Optimizing Policy Design—Evidence from a Large-scale Staged Digital Coupon Program in the Field*

Qiao Liu Qiaowei Shen Tenghui Wang Zhenhua Li Shu Chen

Abstract

Using iterative experiments on digital payment platform to uncover causal links between critical policy details and outcomes helps to optimize policy design. This paper studies a large-scale staged fiscal stimulus program conducted during the COVID-19 pandemic, in which a provincial government in China disbursed digital coupons via Alipay to 8.4 million individual accounts in consecutive waves and updated the program design each time. We find that ruling out unproductive program features leads to a pattern of increasing treatment effects over the waves and that program design matters more than the size of the fiscal stimulus in boosting spending. Our results show that (i) general coupons with no constraints on where the vouchers can be redeemed are more effective than specialized coupons in stimulating consumption in the targeted sectors; (ii) coupon packets with fewer denominations and shorter redemption windows tend to be more effective; and (iii) low-income residents and non-local residents are equally or even more responsive to the coupon program than other groups. Our results illustrate that generating variations in iterative policy experiments, combined with a timely assessment of individuals' responses to marginal incentives leveraging digital technology optimizes program design.

JEL Codes: C93, D12, D90, E61, E62, H71

Keywords: Policy experiments; Optimizing policy design; Digital consumption coupon; Fiscal stimulus program; Behavioral interventions

*Qiao Liu is Professor of Finance (qiao_liu@gsm.pku.edu.cn), Qiaowei Shen is Professor of Marketing (qshen@gsm.pku.edu.cn) at Guanghua School of Management, Peking University. Tenghui Wang is Research Assistant Professor of Finance (tenghuiwang@ln.edu.hk) at Faculty of Business, Lingnan University. Both Liu and Shen acknowledge financial support from the China Education Ministry's Chang Jiang Scholars Program. Shen also acknowledges financial support from the National Natural Science Foundation of China (71722013). Zhenhua Li and Shu Chen are researchers from the Research Institute of Ant Financial.

1 Introduction

Optimizing program details leads to maximizing the policy outcome, while it is not easy since the treatment effects vary with policy designs. Often, governments lack knowledge about the causal links between program details and policy outcomes, leading to unproductive policy features being built into programs. To address this, policymakers can leverage iterative policy experiments to exogenously generate variations in policy, evaluate individuals' responses to marginal incentives, and hence optimize policy design (see, e.g., **???**). However, evidence of governments using iterative experiments to evaluate multifaceted programs and optimize design remains limited, and this is due mainly to the lack of such large-scale iteratively designed programs in the real world. Moreover, it is challenging to track the timely policy outcomes, assess the treatment effects and give the policymakers feedback to optimize the policy design further.

Various coupon programs have been implemented by governments to stimulate consumption during the COVID-19 pandemic (**??**). Notably, cooperating with digital platform companies, many local governments in China have experimented with digital coupons featuring small face value and moderate government subsidy (e.g. "Spend RMB 50 and get RMB 10 off" where the saving is from government subsidy) as a policy tool to revive local retail. Leveraging digital mobile payment platforms, this innovative tool enables the simultaneous distribution and tracking of millions of coupon packets, and conducting timely assessment, further facilitating governments to optimize policy design in later round. These programs varied in coupon design and resulted in varying degree of success in promoting local consumption.

In this paper, we study a large scale staged digital coupon program implemented by a provincial government in China, Guangxi, during the COVID-19 pandemic. This program spanned all the 14 prefecture-level municipal cities within Guangxi province and involved several rounds of coupon distribution within a few months of 2020. For each round, 2 to 3 million coupon packets were distributed through Alipay on a on a "first-come-first-serve" basis. At the every beginning, Guangxi issued *specialized coupons* with designated use in a specific industry or category on March 26, 2020, the same time as Hangzhou that issued *general coupons* involving no constraints on where the vouchers could be redeemed. By studying the three waves of digital coupon issuance in Hangzhou, which the capital city of Zhejiang province with more than 10 million residents, **?** find that an effective government subsidy of RMB 1 can drive excess spending of RMB 3.4 to RMB 5.8, and that the effect persists across multiple coupon issuance waves. The primary goal of Guangxi government

is outcome-oriented and really simple: stimulate consumption as much as possible and revive local retail. Having perceived the conclusions of Hangzhou research and the different design between Guangxi and Hangzhou, Guangxi government chose to cooperate with Alipay and our research team to discuss the coupon design, leverage the platform's technological advantages, and optimize the policy design round by round.

Having solved the technology issue by leveraging the digital payment platform, an important question facing government is how to optimize coupon design to improve marginal propensity to consume (MPC) since the design of stimulus policies matters significantly for MPC improvement (**?**). Several hypothesis or debate about coupon design have emerged. The first is about the use of general versus specialized coupon. The prevailing view is that specialized coupon can better support the target sectors than the general coupons. For example, many local governments have distributed specialized coupons for restaurants, travel or entertainment during the pandemic. The specialized usage, however, may hurt the coupon redemption rate. General coupons, on the other hand, do not impose usage restrictions and are likely to be associated with higher redemption rate. Therefore, when the policy goal is to promote the overall consumption, it is not clear that the specialized coupon targeting specific retail categories can outperform the general coupon. The second is related to the features of coupon packets. Specifically, whether increasing the number and the variety of vouchers (e.g., different face value) and lengthening the redemption period would enhance the effectiveness of coupon? In addition, when the initial rounds of coupon packets involve large government subsidy, would a downsized government subsidy still be effective in driving consumption? This is a relevant issue as local governments are oftentimes financially constrained.

Specifically, we study three consecutive waves of digital coupons issued in Guangxi autonomous region, with three key features that make them well-suited to answering our research questions. The first wave of coupon disbursement included specialized coupons in eight denominations, designated for specific sectors and valid for two weeks. The second wave featured general coupons in four denominations valid for one week, with the same amount of government subsidy per coupon packet as the first wave. The third wave maintained the general coupon design but reduced the number of voucher denominations and the government subsidy per packet. The incremental changes in design allow us to compare treatment effects between general and specialized coupons and between more complex and simpler designs, as well as to assess the relevance and value of critical policy details. Further, the coupon disbursement through the mobile payment platform Alipay enabled us to track coupon disbursement and redemption and promptly evaluate the treatment effect.

For each wave of coupon distribution, we randomly selected 200,000 coupon recipients as the treatment group, and another 200,000 individuals who attempted to claim the coupons but failed due to quantity limit as the control group. Leveraging the transaction level data on the digital platform before and after each coupon disbursement wave, we employ the difference-in-differences approach to identify the causal impact of coupon acquisition on consumption under different coupon designs. Together with information on redemption rate in each wave, we evaluate the effects of coupon design on the effectiveness of coupon program. We are also interested in the heterogeneous effect of coupon acquisition across consumer groups.

Admittedly, this coupon program is not a randomized control trial as consumers are not randomly assigned into either the coupon or non-coupon condition. While all the local residents are eligible to participate, only those who claimed the coupon packets on the platform before they ran out received the coupon. Therefore, we need to address the self selection issue. Furthermore, several features of a coupon packet were altered simultaneously from one round to the next given the feedback from previous rounds, rather than changing one aspect at a time as would be ideal in an experiment setting. This necessitates cautious interpretation of the coupon design results. Nevertheless, empirical evidence of such large-scale, iteratively designed coupon program by the government is scarce. The granularity of the data offers invaluable opportunity to explore the effects and provides insights into optimal coupon design.

Several aspects of the results are remarkable. First, we find that the digital coupon is highly effective in driving additional spending, consistent with the findings in previous research (**???**). Moreover, there is a pattern of increasing policy effectiveness over time. These improvements are notable given that they arose from simple changes in program design.[1]

Second, we find that general coupons with no constraints are more effective than the sector-specific coupons in stimulating consumption in the targeted sectors. We compare the results in the first wave, when sector-specific coupons were disbursed to support spending on food and dining, shopping, tourism, and transportation, with the results in the second wave, when the general coupons were issued. The results indicate that the general coupons stimulated more spending in all the categories than the sector-specific coupons did, even in the categories that the latter were designed to stimulate. Thus, simplifying the set of options given to coupon claimants can

[1]We interpret the treatment effects based on traditional (unadjusted) p-values throughout the research. Furthermore, considering that there is more than one wave of survey we analyse and heterogeneous test of treatment effect, we adjust the p-values of treatment effects using **?** method where False Discovery Rate (FDR) is controlled when necessary. By doing so, we conduct multiple hupothesis testing as robustness test. Overall, our findings are robust to the adjustments.

potentially improve their decision-making and lead to better policy outcomes than a complex set of coupon options.

Third, we find that coupon packets including vouchers with fewer denominations (e.g., three in the third wave) and shorter valid redemption times (e.g., 7 days in the second and the third waves) stimulated more spending than coupons with more denominations (e.g., four in the second wave and eight in the first wave) and longer periods of validity (e.g., 14 days in the first wave). Therefore, ruling out unproductive features (e.g., complexity in coupon design, more denominations, and longer windows of validity) significantly improves policy effectiveness.

Fourth, compared with program design, the size of the government subsidy plays only a secondary role in stimulating consumption. We find that the third wave of coupons was more effective than the second wave despite the size of the government subsidy being far smaller during the third wave (RMB 66 per coupon packet) than the second wave (RMB 268 per coupon packet). These findings illustrate that by deliberately choosing an optimized policy design, the government can achieve a more desirable policy outcome at a much lower budget than is possible with a less effective design.

Fifth, the effectiveness of the consumption coupon as a policy tool was highly heterogeneous across consumer segments. We find that low-income individuals were more responsive to the coupon program, and that non-local residents, who were eligible to participate in the program, responded to the coupon in the same way as local residents.

Despite uncertainties of the underlying economic and psychological processes, some behavioral models aligns closely with our results. One is the mental accounting model (**??**). Consumers may treat the digital coupon as a separate account from their regular weekly budget and therefore exhibit an increased propensity to consume when receiving the digital coupon packet. Furthermore, the findings that general coupons perform much better than the segment-specific coupons and that fewer voucher denominations work better than more vouchers may arise because complexity causes suboptimal choices. Psychological mechanisms such as reciprocity (**?**) may explain why the residents from low-income cities and people with lower income level showed a higher level of responsiveness to the coupon program. And since the program emphasized that non-residents were eligible for the coupons, the salience of their identity may have been primed (**?**), which then may have generated equally or even more positive responses to the coupon program among non-residents. Despite the theories built into the coupon program being implicit, the iterative experimental approach taken by the local government has pragmatic value in terms of improving

the policy outcomes that policymakers can achieve.

Our results have concrete policy importance. The literature shows that one-off blanket fiscal stimulus payments in the form of cash payments or tax rebates typically yield a MPC in a range from 0.1 to 0.8. In a cost and benefit framework, such programs may fail to achieve an outcome at an optimized level because of certain unproductive program features. Our results show that by proactively gathering policy-relevant information through iterative policy experimentations, policymakers can optimize policy design and achieve improved results. This paper also shows that understanding the psychological process can play an important role in designing effective behavioral interventions.

This paper builds on several related literatures. It contributes to the literature showing the value and limitations of experimental methods for informing policy (**???**). **?** point out that experiments unavoidably face design challenges because policymakers usually do not have great knowledge about potential policy outcomes and the causal mechanisms behind crucial policy details and outcomes. Often, as a result, policy experiments are not optimally designed and implemented. These authors suggest using experiments to test the causal mechanism underlying a policy. Our paper offers a novel case study that policy makers break what would ordinarily be a one-off, large-scale program into several staged programs, and leverage information and knowledge from experiments to obtain more accurate and robust prescriptions for optimal policies.

In addition, we add to this body of work by finding that digital platform can facilitate government to evaluate policy outcomes timely and optimize design accordingly. Coupons have long been utilized as an effective marketing tool to boost sales (**??**). Extensive research has examined the effects of various types and designs of coupon issued by brands or retailers to promote specific products or product lines (**?????**), and we add to the studies on the limited designs of the government-issued digital coupons. Moreover, digital platform's low marginal cost of administration and outstanding data and technology-based precise target ability contribute to the research emphasizing the advantages of FinTech adoption and application in finance (e.g., **?????**). We focus on its advantages of helping government design and distribute digital coupons, track the usage and and consume automatically, and assess the effects timely, and the most importantly, it can further help optimize policy design.

Finally, this paper contributes to a broad literature that studies consumers' responses to fiscal stimulus payments (**?????**) as well as the ongoing research on the effects of fiscal stimulus programs in a variety of economies during the COVID-19 era (**????**), and in particular, the studies on

the digital coupon program in China (**???**). Different from the previous studies which focus on evaluating the policy effect of the coupon programs on consumers and small firms, this paper focuses on policy design and offers insights about how to improve policy effect through improved coupon design.

The remainder of the paper is organized as follows. Section 2 overviews the coupon program implemented by the Guangxi government during the pandemic. Section 3 describes our data, econometric approach and the estimated effects of three waves of coupon disbursement. Section 4 analyzes the features of coupon design and the improvement in program effectiveness. We document the heterogeneity across consumer groups and cities in Section 5 and conclude the paper in Section 6.

2 Experiment Overview

The outbreak of the COVID-19 pandemic in early 2020 led to a 19% reduction in retail sales of consumer goods and services in China in the first quarter of 2020 compared with the first quarter of 2019.[2] **?** document an average decline of 32% using data on consumer spending transactions and find that dining, entertainment, and travel-related categories were the most affected sectors. How to revive the retail sector and boost the economy became a critical question facing policymakers in China. Beginning in March 2020, many local governments in China rolled out digital coupon programs as a stimulus policy to spur local consumption.

Guangxi was one of the first regions to launch the digital coupon program. From late March to June 2020, five waves of digital coupons were disbursed. Guangxi is a province-equivalent autonomous region located in southwest China, with 14 prefecture-level municipal cities and 50.1 million residents. The per capita GDP in Guangxi was approximately RMB 43,000 ($6,200) in 2019, lower than the national average. In contrast to many coupon programs that ran at the city level, Guangxi's program was implemented at the provincial level and covered all 14 of its municipal cities. Thus, it offers a unique opportunity to investigate the heterogeneous policy effects across cities.

The five waves of coupon disbursements were distributed through the third-party platform Alipay, which is the dominant mobile payment platform in China.[3] All residents, including visitors,

[2] Report from National Bureau of Statistics, China.

[3] A survey by PricewaterhouseCoopers shows that the mobile payment penetration rate in China was 86% in 2019, the highest in the world; see https://www.pwc.com/gx/en/consumer-markets/consumer-insights-survey/2019/report.pdf.

in Guangxi's 14 cities were eligible to claim the coupons. When the digital coupons were released, one can claim a coupon packet by logging onto the Alipay mobile app (**Figure 1** shows a screenshot of the coupon acquisition page from the first wave). The coupon distribution occurred on a "first-come-first-serve" basis and was limited to one packet per person.[4] Once successfully claimed, a coupon packet would appear in one's Alipay account; it included multiple vouchers, all immediately redeemable at participating merchants. The digital vouchers would be automatically applied to qualified transactions through Alipay, with the restriction that only one voucher could be used per transaction.

[Insert Figure 1: Screenshots of the Coupon Acquisition Page and Redemption Receipt]

The design of the coupon packets, including the number of vouchers, the denominations, whether a voucher was for a designated or general usage, and the valid redemption time, varied across waves. One common feature across waves was that all vouchers were in the form of savings or a discount for a specified amount of spending, e.g., "spend RMB 20, get RMB 4 off," where the RMB 4 deduction was financed by a government subsidy. The details of the coupon packets in each wave were released to the public one or two days before that wave. **Table 1** summarizes the coupon design, along with the overall redemption rate in each wave. The vouchers were redeemable at many local merchants as well as online sellers.[5] Although we do not have the exact numbers of participating merchants in each city of each program wave, the platform provides an index as an approximation, for which we control in the later analysis.

[Insert Table 1: Summary of Guangxi Coupon Program]

The first wave of digital coupon disbursement spanned a two-week period from March 26 to April 8, 2020 and involved the distribution of 2.7 million coupon packets. During the two-week period, a varying number of coupon packets was released at 12 am each day,[6] with the most coupons per day released on the first day. A coupon packet remained valid for 14 days starting from the claim date. As shown in **Table 1**, each coupon packet included 50 vouchers of five types: dining (22 vouchers), shopping (13 vouchers), travel (3 vouchers), transport (10 vouchers), and general coupons (2 vouchers). The total government subsidy for each coupon packet was RMB 268.

[4] As we explain, the coupon packets in each wave were distributed across multiple days. Individuals who successfully claimed a coupon packet were eligible to claim the coupon in the same wave again only if they had redeemed all the vouchers in the received packet. We exclude the repeated claimants within a wave from our analysis. Individuals could claim coupons in a new wave regardless of their redemption activity in the previous wave(s).

[5] All the merchants with operations located in Guangxi were eligible to participate in the program. Additional store promotions offered by merchants could be applied on top of the government coupons.

[6] April 4, which is Tomb Sweeping Day in China, was excluded.

The coupons in the first wave were predominantly specialized coupons, which could only be applied to qualified transactions in the designated categories. Only two vouchers were for general use, and they had small face values (e.g., "spend RMB 5, get RMB 1 off"). The presumption was that specialized coupons would be more effective in driving spending in the corresponding sector. Another feature of the first-wave coupon packet was that it involved many denominations. For example, for dining coupons only, there were two denominations: "spend RMB 20, get RMB 4 off" and "spend RMB 100, get RMB 10 off." The coupons in the first wave involved eight different denominations in total.

Although all 2.7 million coupon packets were claimed, the redemption rate was low. Only 9.27% of the vouchers were used during the redemption period. The lower-than-expected performance of the first wave of the coupon program led to substantial changes in the coupon design in the second wave. The most significant change was a switch from specialized to general coupons, with no constraints regarding the spending category to which they could be applied. In addition, the denominations of the vouchers in a coupon packet were streamlined, being reduced from eight to four denominations. In the second wave, the total government subsidy remained at RMB 268 per coupon packet, as for the first wave.

The second-wave coupons were disbursed through the same platform during a two-week period from April 9 to April 22, 2020 at 12 am each day. In this wave, 2.4 million coupon packets were released. Each packet contained 27 general coupons with four denominations, "spend RMB 100, get RMB 20 off" (3 vouchers), "spend RMB 60, get RMB 15 off" (10 vouchers), "spend RMB 25, get RMB 5 off" (11 vouchers) and "spend RMB 5, get RMB 1 off" (3 vouchers). The vouchers could be applied to any transaction, conditional on meeting the minimum spending amount required. A coupon packet remained valid for 7 days upon claim (rather than the 14 days in the first wave). The coupon redemption rate increased substantially during this wave compared with the first wave, rising from 9.27% to 36.81%.

The coupon features were further modified in the third wave of disbursement, which occurred from April 23 to May 10, 2020. There were two noticeable changes compared with the second wave: (1) the coupon denominations were reduced to three; and (2) the number of vouchers was reduced to eight. The total number of coupon packets distributed in this wave was increased to 3.3 million, and the government subsidy was reduced to RMB 66 per packet. The overall redemption rate for the third wave was 63.32%.

The fourth wave was rolled out between May 11 and May 20, 2020. Only four general vouchers

with a government subsidy of RMB 16 were included in the coupon packet. A total of 2.44 million packets were distributed, and a redemption rate of 54.34% was achieved. The fifth wave coupon disbursement, from May 28 to June 30, 2020, had a similar design to the third wave, with a smaller government subsidy of RMB 33 per packet. The overall redemption rate was 56.19%.

Our analysis below focuses on the first three waves of the coupon distribution, for which we have access to de-identified individual-level data on coupon acquisition and redemption. Another reason for our focus is that the first three waves involved the most substantial changes in program design.

3 Analysis of Waves I–III of the Coupon Program

The first objective of this paper is to examine whether the iterative experiments of digital coupon issuance, which enabled deliberate variations in policy design, led to improvements in policy effectiveness over time. As the policy goal was to spur consumption, we measure the effectiveness of the program by the extent of the increase in consumers' unplanned spending induced by the coupons in comparison with the spending of the control group.

3.1 Empirical Strategy

To evaluate the effectiveness of the digital coupon program in each wave, we employ a difference-in-differences (DID) approach, comparing the spending between those who successfully acquired the coupons (the treatment group) with that of the consumers who failed to obtain the coupons (the control group) during both the coupon redemption period and the period before it. The identification challenge lies in the fact that the coupon packets were not randomly assigned, and therefore the group of consumers that successfully obtained the coupons could be different from those who did not. To alleviate any potential self-selection bias, for each wave, we construct the treatment group as the coupon recipients on the first day of that wave, and the control group as those who logged onto the coupon claim page on the same day but failed to acquire the coupon packets due to limited supply and did not obtain any coupon packets during that wave.[7] These latter consumers are presumably more similar to the coupon recipients than those who did not attempt to participate in the coupon program.

[7]This group of individuals is identifiable because all visits to the coupon claim page and activities on the platform can be tracked, which is a benefit of the digital disbursement of coupons.

We estimate the average treatment effect of coupon acquisition in each wave using the following regression model:

$$y_{it} = \alpha_0 + \alpha_1 Treat_i + \alpha_2 Post_t + \alpha_3 Treat_i \times Post_t + \gamma X_i + \varepsilon_{it} \quad (1)$$

The dependent variable is individual i's total spending in period t, y_{it}. The independent variables include the treatment indicator $Treat_i$, which equals 1 if individual i is in the treatment group in a given wave and 0 otherwise; the time period dummy $Post_t$, which equals 1 if the period is the coupon redemption period and 0 if it is the pre-treatment period; and the interaction of these two independent variables. Taking the first wave as an example, the pre-treatment (post-treatment) period is the two weeks before (commencing from) March 26, 2020. We control for individual characteristics X_i, including age, gender, and the baseline consumption level (the average monthly spending over the six months before the coupon program), as well as city fixed effects.

The key parameter of interest is α_3, which captures the average treatment effect of coupon acquisition on consumption during the redemption period relative to the control group. Because the individuals in the treatment group could redeem varying numbers of vouchers (including zero) after claiming the coupon packets, the size of the effect is influenced by the consumption pattern conditional on coupon usage, as well as by the coupon redemption rate. The parameter identifies the average spending (if any) induced by coupon acquisition compared with the baseline spending in the same period as inferred from the control group.

In addition to the model above, we estimate a model using logarithmic spending, $ln(y_{ijt})$, as the dependent variable. The estimated parameter indicates the percentage increase in spending of the treatment group induced by the coupon program. As the coupon design differed across waves, the treatment effect is more comparable when using the percentage increase in spending rather than the average spending amount.

3.2 Data Description

All three waves of coupon packets were distributed via the mobile payment platform Alipay, which allows accurate tracking of coupon claiming and redemption behavior.[8] For each wave of coupon disbursement, we randomly draw 200,000 coupon recipients on the first day of coupon issuance as the treatment group, and another 200,000 individuals from those who attempted to obtain coupons

[8] Alipay allows users to transfer money and make payments for transactions online and offline (through scanning QR codes).

on the same day but failed to do so in that wave to form the control group. Therefore, the sample size is 400,000 persons for each wave.

The sample is anonymized and contains two types of information:[9] first, the anonymous account holder's demographic information, such as gender, age, city, and resident status, i.e., whether the individual is a local resident or not; and second, the account transaction records for our study period, including information on the transaction amount, merchant category, usage of coupons (if any), and whether the transactions occurred online or offline. The transaction record commences in September 2019, six months before the start of the coupon program, to control for the baseline consumption pattern.

Table 2 presents the summary statistics of the treatment and control groups for each wave. The statistics show that although the two groups are largely comparable, the control group generally has lower baseline spending than the treatment group. Therefore, we use a propensity score matching (PSM) method to address this. Specifically, we run a logistic regression using age, gender, the average monthly spending in the last six months, the total spending during early March (March 1-11, 2020) and the city of residence as control variables. We compute the propensity score for each individual and perform nearest-neighbor matching. To reduce the computational burden, the matching exercise is performed on a random sample of the treatment and control group that is half of the original sample size. Some observations in the treatment group that fail to be matched are dropped. The difference between the treatment and control groups is substantially reduced after matching.

[Insert Table 2: Summary Statistics]

Figure 2 plots the average daily spending of the treatment and control groups (the matched sample) for each wave for the period before the coupon distribution date and the following coupon redemption period. For all three waves, we can see that although the spending patterns of the treatment and control groups are almost identical before the coupon issuance date, spending by the treatment group increases substantially relative to the control group after acquiring the coupons, which provides model-free evidence of the effectiveness of the coupons in stimulating consumption.

[Insert Figure 2: The Average Daily Spending of Two Groups (PSM Sample) of Each Wave]

[9] All data used in this paper were stored and all analyses were conducted within Ant Group's intranet.

3.3 Empirical Results for Coupon Program Effectiveness

In **Table 3**, we present the DID estimation results for the first three waves of the digital coupon program, using both the full sample (Table 3A) and the matched sample (Table 3B). Under each wave, we estimate two models, using spending or its logarithm transformation as dependent variables respectively.

We first focus on the results of using spending as the dependent variable in Table 3A. The results indicate that the digital coupon program effectively increased consumers' total spending, as reflected in the positive and significant coefficient estimates of the interaction term $Treat_i \times Post_t$. For Wave I, on average, the individuals in the treatment group spent RMB 122.91 more than those in the control group during a two-week redemption period. The treatment effect was RMB 389.62 in Wave II and RMB 204.06 in Wave III, both during a one-week redemption period.

The results of using $ln(y_{ijt})$ as the dependent variable provide consistent evidence that the coupon program effectively increased unplanned spending. The percentage increase in consumption of the treatment group relative to the control group, as reflected by the estimated coefficient of the interaction term, steadily increased from 36.6% (Wave I) to 117.2% (Wave II) and 206.5% (Wave III). The results indicate that the effectiveness of the coupon program improved over the waves. Table 3B reports the estimates using the PSM matched sample, which are highly consistent with the results from the full sample. The discussion below focuses on the results in Table 3A, where we use the full sample.

To evaluate the cost-effectiveness of the government subsidy, we compute an efficiency ratio, defined as the amount of unplanned spending (the average treatment effect α_3) relative to the realized government subsidy per coupon recipient, following **?**. This measures how effective an RMB 1 government subsidy is in spurring unplanned spending. This measure is analogous to the MPC, which is defined as the proportion of additional income that individuals spend on consumption. We do not use the term MPC in this paper, as the digital coupon program in our context is starkly different from conventional fiscal stimulus payments.

Based on the actual coupon redemption data, we can infer the total government subsidy realized in each wave; then, dividing this by the total number of coupon packets distributed, we can obtain the effective government subsidy per coupon recipient. For Waves I, II, and III, the effective government subsidy per coupon recipient was RMB 24.84, RMB 98.65, and RMB 41.79, respectively. These numbers are affected by both the redemption rate and the amount of subsidy

committed per coupon packet. In Wave I, a low redemption rate leads to a low effective subsidy per person. In Wave III, although the redemption rate is highest among the three waves, the realized government subsidy is modest because the coupon design for this wave involves a lower subsidy per coupon packet than those for the first two waves. Based on the realized government subsidy per coupon recipient, we can compute the efficiency ratio for each wave. For Waves I, II, and III, the ratios are 4.95, 3.95, and 4.88, respectively, as shown at the bottom of **Table 3**.[10] These numbers confirm that the government subsidy contained in the small-value digital coupons is highly effective in driving additional consumption. To evaluate the overall effectiveness of the program, we must consider both the cost-effectiveness ratio and the coupon redemption rate. While the cost-efficiency ratio is highest for Wave I of the three waves, the low redemption rate renders the overall consumption increase small for this wave, as reflected in the average treatment effect reported above. Conversely, Waves II and III are characterized by increasing redemption rates and policy effectiveness.

Finally, we define the implied elasticity of government subsidy-driven consumption as the ratio of the percentage increase in spending to the effective government subsidy rate. In Waves I, II, and III, the government subsidy rates are 14.2% (268/1,890), 22.5% (268/1,190), and 20% (66/330) respectively.[11] As shown in Table 3A, the percentage increases in spending are 36.6%, 117.2%, and 206.5% for the three waves, respectively. Hence, the implied elasticity of consumption driven by the government subsidy is 2.58 for Wave I, 5.20 for Wave II, and 10.33 for Wave III. The increasing pattern indicates that the treatment effect of the coupon program increases over waves.

[Insert Table 3: Regression Results of Coupon Effects]

3.4 Robustness Checks

We perform several analyses to check the robustness of our results and discuss additional issues that arise. One concern is that consumers may use multiple payment methods for daily transactions that are not fully captured by our data. In common with other studies on government stimulus programs, which utilize transaction-level data from one financial institution (e.g., **??**), our transaction-level data only include transactions on Alipay. This is not an issue if consumers' payment habits are

[10] **?** report similar efficiency ratios for the Hangzhou digital coupon program, ranging from 3.4 to 5.8. The redemption rate for Hangzhou digital coupons ranged from 70% to 80% across waves.

[11] Taking Wave I as an example, the face value of all vouchers adds up to RMB 1,890, with the total government subsidy accounting for RMB 268. Therefore, the government subsidy rate is 14.2%, suggesting that the government subsidizes RMB 14.2 of every RMB 100 of spending.

relatively stable, as the identification of the coupon program's effect on spending is based on the DID approach. However, one concern is that coupon recipients may shift their transactions from other payment methods, such as cash or other mobile platforms, to Alipay to redeem their coupons. In other words, the identified treatment effect could be merely a substitution for other payment methods.

To address this concern, we re-run the DID analysis for each wave, excluding individuals with low transaction volumes before the program. The rationale is that low transaction volumes indicate that an individual's primary payment method was not Alipay before the experiment and that payment switching is more likely to occur among this group of consumers. By excluding these individuals in both the treatment and control groups, we focus on the more frequent users of Alipay with less room for substitution. Specifically, we exclude individuals with transaction amounts of less than RMB 100 per month for the six months before the coupon program. The estimation results are reported in **Table 4**. The estimated treatment effects using the restricted sample for Waves I, II, and III are RMB 127.12, RMB 379.57, and RMB 214.09, respectively, results that are highly consistent with those from the unrestricted sample, and the differences are statistically insignificant. This suggests that substitution of Alipay for other payment methods is not the main reason for the additional consumption. While the estimated percentage increase of spending becomes smaller after excluding the individuals with low transaction volumes, our main conclusion—that the effectiveness of the program, as measured by the percentage increase of consumption, improves over the waves—continues to hold.

[Insert Table 4: Restricted Sample Regression (Spending $\geq$ 100)]

Furthermore, we examine whether the penetration rates of Alipay in each city influence the effectiveness of the digital coupon program. Consumers living in cities with high penetration rates of Alipay may be more familiar with coupon acquisition and redemption on this platform, and therefore exhibit higher responsiveness toward the coupon program than consumers in cities with lower penetration rates. To test this hypothesis, we add an interaction term for the penetration rate in each city and the treatment effect to Equation (1), leveraging the variation in penetration rates across cities. The results of each wave are presented in **Table 5**. As shown in Columns (1), (3), and (5), where total spending y_{ijt} is used as the dependent variable, the interaction of the penetration rate and treatment effect is not significant in Waves I and III, and it is negative in Wave II. In Columns (2), (4), and (6), where $ln(y_{ijt})$ is used as the dependent variable, we find that

the interaction of the penetration rate and treatment effect is significantly negative for all three waves. The results in **Table 5** suggest that responsiveness to the digital coupon program does not increase with the penetration rate for mobile payments.

[Insert Table 5: Interaction Effect with Penetration Rate]

4 Optimizing Coupon Design

Generating variations in policy design and comparing changes in the treatment effect allow policymakers to determine effective policy features and update the program at a re-optimized level. The improvement in policy effectiveness over the waves provides preliminary evidence that policymakers used the staged policy experimentations to promptly take in feedback from the field and modify the program design.[12] As mentioned previously, the coupon packet in Wave I involved 50 vouchers of eight denominations, and the coupons were predominantly specialized rather than general in nature. In Wave II, the coupon packet had fewer vouchers and denominations, and restrictions on coupon usage were removed, i.e., there was a switch from specialized to general coupons while retaining the same total government subsidy per packet as in Wave I. In Wave III, the coupon design was further streamlined, with fewer vouchers and denominations offered. Researchers must consider the practical dimensions of optimally designing and implementing policies (**?**). In this section, we examine how varying the features of the coupon design led to variation in the performance of the coupon program to provide insights into leveraging iterative experiments to upgrade program design and improve policy effectiveness.

4.1 Specialized vs. General Coupons

As shown in **Table 3**, suboptimal program features can drive a wedge between ex ante estimates of program benefits and true ex post benefits. The most salient feature of the first wave of the coupon program was the adoption of specialized coupons. The government's intention was to support the sectors most severely affected by the pandemic, including the dining, shopping, travel (including hotels and attractions), and transportation sectors. Government coupon programs with vouchers designated for use in a specific industry or category are common in practice, based on the

[12] Our interview with government officials and experts from Alipay who participated in the design and rollout of the digital coupon program in Guangxi shows that the local government kept track of redemption rates and coupon-induced spending. The data provided by Alipay, which were immediately available to the government, played a critical role in changing policy designers' presumptions and led to them to altering the implementation of the subsequent waves.

presumption that specialized coupons better support target sectors than general coupons. Examples of such programs include the Supplemental Nutrition Assistance Program (SNAP) and the Food Stamp Program in the US. Research indicates that the MPC for food vouchers is much larger than that for cash (**?**), indicating the effectiveness of such voucher programs in stimulating consumption in the targeted sector.

However, there is a paucity of research that directly studies the effectiveness of specialized coupons relative to general coupons. The common belief is that specialized coupons are more effective in driving spending in the designated sectors than general coupons, which have no constraints on where they can be used. The shift of coupon design from specialized to general coupons in the second wave offers an opportunity to test whether this is the case. Our results in the previous section indicate that the general coupons of Wave II were more effective in driving overall spending than the specialized coupons of Wave I. However, it is possible that the specialized coupons performed better than general coupons in driving consumption in specific categories.

To examine this issue, we focus on increased spending in the designated sectors in the first two waves of the coupon program, in which the government subsidy remained constant at RMB 268 per coupon packet. To estimate the category-level spending induced by the program, we first decompose the total expenditure of an individual in a given period into the following categories, based on those that are most relevant and identifiable from the transaction records: dining, food and drinks, clothing, other nondurables, services, travel, and transportation. Three categories, i.e., dining, travel, and transportation, can be directly matched with the sectors targeted by the specialized coupons. We collate purchases of food and drinks, clothing, and other nondurables and categorize them as "shopping" spending. Then, we estimate the effect of coupon acquisition on spending in each category using Equation (1).

Tables 6 and 7 present the analysis at the spending category level for Waves I and II, respectively, using the full sample.[13] In each table, Panel A (B) shows the results using the total (log) spending of a specific category as the dependent variable. The results in Panel A of **Table 6** show that in Wave I, the incremental spending of the treatment group relative to the control group in the dining, travel, and transportation sectors is RMB 35.43, RMB 1.5, and RMB 0.06, respectively, and the incremental spending in the "shopping" category is RMB 9.34. The estimates from Wave II, shown in **Table 7**, indicate that the corresponding incremental spending amounts are RMB 96.69,

[13]The estimation results using the matched sample are quite similar. They are reported in Tables A1 and A2 of the Appendix.

RMB 1.62, RMB 0.06, and RMB 44.13, respectively. The category-level comparison reveals that the general coupons stimulated substantially more additional spending in almost all of the categories to which the specialized coupons applied, particularly the dining and shopping categories. The results in Panel B of both tables further indicate that the percentage spending increases in the dining, food and drink, clothing, and nondurables categories are substantially higher in Wave II than in Wave I. The percentages of spending increases in travel and transportation are negligible in both waves. Overall, the analysis shows that the general coupons outperformed the specialized coupons in almost all categories to which the specialized coupons were targeted.

[Insert Tables 6 & 7: Specialized Coupons vs. General Coupons]

We conjecture that the seemingly counterintuitive results are due to the complexity of coupon design. When individuals confront a menu of vouchers varying both in categories of purchase and denominations, they may find the coupon package less appealing and redeem less than they would when faced with a simpler coupon design. In the first wave of the coupon program, as noted previously, there were 48 specialized coupons in four categories with seven denominations, whereas the second wave involved 27 general coupons with four denominations. Only 9.27% of vouchers were redeemed in the first wave, compared with 36.81% in the second wave. Thus, simplifying the set of options given to individuals can potentially facilitate their decision-making process and, in this case, stimulate greater spending.

Table 8 provides direct evidence that general coupons are more appealing than specialized coupons by presenting the redemption rate of each voucher category. Among all of the vouchers in the coupon packet of the first wave, the two general coupons "Spend RMB 5, get RMB 1 off" had the highest redemption rate, at 57.59%. The specialized coupons had much lower redemption rates, and only the "shopping" vouchers had a redemption rate above 10%.

[Insert Table 8: Redemption Rate by Voucher Type]

The government subsidy rate for the first wave was 14.2%, increasing to 22.5% in the second wave. One argument is that the lower average subsidy rate per voucher in the first wave was the main reason for the different performances of the two waves. We do not agree with this argument for three reasons. First, 22 vouchers in the first wave had a face value of no less than RMB 20 and a subsidy rate of no less than 20% (12 "spend RMB 20, get RMB 4 off" dining vouchers; eight "spend RMB 20, get RMB 5 off" shopping vouchers, and two "spend RMB 20, get RMB 4 off" travel vouchers), which are comparable with the 24 vouchers in the second wave that meet the

same criteria. In other words, the offerings to consumers in Wave I were as good as those in Wave II. Nevertheless, the specialized coupons in the first wave generated much less consumption than the general coupons in the second wave.

Second, we assess the impact of the subsidy rate on the effectiveness of the coupon program by calculating the implied elasticity of consumption driven by government subsidy. As shown in **Table 3**, the implied elasticity, which controls for the influence of the government subsidy rate, is 2.58 for the first wave and much higher, at 5.20, for the second wave. This suggests that the difference in policy outcomes was associated with the program design rather than the subsidy rate.

Third, **?** study similar digital programs, and their results from multiple coupon waves indicate that lower government subsidy rates do not compromise the effectiveness of coupon programs.[14]

To summarize, we consider that removing the designated usage coupons for specific categories is the primary reason for the improvement in program outcome observed in the second wave. The experimental evidence shows that general coupons are more effective in driving additional spending than specialized coupons, at least in the categories of daily consumption. It appears that the government understood this and immediately responded to the results regarding the coupon design in the first wave by adopting general coupons for all subsequent coupon waves.

4.2 The Length of the Redemption Period

Another notable change in the coupon design is the coupon redemption period. The redemption period was two weeks in Wave I, but this was reduced to one week in Waves II and III. While a longer redemption period offers more opportunity to use the vouchers and therefore does not explain the poorer performance of Wave I relative to later waves, the length of the redemption window is a practical dimension that policy design can potentially optimize upon.

Using data from Wave I, we separately estimate the treatment effect for the first and second weeks of the redemption period. We track individuals in both the treatment and control groups for four weeks, covering two weeks before the first wave coupon distribution date and two weeks after (the whole redemption period for Wave I coupons). The results in **Table 9** show that the spending increase of the treatment group relative to the control group was 59% during the first week of the redemption period, but this fell to 28.2% for the second week. This suggests that coupon receivers

[14] **?** study three waves of coupon disbursement in Hangzhou. The government subsidy rate was 25% in Waves I and III (5 "spend RMB 40, get RMB 10 off" vouchers), while the average subsidy rate was 16.67% in Wave II (with vouchers for "spend RMB 100 and get RMB 20 off", "spend RMB 200 and get RMB 35 off", and "spend RMB 300 and get RMB 45 off"). The total incremental spending was in fact largest in Wave II due to the higher denominations of the coupons compared with the other waves.

tend to use the coupons early.

This pattern may have supported the decision to reduce the redemption window to one week in the second, third, and subsequent waves. Shorter redemption windows also mean that more waves of coupon disbursements could occur during a fixed period than was the case for longer redemption windows, and the total consumption increase is likely to be larger. Relatedly, **?** find that the effectiveness of small value coupons in driving consumption is sustained over multiple rounds of coupon disbursements.

[Insert Table 9: Weekly Treatment Effect-Wave I]

4.3 Discussion on Coupon Design

The iterative experiments of the coupon program provide strong evidence that policymakers can improve the effectiveness of coupon programs by reducing the complexity of coupon design. This is consistent with the abundant literature showing that individuals have limited cognitive resources and that complexity may cause choice avoidance or under-reaction to incentives (e.g., **????**). Empirical policy evaluations find that program confusion, informational complexity, and small "hassles" involved in the claiming procedure contribute substantially to the low take-up rate of many government benefit programs (e.g., **???**).

In our context, the complexity of the coupon design in the first wave, including the restrictions on purchase categories imposed by the specialized coupons and the variety of denominations under each category, could have discouraged consumers from redemption. Although coupon redemption is automatically executed through the digital platform when a transaction meets the requirements (being in a designated category and meeting a certain minimum purchase amount), studies on similar coupon programs show that coupon receivers often "bunch" their spending right above the minimum purchase level to qualify for the coupon subsidy (**?**), suggesting that consumers exert mental effort in utilizing the coupons. **Figure 3** shows the distribution of purchase amounts by transaction during the coupon redemption period of Wave I for both the treatment and control groups. If the treatment group consumers deliberately bunch their spending to take advantage of coupons, we should see spikes in the distribution near the voucher denominations. However, the purchase size distribution of the treatment group is quite similar to that of the control group in this wave, suggesting that bunching behavior is not obvious in Wave I. This is consistent with the argument that the mental effort to categorize a transaction into a designated coupon spending category and consider the best voucher to use were regarded as overwhelming, leading to the low

redemption rate in Wave I. Conversely, we see substantial bunching behavior by the treatment group in Waves II and III, as indicated in **Figures 4 and 5**.

[Insert Figures 3, 4, and 5: Purchase Size Distributions]

We find that turning the specialized coupons into general coupons, coupled with offering fewer denominations, contributed substantially to the increase in the coupon redemption rate in the second wave. The coupon design was further simplified in the third wave, with even fewer denominations than in Wave II. In **Table 8**, we can observe that the redemption rates for vouchers marked as "spend RMB 100, get RMB 20 off" and "spend RMB 5, get RMB 1 off" were above 60%, whereas those for the other two voucher types, "spend RMB 25, get RMB 5 off" and "spend RMB 60, get RMB 15 off," were substantially lower. The statistics support the decision to eliminate one of the low-performing vouchers in the third wave, which led to a further increase in the redemption rate in Wave III.

Our analysis unavoidably faces the identification problem. The process of iteratively implementing the digital coupon program is viewed as experimental, not necessarily as randomization of an experiment. In our context, the program designers deliberately generated variations in the coupon design, which helped to form (imperfect) counterfactuals between waves, but they did not change one program feature at a time. Hence, it is difficult to separate out the impact of each feature improvement. Although various features were packed and incorporated into new wave of coupon disbursement, the overall coupon design moved toward becoming less restrictive and simpler over time. For local government, rolling out the coupon program was an important learning process, through which program designers either ruled out unproductive features or included valuable levers over which they had control based on their evaluation of each wave's coupon disbursement. We find suggestive evidence that fewer restrictions, a simpler design, and an appropriate length of redemption window significantly enhanced the program effectiveness.

Another coupon feature that is worth discussing is the size of the government subsidy. Both the life cycle/permanent income model (e.g., **?**) and a liquidity constraints hypothesis would suggest an inconsequential effect on spending from the coupons studied because of their small size. However, we document treatment effects that are significantly larger than those reported in conventional fiscal stimulus programs under which cash or cash-equivalent coupons were distributed to individuals. The evidence shows that a ***well-designed*** consumption coupon program, despite its small government subsidy value, may still be effective in spurring consumer spending. Program

design matters more than the size of the income shock in improving policy outcomes.

In addition, classic models of consumption (e.g., **??**) predict that the MPC out of unexpected, transitory income shocks declines with the level of liquidity and the size of the income shock. Consistent with this prediction, we find that the size of the government subsidy (i.e., the size of the income shock in our context) plays a secondary role in stimulating spending. Specifically, we find that the third wave of coupon issuance was more effective than the second wave, as shown in the magnitudes of both the efficiency ratio and the implied elasticity, despite the fact that the size of the government subsidy per coupon packet being far smaller in the third wave (RMB 66) than in the second wave (RMB 268). Again, this finding illustrates that program design is more important than the size of the government subsidy in stimulating additional spending in our setting. By deliberately choosing an optimized policy design, the government can achieve a more desirable policy outcome at a much lower budget than would otherwise be possible.

5 Consumer Heterogeneity

The effectiveness of a coupon program also depends on whether it targets the most responsive groups. Although the digital coupon that we examine did not consider the responsiveness of different consumer groups in its design, understanding the heterogeneity in coupon responses can aid future program design when more targeted outcomes are desirable.

5.1 Heterogeneity Across Cities

The coupon program covered all 14 prefecture-level municipal cities in Guangxi autonomous region, which varied in terms of economic development and growth. Policymakers would tend to presume that the coupon program is less effective in spurring consumption in less developed areas, where consumers have relatively lower disposable incomes than those in more economically developed areas. To test this hypothesis and examine the relationship between local characteristics and program effectiveness, we perform the following two-step analysis.

In the first step, we estimate the treatment effect of the coupon program in each wave for each city, using Equation (1) (without city fixed effects). We estimate two models. The first estimates the total unplanned spending during the redemption period for a given wave and uses y_{ijt} as the dependent variable. The second estimates the percentage increase in spending, using $ln(y_{ijt})$ as the dependent variable. In the second step, we use the estimated city-level treatment effect in wave l

as the dependent variable and regress it on several variables of interest as follows:

$$g_{jl} = \theta_0 + \theta_1 GDP_j + \theta_2 Light_j + \theta_3 Penetrate_j + \theta_4 \%Local_j + \theta_5 Merchants_{jl} + W_l + \epsilon_{jl} \quad (2)$$

The first two independent variables capture the level of economic development of city j, including the city's GDP (in log terms) in 2019 and the average night light index of the city in March 2019, the same month but a year before the outbreak of the pandemic. The night light development index of a region reflects the amount of local economic activity and can serve as a useful proxy for the level of economic development. The advantage of using satellite images and night lights as an alternative source of data to measure economic activities is well established in the literature (e.g., **???**).

We further control for the penetration rate of Alipay ($Penetrate_j$) and the ratio of permanent residents among the total population of the city ($\%Local_j$). Consumers in regions with high Alipay penetration rates might be more responsive to the digital coupons, and the population composition may also affect policy efficacy. To account for the fact that the number of participating merchants varies across cities, which may also potentially affect coupon usage, we include a scaled variable that reflects the number of participating merchants in the city ($Merchants_{jl}$) in the regression. Finally, we control for the coupon wave fixed effect.

We report the regression results in **Table 10**. Column (1) uses the overall coupon redemption rate in city j at wave l as the dependent variable. Columns (2) and (3) use the estimated treatment effect in terms of total unplanned spending or its percentage increase as dependent variables. The results indicate that although the economic development measured by GDP is not significant, that measured by the night light index is negative and significant across all outcome variables. The coupon redemption rates, as well as the coupon effectiveness measured by the increase in unplanned spending or its percentage increase, are higher in less developed cities (i.e., cities with smaller night light index values) than in more developed cities. This implies that less economically developed cities respond more vigorously to the coupon program than economically developed cities, which is contrary to what policymakers may have expected.

The results also reveal that the Alipay penetration rate in a city, the ratio of permanent residents, and the number of participating merchants are positively associated with the coupon redemption rate in a city. However, their impact on coupon effectiveness in terms of the amount or percentage of spending increase is limited.

[Insert Table 10: Across City Analysis]

5.2 Heterogeneity Across Consumer Segments

We further examine the heterogeneous treatment effects among consumers using individual-level data from the third wave.

First, we examine how the general expenditure level before the program, a proxy for an individual's income level or purchasing power, affects consumers' responsiveness to the coupon program. We split our sample into terciles according to the individual expenditure level during the six months before the coupon program. An individual falls in the low expenditure group if their average monthly expenditure is below the 33.3 percentile of the distribution, whereas the high expenditure group is defined as those with expenditure above the 66.7 percentile. For each tercile, we run a separate DID analysis using log-spending as the dependent variable and present the estimates in **Table 11**. We find that, on average, the low-income treatment group increased their spending by 371.5% during the redemption period compared with their counterparts in the control group, whereas the corresponding figures for the middle- and high-income treatment groups were 166.6% and 98.5%, respectively. The implied elasticities for the low-, middle- and high-income groups were 18.6, 8.3, and 4.9, respectively. Thus, the results indicate that the low-income group was more responsive to the coupon program than the other groups. This is consistent with the results based on the city-level analysis reported in Section 5.1.

[Insert Table 11: Heterogeneous Effects by Income Level]

Furthermore, we examine whether identity status influences responsiveness to the coupon program. In particular, we are interested in whether non-residents, most of whom were working immigrants, were less responsive than local residents. As the coupon program was region specific and clearly titled the "Guangxi consumption coupon," if the migrants did not identify themselves as locals, they might have been less responsive to the program than residents. To test this hypothesis, we split the matched sample into local and non-local groups. For each group, we separately run the DID analysis to evaluate the treatment effect of the coupon program. The results are presented in **Table 12**. We find that the treatment effects of the two groups are not statistically different. In other words, the migrants are as responsive to the program as the locals.

[Insert Table 12: Heterogeneous Effects: Locals vs. Non-Locals]

5.3 Discussions of Heterogeneous Treatment Effects

The findings in **Tables 10–12** are surprising, and run counter to policymakers' presumptions.[15] However, psychological mechanisms such as reciprocity (**?**) and social identity priming (e.g., **?**) may explain why low-income individuals were more responsive to the coupon program, and why non-residents, once their identity was primed to become salient, responded to the coupon program as responsively as local residents.

On the first point, the government framed the coupon program in the media as a means of supporting local businesses during the pandemic and emphasized that it was the first time that the government had directly used fiscal revenue to subsidize individuals, which may have triggered psychological responses such as reciprocity. The redemption rate of the coupon program was substantially higher than that of the merchant coupon promotion rates reported in the literature (see, e.g., **?**). As the coupon design incorporated various behavioral factors, the effectiveness of the coupon program was a function of consumers' awareness of these behavioral considerations and whether they learned to counteract this consideration (**?**). Consumers with lower incomes and consumers from low-income cities are more likely to be influenced by behavioral factors and, as such, they tended to react more responsively to the coupon program.

On the second point, the government particularly emphasized that all individuals, regardless of whether they were local residents, were eligible for the coupon program. This may have primed non-residents' identity to become salient. This identity treatment may have generated equally or even more positive responses to the coupon program among non-residents than among residents.

Despite the fact that the theories built into the coupon program are implicit, the outcome-driven pragmatic approach combined with iterative experimentation allowed the local government to be mindful of various subtle program details that helped achieve policy outcomes that would otherwise have been unattainable.

6 Conclusion

This paper studies a large-scale, staged fiscal stimulus program leveraging digital platform during the COVID-19 pandemic, in which a provincial government in China disbursed digital consumption coupon packets via Alipay to 8.4 million individual accounts in three consecutive waves and updated the program design in each wave. Leveraging large samples of individual-level data, we estimate

[15] Our interviews with government officials and experts from Alipay confirm this point.

the treatment effect of coupon acquisition during the redemption period of each coupon wave. We document a pattern of increasing treatment effect over waves, suggesting that the government learned from the outcomes of previous experiments and optimized policy design accordingly. In addition, we find that program design matters more than the size of the fiscal stimulus in boosting spending. A ***well-designed*** consumption coupon program, despite its small government subsidy value, may still be effective in spurring consumer spending.

Our empirical results uncover several program features that lead to more desirable outcomes that are otherwise unattainable. First, against conventional wisdom and common practice, we find that the general coupons, with no constraints on where the vouchers could be redeemed, performed much better than the specialized coupons in stimulating consumption in the targeted sectors. Second, we find that the coupon packets with fewer denominations and shorter redemption windows stimulated more additional spending. These two findings suggest that coupons with simpler designs and fewer restrictions are more effective in spurring unplanned spending. Third, we find that the treatment effect is significantly larger among low-income people, and that non-local residents respond to the coupon program in the same way as local residents. Behavioral explanations such as reciprocity and identity priming potentially account for this result.

Many well-intentioned policies fail in practice because governments choose suboptimal designs. Using iterative experiments to uncover causal links between crucial policy details and outcomes, the government can rule out unproductive features and enhance details that achieve a more effective policy. Digital platforms' high technology skill and large customer-based abilities can benefit the distribution of digital coupons, the tracking of redemption automatically, the assessment of treatment effects, and finally help to optimize the design. This study provides empirical evidence that a government can optimize program design through iterative policy experimentation. In this process, implementing the program through a digital platform plays an important role because it enables the collection and analysis of massive amounts of data in real time.

This paper also indicates that paying attention to subtle program details and understanding the psychological mechanisms are important in program design. In contrast with policymakers' presumptions, our results show that general coupons with fewer denominations and shorter redemption periods are more effective than specialized coupons, and that a smaller government subsidy may perform better than a larger government subsidy if unproductive policy features are eliminated. These findings suggest that evidence-based modification of the structure of the interventions can help to target resources by better meeting the needs of the treated group.

Furthermore, our empirical estimates show that the treatment has a more significant impact among residents from low-income cities and those with lower income levels, possibly triggered by behavioral factors. Follow-up experiments could be carefully designed to incorporate this unexpected finding to target an equally or even more responsive group. This study also implies that a government may benefit from involving researchers in the early stage of a program to achieve better policy outcomes through design optimization.

References

Figure 1. Screenshots of the Coupon Acquisition Page and Redemption Receipt



Guangxi Consumption Coupon
The total subsidy is RMB 268.

Subsidy: *Dining* (RMB 148), *Shopping* (RMB 90), *Travel* (RMB 18), *Transportation* (RMB 10), *General* (RMB 2).

Click to check available merchants nearby.
The redemption window is 14 days.



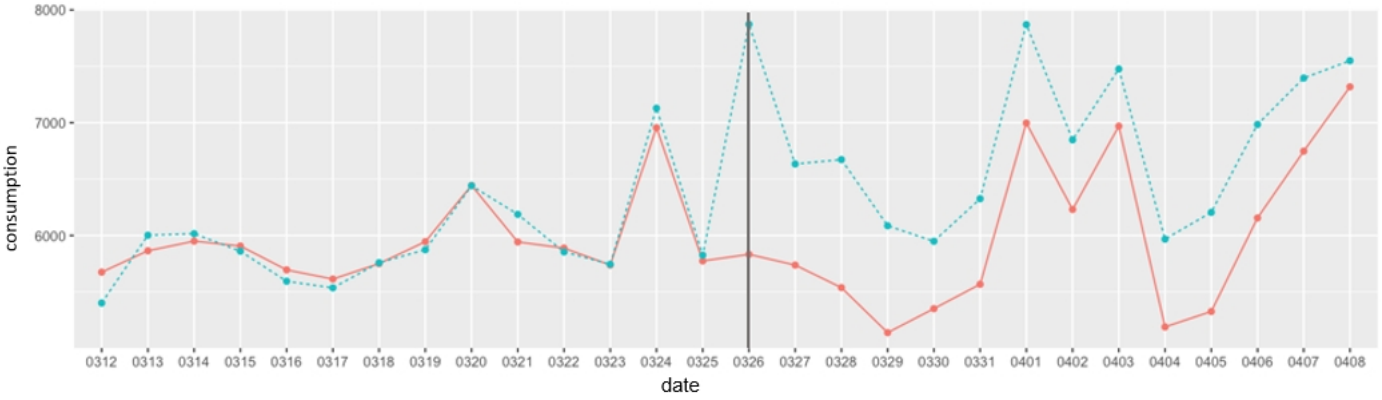
Guangxi Consumption Coupon
The total subsidy is RMB 268.

The packet includes:
3 x "100-20", 10 x "60-15", 11 x "25-5", 3 x "5-1".

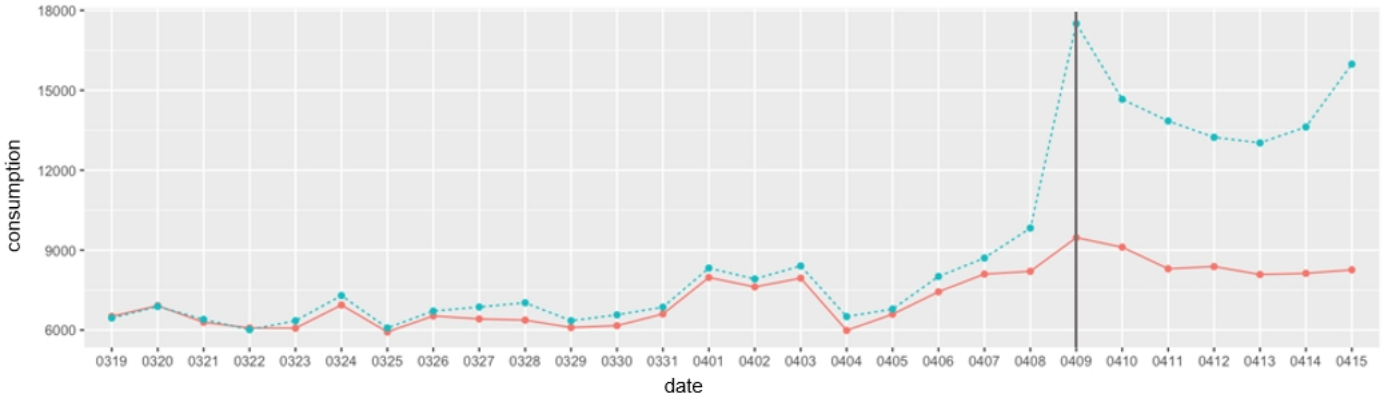
Click to get a coupon packet.
Start at 0 am. The number of packers is limited.

Figure 2. The Average Daily Spending of Two Groups (PSM Sample) of Each Wave

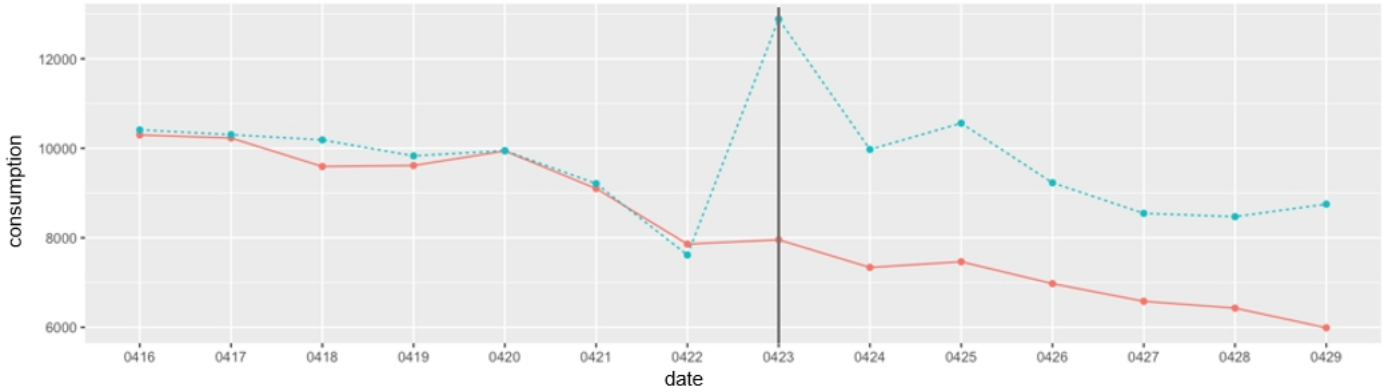
The figures show the time series of daily spending of the treatment group (in dotted blue line) and the control group (in solid red line) with matching. The y-axis refers to the spending amount (in RMB yuan) and x-axis is date.



(a) Wave I coupon



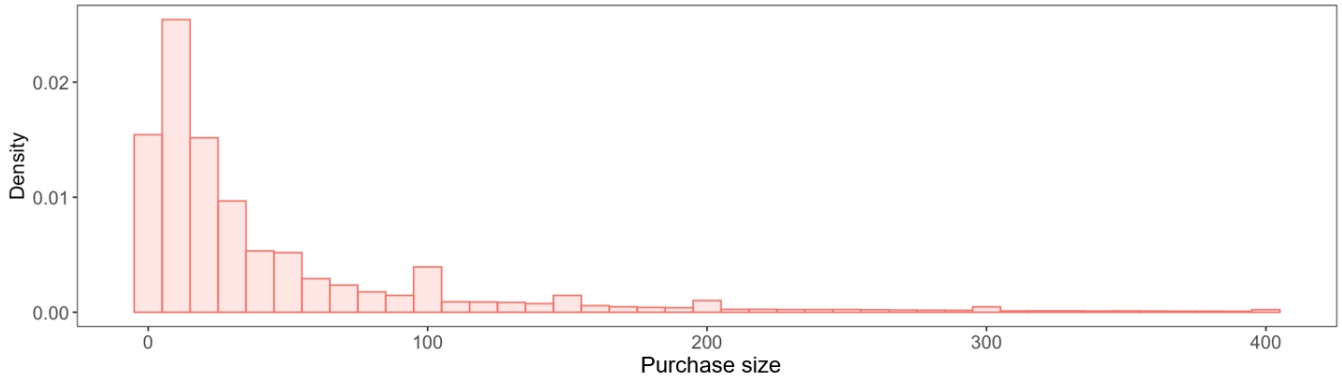
(b) Wave II coupon



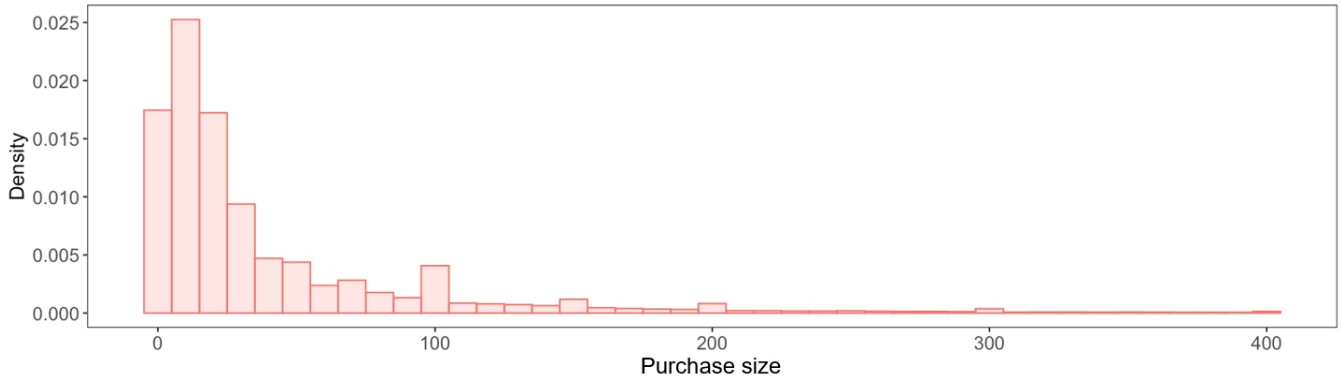
(c) Wave III coupon

Figure 3. The Distribution of Purchase Amount by Transaction -Wave I

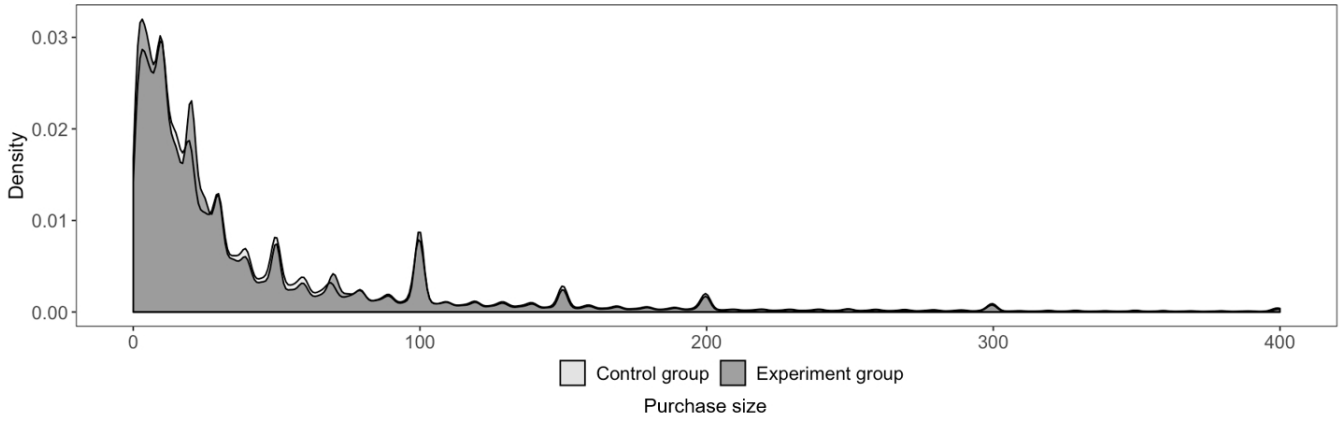
These figures compare the distribution of purchase size (between 0 and RMB 400) by the treatment and control groups in Wave I. The y-axis denotes density; the x-axis is purchase size (in RMB yuan). The upper figure shows the histogram of purchase size of the control group. The middle figure shows the histogram of the experiment group. The bottom figure shows the kernel density of the two, where the dark shaded distribution refers to the experiment group and the light shaded distribution refers to the control group.



(a) Control group



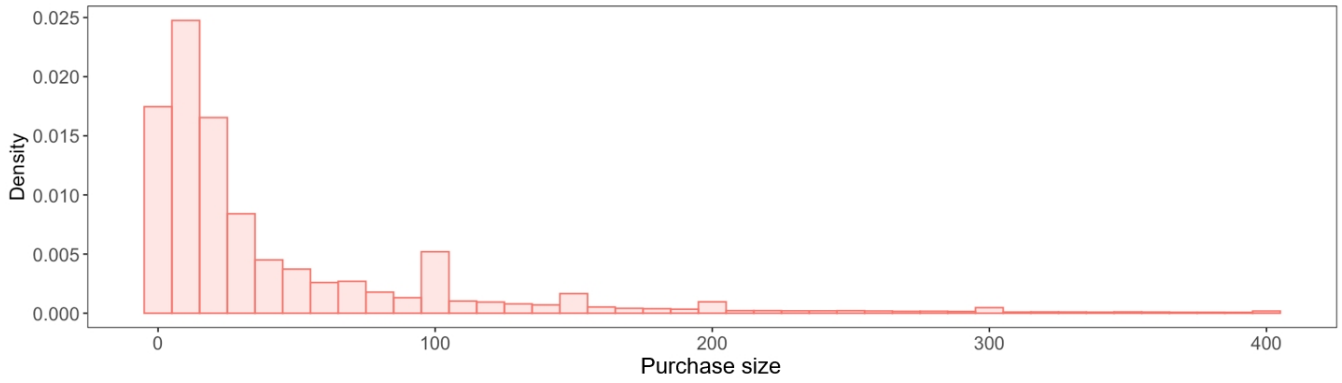
(b) Experiment group



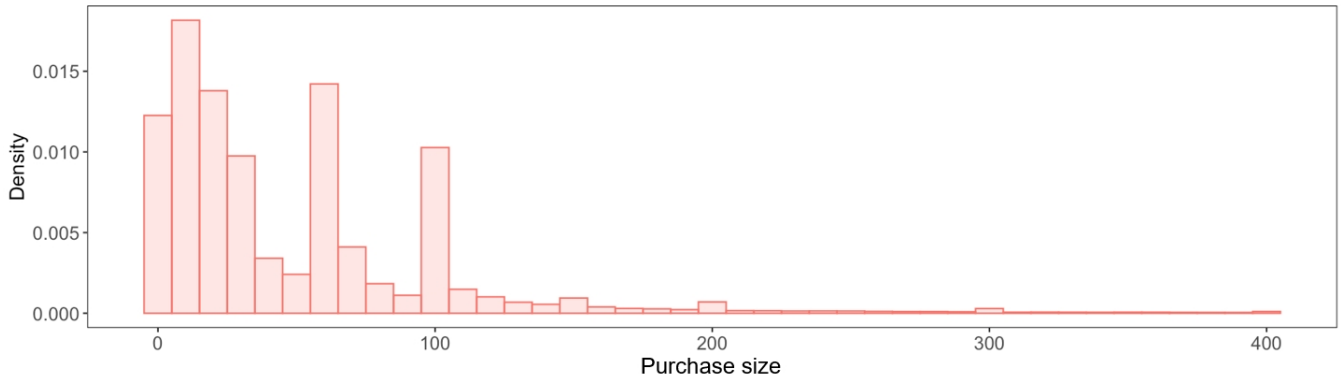
(c) Kernel density of the two groups

Figure 4. The Distribution of Purchase Amount by Transaction -Wave II

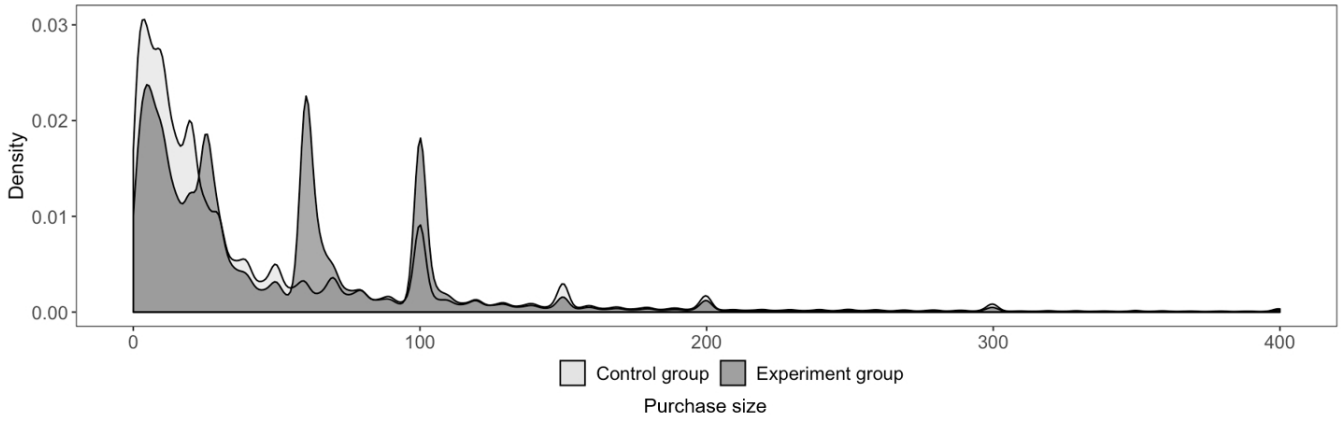
These figures compare the distribution of purchase size (between 0 and RMB 400) by the treatment and control groups in Wave II. The y-axis denotes density; the x-axis is purchase size (in RMB yuan). The upper figure shows the histogram of purchase size of the control group. The middle figure shows the histogram of the experiment group. The bottom figure shows the kernel density of the two, where the dark shaded distribution refers to the experiment group and the light shaded distribution refers to the control group.



(a) Control group



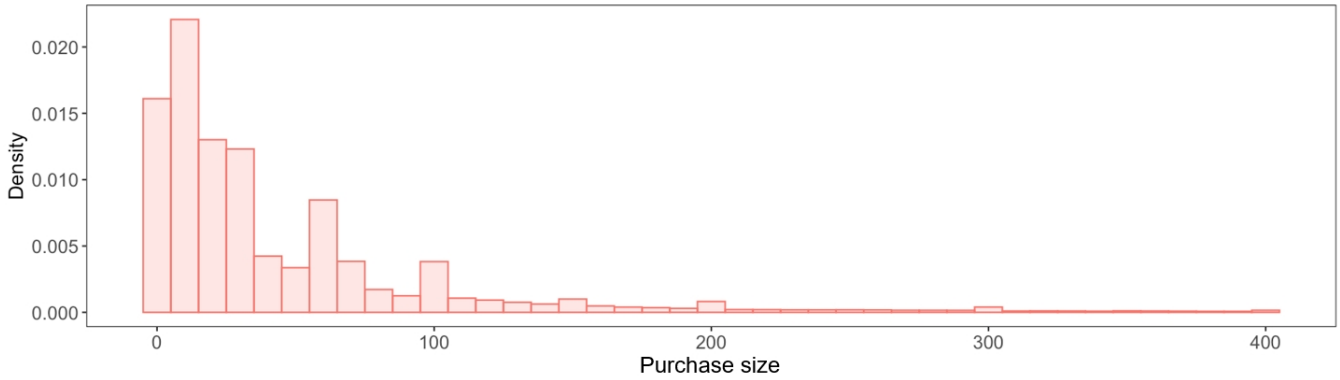
(b) Experiment group



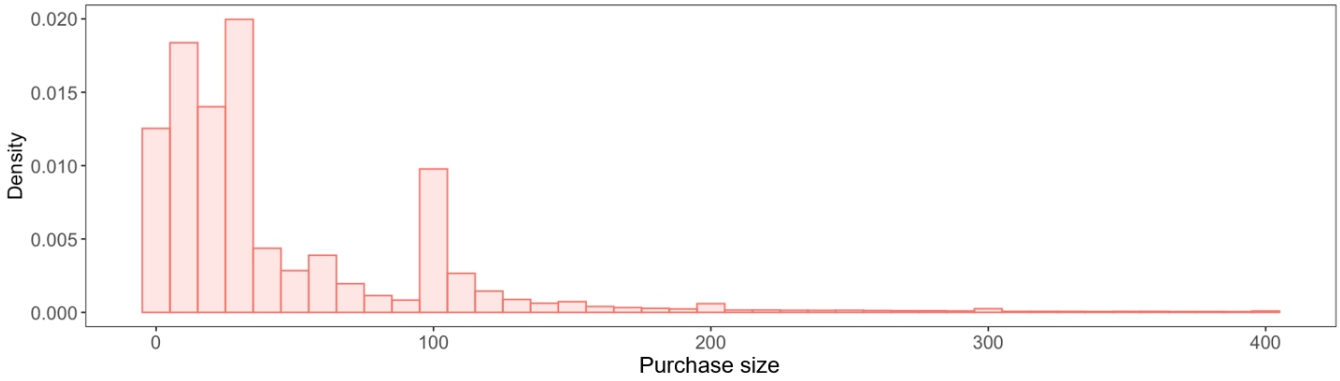
(c) Kernel density of the two groups

Figure 5. The Distribution of Purchase Amount by Transaction -Wave III

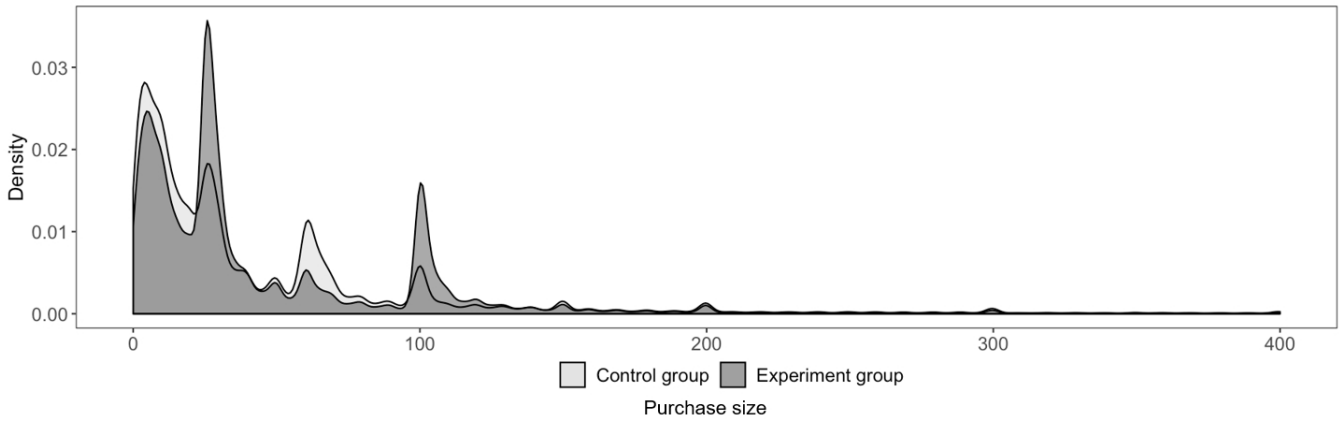
These figures compare the distribution of purchase size (between 0 and RMB 400) by the treatment and control groups in Wave III. The y-axis denotes density; the x-axis is purchase size (in RMB yuan). The upper figure shows the histogram of purchase size of the control group. The middle figure shows the histogram of the experiment group. The bottom figure shows the kernel density of the two, where the dark shaded distribution refers to the experiment group and the light shaded distribution refers to the control group.



(a) Control group



(b) Experiment group



(c) Kernel density of the two groups

Table 1. Summary of Guangxi Coupon Program

Wave	Date	Subsidy/packet (RMB)	Quantity (10,000)	Coupon details	Redemption window	Overall redemption rate
I	3/26-4/8 (Except 4/4)	268	270	Dining: 12x"20-4" Dining: 10x"100-10" Shopping: 8x"20-5" Shopping: 5x"70-10" Travel: 1x"80-10" Travel: 2x"20-4" Transportation: 10x"1-1" General: 2x"5-1"	14 days	9.27%
II	4/9-4/22	268	240	General: 3x"100-20" General: 10x"60-15" General: 11x"25-5" General: 3x"5-1"	7 days	36.81%
III	4/23-5/10	66	330	General: 2x"100-20" General: 5x"25-5" General: 1x"5-1"	7 days	63.32%
IV	5/11-5/20	16	244	General: 3x"25-5" General: 1x"5-1"	7 days	54.34%
V	5/28-6/30	33	385	General: 2x"40-10" General: 2x"25-5" General: 3x"5-1"	7 days	56.19%

Table 2. Summary Statistics

This table shows the summary statistics of the treatment group and control group in three waves. Every panel includes the statistics of the full sample before matching in a given wave as well as the statistics after propensity score matching. The matching is performed based on a random sample of 100,000 individuals in the treatment or control group to reduce computational burden. The variable *Gender* is a dummy that equals one for male and zero for female, and *Pay per month* means the average monthly spending over the six months before the program. The unit of spending is in RMB yuan. T-test is used to compare the group mean of the treatment and control group of continuous variables, and rank test is used to test the mean difference between two groups of dummy variable, i.e., *Gender*.

	Treatment Group		Control Group		
	Mean	SD	Mean	SD	p-value
Panel A: Wave I					
Before Matching					
Age	32.434	9.673	33.962	9.628	0.000
Gender	0.467	0.499	0.418	0.493	0.000
Pay per month	2,905.359	4,136.055	2,757.831	4,263.012	0.000
Obs.	200,000		200,000		
After Matching					
Age	32.636	9.678	32.591	8.902	0.349
Gender	0.440	0.496	0.445	0.497	0.052
Pay per month	2,651.671	3,786.563	2,641.816	3,747.835	0.613
Obs.	74,879		74,879		
Panel B: Wave II					
Before Matching					
Age	33.939	9.395	34.474	10.003	0.000
Gender	0.399	0.490	0.449	0.497	0.000
Pay per month	3,130.734	4,699.451	2,602.243	4,425.939	0.000
Obs.	200,000		200,000		
After Matching					
Age	34.081	9.825	33.955	9.664	0.013
Gender	0.428	0.495	0.411	0.492	0.000
Pay per month	2,706.082	4,226.691	2,696.283	4,369.088	0.661
Obs.	74,041		74,041		
Panel C: Wave III					
Before Matching					
Age	35.126	9.732	35.625	9.891	0.000
Gender	0.402	0.490	0.416	0.493	0.000
Pay per month	2,573.883	4,109.197	2,428.901	4,302.440	0.000
Obs.	200,000		200,000		
After Matching					
Age	35.307	10.077	35.330	9.804	0.653
Gender	0.412	0.492	0.407	0.491	0.048
Pay per month	2,177.772	3,523.309	2,242.100	3,734.247	0.001
Obs.	75,734		75,734		

Table 3. Regression Results of Coupon Effects

Panel A (B) reports the results using the full (PSM) sample of three waves. The dependent variables include total spending (Pay) and the log transformation of the total spending $log(1+Pay)$ at individual level. The unit of regression coefficient is cent (RMB 0.01). All regressions control city level fixed effects. P-values of treatment effects ($Treat \times Post$) are adjusted using **?** method. Standard errors (in parentheses) are clustered at individual level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Full Sample

	Wave I		Wave II		Wave III	
	(1)	(2)	(3)	(4)	(5)	(6)
	Pay	log(1+Pay)	Pay	log(1+Pay)	Pay	log(1+Pay)
Treat x Post	12,291.230***	0.366***	38,961.910***	1.172***	20,406.260***	2.065***
	(479.979)	(0.011)	(403.377)	(0.013)	(366.826)	(0.012)
Treat	1,748.784***	0.450***	3,663.214***	0.901***	1,604.959***	0.606***
	(460.172)	(0.010)	(326.801)	(0.012)	(347.332)	(0.012)
Post	1,352.788***	0.267***	7,268.398***	0.352***	−18,614.360***	−0.864***
	(337.581)	(0.008)	(265.653)	(0.009)	(260.536)	(0.010)
Age	2,460.003***	0.050***	2,253.055***	0.044***	1,554.907***	0.033***
	(110.663)	(0.003)	(75.908)	(0.003)	(75.240)	(0.003)
Age squared	−34.706***	−0.001***	−25.629***	−0.001***	−19.165***	−0.001***
	(1.515)	(0.00004)	(1.006)	(0.00004)	(0.982)	(0.00004)
Gender	2,161.887***	−0.235***	1,453.197***	−0.243***	−568.085**	−0.265***
	(395.214)	(0.008)	(288.975)	(0.009)	(284.733)	(0.009)
Pay per month	0.224***		0.139***		0.133***	
	(0.002)		(0.001)		(0.001)	
log(1+Pay per month)		0.819***		0.713***		0.499***
		(0.003)		(0.002)		(0.002)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.000	0.000
Obs.	800,000	800,000	800,000	800,000	800,000	800,000
Adj. R^2	0.303	0.321	0.277	0.291	0.231	0.235
Efficiency ratio	4.947		3.949		4.883	
Implied elasticity		2.581		5.204		10.325

Panel B: PSM Sample

	Wave I		Wave II		Wave III	
	(1)	(2)	(3)	(4)	(5)	(6)
	Pay	log(1+Pay)	Pay	log(1+Pay)	Pay	log(1+Pay)
Treat x Post	11,656.230***	0.405***	37,869.850***	1.289***	18,818.000***	2.191***
	(764.256)	(0.018)	(651.563)	(0.021)	(563.341)	(0.021)
Treat	713.714	0.345***	4,148.441***	0.772***	1,744.615***	0.505***
	(706.825)	(0.017)	(512.236)	(0.019)	(526.624)	(0.020)
Post	956.534*	0.254***	7,851.285***	0.354***	−17,903.420***	−0.866***
	(546.480)	(0.013)	(441.678)	(0.015)	(404.953)	(0.016)
Age	2,908.657***	0.035***	2,298.623***	0.034***	1,535.250***	0.025***
	(174.956)	(0.005)	(122.872)	(0.005)	(107.241)	(0.005)
Age squared	−41.161***	−0.001***	−26.251***	−0.0005***	−19.246***	−0.0004***
	(2.420)	(0.0001)	(1.643)	(0.0001)	(1.390)	(0.0001)
Gender	1,524.241**	−0.298***	857.088*	−0.221***	−127.937	−0.274***
	(609.714)	(0.014)	(466.034)	(0.015)	(431.707)	(0.015)
Pay per month	0.218***		0.141***		0.133***	
	(0.003)		(0.002)		(0.002)	
log(1+Pay per month)		0.819***		0.706***		0.473***
		(0.004)		(0.004)		(0.003)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.000	0.000
Obs.	299,516	299,516	296,164	296,164	302,936	302,936
Adj. R^2	0.270	0.299	0.265	0.254	0.202	0.210
Efficiency ratio	4.692		3.839		4.503	
Implied elasticity		2.856		5.724		10.955

Table 4. Restricted Sample Regression (Spending $\geq$ 100)

This table shows the DID results based on the restricted sample in which only the individuals with transaction amount no less than RMB 100 per month are retained. The dependent variables include total spending (Pay), and the log transformation of the total spending $log(1 + Pay)$ at individual level. The unit of regression coefficient is cent (RMB 0.01). All regressions control city fixed effects. P-values of treatment effects ($Treat \times Post$) are adjusted using **?** method. Standard errors (in parentheses) are clustered at the individual level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Wave I		Wave II		Wave III	
	(1)	(2)	(3)	(4)	(5)	(6)
	Pay	log(1+Pay)	Pay	log(1+Pay)	Pay	log(1+Pay)
Treat x Post	12,711.680***	0.314***	37,956.870***	1.013***	21,408.900***	1.636***
	(522.660)	(0.011)	(446.713)	(0.013)	(429.606)	(0.013)
Treat	1,096.428**	0.374***	2,910.179***	0.713***	−290.624	0.411***
	(499.001)	(0.010)	(361.315)	(0.012)	(405.059)	(0.011)
Post	1,124.491***	0.216***	8,073.427***	0.296***	−19,528.900***	−0.742***
	(376.539)	(0.008)	(308.257)	(0.010)	(315.470)	(0.010)
Age	2,702.684***	0.036***	2,346.049***	0.031***	1,720.726***	0.020***
	(132.417)	(0.003)	(91.188)	(0.003)	(97.662)	(0.003)
Age squared	−36.746***	−0.001***	−25.726***	−0.0004***	−20.778***	−0.0004***
	(1.860)	(0.00004)	(1.245)	(0.00004)	(1.327)	(0.00004)
Gender	2,428.032***	−0.261***	1,819.097***	−0.276***	−387.201	−0.319***
	(429.729)	(0.008)	(321.557)	(0.009)	(335.454)	(0.009)
Pay per month	0.222***		0.137***		0.131***	
	(0.002)		(0.001)		(0.001)	
log(1+Pay per month)		1.176***		1.084***		0.923***
		(0.004)		(0.006)		(0.004)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.000	0.000
Obs.	730,258	730,258	712,872	712,872	669,246	669,246
Adj. R^2	0.293	0.196	0.265	0.188	0.219	0.161

Table 5. Interaction Effect with Penetration Rate

This table shows the DID results of the coupon interaction effects with the penetration rate of Alipay in a city ($Penetrate$). The dependent variables include total spending (Pay), the log transformation of the total spending $log(1 + Pay)$ at individual level. The unit of regression coefficient is cent (RMB 0.01). All regressions control city fixed effects. P-values of variable $Treat \times Post \times Penetrate$ are adjusted using **?** method. Standard errors (in parentheses) are clustered at consumer level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Wave I		Wave II		Wave III	
	(1)	(2)	(3)	(4)	(5)	(6)
	Pay	log(1+Pay)	Pay	log(1+Pay)	Pay	log(1+Pay)
Treat x Post x Penetrate	−1,244.088	−0.825***	−13,402.430***	−1.307***	2,953.277	−1.189***
	(7,250.747)	(0.151)	(4,254.586)	(0.119)	(3,379.676)	(0.118)
Post x Penetrate	4,957.509	0.351***	5,374.768*	0.183*	−3,205.792	0.452***
	(5,284.398)	(0.121)	(2,921.402)	(0.100)	(2,574.657)	(0.102)
Treat x Post	12,099.480***	0.682***	42,596.790***	1.735***	17,717.140***	2.601***
	(2,511.693)	(0.055)	(1,617.225)	(0.048)	(1,290.564)	(0.047)
Treat	701.663	0.345***	4,177.644***	0.782***	1,780.452***	0.507***
	(706.740)	(0.017)	(513.545)	(0.019)	(528.917)	(0.020)
Post	−724.125	0.136***	5,815.109***	0.285***	−16,719.280***	−1.033***
	(1,831.300)	(0.044)	(1,166.988)	(0.041)	(1,011.950)	(0.041)
Age	2,908.823***	0.036***	2,300.380***	0.035***	1,534.901***	0.025***
	(174.966)	(0.005)	(122.863)	(0.005)	(107.250)	(0.005)
Age squared	−41.162***	−0.001***	−26.274***	−0.0005***	−19.241***	−0.0004***
	(2.420)	(0.0001)	(1.643)	(0.0001)	(1.390)	(0.0001)
Gender	1,523.450**	−0.299***	850.055*	−0.222***	−126.026	−0.275***
	(609.657)	(0.014)	(466.075)	(0.015)	(431.715)	(0.015)
Pay per month	0.218***		0.141***		0.133***	
	(0.003)		(0.002)		(0.002)	
log(1+Pay per month)		0.819***		0.707***		0.473***
		(0.004)		(0.004)		(0.003)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.864	0.000	0.005	0.000	0.573	0.000
Obs.	299,516	299,516	296,164	296,164	302,936	302,936
Adj. R^2	0.270	0.299	0.265	0.254	0.202	0.211

Table 6. Category Analysis of Specialized Coupon (Wave I)-Full Sample

Panel A and Panel B report the DID results using the total category spending (Pay) and the log transformation of the category spending $log(1 + Pay)$ as the dependant variable respectively. The unit of regression coefficient is cent (RMB 0.01). All regressions control for city fixed effects. P-values of treatment effects ($Treat \times Post$) are adjusted using **?** method. Standard errors (in parentheses) are clustered at individual level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Dependent variable: Pay

		Shopping				
	(1)	(2)	(3)	(4)	(5)	(6)
	Dining	Food	Dress	Nondurable	Transportation	Travel
Treat x Post	3,542.897***	544.718***	352.062***	37.352***	5.692***	150.284***
	(128.806)	(16.746)	(27.459)	(3.402)	(2.005)	(37.839)
Treat	1,232.696***	333.130***	173.334***	40.571***	0.129	224.157***
	(129.120)	(17.912)	(25.270)	(2.830)	(1.787)	(31.243)
Post	45.250	14.664	−179.032***	−8.041***	2.119	233.279***
	(88.132)	(11.226)	(18.739)	(2.251)	(1.362)	(25.333)
Age	−293.524***	225.844***	25.044***	9.997***	0.015	−33.838***
	(31.770)	(4.620)	(5.725)	(0.628)	(0.371)	(6.621)
Age squared	1.508***	−2.391***	−0.809***	−0.145***	−0.004	0.095
	(0.425)	(0.066)	(0.077)	(0.008)	(0.005)	(0.087)
Gender	2,219.385***	−173.195***	−2,178.641***	−17.550***	9.101***	585.011***
	(112.993)	(16.434)	(20.191)	(2.317)	(1.495)	(27.217)
Pay per month	0.038***	0.002***	0.004***	0.0002***	0.00005***	0.002***
	(0.0004)	(0.00004)	(0.0001)	(0.00001)	(0.00000)	(0.0001)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.005	0.000
Obs.	800,000	800,000	800,000	800,000	800,000	800,000
Adj. R^2	0.143	0.050	0.066	0.013	0.001	0.011

Panel B. Dependent variable: $log(1 + Pay)$

		Shopping				
	(1)	(2)	(3)	(4)	(5)	(6)
	Dining	Food	Dress	Nondurable	Transportation	Travel
Treat x Post	0.551***	0.414***	0.188***	0.091***	0.022***	0.085***
	(0.014)	(0.012)	(0.014)	(0.009)	(0.003)	(0.008)
Treat	0.591***	0.375***	0.146***	0.111***	0.004*	0.072***
	(0.013)	(0.011)	(0.012)	(0.007)	(0.002)	(0.006)
Post	0.118***	0.054***	−0.085***	−0.022***	0.007***	0.085***
	(0.010)	(0.008)	(0.010)	(0.006)	(0.002)	(0.005)
Age	−0.021***	0.147***	0.003	0.024***	−0.002***	−0.016***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)
Age squared	−0.0001***	−0.002***	−0.0003***	−0.0003***	0.00002***	0.0001***
	(0.00005)	(0.00004)	(0.00004)	(0.00002)	(0.00001)	(0.00002)
Gender	0.025**	−0.225***	−1.239***	−0.054***	0.026***	0.179***
	(0.011)	(0.010)	(0.009)	(0.005)	(0.002)	(0.005)
log(1+Pay per month)	0.712***	0.330***	0.300***	0.111***	0.012***	0.075***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.0003)	(0.001)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.000	0.000
Obs.	800,000	800,000	800,000	800,000	800,000	800,000
Adj. R^2	0.182	0.073	0.071	0.017	0.002	0.017

Table 7. Category Analysis of General Coupon (Wave II)-Full Sample

Panel A and Panel B report the DID results using the total category spending (Pay) and the log transformation of the category spending $log(1 + Pay)$ as the dependant variable respectively. The unit of regression coefficient is cent (RMB 0.01). All regressions control for city fixed effects. P-values of treatment effects ($Treat \times Post$) are adjusted using **?** method. Standard errors (in parentheses) are clustered at individual level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Dependent variable: Pay

		Shopping				
	(1)	(2)	(3)	(4)	(5)	(6)
	Dining	Food	Dress	Nondurable	Transportation	Travel
Treat x Post	9,669.262***	2,862.111***	1,422.988***	128.135***	5.785***	161.932***
	(118.083)	(29.225)	(26.721)	(3.787)	(1.773)	(27.899)
Treat	808.364***	572.169***	102.910***	42.349***	0.815	183.720***
	(97.152)	(19.525)	(19.057)	(2.670)	(1.287)	(21.155)
Post	1,830.982***	332.423***	337.412***	18.241***	4.530***	136.245***
	(77.199)	(13.936)	(15.557)	(2.178)	(1.225)	(17.578)
Age	−71.401***	249.088***	47.764***	12.022***	0.281	−36.484***
	(25.786)	(6.530)	(4.956)	(0.701)	(0.324)	(4.766)
Age squared	0.656*	−2.151***	−0.690***	−0.150***	−0.005	0.275***
	(0.340)	(0.091)	(0.066)	(0.009)	(0.004)	(0.062)
Gender	1,550.102***	−242.485***	−1,352.227***	−35.489***	7.085***	293.683***
	(89.164)	(21.986)	(17.254)	(2.355)	(1.148)	(18.567)
Pay per month	0.025***	0.002***	0.002***	0.0001***	0.00003***	0.001***
	(0.0003)	(0.00004)	(0.00005)	(0.00000)	(0.00000)	(0.0001)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.001	0.000
Obs.	800,000	800,000	800,000	800,000	800,000	800,000
Adj. R^2	0.127	0.062	0.047	0.012	0.001	0.007

Panel B. Dependent variable: $log(1 + Pay)$

		Shopping				
	(1)	(2)	(3)	(4)	(5)	(6)
	Dining	Food	Dress	Nondurable	Transportation	Travel
Treat x Post	1.406***	1.055***	0.615***	0.246***	0.011***	0.059***
	(0.015)	(0.013)	(0.013)	(0.008)	(0.002)	(0.007)
Treat	0.866***	0.527***	0.114***	0.082***	0.009***	0.078***
	(0.014)	(0.011)	(0.010)	(0.006)	(0.002)	(0.005)
Post	0.514***	0.286***	0.240***	0.054***	0.012***	0.038***
	(0.011)	(0.008)	(0.008)	(0.005)	(0.002)	(0.004)
Age	−0.017***	0.150***	0.011***	0.024***	−0.002***	−0.023***
	(0.004)	(0.003)	(0.002)	(0.001)	(0.0005)	(0.001)
Age squared	0.0001	−0.001***	−0.0002***	−0.0003***	0.00002***	0.0002***
	(0.00005)	(0.00004)	(0.00003)	(0.00002)	(0.00001)	(0.00002)
Gender	0.092***	−0.313***	−0.817***	−0.059***	0.016***	0.118***
	(0.011)	(0.010)	(0.008)	(0.005)	(0.002)	(0.004)
log(1+Pay per month)	0.586***	0.246***	0.195***	0.068***	0.008***	0.048***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.0002)	(0.001)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.000	0.000
Obs.	800,000	800,000	800,000	800,000	800,000	800,000
Adj. R^2	0.184	0.090	0.055	0.016	0.002	0.012

Table 8. Redemption Rate by Voucher

Wave	Voucher	Redemption rate
I	Transportation: "1-1"	0.99%
	Shopping: "20-5"	14.14%
	Shopping: "70-10"	11.86%
	Travel: "20-4"	0.37%
	Travel: "80-10"	0.88%
	Dining: "20-4"	3.82%
	Dining: "100-10"	1.49%
	General: "5-1"	57.59%
II	General: "100-20"	64.33%
	General: "25-5"	13.32%
	General: "5-1"	63.34%
	General: "60-15"	26.20%
III	General: "100-20"	67.01%
	General: "25-5"	50.73%
	General: "5-1"	77.43%

Table 9. Weekly Treatment Effect-Wave I

We separately estimate the treatment effect for the first and second week of the redemption period. We regard one week before the distribution date as baseline. $Pre2w$, $Post1w$ and $Post2w$ are two weeks before the coupon distribution date, the first and the second week of redemption period, respectively. The dependent variables include total spending (Pay) and the log transformation of the total spending $log(1 + Pay)$ in a period at individual level. Column(1)-(2) and (3)-(4) are DID results of using full sample and PSM sample, respectively. The unit of regression coefficient is cent (RMB 0.01). All regressions control individual and week level fixed effects. Standard errors (in parentheses) are clustered at consumer level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Full Sample		PSM Sample	
	(1)	(2)	(3)	(4)
	Pay	log(1+Pay)	Pay	log(1+Pay)
Treat x Pre2w	−1,538.911***	−0.090***	−653.637	−0.083***
	(359.606)	(0.015)	(576.242)	(0.025)
Treat x Post1w	6,473.557***	0.590***	6,879.885***	0.630***
	(355.157)	(0.015)	(565.411)	(0.025)
Treat x Post2w	4,278.759***	0.282***	4,122.706***	0.300***
	(386.604)	(0.016)	(617.809)	(0.026)
Individual FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Obs.	1,600,000	1,600,000	599,032	599,032
Adj. R^2	0.476	0.500	0.438	0.483

Table 10. Across City Analysis

The dependent variables are redemption rate, the estimated treatment effect in terms of total unplanned spending and percentage increase at the city level in each wave. The independent variables include the penetration rate of Alipay (*Penetrate*), the ratio of local residents among the total population of the city (%*Local*), scaled index reflecting the number of participating merchants of the city (*Merchants*), city's GDP in year 2019 ($log(1 + GDP)$) and the average nightlight index of the city in March of 2019 (*Light*). The unit of regression coefficient is cent (RMB 0.01). All regressions control coupon wave fixed effects. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
	Redemption rate	Treatment effect (spending)	Percentage increase (log-spending)
Penetrate	0.387*	48,820.190	1.855**
	(0.206)	(31,234.480)	(0.867)
%Local	0.059*	7,764.808	0.213
	(0.034)	(5,124.605)	(0.142)
Merchants	0.021*	2,365.263	0.061
	(0.011)	(1,634.580)	(0.045)
log(1+GDP)	−0.036	−4,932.981	−0.206
	(0.038)	(5,755.781)	(0.160)
Light	−0.079***	−8,813.335**	−0.410***
	(0.023)	(3,464.650)	(0.096)
Wave FE	Yes	Yes	Yes
Obs.	42	42	42
Adj. R^2	0.963	0.740	0.943

Table 11. Heterogeneous Effects by Income Level

We split the matched sample of Wave III into terciles according to an individual's expenditure level during the six months before the coupon program and run analysis separately. Columns (1)-(3) report the results from the low, middle and high level income group. The dependant variable is the log transformation of the total spending $log(1 + Pay)$. The unit of regression coefficient is cent (RMB 0.01). All regressions control for city fixed effects. P-values of treatment effects ($Treat \times Post$) are adjusted using **?** method. Standard errors (in parentheses) are clustered at individual level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
	low	middle	high
Treat x Post	3.715***	1.666***	0.985***
	(0.027)	(0.020)	(0.016)
Treat	1.054***	0.383***	0.251***
	(0.027)	(0.018)	(0.014)
Post	−1.234***	−0.744***	−0.556***
	(0.019)	(0.017)	(0.013)
Age	0.040***	0.023***	0.045***
	(0.006)	(0.005)	(0.005)
Age squared	−0.0004***	−0.0004***	−0.001***
	(0.0001)	(0.0001)	(0.0001)
Gender	−0.275***	−0.347***	−0.278***
	(0.020)	(0.015)	(0.012)
log(1+Pay per month)	0.274***	1.001***	0.754***
	(0.003)	(0.019)	(0.009)
City FE	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000
Obs.	264,000	264,000	272,000
Adj. R^2	0.189	0.072	0.068
Implied elasticity	18.575	8.330	4.925

Table 12. Heterogeneous Effects: Locals vs. Non-locals

We split the matched sample of Wave III into two groups according to whether the individual is a local resident. For each group, we run a separate DID analysis using the log transformation of the total spending $log(1+Pay)$ as the dependent variable. The unit of regression coefficient is cent (RMB 0.01). All regressions control for city fixed effects. P-values of treatment effects ($Treat \times Post$) are adjusted using **?** method. Standard errors (in parentheses) are clustered at individual level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)
	Non-local	Local
Treat x Post	2.196***	2.180***
	(0.028)	(0.031)
Treat	0.499***	0.509***
	(0.026)	(0.030)
Post	−0.829***	−0.912***
	(0.021)	(0.023)
Age	0.027***	0.017**
	(0.006)	(0.008)
Age squared	−0.0004***	−0.0004***
	(0.0001)	(0.0001)
Gender	−0.306***	−0.233***
	(0.021)	(0.023)
log(1+Pay per month)	0.485***	0.460***
	(0.004)	(0.005)
City FE	Yes	Yes
Adj. p-value	0.000	0.000
Obs.	170,026	132,910
Adj. R^2	0.211	0.210
Implied elasticity	10.980	10.900

Table A1. Category Analysis of Specialized Coupon (Wave I)-PSM Sample

This table shows the DID results when decomposing the total spending into categories in Wave I (specialized coupon) using the PSM sample. Panel A and Panel B report the results using the total spending (Pay) and the log transformation of the total spending $log(1 + Pay)$ as the dependant variable respectively. The unit of regression coefficient is cent (RMB 0.01). All regressions control for city fixed effects. P-values of treatment effects ($Treat \times Post$) are adjusted using **?** method. Standard errors (in parentheses) are clustered at individual level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Dependent variable: Pay

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dining	Food	Dress	Nondurable	Service	Transportation	Travel
Treat x Post	3,464.678***	580.434***	327.090***	29.847***	669.508***	3.932	98.478*
	(206.756)	(27.158)	(44.035)	(5.574)	(145.755)	(3.112)	(59.433)
Treat	995.603***	298.076***	−22.747	36.663***	−24.111	−0.143	101.708**
	(199.744)	(28.177)	(39.721)	(4.542)	(143.302)	(2.553)	(49.079)
Post	−65.398	8.386	−175.942***	−2.273	1,192.796***	4.059*	232.548***
	(144.723)	(17.966)	(30.493)	(3.766)	(104.542)	(2.140)	(42.255)
Age	−201.244***	212.128***	4.481	9.151***	478.867***	−0.026	−44.303***
	(51.562)	(7.776)	(10.265)	(1.102)	(36.327)	(0.603)	(11.174)
Age squared	0.155	−2.175***	−0.452***	−0.129***	−7.866***	−0.005	0.228
	(0.697)	(0.113)	(0.144)	(0.015)	(0.507)	(0.008)	(0.154)
Gender	2,475.192***	−188.434***	−2,123.743***	−20.954***	−4,733.396***	5.015**	523.124***
	(176.276)	(26.238)	(31.522)	(3.726)	(122.776)	(2.276)	(40.563)
Pay per month	0.038***	0.002***	0.004***	0.0002***	0.025***	0.00005***	0.002***
	(0.001)	(0.0001)	(0.0001)	(0.00001)	(0.0005)	(0.00000)	(0.0002)
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.000	0.241	0.124
Obs.	299,516	299,516	299,516	299,516	299,516	299,516	299,516
Adj. R^2	0.128	0.051	0.061	0.012	0.111	0.001	0.010

Panel B. Dependent variable: $log(1 + Pay)$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dining	Food	Dress	Nondurable	Service	Transportation	Travel
Treat	0.498***	0.352***	0.064***	0.110***	0.313***	0.006*	0.043***
	(0.022)	(0.018)	(0.019)	(0.011)	(0.021)	(0.003)	(0.010)
Post	0.099***	0.048***	−0.096***	−0.008	0.494***	0.009***	0.081***
	(0.017)	(0.014)	(0.016)	(0.010)	(0.017)	(0.003)	(0.008)
Age	−0.045***	0.136***	−0.010**	0.020***	−0.017***	−0.002**	−0.017***
	(0.006)	(0.005)	(0.005)	(0.003)	(0.006)	(0.001)	(0.002)
Age squared	0.0002***	−0.001***	−0.00004	−0.0002***	−0.0002***	0.00002*	0.0001***
	(0.0001)	(0.0001)	(0.0001)	(0.00004)	(0.0001)	(0.00001)	(0.00003)
Gender	0.001	−0.251***	−1.254***	−0.061***	−0.917***	0.023***	0.166***
	(0.018)	(0.016)	(0.015)	(0.009)	(0.017)	(0.003)	(0.009)
log(1+Pay per month)	0.723***	0.338***	0.302***	0.113***	0.671***	0.011***	0.076***
	(0.005)	(0.003)	(0.003)	(0.001)	(0.004)	(0.0005)	(0.001)
Treat x Post	0.579***	0.444***	0.190***	0.074***	0.181***	0.017***	0.088***
	(0.024)	(0.020)	(0.023)	(0.015)	(0.024)	(0.005)	(0.013)
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.001	0.000	0.000	0.046	0.000
Obs.	299,516	299,516	299,516	299,516	299,516	299,516	299,516
Adj. R^2	0.168	0.073	0.069	0.016	0.167	0.002	0.016

Table A2. Category Analysis of General Coupon (Wave II)-PSM Sample

This table shows the DID results when decomposing the total spending into categories in Wave II (general coupon) using the PSM sample. Panel A and Panel B report the results using the total spending (Pay) and the log transformation of the total spending $log(1+Pay)$ as the dependant variable respectively. The unit of regression coefficient is cent (RMB 0.01). All regressions control for city fixed effects. P-values of treatment effects ($Treat \times Post$) are adjusted using **?** method. Standard errors (in parentheses) are clustered at individual level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Dependent variable: Pay							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dining	Food	Dress	Nondurable	Service	Transportation	Travel
Treat	1,005.102***	505.344***	114.154***	34.998***	−23.650	−0.950	134.924***
	(156.928)	(31.797)	(30.375)	(4.369)	(96.724)	(1.960)	(33.130)
Post	1,780.323***	349.950***	406.230***	18.898***	−230.277***	3.780**	131.476***
	(127.237)	(23.153)	(25.988)	(3.650)	(75.891)	(1.843)	(29.003)
Age	−104.761**	243.803***	49.409***	10.729***	175.504***	−0.381	−35.281***
	(41.951)	(11.012)	(8.182)	(1.173)	(23.501)	(0.503)	(7.302)
Age squared	1.179**	−2.054***	−0.713***	−0.133***	−3.346***	0.005	0.233**
	(0.558)	(0.155)	(0.110)	(0.016)	(0.311)	(0.007)	(0.092)
Gender	1,374.145***	−240.346***	−1,306.194***	−35.146***	−1,850.741***	6.718***	317.663***
	(146.013)	(36.580)	(28.597)	(3.898)	(84.646)	(1.751)	(29.761)
Pay per month	0.027***	0.002***	0.003***	0.0001***	0.013***	0.00004***	0.001***
	(0.001)	(0.0001)	(0.0001)	(0.00001)	(0.0003)	(0.00001)	(0.0001)
Treat x Post	9,581.859***	2,996.139***	1,346.208***	124.025***	1,051.565***	6.676**	135.135***
	(193.417)	(49.057)	(43.560)	(6.198)	(107.981)	(2.711)	(44.048)
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.807	0.676	0.000
Obs.	296,164	296,164	296,164	296,164	296,164	296,164	296,164
Adj. R^2	0.126	0.063	0.042	0.010	0.083	0.002	0.006
Panel B. Dependent variable: $log(1+Pay)$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dining	Food	Dress	Nondurable	Service	Transportation	Travel
Treat x Post	1.466***	1.085***	0.579***	0.236***	0.386***	0.009**	0.054***
	(0.025)	(0.021)	(0.022)	(0.013)	(0.027)	(0.004)	(0.011)
Treat	0.754***	0.487***	0.114***	0.072***	0.224***	0.007**	0.056***
	(0.023)	(0.018)	(0.016)	(0.009)	(0.022)	(0.003)	(0.008)
Post	0.528***	0.304***	0.281***	0.056***	−0.307***	0.012***	0.038***
	(0.018)	(0.014)	(0.014)	(0.008)	(0.019)	(0.003)	(0.007)
Age	−0.035***	0.145***	0.006	0.022***	−0.028***	−0.004***	−0.025***
	(0.006)	(0.005)	(0.004)	(0.002)	(0.006)	(0.001)	(0.002)
Age squared	0.0003***	−0.001***	−0.0001**	−0.0002***	−0.0001	0.00005***	0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.00003)	(0.0001)	(0.00001)	(0.00003)
Gender	0.110***	−0.294***	−0.789***	−0.055***	−0.589***	0.016***	0.124***
	(0.018)	(0.016)	(0.013)	(0.008)	(0.018)	(0.002)	(0.007)
log(1+Pay per month)	0.594***	0.255***	0.201***	0.069***	0.567***	0.008***	0.048***
	(0.004)	(0.003)	(0.002)	(0.001)	(0.003)	(0.0004)	(0.001)
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. p-value	0.000	0.000	0.000	0.000	0.000	0.026	0.000
Obs.	296,164	296,164	296,164	296,164	296,164	296,164	296,164
Adj. R^2	0.164	0.086	0.049	0.014	0.116	0.002	0.012