Financial Conduct Authority

69

Occasional Paper

July 2025

Distress deferred? The impact of buy-now-pay-later credit on consumer indebtedness and arrears

Edwin Chan, Andrea de Mauro, Jonathan Shaw, Sam Hainsworth, Mark Mitchell, Helena Norrgren, Christopher Firth, & Jesal D. Sheth

FCA occasional papers in financial regulation

The FCA occasional papers
The FCA is committed to encouraging debate on all aspects of financial regulation and to creating rigorous evidence to support its decision-making. To facilitate this, we publish a series of Occasional Papers, extending across economics and other disciplines.

The main factor in accepting papers is that they should make substantial contributions to knowledge and understanding of financial regulation. If you want to contribute to this series or comment on these papers, please contact David Stallibrass David.Stallibrass@fca.org.uk.

Disclaimer
Occasional Papers contribute to the work of the FCA by providing rigorous research results and stimulating debate. While they may not necessarily represent the position of the FCA, they are one source of evidence that the FCA may use while discharging its functions and to inform its views. The FCA endeavours to ensure that research outputs are correct, through checks including independent referee reports, but the nature of such research and choice of research methods is a matter for the authors using their expert judgement. To the extent that Occasional Papers contain any errors or omissions, they should be attributed to the individual authors, rather than to the FCA.

Authors
Edwin Chan (FCA), Andrea de Mauro (FCA), Jonathan Shaw (FCA), Sam Hainsworth (FCA), Mark Mitchell (FCA), Helena Norrgren (FCA), Christopher Firth (during his time at the FCA), & Jesal D. Sheth (FCA).

Acknowledgements
We would like to thank Witold Więcek, Aureo de Paula (University College London), and Daniel Gibbons (FCA) for helpful comments and econometrics guidance. John Gathergood (University of Nottingham) kindly provided academic advice, reviewed the paper, and provided insightful feedback. We are also grateful to Kieran Keohane, Robert Rosenberg, Terry Denness, Jonathan O'Bannon, James Reah, Haris Irshad, David Stallibrass, and Kate Collyer (all FCA) for comments and guidance. Finally, we would like to thank the DPC firms who voluntarily provided detailed transaction level data that enabled this analysis.

All our publications are available to download from www.fca.org.uk. If you would like to receive this paper in an alternative format, please call 020 7066 9644 or email publications_graphics@fca.org.uk or write to Editorial and Digital Department, Financial Conduct Authority, 12 Endeavour Square, London E20 1JN.

Contents

Executive Summary

Deferred Payment Credit (DPC), also known as unregulated Buy Now Pay Later, has grown rapidly in recent years and concerns have been raised about potential consumer harm associated with its use. We use detailed transaction-level data and credit assignment rules provided by the four largest UK DPC providers at the time to show descriptive statistics on DPC users and estimate the causal impact of DPC credit on some consumer outcomes.

These outcomes represent a subset of the potential harms from DPC. The full set of harms that the FCA is seeking to address are set out in the accompanying Consultation Paper and Cost Benefit Analysis. We have not assessed all these harms in this research.

This analysis finds DPC users are, on average, younger, less creditworthy, have higher levels of unsecured debt, and have higher levels of financial difficulty compared to the UK population. They are also almost twice as likely to be in serious financial distress than the wider UK population.

We find some evidence that additional DPC borrowing modestly increases the likelihood of falling into arrears on DPC itself. However, we do not find consistent evidence that DPC borrowing *causes* medium-term indebtedness on other credit products, nor evidence that it leads to higher rates of arrears (missed payments) on other credit products or financial distress.

These findings demonstrate that users of DPC are more vulnerable than the wider UK population. These consumers may require additional protections and support given DPC is currently unregulated. However, the research also indicates that DPC borrowing may not be the central cause of widespread financial challenges.

1. Introduction

Buy Now Pay Later (BNPL) has grown significantly in recent years, achieving substantial penetration in North America, Europe, parts of Asia (particularly China) and Australasia (Cornelli et al, 2023).[1] In the UK, unregulated BNPL – what the FCA refers to as Deferred Payment Credit (DPC) – has grown from 0.3 million active customers in 2017 to nearly 9 million customers in 2023[2].

There have been concerns about potential harms associated with this type of lending by consumer groups such as Which?[3], Money Saving Expert[4], and Citizens Advice[5] and in the FCA's Woolard Review[6] published in 2021. Some potential harms from DPC identified by consumer groups as well as those raised in the Woolard Review include:

1. Information asymmetries – consumers misunderstand the product as being a form of payment rather than a form of credit or do not understand the terms of their DPC agreement.
2. Behavioural distortions – products are sold in a way that exploits consumers behavioural biases, such as DPC being set as the default payment method at checkouts.
3. Consumer protections – consumers are unaware of the lack of protections with this form of credit.
4. Unnecessary late fees – consumers can face unexpected late fees that are disproportionate to the borrowing or collection activities.
5. Additional spending – consumers can spend more than they had intended.
6. Misaligned incentives – firms may not be incentivised to assess affordability in creditworthiness assessments and provide appropriate treatment of customers in arrears.
7. Financial distress – DPC increases overall levels of indebtedness and financial distress.
8. Unaffordable borrowing – DPC is repaid using other interest-bearing forms of credit, such as credit cards, overdrafts, high-cost credit, and informal sources of lending.

This paper aims to investigate the last two potential channels of harm. The accompanying CBA, annexed to Consultation Paper 25/23 (CP25/23),[7] discusses the full set of harms in detail.

---

[1] Cornelli, G., Gambacorta, L., & Pancotto, L. (2023). Buy now, pay later: a cross-country analysis. BIS Quarterly Review, 61-75.

[2] Based on FCA calculations using data received from DPC firms.

[3] Under Pressure: Who uses Buy Now, Pay Later? (2023). https://www.which.co.uk/policy-and-insight/article/under-pressure-who-uses-buy-now-pay-later-aiFRV3f8zAI8

[4] Buy now, pay later firms to be regulated – and all shoppers will face affordability checks. (2021). https://www.moneysavingexpert.com/news/2021/02/bnpl-industry-to-be-regulated/

[5] Martin Lewis: 'Dear Government, don't be yankers' – MSE, Citizens Advice and Which? renew calls for buy now, pay later regulation. (2023). https://www.moneysavingexpert.com/news/2023/11/buy-now-pay-later-protection-needed-moneysavingexpert-government/

[6] The Woolard Review – A review of change and innovation in the unsecured credit market. (2021). https://www.fca.org.uk/publication/corporate/woolard-review-report.pdf

[7] Financial Conduct Authority (2025). Deferred Payment Credit (unregulated Buy Now Pay Later): Proposed approach to regulation (CP25/23). https://www.fca.org.uk/publications/consultation-papers/cp25-23-deferred-payment-credit-proposed-approach-regulation

As deHaan et al. (2022)[8] point out, the effects of BNPL on consumers' financial health are theoretically ambiguous. On the one hand, BNPL is a new line of credit that is interest-free, so may provide an affordable way for consumers to spread the cost of important purchases that would otherwise be out of reach. On the other hand, easy access to cheap credit may encourage new spending, potentially leading to declines in consumer financial health. Without the same regulatory protections as other forms of credit, BNPL providers may also extend credit to riskier borrowers than other lenders, resulting in excessive borrowing and greater risk of financial distress.

In this paper, we exploit detailed transaction-level data from 1 January 2017 to 31 March 2023 and precise documentation of credit assignment rules provided directly by the four largest DPC providers in the UK to estimate the causal impact of DPC credit on consumer financial outcomes, addressing two potential channels of harm. This combination of granular administrative data and institutional detail is rare and allows for a more credible identification strategy than typically available for the UK market. Our identification strategy relies on exploiting threshold-based cut-offs that DPC lenders use to set consumers' credit limits for different credit score values. We locate 580 such discontinuities and estimate the impact of additional DPC borrowing at each threshold using a fuzzy regression discontinuity design. We then aggregate these 580 threshold-specific parameters using Bayesian hierarchical methods to derive market-wide estimates and pin down the degree of heterogeneity across borrower risk and income groups. By linking the DPC transaction data to individual credit files, we trace the impact of DPC borrowing on the consumers' full credit portfolio, including debt levels, arrears and default.

At an aggregate level, we find no clear evidence that additional DPC borrowing causes an increase in medium-term (interest-bearing) indebtedness, higher rates of arrears or financial distress on other credit products. We find some evidence that DPC use increases late DPC repayments 7-12 months after usage. DPC usage increases short-term indebtedness. Estimates by borrower subgroup are less precisely estimated. We find that, on average, the estimated effects of DPC usage for most income and risk score groups are close to zero. However, because there is statistical uncertainty in these estimates, we cannot rule out the possibility that there may be non-negligeable effects for some groups.

One important caveat is that few DPC providers charged late fees during the period covered in this analysis so we have not been able to assess their implications for consumer outcomes; it will be important to monitor this given that late and missed repayment fees are now a more common feature and DPC firms' business models are evolving.[9]

Further, we have been able to assess implications of DPC by credit risk score and income groups but have not been able to assess implications for consumer groups that are otherwise vulnerable, such as those with mental health problems. Moreover, our DPC lending data does not cover 2022-2023, a period when DPC usage and the potential for harms may have increased due to cost-of-living pressures. Our analysis includes the four largest providers of DPC in the UK (for the time period covered) which make up ~90% of the market but we omit smaller providers, results for whom might not generalise.

Our paper contributes to a small but growing literature on the use of DPC and its implications for consumer outcomes. DiMaggio et al. (2022)[10], and Consumer Financial

[8] deHaan, E., Jungbae, K., Lourie, B., Zhu, C. (2023). Buy Now Pay (Pain?) Later. SSRN Working Paper.
[9] See for example the press release by Klarna: https://www.klarna.com/international/press/klarna-launches-customer-first-late-payments-programme-and-financial-support-package-for-those-who-fall-behind/
[10] Di Maggio, M., Williams, E., Katz, J. (2022). Buy now, pay later credit: user characteristics and effects on spending patterns. NBER Working Paper 30508. http://www.nber.org/papers/w30508

Protection Bureau (2023)[11] provide an overview of BNPL user characteristics and spending patterns in the US BNPL market. Bian et al. (2023)[12] provide insights on how the Chinese BNPL market has expanded since Covid-19. For the UK, Guttman-Kenney et al. (2023) find that 19.5% of active credit card users had a transaction with a BNPL firm on their credit cards in 2021, with the proportion even higher for younger consumers and those living in deprived regions with limited repayment capability. More recently, Ashby et al. (2025) add behavioural insights, showing how instalment pricing presentation increases spending by lowering perceived expensiveness, which may affect credit risk and financial health.

In terms of causal analysis, a common identification strategy involves exploiting the rollout of BNPL by merchants to provide exogenous variation in exposure to BNPL based on consumers' prior shopping habits. Di Maggio et al. (2022) use this approach to estimate the impact of BNPL access on consumer spending in the US, finding that BNPL access increases both total spending and the proportion of spending that goes on retail goods. They argue that these results are more consistent with a "liquidity flypaper effect" – where additional liquidity through BNPL causes consumers to increase up-front consumption of similar goods – than a standard lifecycle model with liquidity constraints. Using a similar strategy, deHaan et al. (2022)[13] find that access to BNPL in the US causes significant increases in overdraft charges, credit card interest and credit card late fees.

We have some concerns about the validity of this analytical identification strategy. In particular, the decision by retailers to adopt BNPL may not be exogenous to consumer spending decisions. For example, BNPL may be more likely to be adopted by merchants who foresee tough trading conditions ahead – or by more tech-savvy merchants as part of a broader strategy to improve their online offering. Or perhaps merchants that have decided to put up their prices attempt to soften the blow by offering BNPL as a payment method. All of these channels would threaten the validity of the exclusion restriction. It also seems likely that retailers who adopt BNPL earlier will tend to be those who foresee significant benefits to doing so, meaning that treatment effect estimates may not be representative across the ultimate set of retailers who adopt BNPL. Instead of exploiting the adoption of DPC by merchants, we rely on precisely specified discontinuities in the lending rules DPC providers employ, helping to address these identification issues. This is possible because our data is directly from DPC providers and includes detailed information about credit assignment rules.

Two recent papers report results from randomised experiments, which do not suffer from these limitations. Bian et al. (2023) exploits a randomised experiment at an e-wallet provider in China that granted early access to BNPL as a payment option for a subset of eligible customers. The authors find that access to this BNPL credit boosted consumer spending by around 5%. Likewise, Berg et al., (2023) analyse a randomised experiment at an online furniture retailer in Germany, finding that offering consumers BNPL increased sales by 20%, with effects at the extensive margin accounting for 60-70% of the total. These papers advance the literature by providing clean estimates of the impact of BNPL on spending, but both may have issues with external validity. The experiment in the former lasted only two months and granted access to BNPL to new customers so it is possible that much of the estimated effect reflects substitution across time. It is also for an eligible population at a single e-wallet provider in China from more than six years ago, very early in the expansion of BNPL, so results may not translate straightforwardly

[11] Consumer Financial Protection Bureau (2023). Consumer Use of Buy Now, Pay Later: Insights from the CFPB Making Ends Meet Survey.
[12] Bian, W., Cong, L.W., Ji, Y. (2023). The rise of e-wallets and buy-now-pay-later: payment competition, credit expansion, and consumer behavior. NBER Working Paper 31202. http://www.nber.org/papers/w31202
[13] deHaan, E., Jungbae, K., Lourie, B., Zhu, C. (2023). Buy Now Pay (Pain?) Later. SSRN Working Paper.

to the current UK or US context. In the latter study, the BNPL credit product on offer seems to have been more like an in-house 14-day invoice, and the merchant is a single online furniture retailer, so it is not clear how applicable findings are to more typical BNPL offerings. The impact estimates also seem very large, particularly in comparison to results in the same paper for payment via PayPal (which presumably itself offered BNPL credit to at least some customers). In contrast, our results are based on consistent data over the period 2017-2023 across the four main specialist DPC providers in the UK.

Our work is also related to the broader literature on the impact of credit access on consumer outcomes. For example, there is a considerable literature on whether payday loans cause bad credit events, including Gathergood et al. (2019)[14] for the UK and Melzer (2011)[15], Bhutta (2014)[16], and Skiba and Tobacman (2019)[17] for the US. These studies find fairly consistent evidence of negative impacts on consumer credit outcomes such as increased likelihood of arrears and default. Other work has focused on the impact of credit on spending. For example, Gross and Souleles (2002)[18] find a spending response to increased credit limits, even for those with low utilisation rates. Likewise, using a regression discontinuity strategy similar to ours, Agarwal et al. (2018) show that increasing the credit limit increases borrowing, particularly among less creditworthy borrowers. Our findings, which differ from these in important respects, demonstrate that the characteristics of the credit on offer are potentially an important determinant of its impact on consumer outcomes.

The rest of this paper is organised as follows. Section 2 sets out the UK policy context, including the current regulatory environment for DPC. Section 3 describes the data we use and provides an overview of the UK DPC market, including its size, users, and composition. Section 4 details the Methodology, while Section 5 sets out our results. Section 6 concludes.

2. Policy context

Buy Now Pay Later (BNPL) is a broad term used in the UK and internationally as that covers a variety of credit agreements, usually consisting of the deferral of some or all of the cost of a purchase and subsequent repayment either at once or over a series of instalments. Some of these agreements fall inside the FCA's remit (so-called 'regulated BNPL'), while other types remain unregulated, as they are able to rely on a legislative exemption. Article 60F(2) of the Financial Services and Markets Act 2000 (Regulated Activities) Order 2001 exempts certain interest-free agreements repayable in under 12 months and in 12 or fewer instalments from regulation. It is this form of currently unregulated credit agreements which were considered in the Woolard review and have been the subject of government consultation in 2021 and 2023, and which are the subject of this paper. We will refer to this type of credit as 'Deferred Payment Credit (DPC)' throughout this paper. These agreements are commonly offered at the online checkout of retailers by a third-party credit provider.

[14] Gathergood, J., Guttman-Kenney, B., & Hunt, S. (2019). How do payday loans affect borrowers? Evidence from the UK market. The Review of Financial Studies, 32(2), 496-523.
[15] Melzer, B. T. (2011). The real costs of credit access: Evidence from the payday lending market. The Quarterly Journal of Economics, 126(1), 517-555.
[16] Bhutta, N. (2014). Payday loans and consumer financial health. Journal of Banking & Finance, 47, 230-242.
[17] Skiba, P. M., & Tobacman, J. (2019). Do payday loans cause bankruptcy?. The Journal of Law and Economics, 62(3), 485-519.
[18] Gross, D. B., & Souleles, N. S. (2002). Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data. The Quarterly Journal of Economics, 117(1), 149-185.

These DPC credit agreements being unregulated means that firms offering DPC do not need to be authorised and regulated by the FCA, nor are they required to comply with most of the requirements of the Consumer Credit Act 1974 (CCA) or Consumer Credit Sourcebook (CONC). This has a few important implications worth highlighting.

There is no requirement for DPC lenders to carry out creditworthiness and affordability assessments to ensure that consumers have the means to repay the debt, which may increase the risk of unaffordable DPC debt. Further, until very recently, DPC firms did not report to credit reference agencies (credit bureaus) and, even now, there is only partial reporting. This means that other lenders have an incomplete view of consumers' outstanding debt when making their own lending decisions – and, as a result, consumers may be granted too much non-DPC debt relative to what would have been approved had lenders been able to see outstanding DPC debt. Additionally, there is a range of consumer protections contained in the CCA, CONC and the Consumer Duty that consumers do not have when they use DPC.

The FCA has already done significant work to gain a better understanding of the harms that exist in the DPC market. In 2021, the FCA published the Woolard Review[19], which examined innovations in the unsecured credit market and reviewed whether more needed to be done to ensure a healthy, sustainable market. The Woolard Review found that the recent rapid growth in DPC usage posed potential harms to consumers. It recommended DPC should be brought within the FCA's remit to ensure that consumers are protected and that the market and product develop in a sustainable way.

Following the publication of the Woolard Review, the Government announced its intention to regulate exempt Buy Now Pay Later products. The initial consultation, published in October 2021[20], sought views on the scope of BNPL regulation, focusing primarily on what activities should be regulated and the regulatory controls that should be applied.

The Government followed up with a second consultation paper in February 2023 with proposed draft legislation to bring DPC into FCA regulation.[21] The consultation proposed exempting DPC lending extended by merchants to customers for the purchase of their own products from the scope of regulation. Subsequently, the Government laid legislation in July2025, formally extending regulatory oversight to certain DPC activities.[22]

In this regulatory context, the FCA has undertaken two important pieces of analysis to understand the DPC market and the impacts of DPC usage on consumer outcomes. The first is our 2023 research note on DPC, setting out findings from the 2022 Financial Lives Survey. This found that adults with characteristics of vulnerability were more likely to report using DPC and to report using it frequently. They were also likely to report higher use of other credit products and signs of falling into difficulty with debt.[23,24] Future waves of the Financial Lives Survey (FLS) will also capture how individuals use DPC. The second is the analysis reported in this paper. Our 2023 research note focused on correlations and descriptive analysis, the aim in this paper is to provide causal evidence on the

[19] The Woolard Review - A review of change and innovation in the unsecured credit market. (2021). https://www.fca.org.uk/publication/corporate/woolard-review-report.pdf
[20] Regulation of Buy-Now Pay-Later: consultation. (2021). Regulation of Buy-Now Pay-Later: consultation - GOV.UK (www.gov.uk)
[21] Regulation of Buy-Now Pay-Later: consultation on draft legislation. (2023). Regulation of Buy-Now Pay-Later: consultation on draft legislation - GOV.UK (www.gov.uk)
[22] The Financial Services and Markets Act 2000 (Amendment) Order 2025 (2025). GOV.UK (www.legislation.gov.uk)
[23] FCA: Research Note: Deferred Payment Credit, https://www.fca.org.uk/publication/research-notes/deferred-payment-credit.pdf
[24] Financial Lives 2022: Key Findings from the FCA's Financial Lives May 2022 survey, https://www.fca.org.uk/publication/financial-lives/financial-lives-survey-2022-key-findings.pdf

impact of DPC use on the levels of indebtedness and the likelihood of getting into arrears on other credit products and on DPC.

3. Data and descriptives

3.1. Data collection and sampling

We worked in collaboration with the largest DPC providers in the UK covering, to the best of our knowledge, over 90% of the UK unregulated DPC market. All data was collected, ingested and analysed in a GDPR compliant manner. Our data covers the unregulated DPC provider market, i.e., excluding in-house merchant provision.

We collected transactions made by a ~10% quasi-random sample of each firms' customers from January 2017 (or from the point the firm entered the UK market if later) to the end of March 2023, a period of just over 6 years. This sample was drawn based on specific customer dates of birth, three in each month (hence why we describe it as "quasi-random"). These dates were common across DPC firms, allowing us to identify all DPC transactions at these firms for a common sample of customers. Our sample includes all DPC users born on these dates – new customers, existing customers, and applicants who were rejected.

For this set of individuals, Table 1 describes the data and key variables that we used. From DPC firms, we requested customer demographics, information about any applications for DPC credit made, application outcomes, transaction details, repayment records, and merchant details. Crucially, we also collected detailed firm-specific documentation of the criteria used to make DPC lending decisions, and the application-specific decisioning variables used to determine whether applications were accepted or rejected. DPC firms are typically willing to lend to a consumer so long as total outstanding credit for that consumer does not exceed an individual-specific credit limit. These credit limits are assigned based on an assessment of consumer creditworthiness and the likelihood of repayment, and it is not uncommon for the algorithms to incorporate cut-offs. For example, credit limits might be assigned based on a credit score, with credit limits increasing discontinuously at specific credit score points. These discontinuities are crucial to our causal identification strategy (see "Method for causal analysis" section).

We matched data from DPC firms to the individual's credit file information obtained from a major Credit Reference Agency (CRA) for the period up to February 2024 using date of birth, postcode, and a privacy-preserving hash constructed consistently across datasets. The data include detailed monthly information about individuals' credit portfolios, such as the types of debt held, outstanding balances and details of any arrears, as well as balances and turnover in current accounts. We obtained a match rate of nearly 80% between the DPC firm data and the CRA data.[25] CRA data covers most credit users, as well as most consumers that have a current account or a contract with a utilities or telecommunications provider. Importantly, however, they do not include any information on student loans, which are a widespread form of borrowing, particularly for younger cohorts.

[25] The reasons for non-matches between DPC firm and CRA data potentially includes a) no credit file on the individual (as DPC firms did not report to CRAs during this period), b) name changes, c) people exiting the CRA panel due to death or moving abroad, d) discrepancy in postcodes reported to the DPC firm and the CRA. The high match rate was consistent across customers at different firms.

Table 1: Outline of the datasets

Dataset	Description
Customer details	Information on all customers who made a credit application and/or attempted to sign up to during the time period, including customer characteristics and variables influencing whether an application is accepted at sign-up
Transaction information	Information on all credit applications (accepted and rejected) that were made during the time period, including transaction type and amounts
Repayment records	Information on repayments for all successful credit applications that were made during the time period, including repayment amounts and missed and late repayments
Merchant details	Information on contracts with retailers that the DPC firm partnered with at any point during the time period, including fees charged and contract details
Credit Reference Agency data	Information on a consumers' wider financial position, including balances on other credit products held and financial difficulty measures

We use the entire sample of individuals submitted by the largest four DPC firms. This amounts to 1,553,757 unique individuals. Our descriptive analysis focuses on the sample of individuals that have had at least one DPC credit application during the period. For our causal analysis on the impact of DPC on consumer financial outcomes, we make further sample restrictions which reduces the sample to 1,192,758 individuals, namely:

1. Since most of our key outcome variables come from the CRA data, we only use the ~80% of individuals that successfully matched to the CRA files. Individuals that failed to match tend on average to have lower monthly DPC borrowing, a lower credit limit and a lower credit score (as measured by the DPC firm). We provide summary statistics on the unmatched individuals in Annex A – Comparing individuals who are matched and unmatched with the Credit Reference Agency data.
2. We restrict to individuals that had at least one application that was assessed based on a score-based cut-off. This is necessary given our causal identification strategy.

Table 2 presents a summary of the sampling restrictions applied for the analysis. Overall, our final sample for the causal analysis consists of 1,192,758 individuals (76.8% of the original sample).

Table 2: Sample restrictions for the causal analysis

Criteria	Number of individuals remaining	Remaining percent
Total sample from DPC firms	1,553,757	100.0%
At least one DPC credit application	1,488,151	95.8%
DPC data matched to CRA data	1,229,317	79.1%
At least one transaction that is subject to a "threshold-type" credit application decision rule for the causal analysis	1,192,758	76.8%

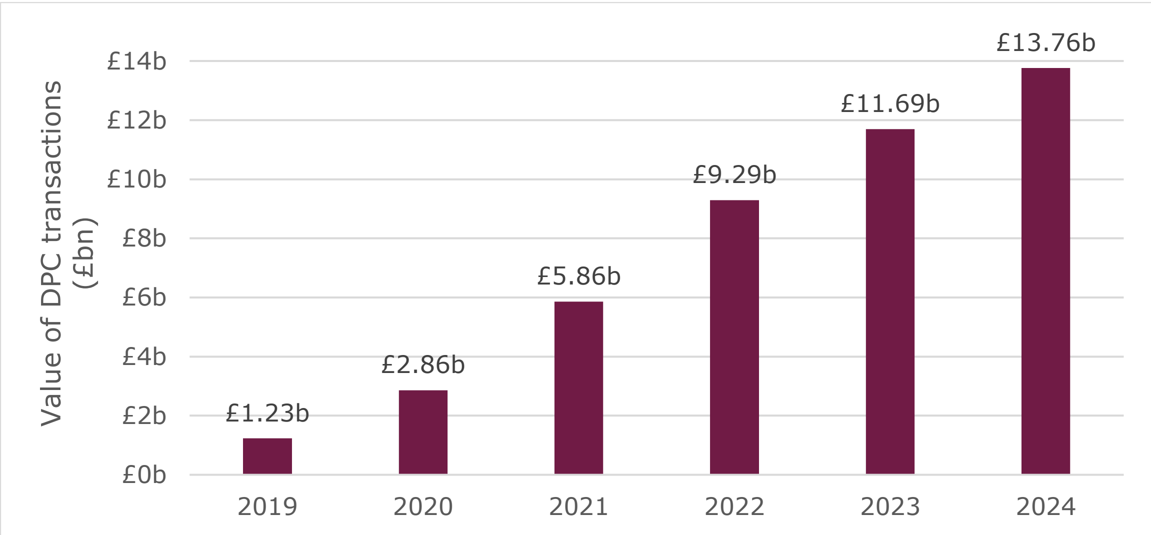
3.2. Background on the DPC market

This section gives an overview of the UK DPC market. The analysis is based on the transaction-level data described in the previous subsection, apart from the 'Growth of the DPC market' section which relies on aggregate data provided by the DPC providers[26]. At the time, these firms were the largest in the market and together are estimated to have made up over 90% of the unregulated sector.[27]

3.2.1. Growth of the DPC market

Figure 1 shows that the DPC market has been growing rapidly. Measured by the value of DPC transactions (i.e., the full value of purchases made using DPC), DPC has grown from £1.23 billion in 2019 to £13.8 billion in 2024. The DPC market has more than quadrupled since the Covid-19 economic crisis. The total value of DPC transactions during March 2023 was approximately £800 million. This corresponds to roughly 2% of total retail sales in the same month (ONS)[28]. When we compare the DPC market to the credit card market, we find that DPC remains a fraction of the credit card market at about 3% of new credit lent.[29] This difference reflects, in part, that DPC providers often only lend part of a transaction's value – for example, when the first instalment is paid upfront. More detail on the size and growth of the DPC market can be found in the accompanying Cost Benefit Analysis and Consultation Paper.

Figure 1: Total value of DPC transactions



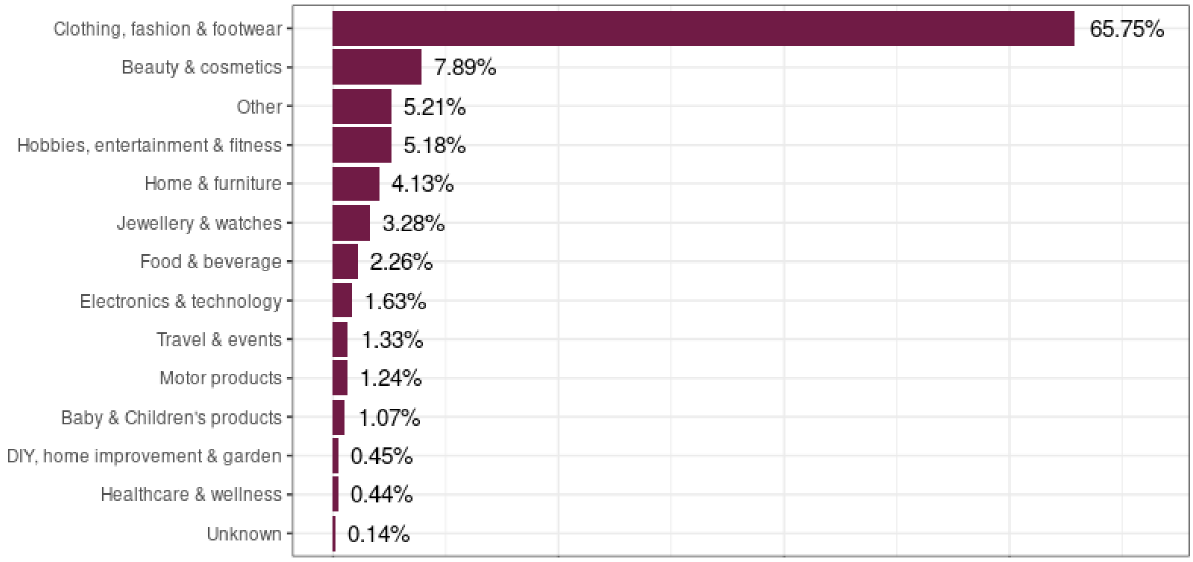
3.2.2. How DPC is used

Figure 2 shows that over 65% of the DPC transactions are in the clothes, fashion, and footwear industry. Indeed, our calculations suggest that DPC accounted for 9.5% of total retail sales in the clothing, fashion and footwear sector in March 2023 (ONS)[30]. Other

[26] The dates provided in the aggregate data (2019-2024) do not overlap completely with the transaction level data (2017-2013).
[27] Estimates in this section might differ quantitatively to estimates using other data sources.
[28] Using FCA and ONS data: https://www.ons.gov.uk/businessindustryandtrade/retailindustry/datasets/poundsdatatotalretailsales
[29] Data from UK Finance on credit card spending: https://www.ukfinance.org.uk/data-and-research/data/card-spending
[30] Using FCA and ONS data: https://www.ons.gov.uk/businessindustryandtrade/retailindustry/datasets/poundsdatatotalretailsales

significant uses of DPC include beauty and cosmetics (7.9%), hobbies, entertainment and fitness (5.2%) and home and furniture (4.1%).

Figure 2: Share of DPC spending by retail sector



Note: Figure uses DPC transactions conducted in 2022.

Figure 3 shows how the average total monthly DPC credit borrowed has changed over time. In March 2023, the average DPC borrowed among users was £168, or £67 per approved transaction. The median borrowed was £43 per transaction, roughly 2/3 of the mean, implying that there are some particularly large transactions. This compares to an average spending of £555 per active credit card in March 2023 (UK Finance)[31].

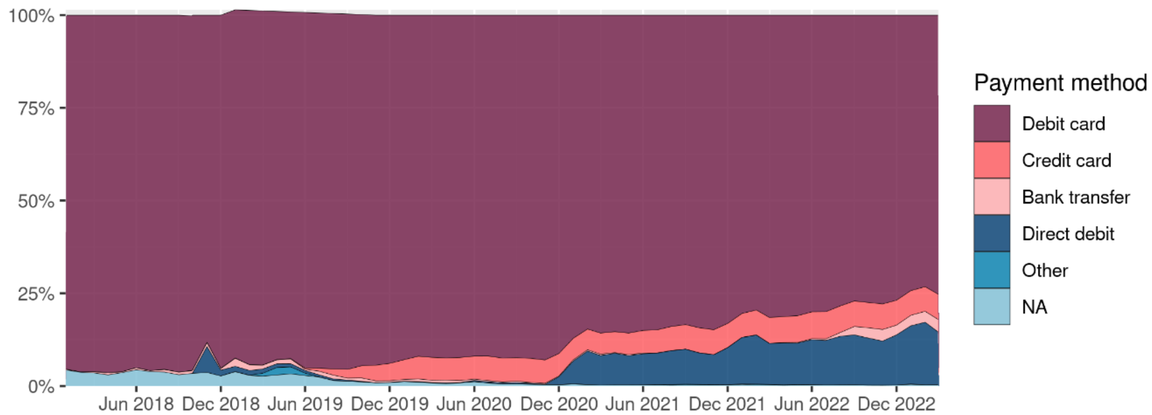
Figure 3: Average monthly borrowing by an active user in that month

Note: Averages are before repayments are made, and before refunds are issued. An active user in a given month is defined as a customer with at least one approved transaction in that month.

[31] UK Finance data: https://www.ukfinance.org.uk/system/files/2023-06/Card Spending Update - March 2023.pdf

There have been concerns raised about potential harms from DPC being exacerbated through repayments made using other forms of credit such as credit cards. Figure 4 sheds light on this and shows the share by value of DPC repayments for each payment method used by consumers. We see that three quarters of repayments are made using debit cards; only 8% are made using credit cards and this share has remained stable since 2020. Most of the remainder are made by direct debit.

Figure 4: Share (by value) of DPC credit repaid by different payment methods



Another potential source of harm could arise from the extent of missed / late repayments on DPC. On average, 10% of transactions in January 2023 had at least one missed or late repayment. 2% of transactions in January 2023 had at least one instalment that incurred a late fee, however not all DPC providers charged late fees during this period. While the extent of missed / late repayments and charges associated might seem low, this might not hold considering changing economic conditions and the fact that more firms began to charge late fees shortly after the end of our dataset in 2023.

3.2.3. Characteristics of DPC users

To understand the characteristics of DPC users and their broader financial circumstances, we can use data from DPC firms matched with consumer credit files. The credit file data also allow us to assess the extent to which DPC users differ from the broader UK population (though acknowledging that some individuals are not captured by credit files).

Table 3 provides summary statistics, comparing DPC users to the UK population. DPC users are on average significantly younger, less creditworthy (based on an FCA-estimated risk score[32]), have higher levels of unsecured debt, and have higher levels of financial difficulty compared to the UK population. The data also suggest that DPC users have a slightly higher annual income than the UK population as a whole, though this is based on a proxy measure so should be treated with caution.

[32] We use a random forest model to predict whether an individual has a new default in the next 12 months, using data from the recent past. The raw predictions of default then are transformed into a standard looking credit scores.

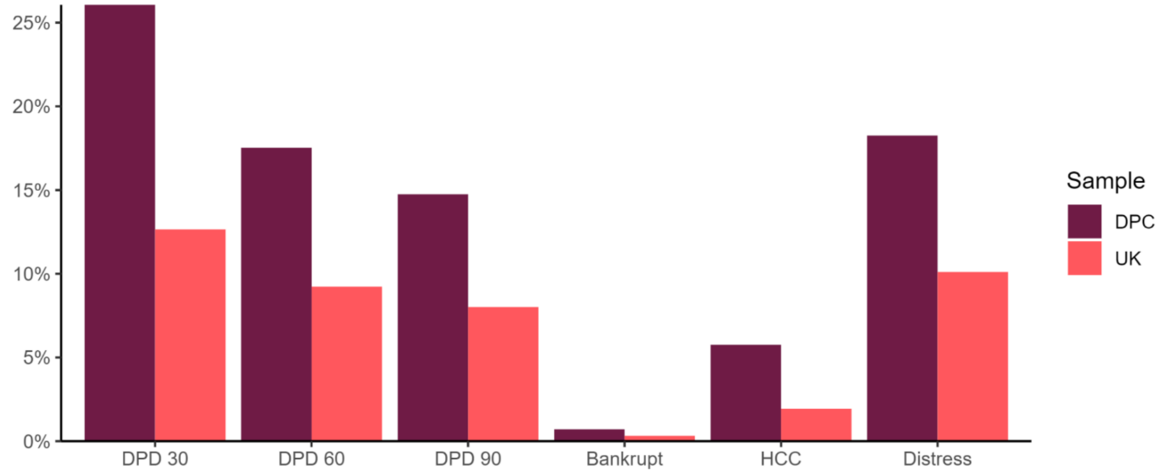
Table 3: Summary statistics for DPC users vs UK population in 2022

	DPC users		Random sample of UK population		Difference
	Mean	Standard Deviation	Mean	Standard Deviation	T-statistic
Age	40	13	49	18	431.39***
Risk score	629	47	658	45	503.44***
Net annual income (£, proxy)	31,123	22,069	29,948	24,069	32.91***
Bankrupt (%)	1.00		0.32		67.04***
Financial distress (%)	23		10		267.71***
Unsecured debt balance					
Total (£)	3,206	6,284	1,674	4,729	214.07***
of which…					
Revolving (£)	1,771	3,733	949	2,813	193.25***
Non-revolving (£)	1,435	4,084	725	3121	151.85***
of which…					
Credit card balance (£)	1,568	3,574	868	2,690	172.02***
Overdraft balance (£)	134	616	57	494	106.26***
Personal loan balance (£)	1,429	4,083	723	3,120	150.86***
Any 30 days arrears (%)	29		13		314.94***
Any 60 days arrears (%)	21		9.2		264.31***
Any 90 days arrears (%)	19		8		246.74***
Users of high-cost credit (%)	6.4		1.9		174.10***
Number of observations	1,189,210		1,240,305		

Note: This table shows summary statistics as an average of 2022 for DPC users and the general UK population. The percentage of observations for which a value is not available in the data: 0.25% (DPC users risk score), 19.23% (DPC users net income), 3.25% (UK population age), 0.95% (UK population risk score), 39.15% (UK population net income). *** denotes significance of $p < 0.01$, ** denotes $p < 0.05$, and * denotes $p < 0.10$.

To investigate the differences in financial health between DPC users and the UK population, Figure 5 compares various measures of financial difficulty across the two groups. DPC users are roughly twice as likely to be in arrears (30, 60 or 90 days past due), over twice as likely to be bankrupt and more than twice as likely to be users of high cost credit. They are also almost twice as likely to be in serious financial distress according to our composite measure (in 90-day arrears on any product, any debt sold to a debt collector, declared bankrupt or subject to a county court judgement).

Figure 5: Financial distress for DPC users vs UK population



Note: Objective vulnerability measures based on January to December 2022 CRA data and a successful DPC transaction in 2022 for those matched to CRA data. UK population includes DPC users. DPD refers to Days Past Due (i.e., arrears). HCC refers to users of high-cost credit which includes high cost short term credit, home credit, rent-to-own, guarantor, logbook, and running account. Distress means 90 Days Past Due on any credit product, has active county court judgement, bankruptcy or debt sold. There are potential definitional issues around how long bankruptcy and CCJ flags should be assumed to persist. A person is coded as being bankrupt if they were declared bankrupt within the previous 18 months. No age or income weightings are used.

3.2.4. Summary

In short, the DPC market is growing rapidly, with usage dominated by clothing, fashion and footwear. DPC tends to be used by younger, less creditworthy, more indebted individuals, and who are also more likely to be in some form of financial difficulty. Of course, this correlation does not mean that DPC use is the cause of consumer financial difficulty – the key issue we investigate in subsequent sections.

4. Methodology

4.1. Overview of econometric approach

Our parameter of interest is the causal effect of £1 additional DPC spending on various measures of consumer financial wellbeing. To estimate this causal effect, we need to compare customers who differ in the amount of DPC credit they have been granted but are otherwise indistinguishable, on average. Clearly, we must avoid comparing customers with different levels of DPC borrowing who also happen to differ across other dimensions (e.g., affluence), since it might be these other dimensions that drive differences in subsequent credit outcome rather than DPC spending itself. Our strategy for making valid comparisons involves exploiting lenders' threshold-based lending rules.

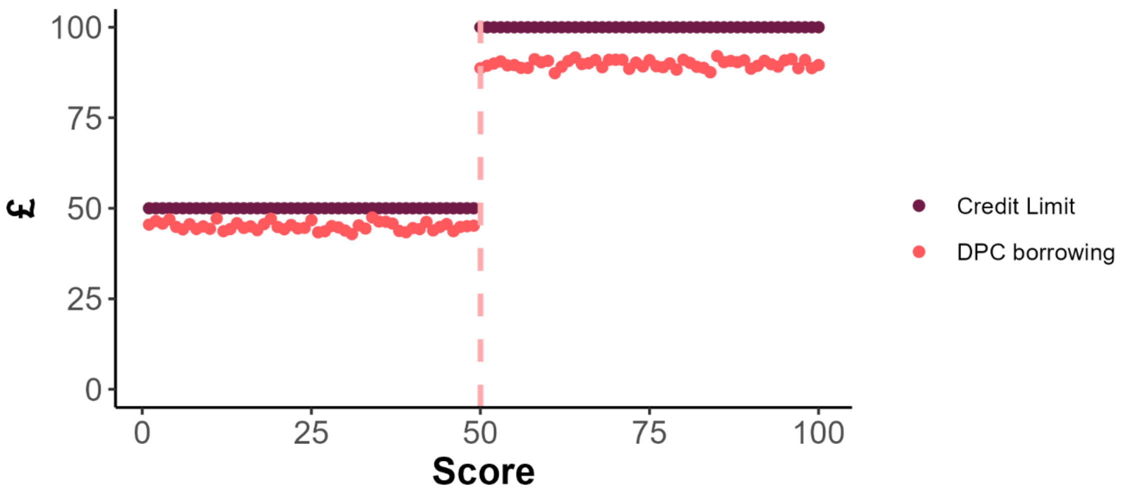
Similar to many lenders[33], DPC providers implement threshold-based lending rules. These rules determine the maximum amount of DPC credit customers are permitted to

[33] Gathergood, J., Guttman-Kenney, B., & Hunt, S. (2019). How do Payday loans affect borrowers? Evidence from the UK market. *The Review of Financial Studies*, *32*(2), 496-523 provides evidence for payday lending. Agarwal, S., Chomsisengphet, S., Mahoney, N., & Stroebel, J. (2018). Do banks pass through credit expansions to consumers who want to borrow? *The Quarterly Journal of Economics*, *133*(1), 129-190 provides evidences for credit cards.

have outstanding at any given point in time, in a similar way to a credit limit for a credit card. Thresholds are set after assessing customers' creditworthiness through either credit checks with a credit reference agency or internally developed credit rating models, typically summarised in terms of a "credit score". We will refer to this as the "score" (or "running variable") from now on. These lending rules often exhibit discrete jumps in credit limits at predetermined values of the score chosen by the firm.

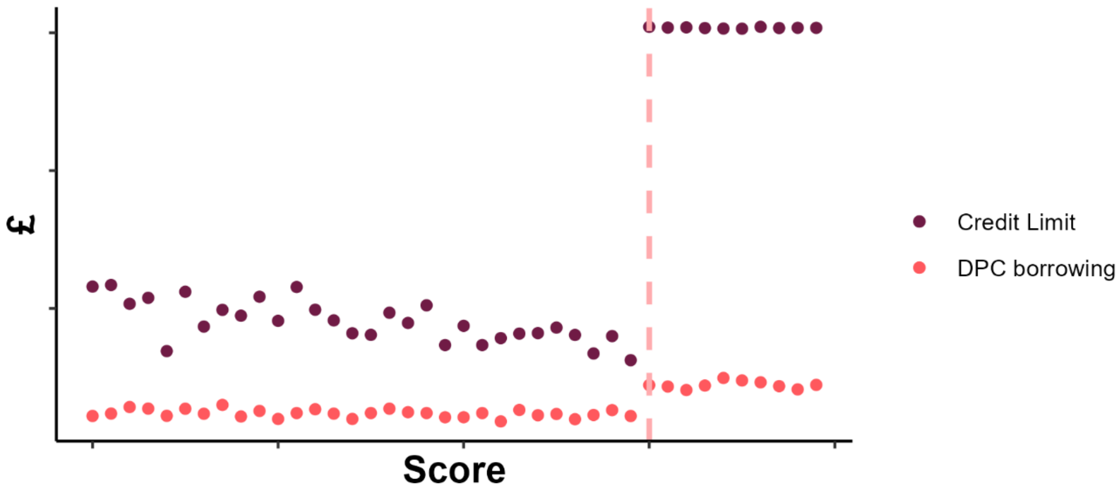
Figure 6 panel A illustrates a fictional threshold lending rule depicting the relationship between credit limit and credit score. In this example, the credit limit is £50 for individuals with a score below 50 and £100 for those with a score of 50 or above.[34] The Figure visually represents the discrete nature of the lending rule, where the credit limit changes sharply as the credit score crosses the 50 threshold. To the extent that the credit limit constraints DPC borrowing on the left (low) side of the threshold, we would expect to see DPC borrowing jump discretely at the threshold, as depicted in the Figure. We attach an actual example of such a rule that we observe in the dataset in panel B.

Figure 6: Hypothetical threshold-type lending rule in panel A, actual threshold-type lending rule in panel B



Panel A: Hypothetical threshold-type lending rule

[34] In reality, credit scores are on the range of 0-999 and there are no thresholds below 100. However, for ease of presentation we use an example of a threshold between 0 and 100.

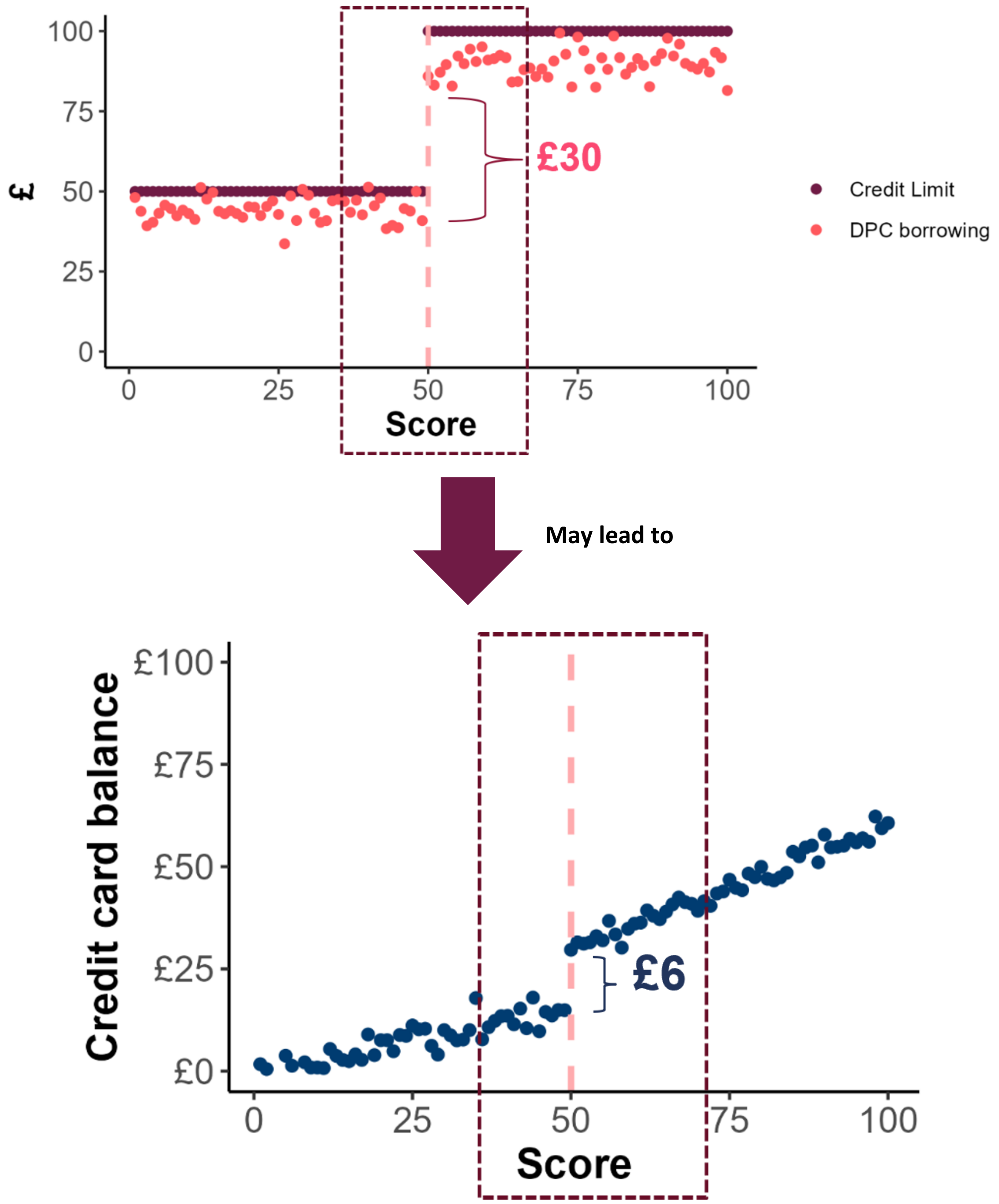


Panel B: Actual threshold-type lending rule that we observe in the dataset

Thresholds are specific to each DPC firm and tend to be assigned separately for different subgroups defined by things like the type of merchant, new vs returning customers, and the score used. The thresholds also change over time. For instance, DPC Firm A might apply a specific set of threshold-based lending rules to new customers between September 2021 and September 2022, utilising a given internal creditworthiness score. We label each of these subgroups a "process".

Our parameter of interest is the causal impact of an additional £1 of DPC borrowing on a given outcome, such as outstanding credit card balances in 12 months' time. We can isolate the effect of DPC borrowing on credit outcomes using a "fuzzy regression discontinuity design" (fuzzy RDD). The idea is to compare individuals just above the credit score threshold who have a higher DPC credit limit (and therefore likely to have higher DPC borrowing) with individuals just below the threshold who have a lower credit limit (and therefore likely to have lower borrowing). We illustrate this in Figure 7.

Figure 7: Fuzzy RDD intuition



In Figure 7, individuals who have a credit score above 50 are granted a credit limit of £100, whereas those situated on the left side below the threshold are restricted to a credit limit of £50. This induces an average increase of £30 in borrowing behaviour among those above the threshold. Consequently, individuals to the right of the threshold exhibit an average credit card balance that is £6 higher than their counterparts who fall below the score of 50. If we focus on the individuals near the threshold, depicted by the dotted box, these individuals are similar barring two key distinctions: (1) minor discrepancies in their credit scores, and (2) variations in their assigned credit limits and DPC borrowing. Under certain assumptions, these individuals are essentially alike across all dimensions except DPC credit limit and borrowing, meaning that any difference in subsequent outcomes can be attributed to the effect of DPC borrowing itself. In this example, on average, £30 additional DPC borrowing causes a £6 increase in credit card balance. This is equivalent to £1 additional DPC borrowing causing a £0.2 increase in credit card balance, on average.

In our final sample, we use a total of 551 thresholds (out of 580 identified) across the four DPC firms. 29 identified thresholds are discarded due to limited data either side of the threshold. Ultimately, we are interested in deriving market-wide measures of the impact of DPC spending on consumer outcomes, meaning we need a way to combine the information at each of these 551 thresholds. We do this in two steps: first we calculate a fuzzy RDD estimate at each of these thresholds; then we aggregate these estimates using Bayesian hierarchical methods. These two steps are discussed in the subsections below, after laying out our outcomes of interest.

To motivate our choice of aggregation method, it is helpful to consider other approaches that have been proposed in the literature for combining multiple RDD estimates. While these alternatives are theoretically useful, we find them difficult to apply in our setting due to specific data and design features. We are aware of four such frequentist approaches. The earliest papers combining information at multiple thresholds typically used a two-stage least squares approach that pools all thresholds together in a single IV regression, for example Angrist and Lavy (1999)[35] who estimate the effect of class size on school achievement, and Chen and Van der Klaauw (2008)[36] who investigate the work disincentive effects of disability insurance. Were we to apply this approach, the natural way to implement it would lead to a proliferation of instruments, and it's also not clear how to adapt the insights from RDD estimation that the fitted polynomial should be treated as a local approximation to the regression function rather than a global approximation. Another common approach involves standardising the credit score (by subtracting the mean and dividing by the standard deviation) and then running a single pooled regression with a single instrument. This approach was taken by Gathergood et al. (2019) in their study of the UK payday lending market. This is hard to implement in our setting because a given running variable often has multiple cutoffs (so which do we standardise to?) and the bandwidth available (due to the proximity of other cutoffs) varies widely from one cutoff to another, making it hard to know how to fit sensible local polynomials.

More recent work has attempted to provide theoretically grounded approaches for combining multiple thresholds. Bertanha (2020)[37] develops frameworks for both sharp and fuzzy regression discontinuity, but unfortunately the fuzzy framework requires a discrete set of possible treatments. This is not tenable in our setting where the average credit limit is a continuous variable. Finally, Cattaneo et al. (2016)[38] and Cattaneo et al. (2021)[39] consider alternative multi-cutoff settings but their framework seems to require that the running variable is comparable across different cutoffs. This condition is frequently violated in our context, due to differences across DPC firms and changes in their credit scoring rules over time.

Taken together, these limitations lead us to adopt the Bayesian hierarchical framework discussed in subsection 4.2 to aggregate our many local RDD estimates. This approach

[35] Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly journal of economics*, *114*(2), 533-575.
[36] Chen, S., & Van der Klaauw, W. (2008). The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics*, *142*(2), 757-784.
[37] Bertanha, M. (2020). Regression discontinuity design with many thresholds. *Journal of Econometrics*, *218*(1), 216-241.
[38] Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., & Keele, L. (2016). Interpreting regression discontinuity designs with multiple cutoffs. *The Journal of Politics*, *78*(4), 1229-1248.
[39] Cattaneo, M. D., Keele, L., Titiunik, R., & Vazquez-Bare, G. (2021). Extrapolating treatment effects in multi-cutoff regression discontinuity designs. *Journal of the American Statistical Association*, *116*(536), 1941-1952.

allows us to flexibly model heterogeneity across thresholds while preserving the local identification strength of each design.

4.2. Outcomes of interest

Given our focus on the impact of using DPC on consumer financial well-being, our primary outcomes of interest are derived from the CRA dataset: changes in non-DPC debt balances and measures of financial difficulty.

Our first outcome of interest is the change in non-DPC debt balances relative to one month before the DPC application. This will enable us to answer whether using DPC leads to the accumulation of other forms of (potentially interest-bearing) debt. We use the *change* in debt relative to a baseline period rather than the *level* of debt because there is a risk that part of the reason the level of debt is unbalanced in any given period is due to pre-existing differences in borrowing either side of the threshold. This could happen if credit scores and discontinuities in the credit limit are persistent over time (which they seem to be).

We focus on types of debt that seem most likely to be affected by additional DPC spending. In particular, we focus on total unsecured debt balance (which includes revolving debt like credit cards, overdrafts, and store cards, and non-revolving debt like high-cost short-term credit, and personal loans) because DPC is most likely to be used for similar purposes as other forms of unsecured debt. We also look at the component parts of total unsecured debt separately (i.e., total revolving debt and total non-revolving debt).

For these outcomes, we consider the following time horizons:

- The month of the transaction: this is the point at which DPC credit has just been taken out. This allows us to study whether DPC credit is being used as a substitute for other forms of credit.
- 3 months after the transaction: by this stage, most if not all repayments will have fallen due. This allows us to investigate whether customers are using other forms of credit to repay their DPC borrowing.
- 6 and 12 months after the transaction: medium-term outcomes, typically some time after the DPC credit should have been repaid in its entirety. This allows us to look for evidence that DPC debt is rolling onto other forms of borrowing including those that incur interest.

In addition to debt balances, we are also interested in outcomes relating to financial difficulty. We consider three such measures:

1. A flag for being in 30-day arrears, which means having missed one payment on any non-DPC credit product.
2. A flag for serious financial distress, meaning that at least one of the following has occurred – reaching 90-day arrears or more (or a default) on a non-DPC credit product, declaration of bankruptcy in the previous 18 months, a credit account being passed to a debt collection agency, or a county court judgment being issued against the consumer (FCA occasional Paper 49, and Federal Reserve Board, 2007)[40, 41].

[40] FCA occasional paper 49: https://www.fca.org.uk/publication/occasional-papers/occasional-paper-49.pdf
[41] Federal Reserve Board, (2007), "Report to the Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit," Federal Reserve Board

These measures of distress are also included in the FCA's description of financial distress (FCA Handbook, 2014)[42]. We focus on financial distress 7-12 months after borrowing.

3. A flag for being in arrears on DPC 7-12 months after borrowing (7-12 months for consistency with the flag for serious financial distress).

One important issue is that the time horizon over which outcomes are measured affects the periods we can use DPC spending data from since we don't observe outcomes over full 12-month periods for more recent DPC spending. The implication of this is that our findings are limited to a specific time-period depending on the outcome variable and does not cover the changing nature of the DPC market. Specifically, for the various outcome horizons, we use DPC borrowing data as follows:

- From the month of the transaction until 6 months after the transaction: DPC borrowing data between 05/2017 and 03/2023 (full sample).
- 3 months after the transaction: DPC borrowing data between 05/2017 and 03/2023 (full sample).
- 6 months after the transaction: DPC borrowing data between 05/2017 and 03/2023 (full sample).
- 12 months after the transaction: DPC borrowing data between 05/2017 and 01/2023.
- 7-12 months after the transaction (arrears outcomes): DPC borrowing data between 05/2017 and 01/2023.

4.3. Threshold-specific regression discontinuities

As described above, we use a total of 551 credit limit thresholds across the four DPC firms. For each threshold, we conduct a fuzzy RDD for each discontinuity to identify the causal impact of DPC spending on various consumer outcomes. In this subsection, we discuss how we calculate our threshold-specific fuzzy RDD estimates and the assumptions required to recover valid causal parameters; in the next subsection we describe how we aggregate across the threshold-specific estimates.

We begin with our fuzzy RDD framework, which we can express using the language of instrumental variables. The first stage is:

$$\sum_{j=0}^{k} DPC_{it+j} = \pi_{0h} + \pi_{1h} 1[S_{it} \geq 0] + \pi_{2h} S_{it} + \pi_{3h}\, 1[S_{it} \geq 0] \times S_{it} + u_{it} \text{ mko0}$$

And the reduced form:

$$Y_{it+k} = \gamma_{0h} + \beta_h \pi_{1h} 1[S_{it} \geq 0] + \gamma_{2h} S_{it} + \gamma_{3h} 1[S_{it} \geq 0] \times S_{it} + v_{it}$$

The coefficient of interest is β_h and:

- DPC_{it}: individual i's total DPC spending in month t less any refunds. For $k > 0$ we take the sum of DPC borrowing from 0 to k (discussed below)
- S_{it}: individual i's credit score (centred around the threshold), as assessed during their first DPC application in month t[43]. This is the "running" variable. The

[42] FCA Handbook: https://www.handbook.fca.org.uk/handbook/CONC/1/3.html
[43] 85% of individuals are assessed by only one process by a firm in a month.

coefficient for this variable is threshold-specific since the scale of the internal score differs across processes.

- Y_{it+k}: the outcome variable for individual i in time $t+k$. For balance-type outcome variables, such as debt amounts, we take the first differences (i.e. $Y_{it+k} - Y_{it-1}$)

The instrument is a dummy for whether the score exceeds the threshold: $1[S_{it} \geq 0]$.

In RDD terms, the specification above is a "local linear regression" (i.e., the polynomial in the S_{it} is linear). We use a triangular kernel and select bandwidths to minimise mean squared error. We cluster standard errors at the customer level[44]. These modelling choices are in line with best practice described in Calonico, Cattaneo, and Titiunik's (2014)[45].

Under the fuzzy RDD setup, the estimate has a causal interpretation under the following assumptions, which seem reasonable in our context:[46]

(1) Individuals near the threshold cannot control their running variable (the score) to locate themselves one side of the threshold rather than the other

(2) Only the credit limit changes at the threshold. No other changes are triggered immediately around the cutoff

(3) The jump in the running variable (the score) at the threshold only affects the outcome of interest through the amount of DPC that the individual borrows

(4) Thresholds affect individuals' DPC spending decisions. That is, for individuals around the threshold, increasing the credit limit would have a meaningful impact on the average level of DPC spending

With regard to (1), while consumers can improve their credit score through a history of responsible credit usage and timely repayment, it seems highly unlikely that consumers know what the score thresholds are or where they are located relative to the threshold, since neither are made known to customers. For (2), information provided by DPC firms suggests that, as required, only the credit limit changes at the threshold.

Assumption (3) merits a little more discussion. It seems reasonable to assume that the current month's DPC credit limit affects outcomes of interest (debt balances and measures of financial difficulty) in the current month only via its effect on DPC borrowing in the current month. But an issue arises when outcomes are measured further in the future, e.g., the credit card balance in 12 months' time. From the data, there is considerable stickiness over time in both the location of thresholds and individuals' credit scores. As a result, if an individual is located one side of a threshold in one month, they are quite likely to be the same side of the threshold in subsequent months. This opens up the possibility that today's credit limit affects future DPC spending – and therefore future outcomes of interest – because the credit limit next month is unchanged relative to today. To address this issue, the treatment variable in all our regressions is the sum of DPC spending (less refunds) between today and the point at which the outcome is

[44] We use cluster-robust variance estimation with degrees-of-freedom weights using nearest neighbour variance estimation. An individual can enter the analysis more than once if they are assessed by the same process in different months. It is worth highlighting the significance of clustering within our setup, particularly concerning outcomes that extend further into the future (e.g., credit card balance after 6 months). This importance arises from the fact that the variable of interest represents the cumulative DPC borrowing between the transaction month and the outcome month. When we treat subsequent months for the same individual as separate observations in the regression, it implies a substantial "reuse" of the same information, so observations of the same individual are no longer independent.
[45] Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica, 82(6), 2295-2326.
[46] Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2024). A practical introduction to regression discontinuity designs: Extensions. Cambridge University Press.

measured (where the outcome is a rolling sum, this is the last month in the rolling window). To illustrate, Table 4 gives an example based on an individual who used DPC in 1/2024, setting out how the outcome variable and treatment variable are defined for different outcome horizons.

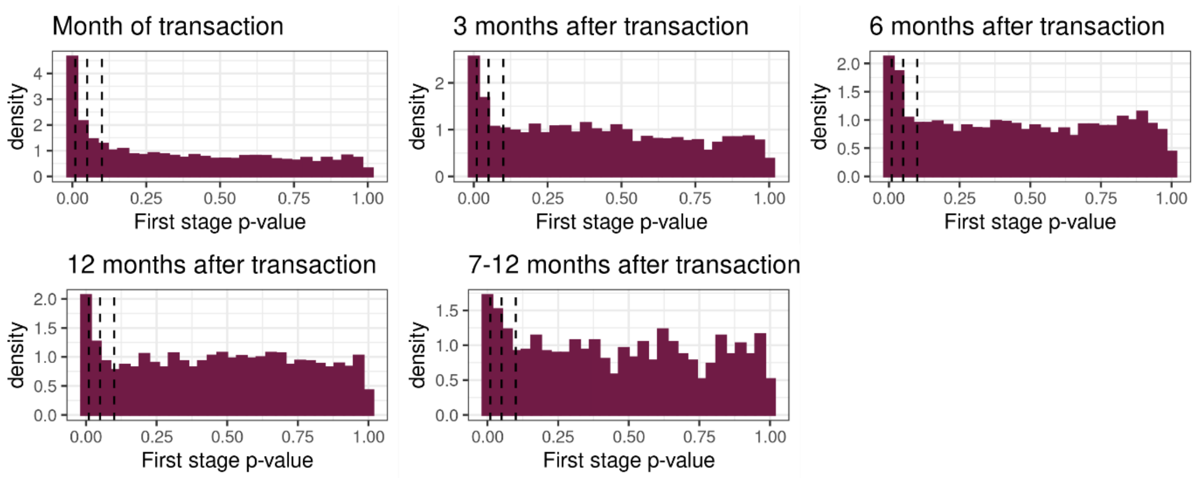
Table 4: Definition of outcome measures and treatment variables

Month	Months since 1/2024	Outcome measure	Treatment variable
1/2024	0	Debt in 1/2024 minus debt in 12/2023	DPC spending (less refunds) in 1/2024
2/2024	1	Debt in 2/2024 minus debt in 12/2023	DPC spending (less refunds) between 1/2024 - 2/2024
3/2024	2	Debt in 3/2024 minus debt in 12/2023	DPC spending (less refunds) between 1/2024 - 3/2024
4/2024	3	Debt in 4/2024 minus debt in 12/2023	DPC spending (less refunds) between 1/2024 - 4/2024
5/2024	4	Debt in 5/2024 minus debt in 12/2023	DPC spending (less refunds) between 1/2024 - 5/2024
6/2024	5	Debt in 6/2024 minus debt in 12/2023	DPC spending (less refunds) between 1/2024 - 6/2024
7/2024	6	Debt in 7/2024 minus debt in 12/2023	DPC spending (less refunds) between 1/2024 - 7/2024

An important consequence of defining the treatment variable in this way is that it affects the interpretation of estimates over longer horizons. For example, for an outcome measured in six months, the interpretation is not "the impact in six months' time of £1 additional DPC spending today"; rather it is "the impact in six months' time of £1 additional DPC spending over the next six months." This should be borne in mind in interpreting the results below.

Assumption (4) also requires some justification. This is an empirical issue: does the jump in the credit limit at each threshold induce sufficient variation in DPC spending to allow a causal estimate to be estimated? To assess this, Figure 8 plots the distribution of first-stage p-values across our 551 thresholds. The first-stage p-value provides a measure of the strength of the relationship between DPC spending and the instrument (which side of the threshold an individual is). Lower p-values imply a stronger relationship. It is clear from this Figure that there is a wide degree of variation in the strength of the relationship, with particular concerns for longer horizons.

Figure 8: First stage p-values distribution, by different type of outcome variables

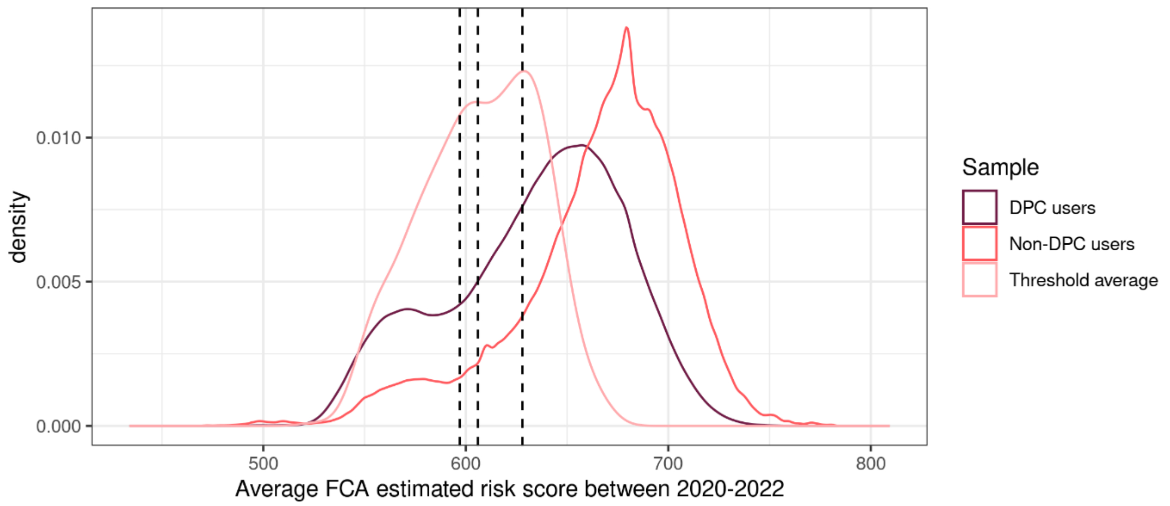


Although there is a clear concentration of thresholds with low p-values, particularly for shorter outcome horizons, many other first stages are weak. The consequence of having a weak first stage can be serious: coefficient estimates may be biased (towards the OLS estimate that ignores endogeneity), standard errors underestimated and the distribution of estimates not well approximated by a normal distribution (see Andrews et al., 2019)[47]. As a result, we have carefully considered the appropriate set of thresholds to use, the sensitivity of results to different decisions, and have explored approaches that are robust to weak instruments. We discuss them in more detail in the next subsection.

If the assumptions above hold, the causal parameter that the fuzzy RDD estimator recovers is a local average treatment effect (LATE) – that is, the average effect of increased borrowing for individuals at the threshold whose borrowing behaviour is actually influenced by the change in the credit limit. One question, therefore, is the extent to which score thresholds affect a representative set of DPC users. This issue is taken up in Figure 9, which compares the distributions of credit scores for DPC users, DPC non-users and averages around our 551 thresholds. (We use the FCA's in-house risk score for this purpose since credit scores used by firms are not comparable). This shows that the thresholds our fuzzy RDD approach exploits are concentrated more towards low credit score DPC users. In turn, DPC users tend to have lower credit scores than non-users. This means that the estimates we calculate below will tend to relate to the less creditworthy end of the distribution of DPC users, who themselves are a less creditworthy subset of the broader population – so not fully representative but a key group for policymaking given that concerns about harms are likely to focus on less creditworthy users.

[47] Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. Annual Review of Economics, 11, 727-753.

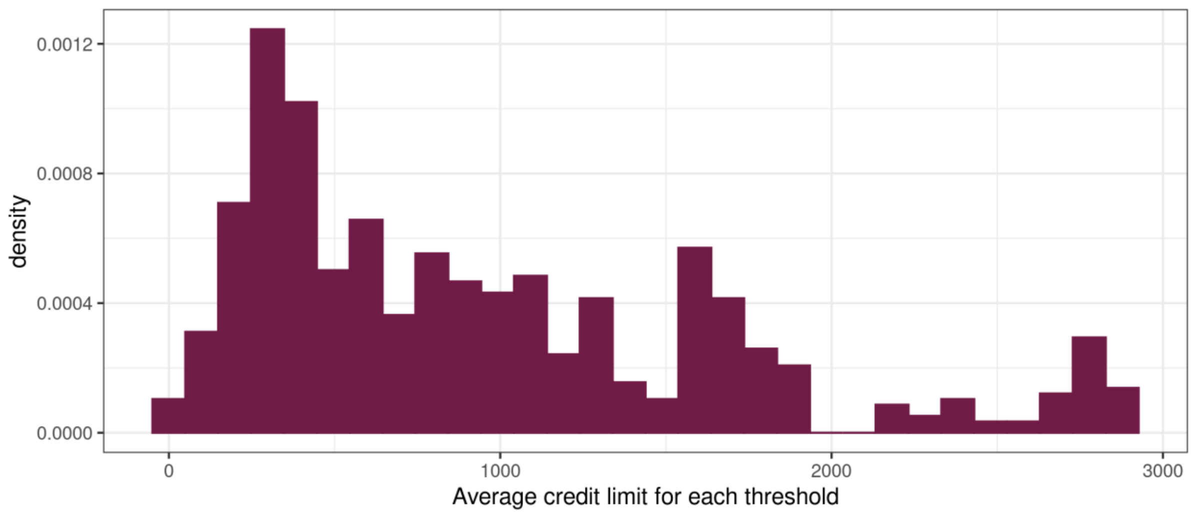
Figure 9: Distribution of FCA credit score for DPC users, non-DPC users and the thresholds



Note: Density represents how likely it is to find data points within a particular range of values on the x-axis (FCA-estimated risk score). The dashed vertical lines represent segments for very poor (<597), poor (597-606), fair (607-628) and good (> 629) credit scores segments.[48]

Figure 10 depicts the distribution of average transaction credit limits across the thresholds used in the analysis. The greatest density is for thresholds with transaction credit limits below 500. We anticipate that these are likely to be the thresholds that bind the most[49], that is, where DPC borrowing differs most across the threshold, so we have a chance of estimating the causal effect of additional DPC borrowing at that threshold. These thresholds are more likely to be binding because the lower the credit limit is, the more likely that it will constrain consumers' spending decisions.

Figure 10: Distribution of average transaction credit limit for each threshold



Note: Density represents how likely it is to find data points within a particular range of values on the x-axis (Average credit limit for each threshold).

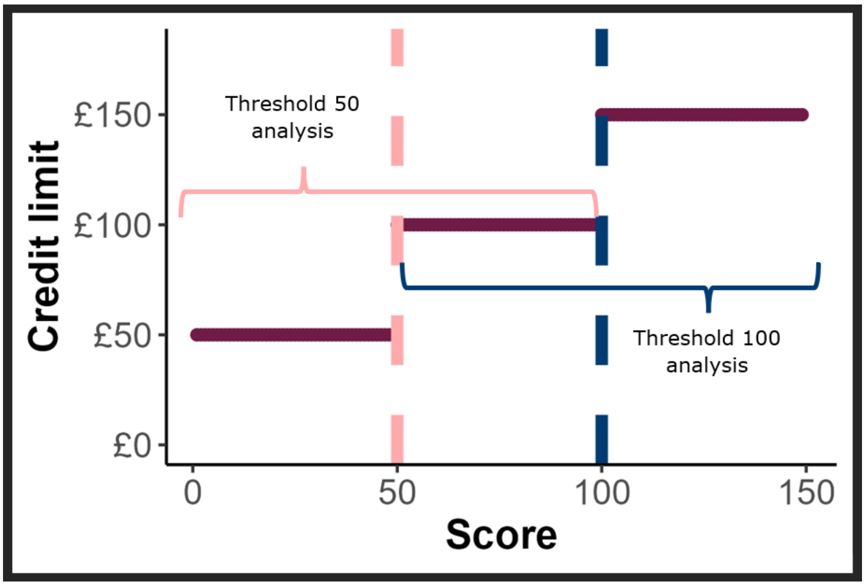
For processes with multiple thresholds, it is possible for a given observation to be close to more than one threshold. We allow observations to be used to estimate the impact at

[48] The bands ("Poor", "Fair", etc) are defined by benchmarking what the CRA does in their banding.
[49] We further investigate whether a threshold binds for customers located close to that threshold. Details in Annex B.

multiple thresholds, but we restrict bandwidths such that they do not include other thresholds (otherwise the outcome function is potentially discontinuous, so we could not justify fitting a polynomial to it). In practice, this means that a given observation is used to estimate the impact at a maximum of two thresholds.

To illustrate, Figure 11 depicts a process with two thresholds, one at score 50 (pink dashed line) and one at score 100 (blue dashed line). The pink and blue braces represent the widest possible bandwidths used to calculate the fuzzy RDD estimates for the two thresholds. In practice, the bandwidth selection procedure may select a narrower set of credit scores around each threshold, to reduce the bias and/or improve the precision of estimates.

Figure 11: Illustration of observations used for each threshold specific fuzzy RDD



4.4. Bayesian Hierarchical Aggregation

Each threshold-specific RDD can be treated as a distinct study, akin to the concept in meta-analysis, where the goal is to aggregate findings from multiple studies to derive an overarching estimate. It is becoming increasingly common in economics to use Bayesian hierarchical methods to perform this sort of aggregation. For example, Meager (2019)[50] uses this approach to aggregate impact estimates from different microcredit expansions, while Havranek and Sokolova (2020)[51] use it to aggregate estimates of whether consumers follow consumption rules of thumb.

We summarise each RDD model by two numbers: the fuzzy RDD "treatment effect" coefficient and its standard error, $(\widehat{\beta_h}, \widehat{se_h^2})$, with studies indexed by h. We assume that the standard error is known, which is easily justified at the number of observations we consider here.

[50] Meager, R. (2019). Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, *11*(1), 57-91.
[51] Havranek, T., & Sokolova, A. (2020). Do consumers really follow a rule of thumb? Three thousand estimates from 144 studies say "probably not". *Review of Economic Dynamics*, *35*, 97-122.

Meta-analysis involves combining results from various studies to characterise the mean results and their variability across studies. We utilise Rubin's (1974)[52] hierarchical model to aggregate threshold-specific estimates. The input is the series of pairs ($\widehat{\beta_h}$, $\widehat{se_h^2}$) and the output is an estimate of two quantities: the "global" mean and "global" variance, also referred to as the hypermean (or average effect) and hypervariance (or heterogeneity).

This approach is also referred to as a random-effect (RE) meta-analysis. The term RE refers to the fact that the model allows for heterogeneity in effects across studies: in this case, thresholds that vary across different firms and processes within firms.

The model is summarised by two equations:

$$\widehat{\beta_h} \sim N\left(\beta_h, \widehat{se_h^2}\right) \forall\, h$$

$$\beta_h \sim N\left(\beta + \gamma X_{h,}\, \sigma^2\right) \forall\, h$$

The first line is the study-specific model: we assume that that the fuzzy RDD coefficients are normally distributed around the true effect. The second line refers to the RE component: each of the true study-specific effects is normally distributed[53] around a common mean (β) plus the effect of any covariates at the study level (X_h), with some variance (σ^2). We only include X_h in our heterogeneity analysis, where we consider heterogeneity with respect to income and risk score groups.

The model can be estimated using standard statistical software. In this instance, we use the *baggr* (Bayesian aggregator) package in R, which implements Bayesian inference for all common meta-analytic models and simplifies processing of data, summaries, and visuals.[54]

To carry out Bayesian inference, priors are needed for β and σ^2. In the absence of any prior knowledge about the effects of the intervention, the typical choice for the mean effect (β) is to centre the prior at no effect (in this case zero) and to allow it to vary widely relative to the input data.[55] Similarly, for heterogeneity (σ^2), it is typical to scale it in accordance with empirical between-study variance. This is done automatically in *baggr* in a way that is specific to meta-analysis of each outcome.

When σ^2=0 (fixed effects, FE), the meta-analysis model has a closed form formula for β, expressed as a simple weighted average of each β_h, with weights equal to inverse variance of each estimate. In the RE model, the weights depend on the additional parameter σ^2, which is unknown and needs to be estimated jointly with β. Intuitively, the RE inference procedure can be seen as finding the amount of heterogeneity which best explains differences between study-specific estimates. In other words, answering the question of whether, or what part of, the variability in RDD coefficients is due to true differences in study-specific effects vs sampling variation in each study.

Another way of phrasing this is through the lens of what happens to individual study estimates. This is referred to as "pooling" of study effects around the common mean:

- If σ^2 is greater than 0 and less than infinity ("partial pooling"), each true RDD coefficient is somewhere between the study-specific estimate and the common

[52] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.
[53] We conduct robustness checks on the normal distribution assumption. We refer readers to Annex C.
[54] Więcek, Witold, and Rachael Meager. 'Baggr: Bayesian Aggregate Treatment Effects Package'. 2024, available at https://cran.r-project.org/web/packages/baggr/index.html
[55] For Bayesian aggregation in our paper, we use the following priors: Hyper-mean ~ N(0,5), Hyper-sd ~ N(0, 100), gamma ~ N(0,100).

mean or in algebra: $\beta_h = p\beta + (1-p)\widehat{\beta_h}$ with $0 < p < 1$. We refer to p as the pooling metric.

- If $\sigma^2 = 0$ ("full pooling", p = 1), each true RDD coefficient is assumed to be identical across studies. All differences are due to sampling variation.
- If σ^2 is infinite ("no pooling", p = 0), each true RDD coefficient is equal to its study-specific estimate, and what happens in other studies does not influence it.

RE and FE models can be estimated using both Bayesian and frequentist inference. While the statistical and philosophical interpretations of these models are very different, their estimates tend to converge when (1) there are a lot of reasonably precise estimates, (2) Bayesian priors are uninformative (as in our case).

However, Bayesian inference has some distinctive advantages that go beyond the possibility of using informative priors. First, various frequentist methods for meta-analysis are known to underestimate heterogeneity in RE models due to difficulty of jointly estimating the "hyperdistribution"[56] (β, σ^2). Secondly, Bayesian inference makes it simple to do model selection and to derive predictive distributions, that is distributions of effects that can be expected in future implementations.

As described above, a significant proportion of the 551 thresholds are associated with weak instruments. Annex F shows that with these weak instruments, the asymptotic distribution of the fuzzy RD estimator and, as a result, the aggregate distribution of our fuzzy RD estimates is not necessarily normal (as we have assumed). To account for this in some way, our main results below are based on aggregating only those thresholds where the first stage value is below 0.1. Depending on the outcome and horizon under consideration, this means using between 59 and 170 discontinuities.

Concerns are sometimes raised about the validity of making such exclusions, so we have also experimented with aggregating using all thresholds (regardless of the strength of the associated first stage) and aggregating using a t-distribution and weak-instrument robust distribution based on Staiger and Stock (1997) and Montiel-Olea and Pfleuger (2013). Results based on these alternative approaches are reported in Annexes E and F respectively.[57]

There is little difference between our baseline results and the results assuming a t-distribution. Our results using all thresholds and either a non-robust or a weak instrument robust distribution suggest that there are some small differences in aggregation but still present a broadly consistent story. However, these results are experimental. As a result, in the next section we focus on results excluding fuzzy RD estimates with a first stage p value of below 0.1 from our aggregations.

[56] A hyper distribution describes a situation where parameters come from a higher level distribution. In our situation, it is a prior distribution on a hyperparameter. In our context, β and σ^2 are the are the priors of the β_h, which is the mean of the $\widehat{\beta_h}$
[57] We do not report results in this paper, however we also tried the approach of Chamberlain and Imbens (1996)[57], in which first stage and reduced form estimates are aggregated separately, rather than aggregation taking place directly on the ratio. Results were broadly similar.

5. Results

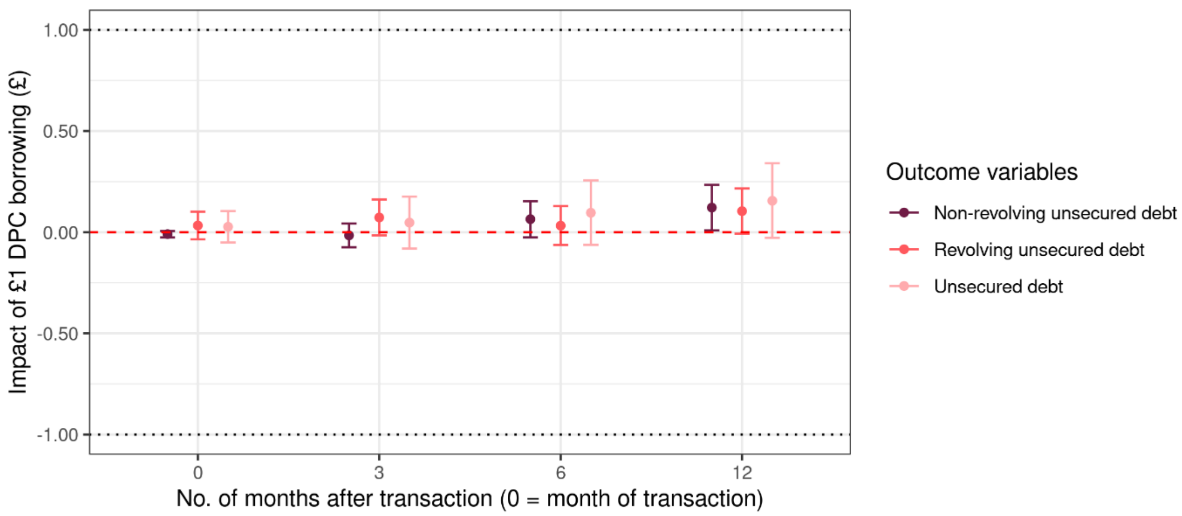
5.1. Aggregate impacts

This section presents our headline findings on the impact of DPC usage on debt balances and financial distress. We present our results graphically where the dot, lower and upper whisker represents the mean, 2.5% and 97.5% of the aggregated mean's posterior distribution respectively.[58] We include the results in tables in Annex DAnnex D. These percentiles form a credible interval, indicating the range within which we believe the true parameter value lies with a certain degree of confidence. If the credible interval includes zero, it suggests that there is a non-negligible probability that the parameter is zero or close to zero. This implies that we cannot be sure that the causal effect we are interested in is different to zero. If, on the other hand, zero falls outside the credible interval, this indicates that we can be confident that the parameter is likely to be different from zero, implying statistical significance[59].

We first investigate debt balance outcomes for total unsecured credit (excluding DPC) and split by revolving and non-revolving credit. Figure 12 shows that there is no clear evidence that additional DPC borrowing has an impact on total unsecured credit (excluding DPC). For month 0, estimates are precise zeros, implying that, in the short run, additional DPC borrowing is additional credit and not a substitute for other forms of unsecured borrowing. The main DPC products in our data are typically paid off over a period of less than six months, so we might worry that DPC debt rolls off onto other forms of unsecured borrowing in the medium run. Figure 12 suggests there is little evidence to support this, at least at an aggregate level: impacts for months 3 and 6 are all close to zero and quite precisely estimated. However, at 12 months, the impacts are slightly larger, and credible intervals only overlap with zero marginally.

While we find little evidence of impact at an aggregate level, an impact may exist for particular subgroups. We explore this issue in the next subsection.

Figure 12: Effect of DPC borrowing on changes in unsecured debt levels



[58] It is worthwhile to note how this differs from a frequentist approach. Under the Bayesian framework, we use prior knowledge about a parameter of interest and combine it with observed data to update our beliefs, resulting in a posterior distribution representing our updated belief of the parameter. As opposed to the frequentist approach, where they could make an inference to what would have happened if many experiments were repeated, this posterior distribution incorporates our uncertainty about the parameter, allowing us to quantify it in terms of probability.
[59] Unlike null hypothesis significance testing approach in frequentist statistics, Bayesian inference does not include a concept of "significance". However, we can simply calculate a related quantity: the probability that the estimated effect is different from zero. We refer to parameters for which the 95% interval does not include zero as "significant".

Note: 95% credible intervals. Unsecured debt balance includes revolving debt (credit cards, overdrafts, and store cards), and non-revolving debt (high-cost short-term credit, and personal loans). Bayesian aggregation is used to produce results.

In Figure 13, we focus on specific types of unsecured debt: overdrafts, credit cards and personal loans. Consistent with our previous findings, at the aggregate level, we do not find evidence that additional DPC borrowing is a short-term substitute for these types of unsecured debt. At longer horizons, we also do not find evidence that DPC borrowing tends to roll off onto overdrafts; there is some evidence that a small fraction of DPC borrowing may roll onto other forms of unsecured credit by the end of the outcome horizon, though credible intervals are close to crossing zero.

These findings are somewhat less pronounced than deHaan et al. (2022)[60], whose results suggest that access to BNPL in the US causes significant increases in overdraft charges, credit card interest and credit card late fees. Some of the differences between the two studies can be explained by the distinct ways overdraft, credit card, and DPC fees operate in the US and UK markets, as well as differences in DPC users across the countries and the samples used. While there are also methodological differences between the studies, the results discussed in Annex F give us confidence in the robustness of our qualitative findings.

Figure 13: Effect of DPC borrowing on changes in unsecured debt levels

Note: 95% credible intervals. Credit cards and overdrafts are components of revolving debt, and personal loan is a component of non-revolving debt. Bayesian aggregation is used to produce results.

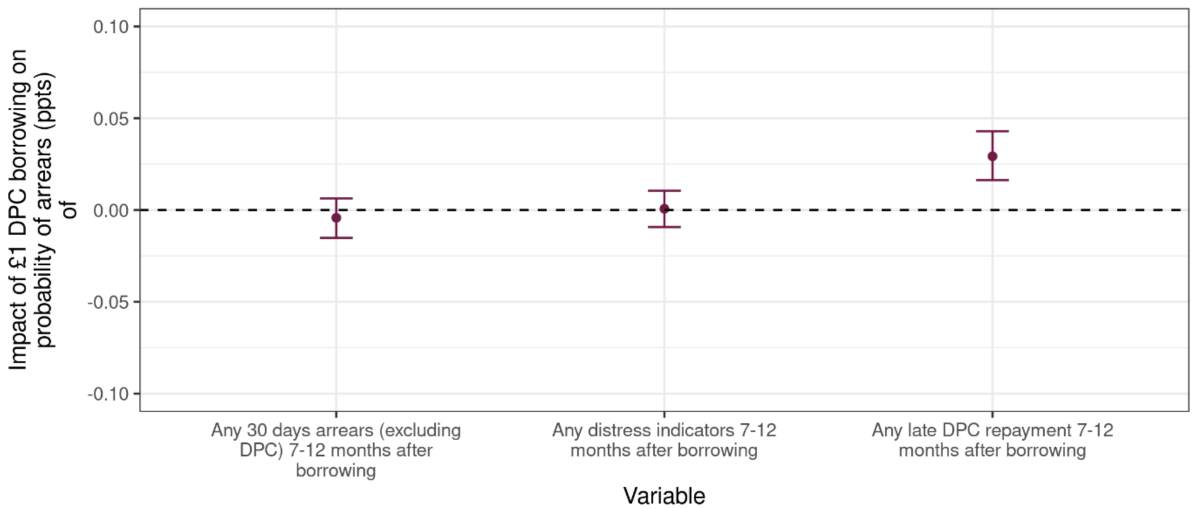
Figure 14 shows the effect of a £1 increase in DPC borrowing on 30 days arrears (excluding DPC) 7-12 months after the DPC borrowing, financial distress, and late repayments on DPC 7-12 months after borrowing. There is some evidence that additional DPC borrowing increases the likelihood of being in arrears on DPC in 7-12 months after the DPC borrowing. While this is an important finding, increasing lending to consumers increases the chances that they don't repay on time. The estimated effect is an increase of around 0.02 percentage points in the probability of being in DPC arrears for each additional £1 of DPC borrowed. For context, this implies for instance that a £110 increase in DPC borrowing – roughly the average amount borrowed – would raise the probability of arrears by around 2.2 percentage points at the midpoint of the credible interval. This suggests the effect, while statistically non-zero, is modest in size – across our sample, roughly 28.7% of transactions are associated with a late repayment. We find no clear effects of additional DPC borrowing on 30-day arrears on other forms of credit nor financial distress in the 7-12 months after borrowing. On aggregate, across all

[60] deHaan, E., Jungbae, K., Lourie, B., Zhu, C. (2023). Buy Now Pay (Pain?) Later. SSRN Working Paper.

outcomes, we do not find statistically robust evidence indicating that additional DPC borrowing causes individuals to experience increased financial distress in the medium term on average.

For Figures 12–14, Annex F shows that when we aggregate our fuzzy RD estimates using a method that accounts for weak instruments – where standard results about estimator accuracy don't hold – our overall qualitative conclusions remain largely the same. This approach, based on established weak instrument theory, helps correct for potential biases and underestimated standard errors that arise when instruments are weak. The consistency of our findings using this alternative aggregation approach gives us confidence in our main conclusions about the current effects of DPC borrowing.

Figure 14: Effect of DPC borrowing on financial distress measures



Note: 95% credible intervals. 'ppts' means percentage points. Distress means 90 Days Past Due on any credit product, has active county court judgement, bankruptcy or debt sold, but excludes distress on DPC credit. DPC repayment outcome is defined as no repayment or any late DPC repayment 7-12 months after any DPC borrowing. Method used is Bayesian aggregation.

5.2. Impacts by subgroup

The previous section shows that there is no statistically significant evidence that additional DPC borrowing leads to higher medium-term debt or arrears, on average, among the group of DPC users for whom credit limits might be binding.

In this section, we investigate whether this "no effects on average" result masks differing effects across different sub-groups of DPC users. We segment the results by dividing thresholds by different credit scores and income brackets.[61] This is important as it allows us to understand whether the benefits or harms disproportionately affect customers with lower incomes (and who therefore may be more vulnerable) and those categorised as subprime (who may have a poor history repaying debts and/or may struggle to borrow from elsewhere).

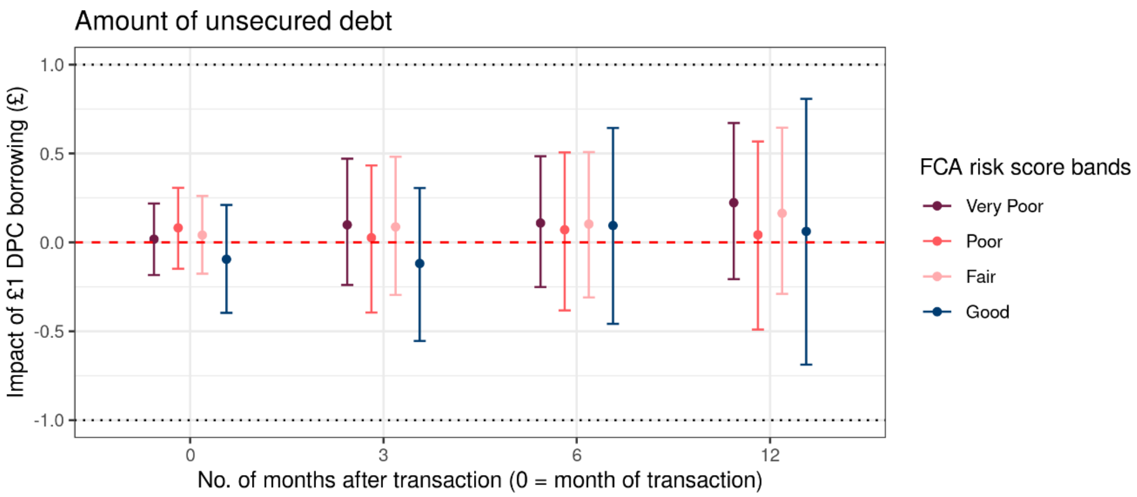
It is important to acknowledge that we anticipate less precision in the estimates compared to the overall results. This is because there are fewer threshold-specific effects available to aggregate within each subgroup band. As a result, it is hard to draw as firm

[61] This arises from the limitations of the models, which only allow us to segregate at the threshold level rather than at the individual level. We compute the average of the variable (such as credit score or income) for individuals who were evaluated under the specific threshold.

conclusions as in the previous section. Nevertheless, we don't find evidence of adverse outcomes affecting particular subgroups and most of the point estimates are fairly close to zero.

We first examine how impact on total unsecured debt varies vary across the credit score distribution. Figure 15 shows that there is little evidence that this impact varies across the credit score distribution: point estimates are at or close to zero across all outcome horizons. But credible intervals are wide, especially at 12 months, meaning that we cannot rule out significant effect sizes across most groups.

Figure 15: Effect of DPC borrowing on changes in unsecured debt levels, by credit score estimated by FCA



Note: 95% credible intervals. Unsecured debt balance includes revolving debt (credit cards, overdrafts, and store cards), and non-revolving debt (high-cost short-term credit, and personal loans). Method used is Bayesian aggregation.

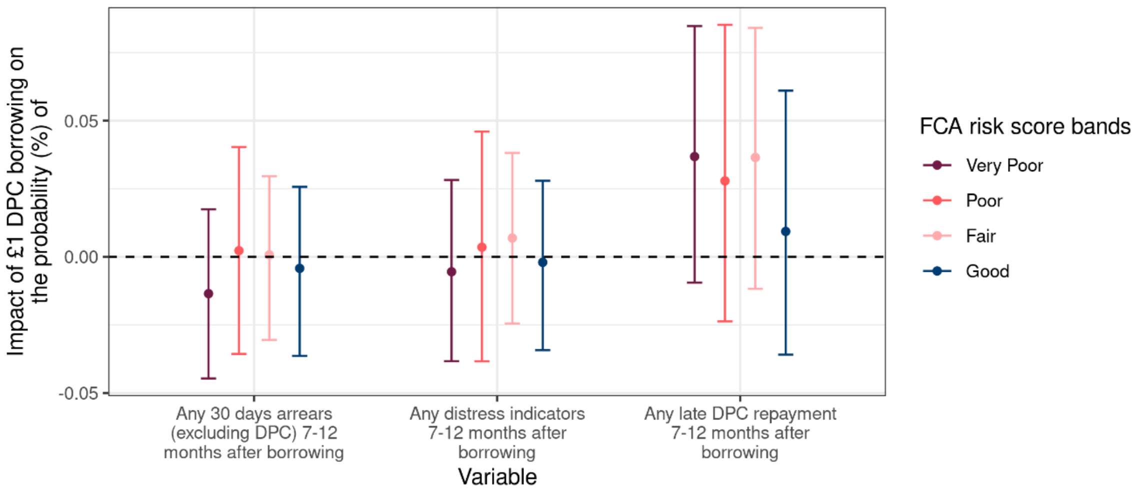
We also examine how impacts vary across the income distribution. Figure 16 shows that we do not find evidence of impacts varying across three net income bands. There is little variation across the bands or time horizons, though again the credible intervals become wide so that we cannot rule out fairly substantial effects. For the highest income band (above £31,000) the estimated hypermean for 6 and 12 months is close to one and credible intervals are very large - this is in part due to heterogeneity in impacts, but is also driven by the small number of thresholds in this income band.

Figure 16: Effect of DPC borrowing on changes in unsecured debt levels, by net income band (FCA estimate)

Note: 95% credible intervals. Unsecured debt balance includes revolving debt (credit cards, overdrafts, and store cards), and non-revolving debt (high-cost short-term credit, and personal loans). Method used is Bayesian aggregation.

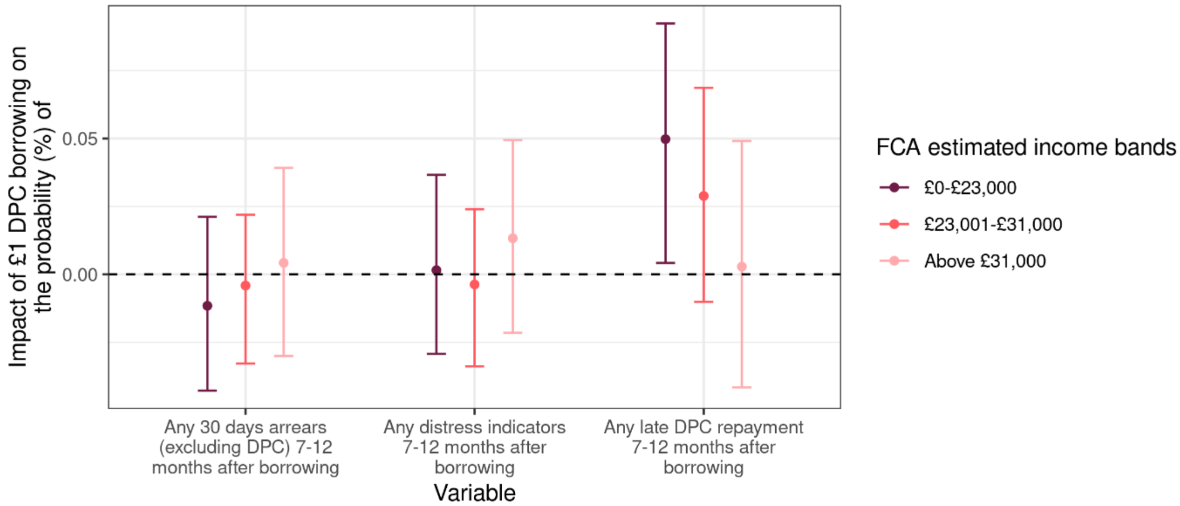
Figures 17 and 18 show how effects on arrears-related outcomes vary by risk score and income bands. There is not much variation of note, except perhaps that impacts on DPC arrears seem to be lower for more creditworthy and higher income groups (though well within the bounds of statistical uncertainty).

Figure 17: Effect of DPC borrowing on financial distress measure, by credit score estimated by the FCA



Note: 95% credible intervals. Distress measure excludes distress on DPC credit. Distress means 90 Days Past Due on any credit product, has active county court judgement, bankruptcy or debt sold. DPC repayment outcome captures no repayment or any late DPC repayment 7-12 months after any DPC borrowing. Method used is Bayesian aggregation.

Figure 18: Effect of DPC borrowing on financial distress measure, by income bands



Note: 95% credible intervals. Distress measure excludes distress on DPC credit. Distress means 90 Days Past Due on any credit product, has active county court judgement, bankruptcy or debt sold. DPC repayment outcome captures no repayment or any late DPC repayment 7-12 months after any DPC borrowing. Method used is Bayesian aggregation.

6. Conclusion

The Deferred Payment Credit (DPC) market has grown substantially in the UK – and around the world – since 2017, leading to concerns about its effects on consumer outcomes. While offering consumers the flexibility to spread the cost of retail goods through interest-free short-term credit, the unregulated nature of DPC raises questions over potential risks to consumers. The lack of regulatory oversight means that firms providing DPC are not obligated to carry out creditworthiness and affordability assessments, potentially exposing consumers to unaffordable credit with a high risk of non-repayment. Absence of regulation also means DPC providers are not required to provide the same protections consumers get when using other regulated credit products – for instance, support from their lender should they experience financial difficulty.

In this paper, we exploited detailed transaction-level data and precise documentation of credit assignment rules provided directly by lenders, including the four largest UK DPC providers at the time. We produce descriptive statistics on the users of DPC and estimate the causal impact of DPC credit on consumer financial outcomes, focusing specifically on consumer indebtedness and the likelihood of arrears and financial distress. Whilst these are key dimensions over which potential DPC harms may be evident, they only represent a subset of all the harms considered in relation to DPC. More details on these harms are set out in the accompanying Consultation Paper and Cost Benefit Analysis. These include unfair treatment of consumers and the impact of fees and charges when payments are missed. These harms have not been assessed in this research.

Consistent with previous work, our descriptive analysis finds that users of DPC users are, on average, younger, less creditworthy, have higher levels of unsecured debt, and have higher levels of financial difficulty compared to the UK population. They are also almost twice as likely to be in serious financial distress than the UK population.

In general, the analysis has not found consistent evidence across time horizons that DPC use has negative impacts on consumer outcomes. While DPC spending does increase total short-term indebtedness (including DPC), we have not found consistent evidence that DPC use leads to greater long-term (potentially interest-bearing) indebtedness, nor to higher arrears on other credit products. However, we do find evidence that additional DPC borrowing modestly increases the likelihood of falling into arrears on DPC itself.

We have not been able to assess the implications of DPC on higher levels of late payment fees given few providers charged late fees during the time period covered by our data, but late payment fees have since become more common. We have been able to look at implications of DPC by credit risk score and income groups but have not been able to assess implications for consumer groups that are otherwise vulnerable such as those with mental health conditions. Other potential channels of harm that are outside the scope of this paper include whether DPC use causes higher levels of spending, consumer misunderstanding of the product, the lack of regulatory protections to consumers, the way offers are presented which might exploit consumers' behavioural biases, treatment of vulnerable consumers, and the expansion of DPC products from online to in-store. Our analysis includes four providers of DPC which, at the time, were the largest in the UK and made up ~90% of the unregulated sector, but we omit smaller providers, results for whom might not generalise. These may be important to assess to determine the consumer welfare implications of DPC use.

The nuanced conclusions from our analysis emphasise the importance of considering both descriptive and causal analysis to unravel complex relationships within the DPC market. It is important to consider the findings from this analysis alongside other evidence relating to DPC.

Annex A – Comparing individuals who are matched and unmatched with the Credit Reference Agency data

Table 5: Comparing individuals who are matched and unmatched with the CRA

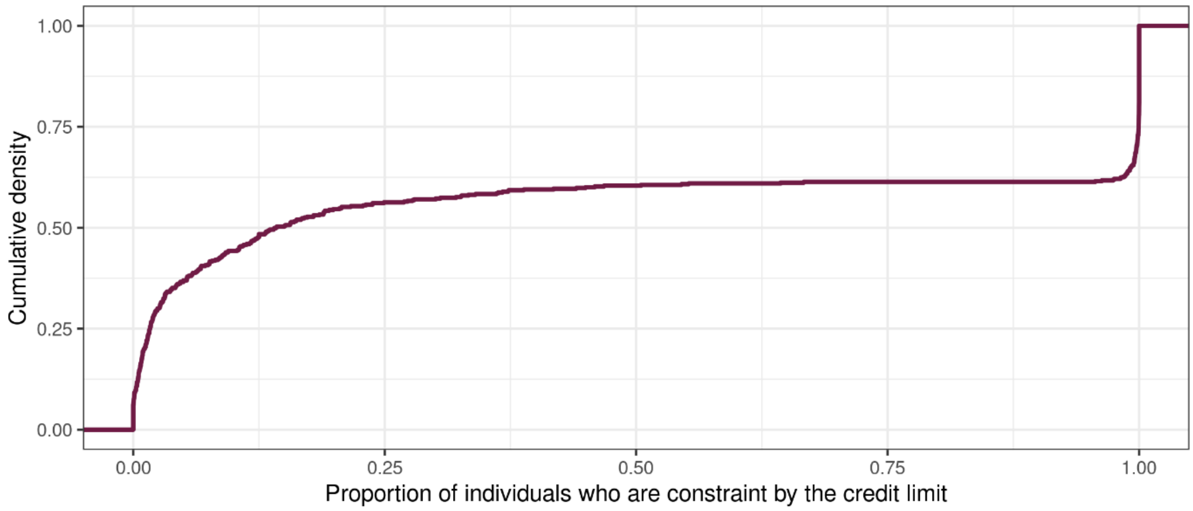
	Not matched with the CRA		Matched with the CRA		Difference
Variable	**Mean**	**SD**	**Mean**	**SD**	**T-Statistics**
Age at the first DPC application	35	15	39	14	111.042***
No. DPC provider applied per month	1.1	0.20	1.1	0.25	177.32***
Average monthly DPC borrowed with at least one application	54	105	90	115	172.239***
Having at least one approved transaction	0.56	0.5	0.87	0.33	343.082**
Average credit limit	702	765	1055	735	236.958***
Average firm assessed credit score (standardised)	-0.092	1	0.15	0.77	75.894***
No. individuals	333,032		1,253,377		

Annex B – Investigating whether individuals are constrained by the credit limit

We assess whether a threshold contains credit-constrained customers by documenting the proportion of customers on the right side whose firm specific outstanding debt exceeds the average credit limit of customers on the left side. If a customer's outstanding debt exceeds the hypothetical credit limit they would have received, had they been on the left side of the threshold, their additional DPC credit application would have been rejected. Thus, being on the right side of the threshold alleviates this credit constraint, enabling them to borrow more. It should be noted that this may not necessarily be a perfect measure, as individuals on the left side of the threshold can borrow from other DPC providers.

Figure 19 shows how the proportion of credit-constrained individuals varies across thresholds. At the median threshold, 14% of individuals appear to be credit constrained – meaning they borrowed more than they would have been allowed to if they had been just below the threshold. In 45% of thresholds, fewer than 10% of individuals are constrained, suggesting that in many cases, the additional credit granted by crossing the threshold isn't essential for most borrowers. However, in 22% of thresholds, all individuals appear to be credit constrained – every one of them borrowed more than they would have been permitted otherwise. This indicates that while many borrowers just above the threshold may not be strongly constrained, there are some thresholds where access to additional credit clearly plays a critical role in enabling borrowing.

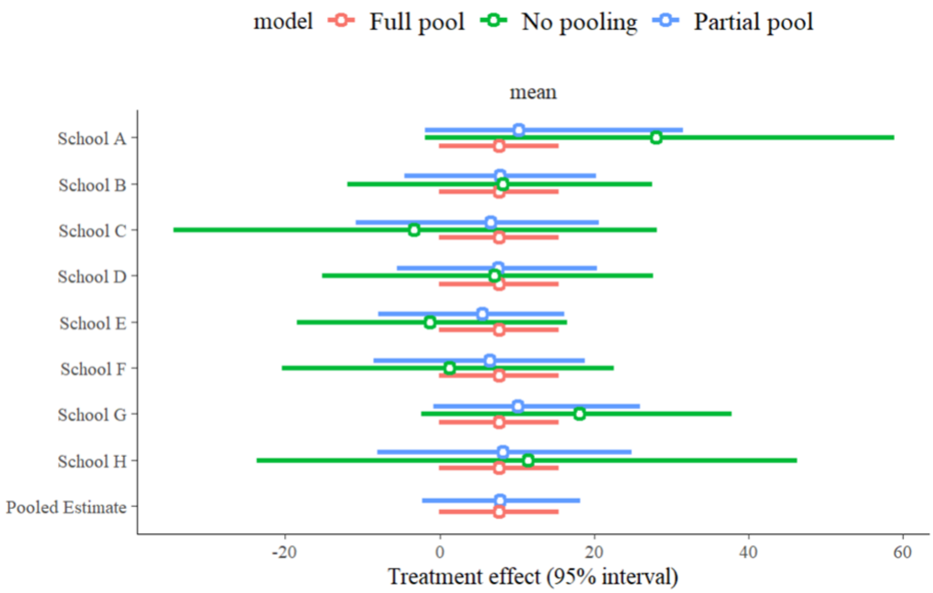
Figure 19: Cumulative density of the proportion of customers who are credit constrained



Annex C – Details on the random effect model

The random effect (RE) model nests all three possibilities of (1) partial pooling, (2) full pooling and (3) no pooling. In Bayesian inference on RE model we can then calculate a summary statistic known as overall model pooling, which is a weighted average of pooling across individual studies.

Graphically, this is best illustrated by the following visual, which recreates the canonical analysis of eight educational experiments by Rubin (1974)[62] under these three distinct models. That is, we use identical Bayesian priors on the mean in each model but change assumptions on σ^2:



Each point is the mean estimate (Bayesian posterior). Each line is a 95% uncertainty interval. We can see that

- Unpooled estimates of individual studies are large and variable. There is no common mean ("pooled estimate") to be calculated, since studies are assumed to be unrelated.
- Fully pooled estimates are all identical: each individual study is equal to the common mean. The estimate is narrow, but this comes at the cost of making a potentially strong assumptions that there are no differences between each study.
- Partially pooled estimates of studies have smaller errors than unpooled ones, but larger estimates than in the fully pooled model. The hypermean is slightly wider than in the fully pooled model, because there is additional uncertainty in the model (about the parameter σ^2)

[62] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

Annex D – Results in tabular form

Table 6: Table of causal results

Variable	2.5% Percentile	Mean	97.50% Percentile
Any 30 days arrears (excluding DPC) 7-12 months after borrowing			
Income - £0-£23,000	-0.00043	-0.00012	0.00021
Income - £23,001-£31,000	-0.00032	-0.00004	0.00022
Income - Above £31,000	-0.00030	0.00004	0.00040
Risk score - Very Poor	-0.00045	-0.00013	0.00018
Risk score - Poor	-0.00036	0.00002	0.00040
Risk score - Fair	-0.00031	0.00001	0.00030
Risk score - Good	-0.00036	-0.00004	0.00026
All estimate	-0.00015	-0.00004	0.00006
Any distress indicators 7-12 months after borrowing			
Income - £0-£23,000	-0.00029	0.00002	0.00037
Income - £23,001-£31,000	-0.00034	-0.00004	0.00024
Income - Above £31,000	-0.00022	0.00013	0.00050
Risk score - Very Poor	-0.00039	-0.00006	0.00028
Risk score - Poor	-0.00039	0.00003	0.00046
Risk score - Fair	-0.00025	0.00007	0.00038
Risk score - Good	-0.00035	-0.00002	0.00028
All estimate	-0.00009	0.00001	0.00011
Credit card balance - 0 months			
Income - £0-£23,000	-0.09620	0.04017	0.17664
Income - £23,001-£31,000	-0.09211	0.00766	0.10526
Income - Above £31,000	-0.25608	0.04278	0.33986
Risk score - Very Poor	-0.11725	-0.00765	0.10042
Risk score - Poor	-0.07050	0.05814	0.18790
Risk score - Fair	-0.08950	0.03138	0.15019
Risk score - Good	-0.22650	-0.04752	0.13167
All estimate	-0.03212	0.01322	0.05796
Credit card balance - 3 months			
Income - £0-£23,000	-0.14633	0.08817	0.31876
Income - £23,001-£31,000	-0.10995	0.04549	0.20274
Income - Above £31,000	-0.31348	-0.00499	0.30467
Risk score - Very Poor	-0.08024	0.10078	0.27349
Risk score - Poor	-0.20173	0.00597	0.22155
Risk score - Fair	-0.14180	0.04089	0.22636
Risk score - Good	-0.22879	-0.00717	0.21114
All estimate	-0.02490	0.04765	0.12045
Credit card balance - 6 months			
Income - £0-£23,000	-0.25949	-0.01545	0.22174
Income - £23,001-£31,000	-0.13151	0.06624	0.25727
Income - Above £31,000	-1.60590	0.06760	1.75219
Risk score - Very Poor	-0.16027	0.07093	0.30680
Risk score - Poor	-0.20232	0.07175	0.33191
Risk score - Fair	-0.25188	0.00099	0.23872
Risk score - Good	-0.26122	0.05412	0.35432
All estimate	-0.02818	0.05232	0.13181

Credit card balance - 12 months			
Income - £0-£23,000	-0.16607	0.04565	0.25238
Income - £23,001-£31,000	-0.03122	0.14348	0.31952
Income - Above £31,000	-1.81396	1.43514	4.68497
Risk score - Very Poor	-0.03015	0.17225	0.38984
Risk score - Poor	-0.12776	0.13233	0.39043
Risk score - Fair	-0.18306	0.03325	0.24528
Risk score - Good	-0.16036	0.13323	0.40882
All estimate	0.03063	0.11651	0.20218
Amount of Arranged/Unarranged overdraft - 0 months			
Income - £0-£23,000	-0.07100	-0.00082	0.06254
Income - £23,001-£31,000	-0.03884	0.00903	0.05962
Income - Above £31,000	-0.07157	0.01711	0.10680
Risk score - Very Poor	-0.03060	0.01748	0.06426
Risk score - Poor	-0.03616	0.01459	0.06498
Risk score - Fair	-0.05552	-0.00851	0.04267
Risk score - Good	-0.05254	0.00839	0.07011
All estimate	-0.00808	0.00864	0.02539
Amount of Arranged/Unarranged overdraft - 3 months			
Income - £0-£23,000	-0.06692	-0.01131	0.04425
Income - £23,001-£31,000	-0.03054	0.00188	0.03536
Income - Above £31,000	-0.12305	0.26391	0.65356
Risk score - Very Poor	-0.04765	-0.00996	0.02893
Risk score - Poor	-0.03932	0.00785	0.05505
Risk score - Fair	-0.03918	0.00173	0.04218
Risk score - Good	-0.03500	0.02098	0.07695
All estimate	-0.01541	0.00164	0.01869
Amount of Arranged/Unarranged overdraft - 6 months			
Income - £0-£23,000	-0.03785	0.01361	0.06460
Income - £23,001-£31,000	-0.02770	-0.00045	0.02882
Income - Above £31,000	-0.08242	0.08996	0.26278
Risk score - Very Poor	-0.03316	0.00559	0.04256
Risk score - Poor	-0.06943	-0.02323	0.02260
Risk score - Fair	-0.02896	0.00272	0.04280
Risk score - Good	-0.03130	0.02148	0.07435
All estimate	-0.01030	0.00082	0.01437
Amount of Arranged/Unarranged overdraft - 12 months			
Income - £0-£23,000	-0.05964	-0.01176	0.03399
Income - £23,001-£31,000	-0.02679	0.00451	0.03823
Income - Above £31,000	-195.41756	-0.20724	194.86788
Risk score - Very Poor	-0.03992	-0.00317	0.03459
Risk score - Poor	-0.05092	-0.00619	0.03711
Risk score - Fair	-0.03988	0.00233	0.04199
Risk score - Good	-0.01008	0.04735	0.10611
All estimate	-0.01277	0.00196	0.01687
Any late DPC repayment 7-12 months after borrowing			
Income - £0-£23,000	0.00004	0.00050	0.00093
Income - £23,001-£31,000	-0.00010	0.00029	0.00069
Income - Above £31,000	-0.00041	0.00003	0.00049
Risk score - Very Poor	-0.00010	0.00037	0.00085
Risk score - Poor	-0.00024	0.00028	0.00085

Risk score - Fair	-0.00012	0.00036	0.00084
Risk score - Good	-0.00036	0.00009	0.00061
All estimate	0.00016	0.00029	0.00043
Total number of 'hard credit searches' 7-12 months after borrowing			
Income - £0-£23,000	-0.00069	0.00010	0.00093
Income - £23,001-£31,000	-0.00051	0.00015	0.00081
Income - Above £31,000	-0.00089	0.00038	0.00163
Risk score - Very Poor	-0.00062	0.00011	0.00086
Risk score - Poor	-0.00037	0.00048	0.00132
Risk score - Fair	-0.00077	0.00000	0.00079
Risk score - Good	-0.00068	0.00011	0.00091
All estimate	-0.00013	0.00015	0.00044
Amount of Personal loan - 0 months			
Income - £0-£23,000	-0.04390	0.00920	0.06653
Income - £23,001-£31,000	-0.05907	-0.01130	0.03532
Income - Above £31,000	-0.07971	0.01086	0.10132
Risk score - Very Poor	-0.04832	0.00115	0.05365
Risk score - Poor	-0.06806	-0.01042	0.04226
Risk score - Fair	-0.05966	-0.00678	0.04651
Risk score - Good	-0.09601	-0.03181	0.03496
All estimate	-0.02150	-0.00678	0.00804
Amount of Personal loan - 3 months			
Income - £0-£23,000	-0.21971	-0.02798	0.18137
Income - £23,001-£31,000	-0.18125	-0.02548	0.12127
Income - Above £31,000	-2.36993	-0.76295	0.84583
Risk score - Very Poor	-0.15090	-0.02298	0.11263
Risk score - Poor	-0.14570	0.01117	0.16272
Risk score - Fair	-0.13287	0.01627	0.15513
Risk score - Good	-0.42490	-0.23299	-0.03788
All estimate	-0.08083	-0.02674	0.02701
Amount of Personal loan - 6 months			
Income - £0-£23,000	-0.14747	0.03181	0.25079
Income - £23,001-£31,000	-0.14758	0.02787	0.20316
Income - Above £31,000	-0.72759	0.48227	1.69484
Risk score - Very Poor	-0.17722	0.03299	0.24372
Risk score - Poor	-0.15497	0.05662	0.27026
Risk score - Fair	-0.16006	0.02084	0.22533
Risk score - Good	-0.31966	-0.00372	0.31605
All estimate	-0.04822	0.02641	0.10380
Amount of Personal loan - 12 months			
Income - £0-£23,000	-0.31369	-0.03673	0.24587
Income - £23,001-£31,000	-0.06905	0.16145	0.39114
Income - Above £31,000	-0.43145	0.05701	0.54280
Risk score - Very Poor	-0.17897	0.07398	0.33208
Risk score - Poor	-0.26100	0.04402	0.34926
Risk score - Fair	-0.02576	0.26857	0.55856
Risk score - Good	-0.39955	-0.01863	0.35421
All estimate	-0.00655	0.10019	0.20688
Amount of unsecured debt - 0 months			
Income - £0-£23,000	-0.15649	0.10556	0.36714
Income - £23,001-£31,000	-0.16053	0.02259	0.20183

Income - Above £31,000	-0.75509	-0.21450	0.32876
Risk score - Very Poor	-0.18559	0.01799	0.21703
Risk score - Poor	-0.14849	0.08244	0.30805
Risk score - Fair	-0.17797	0.04076	0.25994
Risk score - Good	-0.39786	-0.09420	0.21208
All estimate	-0.05109	0.02676	0.10425
Amount of unsecured debt - 3 months			
Income - £0-£23,000	-0.39649	0.03794	0.46477
Income - £23,001-£31,000	-0.26446	0.04348	0.36163
Income - Above £31,000	-0.46123	0.11717	0.68861
Risk score - Very Poor	-0.23788	0.09994	0.47607
Risk score - Poor	-0.39392	0.02663	0.43753
Risk score - Fair	-0.30134	0.08810	0.48500
Risk score - Good	-0.55806	-0.11901	0.31262
All estimate	-0.08102	0.04730	0.17617
Amount of unsecured debt - 6 months			
Income - £0-£23,000	-0.30815	0.15151	0.60755
Income - £23,001-£31,000	-0.22542	0.08215	0.38792
Income - Above £31,000	-1.34249	0.98995	3.31961
Risk score - Very Poor	-0.24745	0.10939	0.48437
Risk score - Poor	-0.38344	0.07136	0.50868
Risk score - Fair	-0.30883	0.10274	0.51208
Risk score - Good	-0.46075	0.09505	0.64787
All estimate	-0.06275	0.09603	0.25596
Amount of unsecured debt - 12 months			
Income - £0-£23,000	-0.39122	0.10653	0.60297
Income - £23,001-£31,000	-0.21177	0.15644	0.53387
Income - Above £31,000	-0.63746	1.06830	2.75860
Risk score - Very Poor	-0.20870	0.22361	0.66984
Risk score - Poor	-0.49059	0.04310	0.56579
Risk score - Fair	-0.29303	0.16351	0.64566
Risk score - Good	-0.69417	0.06038	0.81526
All estimate	-0.02806	0.15525	0.34066
Amount of non-revolving unsecured debt - 0 months			
Income - £0-£23,000	-0.05944	-0.00741	0.04564
Income - £23,001-£31,000	-0.05310	-0.01057	0.02998
Income - Above £31,000	-0.09464	-0.00026	0.09326
Risk score - Very Poor	-0.04938	-0.00386	0.04312
Risk score - Poor	-0.06873	-0.01603	0.03354
Risk score - Fair	-0.05529	-0.00522	0.04326
Risk score - Good	-0.09658	-0.03198	0.03334
All estimate	-0.02566	-0.00959	0.00623
Amount of non-revolving unsecured debt - 3 months			
Income - £0-£23,000	-0.18025	0.02975	0.26898
Income - £23,001-£31,000	-0.20534	-0.02678	0.14778
Income - Above £31,000	-0.31035	0.05786	0.42360
Risk score - Very Poor	-0.15593	0.00431	0.17619
Risk score - Poor	-0.19242	-0.00230	0.18601
Risk score - Fair	-0.16945	0.01831	0.19071
Risk score - Good	-0.35389	-0.15157	0.05273
All estimate	-0.07472	-0.01590	0.04306

Amount of non-revolving unsecured debt - 6 months			
Income - £0-£23,000	-0.09055	0.13999	0.37678
Income - £23,001-£31,000	-0.13932	0.04319	0.22527
Income - Above £31,000	-0.80027	0.36937	1.53893
Risk score - Very Poor	-0.14690	0.07382	0.29642
Risk score - Poor	-0.19166	0.03345	0.25511
Risk score - Fair	-0.10958	0.11885	0.34777
Risk score - Good	-0.31074	0.01402	0.34242
All estimate	-0.02545	0.06434	0.15297
Amount of non-revolving unsecured debt - 12 months			
Income - £0-£23,000	-0.34240	-0.02439	0.29619
Income - £23,001-£31,000	-0.06912	0.17361	0.41684
Income - Above £31,000	-0.44801	0.04754	0.54249
Risk score - Very Poor	-0.17710	0.09242	0.36286
Risk score - Poor	-0.23284	0.08124	0.39542
Risk score - Fair	-0.01022	0.29317	0.59554
Risk score - Good	-0.40971	-0.02419	0.35451
All estimate	0.00874	0.12155	0.23376
Amount of revolving unsecured debt - 0 months			
Income - £0-£23,000	-0.16088	0.06991	0.29294
Income - £23,001-£31,000	-0.12824	0.03346	0.19770
Income - Above £31,000	-0.74062	-0.21550	0.31278
Risk score - Very Poor	-0.18617	0.00910	0.19562
Risk score - Poor	-0.15003	0.06804	0.28672
Risk score - Fair	-0.14086	0.06093	0.27253
Risk score - Good	-0.27910	-0.00468	0.27325
All estimate	-0.03525	0.03305	0.10124
Amount of revolving unsecured debt - 3 months			
Income - £0-£23,000	-0.24495	0.04722	0.34558
Income - £23,001-£31,000	-0.11492	0.08834	0.29348
Income - Above £31,000	-0.48685	-0.08086	0.32821
Risk score - Very Poor	-0.11183	0.11123	0.33196
Risk score - Poor	-0.19916	0.08131	0.36220
Risk score - Fair	-0.15319	0.08561	0.33951
Risk score - Good	-0.36350	-0.07634	0.21485
All estimate	-0.01598	0.07266	0.16157
Amount of revolving unsecured debt - 6 months			
Income - £0-£23,000	-0.34448	-0.03758	0.26217
Income - £23,001-£31,000	-0.18162	0.03804	0.26228
Income - Above £31,000	-1.52127	-0.06554	1.38972
Risk score - Very Poor	-0.21109	0.03485	0.29186
Risk score - Poor	-0.20457	0.09708	0.39094
Risk score - Fair	-0.25572	0.00573	0.27703
Risk score - Good	-0.45153	-0.08982	0.26665
All estimate	-0.06349	0.03237	0.12894
Amount of revolving unsecured debt - 12 months			
Income - £0-£23,000	-0.26484	0.02042	0.30949
Income - £23,001-£31,000	-0.10888	0.12461	0.36192
Income - Above £31,000	-0.91199	2.83768	6.59814
Risk score - Very Poor	-0.09702	0.17194	0.46416
Risk score - Poor	-0.18041	0.16939	0.51398

Risk score - Fair	-0.24850	0.03410	0.31196
Risk score - Good	-0.39564	-0.01239	0.37167
All estimate	-0.00688	0.10409	0.21649

Annex E – Robustness check using different Bayesian modelling assumptions

We conduct a robustness check on whether our Bayesian aggregation results are sensitive to distributional assumptions. This robustness check is motivated by the weak instrument literature. When the instruments are weak, two stage least square is biased towards OLS and the usual standard errors are typically too small (Staiger & Stock, 1994)[63]. Hence, instead of specifying the study estimate the $\widehat{\beta_h}$ as normally distributed, we use a Student's t distribution to capture the fat tail behaviour due to weak instruments. This new model is summarised by the following equations:

$$\begin{pmatrix} \hat{\mu}_h \\ \hat{\tau}_h \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_h \\ \tau_h \end{bmatrix}, \begin{bmatrix} \widehat{se}_\mu^2 & \widehat{covar}_{\mu,\tau} \\ \widehat{covar}_{\mu,\tau} & \widehat{se}_\sigma^2 \end{bmatrix}_h\right)$$

$$\mu_h \sim N(\mu, \sigma_\mu^2)\ \forall h$$

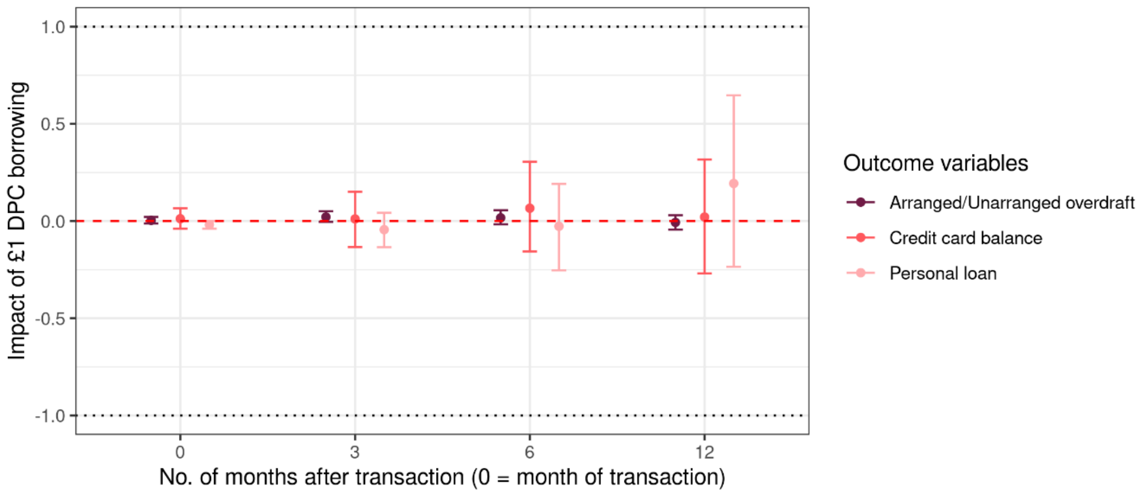
$$\tau_h \sim N(\tau, \sigma_\tau^2)\ \forall h$$

- μ: Reduced form parameter
- τ: First stage parameter
- $\widehat{se}_\mu^2, \widehat{se}_\sigma^2$: The standard error of the reduced form and first stage biased-corrected RDD estimates
- $\widehat{covar}_{\mu,\tau}$: The adjusted correlation between the two estimates. We derive this by estimating a seemingly unrelated regression with the first stage and reduced form RDD. However, this yields the correlation of the *conventional* RDD estimates. We adjust this by first dividing by the standard errors of the conventional RDD estimates, then dividing by the standard error of the reduced form and first stage *biased-corrected* RDD estimates

Figure 20: Effect of DPC borrowing on unsecured debt (aggregation with T-distribution)

[63] Staiger, D. O., & Stock, J. H. (1994). Instrumental variables regression with weak instruments.

Figure 21: Effect of DPC borrowing on components of unsecured debt (aggregation with T-distribution)



Note: Figure 21 show two examples of aggregation results assuming a T-distribution. We note that, unlike the other aggregations in this paper, these aggregations exclude some thresholds that came into effect later in the period our DPC data covers. However, we consider it unlikely that these exclusions are affecting our conclusions.

Annex F – Bayesian aggregation with weak instruments

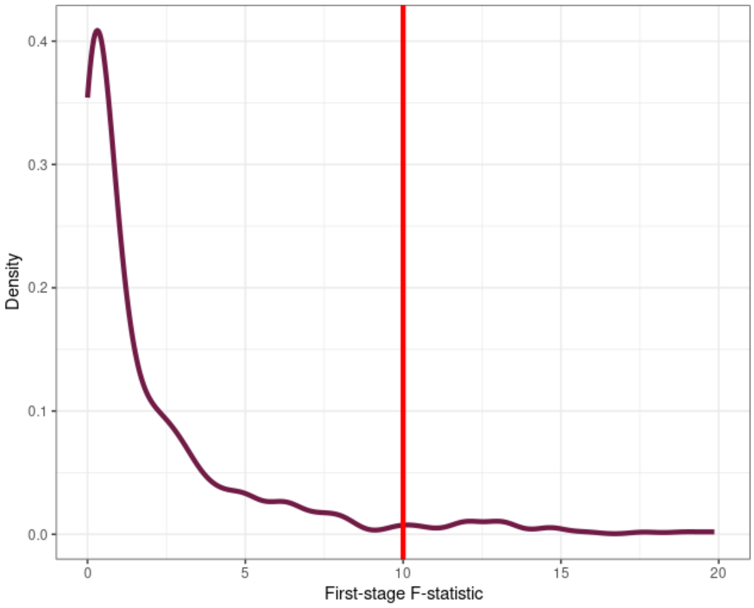
One concern with the standard approach to implementing Bayesian hierarchical aggregation is that the distributional assumptions made are not appropriate in the context of weak instruments. Specifically, the standard approach assumes that study level estimates are normally distributed. This assumption is usually justified on the basis of large sample asymptotic results that indicate that the study level parameters of interest are normally distributed. But this reasoning doesn't hold when instruments are weak. In that case, study level estimates and their standard errors will tend to be biased. Failing to take this into account in the aggregation could lead to inaccurate results.

The theory about how study level estimates will be distributed under weak instrument asymptotics is set out in Staiger and Stock (1997) and Montiel-Olea and Pfleuger (2013), the former are assuming spherical errors and the latter allowing for heteroskedasticity and clustering. In this Annex, we set out the extent to which our study level analyses suffer from a week instrument problem, describe the theory, set out some simulations to demonstrate that it effectively captures the biases when instruments are weak, and finally provide aggregation estimates appropriate in the weak instrument context.

Study level estimates: instrument strength

Figure 22 below plots the distribution of first stage F statistics across all thresholds for a single outcome, revolving debt in 12 months (distributions for other outcomes are almost identical since the first stage is basically the same). This graph makes clear that we have a significant weak instruments problem: the informal rule of thumb that instruments are strong only if the F statistic is above 10 is failed by the vast majority (93%) of the thresholds we use. Over half the thresholds have an F statistic of below 1. The potential, therefore, for biases in our estimates is substantial.

Figure 22: Distribution of first stage F statistics across all thresholds



Notes: distribution for revolving debt in 12 months. x-axis truncated at an F statistic of 20 (2.8% of F statistics are above this level)

Previous work has indicated that Bayesian aggregation can be fairly effective at extracting information from precise study level estimates and ignoring imprecise ones. This is important in our context because precise study level estimates will tend to occur when the instrument is strong (or stronger) and therefore less subject to biases, and imprecise ones when the instrument is weak. But it is important to establish this for our context and ensure that biases in point estimates and standard errors do not carry over to aggregated results.

Asymptotic distributions under weak instruments: theory

The model considered by Staiger and Stock (1997) is as follows.

$$\begin{aligned} y &= Y\beta + X\gamma + u \\ Y &= Z\Pi + X\Phi + V \end{aligned}$$

where y is a vector of outcomes $(\mathrm{T} \times 1)$, Y is a matrix of endogenous variables (excl. outcome) $(\mathrm{T} \times \mathrm{n})$, X is a matrix of exogenous regressors incl. any constant $(\mathrm{T} \times \mathrm{K}_1)$, Z is a matrix of instruments $(\mathrm{T} \times \mathrm{K}_2)$, u is a vector of errors in outcome equation $(\mathrm{T} \times 1)$, and V is a matrix of errors in first stage equations $(\mathrm{T} \times 1)$. In this, T is the number of observations, n is the number of endogenous variables, K_1 is the number of endogenous regressors and K_2 is the number of instruments.

In our setting, we have one endogenous variable $(n = 1)$ and can partial out any exogenous covariates so that we can run as if there were none of these. In this situation, we can write

$$\hat{\beta}_{2SLS} = (Y'P_Z Y)^{-1}(Y'P_Z\, y)$$

where

$$P_Z = Z(Z'Z)^{-1}Z'$$

The key assumption for weak instrument asymptotics is that the first-stage parameters are local to zero:

$$\Pi = \Pi_T = \frac{C}{\sqrt{T}}, \qquad C \text{ fixed}$$

In this case, simplifying equation (2.5) in the paper gives 2SLS bias as

$$\hat{\beta}_{2SLS} - \beta_0 = (Y'P_Z\, Y)^{-1}(Y'P_Z u) \xrightarrow{d} \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}} v_1^{-1} v_2 \sim \hat{\beta}^*_{2SLS}$$

where

$$\begin{aligned} \sigma_{uu} &= E[u_t^2] \quad \text{(scalar)} \\ \sigma_{vv} &= E[v_t^2] \quad \text{(scalar)} \\ \lambda &= \Omega^{1/2}\, C\, \sigma_{vv}^{-1/2} \quad (K_2 \times 1) \\ C &= \Pi\sqrt{T} \\ C & \quad \text{fixed} \\ v_1 &= (\lambda + z_v)'(\lambda + z_v) \quad \text{(scalar)} \\ v_2 &= (\lambda + z_v)'z_u \quad \text{(scalar)} \\ \Omega &= Q_{ZZ} = E[Z_t\, Z_t'] \quad (K_2 \times K_2) \\ \begin{pmatrix} z_u \\ z_v \end{pmatrix} &\sim N_{2K_2}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \otimes I_{K_2}\right) \end{aligned}$$

The paper states, "This expresses $\hat{\beta}_{2SLS} - \beta_0$ as, asymptotically, the ratio of quadratic forms in the $K_2 \times 1$ jointly normal random variables, z_u and z_v." This limiting distribution can be expressed as the following random mixture of normal:

$$\int N(\beta_0 + \theta\, m(z\_v), \text{Var}(z_v))\, dF(z_v)$$

where

$$\begin{aligned} \theta &= \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\,\rho, \quad \text{i.e. the bias of OLS} \\ m(z_v) &= (\lambda + z_v)'z_v \,/\, (\lambda + z_v)'(\lambda + z_v) \\ \text{Var}(z_v) &= \tau^2 \,/\, (\lambda + z_v)'(\lambda + z_v) \\ \tau &= [(1 - \rho^2)\sigma_{uu}\,\sigma_{vv}^{-1}]^{1/2} \\ \rho &= \sigma_{vv}^{1/2\prime}\,\sigma_{vu}\,\sigma_{uu}^{-1/2} \end{aligned}$$

This is the limiting distribution we assume below for aggregating the weak instrument estimates. Montiel-Olea and Pfleuger (2013) extend these results to the robust case, i.e. allowing for nonspherical errors. The model they consider is:

$$\begin{aligned} y &= Z\Pi\beta + v_1 \\ Y &= Z\Pi + v_2 \end{aligned}$$

where y is a vector of outcomes $(\mathrm{S} \times 1)$, Y is a vector of a single endogenous regressor $(\mathrm{S} \times 1)$, Z is a matrix of instruments $(\mathrm{S} \times \mathrm{K})$, v_1 is a vector of reduced form errors $(\mathrm{S} \times 1)$, and v_2 is a vector of first-stage errors $(\mathrm{S} \times 1)$. In this, S is the number of observations

and K is the number of instruments. Note: in this, a single endogenous regressor is assumed, the matrix of instruments has been orthonormalised, $Z'Z/S = I_K$ and, if there are exogenous Xs, all variables have been replaced by projection errors.
Under a local to zero assumption, $\Pi = \Pi_S = C/\sqrt{S}$, Lemma 1 in the paper states

$$\hat{\beta}_{2SLS} - \beta_0 = (\gamma_2'\gamma_2)^{-1}\gamma_2'\gamma_1 - \beta_0 \xrightarrow{d} (\gamma_2'\gamma_2)^{-1}\gamma_2'(\gamma_1 - \gamma_2\beta_0) \sim \hat{\beta}^*_{2SLS}$$

where

$$\begin{aligned} \begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix} &\sim N_{2K}\left(\begin{pmatrix}\beta C\\ C\end{pmatrix}, W\right) \\ C &= \Pi\sqrt{S} \\ \begin{pmatrix}Z'v_1\\ Z'v_2\end{pmatrix} &\xrightarrow{d} N_{2K}(0, W) \end{aligned}$$

and where γ_1 and γ_2 are $K \times 1$, C is $K \times 1$, β is scalar coefficient of interest and W is $2K \times 2K$ (allowing for a general \$W\$ matrix is what makes it robust).
Following the same logic as above, this implies we have

$$\hat{\beta}_{2SLS} - \beta_0 \sim \hat{\beta}^*_{2SLS} \xrightarrow{d} \int N\big(m(\gamma_2), \mathrm{Var}(\gamma_2)\big) f_{\gamma_2}(\gamma_2) d\gamma_2$$

where

$$\begin{aligned} m(\gamma_2) &= (\gamma_2'\gamma_2)^{-1}\gamma_2'(\beta C + W_{12}W_{22}^{-1}(\gamma_2 - C) - \gamma_2\beta) \\ \mathrm{Var}(\gamma_2) &= (\gamma_2'\gamma_2)^{-1}\gamma_2'[W_{11} - W_{12}W_{22}^{-1}W_{21}]\gamma_2(\gamma_2'\gamma_2)^{-1} \end{aligned}$$

This is the week instrument distribution we would implement for the robust case.

Returning to the Staiger-Stock setup, this means that the hierarchical model we would use for aggregation would be

$$\widehat{\beta_h} \sim F\big(\beta_h, \hat{\theta}_h, \hat{\lambda}_h, \hat{\tau}_h\big)\ \forall\, h$$

$$\beta_h \sim N\big(\beta + \gamma X_{h,}\, \sigma^2\big)\ \forall\, h$$

The second line is unchanged relative to the standard Bayesian aggregation model, but the first line reflects the week instrument distribution having replaced the usual Gaussian assumption and the arguments are as defined above.

Validating weak instruments theory against simulations

To assess whether the weak instrument asymptotic theory is effectively capturing the bias is displayed by 2SLS estimates, in this section we compare the predictions of the theory against simulations.
Our simulations follow the structure of Keane and Neal (2024). Specifically, we have

$$\begin{aligned} y_i &= \beta x_i + u_i \\ x_i &= \pi z_i + v_i \quad \text{where } v_i = \rho u_i + \sqrt{(1-\rho^2)}\eta_i \\ u_i &\sim \text{iid } N(0,1) \\ \eta_i &\sim \text{iid } N(0,1) \\ z_i &\sim \text{iid } N(0,1) \end{aligned}$$

In this model, $\rho \in [-1,1]$ controls the degree of endogeneity, while π determines the strength of the instrument. We normalise $\sigma_u = \sigma_\eta = \sigma_z = 1$ so that the OLS bias is $\rho/(1+\pi^2) \approx \rho$ for small π. We set true $\beta = 0.25$ and allow the strength of the instrument to vary, $\pi = 0.01, 0.05, 0.1, 0.2, 0.4$ For each of these values of π, we generate artificial datasets of size $N = 1{,}000$ and run 2SLS on them, as well as using the theoretical results to calculate the biases according to theory. We then replicate 10,000 times.

Tables 7 and 8 below present the results. One feature of the week instrument asymptotic distribution is that its mean (expected value) – and therefore its variance – doesn't exist. This implies we can't compare the theory and our simulations based on these statistics. Instead, we focus on the median and the median absolute deviation (MAD).

Table 7 is for the point estimate, $\hat{\beta}_{2SLS}$. Each row corresponds to a different value of π. Rows are ordered from week instruments to strong. The 'Median F-stat' column reports the median F statistic in the 10,000 replications for each value of π. The informal rule of thumb is that an instrument is only judged as being strong if the F statistic is above 10. The 'Median empirical bias' column reports the median difference between $\hat{\beta}_{2SLS}$ and the truth; for weak instruments, this bias is considerable. The remaining columns of the table report alternative estimates of the theoretical bias for Staiger-Stock (columns 4-5) and Montiel-Olea-Pflueger (6-7). These theoretical corrections do a fairly good job of matching the empirical bias, though there is evidence that the bias is understated in the MOP case when the instrument is very weak ($\pi = 0.01$).

Table 7: fit of weak instrument asymptotic distribution for point estimate, $\hat{\beta}_{2SLS}$

			Staiger-Stock		Montiel-Olea-Pflueger	
π	Median F-stat	Median empirical bias	Median mode bias	Median median bias	Median mode bias	Median median bias
0.01	0.52	0.661	0.502	0.670	0.301	0.487
0.05	2.51	0.087	0.260	0.153	0.203	0.175
0.1	10.01	0.001	0.100	0.017	0.102	0.020
0.2	39.98	-0.002	0.038	0.006	0.038	0.006
0.4	160.01	0.001	0.012	0.002	0.012	0.002

Table 8 is for the standard error of $\hat{\beta}_{2SLS}$. 2SLS estimates usually report the standard error of the estimate as a measure of its dispersion but, as described above, this quantity doesn't exist under week instrument asymptotics. We use this information to back out and implied median absolute deviation based on the theoretical formula for the normal distribution; this is reported in the column 'Empirical median absolute deviation.' In the column 'Simulated median absolute deviation,' we report this statistic calculated across the 10,000 replications of $\hat{\beta}_{2SLS}$. There is some difference relative to the previous column, but not as significant as with the point estimate. The final two columns report median values for the median absolute deviation under the Staiger-Stock and Montiel-Olea-Pflueger cases. As with the point estimates, these theoretical corrections do a fairly good job of matching the empirical distribution, but Montiel-Olea-Pflueger does less well when instruments are very weak.

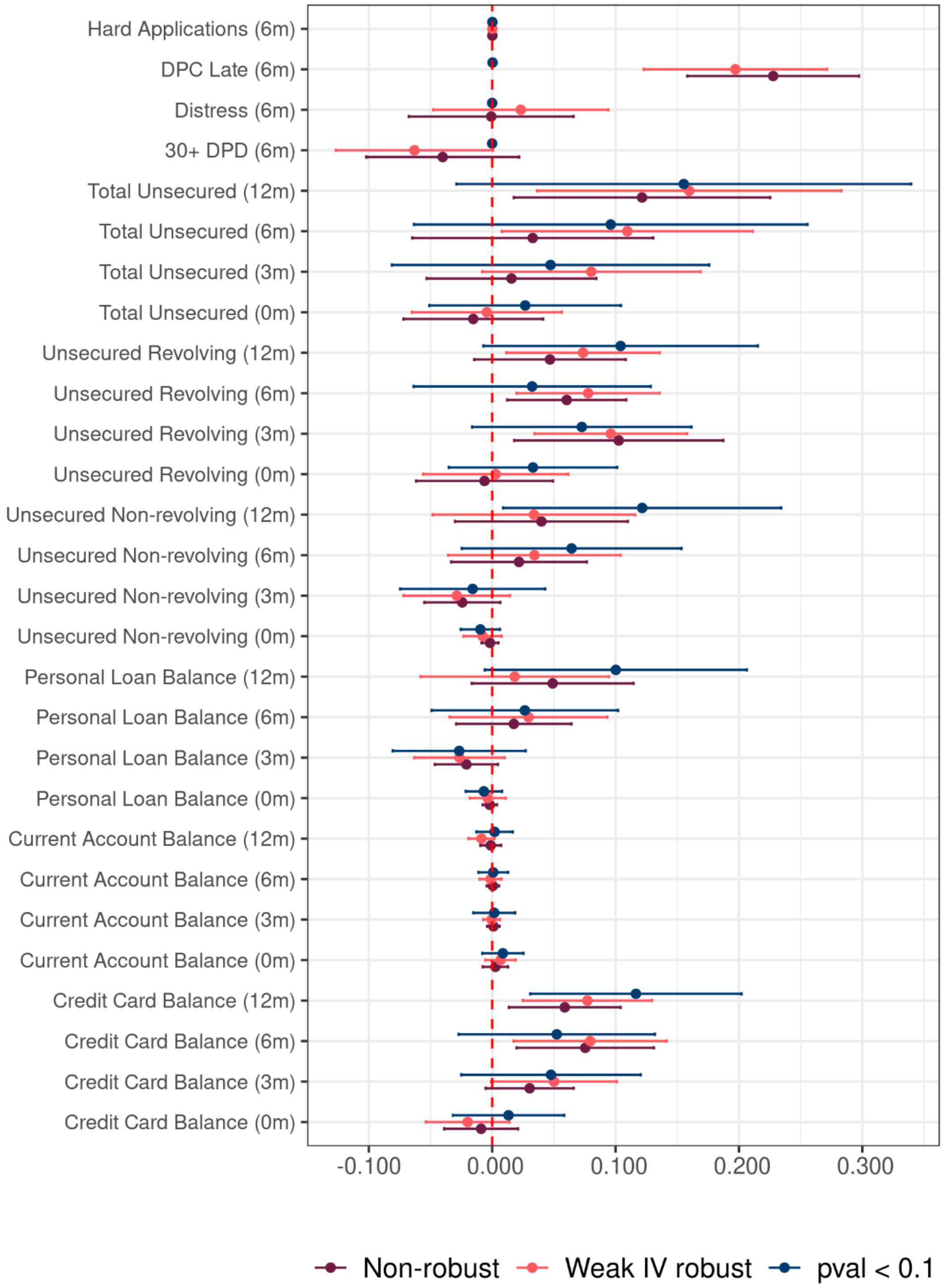
In unreported simulations, we attempted to compare the ability of standard Bayesian aggregation approaches with these alternative weak instrument assumptions in terms of their ability to recover the true parameter values. The week instrument results produced aggregations closer to the true underlying parameters but when instruments were very weak, did not fully correct for the biases. However, in cases with a mixture of weak and strong instruments, both approaches worked reasonably well. As a consequence, our main result above rely on excluding the majority of cases with very weak instruments.

Table 8: fit of weak instrument asymptotic distribution for standard error of $\hat{\beta}_{2SLS}$

				Staiger-Stock	Montiel-Olea-Pflueger
π	Median F-stat	Empirical median absolute deviation	Simulated median absolute deviation	Median median absolute deviation	Median median absolute deviation
0.01	0.52	0.973	1.041	0.912	0.586
0.05	2.51	0.425	0.581	0.572	0.434
0.1	10.01	0.211	0.322	0.316	0.258
0.2	39.98	0.107	0.159	0.159	0.134
0.4	160.01	0.053	0.077	0.077	0.066

Comparing results under weak instrument asymptotics

Figure 23: Comparison of Bayesian aggregation results across our methods accounting for weak instruments



Note: 'pval < 0.1' are the results we use in the main body of the report; 'non-robust' is using all threshold fuzzy RD estimates regardless of the strength of their first stage; 'weak-iv robust' is using the aggregation that assumes a prior distribution that is robust to weak instruments, described above. Convergence for Distress (6m), DPC Late (6m), 30+ DPD (6m), and Current Account Balance (3m) converged poorly in our aggregations, so we interpret them with caution.

Table 9: Comparison of Bayesian aggregation results across our methods accounting for weak instruments

Variable	2.5% Percentile	Mean	97.50% Percentile
Amount of Arranged/Unarranged overdraft - 0 months			
Non-robust	-0.00779	0.00239	0.01257
Report (p<0.1)	-0.00808	0.00864	0.02535
Weak IV robust	-0.00547	0.00672	0.0189
Amount of Arranged/Unarranged overdraft - 12 months			
Non-robust	-0.00936	-0.00116	0.00704
Report (p<0.1)	-0.01279	0.00196	0.01671
Weak IV robust	-0.01921	-0.00866	0.00188
Amount of Arranged/Unarranged overdraft - 3 months			
Non-robust	-0.00413	0.00108	0.00629
Report (p<0.1)	-0.01532	0.00164	0.01859
Weak IV robust	-0.00735	-0.00049	0.00636
Amount of Arranged/Unarranged overdraft - 6 months			
Non-robust	-0.00525	0.00116	0.00756
Report (p<0.1)	-0.01117	0.00082	0.01282
Weak IV robust	-0.0102	-0.00133	0.00754
Amount of Personal loan - 0 months			
Non-robust	-0.00701	-0.00168	0.00366
Report (p<0.1)	-0.02154	-0.00678	0.00798
Weak IV robust	-0.01815	-0.0036	0.01094
Amount of Personal loan - 12 months			
Non-robust	-0.01845	0.04878	0.11601
Report (p<0.1)	-0.00599	0.10019	0.20637
Weak IV robust	-0.05803	0.01826	0.09455
Amount of Personal loan - 3 months			
Non-robust	-0.04679	-0.02065	0.00548
Report (p<0.1)	-0.08065	-0.02674	0.02716
Weak IV robust	-0.06311	-0.02659	0.00993
Amount of Personal loan - 6 months			
Non-robust	-0.02869	0.01738	0.06345
Report (p<0.1)	-0.04915	0.02641	0.10197
Weak IV robust	-0.03444	0.02944	0.09333
Amount of non-revolving unsecured debt - 0 months			
Non-robust	-0.00821	-0.00173	0.00475
Report (p<0.1)	-0.02545	-0.00959	0.00627
Weak IV robust	-0.02328	-0.00781	0.00767
Amount of non-revolving unsecured debt - 12 months			
Non-robust	-0.02997	0.04023	0.11043
Report (p<0.1)	0.00884	0.12155	0.23427
Weak IV robust	-0.04815	0.03383	0.1158
Amount of non-revolving unsecured debt - 3 months			
Non-robust	-0.05368	-0.0246	0.00447
Report (p<0.1)	-0.07468	-0.0159	0.04289
Weak IV robust	-0.07186	-0.0287	0.01446
Amount of non-revolving unsecured debt - 6 months			
Non-robust	-0.03145	0.02129	0.07404
Report (p<0.1)	-0.02468	0.06434	0.15337

Weak IV robust	-0.03563	0.03422	0.10407
Amount of revolving unsecured debt - 0 months			
Non-robust	-0.06092	-0.00565	0.04962
Report (p<0.1)	-0.03522	0.03305	0.10132
Weak IV robust	-0.05601	0.00288	0.06178
Amount of revolving unsecured debt - 12 months			
Non-robust	-0.01081	0.04758	0.10598
Report (p<0.1)	-0.0071	0.10409	0.21528
Weak IV robust	0.01145	0.07361	0.13578
Amount of revolving unsecured debt - 3 months			
Non-robust	0.01706	0.10211	0.18716
Report (p<0.1)	-0.01624	0.07266	0.16156
Weak IV robust	0.03431	0.09605	0.15778
Amount of revolving unsecured debt - 6 months			
Non-robust	0.01339	0.0601	0.10681
Report (p<0.1)	-0.06378	0.03237	0.12853
Weak IV robust	0.01991	0.07779	0.13567
Amount of unsecured debt - 0 months			
Non-robust	-0.07295	-0.0147	0.04354
Report (p<0.1)	-0.05089	0.02676	0.1044
Weak IV robust	-0.06514	-0.0044	0.05633
Amount of unsecured debt - 12 months			
Non-robust	0.01849	0.12333	0.22818
Report (p<0.1)	-0.02902	0.15525	0.33952
Weak IV robust	0.03634	0.1597	0.28307
Amount of unsecured debt - 3 months			
Non-robust	-0.05275	0.01643	0.08561
Report (p<0.1)	-0.08126	0.0473	0.17585
Weak IV robust	-0.00816	0.08034	0.16884
Amount of unsecured debt - 6 months			
Non-robust	-0.06574	0.03237	0.13048
Report (p<0.1)	-0.06344	0.09603	0.25551
Weak IV robust	0.00782	0.10951	0.2112
Any 30 days arrears (excluding DPC) 7-12 months after borrowing			
Non-robust	-0.10094	-0.04124	0.01845
Report (p<0.1)	-0.00015	-0.00004	0.00007
Weak IV robust	-0.12666	-0.06301	0.00064
Any distress indicators 7-12 months after borrowing			
Non-robust	-0.0671	0.00007	0.06723
Report (p<0.1)	-0.00009	0.00001	0.00011
Weak IV robust	-0.04773	0.02318	0.09408
Any late DPC repayment 7-12 months after borrowing			
Non-robust	0.15634	0.22728	0.29821
Report (p<0.1)	0.00016	0.00029	0.00042
Weak IV robust	0.12284	0.19712	0.2714
Credit card balance - 0 months			
Non-robust	-0.03754	-0.00889	0.01975
Report (p<0.1)	-0.03182	0.01322	0.05826
Weak IV robust	-0.05375	-0.01988	0.01399
Credit card balance - 12 months			
Non-robust	0.01494	0.05894	0.10294

	Report (p<0.1)	0.03087	0.11651	0.20216
	Weak IV robust	0.02483	0.07714	0.12945
Credit card balance - 3 months				
	Non-robust	-0.00549	0.0306	0.06668
	Report (p<0.1)	-0.02504	0.04765	0.12034
	Weak IV robust	-0.00084	0.0501	0.10104
Credit card balance - 6 months				
	Non-robust	0.0196	0.07474	0.12988
	Report (p<0.1)	-0.02725	0.05232	0.13189
	Weak IV robust	0.01738	0.07939	0.14141
Total number of 'hard credit searches' 7-12 months after borrowing				
	Non-robust	-0.00003	0.00011	0.00026
	Report (p<0.1)	-0.00013	0.00015	0.00043
	Weak IV robust	-0.00003	0.00014	0.00031

Note: 'pval < 0.1' are the results we use in the main body of the report; 'non-robust' is using all threshold fuzzy RD estimates regardless of the strength of their first stage; 'weak-iv robust' is using the aggregation that assumes a prior distribution that is robust to weak instruments, described above. Convergence for Distress (6m), DPC Late (6m), 30+ DPD (6m), and Current Account Balance (3m) converged poorly in our aggregations, so we interpret them with caution.