at *some* point, in any given month many will not be eligible. A decline in the coefficients β_1 through β_8 then captures a differential decline in revenue of advance takers, providing an estimate of the Gap. Figure C.6 shows that such a differential decline exists, consistent with the existence of the Gap. The difference between the highest and lowest coefficients is around 6,000, a magnitude roughly equal to our baseline result in Figure 2.

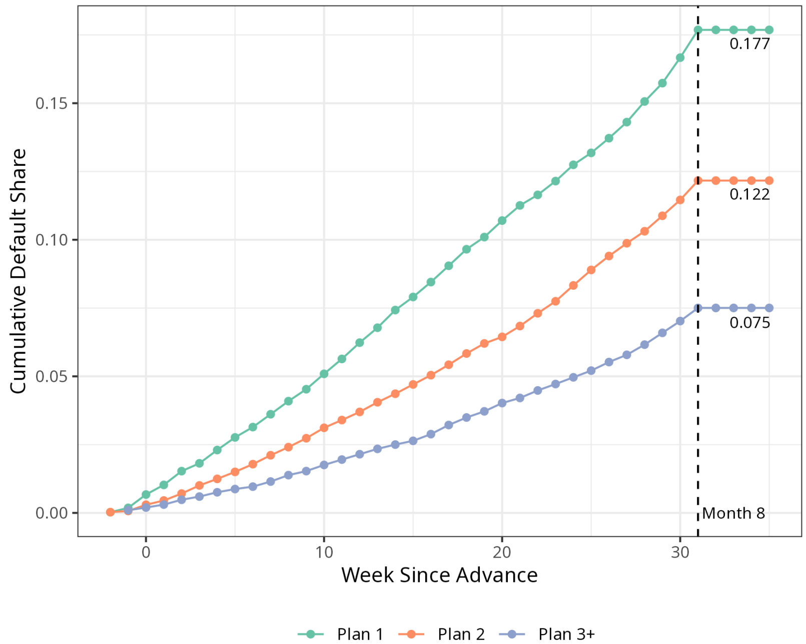
Machine Learning Approach

We use a panel of business-by-quarter observations for non-advance-taking businesses combined with observations for each advance taker in the quarter of their advance. We use each quarter-by-business observation to train random forests to predict the revenue of each taker and non-taker in the next eight months.[49] For each model we use revenue and transaction months in the prior three months, months since joining the platform, month, industry, and a taker indicator as predictors. We then use each model to make revenue predictions for the capital takers and, counterfactually, the capital takers if they did not take an advance. Figure C.7 shows the resulting estimates. The Gap between the in-sample (solid green line) and counterfactual (solid orange line) predictions is around 4,000 ZAR, roughly two-thirds of the magnitude of our baseline result in Figure 2.

[49]The algorithm allows us to find non-linear relationships between variables, without overfitting, by aggregating mean predictions from a number of regression trees generated over sample subsets of both observations and input variables. See Breiman (2001).

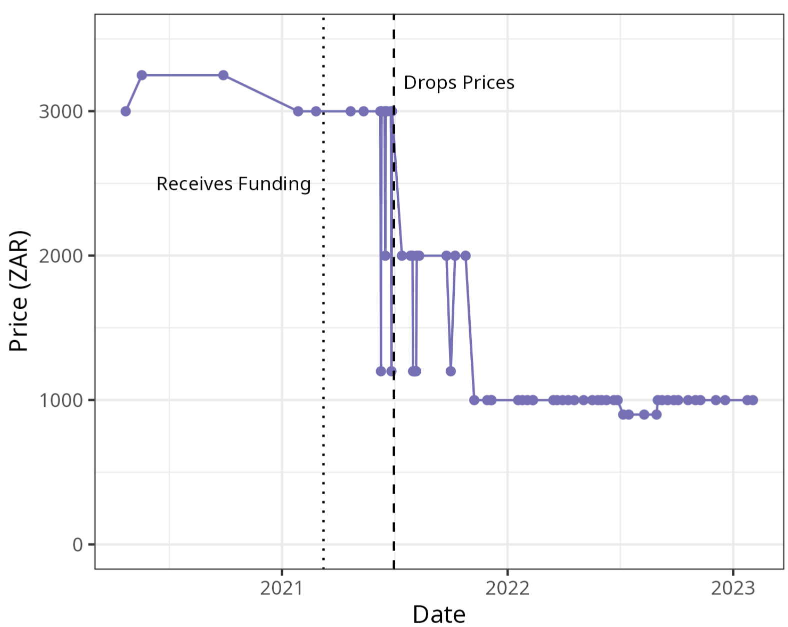
C Additional Figures

Figure C.1: *Hazard Plot for 8 Month Default*



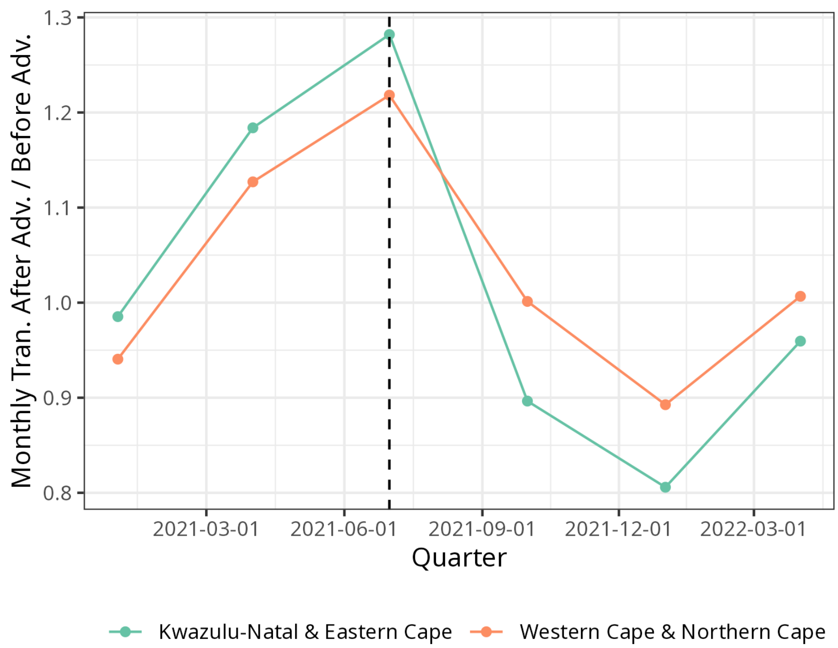
Note: Figure shows a hazard plot of default. Default is whether the business (on the advance) has an open advance and no transactions 8 months after the start of the advance. The cumulative default share is the fraction of businesses that had last transacted more than x weeks ago.

Figure C.2: *Rival Pricing Over Time*



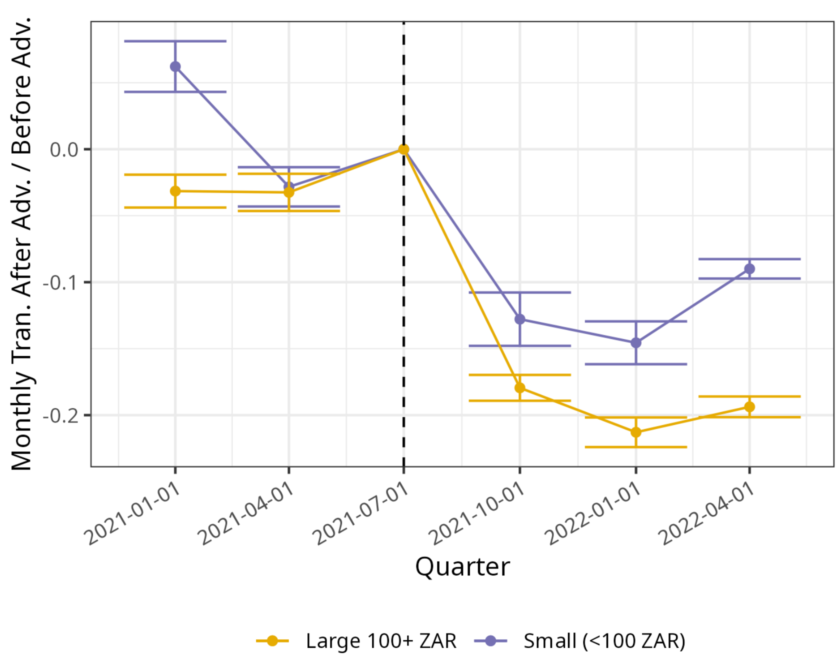
Note: Figure shows the up-front price of the Competitor's flagship product over time. Observations are archived pages from the Internet Archive and the Competitor's Facebook posts.

Figure C.3: *Post-Capital Transactions When Rival Lowers Price: Raw Averages*



Note: Figure shows averages of Y_{it} from Equation 8 by province group.

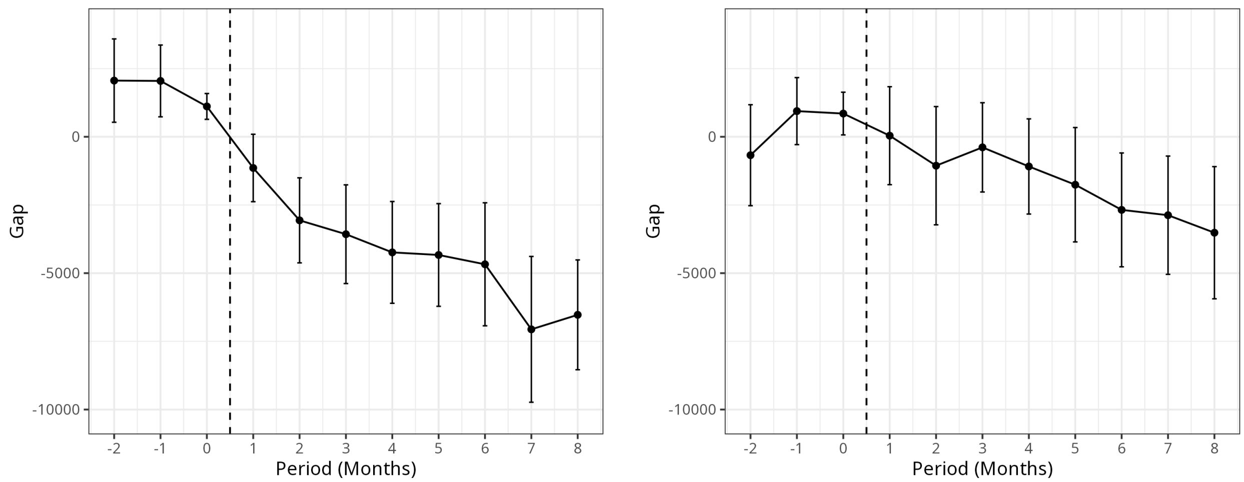
Figure C.4: *Post-Capital Transactions When Rival Lowers Price: By Transaction Size*



Note: Figure shows estimates of β_t from Equation 8 separately for large and small transactions. Bars display 95% confidence intervals. The figure includes only businesses who had more than five transactions of each size in the quarter before taking the advance. The top two percent of outcomes have been winsorized.

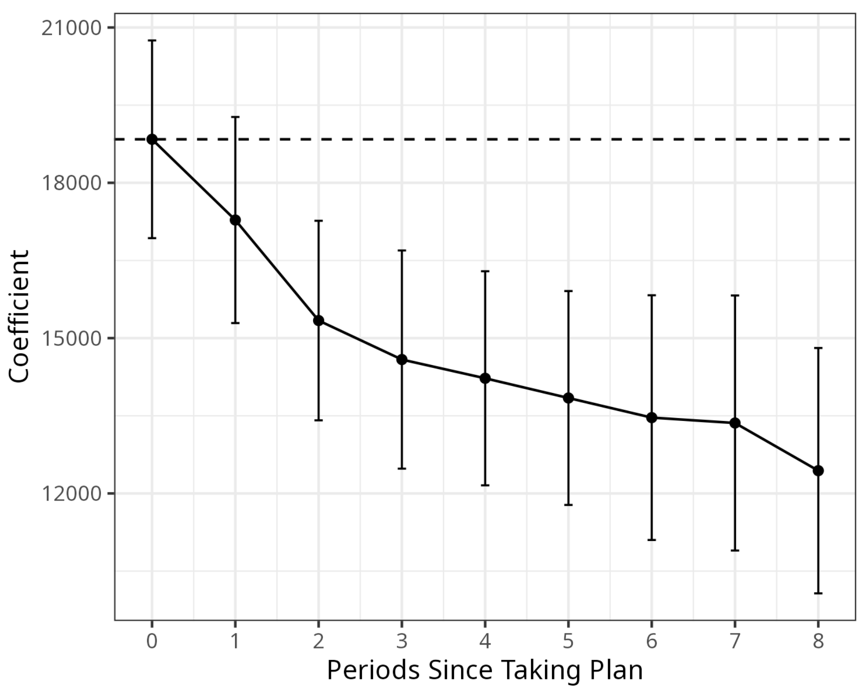
Figure C.5: *Gap Between Capital Takers and Non-Takers by Time on Platform at Advance*

(A) 6 or Fewer Months on Platform *(B) More than 6 Months on Platform*



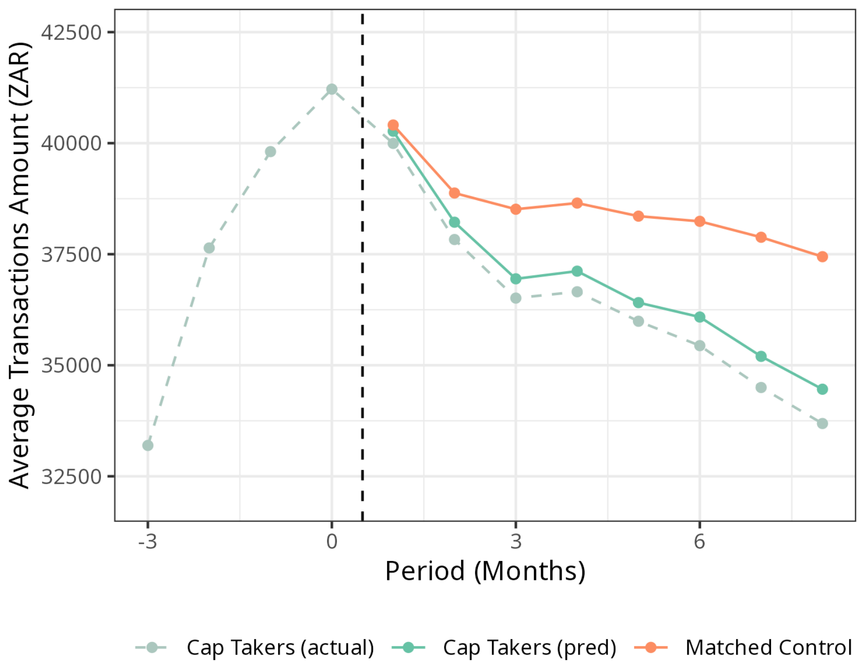
Note: Figure shows average monthly transactions of capital takers and matched control businesses split by time-on-platform. Panel (A) only includes businesses that had 6 or fewer months on the platform, panel (B) only includes businesses that had more than 6 months on the platform. The matching was done using the same methodology as in Figure 2. Advances were taken in month 0. Bars display 95% confidence intervals of the means.

Figure C.6: *Gap Between Capital Takers and Non-Takers - Panel Regression*



Note: Figure shows estimates of β_0 through β_8 from regression Equation B.1.

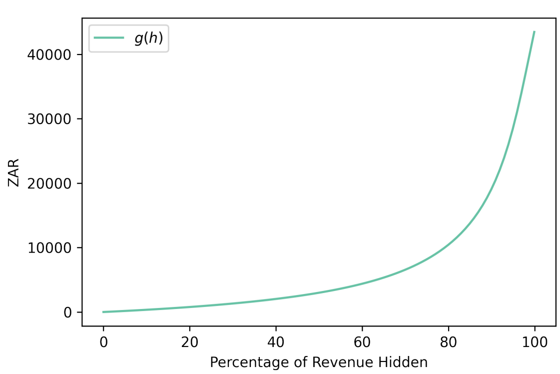
Figure C.7: *Gap Between Capital Takers and Non-Takers - Random Forest*



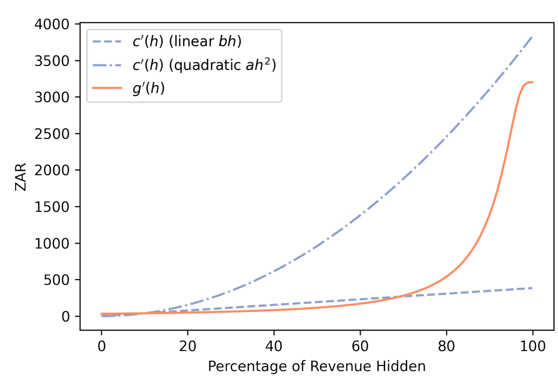
Note: Figure shows the average actual and predicted monthly transactions of first time capital takers. Advances were taken in month 0. The dashed green line shows the average monthly transactions amount. The solid green line shows the average in-sample predicted amount using a random forest model. The solid orange line shows the average of the random forest's counterfactual predictions if takers were instead not takers.

Figure C.8: *Simulation for Valuing the Processor*

(A) Gain from Hiding



(B) Marginal Cost vs. Marginal Gain from Hiding



Note: Figure shows the gains from hiding, and marginal cost and value curves for the average advance. We assume that daily revenue is constant, given by $103,570/90$ as in Figure 8. We assume an annual discount rate of $r = 30\%$, which implies a daily discount rate of $x_d = (1.3)^{\frac{1}{365}} - 1$. Then, daily repayments are discounted by $\frac{1}{(1+x_d)^t}$ where t is the number of days since the advance was opened. We use a daily discount rate because repayments are collected at the end of each day. We finally assume a factor rate of 1.3, charge rate of 20%, and principal of $36,224$ (Table 1). Panel (A) plots $g(h)$, the difference in the NPV of repayments between when the business hides 0% of transactions and h% of revenue. Panel (B) plots $g'(h)$ and two marginal cost curves. Linear of the form bh and quadratic of the form ah^2, where a and b are determined by the fact that both curves start at $(0, 0)$ and pass through $(10, g'(10))$.