Designing Discontinuities

Ibtihal Ferwana [1 2] **Suyoung Park** [1] **Ting-Yi Wu** [1] **Lav R. Varshney** [1 3]

Abstract

Discontinuities can be fairly arbitrary but also cause a significant impact on outcomes in social systems. Indeed, their arbitrariness is why they have been used to infer causal relationships among variables in numerous settings. Regression discontinuity from econometrics assumes the existence of a discontinuous variable that splits the population into distinct partitions to estimate causal effects. Here we consider the *design* of partitions for a given discontinuous variable to optimize a certain effect. To do so, we propose a quantization-theoretic approach to optimize the effect of interest, first learning the causal effect size of a given discontinuous variable and then applying dynamic programming for optimal quantization design of discontinuities that balance the gain and loss in the effect size. We also develop a computationally-efficient reinforcement learning algorithm for the dynamic programming formulation of optimal quantization. We demonstrate our approach by designing optimal time zone borders for counterfactuals of social capital.

1. Introduction

Whether one earns admission to a particular school on the basis of a test score is often linked to significant educational and life outcomes (Park et al., 2015), but the admissions threshold may be quite arbitrary. Those on one side of the threshold may otherwise be quite similar to those on the other side. There are similarly arbitrary discontinuities in numerous settings in public policy, economics, healthcare, and elsewhere that may cause significant impacts. Indeed, categorization on the basis of fairly arbitrary partitioning of certain attributes abounds in social life (Varshney & Varshney, 2017).

Discontinuity—the presence of a discrete set of partitions—has been used to learn causal relationships among variables. Indeed a leading method for causal inference in econometrics is regression discontinuity design (RDD), which assumes a pre-known threshold dividing the population into two discontinuous groups (Angrist & Pischke, 2009). Comparing samples on each side of the threshold point then allows inference of causality.

RDD has been applied in various domains with discontinuities, e.g. in national security (Dell & Querubin, 2018), in estimating prices of used cars (Englmaier et al., 2018), and in faculty performance reviews (Rivera & Tilcsik, 2019).

Here, we take up the challenge of partition design in the context of causal effects of discontinuities, which as far as we know has not been studied in any systematic manner. Indeed, there appears to be a lacuna at this intersection of causal inference and mechanism design. To do so, we first establish theory to address the counterfactuals of redesigning discontinuities. Then we develop new techniques in optimal quantization theory for partition design. Although there is a large literature in quantization theory (Gray & Neuhoff, 1998; Gersho & Gray, 1991)—including in settings of statistical inference (Poor & Thomas, 1977; Varshney & Varshney, 2008; Misra et al., 2011; Shlezinger et al., 2019)—there appears to be no prior work on quantization in the context of causal inference.

1.1. Time Zone Discontinuity

Time zones are an example of discontinuities that affect daily life. Normatively, the 720 longitude lines of the Earth are divided into 24 time zones. However, there are countries that geographically lie in certain time zones but actually follow other time zones, such as France and Spain, which use Western European Time. China geographically spans five time zones but uses just a single China Standard Time throughout. Time zone borders may not run straight from north to south, but may follow certain political boundaries.

Time zone systems along with globalized social systems led to uniform start times of school, work, and sleep, which may dictate wake-up times not aligned with the sun (Hamermesh et al., 2008). Such schedules can disrupt human circadian

[1]Coordinated Science Laboratory, University of Illinois Urbana-Champaign [2]Department of Computer Science, University of Illinois Urbana-Champaign [3]Electrical and Computer Engineering, University of Illinois Urbana-Champaign. Correspondence to: <{iferwna2, varshney}@illinois.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

rhythms and have consequences on health and productivity (Cappuccio et al., 2010; Gibson & Shrader, 2018). By exploiting the discontinuity in the timing of natural light at time zone boundaries, (Giuntella & Mazzonna, 2019) found that an extra hour of natural light in the evening reduces sleep duration by 20 minutes. Such results imply sunset time, as linked to geographic location, may contribute in other ways to health and well-being.

In this paper, we ask how to partition the world into time zones that optimize social capital, rather than the arbitrary time zone system in place today. For this, we need two things: first, the counterfactual prediction using regression discontinuity design to measure the effect of current time zone borders, and second, quantization to design the optimal time zone borders.

2. Causal Effects of Time Zones on Human Well-being

Let us consider social capital as an indicator of human wellbeing. To measure the discrepancy on social capital, when given standard time zone borders, we apply RDD (Lee & Lemieux, 2010). We fit a linear model to measure the causal effects of time zone borders by modeling the distance to borders as discontinuous. In RDD, we have two groups based on a cut-off point—treated and control—and we aim to find the average treatment effect. Since we cannot observe two outcomes simultaneously for a given region at a given point of time, we estimate the average treatment effect to derive the causal effects on the outcome of interest.

We fit the following RD model:

$$Y_c = \beta_0 + \beta_1 h_c + \beta_2 f(d_c) + h_c \cdot \beta_3 f(d_c) + \beta_4 \delta_c + \varepsilon_i, \quad (1)$$

where Y_c is the outcome variable for a region c, $h_c \in \{0, 1\}$ is a dummy variable where 1 indicates the eastern side of a time zone boundary and 0 indicates the western side based on the distance d_c from the time zone boundary (positive d_c value is for the eastern side and negative d_c value is for the western side). The function $f(\cdot)$ is taken as polynomial (Gelman & Imbens, 2019). The control variable δ_c includes socio-demographic variables to improve the accuracy of estimates, and ε_i is the error term in the regression model. The treatment effect is β_1 which measures the effect size of the distance from time zone borders on the variable of interest. We pick the coefficient β_1 to refelct the treatment effect because β_1 captures the treatment at the discontinuity $d_c = c$.

3. Quantizer Design

To design optimal time zone borders, we consider the problem of partitioning the geographic distribution of the human population through the choice of time zone borders of maximum human well-being. Basically, we want as many people living at the eastern edges of their time zones so the sun sets earlier, they sleep more, and therefore have better wellbeing. So as not to have adverse circadian rhythm effects of mismatch between clock and sun, we also want to minimize the mismatch between the clock time and the solar time. Here, we measure the gain of optimal time zone regions on well-being by using social capital data.

Let us consider a quantizer that partitions the set of longitude lines $\mathcal{B}$ into K subsets, $\mathcal{B}_1, \ldots, \mathcal{B}_K$, called quantization regions. The regions are intervals bounded by time zone borders $b_i^{(k)}$, for K time zone borders, $k \in [0, K]$ and N longitude lines, $i \in [0, N]$. For example, the interval for region $\mathcal{B}_1 = (-\infty, b_i^{(1)}], \mathcal{B}_2 = (b_i^{(1)}, b_{i'}^{(2)}], \ldots, \mathcal{B}_K = (b_{i''}^{(K)}, \infty)$. For each quantization region $\mathcal{B}_k$, there is a representation point $r_i^{(k-1)}$ to which elements are mapped. The values $r_i^{(k-1)}$ are in the middle of the region k, and the values $b_i^{(k)}$ are at the boundaries of the region $\mathcal{B}_k$.

To achieve our goal of having optimal partitions of the human population, we measure the effect of a given time zone border position on the rest of the longitude lines. Let $\mathcal{D}(b_j, b_i^{(k)})$ be a function that measures the distortion on the population (at longitude line b_j) being at the eastern side of the time zone border $b_i^{(k)}$. Similarly, let $\mathcal{D}(b_j, r_i^{(k)})$ be the distortion from the circadian mismatch, such that the population at longitude line b_j are distant from the middle of the region. Therefore, we aim to minimize the distortion brought by the eastern edge effect $\mathcal{D}(b_j, b_i^{(k)})$ and the circadian rhythm effect $\mathcal{D}(b_j, r_i^{(k)})$ for the population at b_j having $b_i^{(k)}$ to be their time zone border. We approach the problem in three formulations of increasing intricacy, as follows.

Prime Meridian Choice The first optimization is that we search for a longitude line $b_i^{(k^o)}$ that acts as the reference time zone border for other borders. That reference time zone border $b_i^{(k^o)}$ minimizes the average distortion $\mathcal{D}(\cdot)$ for all other longitude lines. A direct optimization is to fix $K = 24$ and fix the quantizer to be uniform-sized such that regions are of size $720/24 = 30$, for 720 longitude lines. Therefore, the objective is to minimize the distortion of the eastern edge effect

$$\arg\min_{b_i^{(k^o)}} \mathbb{E}\left[\sum_{j=0}^{N} \mathcal{D}(b_j, b_i^{(k)})\right]. \quad (2)$$

Timezone Boundaries Choice In the second optimization, we allow regions to be of non-uniform size and search for the optimal boundaries of each time zone region. Given that the Earth is divided into 24 regions, each belonging to a time zone, we set $K = 24$ to find the optimal K par-

titions. For each $k \in K$, we find the time zone border $b_i^{(k)}$ and the corresponding representation point $r_{i'}^{(k)}$ that minimize the distortion at longitude lines in partition $\mathcal{B}_k$. Therefore, the objective is to find the set of optimal K time zone borders, $\{b_i^{(1)}, \ldots, b_{i'}^{(K)}\}$, and their corresponding representation (middle) points $\{r_n^{(0)}, \ldots, r_{n'}^{(K-1)}\}$.

$$\begin{aligned} \underset{\{b_i^{(1)},\ldots,b_{i'}^{(K)}\},\{r_n^{(0)},\ldots,r_{n'}^{(K-1)}\}}{\arg\min} & \mathbb{E}\left[\sum_{j=0}^{N} \mathcal{D}(r_j, r_n^{(k)})\right] \\ & + \lambda \mathbb{E}\left[\sum_{j=0}^{N} \mathcal{D}(b_j, b_i^{(k)})\right], \end{aligned} \tag{3}$$

where $i, i', j, n, n' \in [0, N]$. The quantity λ is meant to reflect the causal impact of current timezone borders on wellbeing. Therefore, the effect estimated β_1 from the current time zone borders on wellbeing is substituted by λ within the formulation.

Numbers and Boundaries of Time Zones The last formulation makes K open to optimization to find the optimal number of time zone borders.

$$\begin{aligned} \underset{K,\{b_i^{(1)},\ldots,b_{i'}^{(K)}\},\{r_n^{(0)},\ldots,r_{n'}^{(K-1)}\}}{\arg\min} & \mathbb{E}\left[\sum_{j=0}^{N} \mathcal{D}(r_j, r_n^{(k)})\right] \\ & + \lambda \mathbb{E}\left[\sum_{j=0}^{N} \mathcal{D}(b_j, b_i^{(k)})\right] \\ & + \eta K. \end{aligned} \tag{4}$$

3.1. Dynamic Programming Formulation

The first optimization in (2) is approached by enumeration while finding optimal partitions using (3) is approached through dynamic programming. Therefore, we extend a dynamic programming (DP) algorithm for optimal quantization due to Sharma (1978). A common approach for quantization is the Lloyd-Max algorithm (Max, 1960; Lloyd, 1982), but this does not work in our setting, since we aim to concentrate the probability mass at the right (eastern) edge of partition regions, b_i rather than largely being near the center r_i.

Recall N is the total number of longitude lines, with K time zone borders. We aim to find the optimal K-level quantizer, where $b_i^{(k)}$ acts as the time zone border of minimum distortion for region $\mathcal{B}_k$.

Specifically, to measure the distortion $\mathcal{D}(\cdot)$, consider two longitude lines, b_i and b_j, where b_i acting as a reference time zone border for b_j. We calculate the eastern edge effect to be the amount of population at longitude lines to the eastern of b_i up to b_j. We start by measuring the eastern edge effect $D_k(b_i, b_j)$ at just one quantization level, $k = 1$, i.e. one time zone region, assuming longitude lines belong to one segment as follows,

$$D_1(b_i, b_j) = \sum_{w=i}^{j} \rho Z_w \,, \tag{5}$$

where Z_w is the weighted population at longitude line b_w, and ρ is a scaling factor, e.g. 0.5, multiplied by the difference between b_i and b_j, to reflect the distance between b_i and b_j.

Next, we define the distortion at $2 \leq k \leq K$, when k segments are placed in the interval (b_i, b_j). For each k, we define a value $M_k = N - 2k$ to represent the end of the interval. For example, let $b_i^{(1)}, \ldots, b_j^{(K)}$ be the optimal solution, (optimal time zone borders), i.e. they represent the optimal K-level quantizer for the interval (i, M_K). With $j' < j$, $b_i^{(1)}, \ldots, b_{j'}^{(K-1)}$ must represent the optimal $(K-1)$-level quantizer for the interval (i, M_{K-1}). Thus, we can split the problem into sub-problems as:

$$\begin{aligned} D_k(b_i, b_{M_k}) = \min_{\alpha_m} & D_1(b_i, b_m) + D_{k+1}(b_m, b_{M_{k+1}}) \\ & \text{s.t.} \quad m \in (i, M_k), \end{aligned} \tag{6}$$

such that the longitude line b_m is between lines b_i and b_{M_k}. The value α_m is the minimum distortion attained within the interval (i, M_k) for time zone $\mathcal{B}_k$. Therefore, the optimal time zone border $b_m^{(k)}$ for time zone region $\mathcal{B}_k$ is

$$b_m^{(k)} = \arg\min_{b_m} D_k(b_i, b_{M_k}) \,. \tag{7}$$

We solve (6) up until $k+1 \leq K$. We follow a similar procedure to fine optimal representation points $r_n^{(k)}$ by replacing b_i with r_{i-1} in (5) and (6).

3.2. Reinforcement Learning Formulation

The computational complexity of the DP approach makes computation challenging for large values of K and N, e.g., $K = 24$. In DP we end up with a worst-case computation of $\mathcal{O}(N(KN^3))$.

Given the dependency in finding partitions in (6), the variables at k-level quantizer depend on future $(k+1)$-level quantizer results. Therefore, the quantization problem for partitioning longitude lines can be cast as a Markov Decision Process (MDP). (Details on MDPs in Appendix B). Commonly, DP solves MDPs by breaking them into smaller sub-problems. However, in large optimization problems, it is challenging to solve MDPs using DP (Powell, 2007). Reinforcement learning (RL) alternatively solves large MDPs in a way that combines optimization with simulation to approximate the optimal solution of MDPs (Mes & Rivera, 2017).

3.2.1. Quantization with Value Iteration

The *value iteration* (VI) algorithm, from RL, guarantees that the algorithm does not get locked into any hopelessly long sub-calculations before it can find the optimal value (Sutton & Barto, 2018).

We convert our modified version of Sharma's DP function (Sharma, 1978) in (6) into the VI function, we consider redefining the states, rewards, and actions based on our setting. Complete derivation in Appendix C. The final formulation we end up with is the following:

$$\mathcal{V}_t(b_i, k, b_{M_k}) = \min_{b_m \in \mathcal{A}(b_i,k,b_{M_k})} \mathcal{R}(b_i, b_m) + \gamma \mathcal{V}_{t-1}(b_m, k+1, b_{M_{k+1}}). \tag{8}$$

After convergence, the optimal policy π^* is the set of longitude lines $\{b_m^{(1)}, \ldots, b_{m'}^{(K)} \mid m, m' \in [0, N]\}$ which are the optimal boundaries for each time zone $\mathcal{B}_k$, at, and their corresponding representation points $\{r_n^{(0)}, \ldots, r_{n'}^{(K-1)} \mid n, n' \in [0, N]\}$.

4. Empirical Investigations

In this section, we demonstrate our approach for the time zone design problem so as to explore the possibility of improving social capital.

Social capital includes five different metrics: family unity, community health, institutional health, efficacy, and an overall index that averages the four sub-indices. Here we investigate the effect of sunlight amount determined by time zone borders on social capital across counties.

Table 1 first row reports the RDD estimates. We find a significant causal effect from sunset time on both the overall social capital index and the community health sub-index. We find that later sunset causes a lower social capital index. Thus, if the western counties near the time zone boundary were moved to the eastern side of the time zone boundary, we should expect a lower counterfactual prediction of the county index.

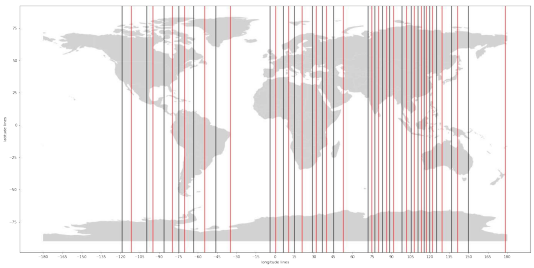
4.1. Quantization

Given our finding of causal significance between sunset and social capital in Section 4, we focus on those indices to demonstrate our quantization approach.

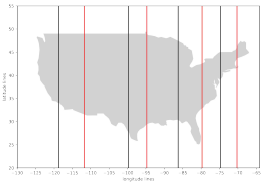
To partition regions, we base our partitioning on the population size at each longitude line (Rankin, 2008). We use the world population estimates at the west and east of each of the 360 longitude lines, with 15 degrees from both sides of a longitude line, hence, population sizes are observed at a total of 720 longitude coordinates/points. We use population data (Data & Center, 2000) calculated at each longitude line of the year 2000. At each time zone region, we aim to have a minimum population size at the east edge of a time zone border to minimize the distortion effect following (2)–(4).

4.1.1. Quantized Time Zone Borders

We apply our method and investigate the quantized partitioning by allowing the time zone regions to be of non-uniform widths. Figure 1a shows the optimal time zone boundaries redesigned by quantization over the world map for $K = 24$. We see that more time zone boundaries appear just to the east of regions with large populations, e.g. India and China. This follows our intuition for the need to redesign optimal time zone boundaries that minimize distortions across populations.



(a) Quantized World partitions



(b) Quantized Continental United States partitions

Figure 1. Time zone quantization for $K = 24$. (Black lines are the boundaries of a time zone and the red lines are the representation points)

4.2. Designed Discontinuity Counterfactual Prediction

After we have optimally redesigned time zone borders as in Figure 1a, with a focus on the United States in Figure 1b, we measure the counterfactual change in the causal impact of the new borders on social capital, as compared to the impact under the standard time zone borders. Focusing on the continental United States, we reuse the RDD model in (1), following the same causal inference analysis in 2. Counties are assigned to a time zone region with the minimum distance from a time zone border. Counties that are to the east of the time zone border have $h_c = 1$, to indicate the treatment, and hence, measure the treatment effect β_1 in (1).

	Effects of:	County Index	Family Unity	Community Health	Institutional Health	Efficacy
Later Sunset Counties	Current Timezone	$-1.194^{***}(0.419)$	$-0.491(0.540)$	$-1.644^{***}(0.558)$	$-0.694^{*}(0.395)$	$-0.519(0.447)$
	Redesigned Timezone	$0.083^{***}(0.024)$	$0.098^{***}(0.022)$	$-0.0443^{**}(0.019)$	$0.073^{**}(0.024)$	$0.047^{**}(0.019)$

Table 1. Counterfactual Prediction on Social Capital of the United States: Local non-parametric regression discontinuity estimates (In the later sunset counties row, first number represents the estimate, and shows the significance levels where $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$, and number in parenthesis displays the standard error).

The outcome of interest, Y_c, is the social capital index for county c, including the aggregated measure, county index, and subindex measures.

Table 1, second row, shows the effect of being at the eastern edge of the redesigned time zone border on social capital indices. In comparison to results in first row in Table 1, we see that the redesigned borders have yielded a stronger positive causal impact on social capital which mitigates the effect of distortion from the standard time zone borders.

5. Discussion

If we want as many people as possible living at the eastern edges of their time zones so the sun sets earlier, they sleep more, and therefore have more social capital (Putnam, 2000) (while ensuring the sun and the clock do not differ too much), can we design time zone boundaries to do so? For this problem and structurally-related ones that involve the design of discontinuities such as in health and education policies, we studied a problem formulation at the intersection of causal inference and quantization theory for the purpose of mechanism design. This led to new mathematical developments in linking regression discontinuity counterfactuals with optimal quantization theory. For the time zone problem specifically, results put time zone boundaries just to the east of large population centers, and we showed the possibility of significant gains in social capital.

There are natural questions of equity that arise through the design of discontinuities of the type we developed here.

References

Agarwal, A., Jiang, N., Kakade, S. M., and Sun, W. *Reinforcement learning: Theory and algorithms*. Working draft, 2019.

Angrist, J. D. and Pischke, J.-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ, USA, 2009.

Bellman, R. Dynamic programming. *Science*, 153(3731): 34–37, 1966.

Cappuccio, F. P., D'Elia, L., Strazzullo, P., and Miller, M. A. Sleep duration and all-cause mortality: A systematic review and meta-analysis of prospective studies. *Sleep*, 33(5):585–592, 05 2010. ISSN 0161-8105. doi: 10.1093/sleep/33.5.585.

Data, S. and Center, A. Gridded population of the world (gpw), 2000. https://sedac.ciesin.columbia.edu/data/set/gpw-v4-admin-unit-center-points-population-estima data-download.

Dell, M. and Querubin, P. Nation building through foreign intervention: Evidence from discontinuities in military strategies. *Quarterly Journal of Economics*, 133(2):701–764, May 2018.

Englmaier, F., Schmöller, A., and Stowasser, T. Price discontinuities in an online market for used cars. *Management Science*, 64(6):2754–2766, June 2018.

Gelman, A. and Imbens, G. Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456, 2019.

Gersho, A. and Gray, R. M. *Vector Quantization and Signal Compression*. Springer, 1991.

Gibson, M. and Shrader, J. Time use and labor productivity: The returns to sleep. *Review of Economics and Statistics*, 100(5):783–798, 2018.

Giuntella, O. and Mazzonna, F. Sunset time and the economic effects of social jetlag: evidence from us time zone borders. *Journal of Health Economics*, 65:210–226, 2019. doi: 10.1016/j.jhealeco.2019.03.007.

Gray, R. M. and Neuhoff, D. L. Quantization. *IEEE Transactions on Information Theory*, 44(6):2325–2383, October 1998. doi: 10.1109/18.720541.

Hamermesh, D. S., Myers, C. K., and Pocock, M. L. Cues for timing and coordination: latitude, letterman, and longitude. *Journal of Labor Economics*, 26(2):223–246, 2008.

Imbens, G. W. and Lemieux, T. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

Lee, D. S. and Lemieux, T. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355, 2010.

Lee, M. The geography of social capital in America. Social Capital Project report, United States Senate Joint Economic Committee, April 2018.

Lloyd, S. P. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982. doi: 10.1109/TIT.1982.1056489. also unpublished Bell Laboratories Memorandum, July 31, 1957.

Max, J. Quantizing for minimum distortion. *IRE Transactions on Information Theory*, 6(1):7–12, 1960.

Mes, M. R. and Rivera, A. P. Approximate dynamic programming by practical examples. In *Markov Decision Processes in Practice*, pp. 63–101. Beta Working Paper Series 495, 2017.

Misra, V., Goyal, V. K., and Varshney, L. R. Distributed scalar quantization for computing: High-resolution analysis and extensions. *IEEE Transactions on Information Theory*, 57(8):5298–5325, August 2011. doi: 10.1109/TIT.2011.2158882.

Park, A., Shi, X., tai Hsieh, C., and An, X. Magnet high schools and academic performance in China: A regression discontinuity design. *Journal of Comparative Economics*, 43(4):825–843, November 2015.

Pearl, J. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009. doi: 10.1214/09-SS057.

Pearl, J. Causal inference. In *Proceedings of Workshop on Causality: Objectives and Assessment*, pp. 39–58, 2010.

Poor, H. V. and Thomas, J. Applications of Ali-Silvey distance measures in the design generalized quantizers for binary decision systems. *IEEE Transactions on Communications*, 25(9):893–900, September 1977. doi: 10.1109/TCOM.1977.1093935.

Powell, W. B. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley, 2007.

Putnam, R. *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, 2000.

Rankin, W. Population histograms. Radical Cartography, 2008. URL http://www.radicalcartography.net/index.html?histpop.

Rivera, L. A. and Tilcsik, A. Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation. *American Sociological Review*, 84(2):248–274, 2019.

Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. doi: 10.1037/h0037350.

Sharma, D. Design of absolutely optimal quantizers for a wide class of distortion measures. *IEEE Transactions on Information Theory*, 24(6):693–702, 1978.

Shlezinger, N., Eldar, Y. C., and Rodrigues, M. R. D. Hardware-limited task-based quantization. *IEEE Transactions on Signal Processing*, 67(20):5223–5238, October 2019. doi: 10.1109/TSP.2019.2935864.

Srikant, R. Lecture notes in MDPs and reinforcement learning, 2022.

Sutton, R. S. and Barto, A. G. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA, 2018.

Szepesvári, C. Lecture notes in theoretical foundations of reinforcement learning, 2020.

Varshney, K. R. and Varshney, L. R. Quantization of prior probabilities for hypothesis testing. *IEEE Transactions on Signal Processing*, 56(10):4553–4562, October 2008. doi: 10.1109/TSP.2008.928164.

Varshney, L. R. and Varshney, K. R. Decision making with quantized priors leads to discrimination. *Proceedings of the IEEE*, 105(2):241–255, February 2017. doi: 10.1109/JPROC.2016.2608741.

A. Background: Causal Inference

In engineering, health, and social science fields, the randomized experiment has played important roles to uncover causal effects under a given intervention on an outcome of interest. For example, to study the efficacy of a new drug, one can randomly assign patients to two groups where one group receives the new drug and the other receives a placebo. By comparing the difference in efficacy between the two groups, one can estimate the treatment effect. Although randomized experiments have been robust in producing estimates and simple interpretations, it is often hard to apply to real-world applications given practical and ethical limitations in randomly assigning groups and interventions. To overcome such limitations, non-experimental designs have been developed to uncover causal effects.

One of the main distinctions between standard statistical analysis and causal inference is dealing with changing conditions. Statistical analysis, represented by regression and hypothesis testing techniques, estimates beliefs from past to future as long as experiment conditions do not change, whereas causal inference infers beliefs under changing conditions to uncover causal relationships among variables (Pearl, 2009). Several frameworks are used for causality analysis, such as structural models (Pearl, 2010) and the potential outcome framework (Rubin, 1974), which we focus on here.

The potential outcome framework assumes effects are tied to a treatment or an intervention. To reveal the causal effects of an intervention, (Rubin, 1974) proposed to measure the difference between two potential outcomes; let us denote them as Y^N and Y^I, for a given unit x. The potential outcome Y^N is the outcome for x without being exposed to an intervention, and Y^I is the outcome after an intervention is applied on x. So, the causal effect is

$$\tau = Y^N - Y^I \,. \tag{9}$$

However in real applications, we can never observe both outcomes for the same unit under the same conditions, only one of the two will take place at a given time. Since one of the potential outcomes will always be unavailable, the core objective of the framework is to estimate it.

Let us introduce the main terms in the potential outcomes literature, which are used throughout. A *unit* is the atomic object in the framework, which can be a city or a county. A *treatment* is the action applied to a unit to change its state. The treatment [1] can be a medicine given to a particular group. The treatment is usually thought of as binary, so one group receives the treatment (the *treated* group), and the other does not (the *control* group). One commonly used design under the potential outcomes framework is regression discontinuity, which we consider here.

A.1. Regression Discontinuity

The RDD has recently gained attention in social science because it provides the most credible analysis of causal effects with relatively mild assumptions compared to other non-experimental designs such as instrumental variables (IVs). In RDD, each observation can be split into two groups based on a known discontinuous variable—a cut-off point for an intervention of interest (Imbens & Lemieux, 2008). Suppose observations cannot perfectly manipulate the intervention; then the difference between two groups near the known cut-off point can be used to measure the local average treatment effect on an outcome of interest. Therefore, RDD does not require researchers to explicitly randomize the treatment, yet it gives the comparative analysis of the causal effects as the randomized experiment.

The causal estimate arises from the comparison between both groups, in which the distribution of samples below and above the threshold is expected to be different if an intervention had an impact on treated samples. However, to ensure the validity of the estimate, the distribution of characteristics around the threshold should not change discontinuously (Lee & Lemieux, 2010).

B. Markov Decision Process

MDP is a controlled Markov chain, in which the transition from one state to another depends on an external control parameter called the *action*. Specifically, the probability of transitioning to state s' from state s upon taking action a is denoted by $P_{s,s'}(a)$ and given by

$$\begin{aligned} \mathcal{P}(s, s', a) := P(s_1 = s' | s_0 = s, A_0 = a), \\ \forall s, s' \in \mathcal{S}, \forall a \in \mathcal{A}(s) \,. \end{aligned} \tag{10}$$

[1] The terms *treatment* and *intervention* are used interchangeably.

In addition to the finite state space $\mathcal{S}$, for any state $s \in \mathcal{S}$, MDP has a finite action space $\mathcal{A}(s)$ of possible actions that can be taken at state s. Another component of an MDP is an underlying one-step *reward* function $\mathcal{R}(s, a)$ that assigns random rewards to each (s, a) pair. Therefore, an MDP is given by a quadruplet $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$. The goal in analyzing an MDP is to find the optimal *policy* π_t^* at different time steps. More formally, a policy is a probability distribution over the action space $\mathcal{A}$:

$$\pi_t(a|\{s_i, a_i\}_{i=0}^{t-1}, s_t) := P(\mathcal{A}(s_t) = a|s_0, a_0, \ldots, s_{t-1}, a_{t-1}, s_t). \tag{11}$$

In words, the current policy π_t at current time t depends on a sequence of previous states and actions from $t = 0$ up to $t = t - 1$ and current state s_t in an MDP.

Commonly, DP solves MDPs by breaking them into smaller sub-problems. However, in large optimization problems, it is challenging to solve MDPs using DP (Powell, 2007). Reinforcement learning (RL) is an alternative to solve large MDPs in a way that combines optimization with simulation to approximate the optimal solution of MDPs (Mes & Rivera, 2017).

C. Value Iteration

One RL technique to solve an MDP is the *value iteration* (VI) algorithm, in which all states are updated in random order at one iteration. It is also based on a one-step look-ahead search from the current state. Therefore, this guarantees that the algorithm does not need to get locked into any hopelessly long sub-calculations before it can find the optimal value (Sutton & Barto, 2018).

The VI function $\mathcal{V}(\cdot)$ basically uses the Bellman optimality equation (Bellman, 1966) with an update rule at each iteration from $t = 1$ to T. Each state $s \in S$ has a value $\mathcal{V}_t(s)$ that is updated using the previous value function of the next state $s' \in S$ with a non-negative reward function $R(s, a)$ to dictate the best action to take over all possible actions $a \in \mathcal{A}(s)$ as in (10). The reward indicates what is the good action to take in an immediate sense, whereas a value function indicates what is the best, in the long run, (Sutton & Barto, 2018). Formally, the value function iterates from t to a total of T and is expressed as follows,

$$\mathcal{V}_t(s) = \max_{a \in \mathcal{A}(s)} \mathcal{R}(s, a) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}(s, s', a)\mathcal{V}_{t-1}(s'). \tag{12}$$

The parameter γ is a *discount* factor, $0 < \gamma < 1$ which is used to make an infinite sum finite when $T \to \infty$. Therefore, this also imposes convergence guarantees as proved in (Sutton & Barto, 2018; Agarwal et al., 2019).

Specifically, running the value function in (12) for T iterations such that

$$T \geq \frac{\log(||\mathcal{V}_1 - \mathcal{V}_0||_\infty) + \log(2) - \log(\epsilon(1 - \gamma))}{\log(\frac{1}{\gamma})} \tag{13}$$

ensures that the optimal value function $\mathcal{V}^*$ is $||\mathcal{V}_T - \mathcal{V}^*||_\infty < \epsilon$, therefore, $\pi_T \to \pi^*$, the optimal policy. For the theorems and derivations that lead to this result, please see (Agarwal et al., 2019; Szepesvári, 2020; Srikant, 2022)

C.0.1. Quantization with Value Iteration

To convert our modified version of Sharma's DP function (Sharma, 1978) in (6) into the VI function, we consider redefining the states, rewards, and actions based on our setting. First, we rewrite (12) with equal transition probabilities and with a minimization objective as follows:

$$\mathcal{V}_t(s) = \min_{a \in \mathcal{A}(s)} \mathcal{R}(s, a) + \gamma \mathcal{V}_{t-1}(s'). \tag{14}$$

Then, we build the MDP using the set of longitude lines $\mathcal{B}$ and the number of time zones K as follows

$$\begin{aligned} s &:= (b_i, k, b_{M_k}) \\ a &:= b_m \\ \mathcal{A}(s) &:= f(b_i, k, b_{M_k}) \\ \mathcal{R}(s, a) &:= D_1(b_i, b_m) \\ s' &:= (b_m, k + 1, b_{M_{k+1}}). \end{aligned} \tag{15}$$

The function $f(b_i, k, b_{M_k})$ returns actions $\mathcal{A}(s)$ that satisfy the constraints on b_m in (6).

Therefore, combining (6) and (14) with change in notation we get

$$\mathcal{V}_t(b_i, k, b_{M_k}) = \min_{b_m \in \mathcal{A}(b_i,k,b_{M_k})} \mathcal{R}(b_i, b_m) + \gamma \mathcal{V}_{t-1}(b_m, k+1, b_{M_{k+1}}). \tag{16}$$

Notice that if V_t converges before T, that is $\mathcal{V}_{T-1} = \mathcal{V}_T$, then it satisfies the DP in (6)

$$\mathcal{V}_T(b_i, k, b_{M_k}) = \min_{b_m \in \mathcal{A}(b_i,k,b_{M_k})} \mathcal{R}(b_i, b_m) + \gamma \mathcal{V}_T(b_m, k+1, b_{M_{k+1}}). \tag{17}$$

D. Data and Pre-processing for Regression Discontinuity Models

Let us initially consider the continental United States. Following (Giuntella & Mazzonna, 2019), we consider the distance between the centroid of a region and the time zone boundary to calculate the daylight hours for a given region. We use Census center of population and time zone boundary data from the Bureau of Transportation Statistics to compute the distance between the centroid of the county and adjacent time zone boundaries.

Census Centers of Population The Census centers of population dataset[2] provides the balance point of various geographic and demographic features. We especially use the centers of the population by county data to obtain coordinates of counties in the continental U.S. With this dataset, we compute the average sunset time given year and distance to the time zone boundary for all counties.

Time Zone Boundary U.S. time zone boundaries and daylight saving time (DST) are managed by the Department of Transportation. There are four time zones in the continental U.S.: Pacific (UTC $-$07:00), Mountain (UTC $-$06:00), Central (UTC $-$05:00) and Eastern (UTC $-$04:00). There is a one hour difference between each time zone. As time zones in the U.S. are not strictly based on mean solar time at the meridian, we use the shape file provided by the Bureau of Transportation Statistics. With coordinates of counties from Census data, we compute the distance between adjacent time zone boundaries and centroids of counties using Euclidean distance. Since not all states observe DST or have a consistent time zone over years, we exclude counties in Arizona, Florida, and Indiana from our analysis. The distance from the time zone boundary and the current year are used to calculate the average sunset time for a given region.

Social Capital The social capital measure is obtained from the Social Capital Project (Lee, 2018) and comprises *family unity*, an indicator of the structure of families in terms of marriage and children; *community health*, an indicator of participation in civic life; *institutional health* that considers confidence in media/corporations/schools and participation in institutions such as elections and census; and *collective efficacy*, an indicator for the converse of social disorganization, operationalized via violent crime rates.

[2] https://www.census.gov/geographies/reference-files/time-series/geo/centers-population.html