The Causal Impact of Credit Lines on Spending Distributions

Yijun Li*[1], Cheuk Hang Leung*[1], Xiangqian Sun[2], Chaoqun Wang[1], Yiyan Huang[1], Xing Yan[3], Qi Wu† [1], Dongdong Wang[4], Zhixiang Huang[4]

[1] School of Data Science, City University of Hong Kong
[2] Department of Financial and Actuarial Mathematics, Xi'an Jiaotong Liverpool University
[3] Institute of Statistics and Big Data, Renmin University of China
[4] JD Digits
{yijunli5-c, cqwang5-c, yiyhuang3-c}@my.cityu.edu.hk, {chleung87, qiwu55}@cityu.edu.hk,
Xiangqian.Sun@xjtlu.edu.cn, xingyan@ruc.edu.cn, {wangdongdong9, huangzhixiang}@jd.com

Abstract

Consumer credit services offered by electronic commerce platforms provide customers with convenient loan access during shopping and have the potential to stimulate sales. To understand the causal impact of credit lines on spending, previous studies have employed causal estimators, (e.g., direct regression (DR), inverse propensity weighting (IPW), and double machine learning (DML)) to estimate the treatment effect. However, these estimators do not treat the spending of each individual as a distribution that can capture the range and pattern of amounts spent across different orders. By disregarding the outcome as a distribution, valuable insights embedded within the outcome distribution might be overlooked. This paper thus develops distribution valued estimators which extend from existing real valued DR, IPW, and DML estimators within Rubin's causal framework. We establish their consistency and apply them to a real dataset from a large electronic commerce platform. Our findings reveal that credit lines generally have a positive impact on spending across all quantiles, but consumers would allocate more to luxuries (higher quantiles) than necessities (lower quantiles) as credit lines increase.

Introduction

"Buy now, pay later" (BNPL) is a FinTech credit product offered by e-commerce platforms that allow consumers to make purchases first and defer payments later. BNPL is becoming increasingly popular due to its convenience in online shopping (Guttman-Kenney, Firth, and Gathergood 2023). In practice, e-commerce platforms assign different credit lines (the total amount of money that the platforms lends to a consumer) to potential customers according to their personal information and the history of purchases, payments, and default behaviors.

The primary goal of e-commerce platforms in introducing BNPL is to alter the consumption behavior of consumers, which is usually characterized as a specific ***spending distribution*** formed by the consumption amounts of the consumer's all orders. The spending distributions of various consumers are different. For example, in Figure 7, the spending distribution of some consumers may exhibit a long tail, indicating a preference for both low-price necessities and high-price luxury items, whereas other consumers focus more on middle-valued products.

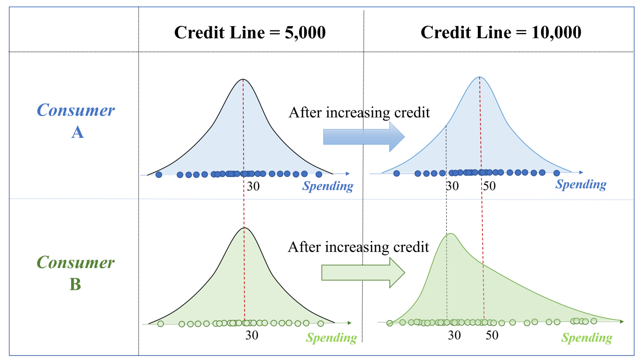

Figure 1: An example for the impact of credit lines change on spending distribution shift (one point stands for spending of one order).

An essential question for e-commerce platforms is whether and how credit lines affect the consumption behavior of consumers. Previous studies have shown that increasing credit lines can lead to increased spending amounts, e.g., (Aydin 2022; Soman and Cheema 2002). Nevertheless, they use a scalar quantity (e.g., ***average spending*** of all orders) to represent the spending of each consumer, which overlooks the complexity of consumption behaviors. For example, consider two consumers (A and B) in Figure 1. When the credit lines of them both equal 5,000, their spending distributions formed by 50 orders are the same, with an average spending of 30 dollars. Supposing the platform increases their credit lines to 10,000, consumer A prefers to increase the spending of all the orders by 20 dollars, and thus the shape of spending distribution does not change but parallelly shifts to the right by 20. On the other hand, consumer B prefers to purchase more luxury goods and remains the spending amounts of orders for necessities unchanged. The shape of consumer B's spending distribution has shifted dramatically, but the average spending is the same as the first consumer (also increased

*These authors contributed equally.
†Corresponding author.

from 30 to 50). Even though these two consumers have the same average spending, their spending behaviors are distinct after the change in credit lines. In this case, focusing only on the average spending loses some of the information of distribution (e.g., the part of quantile information). To this end, we propose to investigate *how the changes of credit lines affect the shift of spending distributions*. However, this raises another question: since classical causal inference literature targets the outcome of each individual as a scalar, *how can we perform causal inference when the outcome of each individual is a distribution?*

In this paper, we employ a novel causal framework to tackle this problem, where the outcome of each unit is a distribution, and the treatment takes multiple values. Based on Rubin's causal framework (Rubin 1977, 1978, 2005), we propose three estimators of target quantities: Direct Regression (DR) estimator, Inverse Propensity Weighting (IPW) estimator, and Doubly Machine Learning (DML) estimator. We first study the statistical asymptotic properties of these estimators. Then, to implement these estimators, we develop a deep-learning-based model named **N**eural **F**unctional **R**egression **Net** (**NFR Net**) to estimate the complex relationship between functional output and scalar input. To assess the effectiveness of our methods, we conduct a simulation study. The results reveal that all three estimators are effective, especially for the DML estimator. We finally apply our approach to investigate the causal impact of credit lines on spending distributions based on a real-world dataset collected from a large e-commerce platform. We find that when credit lines increase, consumers' spending tends to rise, which aligns with previous literature. Additionally, we reveal that the impact of credit lines is more significant in the high-quantile range of spending distribution, suggesting that the increase in credit lines is associated with greater demands for luxury goods rather than necessities.

Our contributions can be summarized as follows:

- We are the first to explore the causal impact of credit lines on spending when each consumer's spending is summarized as a distribution, and discover more detailed findings on the distribution quantiles compared to the literature.
- We consider multiple treatment and propose three estimators (i.e., DR, IPW, and DML estimators) for the target quantities. We further study their statistical properties and compare them in a simulation experiment.
- The relation between functional output and scalar input is always non-linear and complex. Existing works captured the relation by a linear or parameterized function, but we develop a deep learning model 'NFR Net' to learn it.

Related Work

Causal inference is a significant challenge in various fields, such as finance (Huang et al. 2021) and health care (Shi, Blei, and Veitch 2019). The key assumption of classical causal inference is that, given the treatment $D = d$, all the units have the same potential outcome distribution (unconditional). As a result, the realization of the outcome for each individual is a *scalar point* drawing from that potential outcome distribution (for instance, in Figure 2 when $D = d$, the blue (red) point is a realization of the i^{th} (j^{th}) unit). Under the assumption, several causal quantities are introduced and studied. For instance, the average treatment effect (ATE) (Chernozhukov et al. 2018) is the difference between the means of any two potential outcome distributions (i.e., $\mathbb{E}[Y(D = \bar{d})] - \mathbb{E}[Y(D = d)]$, or see the left half of Figure 2). Another quantity is the quantile treatment effect (QTE) (Chernozhukov and Hansen 2005) that studies the difference between two potential outcome distributions at τ-quantiles (i.e., $Q(\tau, Y(D = \bar{d})) - Q(\tau, Y(D = d))$), or see the right half of Figure 2).

Various methods have been proposed to estimate the causal effect between treatment and outcome. A common approach is constructing the estimators for the target quantities. For example, Direct Regression (DR) incorporates all confounding factors into a single regression function. The inverse propensity weighting (IPW) method (Rosenbaum and Rubin 1983; Hirano, Imbens, and Ridder 2003), on the other hand, assigns weights to the units based on their propensity scores which mimic RCTs in the pseudo population. However, both of them require accurate estimations of the nuisance parameters, such as the regression function and propensity scores. Doubly Machine Learning (DML) (Chernozhukov et al. 2018) method overcomes the shortcomings. It has the doubly robust property such that the accuracy of estimating nuisance parameters can be loosened.

The above methods are restricted when the outcome of each unit includes many observations or points and they constitute a *distribution*. For example, the shopping spending of a consumer may differ each time, and all the spending amounts form a distribution. In this case, it is impossible to infer the causal relationship via the standard framework unless we reduce the distributions to points (e.g., take the mean). Thus, it is necessary to seek alternative frameworks for distributional outcomes.

The distributional outcome can be treated as a continuous function. It is closely related to the field of functional data analysis that analyzes data under information varying over a continuum (Ramsay and Silverman 2005; Wang, Chiou, and Müller 2016; Cai et al. 2022; Chen, Goldsmith, and Ogden 2016). (Jacobi, Wagner, and Frühwirth-Schnatter 2016) and (Chib and Jacobi 2007) apply the functional data analysis to study the relationship between functional outcomes and independent variables based on the panel dataset. Nevertheless, they do not focus on the causal studying. (Ecker, de Luna, and Schelin 2023) considers a causal framework to study the impact of treatment on the functional outcome. However, their work conducts causal inference in Euclidean space. It is believed that the random structure of the distributional outcome is destroyed in the Euclidean space (Verdinelli and Wasserman 2019; Panaretos and Zemel 2019). As such, (Lin, Kong, and Wang 2023) considers the causal study in the Wasserstein space, and we extend their framework to study the causal effect on distributional outcomes under multiple treatments and with a deep learning model NFR Net (statistical properties can be ensured as well). In this case, the realization of the outcome for each unit is a *distribution* (for example, in Figure 3 when $D = d$, the blue (red) distribution is a realization of the i^{th} (j^{th}) unit).

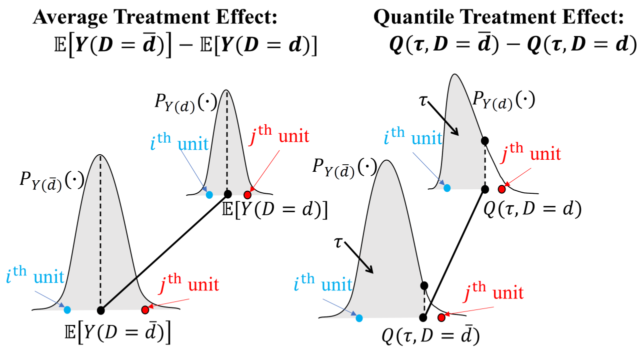

Figure 2: ATE and QTE in the literature.

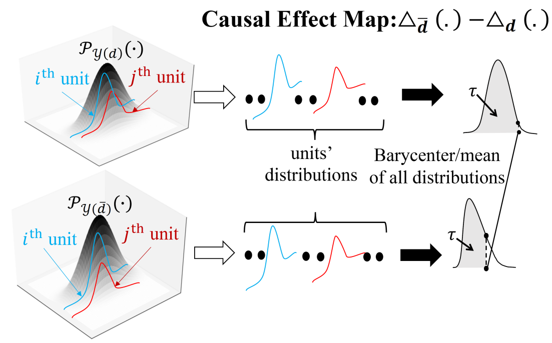

Figure 3: Causal Effect Map in our paper.

Causal Inference Framework

We denote D as the treatment and $D \in \mathfrak{D} = \{d^1, \dots, d^r\}$, $\mathbf{X} = [X^1, \cdots, X^n] \in \mathcal{X} \subseteq \mathbb{R}^n$ as the *covariates/confounders* with distribution $F_{\mathbf{X}}$ and $\mathcal{X}$ is bounded. With scalar outcomes, prior literature defines Y as the outcome variable and $Y(d)$ as the potential outcome variable when receiving treatment $D = d$ such that $Y = \sum_{i=1}^{r} Y(d^i) \cdot \mathbf{1}_{\{D=d^i\}}$. The *potential outcome distribution and density* of $(Y, Y(d))$ are $(F_Y, F_{Y(d)})$ and $(P_Y, P_{Y(d)})$. In our framework, the outcome of each unit is described as a distribution and varies across units. We distinguish the differences by using $\mathcal{Y}$ as the outcome variable and $\mathcal{Y}(d)$ as the potential outcome variable when receiving treatment $D = d$ such that $\mathcal{Y} = \sum_{i=1}^{r} \mathcal{Y}(d^i) \cdot \mathbf{1}_{\{D=d^i\}}$. The *potential outcome distribution and density* of $(\mathcal{Y}, \mathcal{Y}(d))$ are $(\mathcal{F}_{\mathcal{Y}}, \mathcal{F}_{\mathcal{Y}(d)})$ and $(\mathcal{P}_{\mathcal{Y}}, \mathcal{P}_{\mathcal{Y}(d)})$. We assume that there are N-independent units, i.e., $\{(D_s, \mathbf{X}_s, \mathcal{Y}_s)\}_{s=1}^{N}$.

Causal Assumptions

The following causal assumptions are standard under Rubin's framework (Rubin 2005): (1) *Consistency* (i.e., if $D = d^i$ occurs, then $\mathcal{Y} = \mathcal{Y}(d^i)$ a.s.); (2) *Ignorability/Unconfoundness* (i.e., $\mathcal{Y}(d^i) \perp\!\!\!\perp D|\mathbf{X}, \forall i \in \{1, \dots, r\}$); (3) *Overlap* (i.e., $\mathbb{P}\{D = d^i|\mathbf{X}\} > 0, \forall i \in \{1, \dots, r\}$).We defer detailed explanations about the causal assumptions to Appendix A.

Causal Quantities on Distributions

Since the realization of $\mathcal{Y}$ for each unit is a distribution, it is inappropriate to conduct causal inference in the Euclidean space as it destroys the structure of distributions. For example, Figure 4 displays the 10 distributions (all are normal distributions with different mean and variance), and the corresponding "mean" distribution using the Wasserstein metric (Barycenter) and the Euclidean metric. We notice that the "mean" distribution cannot preserve the Gaussian structure unless the Wasserstein metric is used. We thus choose to conduct causal inference in the *Wasserstein space* (Villani 2021; Panaretos and Zemel 2019; Feyeux, Vidard, and Nodet 2018). Here, we use the p-Wasserstein metric to characterize the "distance" between two distributions (see Definition 1). In the sequel, we let the realizations of $\mathcal{Y}$, $\mathcal{Y}(d)$ reside in $\mathbb{R}$.

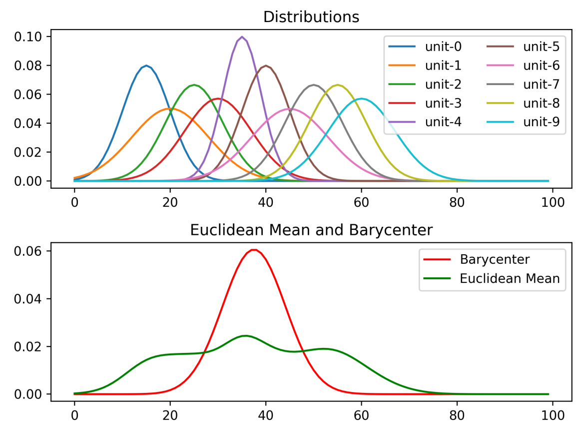

Figure 4: The Euclidean mean and Barycenter of 10 distributions.

Definition 1. *Let $\mathcal{I} \subset \mathbb{R}$, $\mathcal{W}_p(\mathcal{I}) = \{\lambda : \int_{\mathcal{I}} s^p d\lambda(s) < \infty\}$ (λ is a distribution), and $\Lambda(\lambda_1, \lambda_2)$ be the set containing the joint distribution $\Pi(\lambda_1(s), \lambda_2(t))$ whose marginals are λ_1 and λ_2. The p-Wasserstein metric between λ_1 and λ_2 is*

$$\mathbb{D}_p(\lambda_1, \lambda_2) = \left\{ \inf_{\Pi \in \Lambda(\lambda_1, \lambda_2)} \int_{\mathcal{I}} |s - t|^p d\Pi(\lambda_1(s), \lambda_2(t)) \right\}^{\frac{1}{p}}.$$

$\mathbb{D}_p(\cdot, \cdot)$ satisfies the *axioms of a metric* (i.e., non-negativity, symmetric, and triangle inequality). Usually, we set $p = 2$. Next, we introduce two quantities - the *causal map* and the *causal effect map*.

Definition 2. *The causal map of treatment d^i is denoted as $\triangle_{d^i}$* [1] *such that*

$$\triangle_{d^i} = \mu_{d^i}^{-1}, \tag{1}$$

where $\mu_{d^i} = \underset{v \in \mathcal{W}_2(\mathcal{I})}{\arg\min}\ \mathbb{E}\left[\mathbb{D}_2(\mathcal{Y}(d^i), v)^2\right]$ is the Wasserstein barycenter/mean of units' distributions when they take the treatment d^i. The superscript "-1" of μ_{d^i} is the inverse of the cumulative distribution function (CDF) or the quantile

[1] $\triangle_{d^i}$ is a function and should be $\triangle_{d^i}(\cdot)$ formally. In the sequel, we use both $\triangle_{d^i}$ and $\triangle_{d^i}(\cdot)$ interchangeably.

	Our framework	Literature framework
Treatment/Covariates variable	$D/\mathbf{X}$	$D/\mathbf{X}$
Outcome variable	$\mathcal{Y}, \mathcal{Y}(d)$	$Y, Y(d)$
Potential outcomes distribution (density)	$\mathcal{F}_{\mathcal{Y}(d)}(\cdot)$ $(\mathcal{P}_{\mathcal{Y}(d)}(\cdot))$	$F_{Y(d)}(\cdot)$ $(P_{\mathcal{Y}(d)}(\cdot))$
Metric	Wasserstein	Euclidean
Space of outcome variable	$\mathcal{W}_2(\mathcal{I})$	$\mathcal{I} \in \mathbb{R}$
Realization of outcome variable	distribution	scalar
Target quantity	$\triangle_{d^i}, \triangle_{d^{ij}}$	$\mathbb{E}[Y(d^i)], \mathbb{E}[Y(d^i)] - \mathbb{E}[Y(d^j)]$

Table 1: Comparisons between our framework and the framework given in the literature.

function. Hence, the causal effect map between treatment d^i and d^j is

$$\triangle_{d^{ij}} = \triangle_{d^i} - \triangle_{d^j} = \mu_{d^i}^{-1} - \mu_{d^j}^{-1}. \tag{2}$$

The *causal effect map* in Eqn. (2) is an analogy to the ATE ($\mathbb{E}[Y(d^i)] - \mathbb{E}[Y(d^j)]$) in the literature. However, $\triangle_{d^i}$, $\triangle_{d^j}$ and $\triangle_{d^{ij}}$ are functions, but $\mathbb{E}[Y(d^i)]$, $\mathbb{E}[Y(d^j)]$, and $\mathbb{E}[Y(d^i)] - \mathbb{E}[Y(d^j)]$ are scalars. In Table 1, we summarize the differences between the framework in our paper and in the literature.

Remark 1.

1. *Classically, the case "distribution over $\mathbb{R}$" means that a realization is a point (scalar or vector) drawing from the distribution of the potential outcome, while the case "distribution over distributions" means that the realization is a distribution. For instance, let μ and σ be the mean and standard deviation of a normal distribution, and $(\mu, \log\sigma) \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_2)$. If the realization $(\mu, \log\sigma)$ of a unit (e.g., a consumer) is $(0.1, -0.5)$, then it means that a collection of observations (e.g., spending amounts of all orders) are drawn from $\mathcal{N}(0.1, e^{-1})$ for this unit.*
2. *$\triangle_{d^i}(\cdot)$ is a quantile function (inverse of CDF), so does $\triangle_{d^{ij}}(\cdot)$. Further, we can explore the impact of between treatment d^i and d^j on the distributional outcome respectively at a specific τ quantile level by $\triangle_{d^{ij}}(\cdot)$, i.e.,*
$$\triangle_{d^{ij}}(\tau) = \triangle_{d^i}(\tau) - \triangle_{d^j}(\tau) = \mu_{d^i}^{-1}(\tau) - \mu_{d^j}^{-1}(\tau).$$
*Note that $\triangle_{d^{ij}}(\tau)$ differs from the quantile treatment effect (QTE) in the literature (e.g., (Machado and Mata 2005; Chernozhukov and Hansen 2005)). $\triangle_{d^{ij}}(\tau)$ is the τ-quantiles difference of the **barycenters** under treatments d^i and d^j, but QTE is the τ-quantiles difference of the **potential outcome distributions** under two treatments. It is thus inappropriate to compare them or study $\triangle_{d^{ij}}(\tau)$ using the approaches in the QTE literature. The visualized difference of the two quantities is given in Figure 2 and 3.*

We need to ensure $\triangle_{d^i}$ is *identifiable* such that we can estimate it from an observed dataset. It is also necessary to simplify the calculation of μ_{d^i} to address the computational complexity of optimal transport. Proposition 1 states an equivalent form of $\triangle_{d^i}$ without computing optimization and guarantees that we can estimate it from the observed dataset:

Proposition 1. *Given the conditions in Definition 1 and 2, and Assumptions (1) - (3) hold, we have (1) $\triangle_{d^i} = \mathbb{E}\big[\mathcal{Y}(d^i)^{-1}\big]$; (2) $\triangle_{d^i}$ is identifiable.*

The first assertion gives a simpler way to compute $\triangle_{d^i}$, while the second assertion ensures that $\triangle_{d^i}$ is identifiable. We defer the proofs to Appendix D.

Estimators

Similar to the causal inference methods in the literature (Horvitz and Thompson 1952; Chernozhukov et al. 2018), we also propose three estimators to compute the causal map $\triangle_{d^i}$, namely (1) *Direct Regression (DR) estimator ($\triangle_{d^i;DR}$)*, (2) *Inverse Probability Weighting (IPW) estimator ($\triangle_{d^i;IPW}$)*, and (3) *Double Machine Learning (DML) estimator ($\triangle_{d^i;DML}$)*. Let $\pi_{d^i}(\mathbf{X}) = \mathbb{P}\{D = d^i|\mathbf{X}\}$ and $m_{d^i}(\mathbf{X}) = \mathbb{E}\big[\mathcal{Y}^{-1}|D = d^i, \mathbf{X}\big]$. Given that there are N units. The estimators $\triangle_{d^i;DR}$, $\triangle_{d^i;IPW}$, and $\triangle_{d^i;DML}$ are given in Eqns. (3), (4), and (5) respectively:

$$\frac{1}{n}\sum_{s=1}^{n} m_{d^i}(\mathbf{X}_s), \tag{3}$$

$$\frac{1}{n}\sum_{s=1}^{n} \frac{\mathbf{1}_{\{D_s=d^i\}}}{\pi_{d^i}(\mathbf{X}_s)}(\mathcal{Y}_s^{-1}), \tag{4}$$

$$\frac{1}{n}\sum_{s=1}^{n}\big[m_{d^i}(\mathbf{X}_s) + \frac{\mathbf{1}_{\{D_s=d^i\}}}{\pi_{d^i}(\mathbf{X}_s)}(\mathcal{Y}_s^{-1} - m_{d^i}(\mathbf{X}_s))\big]. \tag{5}$$

Theory and Algorithm

In practical scenarios, when using all the available units to train the regression function $m_{d^i}(\mathbf{X}_s)$ and propensity score function $\pi_{d^i}(\mathbf{X}_s)$, there is a risk of over-fitting. To mitigate this issue, a cross-fitting technique, as introduced by (Chernozhukov et al. 2018), is commonly employed. Along this way, we also need to obtain the cross-fitting estimators of $\triangle_{d^i}$ according to Eqns. (3), (4), and (5).

We split the N units into K disjoint groups. Let the k^{th} group be $\mathcal{D}_k$ of size N_k, $\forall k = 1, \cdots, K$. Denoting $\mathcal{D}_{-k} = \cup_{r=1, r\neq k}^{K}\mathcal{D}_r$, we use $\mathcal{D}_{-k}$ to obtain $\hat{m}_{d^i}^k(\mathbf{X})$, $\hat{\pi}_{d^i}^k(\mathbf{X})$, which are the estimations of $m_{d^i}^k(\mathbf{X})$, $\pi_{d^i}^k(\mathbf{X})$ for the k^{th} group. $\hat{\mathcal{Y}}$ is the empirical estimation of $\mathcal{Y}$. We then use $\mathcal{D}_k$ to compute the estimation of $\triangle_{d^i}^k$ (i.e., $\hat{\triangle}_{d^i;DR}^k$, $\hat{\triangle}_{d^i;IPW}^k$, and $\hat{\triangle}_{d^i;DML}^k$) according to Eqns. (6), (7), and (8) respectively. We thus define $\hat{\triangle}_{d^i;DR}^k$, $\hat{\triangle}_{d^i;IPW}^k$, and $\hat{\triangle}_{d^i;DML}^k$ in Eqns. (6), (7), and (8) respectively

$$\frac{1}{N_k}\sum_{s\in\mathcal{D}_k}\hat{m}^k_{d^i}(\mathbf{X}_s) \tag{6}$$

$$\frac{1}{N_k}\sum_{s\in\mathcal{D}_k}\frac{\mathbf{1}_{\{D_s=d^i\}}}{\hat{\pi}^k_{d^i}(\mathbf{X}_s)}\hat{\mathcal{Y}}^{-1}_s \tag{7}$$

$$\frac{1}{N_k}\sum_{s\in\mathcal{D}_k}\left[\hat{m}^k_{d^i}(\mathbf{X}_s)+\frac{\mathbf{1}_{\{D_s=d^i\}}}{\hat{\pi}^k_{d^i}(\mathbf{X}_s)}(\hat{\mathcal{Y}}^{-1}_s-\hat{m}^k_{d^i}(\mathbf{X}_s))\right]. \tag{8}$$

Denoting $w \in \{DR, IPW, DML\}$, the cross-fitting estimators are $\hat{\triangle}_{d^i;w}$ such that

$$\hat{\triangle}_{d^i;w}=\sum_{k=1}^{K}\frac{N_k}{N}\hat{\triangle}^k_{d^i;w}. \tag{9}$$

We study the consistency of $\hat{\triangle}_{d^i;w}$. When $w = DR$ or IPW, the results are deferred to Appendix C. When $w = DML$, the consistency result is given in Theorem 1 while the proofs and the notational meanings are deferred to Appendix D.

Theorem 1. *Let $\tilde{m}^k_{d^i}(\mathbf{X})$ ($\hat{m}^k_{d^i}(\mathbf{X})$) be the estimate of $\mathbb{E}[\mathcal{Y}^{-1}|D = d^i, \mathbf{X}]$ using the true $\mathcal{Y}$ (estimated $\hat{\mathcal{Y}}$) based on $\mathcal{D}_{-k}$. Suppose that, for any k, $\rho^4_\pi = \mathbb{E}[|\hat{\pi}^k_{d^i}(\mathbf{X}) - \pi_{d^i}(\mathbf{X})|^4]$, $\rho^4_m = \max\{\|\tilde{m}^k_{d^i} - m_{d^i}\|^4,\ 1 \le i \le r\} = \max\{[\int \|\tilde{m}^k_{d^i}(\mathbf{x}) - m_{d^i}(\mathbf{x})\|^2 dF_{\mathbf{X}}(\mathbf{x})]^2,\ 1 \le i \le r\}$. Under the convergence assumptions in Appendix D, we have*

1. *$\|\hat{\triangle}_{d^i;DML} - \triangle_{d^i}\| = O_P(N^{-\frac{1}{2}} + N^{-\frac{1}{2}}\rho_\pi + N^{-\frac{1}{2}}\rho_m + \rho_\pi\rho_m)$.*
2. *If $\rho_m\rho_\pi = o(N^{-\frac{1}{2}})$, $\rho_m = o(1)$ and $\rho_\pi = o(1)$, then $\sqrt{N}(\hat{\triangle}_{d^i;DML} - \triangle_{d^i})$ converges weakly to a centred Gaussian process.*

Theorem 1 not only gives the consistency of $\hat{\triangle}_{d^i;DML}$, but also gives the convergence speed of $\hat{\triangle}_{d^i;DML}$. It is indeed a $\sqrt{N}$-consistent estimator.

We can also investigate the $\sqrt{N}$-consistency of the DR or IPW estimators. In fact, we can obtain the desired results by setting $\mathbf{1}_{\{D=d^i\}} = 0$ and $(m_{d^i}, \hat{m}^k_{d^i}, \tilde{m}^k_{d^i}) = (0, 0, 0)$ in the proofs of Theorem 1 respectively. Last but not least, we summarize the steps of computing $\hat{\triangle}_{d^i;w}$ in Algorithm 1.

Models

To estimate the target quantity $\triangle_{d^i}$, we need to estimate several nuisance parameters accurately, e.g., $\mathcal{Y}^{-1}$, $\pi_{d^i}(\mathbf{X})$, and $m_{d^i}(\mathbf{X})$. First, to estimate $\mathcal{Y}^{-1}$, we can estimate $\mathcal{Y}$ empirically and invert the estimated $\mathcal{Y}$ (CDF) for each unit to get the $\hat{\mathcal{Y}}^{-1}$. Second, $\pi_{d^i}(\mathbf{X})$ is the *propensity score* that can be estimated using the multi-class logistic regression, random forest classifier, or feed-forward networks. Finally, we can estimate the regression function $m_{d^i}(\mathbf{X})$ by regressing the outcome $\mathcal{Y}^{-1}$ on treatment D and covariates $\mathbf{X}$ via a *functional-on-scalar regression*. The first two quantities can be well estimated using the classical approaches. On the other hand, the third quantity, $m_{d^i}(\mathbf{X})$, is difficult to estimate accurately using the classical functional regression approach. Specifically, the classical functional regression

Algorithm 1: Computations of $\hat{\triangle}_{d^i;w}$

Require: The observations of $(D_s, \mathbf{X}_s, \mathcal{Y}_s)^N_{s=1}$.
Ensure: $\hat{\triangle}_{d^i;w}$ for $w \in \{DR, IPW, DML\}$.

1: Split $(D_s, \mathbf{X}_s, \mathcal{Y}_s)^N_{s=1}$ to K disjoint units groups $\mathcal{D}_k$ of size N_k and form $\mathcal{D}_{-k}$.
2: Estimate $\hat{\mathcal{Y}}^{-1}_s$ for each unit s.
3: **for** $k \leftarrow 1$ to K **do**
4:     Regress D w.r.t. $\mathbf{X}$ based on $\mathcal{D}_{-k}$ and obtain $\hat{\pi}^k_{d^i}$.
5:     Regress $\hat{\mathcal{Y}}^{-1}$ w.r.t. $(D, \mathbf{X})$ based on $\mathcal{D}_{-k}$ and obtain $\hat{m}^k_{d^i}$.
6:     Compute $\hat{\triangle}^k_{d^i;w}$ using Eqns. (6), (7) and (8) according to w.
7: **end for**
8: Compute $\hat{\triangle}_{d^i;w}$ using Eqn. (9).

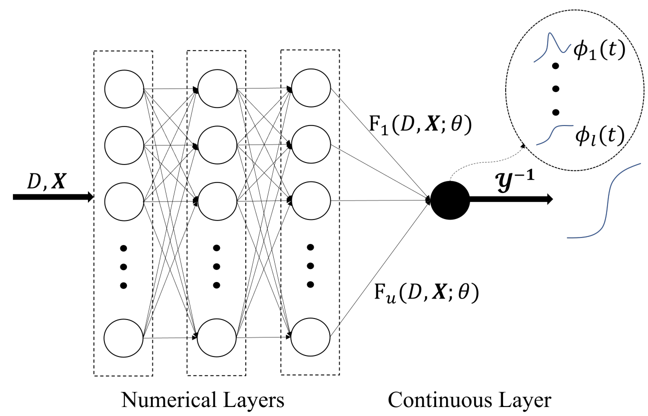

Figure 5: The proposed NFR Net.

(Ramsay and Silverman 2005) assumes that the regression equation between outcome $\mathcal{Y}^{-1}$ and predictors $(D, \mathbf{X})$ can be approximated by a finite series of some pre-determined basis functions, i.e., the response function $\mathcal{Y}^{-1}(t)$ equals

$$D\sum_{l=1}^{v}\left[\gamma_{0l}\phi_l(t)+\sum_{j=1}^{n}X^j\gamma_{jl}\phi_l(t)\right]+\epsilon(t), \tag{10}$$

where $(D, \mathbf{X}) = [D, X^1, \cdots, X^j, \cdots, X^n]$ are predictors; $\{\phi_1, \ldots, \phi_v\}$ are basis functions, e.g., B-spline basis; γ_{jl} with $0 \le j \le n$ and $1 \le l \le v$ are regression parameters; and $\epsilon(t)$ is the noise term.

However, the relation between $\mathcal{Y}^{-1}(t)$ and $(D, \mathbf{X})$ may not be additive as in Eqn. (10). Generally, the relationship is non-linear and complex. To this end, we design **N**eural **F**unctional **R**egression (**NFR**) Net to address this issue. The NFR Net consists of two parts: (1) the *numerical layers*, and (2) the *continuous layer* (see Figure 5). Under our framework and settings, the numerical layers aim to learn the u representations $\mathsf{F}(D, \mathbf{X}; \theta) = [\mathsf{F}_1(D, \mathbf{X}; \theta), \cdots, \mathsf{F}_u(D, \mathbf{X}; \theta)]^\top$, where each $\mathsf{F}_i(D, \mathbf{X}; \theta), 1 \le i \le u$ is a linear coefficient to constitute the target distribution. The representations $\mathsf{F}(D, \mathbf{X}; \theta)$ is then processed by a continuous layer to output

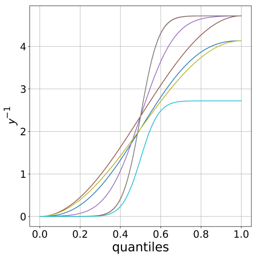

Figure 6: 10 instances of simulated quantile function ($\mathcal{Y}^{-1}$).

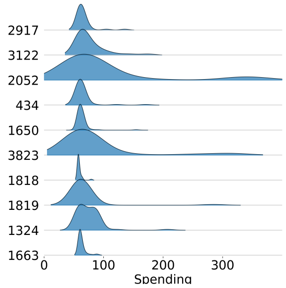

Figure 7: The spending distribution of 10 consumers from a e-commerce platform.

	DR	IPW	DML
Lasso	0.124 ± 0.053	0.044 ± 0.021	0.038 ± 0.032
Ridge	0.118 ± 0.050	0.044 ± 0.021	0.034 ± 0.027
Elastic net	0.124 ± 0.053	0.045 ± 0.021	0.038 ± 0.032
MCP	0.118 ± 0.050	0.044 ± 0.021	0.034 ± 0.027
D/DML	0.300 ± 0.240	0.044 ± 0.021	0.037 ± 0.026
NFR	$\mathbf{0.052 \pm 0.027}$	0.044 ± 0.021	$\mathbf{0.034 \pm 0.026}$

Table 2: MAE between true and estimated causal effect maps under various methods of regressing $\hat{\mathcal{Y}}^{-1}$ w.r.t. $(D, \mathbf{X})$ (mean $\pm$ standard deviation with 50 trials). Best results are in bold. The IPW results are similar because the same classifier (random forest) is used to get the propensity scores.

a function $\tilde{\mathcal{Y}}^{-1}$, i.e.,

$$\tilde{\mathcal{Y}}^{-1}(t; \theta, \{\gamma_{ij}\}) = \sum_{i=1}^{u} \mathsf{F}_i(D, \mathbf{X}; \theta) \sum_{j=1}^{v} \gamma_{ij}\phi_j(t), \tag{11}$$

where $\{\gamma_{ij}\}$ now are trainable parameters and $\{\phi_j(t)\}$ are pre-defined basis functions.

The model can be trained as follows: let L be the loss metric (e.g., L_1 or L_2 loss), and our task is finding the optimal $\theta, \{\gamma_{ij}\}$ by minimizing the loss function $\mathcal{L}(\theta, \{\gamma_{ij}\})$:

$$\min_{\theta, \{\gamma_{ij}\}} \mathcal{L}(\theta, \{\gamma_{ij}\}) := \int L(\tilde{\mathcal{Y}}^{-1}(t; \theta, \{\gamma_{ij}\}), \hat{\mathcal{Y}}^{-1}(t))dt. \tag{12}$$

In practice, we can estimate the integral using the trapezoidal rule/Simpson's rule by taking any number of discrete quantile points t.

Synthetic Experiment[2]

Data Generation Process Since the ground truth is unavailable in the real dataset, we simulate data using the following data generation process for the s^{th} unit to test our proposed model similar to many other causal inference studies:

$$\mathcal{Y}_s^{-1}(D_s) = c + (1-c)(\mathbb{E}[D] + \sqrt{D_s})\times \sum_{j=1}^{\frac{n}{2}} \frac{\exp(X_s^{2j-1}X_s^{2j})}{\sum_{k=1}^{\frac{n}{2}} \exp(X_s^{2k-1}X_s^{2k})} \mathbf{B}^{-1}(\alpha_j, \beta_j) + \epsilon_s, \tag{13a}$$

$$\mathbb{P}\{D_s = d \mid \mathbf{X}_s\} = \frac{\exp(\gamma_d^\top \mathbf{X}_s)}{\sum_{w=1}^{r} \exp(\gamma_w^\top \mathbf{X}_s)}. \tag{13b}$$

In our experiment, we set $n = 10$. We assume that covariates $X^1, X^2 \sim \mathcal{N}(-2, 1)$, $X^3, X^4 \sim \mathcal{N}(-1, 1)$, $X^5, X^6 \sim \mathcal{N}(0, 1)$, $X^7, X^8 \sim \mathcal{N}(1, 1)$, $X^9, X^{10} \sim \mathcal{N}(2, 1)$, and $\epsilon_s \sim \mathcal{N}(0, 0.05)$. $\mathbf{B}^{-1}(\alpha, \beta)$ is the inverse CDF of Beta distribution with the shape parameters α and β. We select 5 inverse Beta CDFs, where each one has different parameters to ensure the complexity of the distribution function.

[2]Our code is available at https://github.com/lyjsilence/The-Causal- Impact-of-Credit-Lines-on-Spending-Distributions.

The treatment D takes the value in $\{d^1, d^2, d^3, d^4, d^5\}$ with a softmax distribution. $c \in [0, 1]$ is the constant that controls the strength of the causal relationship between treatment D and outcome distribution $\mathcal{Y}^{-1}$. In one experiment, $5,000$ instances are generated according to Eqns. (13a) - (13b). For each unit s, 100 observations are sampled from the inverse CDF using the inverse transform sampling method. Figure 6 summarizes 10 simulated instances, indicating that the inverse CDF of each instance varies widely.

Baselines In our experiment, we consider two aspects of potential baseline methods. The first aspect is from the statistical field, where approaches such as those presented in (Lin, Kong, and Wang 2023) assume a linear relationship between the functional output and the scalar input. They utilize regularization techniques like lasso, ridge, and elastic net to estimate the causal effect map. Additionally, (Chen, Goldsmith, and Ogden 2016) addresses situations with a large number of covariates by using the group minimax concave penalty (MCP) for variable selection and fitting. However, these methods inherently assume a linear form between the functional output and scalar input, possibly overlooking the presence of nonlinear relationships in the data. The second aspect is from the deep learning field, where we compare our model with classical Double/debiased machine learning (D/DML) proposed in (Chernozhukov et al. 2018). This approach introduces a DML estimator to investigate the causal impact of scalar input on scalar outcome. To model the functional outcome, we conduct independent regressions at interesting quantiles using a standard MLP. Subsequently, we concatenate all the quantile counterfactuals to form a distribution.

Experiment Setting The classification and functional regression models are trained separately. $5,000$ generated instances are trained using 5-fold cross-fitting, i.e., $4,000$ instances are used to train, and $1,000$ instances are used to obtain the three estimators (i.e., DR, IPW, and DML estimator). At last, we average the obtained estimators from the 5 folds as the final results. In the classification task, we use the same classifier (i.e., random forest) to compute IPW for all the estimators. The training details are given in Appendix E.

Evaluation Metric Since $\mathcal{L}(\theta, \{\gamma_{ij}\})$ in (12) is continuous, we discretize it and compare the mean absolute error (MAE) between true causal effect map $\triangle_{d^{ij}} (1 \le i, j \le 5)$

Quantiles	Low	Middle	High	Low→Middle	Low→High
10%	28.0 (27.8, 28.3)	29.9 (29.8, 30.0)	30.6 (30, 31.2)	6.79%↑	9.29%↑
20%	43.6 (43.4, 43.9)	47.4 (47.1, 47.8)	48.8 (47.9, 49.5)	8.72%↑	11.93%↑
30%	58.7 (58.3, 59.0)	65.5 (65.1, 65.9)	67.5 (66.5, 68.4)	11.58%↑	14.99%↑
40%	75.2 (74.7, 75.6)	86.8 (86.3, 87.2)	91.0 (90.0, 92.1)	15.43%↑	21.01%↑
50%	94.9 (94.3, 95.6)	115.8 (114.9, 116.9)	122.4 (121.1, 123.8)	22.02%↑	28.98%↑
60%	119.0 (118.2, 119.7)	150.8 (149.7, 152.0)	170.8 (167.4, 174.7)	26.72%↑	43.53%↑
70%	155.1 (153.7, 156.4)	207.0 (205.6, 208.5)	256.0 (251.8, 261.5)	33.46%↑	65.05%↑
80%	212.9 (210.8, 214.6)	325.6 (323.2, 328.3)	433.0 (424.4, 442.7)	52.94%↑	103.38%↑
90%	381.0 (374.1, 386.7)	654.5 (650.7, 658.4)	1020.3 (1003.8, 1036.9)	71.78%↑	167.80%↑

Table 3: The results of the causal map of three treatments at 9 quantiles (mean and 95% CI).

(computed from Eqns. (13a) - (13b)) and estimated causal effect map $\hat{\triangle}_{dij}$ on 5 quantiles with levels ranging from 10% to 90%. We repeat the experiment 50 times to report the mean and standard deviation of MAE.

Experiment Results Table 2 presents a summary of the experiment results (A table of quantiles 10%, 30%, 50%, 70%, 90%, and Average is given in Appendix). We observe several key findings: Firstly, NFR Net demonstrates superior performance compared to all statistical models, particularly on the DR methods. This result can be attributed to the capability of our proposed model to capture non-linear patterns between covariates and the outcome distribution effectively. Secondly, NFR Net outperforms the D/DML method. The advantage stems from our ability to model the outcome as a function. In contrast, D/DML treats each quantile as independent scalar points, overlooking the continuous structure of the distribution. Lastly, DML can utilize the IPW estimator to correct most of the bias in the DR estimator, and the DML estimator demonstrates improved robustness compared to both the DR and IPW estimators.

Empirical Experiment

E-commerce platforms face a significant challenge in comprehending the impact of credit lines on consumer spending patterns, particularly in terms of the shift in spending distribution caused by changes in the credit lines. To address this issue, we employ our approach by leveraging data collected from a large e-commerce platform. The platform assigns distinct credit lines to users based on various factors such as income, age, and past behaviors like shopping and default behaviors. Besides, the platform provides users with an interest-free, one-month loan option for their purchases, with the condition that the total loan amount must not exceed their assigned credit lines.

We collect data from 4,043 platform users. The data comprises various variables, such as demographic information (e.g., age, income, and location), purchasing behaviors (e.g., the total number of orders, the amount paid for each order), and financial information (e.g., credit lines assigned by the platform, the total number of loans, and the presence of default records). Appendix F displays a detailed statistical description. All the paid amounts of orders constitute a unique spending distribution for each user (e.g., Figure 7). In our empirical study, we investigate the causal maps when the credit lines take values as Low (from 0 to 9,000), Middle (from 9,000 to 15,000), and High (higher than 15,000).

In Table 3, we give 9 percentiles of the causal map $\triangle_{High}$, $\triangle_{Middle}$, and $\triangle_{Low}$ of all the consumers' spending distributions if they are assigned to High, Middle, and Low credit lines, respectively. Generally, the lower quantile of spending distribution stands for life necessities, while the higher quantiles represent luxury goods. Our findings support prior research (Aydin 2022; Soman and Cheema 2002), revealing a positive correlation between credit lines and spending since the causal effect maps $\triangle_{High} - \triangle_{Low}$ and $\triangle_{Middle} - \triangle_{Low}$ are always positive at all quantiles. Additionally, we observe that such an effect is heterogeneous across different quantiles. Specifically, when the credit lines increase, the spending on higher quantiles (e.g., higher than 70%) grows significantly while the spending on lower quantiles increases relatively slowly. For example, when credit lines change from Low to High, the spending at 90% quantile increases from 381.0 to 1020.3 (increasing about 167.8%) while the spending at 10% quantile only increases from 28.0 to 30.6 (increasing about 9.3%). This suggests that users tend to increase their spending on luxury goods or services when they are able to access higher credit.

Conclusion

We study the causal inference on distributional outcomes with multiple treatments in the Wasserstein space. Our target quantity, the causal effect map, is the analogy to ATE in classical causal inference literature. We then propose three estimators, i.e., DR, IPW, and DML estimators, and study their asymptotic properties. Our proposed NFR Net captures complex patterns among covariates, treatments, and functional outcomes, which is verified by the synthetic experiment. Moreover, we apply it to a credit dataset and explore the causal relationship between credit lines and spending distributions. We find that when credit lines increase, the spending at every quantile level increases, with a more significant change at higher quantiles.

Generally, the credit lines is measured continuously, and a potential future research direction involves investigating causal inference in the context of continuous treatment. Additionally, the realized distribution can be multivariate, such as the joint distribution of spending behavior and credit risk, providing an opportunity to explore such scenarios.

Acknowledgments

Qi Wu acknowledges the support from The CityU-JD Digits Joint Laboratory in Financial Technology and Engineering; The Hong Kong Research Grants Council [General Research Fund 11219420/9043008 and 11200219/9042900]; The HK Institute of Data Science. The work described in this paper was partially supported by the InnoHK initiative, The Government of the HKSAR, and the Laboratory for AI-Powered Financial Technologies.

References

Aydin, D. 2022. Consumption response to credit expansions: Evidence from experimental assignment of 45,307 credit lines. *American Economic Review*, 112(1): 1–40.

Cai, X.; Xue, L.; Cao, J.; and Initiative, A. D. N. 2022. Robust estimation and variable selection for function-on-scalar regression. *Canadian Journal of Statistics*, 50(1): 162–179.

Chen, Y.; Goldsmith, J.; and Ogden, R. T. 2016. Variable selection in function-on-scalar regression. *Stat*, 5(1): 88–101.

Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; and Robins, J. 2018. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1).

Chernozhukov, V.; and Hansen, C. 2005. An IV model of quantile treatment effects. *Econometrica*, 73(1): 245–261.

Chib, S.; and Jacobi, L. 2007. Modeling and calculating the effect of treatment at baseline from panel outcomes. *Journal of Econometrics*, 140(2): 781–801.

Ecker, K.; de Luna, X.; and Schelin, L. 2023. Causal inference with a functional outcome. *arXiv preprint arXiv:2304.07113*.

Feyeux, N.; Vidard, A.; and Nodet, M. 2018. Optimal transport for variational data assimilation. *Nonlinear Processes in Geophysics*, 25(1): 55–66.

Guttman-Kenney, B.; Firth, C.; and Gathergood, J. 2023. Buy now, pay later (bnpl)... on your credit card. *Journal of Behavioral and Experimental Finance*, 100788.

Hirano, K.; Imbens, G. W.; and Ridder, G. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4): 1161–1189.

Horvitz, D. G.; and Thompson, D. J. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260): 663–685.

Huang, Y.; Leung, C. H.; Yan, X.; Wu, Q.; Peng, N.; Wang, D.; and Huang, Z. 2021. The Causal Learning of Retail Delinquency. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1): 204–212.

Jacobi, L.; Wagner, H.; and Frühwirth-Schnatter, S. 2016. Bayesian treatment effects models with variable selection for panel outcomes with an application to earnings effects of maternity leave. *Journal of Econometrics*, 193(1): 234–250.

Lin, Z.; Kong, D.; and Wang, L. 2023. Causal inference on distribution functions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(2): 378–398.

Machado, J. A.; and Mata, J. 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4): 445–465.

Panaretos, V. M.; and Zemel, Y. 2019. Statistical aspects of Wasserstein distances. *Annual review of statistics and its application*, 6: 405–431.

Ramsay, J. O.; and Silverman, B. W. 2005. *Fitting differential equations to functional data: Principal differential analysis*. Springer.

Rosenbaum, P. R.; and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1): 41–55.

Rubin, D. B. 1977. Assignment to treatment group on the basis of a covariate. *Journal of educational Statistics*, 2(1): 1–26.

Rubin, D. B. 1978. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 34–58.

Rubin, D. B. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469): 322–331.

Shi, C.; Blei, D.; and Veitch, V. 2019. Adapting neural networks for the estimation of treatment effects. *Advances in neural information processing systems*, 32.

Soman, D.; and Cheema, A. 2002. The effect of credit on spending decisions: The role of the credit limit and credibility. *Marketing Science*, 21(1): 32–53.

Verdinelli, I.; and Wasserman, L. 2019. Hybrid Wasserstein distance and fast distribution clustering. *Electronic Journal of Statistics*, 13: 5088–5119.

Villani, C. 2021. *Topics in optimal transportation*, volume 58. American Mathematical Soc.

Wang, J.-L.; Chiou, J.-M.; and Müller, H.-G. 2016. Functional data analysis. *Annual Review of Statistics and its application*, 3: 257–295.