

Free Discontinuity Regression

With an Application to the Economic Effects of Internet Shutdowns

Florian Gunsilius¹ and David Van Dijke^{2,3}

¹Department of Economics, Emory University

²Department of Economics, University of Michigan

³Risk Analytics Division, Ipsos Public Affairs

September 16, 2024

Abstract

Discontinuities in regression functions can reveal important insights. In many contexts, like geographic settings, such discontinuities are multivariate and unknown a priori. We propose a non-parametric regression method that also estimates the location and size of discontinuities by segmenting the regression surface. In contrast to thresholding approaches, which are well-known to misclassify discontinuities in settings with noise, we prove identification of and convergence to the true location and size of the discontinuities for the proposed method. We use the estimator to show that an internet shutdown in India resulted in a reduction of economic activity by 25–35%, greatly surpassing previous estimates and shedding new light on the true cost of such shutdowns for digital economies globally.

Keywords: bounded variation, calibration, free discontinuity problem, Γ -convergence, internet shutdowns, Mumford-Shah functional, nonparametric regression, segmentation.

JEL Codes: C14, C21, L86, O33

{ffg;dvdijke}@umich.edu. Authors listed alphabetically. We thank Austin Wright as well as audiences at the Bank of England, Emory, MIT, Northwestern, Princeton, Simon Fraser, UIUC, the University of Michigan, Yale, the 2023 Munich Econometrics conference, the 2023 European Winter Meetings of the Econometric Society, and the 2024 ESIF Economics and AI+ML Meeting for helpful comments and discussions. The majority of this work was finished while the first author was affiliated with the University of Michigan and Princeton University. The first author gratefully acknowledges support from a MITRE award at the University of Michigan. This research was supported in part through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor. All errors are the authors'. Python code for this paper can be found at <https://github.com/Davidvandijke/fdr>.

1 Introduction

In many settings of interest, discontinuous changes in regression surfaces can reveal fundamental insights into the underlying problem structure. In such cases, the location and size of these discontinuities are fundamentally unknown to the researcher. Often, these discontinuities are functions of many variables. Examples include, but are not limited to, discontinuities induced by confidential or proprietary treatment algorithms such as loan or financial aid disbursement (Argyle et al., 2020; Carneiro et al., 2019), school admissions (Brunner et al., 2021), customer segmentation for marketing (Kuo et al., 2002; Hartmann et al., 2011), or the bandit algorithms that are used widely in the recommendation systems of tech companies (Li et al., 2010; Narita et al., 2023); by tipping points (Lamberson and Page, 2012) in racial segregation (Card et al., 2008), hospital stress (Kuntz et al., 2015), and complex systems such as the climate (Scheffer et al., 2009) or financial markets (Hansen, 2017); as well as geographic discontinuities induced by, for example, neighborhood-based segregation (Caetano and Maheshri, 2017), broadcast signal penetration (Sonin and Wright, 2022; Kern and Hainmueller, 2009; Olken, 2009) or epidemic spread (Ambrus et al., 2020).

This article introduces a disciplined statistical method that estimates a regression surface together with the location as well as the size of the discontinuities if they exist. It does so by estimating and segmenting the regression surface into smooth and discontinuous parts. Crucially, it does not require pointwise smoothness assumptions on the shape of the discontinuity set. For this approach, neither the location nor the size of the discontinuity set need to be known a priori. In particular, the segmentation is what sets this approach apart from other estimation procedures like wavelet estimation or thresholding approaches. The segmentation not only provides an approximation to the true regression, but does so while explicitly estimating the location and size of the discontinuities, balancing smoothing and thresholding.

To this end, we cast ideas from the theory of free discontinuity problems (e.g. Ambrosio et al., 2000) and computer vision, in particular the Mumford-Shah (MS) functional (Mumford and Shah, 1989), in a statistical setting. The MS functional was developed in the image processing literature to segment images into smooth and discontinuous parts. As such, the functional addresses the shortcomings of more

simplistic edge detection (gradient thresholding) (Canny, 1986) or “active contours” approaches (Kass et al., 1988). It does so by jointly segmenting and denoising the image, while using global information to incorporate spatial regularity into the segmentation and avoid the spurious artifacts often generated by these simpler thresholding methods (Chan and Shen, 2005, Ch. 7.2). Indeed, in the presence of noise, a simple thresholding approach will almost surely mis-classify the discontinuities (Chan and Shen, 2005, §7.2.1), which leads to severely biased estimates in a statistical setting. Thus, such an approach would always require a pre-processing step to smooth out noise. This is rather difficult, as one needs to pick the “correct” amount of smoothing in finite samples. By contrast, the proposed estimates and segments the regression surface in one step, which is more amenable to statistical analysis. For empirical results illustrating the superior performance of the MS functional compared to simpler edge detection or denoising methods, see e.g. Wang et al. (2012); Chan and Vese (2001); Lucas et al. (2022); Strelakovski et al. (2012).

The MS functional is not convex, so its critical points need not be global optima. This poses a threat to the reproducibility of estimates of economic interest. To circumvent this, we convexify the MS functional using the method of calibrations (Alberti et al., 2003), which lifts the problem to a higher-dimensional space. This convex relaxation provides global solutions and is efficiently implementable via a primal-dual algorithm (Chambolle and Pock, 2011). It has been successfully applied in a variety of settings in image recognition and computer vision (Pock et al., 2009, 2010), see Vogt et al. (2020) for a recent overview.

We extend this convexification of the MS functional to a statistical setting where both the location as well as the outcome values of observations are random, which we call the “Free Discontinuity Regression” (FDR) estimator. We define an empirical estimator for this setting and provide identification as well as convergence results in any dimension. These complement and extend recent mathematical results for discrete approximations of the classical non-convex MS functional (Caroccia et al., 2020; Ruf, 2019; Morini, 2002; Richardson, 1992) and related functionals (Chambolle and Pock, 2021; García Trillos and Slepčev, 2016; García Trillos and Murray, 2017; Hütter and Rigollet, 2016; Hu et al., 2022).

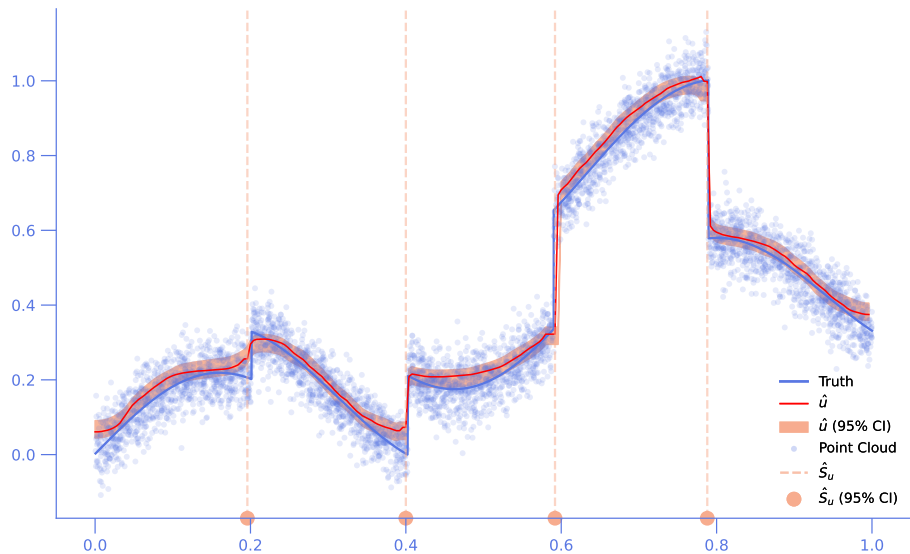
The resulting estimator is illustrated for simulated examples in 1D to 3D in Figure

1. The figure shows the true underlying functions and randomly sampled, noisy point clouds together with our resulting estimated functions and jump locations. In all examples, our method accurately reproduces the true function and jump set. We discuss these examples in more depth in Section 3.4 below.

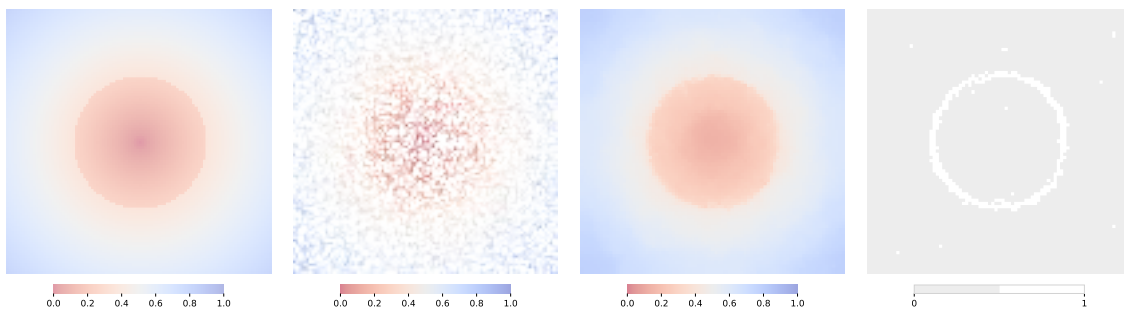
We apply the proposed method to estimate the economic effects of an internet shutdown in the state of Rajasthan in 2021. This one-day shutdown was implemented across the state to prevent cheating on a state exam. The shutdown was induced directly at the device level, which leads to sharp discontinuities in the observed internet connection signal. Its arbitrary nature provides a setting for estimating the effects of internet shutdowns on economic activity. To proxy for economic activity, we estimate anonymized mobile activity around economic points of interest like amenities, shops, and offices. For that, we rely on the cached location data that mobile apps with background location permissions collect from satellite data so that we can observe limited mobile activity even inside the shutdown region. Along the boundary of the shutdown region, we estimate effects that suggest a 25–35% reduction in economic activity, greatly exceeding previous estimates. These results highlight the asymmetry between internet expansion and internet shutdowns and call attention to the critical importance of a reliable digital communications infrastructure in modern economies.

The proposed method is most closely related to change point detection literature. Generally, existing detection methods have focused on estimating the location of the unknown discontinuities and are less concerned with estimating jump sizes (e.g. Page, 1954; Killick et al., 2012; Porter and Yu, 2015; Donoho and Johnstone, 1994; Harchaoui et al., 2008). The estimation of structural breaks in time series is a well-developed research area (Andrews, 1993; Bai and Perron, 2003; Delgado and Hidalgo, 2000), which has largely focused on functions with time as an input variable.

Recent papers have proposed extensions to functions on a multivariate domain, under various assumptions. Park (2022) and Herlands et al. (2019) model the regression surface using Gaussian processes; Zhu et al. (2014) consider a spatially varying coefficient model with jumps discontinuities without explicitly estimating those discontinuities. Madrid Padilla et al. (2022) propose a functional binary seeded segmentation algorithm for change point detection in multivariate domains but do not estimate the regression function. Herlands et al. (2018) take a data-mining approach



(a) 1D: Truth + Point Cloud + \hat{u} + Boundary

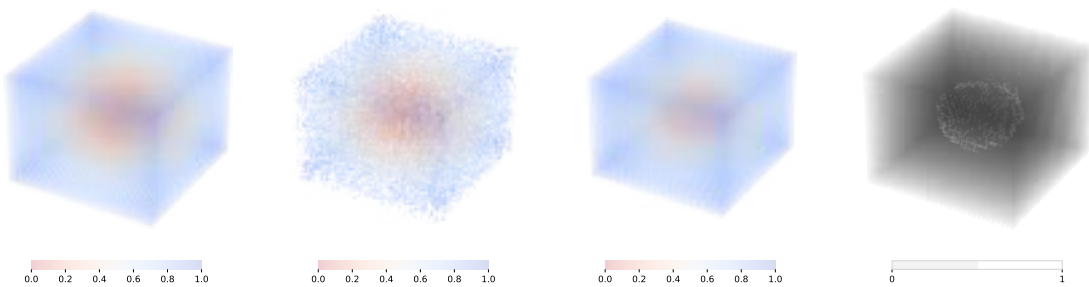


(b) 2D: Truth

(c) 2D: Point Cloud

(d) 2D: \hat{u}

(e) 2D: Boundary



(f) 3D: Truth

(g) 3D: Point Cloud

(h) 3D: \hat{u}

(i) 3D: Boundary

Figure 1: Simulations: 1D to 3D

Notes: figures depict, for 1D–3D; the true underlying function without noise, the random point cloud with Gaussian noise ($\sigma = 0.05$), the estimated response function \hat{u} and the estimated boundary locations. 1D case also includes red shaded 95% confidence bands estimated by subsampling with an estimated rate of convergence (Politis et al., 1999), and red dots below the plot to indicate jump sizes that are different from 0 at 95% level. Hyperparameters for 1D, 2D chosen by SURE. *Details:* **1D** jump sizes: 0.1286, 0.2133, 0.3192, -0.4220 ($d = 0.2823, 0.4682, 0.7005, -0.9262$), $n=5,000$, $\lambda = 98.6712, \nu = 0.0001$. **2D** jump size $\alpha = 0.1126$ ($d = 0.75$), $n=10,000$; $\lambda = 48.3921, \nu = 0.0012$; **3D** jump size $\alpha = 0.0738$ ($d = 0.5$), $n=50,000$; $\lambda = 10, \nu = 0.008$.

to search for discontinuities in units’ treatment status in any dimension, but require the treatment status to be observed, and do not establish the statistical consistency of their algorithm. Several papers in economics have also proposed estimators for multivariate discontinuities when their location is known (Cheng, 2023; Narita and Yata, 2021; Abdulkadiroğlu et al., 2022; Cattaneo and Titiunik, 2022; Keele and Titiunik, 2015).

A more classical literature deals with estimating regression surfaces with jumps (Qiu and Yandell, 1997; Korostelev and Tsybakov, 1993; O’Sullivan and Qian, 1994; Muller and Song, 1994; Donoho, 1999; Li and Ghosal, 2017; Qiu, 1998). These papers analyze multidimensional jump estimation under specific assumptions on the underlying process, such as that the number of partitions of the domain is known, or that the domain is of a specific dimension (generally $d \leq 3$). Finally, the “fused lasso” literature deals with approximating piecewise constant functions on potentially multivariate domains by imposing sparsity of both the coefficients and their differences in a high-dimensional means problem (Tibshirani et al., 2005; Rinaldo, 2009; Harchaoui and Lévy-Leduc, 2010). Our approach focuses on segmentation and estimates both the jump locations and size under no additional smoothness assumptions on the discontinuity set.

2 Free Discontinuity Regression

In this section, we introduce the regression framework and the statistical estimator for our method. All mathematical notation is explained in Appendix A.

2.1 Regression framework

We assume we have n randomly sampled units indexed by $i = 1, 2, \dots, n$ for which we observe a potentially multivariate regressor $X_i \in \mathbb{R}^d$, and an outcome of interest $Y_i \in \mathbb{R}$. The regression model is

$$(1) \quad Y_i = f(X_i) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i | X_i] = 0,$$

where the ε_i are standard independent and identically distributed unobservable error terms.

Our goal is to estimate the regression surface $f(x)$ including the location and jump sizes of the discontinuity set S_f of f . Since the regressors can be multivariate, we need to extend the notion of a univariate discontinuity to a multivariate setting; the following definition (e.g. [Ambrosio et al., 2000](#), Definition 3.67) provides this.

Definition 1 (Approximate jump points). *Let u be locally integrable on \mathcal{X} , that is $u \in L^1_{loc}(P_X)$, where P_X denotes the law of X . We say that a point $x \in \mathcal{X}$ is an approximate jump point of u if there exist constants $a, b \in \mathbb{R}$ and an orientation $\rho \in \mathbb{S}^{d-1}$ such that $a \neq b$ and*

$$(2) \quad \lim_{\varepsilon \downarrow 0} \frac{\int_{B_\varepsilon^+(x, \rho)} |u(x') - a| dP_X(x')}{P_X(B_\varepsilon^+(x, \rho))} = 0, \quad \lim_{\varepsilon \downarrow 0} \frac{\int_{B_\varepsilon^-(x, \rho)} |u(x') - b| dP_X(x')}{P_X(B_\varepsilon^-(x, \rho))} = 0,$$

where

$$\begin{cases} B_\varepsilon^+(x, \rho) := \{x' \in B_\varepsilon(x) : \langle x' - x, \rho \rangle > 0\} \\ B_\varepsilon^-(x, \rho) := \{x' \in B_\varepsilon(x) : \langle x' - x, \rho \rangle < 0\} \end{cases}$$

denote the two half balls, oriented by ρ , contained in the ε -ball $B_\varepsilon(x)$, and \mathbb{S}^{d-1} is the unit sphere in \mathbb{R}^d . The triplet (a, b, ρ) , uniquely determined by (2) up to a permutation of (a, b) and a change of sign of ρ , is denoted by $(u^+(x), u^-(x), \rho_u(x))$; with u^+, u^- called the traces of u . The approximate discontinuity set contains all approximate jump points and is denoted by S_u .

In this definition, the univariate idea of approaching a function from the “left” or “right” is generalized by means of an orientation ρ which is aligned with the boundary and which orients two epsilon half-balls according to the angle between the balls and the boundary, as illustrated in [Figure 3](#).

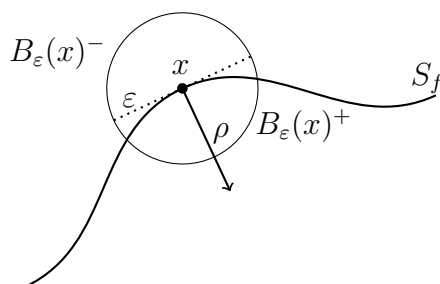


Figure 3: Discontinuity Set S_f for a Function f

2.2 FDR as an estimator based on the Mumford-Shah Functional

We now introduce the statistical estimator. Recall that the idea is to estimate the entire regression surface $f(x) = \mathbb{E}[Y_i|X_i = x]$ including the location and size of the discontinuities, if they exist. For this, standard smoothing and thresholding approaches are not well-suited for several reasons. First, smoothing approaches do not estimate the size or location of the discontinuity, they simply provide an approximation to the regression surface that preserves discontinuities as they would smooth over discontinuities. Second, they yield downward-biased jump sizes in practice (e.g. [Caroccia et al., 2020](#), Fig. 7.1f). Third, simple thresholding approaches that estimate jump locations by considering the steepness of the gradient are known to almost surely misclassify discontinuities, which leading to severely biased estimates ([Chan and Shen, 2005](#)). The MS functional approach circumvents these issues.

2.2.1 The Mumford-Shah Functional

The proposed estimator is based on the MS functional ([Mumford and Shah, 1989](#)). Throughout, we assume that the random variable X has a density f_X . Under this assumption, the MS functional takes the following form:

$$(3) \quad E(u) = \lambda \underbrace{\int_{\mathcal{X}} (f(x) - u(x))^2 f_X(x) dx}_{\text{Regression}} + \underbrace{\int_{\mathcal{X} \setminus S_u} |\nabla u(x)|^2 f_X(x) dx}_{\text{Roughness penalty away from discontinuity}} + \underbrace{\nu \mathcal{H}^{d-1}(S_u)}_{\text{Discontinuity regularity penalty}},$$

which consists of three terms: 1) a non-parametric regression term, which minimizes the mean-squared error between the true function f and its approximation u ; 2) a “roughness penalty”, which penalizes the size gradient of the estimated function u away from the discontinuity set S_u ; 3) a “boundary regularity penalty”, which penalizes the “size”, more formally the $(d-1)$ -dimensional Hausdorff measure \mathcal{H}^{d-1} , of the discontinuity set. The parameters $\lambda, \nu \geq 0$ control the smoothness of u and the expressiveness of the boundary S_u . The weight on the gradient term is normalized to 1; $|\cdot|^2$ denotes the squared Euclidean norm in \mathbb{R}^d .

2.2.2 Convexification Through Calibrations

The MS functional (3) is non-convex due to the inclusion of the second penalty term. This is a threat to consistent estimation of the regression function, as minimization routines need not converge to global optima. In practice, this expresses itself by artifacts, meaning spuriously estimated boundaries and discontinuities (Brown et al., 2012, Fig. 4-5; Pock et al., 2009, Fig. 5). A common approach for non-convex objective functions is to re-estimate the model many times for various initializations of the solution and pick the solution that delivers the smallest objective function or some other goodness of fit measure. Apart from being computationally intensive, such an approach is likely to fail for functional minimands or real-valued vectorial minimands with dimension larger than 1, as then the set of potential initializations is too large.

To circumvent this, we rely on the method of calibrations introduced in Alberti et al. (2003) to convexify the MS functional. The idea is to lift the problem to a higher dimension by considering the graph Γ_f of f and finding a vector field $p : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^{d+1}$ on the lifted space which maximizes the flux through the graph.

Formally, the convex relaxation in our setting reads

$$E(u) = \sup_{p \in K} \int_{\mathcal{X} \times \mathbb{R}} p \cdot D\mathbb{1}_u$$

with a convex set

$$K = \left\{ p \in C_0(\mathcal{X} \times \mathbb{R}, \mathbb{R}^{d+1}) : \right. \\ \left. p^t(x, t) \geq \frac{|p^x(x, t)|^2}{4f_X(x)} - \lambda f_X(x)(t - f(x))^2, \quad \left| \int_{t_1}^{t_2} p^x(x, s) ds \right| \leq \nu, \quad \forall t_1, t_2 \in \mathbb{R} \right\}.$$

The derivation is similar to the derivation of the calibration under Lebesgue measure in Alberti et al. (2003) and is therefore omitted.

Here, t indexes the lifted dimension, $\mathbb{1}_u(x, t)$ is the indicator function of the subgraph of $u(x)$, which is 1 if $t < u(x)$ and 0 otherwise; $D\mathbb{1}_u := (D_1\mathbb{1}_u, \dots, D_{d+1}\mathbb{1}_u)$ is a $(d + 1)$ -dimensional Radon measure, $C_0(\mathcal{X} \times \mathbb{R}, \mathbb{R}^{d+1})$ denotes the space of all continuous functions from $\mathcal{X} \times \mathbb{R}$ to \mathbb{R}^{d+1} that vanish at infinity, and $p \cdot D\mathbb{1}_u$ denotes the scalar product. $p^x(x, t)$ denotes the first d dimensions of the vector p indexed by

x and $p^t(x, t)$ denotes the last dimension.

Although the objective function is now convex, the set of indicator functions of subgraphs of u over which we optimize is not. To make the problem fully convex, we follow [Pock et al. \(2009\)](#) and further relax the problem by optimizing over general functions mapping to the unit interval

$$v(x, t) \in C := \left\{ v \in SBV(\mathcal{X} \times \mathbb{R}, [0, 1]) : \lim_{t \rightarrow -\infty} v(x, t) = 1, \lim_{t \rightarrow +\infty} v(x, t) = 0 \right\}.$$

Here, $SBV(\mathcal{X} \times \mathbb{R}, [0, 1])$ denotes the space of special functions of bounded variation on $\mathcal{X} \times \mathbb{R}$ mapping to the unit interval.

The class of functions of bounded variation is large and in particular includes functions with discontinuities, which makes it well-suited for the purpose of estimating regressions surfaces with discontinuities. With these definitions, we can state the following optimization problem, which we define as the Free Discontinuity Regression estimator.

Definition 2. *The Free Discontinuity Regression estimator is the 0.5-level set of the optimizer v^* of the optimization problem*

$$(4) \quad \inf_{v \in C} E(v) := \inf_{v \in C} \sup_{p \in K} \langle p, Dv \rangle \equiv \inf_{v \in C} \sup_{p \in K} \int_{\mathcal{X} \times \mathbb{R}} p \cdot Dv.$$

The optimization problem (4) is well-behaved in the sense that it is a convex problem; moreover, we now show that it always admits a solution v^* if we assume that $|Dv| \leq c$ for some c , i.e. that we only consider bounded variation functions. This assumption is weak as v is an approximation of an indicator function, which by definition has a variation of 1, see also the discussion below after [Theorem 2](#).

Proposition 1. *The optimization problem (4) admits a global solution v^* in $SBV(\mathcal{X} \times \mathbb{R}) \cap \{v : |Dv| \leq c\}$ for fixed $c < +\infty$ if $f \in SBV(\mathcal{X})$.*

The solution obtained will be a general function v that need not take the form of an indicator function, although we prove below that it will, in a certain topology, do so as $\lambda \rightarrow +\infty$ ([Theorem 1](#)).

We also observe in practice that, at least for the hyperparameters selected by our automatic method introduced below—which usually sets λ large—the solution

obtained is an indicator function. To obtain the underlying function u , we follow Pock et al. (2009) and threshold v at the 0.5-level set. The choice of 0.5 is arbitrary insofar as, if v is indeed an indicator function, we can use any value $t \in (0, 1)$ to threshold it; choosing the midpoint is natural in cases where the solution is not exactly equal to an indicator function.

3 Statistical Properties

In this section, we provide mathematical and statistical properties of the optimization problem (4). We only state the results and defer all proofs to Appendix B. First, we show that for fixed $\nu > 0$ and in the limit as $\lambda \rightarrow +\infty$ the solution to (4) picks out the correct discontinuity set and jump sizes, under mild assumptions. Then we introduce the empirical analogue of our population estimator in (4) and prove consistency as the number of data points grows, in terms of Γ -convergence (Dal Maso, 2012) of the functional defined by our estimator.

We also introduce a disciplined way to choose the tuning parameters λ, ν in practice based on Stein’s unbiased risk estimate (SURE). Finally, while the focus of this article is on estimation, we quantify uncertainty in our setting by using a standard subsampling routine (Politis et al., 1999) including an estimation of the rate of convergence; this approach is often costly in terms of computational and data requirements, especially in higher dimensions. Therefore, we also provide uncertainty quantification via split conformal inference (Lei et al., 2018) which provides rather conservative confidence intervals but is efficiently implementable in high dimensions. Theoretical inference results for the proposed estimator are beyond the scope of this article: the space of functions of bounded variations we consider implies a curse of dimensionality even for significantly simpler estimators than ours (Hu et al., 2022) without strong regularity restrictions on the shape of the discontinuity set, which is what we want to explicitly avoid in this paper.

3.1 Identification of the Discontinuity Sets and Jump Sizes

By definition, the proposed functional provides a regularized approximation to the true function f , even in the population. We now provide an identification result,

under mild regularity assumptions.

This result in particular implies that for fixed $\nu > 0$ and as $\lambda \rightarrow +\infty$, the solution to (4) picks out the correct discontinuity set as well as the correct jump sizes. It complements fundamental results (Richardson, 1992; Morini, 2002) from the free discontinuity literature that prove convergence of the estimated discontinuity set via the classical Mumford-Shah problem towards the true discontinuity set. Our proposed result is stronger as it guarantees not just convergence towards the correct discontinuity set but also the corresponding jump sizes.

We require the following assumptions, which are the same as in Richardson (1992). In the following, S_f denotes the discontinuity set of f , which includes the jump set J_f and coincides with it \mathcal{H}^{d-1} -almost everywhere by the Federer-Vol’pert theorem (Ambrosio et al., 2000, Theorem 3.78).

Assumption 1. *The density f_X is bounded above and below everywhere on its support \mathcal{X} , i.e. there is some $0 < k < +\infty$ such that $\frac{1}{k} \leq f_X(x) \leq k$ for all $x \in \mathcal{X}$. Further, $f \in SBV(\mathcal{X})$ and there is a constant $c > 0$ such that $|f(x)| \leq c$ for \mathcal{L}^d -almost every $x \in \mathcal{X}$. Moreover, $\int_{\mathcal{X}} |\nabla f|^2 dx + \mathcal{H}^{d-1}(S_f) < +\infty$.*

Assumption 2. *For any $x \in S_f$ it holds $\mathcal{H}^{d-1}(S_f \cap B_\rho(x)) > 0$ for all $\rho > 0$. Moreover, for any set $A \subset \mathcal{X}$ with $\text{dist}(A, S_f) > 0$ there exists a constant $L > 0$ such that $|f(x) - f(y)| \leq L|x - y|$ for any $x, y \in A$.*

Assumption 1 includes standard boundedness assumptions of the functions. The bound on f also puts a bound on the jump sizes of f . Assumption 2 imposes some regularity on f away from and at the discontinuity set. The first part intuitively implies that the graph discontinuity set is connected when measured with \mathcal{H}^{d-1} on \mathcal{X} ; in a 2-dimensional setting it implies that S_f does not have isolated points. Importantly, it directly implies that $\mathcal{H}^{d-1}((\mathcal{X} \cap \bar{S}_f) \setminus S_f) = 0$; in words, it implies that the closure \bar{S}_f coincides with S_f \mathcal{H}^{d-1} -almost everywhere (e.g. Ambrosio et al., 2000, p. 337). Both parts are the same as the assumptions in Richardson (1992).

Theorem 1. *Let Assumptions 1 and 2 hold. Then for fixed $\nu > 0$ and in the limit as $\lambda \rightarrow +\infty$ every sequence of solutions $v^*(\lambda)$ to (4) satisfies $\lim_{\lambda \rightarrow +\infty} \nabla v^*(\lambda) = 0$ \mathcal{L}^{d+1} -almost everywhere. Moreover, the jump set $J_{v^*}(\lambda)$ converges in Hausdorff distance*

d_H to the graph Γ_f , i.e.

$$\lim_{\lambda \rightarrow +\infty} d_H(J_{\nu^*}(\lambda), \Gamma_f) = 0.$$

In the statement, λ has to always diverge “fast enough”, meaning that for any sequence $\{r_n\}$ converging to 0, we always pick a sequence $\{\lambda_n\}$ such that $\lambda_n r_n \rightarrow +\infty$. This is not a restriction as we can set λ as large as we want. In practice, this means that λ will in general need to be set large to identify the correct jump sizes, especially in comparison to ν . This is exactly what our hyperparameter selection procedure finds: λ is consistently set many orders of magnitude larger than ν , and the estimated jump set and jump sizes converge to the truth as N grows large. We now turn to the practical implementation and the corresponding estimator.

3.2 The Empirical Estimator and Convergence

In the following, we assume that we observe a random sample of independent and identically distributed points $\{X_i, Y_i\}_{i=1, \dots, n}$ from the joint distribution $P_{Y, X}$. Moreover, we assume that ε_i are independent and identically distributed draws from an idiosyncratic error term ε .

Assumption 3. *Let (Ω, \mathcal{A}, P) be a probability space and let $(X, Y) : \Omega \rightarrow \mathbb{R}^d \times \mathbb{R}$, and $\varepsilon : \Omega \rightarrow \mathbb{R}$ be random variables with $\mathbb{E}[\varepsilon|X] = 0$. Furthermore, let $(X_i, Y_i, \varepsilon_i)_{i=1, \dots, n}$ be independent and identically distributed copies of (X, Y, ε) .*

The random locations of the draws are the main complication for the estimation procedure. Instead of having all the values arranged on a fixed lattice, like in computer vision applications (Pock et al., 2009), observations can be located at arbitrary points. Recent statistical approaches for the non-convex Mumford-Shah functional that deal with this setting (Caroccia et al., 2020; Chambolle et al., 2017; Chambolle and Pock, 2021) are not applicable, as the constraint set K requires a distinction between the different dimensions in \mathbb{R}^{d+1} ; in particular, it requires summation over one dimension, which is not feasible to do in practice with Voronoi- or graph partitions, but straightforward to accomplish with a rectangular grid setup.

We therefore discretize (4) on a grid to construct the estimator. For the discretization we follow Pock et al. (2009) and use a regular $(N_1 \times N_2 \times \dots \times N_d) \times S$ pixel

grid which we overlay on top of the random point cloud $\{X_i\}_{i=1,\dots,n}$. We define the grid that discretizes $\mathcal{X} \times \mathbb{R}$ by

$$\mathcal{Q}_N = \{(k_1, \dots, k_d, s) : k_i = 1, 2, \dots, N_i; \dots; k_d = 1, \dots, N_d; s = 1, 2, \dots, S\}.$$

where k_1, \dots, k_d, s are the discrete locations of each hyper-rectangle. To simplify notation, we denote $s := k_{d+1}$ and set $N_1, \dots, N_{d+1} = N$, hence assuming without loss of generality that all dimensions, including the lifted one, are discretized in the same manner. We also denote $k := (k_1, \dots, k_{d+1})$ and $k_- := (k_1, \dots, k_d)$. We define the empirical analogue of $v \in C$ as

$$\hat{v}_{Nn}(x) \equiv \hat{v}_{Nn}(\tilde{x}, t) := \sum_{1 \leq k_1, \dots, k_{d+1} \leq N} v_{k_1, \dots, k_{d+1}} \mathbb{1}\{x \in Q_{k_1, \dots, k_{d+1}}\}$$

where $v_{k_1, \dots, k_{d+1}} \equiv v_k$ is a value that we assign to the center of the respective cube $Q_{k_1, \dots, k_{d+1}} \equiv Q_k$. We define the empirical analogue p_N of $p \in K$ to take values on the boundary of the respective cubes. That is, for each $j = 1, \dots, d+1$, the corresponding value $p_{k_1, \dots, k_j + \frac{1}{2}, \dots, k_{d+1}}$ lies on the boundary $\partial Q_{k_1, \dots, k_j, \dots, k_{d+1}} \cap \partial Q_{k_1, \dots, k_j + 1, \dots, k_{d+1}}$. It is in this sense that duality is preserved in the empirical setting, as the empirical analogues of p are defined on boundaries of the cubes, while the analogues of v are defined in the center (Chambolle and Pock, 2021). Problem (4) becomes

(5)

$$\begin{aligned} \min_{v \in \tilde{C}_N} \hat{E}_{Nn}(v) &:= \min_{v \in \tilde{C}_N} \max_{p \in \hat{K}_{Nn}} \langle p, D_N v \rangle_N \\ \tilde{C}_N &= \{v : v(k) \in [0, 1], v(k_1, \dots, k_d, 1) = 1, v(k_1, \dots, k_d, N) = 0\} \\ \hat{K}_{Nn} &= \left\{ p = (p^x, p^t)^T : p^t(k) \geq \frac{|p^x(k)|^2}{4 \hat{f}_{X, Nn}(k_-)} - \lambda \hat{f}_{X, Nn}(k_-) \left(\frac{k_{d+1}}{N} - \hat{f}_{Nn}(k_-) \right)^2, \right. \\ &\quad \left. \left| \frac{1}{N} \sum_{s_1 \leq k_{d+1} \leq s_2} p^x(k) \right| \leq \nu \right\}, \end{aligned}$$

where $D_N v$ is the forward difference normalized by the grid side length; $\langle p, D_N v \rangle_N$ denotes the scalar product between two vectors,

$$\langle p_N, D_N v_N \rangle_N = \sum_{0 \leq k_1, \dots, k_{d+1} \leq N} N \left(v_{k_1, \dots, k_{d+1}}^\uparrow - v_{k_1, \dots, k_{d+1}} \right) p_{k_1, \dots, k_{d+1}},$$

with $v_{k_1, \dots, k_j, \dots, k_{d+1}}^\uparrow := v_{k_1, \dots, k_{j+1}, \dots, k_{d+1}}$ the forward value of $v_{k_1, \dots, k_j, \dots, k_{d+1}}$ for a given dimension k_j and $p_k^\uparrow = p_k \cdot e^\uparrow$, where e^\uparrow is the unit vector in the forward direction for respective j . The difference to the deterministic case is that the constraint \hat{K}_{Nn} is now random and depends on the number of data points $n = N^d$. Note that these constraints hold for all possible combinations of grid points $k = (k_1, \dots, k_{d+1})$ since they are discrete equivalents of functional constraints. Finally, the index $s_1 \leq k_{d+1} \leq s_2$ iterates over all possible combinations of indices s_1, s_2 with $1 \leq s_1 \leq k_{d+1} \leq s_2 \leq S$, along the lifted dimension. In practice, we implement 5 using the primal-dual algorithm of [Chambolle and Pock \(2011\)](#). Details on the implementation are provided in the Online Appendix.

Since we want to estimate jump locations, we cannot use a standard smoothing approach. We therefore estimate the function values $f : \mathbb{R}^d \rightarrow \mathbb{R}$ inside each grid cube via

$$\hat{f}_{Nn}(\bar{x}_{k_-}) = \sum_{i: X_i \in Q_{k_-}} w_i Y_i$$

where \bar{x}_{k_-} is the center point of the d -dimensional cube Q_{k_-} and w_i are the weights, which satisfy $\sum_{i: X_i \in Q_{k_-}} w_i = 1$ for all $k_- = (k_1, \dots, k_d)$; recall that Y_i are defined in the $(d+1)$ -st dimension. In practice, uniform weights deliver excellent results. We estimate the density f_x via a standard smoothing or histogram estimator $\hat{f}_{X, Nn}$.

The following result provides the consistency of the proposed approach. It is the first result in the mathematical and statistical literature that provides consistency of the finite approximation (either deterministic or random) to the population problem of the convexified Mumford-Shah ([Pock et al., 2009, 2010](#)). We uphold the following condition:

Assumption 4. *The density f_X has compact support \mathcal{X} and is bounded away from zero. Also, $f(x) \in BV(\mathcal{X})$ is bounded on \mathcal{X} .*

Note that the assumption on f is very weak, in particular, it does not require smoothness of the shape of the discontinuity set. We now prove convergence of the estimated \hat{v}_{Nn} towards its population counterpart v . The topology in the convergence result for the functions v_N is the weak*-convergence of functions $v \in SBV(\mathcal{X} \times \mathbb{R})$ mapping to $[0, 1]$. The estimator depends on both the number of data points n and the sidelength N^{-1} of each cube Q_k in the grid \mathcal{Q}_N .

Theorem 2. *Let Assumptions 3 and 4 hold and suppose that $N(n) \rightarrow +\infty$ with $\frac{n}{N} \rightarrow +\infty$ as $n \rightarrow +\infty$ and $\lambda, \nu > 0$ are fixed. Then $\hat{E}_{Nn}(v)$ Γ -converges in the weak*-topology in P probability to*

$$E(v) := \begin{cases} \sup_{p \in K} \int_{\mathcal{X} \times \mathbb{R}} p \cdot Dv & \text{if } v \in C \\ +\infty & \text{else.} \end{cases}$$

The limit is deterministic, which follows from the fact that we smooth out the randomness in each grid element in the limit. To obtain convergence of the minimizer, Γ -convergence needs to be complemented with the following compactness result. It follows by restricting $v \in C \cap \{v \in SBV(\mathcal{X} \times \mathbb{R}) : |Dv| \leq c\}$ for some constant c , which is natural since v is an approximation of an indicator function, which implies $c = 1$.

Corollary 1. *Let Assumptions 3 and 4 hold with $N(n) \rightarrow +\infty$ and $\frac{n}{N} \rightarrow +\infty$ as $n \rightarrow +\infty$ and let \hat{v}_{Nn} be minimizers of $\hat{E}_{N,n}(v)$ with $v \in C \cap \{v \in SBV(\mathcal{X} \times \mathbb{R}) : |Dv| \leq 1\}$. Then $\hat{v}_{Nn} \rightarrow v$ in the weak*-topology in P probability, where v is a minimizer of the deterministic $E(v)$.*

Corollary 1 implies that the estimated functions \hat{v}_{Nn} which we obtain by computing the 0.5-isosurface of \hat{v}_{Nn} converge to u^* in P probability, which is the 0.5-level set of the optimal v^* in the population. This follows from the fact that $\hat{v}_{Nn} \rightarrow v$ in $L^1(\mathcal{X} \times \mathbb{R})$ in P probability by Corollary 1 and the definition of weak*-convergence. Proposition 3.1 in Camilli (1999) then implies that the level sets converge in probability in the sense that if $|\hat{v}_{Nn} - v|_{L^1(\mathcal{X} \times \mathbb{R})} \leq \eta_n$ in P probability with $\eta_n \rightarrow 0$ as $n \rightarrow \infty$, then

$$\mathcal{L}^d \left(\mathbb{1} \left\{ \hat{v}_{Nn} = \frac{1}{2} \right\} \Delta \mathbb{1} \left\{ v = \frac{1}{2} \right\} \right) \leq \rho_n$$

in P probability with $\frac{\eta_n}{\rho_n} \rightarrow 0$. Here $A \Delta B$ is the symmetric difference of the sets A and B .

Theorem 2 implies that for fixed $\lambda, \nu > 0$, the regression surface converges to a pseudo-true value, which approximates the true regression function $f(x)$ by a piecewise smooth version $u(x)$. This is one of the main uses of the Mumford-Shah functional in image recognition, to approximate images by a “cartoon version”. That is

why Theorem 1 is crucial, as it shows that as $\lambda \rightarrow +\infty$, the estimator recovers the true $f(x)$, including the discontinuities of the correct location and size.

We estimate the jump set \hat{S}_u in practice by thresholding the estimated function u_{Nn} as follows,

$$\hat{S}_u := \{x : |\nabla \hat{u}_{Nn}(x)| \geq \sqrt{\nu}\},$$

which at the limit has to be satisfied by the true u, S_u that solve (3) (Stekalovskiy and Cremers, 2014).

We construct confidence bands for the regression function \hat{u} and the jump set \hat{S}_u by subsampling (Politis et al., 1999, Ch.8). We opt for subsampling because the inherently non-smooth nature of our estimator suggests that standard consistency results for the bootstrap need not hold. For more details on our subsampling implementation as well as a computationally efficient but more conservative alternative via conformal prediction, see the Online Appendix.

3.3 Data-Driven Choice of Hyperparameters by SURE

Theorem 1 establishes that the solution to the population problem (4) estimates the correct jump locations and sizes as λ grows for fixed ν . In finite samples, the optimal choice of hyperparameters depends on the data. Therefore, we now provide a data-driven foundation for this choice.

We pick λ, ν to minimize Stein’s unbiased risk estimate (SURE) (Stein, 1981), which under some regularity conditions is an asymptotically unbiased estimate of the mean-squared error (MSE) of the true function $f(x)$. In the image processing literature, SURE has recently been applied to a variety of settings, including the classical (non-convex) Mumford-Shah problem (Lucas et al., 2022). The benefit of using SURE for hyperparameter selection in our setting is that it has lower computational and data requirements compared to other methods such as cross-validation.

The main condition for the unbiasedness of SURE is that the error terms ε_i in (1) are Gaussian, i.e. $\varepsilon_i \sim N(0, \sigma^2) \in \mathbb{R}^n$, such that $E[Y_i|X_i] \sim N(0, \sigma^2)$ as well. Given an estimator $\hat{u}_\theta(Y)$ for the image u , which depends on the data $Y := (Y_1, \dots, Y_n)$ and hyperparameters $\theta = (\lambda, \nu)$, the Stein estimator of the mean squared error (MSE) is,

$$(6) \quad \eta(\hat{u}_\theta(Y)) = \frac{1}{N} |Y - \hat{u}_\theta(Y)|^2 - \sigma^2 + 2\sigma^2 \operatorname{div}_Y \hat{u}_\theta(Y),$$

where $\text{div}_Y \hat{u}_\theta(Y) := \sum_{i=1}^N \frac{\partial \hat{u}_{\theta,i}(Y_i)}{\partial Y_i}$ is the (weak) divergence operator of the estimator with respect to the data. Stein’s lemma famously proves that if \hat{u}_θ is a continuous and bounded operator with a well-defined divergence (Ramani et al., 2008, p.1542), then $\eta(\hat{u}_\theta)$ is an unbiased estimator of

$$MSE(\hat{u}_\theta(Y)) := \frac{1}{N} |\hat{u}_\theta(Y) - f|^2.$$

Thus, we choose λ, ν to minimize (6). Since our estimator does not have a closed-form expression for $\text{div}_Y \hat{u}_\theta(Y)$, we compute it using an asymptotically unbiased (Ramani et al., 2008, Theorem 2) Monte-Carlo approximation by perturbing the data with a vector of random noise $\mathbf{b} \sim N(0, 1)$,

$$\text{div}_Y \hat{u}_\theta(Y) = \lim_{\delta \rightarrow 0} E_{\mathbf{b}} \left\{ \mathbf{b}' \left(\frac{\hat{u}_\theta(Y + \delta \mathbf{b}) - \hat{u}_\theta(Y)}{\delta} \right) \right\}.$$

We follow Lucas et al. (2022) in further averaging the resulting SURE estimator $\eta_{\delta, \mathbf{b}^{(r)}}(\hat{u}_\theta(Y))$ out over R draws of $\mathbf{b}^{(r)}$, leading to the “Monte Carlo averaged SURE”, $\bar{\eta}^R(\hat{u}_\theta(Y)) := \frac{1}{R} \sum_{r=1}^R \eta_{\delta, \mathbf{b}^{(r)}}(\hat{u}_\theta(Y))$. Following Ramani et al. (2008), we set $\delta = 0.01$ in practice and note that the results are robust to choosing δ smaller. Next, we show with Monte Carlo simulations that the hyperparameters chosen by SURE give correct results.

3.4 Simulations

We illustrate the theoretical results from the previous sections with various simulations. Throughout, we scale the jumps by the standard deviation of the raw data away from the jumps, also known as Cohen’s d . In line with the literature, we consider a jump with a Cohen’s d of 0.25 to be small, 0.5 medium, and 0.75 large (Stommes et al., 2023). Throughout, we set the noise standard deviation to $\sigma = 0.05$ and estimate the hyperparameters λ, ν by doing a grid search over the averaged MC-SURE with $R=3$ on a 20×20 grid of values sampled from two uniform distributions. In line with Lucas et al. (2022), we search for λ on the interval $[1, 500]$, and for ν on the interval $[5e10^{-4}, 0.1]$, but our results are robust to the choice of these intervals. Following Pock et al. (2009), we discretize the lifted dimension at 32 points. The

convergence tolerance for the algorithm is set to $5e10^{-5}$.

Table 1: Monte Carlo Simulations

(a) 1D

n	MSE	MSE τ_{FD}	Bias τ_{FD}	FNR	FPR
500	0.0789	0.0358	-0.084	-0.0437	0.1933
1000	0.0461	0.0305	-0.0659	0	0.1099
5000	0.0012	0.0099	-0.053	0	0.018

SURE: $\lambda = 98.6712$, $\nu = 0.0001$

(b) 2D

d = 0.25							
n	α	$\hat{\alpha}$	MSE	MSE τ_{FD}	Bias τ_{FD}	FNR	FPR
1000	0.0375	0.0698	0.0022	0.0027	0.0323	0.0098	0.6757
5000	0.0375	0.0351	0.001	0.0002	-0.0024	0.0735	0.2319
10000	0.0375	0.0297	0.0009	0.0001	-0.0078	0.269	0.0391

SURE: $\lambda = 65.6470$, $\nu = 0.0010$

d = 0.50							
n	α	$\hat{\alpha}$	MSE	MSE τ_{FD}	Bias τ_{FD}	FNR	FPR
1000	0.075	0.0859	0.0025	0.0021	0.0108	0.0159	0.561
5000	0.075	0.056	0.0011	0.001	-0.0191	0.0848	0.0837
10000	0.075	0.0545	0.001	0.0007	-0.0206	0.1476	0.0072

SURE: $\lambda = 107.6274$, $\nu = 0.0019$

d = 0.75							
n	α	$\hat{\alpha}$	MSE	MSE τ_{FD}	Bias τ_{FD}	FNR	FPR
1000	0.1126	0.0953	0.003	0.0027	-0.0173	0.0093	0.5902
5000	0.1126	0.0732	0.0016	0.0036	-0.0394	0.0588	0.1676
10000	0.1126	0.078	0.0012	0.0027	-0.0346	0.0901	0.0181

SURE: $\lambda = 209.4585$, $\nu = 0.0020$

Note: table shows averaged results from Monte Carlo simulations of 1D and 2D (300 and 100 per row, respectively) smoothly varying functions with jumps and additive Gaussian noise ($\sigma = 0.05$) in Figures 2a and 2c. The functions have varying true jump sizes for 1D (see the notes in Figure 1) and jump sizes of Cohen’s d 0.25, 0.5, 0.75 with true jump sizes indicated by α for 2D. We uniformly sample random point clouds from these functions, where N denotes the sample size, and estimate the function using our FDR estimator based on those point clouds. $\hat{\alpha}$ denotes the estimated jump size, MSE is the mean squared error with respect to the true noise-free image, $MSE_{\tau_{FD}}$ is the MSE with respect to the true jump sizes, Bias τ_{FD} is the bias of the estimated jump sizes, FNR and FPR are the false negative and false positive rate of the estimated jump locations. Hyperparameters λ , ν are estimated using a finite-difference Monte-Carlo approximation of Stein’s unbiased risk estimate (SURE) with $R = 3$ simulations on a 20×20 grid (Ramani et al., 2008), as in Eq. (6). $N = \frac{1}{20}n$ for 1D simulations, $N = \frac{2}{3}n$ for 2D simulations, where n is the raw sample size and N the number of grid cells along each dimension.

Figure 1 further above depicts the simulated examples for the 1D to 3D case. In the top panel, we show the raw data as a scatterplot, with the true underlying function in a solid blue line and our estimated function \hat{u} in solid red. We simulated 5,000 noisy samples of a smoothly varying function with various jumps ranging from small to large in terms of Cohen’s d , reported in the figure notes. The estimated jump locations are indicated by vertical dashed lines. The 95% confidence bands on \hat{u} are depicted in transparent red, while the jump locations that are significant at the 95% level are indicated by red dots on the x-axis. The confidence band were obtained from 500 subsamples for each of 4 different subsample sizes, which were used to estimate a rate of convergence of close to \sqrt{n} (Politis et al., 1999), see the Online Appendix for more details. As shown, the estimated function closely tracks the true underlying function, and the estimated jump locations overlap exactly with the true jump locations. Moreover, all estimated jumps are significant at the 95% level.

Next, we depict the 2D and 3D simulated examples in the middle and lower panels. From left to right, we show in; (b/f) the true, smoothly varying functions that jump on the perimeter of a circle and sphere, respectively, with a jump size of 0.75 and 0.5 in terms of Cohen’s d , and the function values indicated by the color intensity; (c/g) the corresponding point clouds with $n = 10,000$ and $n = 50,000$; (d/h) the corresponding estimated function \hat{u} ; (e/i) and the corresponding jump sets in white. As for the 1D example, the estimated functions closely reproduce the true underlying functions, and the estimated jump sets almost perfectly overlap with the true jump sets.

Finally, we report Monte Carlo simulations for the 1D and 2D cases in Table 1, for various jump sizes. We report the mean squared error (MSE) of the entire function as well as the jump sizes ($\text{MSE}_{\tau_{\text{FD}}}$), the bias of the jump sizes, and the false negative and positive rates (FNR and FPR) of the jump locations. Across simulations, the measures converge to zero, suggesting convergence of the true function as well as the jump sizes and locations to the truth, as we established theoretically in Sections 3.1 and 3.2. The FNR increases slightly, but this is mechanical as pixels are more likely to overlap only marginally with the boundary as the pixel size decreases. Lastly, as mentioned above, the selected hyperparameters align well with the result in Theorem 1, since the selected λ is orders of magnitude larger than ν .

4 Application: The Economic Effects of Internet Shutdowns in India

Internet shutdowns, which are deliberate disruptions of internet or electronic communications, rendering them inaccessible or effectively unusable, have recently attracted international attention, as their number reached a record high in 2022 (Rosson et al., 2023). India, with its burgeoning digital economy, has shut down the internet at least 646 times between 2018 and 2023 (Software Freedom Law Center, 2024), more than any other country in that period, mainly to prevent or shut down protests and communal violence and to prevent cheating in examinations (Human Rights Watch, 2023). In this section, we exploit the geographic discontinuities induced by such internet shutdowns to estimate their effect on economic activity.

The FDR estimator is ideally suited for this task: the precise areas exposed to the shutdown, which induced a discontinuous drop in internet connectivity, were *a priori* unknown. Furthermore, the multidimensional nature of the method allows us to estimate the discontinuities directly, circumventing the common practice of collapsing the setting to a univariate setting using “distance from the boundary”.

We study a shutdown imposed by the Rajasthan state government on September 26, 2021, that aimed to prevent cheating on the Rajasthan Eligibility Exam for Teachers. This exam paves the way for employment as a public school teacher. Such teaching positions are coveted as they come with generous benefits in many Indian states. Moreover, the exam had not been held since 2018, as it was scheduled on a discretionary basis before a change of rules led it to be held yearly from 2023 onward (Yeung et al., 2021). As a result, hundreds of thousands of candidates sat for the exam, increasing the likelihood of cheating, which had become a high-profile issue after several cheating scandals received nationwide media attention in previous years (Purohit, 2022). In response to these concerns, various Rajasthan district governments announced, in the days prior to September 26, a shutdown of mobile internet connectivity between 6 am and 6 pm.¹

In addition, several other Rajasthan districts shut down their telecom services

¹Specifically, the District Magistrates or Divisional Commissioners of the following districts issued official orders mandating the temporary suspension of telecom services: Ajmer, Jhunjhunu, Kota, Bundi, Baran, Jhalawar, and Udaipur (other than Lasadiya and Kotda) (Mishra, 2021).

without official warning, leaving an estimated 25 million people affected (Yeung et al., 2021). The lack of clarity around which areas would be affected induced jumps in the mobile signal at unknown locations. Furthermore, while the telecom circle (also known as License Service Area) for Rajasthan state does roughly overlap with the state boundaries, mobile device connectivity is a diffuse process that is in no way guaranteed to overlap with administrative boundaries, as it is determined by the interplay of service providers, points of connection (telecom masts), and points of contact (mobile phones). As a result, the September 26 shutdown is marked by uncertainty in both the precise set of districts as well as which areas within each district were exposed to the shutdown.

At the same time, those areas that were exposed to the shutdown were disconnected completely, as local governments effectively exercised total control over the service providers. Those, in turn, can implement the shutdown at the point of contact (the mobile phone), not just the transmission points (the masts). This implies that the signal disruption induced discontinuities because individuals could not connect to other masts outside the shutdown area. Together, these factors generated unknown, discontinuous jumps in mobile connectivity.

Although the shutdown only targeted mobile connectivity, this was in effect equivalent to a near-complete shutdown of the internet, as Wi-Fi made up only 0.08% of wireless connectivity in India in 2022, while only 3.74% of total internet subscribers had wired connections (Telecom Regulatory Authority of India, 2023). This near-total internet shutdown affected the economy through a variety of channels. It was estimated to have led to the closure of 80,000 shops in Jaipur, Rajasthan’s capital, alone (The Economist, 2021). Many shops and restaurants rely on United Payments Interface, an Indian mobile payments platform, or similar platforms such as Google Pay, to process payments. Cash payments are commonly accepted, though less preferred: the ubiquity of mobile payments means few people carry cash, and even very small transactions are settled with mobile payments (Mashal and Kumar, 2023). As a result, many ATMs were reported to have run dry during similar shutdown episodes (The Times of India, 2023). Moreover, a large share of retail and hospitality sales that happen through mobile applications get disrupted by shutdowns.

More generally, internet shutdowns directly disrupt the digital economy – the

share of the economy dependent on digital services – which was estimated to make up 22% of Indian GDP in 2019 while continuing to grow rapidly. While wired connections were not disrupted, a disruption of mobile connectivity still affects business operations outside of the office setting, such as supply-chain tracking systems, process automation systems, distribution networks, remote working, and customer support services. Moreover, small businesses and businesses in rural areas often rely entirely on mobile internet for their in-office operations (Kathuria et al., 2018, p.55).

The prevalent framework for quantifying the cost of internet shutdowns was developed by West (2016).² It relies on a back-of-the-envelope quantification that combines estimates from statistical offices and the academic literature of the size of the digital economy, the degree of mobile penetration, and the “digital multiplier” (Quelch, 2009). While this approach provides a transparent way to illustrate the costs associated with internet shutdowns, it cannot claim to capture true local economic effects. To our knowledge, there exist no estimates of the economic costs of internet shutdowns in the academic literature, despite their global prominence. There is, however, a rich literature in development economics that has studied the economic effects of digital communication infrastructure rollouts. A seminal paper by Röller and Waverman (2001) used a cross-country production function approach to estimate that every percent increase in mobile penetration led to around 0.15% increase in economic growth on average. Other papers have found that mobile phone expansion and subsidies led to welfare improvements, cost of living reductions, and increased employment (Jensen, 2007; Björkegren and Karaca, 2022; Couture et al., 2021; Zuo, 2021; Hjort and Poulsen, 2019). More generally, the digitization of the economy has been found to lower search, replication, transportation, tracking, and verification costs (Goldfarb and Tucker, 2019).

An internet shutdown can disrupt all of these cost-reduction and welfare-improvement channels and is thus likely to have ripple effects across various dimensions of the economy, not just its explicitly digitized parts. On the other hand, there is clearly a distinction between the initial expansion of connectivity and its disruption once telecommunications channels have been put in place: the former expands the production possibilities set, while the latter tends to disrupt certain modes of production

²See also Top10VPN (2023) and <https://netblocks.org/projects/cost>.

altogether, to the extent that digital communications technologies, after they have been rolled out, play “an ‘enabling function’ across all critical infrastructure sectors” (CISA, n.d.). Thus, the effects of an internet shutdown will not be symmetric to the effects of an internet expansion. In that sense, our findings are complementary to the existing literature.

Finally, the use of mobile device data to proxy socioeconomic activity has been well-established in various literatures that study human mobility, such as remote sensing, network science, and complex systems (Šćepanović et al., 2015; Kung et al., 2014; Frias-Martinez and Virseda, 2012; Eagle et al., 2010; Blumenstock et al., 2015), as well as in industry (Naef et al., 2014) and government (World Bank Group, 2022). Recently, several papers in the social sciences have applied similar methods to study questions in economics (Kreindler and Miyauchi, 2021) and public policy (Van Dijke et al., 2023).

4.1 Data

Anonymized device-level location data from data provider Veraset allow us to assess the effects of the shutdown at a fine spatio-temporal granularity. The data consist of “pings”, timestamped GPS locations shared by the device with a mobile app. Veraset cleans and aggregates such data from thousands of “Software-Development Kits” (SDK), packages of tools that supply the infrastructure for most mobile applications. Location data from the same device can be recombined from various SDKs by use of an anonymized device ID. Inside our sample area, which is a bounding box around the state of Rajasthan, we observe 126 million unique pings on the four Sundays of September 2021 between 6 am and 6 pm, emanating from 3.8 million unique devices. With an estimated smartphone penetration of 60.63% in India in 2021 and a population in Rajasthan of around 80 million, this means we capture a little under 10% of all mobile devices in the area.

Mobile Signal. To capture the disruption to all mobile signal, we define,

$$(7) \quad \overline{Pings}_i := \frac{Pings_{it_0}}{\frac{1}{3} \sum_{t=t_0-3}^{t_0-1} Pings_{it}}$$

where i indexes a 5×5 km grid cell in a grid overlaid on top of a bounding box around Rajasthan; t_0 indexes September 26, 2021, between 6 am and 6 pm, the shutdown period; and $t_0 - 3, \dots, t_0 - 1$ index the same time window on the three previous Sundays. $Pings_{it}$ is a count of the number of mobile device pings observed in grid cell i in period t . The normalization by the average number of pings on Sunday in the previous month should remove regional differences in absolute activity levels, which should help make the underlying process continuous away from the jumps, as required by our assumptions. This measure of mobile activity combines the effects of the shutdown on both the mobile signal and the underlying activity it is supposed to transmit, which likely also drops in response to the shutdown. Conversely, the signal disruption is possibly understated due to satellite-based location caching, as further discussed below.

Finally, the reason some devices still emit a signal from within Rajasthan even during the shutdown is that mobile devices do not, in fact, require a mobile connection to measure their geographic location, but only a satellite connection, which was not shut down. Some mobile applications will cache users' locations when the device is offline and upload them when the internet connection is restored. This requires that users grant the applications permission to continuously collect background location.

Economic Activity. To further estimate the effect of the shutdown on economic activity, we recompute our \overline{Pings}_i measure for the subset of devices that emitted a signal during both the shutdown period and the three Sundays prior. This separates the effect of the signal disruption from that of the disruption to daily activities caused by the shutdown, both of which are captured by the raw \overline{Pings}_i measure. As we discuss below, devices that emit a signal during the shutdown period almost certainly have continuous background location collection activated, which means their signal should not be meaningfully disrupted by the shutdown. As a result, by restricting the pre-shutdown sample to those devices and normalizing by the average signal in that sample, we can remove the effects of the signal disruption.

In order to capture “economic activity”, we focus on the activity of the restricted sample of devices around “economic areas”. That is, we classify a ping as being related to economic activity if it originates from within a 150-meter radius around a Point of Interest (POI) in SafeGraph Places, or an OpenStreetMaps (OSM) POI related

to commercial or public use. SafeGraph Places is a proprietary machine-generated, human-verified data product that contains precise building footprints of millions of commercial POIs around the world. In India, it mostly consists of global brands such as McDonald’s, Toyota, or Adidas. We complement this dataset with a snapshot of OSM for India in September 2021, filtering on commerce-related POIs.³ OSM is an open-source geographic platform that aggregates POI information from user contributions. It has broader coverage of local, non-brand stores than SafeGraph Places, but is generally less accurate, often providing only the lat-long of a place rather than its precise footprint. By combining both datasets, we obtain a large and accurate sample of around 108 thousand “economic” POIs in our bounding box around Rajasthan.

We cast our measure of economic activity within each grid cell i , \overline{Econ}_i , to a coarser grid than the mobile signal data, of 40×40 km, to increase the number of observations within each cell and avoid artificial left-censoring of the data due to the lower signal. We also further smooth the data by intersecting the grid cells with the 200+ municipalities (Urban Local Bodies) in our sample area and taking the average of our economic measure across the intersected areas.

Few other data sources can proxy for economic activity at the fine spatiotemporal granularity we consider: night lights data are not suited for such rapid temporal variations in economic activity that happen during daytime; credit card penetration is low in India (around 5%); and data from neither Google Pay nor UPI were being sold to third parties at the time of this paper’s writing.

Our device-based measure should be a good proxy of economic activity under two assumptions: 1) the economic activity of the restricted device sample is representative of the wider population; 2) the set of apps that cached devices’ background location for the duration of the shutdown was not in some way skewed so as to collect geolocation information at different rates in economic areas during the shutdown. To point 1), though information on location-sharing behavior by Indian smartphone users is sparse, one recent poll of Android users suggests 66.78% have background location enabled, and Android makes up 95.21% of the smartphone market share in India ([An-](#)

³In particular, we consider all OSM POIs classified as amenity, shop, office, industrial area, craft, and all areas classified to have a primary land use related to retail, commercial, or industrial activity.

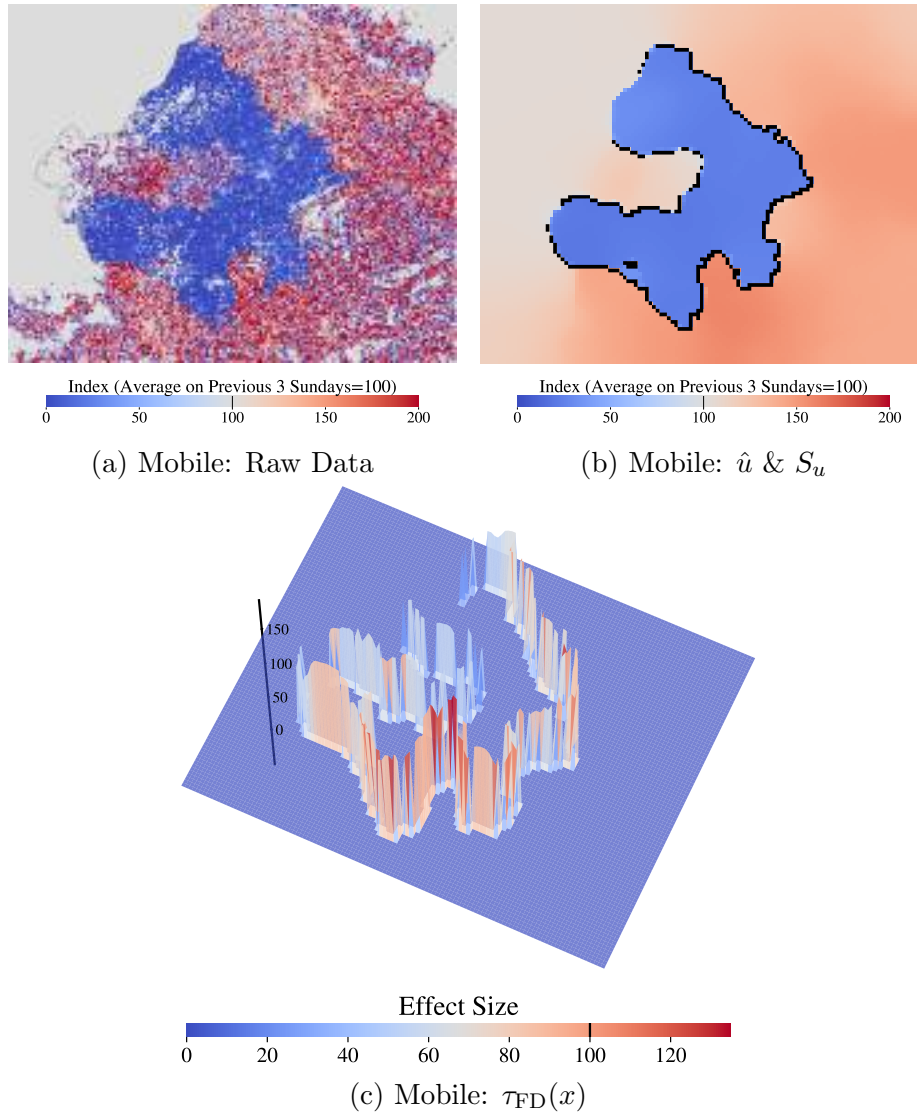


Figure 4: Internet Shutdown: Mobile Data Effects

Notes: Plots depict in- and outputs of FDR estimation of effects of internet shutdown in Rajasthan state, India, on September 26, 2021, on mobile device signal, as measured by \overline{Pings}_i (see (7)) which is a measure of the total mobile device pings per 5×5 km between 6 am and 6 pm on the day of the shutdown relative to the average in the same time window on the preceding 3 Sundays. $\lambda = 91.2474, \nu = 0.0656$ selected by SURE. (a) Shows the raw input data with the fill color of each cell indicating the value of \overline{Pings}_i . The outline of Rajasthan state is indicated by black lines. (b) shows the estimated regression function in color, with the estimated jump set S_u indicated in black. (c) shows the effect curve $\tau_{FD}(x)$ for those areas that have a jump induced by the shutdown (dropping the discontinuity in the north-east), with the z-axis indicating the magnitude of the drop in terms of \overline{Pings}_i .

droid Authority, 2022; Statcounter, 2023).⁴ This is further supported by the fact that even after filtering on economic areas for our restricted device sample, we still retain around 68 million unique pings compared to the 126 million total pings. To point 2), our location data comes from 1,000+ different mobile applications which allows location tracking across a wide variety of platforms and thus areas and behaviors. Additionally, as mentioned, background location collection happens continuously so it should not be meaningfully affected by people’s differential app usage during the shutdown.

Finally, one might expect there to be selection effects on economic activity around the discontinuity, if people near the border cross state lines to avoid the shutdown. We can, in fact, test for the presence of such selection effects, by counting the number of devices that are observed crossing the state boundaries during the shutdown and comparing it to the average in the preceding month.⁵ The idea is illustrated in Figure 2 in the Online Appendix, which plots the pings emanating from a 40km band around the Rajasthan state boundary. We then calculate the share of these devices that appeared in one of the neighboring states during the shutdown period and compare this share to its average in the month prior.

That way, we find that 2.66% of devices near the discontinuity cross it between 6 am and 6 pm on the day of the shutdown, while on average 3.57% do so in the same time window the month prior. Thus, there do not appear to be any self-selection effects associated with the localized internet shutdown. In fact, the lower share of devices that cross suggests that some crossing behavior was disrupted due to the shutdown, likely because of the associated disruption of mobile-based navigation services. This does not contradict the fact that location apps may still be caching devices’ location, as route calculation and live navigation require full mobile connection, not just satellite-based geo-positioning. Relatedly, we do not expect mobile phones near the Rajasthan border to connect to telecommunications masts across the border, as the shutdown was implemented at the point of contact (the mobile phone) and not just the transmission points (the masts), as mentioned above.

⁴This poll was conducted after the release of Android 10 in September 2019, which introduced explicit permission prompts for background location tracking.

⁵We focus on self-selection around the state boundary as it was not announced prior to the shutdown that the island of connectivity we discovered inside the Rajasthan borders would remain connected, so there was limited room for self-selection there.

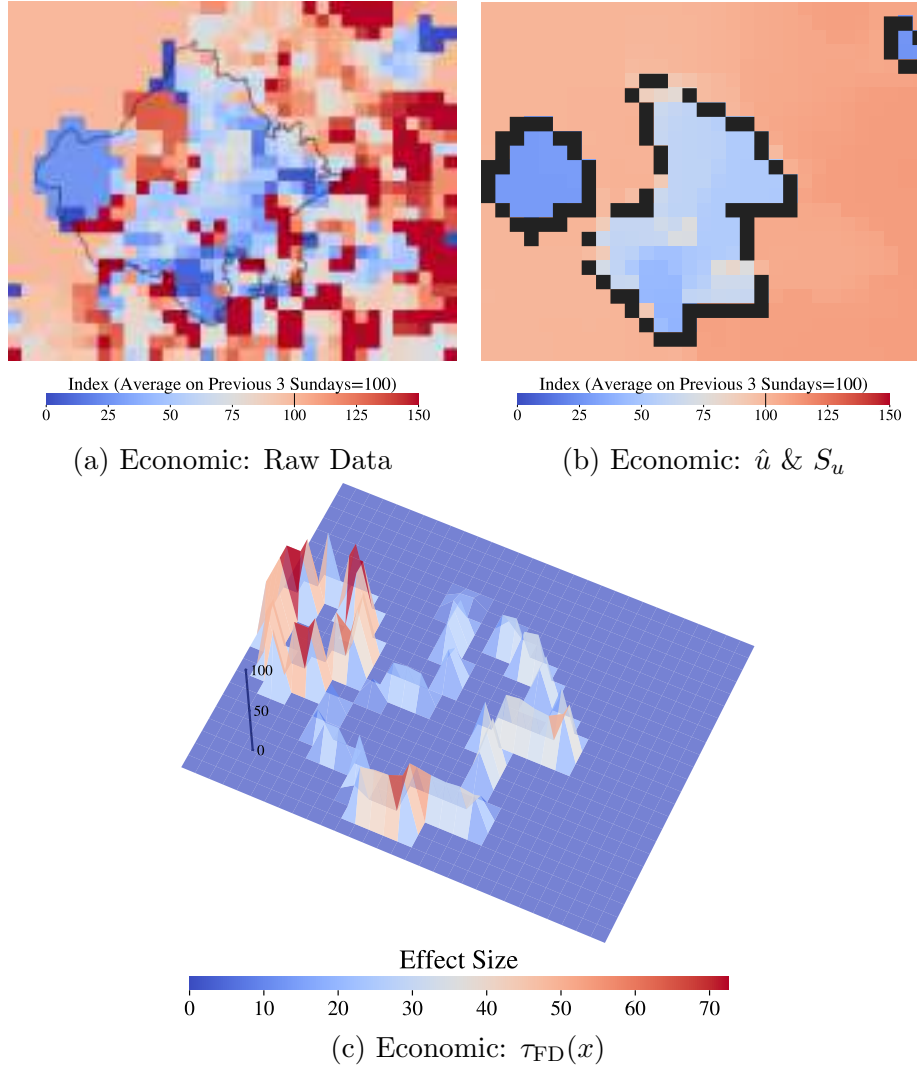


Figure 5: Internet Shutdown: Economic Effects

Notes: Plots depict in- and outputs of FDR estimation of effects of internet shutdown in Rajasthan state, India, on September 26, 2021, on economic activity, as measured by \overline{Econ}_i (see the “Economic Activity“ paragraph) which is a measure of the share of mobile device pings per 40×40 km that fell within a 150m radius of a commerce-related Point of Interest between 6 am and 6 pm on the day of the shutdown relative to the average in the same time window on the preceding 3 Sundays. $\lambda = 4.9813, \nu = 0.0754$ selected by SURE.

(a) Shows the raw input data with the fill color of each cell indicating the value of \overline{Econ}_i . The outline of Rajasthan state is solid black. (b) shows the estimated regression function in color, with the estimated jump set S_u indicated in black if the jump size is significantly different from 0 at the 95% level based on 4×200 subsamples with estimated rate of convergence of approx. \sqrt{n} (Politis et al., 1999) and grey otherwise. (c) shows the effect curve $\tau_{FD}(x)$ for those areas that have a jump induced by the shutdown (dropping the discontinuity in the north-east), with the z-axis indicating the magnitude of the drop in terms of \overline{Econ}_i .

4.2 Results

We apply the FDR estimator to the data described above, starting with the mobile device signal in Figure 4.

Mobile signal The raw data is shown in Figure 4a, with the outline of Rajasthan depicted in black. The 5×5 km grid cells provide a highly granular picture of the signal. Warmer colors indicate that the signal was stronger relative to its average on the previous 3 Sundays in the same time window, with an index of 100 corresponding to a greyish blue color. Areas without signal before the shutdown are normalized to 100 (no change).⁶ The figure shows a stark drop-off in the signal which almost perfectly overlaps with the state boundary. Moreover, there is an unexpected pocket of activity in the northwest of the state that intersects, but does not overlap with, the districts of Jaisalmer, Jodhpur, and Nagaur.⁷ It is unexpected in the sense that there are no official documents or news reports that suggested the shutdown would not affect this area. In the areas with a signal that is not affected by the shutdown, it is, on average, 130% of the September mean. This higher-than-average signal is partly mechanical since the low spatial granularity naturally leads to noisy data, as also evidenced by the dark blue dots scattered throughout the non-shutdown area. Moreover, the Indian economy was seeing a post-COVID rebound at the time, with the Indian chief scientist of the World Health Organization declaring that India is “learning to live with the virus” on August 25, 2021 (Bhaduri, 2021). Additionally, preparations for the festive season, which lasts from October to December, usually kick off around late September.

In Figure 4b, we depict the estimated regression function \hat{u} , using the same color scale as for the raw data, and the estimated jump set S_u in black. For the hyperparameter selection, we searched over $\lambda \in [1, 100]$, $\nu \in [0.001, 0.1]$ to account for the large amount of noise. We also winsorized the mobile and economic data at the 90th percentile to smooth out outliers. Some areas see an unusually high number of pings due to an infrequent congregation of people (e.g. sports events or concerts), such

⁶These are the country of Pakistan to the west of the Rajasthan border, the Great Rann of Kutch salt marsh in the southwest, the Kuno national park in the east, and China and Nepal in the northeast.

⁷For reference, we provide a street map and satellite view of Rajasthan in Figure 4 in the Online Appendix.

that our normalization with respect to the monthly baseline cannot account for it. Finally, we set f_X to the uniform distribution for simplicity as explicitly estimating it delivers similar results.

The resulting regression function is smooth on the two partitions of the space: the shutdown area and elsewhere. The estimated jump set is highly detailed and overlaps almost perfectly with the Rajasthan state boundary. It also picks out the island of activity in the northwest of the state, which extends into Pakistan in the west, where we normalized the index to 100. Comparing the estimated jump set with the raw data, we can see that it correctly picks out the entire shutdown region.

Looking at the effect curve, we estimate an average drop in the mobile device signal of 100 percent relative to the average in the month before, with the areas bordering the states of Gujarat in the south and Madhya Pradesh in the southeast seeing relatively higher drops in connectivity. Nonetheless, approximately 25% of the monthly average signal remains in the shutdown area because of satellite-based location caching, as explained above. Overall, we estimate that the shutdown was highly effective in its goal of disrupting connectivity, inducing a large drop in the mobile device signal throughout the shutdown area.

Economic activity Turning to the effect on economic activity, we depict the input data, estimated regression function and jump set, and effect curve in Figure 5. As in the 1D simulations, we construct 95% confidence bands around the function gradient using a subsampling approach, where we resample the raw pings data 200 times for each of 4 different subsample sizes, which we use to estimate a rate of convergence of approximately \sqrt{n} . Then, we say that a jump size is significantly different from zero whenever either the largest lower bound on the two elements of the gradient vector is above 0, or the smallest upper bound is below 0. To pick the hyperparameters, we search over $\lambda \in [1, 5]$, $\nu \in [0.075, 0.15]$ to account for the non-smooth data as well as the coarser grid.

Though the grid is coarser, economic activity is disrupted in almost exactly the same locations as mobile device activity: the jump set overlaps almost perfectly with the Rajasthan border, barring some missed curvature due to the coarse signal. It also picks out the island inside the state. In addition, one small patch of reduced signal near the Nepalese border is picked out as well, which is likely related to it falling in

the Himalayas region.

Looking at magnitudes, we find that, outside the shutdown area, economic activity had increased less than overall activity, relative to the monthly average, by approximately 0-5%. This aligns with India-wide estimates of weekly GDP growth in the same period (Woloszko, 2020). The corresponding average drop in economic activity caused by the shutdown is estimated to be around 35% of the monthly average in the eastern part of the shutdown area and 60% in the western part. The signal in the western part is sparse and it borders Pakistan. Moreover, the effects in the eastern part do not exhibit substantial heterogeneity. As a result, we conclude that our economic index dropped by around 35% in response to the shutdown. We can account for the fact that mobility data is not perfectly predictive of economic activity by scaling this estimate by the correlation between mobility data and GDP. The lowest estimate of this correlation from studies in countries around the world is 0.7 (Dong et al., 2017; Frias-Martinez and Virseda, 2012; Spelta and Pagnottoni, 2021), which gives an estimated drop in economic activity of around 25%. Even this lower bound is still 50% larger than the largest prevalent estimate of the costs of a day-long internet shutdown in India, which is around 16.4% of annualized GDP.⁸

One possible explanation for this large discrepancy is that these previous estimates are primarily based on estimates of the share of the economy that is digitized and its multiplier effects, but disruptions to communications technology can affect critical infrastructure across many non-digitized sections of the economy as well. Relatedly, the methodology behind these estimates relies on a single global digital multiplier estimate that dates from 2009 (Quelch, 2009). It seems very likely that 1) this multiplier has drastically increased in the last decade, and 2) it is larger for India, one of the world’s largest developing digital economies in 2021.

Implications of these results Taken together, our estimates reveal that the short-term economic damage inflicted by arbitrary internet shutdowns is much larger than previously thought. Of course, one can expect part of the damage induced by a day-long shutdown to be mitigated by a temporary economic overshooting in the following days. Nonetheless, while there appears to be some moderate increase in

⁸Calculated using the interactive application at <https://netblocks.org/cost/>, which is based on West (2016) using updated estimates.

mobile activity in the hours after the shutdown in Figure 3 in the Online Appendix, there is no visual evidence of such an increase in the day following the shutdown.

Similarly, one may expect the long-term costs of extended shutdowns to be lower as people adapt to the new reality. This does not undermine the fact that our estimates are much larger than comparable short-term damage estimates. That, in turn, further underscores the vital importance of digital communications infrastructure to the well-being of citizens in digitized economies and the need for states to refrain from shutting this infrastructure down ([United Nations High Commissioner for Human Rights, 2022](#)).

Finally, our findings highlight the asymmetry between internet expansion and internet shutdowns. They confirm the critical role digital communications infrastructure plays in a digitized economy *after* it has been rolled out, and emphasize the urgency to not just expand internet access to citizens across the globe, but also to safeguard its continued reliability.

5 Conclusion

This article introduces a new statistical method for estimating a regression function with unknown and potentially multivariate discontinuities. The idea is to non-parametrically estimate the entire regression surface and segment it into smooth and discontinuous parts. The method is an extension of a convex relaxation of the Mumford-Shah functional to a statistical setting with random data points. We prove identification and statistical convergence results. For the practical implementation, we use the primal-dual algorithm from [Chambolle and Pock \(2011\)](#) and pick the hyperparameters via Stein’s unbiased risk estimate.

The focus of this article is on estimation under minimal regularity assumptions. For uncertainty quantification in low-dimensional settings with many data points, we use subsampling methods while estimating the rate of convergence of the estimator. In higher-dimensional settings, we recommend conformal prediction bands. Those are larger than classical confidence bands, as they focus on the randomness of the outcome variable and not the conditional mean, but are efficiently implementable in any dimension. Other inference methods for this setting are of interest but require strong assumptions on the regularity of the discontinuity set, which we explicitly

avoid in this article.

We apply the method to estimate the effect of internet shutdowns on economic activity in the state of Rajasthan, India. We find that the shutdown resulted in a short-term reduction of economic activity by 25 – 35%, considerably exceeding previous estimates. This points to a large asymmetry in the effects of internet expansions and shutdowns, calling attention to the critical role of a robust and reliable communications infrastructure in a digitized economy.

References

- Abdulkadirođlu, A., Angrist, J. D., Narita, Y. and Pathak, P. (2022), ‘Breaking ties: Regression discontinuity design meets market design’, *Econometrica* **90**(1), 117–151.
- Alberti, G., Bouchitté, G. and Dal Maso, G. (2003), ‘The calibration method for the Mumford–Shah functional and free-discontinuity problems’, *Calculus of Variations and Partial Differential Equations* **16**(3), 299–333.
- Alberti, G., Bouchitté, G. and Maso, G. D. (2001), ‘The calibration method for the Mumford–Shah functional and free-discontinuity problems’, *arXiv math/0105013*.
- Aliprantis, C. D. and Border, K. C. (2006), *Infinite dimensional analysis—A hiker’s guide*, Springer.
- Ambrosio, L., Fusco, N. and Pallara, D. (2000), *Functions of bounded variation and free discontinuity problems*, Oxford University Press.
- Ambrus, A., Field, E. and Gonzalez, R. (2020), ‘Loss in the time of cholera: Long-run impact of a disease epidemic on the urban landscape’, *American Economic Review* **110**(2), 475–525.
- Andrews, D. W. (1993), ‘Tests for parameter instability and structural change with unknown change point’, *Econometrica: Journal of the Econometric Society* pp. 821–856.
- Android Authority (2022), ‘We asked, you told us: Most of you keep location tracking enabled’, <https://tinyurl.com/55f5mp2s>.
- Argyle, B. S., Nadauld, T. D. and Palmer, C. J. (2020), ‘Monthly payment targeting and the demand for maturity’, *The Review of Financial Studies* **33**(11), 5416–5462.

- Bai, J. and Perron, P. (2003), ‘Computation and analysis of multiple structural change models’, *Journal of applied econometrics* **18**(1), 1–22.
- Bauer, M. (2016), Solving Convex and Non-convex Functionals in Image Processing Using a Primal-dual Algorithm, PhD thesis, Universität Regensburg.
- Bhaduri, A. (2021), ‘Learning to live with covid? India may be entering endemic stage, says top doc’, *Hindustan Times* .
URL: <https://tinyurl.com/2m3rfdum>
- Björkegren, D. and Karaca, B. C. (2022), ‘Network adoption subsidies: A digital evaluation of a rural mobile phone program in Rwanda’, *Journal of Development Economics* **154**, 102762.
- Blumenstock, J., Cadamuro, G. and On, R. (2015), ‘Predicting poverty and wealth from mobile phone metadata’, *Science* **350**(6264), 1073–1076.
- Brown, E. S., Chan, T. F. and Bresson, X. (2012), ‘Completely convex formulation of the Chan-Vese image segmentation model’, *International Journal of Computer Vision* **98**(1), 103–121.
- Brunner, E. J., Dougherty, S. M. and Ross, S. L. (2021), ‘The effects of career and technical education: Evidence from the connecticut technical high school system’, *Review of Economics and Statistics* pp. 1–46.
- Caetano, G. and Maheshri, V. (2017), ‘School segregation and the identification of tipping behavior’, *Journal of Public Economics* **148**, 115–135.
- Camilli, F. (1999), ‘A note on convergence of level sets’, *Zeitschrift für Analysis und ihre Anwendungen* **18**(1), 3–12.
- Canny, J. (1986), ‘A computational approach to edge detection’, *IEEE Transactions on pattern analysis and machine intelligence* (6), 679–698.
- Card, D., Mas, A. and Rothstein, J. (2008), ‘Tipping and the dynamics of segregation’, *The Quarterly Journal of Economics* **123**(1), 177–218.
- Carneiro, P., Galasso, E. and Ginja, R. (2019), ‘Tackling social exclusion: evidence from Chile’, *The Economic Journal* **129**(617), 172–208.
- Caroccia, M., Chambolle, A. and Slepčev, D. (2020), ‘Mumford–Shah functionals on graphs and their asymptotics’, *Nonlinearity* **33**(8), 3846.
- Cattaneo, M. D. and Titiunik, R. (2022), ‘Regression discontinuity designs’, *Annual Review of Economics* **14**, 821–851.

- Chambolle, A. and Pock, T. (2011), ‘A first-order primal-dual algorithm for convex problems with applications to imaging’, *Journal of Mathematical Imaging and Vision* **40**(1), 120–145.
- Chambolle, A. and Pock, T. (2021), ‘Learning consistent discretizations of the total variation’, *SIAM Journal on Imaging Sciences* **14**(2), 778–813.
- Chambolle, A., Tan, P. and Vaiter, S. (2017), ‘Accelerated alternating descent methods for dykstra-like problems’, *Journal of Mathematical Imaging and Vision* **59**, 481–497.
- Chan, T. F. and Shen, J. (2005), *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*, SIAM.
- Chan, T. F. and Vese, L. A. (2001), ‘Active contours without edges’, *IEEE Transactions on Image Processing* **10**(2), 266–277.
- Cheng, A. (2023), ‘Regression discontinuity designs with multiple running variables’, *MIT working paper* .
- CISA (n.d.), ‘Communications sector’, <https://tinyurl.com/ycktyu7n>.
- Couture, V., Faber, B., Gu, Y. and Liu, L. (2021), ‘Connecting the countryside via e-commerce: evidence from China’, *American Economic Review: Insights* **3**(1), 35–50.
- Dal Maso, G. (2012), *An introduction to Γ -convergence*, Vol. 8, Springer Science & Business Media.
- Delgado, M. A. and Hidalgo, J. (2000), ‘Nonparametric inference on structural breaks’, *Journal of Econometrics* **96**(1), 113–144.
- Dong, L., Chen, S., Cheng, Y., Wu, Z., Li, C. and Wu, H. (2017), ‘Measuring economic activity in China with mobile big data’, *EPJ Data Science* **6**, 1–17.
- Donoho, D. L. (1999), ‘Wedgelets: Nearly minimax estimation of edges’, *The Annals of Statistics* **27**(3), 859–897.
- Donoho, D. L. and Johnstone, I. M. (1994), ‘Ideal spatial adaptation by wavelet shrinkage’, *Biometrika* **81**(3), 425–455.
- Eagle, N., Macy, M. and Claxton, R. (2010), ‘Network diversity and economic development’, *Science* **328**(5981), 1029–1031.
- Federer, H. (2014), *Geometric measure theory*, Springer.
- Frias-Martinez, V. and Virseda, J. (2012), On the relationship between socio-economic

- factors and cell phone usage, *in* ‘Proceedings of the Fifth International Conference on Information and Communication Technologies and Development’, pp. 76–84.
- García Trillos, N. and Murray, R. (2017), ‘A new analytical approach to consistency and overfitting in regularized empirical risk minimization’, *European Journal of Applied Mathematics* **28**(6), 886–921.
- García Trillos, N. and Slepčev, D. (2016), ‘Continuum limit of total variation on point clouds’, *Archive for Rational Mechanics and Analysis* **220**, 193–241.
- Goldfarb, A. and Tucker, C. (2019), ‘Digital economics’, *Journal of Economic Literature* **57**(1), 3–43.
- Hansen, B. E. (2017), ‘Regression kink with an unknown threshold’, *Journal of Business & Economic Statistics* **35**(2), 228–240.
- Harchaoui, Z. and Lévy-Leduc, C. (2010), ‘Multiple change-point estimation with a total variation penalty’, *Journal of the American Statistical Association* **105**(492), 1480–1493.
- Harchaoui, Z., Moulines, E. and Bach, F. (2008), ‘Kernel change-point analysis’, *Advances in Neural Information Processing Systems* **21**.
- Hartmann, W., Nair, H. S. and Narayanan, S. (2011), ‘Identifying causal marketing mix effects using a regression discontinuity design’, *Marketing Science* **30**(6), 1079–1097.
- Herlands, W., McFowland III, E., Wilson, A. G. and Neill, D. B. (2018), Automated local regression discontinuity design discovery, *in* ‘Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining’, pp. 1512–1520.
- Herlands, W., Neill, D. B., Nickisch, H. and Wilson, A. G. (2019), ‘Change surfaces for expressive multidimensional changepoints and counterfactual prediction.’, *J. Mach. Learn. Res.* **20**, 99–1.
- Hjort, J. and Poulsen, J. (2019), ‘The arrival of fast internet and employment in Africa’, *American Economic Review* **109**(3), 1032–1079.
- Hu, A. J., Green, A. and Tibshirani, R. J. (2022), ‘The voronoigram: Minimax estimation of bounded variation functions from scattered data’, *arXiv preprint arXiv:2212.14514* .
- Human Rights Watch (2023), India: Internet shutdowns hurt vulnerable communities,

- Technical report, Human Rights Watch.
URL: <https://tinyurl.com/5y4nv9ph>
- Hütter, J.-C. and Rigollet, P. (2016), Optimal rates for total variation denoising, in ‘Conference on Learning Theory’, PMLR, pp. 1115–1146.
- Jensen, R. (2007), ‘The digital divide: Information (technology), market performance, and welfare in the south Indian fisheries sector’, *The Quarterly Journal of Economics* **122**(3), 879–924.
- Kass, M., Witkin, A. and Terzopoulos, D. (1988), ‘Snakes: Active contour models’, *International journal of computer vision* **1**(4), 321–331.
- Kathuria, R., Kedia, M., Varma, G., Bagchi, K. and Sekhani, R. (2018), The anatomy of an internet blackout: measuring the economic impact of internet shutdowns in india, Technical report, Indian Council for Research on International Economic Relations.
- Keele, L. J. and Titiunik, R. (2015), ‘Geographic boundaries as regression discontinuities’, *Political Analysis* **23**(1), 127–155.
- Kern, H. L. and Hainmueller, J. (2009), ‘Opium for the masses: How foreign media can stabilize authoritarian regimes’, *Political Analysis* **17**(4), 377–399.
- Killick, R., Fearnhead, P. and Eckley, I. A. (2012), ‘Optimal detection of changepoints with a linear computational cost’, *Journal of the American Statistical Association* **107**(500), 1590–1598.
- Korostelev, A. P. and Tsybakov, A. B. (1993), *Minimax linewise algorithm for image reconstruction*, Springer.
- Kreindler, G. E. and Miyauchi, Y. (2021), ‘Measuring commuting and economic activity inside cities with cell phone records’, *Review of Economics and Statistics* pp. 1–48.
- Kung, K. S., Greco, K., Sobolevsky, S. and Ratti, C. (2014), ‘Exploring universal patterns in human home-work commuting from mobile phone data’, *PLOS One* **9**(6), e96180.
- Kuntz, L., Mennicken, R. and Scholtes, S. (2015), ‘Stress on the ward: Evidence of safety tipping points in hospitals’, *Management Science* **61**(4), 754–771.
- Kuo, R., Ho, L. and Hu, C. M. (2002), ‘Integration of self-organizing feature map and k-means algorithm for market segmentation’, *Computers & Operations Research*

- 29**(11), 1475–1493.
- Lamberson, P. and Page, S. E. (2012), ‘Tipping points’, *Quarterly Journal of Political Science* **7**(2), 175–208.
- Lei, J., G’Sell, M., Rinaldo, A., Tibshirani, R. J. and Wasserman, L. (2018), ‘Distribution-free predictive inference for regression’, *Journal of the American Statistical Association* **113**(523), 1094–1111.
- Li, L., Chu, W., Langford, J. and Schapire, R. E. (2010), A contextual-bandit approach to personalized news article recommendation, in ‘Proceedings of the 19th international conference on World wide web’, pp. 661–670.
- Li, M. and Ghosal, S. (2017), ‘Bayesian detection of image boundaries’, *The Annals of Statistics* **45**(5), 2190–2217.
URL: <https://doi.org/10.1214/16-AOS1523>
- Lucas, C.-G., Pascal, B., Pustelnik, N. and Abry, P. (2022), ‘Hyperparameter selection for discrete Mumford–Shah’, *Signal, Image and Video Processing* pp. 1–8.
- Madrid Padilla, C. M., Wang, D., Zhao, Z. and Yu, Y. (2022), ‘Change-point detection for sparse and dense functional data in general dimensions’, *Advances in Neural Information Processing Systems* **35**, 37121–37133.
- Mashal, M. and Kumar, H. (2023), ‘Where digital payments, even for a 10-cent chai, are colossal in scale’, *The New York Times* .
URL: <https://tinyurl.com/4xyfd846>
- Mattila, P. (1999), *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*, number 44, Cambridge University Press.
- McKelvey, J. (1984), ‘Simple transcendental expressions for the roots of cubic equations’, *American Journal of Physics* **52**(3), 269–270.
- Mishra, A. (2021), ‘The internet cannot be suspended in entire districts to prevent cheating in exams - IFF writes to the Rajasthan government’, *Internet Freedom Foundation* .
URL: <https://tinyurl.com/487zbf67>
- Mollenhoff, T. and Cremers, D. (2017), Sublabel-accurate discretization of nonconvex free-discontinuity problems, in ‘Proceedings of the IEEE International Conference on Computer Vision’, pp. 1183–1191.
- Morini, M. (2002), ‘Global calibrations for the non-homogeneous Mumford–Shah

- functional’, *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze* **1**(3), 603–648.
- Muller, H.-G. and Song, K.-S. (1994), ‘Maximin estimation of multidimensional boundaries’, *Journal of Multivariate Analysis* **50**(2), 265–281.
- Mumford, D. B. and Shah, J. (1989), ‘Optimal approximations by piecewise smooth functions and associated variational problems’, *Communications on Pure and Applied Mathematics* .
- Naef, E., Muelbert, P., Raza, S., Frederick, R., Kendall, J. and Gupta, N. (2014), Using mobile data for development, Technical report, Cartesian and the Bill & Melinda Gates Foundation.
URL: <https://tinyurl.com/2p9e2wcm>
- Narita, Y., Okumura, K., Shimizu, A. and Yata, K. (2023), Counterfactual learning with general data-generating policies, in ‘Proceedings of the AAAI Conference on Artificial Intelligence’, Vol. 37, pp. 9286–9293.
- Narita, Y. and Yata, K. (2021), ‘Algorithm is experiment: Machine learning, market design, and policy eligibility rules’, *arXiv:2104.12909* .
- Olken, B. A. (2009), ‘Do television and radio destroy social capital? Evidence from Indonesian villages’, *American Economic Journal: Applied Economics* **1**(4), 1–33.
- O’Sullivan, F. and Qian, M. (1994), ‘A regularized contrast statistic for object boundary estimation-implementation and statistical evaluation’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(6), 561–570.
- Page, E. S. (1954), ‘Continuous inspection schemes’, *Biometrika* **41**(1/2), 100–115.
- Park, C. (2022), ‘Jump gaussian process model for estimating piecewise continuous regression functions’, *The Journal of Machine Learning Research* **23**(1), 12737–12773.
- Pock, T., Cremers, D., Bischof, H. and Chambolle, A. (2009), An algorithm for minimizing the Mumford–Shah functional, in ‘2009 IEEE 12th International Conference on Computer Vision’, IEEE, pp. 1133–1140.
- Pock, T., Cremers, D., Bischof, H. and Chambolle, A. (2010), ‘Global solutions of variational models with convex regularization’, *SIAM Journal on Imaging Sciences* **3**(4), 1122–1145.
- Politis, D. N., Romano, J. P. and Wolf, M. (1999), *Subsampling*, Springer Science &

- Business Media.
- Porter, J. and Yu, P. (2015), ‘Regression discontinuity designs with unknown discontinuity points: Testing and estimation’, *Journal of Econometrics* **189**(1), 132–147.
- Purohit, D. (2022), ‘237 arrested in 4 years, yet exam mafia on a high in Rajasthan’, *The Times of India* .
URL: <https://tinyurl.com/4vr5bs83>
- Qiu, P. (1998), ‘Discontinuous regression surfaces fitting’, *The Annals of Statistics* **26**(6), 2218–2245.
- Qiu, P. and Yandell, B. (1997), ‘Jump detection in regression surfaces’, *Journal of Computational and Graphical Statistics* **6**(3), 332–354.
- Quelch, J. (2009), Quantifying the economic impact of the internet, Technical report, Harvard Business School.
- Ramani, S., Blu, T. and Unser, M. (2008), ‘Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms’, *IEEE Transactions on Image Processing* **17**(9), 1540–1554.
- Richardson, T. J. (1992), ‘Limit theorems for a variational problem arising in computer vision’, *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze* **19**(1), 1–49.
- Rinaldo, A. (2009), ‘Properties and refinements of the fused lasso’, *The Annals of Statistics* **37**(5B), 2922–2952.
URL: <https://doi.org/10.1214/08-AOS665>
- Röller, L.-H. and Waverman, L. (2001), ‘Telecommunications infrastructure and economic development: A simultaneous approach’, *American Economic Review* **91**(4), 909–923.
- Rosson, Z., Anthonio, F. and Tackett, C. (2023), ‘Weapons of control, shields of impunity: Internet shutdowns in 2022’, *Access Now* .
URL: <https://tinyurl.com/3pf9dt3r>
- Ruf, M. (2019), Discrete stochastic approximations of the Mumford–Shah functional, in ‘Annales de l’Institut Henri Poincaré C, Analyse non linéaire’, Vol. 36–4, Elsevier, pp. 887–937.
- Šćepanović, S., Mishkovski, I., Hui, P., Nurminen, J. K. and Ylä-Jääski, A. (2015), ‘Mobile phone call data as a regional socio-economic proxy indicator’, *PLOS One*

- 10(4), e0124160.
- Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M. and Sugihara, G. (2009), ‘Early-warning signals for critical transitions’, *Nature* **461**(7260), 53–59.
- Software Freedom Law Center (2024), ‘Internet shutdowns in india’, <https://internetshutdowns.in/>.
- Sonin, K. and Wright, A. L. (2022), ‘Information operations increase civilian security cooperation’, *The Economic Journal* **132**(643), 1179–1199.
- Spelta, A. and Pagnottoni, P. (2021), ‘Mobility-based real-time economic monitoring amid the COVID-19 pandemic’, *Scientific Reports* **11**(1), 13069.
- Statcounter (2023), ‘Mobile operating system market share India’, <https://gs.statcounter.com/os-market-share/mobile/india>.
- Stein, C. M. (1981), ‘Estimation of the mean of a multivariate normal distribution’, *The Annals of Statistics* pp. 1135–1151.
- Stommes, D., Aronow, P. and Sävje, F. (2023), ‘On the reliability of published findings using the regression discontinuity design in political science’, *Research & Politics* **10**(2).
- Stekalovskiy, E., Chambolle, A. and Cremers, D. (2012), A convex representation for the vectorial Mumford–Shah functional, in ‘2012 IEEE Conference on Computer Vision and Pattern Recognition’, IEEE, pp. 1712–1719.
- Stekalovskiy, E. and Cremers, D. (2014), Real-time minimization of the piecewise smooth Mumford–Shah functional, in ‘European Conference on Computer Vision’, Springer, pp. 127–141.
- Telecom Regulatory Authority of India (2023), ‘The Indian telecom services performance indicators’, <https://tinyurl.com/yj932ths>.
- The Economist (2021), ‘Internet shutdowns have become a weapon of repressive regimes’, <https://tinyurl.com/yc2jwbjz>.
- The Times of India (2023), ‘Ban on mobile internet brings life in Jaipur to virtual standstill’, <https://tinyurl.com/pa7263z9>.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005), ‘Sparsity and smoothness via the fused lasso’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **67**(1), 91–108.

- Top10VPN (2023), ‘The global cost of internet shutdowns’, <https://www.top10vpn.com/research/cost-of-internet-shutdowns/>.
- United Nations High Commissioner for Human Rights (2022), Internet shutdowns: trends, causes, legal implications and impacts on a range of human rights, UN document, United Nations Human Rights Council. A/HRC/50/55.
- van der Vaart, A. W. and Wellner, J. A. (1996), *Weak convergence and empirical processes*, Springer.
- Van Dijke, D., Wright, A. L. and Polyak, M. (2023), ‘Public response to government alerts saves lives during Russian invasion of Ukraine’, *Proceedings of the National Academy of Sciences* **120**(18), e2220160120.
- Vogt, T., Strelakovsky, E., Cremers, D. and Lellmann, J. (2020), ‘Lifting methods for manifold-valued variational problems’, *Handbook of Variational Methods for Nonlinear Geometric Data* pp. 95–119.
- Wang, L.-L., Shi, Y. and Tai, X.-C. (2012), Robust edge detection using Mumford-Shah model and binary level set method, *in* ‘Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3’, Springer, pp. 291–301.
- West, D. M. (2016), ‘Internet shutdowns cost countries \$2.4 billion last year’, *Center for Technological Innovation at Brookings, Washington, DC*.
- Woloszko, N. (2020), Tracking activity in real time with Google Trends, Technical report, OECD.
- World Bank Group (2022), Ukraine, rapid damage and needs assessment August 2022, Technical report, World Bank Group.
URL: <https://tinyurl.com/4ds3sdu6>
- Yeung, J., Gupta, S. and Jha, M. (2021), ‘Internet goes dark for millions in Indian state’s bid to stop exam cheats’, *CNN Business*.
URL: <https://tinyurl.com/2p9a62dz>
- Zhu, H., Fan, J. and Kong, L. (2014), ‘Spatially varying coefficient model for neuroimaging data with jump discontinuities’, *Journal of the American Statistical Association* **109**(507), 1084–1098.
- Zuo, G. W. (2021), ‘Wired and hired: Employment effects of subsidized broadband internet for low-income Americans’, *American Economic Journal: Economic Policy*

A Notation and Definitions

This section contains notation and definitions of mathematical objects used in the main text. These are standard and we refer to [Ambrosio et al. \(2000\)](#) for further reading.

A Radon measure μ is an inner regular and locally finite Borel measure. We denote

$$\int_A f(x) d\mu(x) := \frac{\int_A f(x) d\mu(x)}{\mu(A)}.$$

We denote the space of (k -times) continuously differentiable functions on \mathcal{X} by $C^k(\mathcal{X})$. $C_0(\mathcal{X})$, the space of all continuous functions that vanish eventually, is the closure in the sup-norm of $C_c(\mathcal{X})$, the space of all continuous functions on \mathcal{X} with compact support.

The k -dimensional Hausdorff-measure $\mathcal{H}^k(A)$ of a set $A \subset \mathbb{R}^d$, for $k \in [0, +\infty)$, is defined as (e.g. [Ambrosio et al., 2000](#), Definition 2.46)

$$\mathcal{H}^k(A) := \lim_{\delta \downarrow 0} \frac{\pi^{k/2}}{2^k \Gamma(1 + k/2)} \inf \left\{ \sum_{i \in I} (\text{diam}(A_i))^k : \text{diam}(A_i) < \delta, A \subset \bigcup_{i \in I} A_i \right\}$$

with $\Gamma(x)$ being Euler's Gamma function, $\text{diam}(A)$ the diameter of the set A , and where the sums are taken over finite or countable covers of A .

We use the following definition of functions of bounded variation.

Definition 3 (Functions of bounded variation and SBV). ([Ambrosio et al., 2000](#), Def. 3.1) *Let $u \in L^1(\mathcal{X})$; we say that u is a function of bounded variation in $\mathcal{X} \subset \mathbb{R}^d$ if the distributional derivative of u , Du , is representable by a finite Radon measure in \mathcal{X} . i.e. if*

$$(8) \quad \int_{\mathcal{X}} u \operatorname{div} \varphi dx = - \sum_{i=1}^d \int_{\mathcal{X}} \varphi_i dD_i u \quad \forall \varphi \in C_i^1(\mathcal{X}).$$

for some \mathbb{R}^d -valued Radon measure $Du = (D_1 u \dots D_d u)$ in \mathcal{X} and all continuously differentiable test functions φ . The vector space of all functions of bounded variation

in \mathcal{X} is denoted by $BV(\mathcal{X})$. If the Cantor part $D_c u$ of the decomposition of Du is zero, then u is called a special function of bounded variation and denoted as $u \in SBV(\mathcal{X})$.

We define the weak*-topology in the standard way (Ambrosio et al., 2000, Definition 3.11): Consider some $u \in BV(\mathcal{X})$ and a sequence $\{u_n\} \subset BV(\mathcal{X})$. Then we say that $\{u_n\}$ converges to u in the weak*-topology, or $u_n \xrightarrow{*} u$, if u_n converges to u in $L^1(\mathcal{X})$ and the corresponding Du_n converge in the weak*-topology to Du , i.e.

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X}} \phi \, dDu_n = \int_{\mathcal{X}} \phi \, dDu \quad \text{for all } \phi \in C_0(\mathcal{X}).$$

For two non-empty subsets $A, B \subset M$ of some metric space (M, d) , we define their Hausdorff distance as

$$d_H(A, B) := \max \left\{ \sup_{a \in A} \text{dist}(a, B), \sup_{b \in B} \text{dist}(b, A) \right\},$$

where $\text{dist}(a, B) := \inf_{b \in B} d(a, b)$ is the distance of a from B .

B Proofs

B.1 Proof of Proposition 1

Proof. We work in the weak*-topology (Ambrosio et al., 2000, Definition 3.11). Since all v have a uniformly bounded variation by $c < +\infty$ and since all v map to $[0, 1]$ and are hence uniformly bounded, it follows from Theorem 4.8 in Ambrosio et al. (2000) that $C \cap \{v : |Dv| \leq c\}$ is compact in the weak*-topology. Since C_0 is a Banach space with respect to the supremum norm, Corollary 6.40 in Aliprantis and Border (2006) implies that the objective function of (4) is jointly continuous as the total variations are bounded by $c < +\infty$. Therefore, Berge's Maximum theorem (Aliprantis and Border, 2006, Theorem 17.31) proves the existence claim. The fact that the solution is a global minimum follows from the convexity of $C \cap \{v : |Dv| \leq c\}$. \square

B.2 Proof of Theorem 1

Proof. We first show that if we set $v^*(x, t) = \mathbf{1}_f(x, t)$, then $E(v^*) < +\infty$ for any $\lambda \geq 0$ and also as $\lambda \rightarrow +\infty$. In fact, since $f \in SBV(\mathcal{X})$ by assumption, fixing the

rotation via the inward normal to the graph, the objective function in (4) can be written as (Alberti et al., 2001, Lemma 2.10)

$$E(\mathbf{1}_f) = \sup_{p \in K} \int_{\mathcal{X}} [p^x(x, f(x)) \cdot \nabla f(x) - p^t(x, f(x))] dx + \int_{S_f} \left[\int_{f^-}^{f^+} p^x(x, t) dt \right] \cdot \rho_f(x) d\mathcal{H}^{d-1}(x),$$

where p^x denotes the first d dimensions of the vector field p and p^t denotes the lifted dimension. $f^-(x)$ and $f^+(x)$ denote the respective traces corresponding to the inward normal orientation $\rho_f(x)$ of the graph of f at the point $x \in S_f$. Note that when using $f(x)$, the constraint set K implies under Assumption 1 that

$$p^t(x, f(x)) \geq \frac{|p^x(x, f(x))|^2}{4f_X(x)} \geq 0$$

independently of λ and ν . Moreover, by Assumption 1 it holds that $\int_{\mathcal{X}} |\nabla f|^2 dx < +\infty$. This, by an application of Hölder's inequality for vector-valued functions in conjunction with the fact that $p \in K$ – and thus p^t cannot be negative – and $\nu < +\infty$ implies that the first term on the right-hand side is finite for all $p \in K$. For the second term on the right-hand side note that by assumption $\mathcal{H}^{d-1}(S_f) < +\infty$. Furthermore, since $p \in K$, the integrand over $p^x(x, t)$ is bounded by $\nu < +\infty$, so that the overall $E(\mathbf{1}_f) < +\infty$ for all $\lambda \geq 0$ and also as $\lambda \rightarrow +\infty$.

We now show that in the limit as $\lambda \rightarrow +\infty$ and for fixed $\nu > 0$, it holds that $\nabla v^*(\lambda) = 0$ \mathcal{L}^{d+1} -almost everywhere. To prove this, consider any $v \in C$. Since $v \in SBV(\mathbb{R}^{d+1})$ it holds by definition that

$$Dv = \nabla v \mathcal{L}^{d+1} + (v^+ - v^-) \rho_v \mathcal{H}^d \llcorner J_v,$$

where J_v is the set where $v(x, t)$ jumps, ρ_v is a corresponding orientation, and \mathcal{L}^{d+1} denotes the $(d+1)$ -dimensional Lebesgue measure. Note that by the Federer-Vol'pert theorem (Ambrosio et al., 2000, Theorem 3.78) J_v coincides with S_v \mathcal{H}^d -almost everywhere, so that we focus on J_v . It therefore holds that

$$\int p \cdot Dv = \int p \cdot \nabla v d\mathcal{L}^{d+1} + \int_{J_v} (v^+ - v^-) p \cdot \rho_v d\mathcal{H}^d.$$

Since $v \in C$, it cannot be constant everywhere because the limits as $t \rightarrow \pm\infty$ are different for all $x \in \mathcal{X}$. On the other hand, as $\lambda \rightarrow +\infty$, the range of p^t converges to \mathbb{R} , which implies that for every $\eta > 0$ and every $v \in C$ such that $\nabla v \neq 0$ on a subset of \mathbb{R}^{d+1} of positive \mathcal{L}^{d+1} -measure, there exists a large enough λ and a corresponding $p \in K$ such that $\int p \cdot \nabla v \, dx > \eta$. This implies that no v for which ∇v is not constant on a set of positive \mathcal{L}^{d+1} measure can be a solution to (4) as $\lambda \rightarrow +\infty$. Therefore, in the limit as $\lambda \rightarrow +\infty$, the only changes in v must lie in J_v .

We now prove the rest of the theorem. First, we prove that if $v \in C$ is such that for some $\eta > 0$

$$(9) \quad \Gamma_f \not\subset J_v^\eta(\lambda),$$

where $A^\eta := \{x' \in \mathcal{X} : \text{dist}(x', A) \leq \eta\}$ is the η -enlargement of A , then it cannot be a solution to (4) in the limit as $\lambda \rightarrow +\infty$. Then we prove that if for some $\eta > 0$

$$(10) \quad J_v(\lambda) \not\subset \Gamma_f^\eta$$

then it cannot be a solution to (4) in the limit as $\lambda \rightarrow +\infty$.

Showing (9) leads to a contradiction.

The idea is to show that $\lim_{\lambda \rightarrow +\infty} E(v(\lambda)) = +\infty$ for any sequence $\{v(\lambda)\} \subset C$ that satisfies (9). So suppose that $\{v(\lambda)\} \subset C$ is some sequence that satisfies (9). This implies that in the limit as $\lambda \rightarrow +\infty$, it holds that there is some $(x', t') \in \Gamma_f$ such that $\text{dist}((x', t'), J_v(\lambda)) > \eta > 0$.

This can happen in two ways. Either, $x' \notin S_f$ or $x' \in S_f$. Recall that under the first part of Assumption 2 it holds that $\mathcal{H}^{d-1}(\bar{S}_f \setminus S_f) = 0$, where \bar{S}_f is the closure. We now consider each case one by one.

First subcase. In the first case, it follows that there is some $\delta > 0$ such that $\text{dist}(x', S_f) > \delta$. We now use the Lipschitz assumption on f to show that $\mathcal{H}^d(\Gamma_f \setminus J_v) > 0$. In fact, it holds that $B_\delta(x') \cap S_f = \emptyset$. Now by the second part of Assumption 2 there exists a Lipschitz constant $L < +\infty$ such that

$$|f(x') - f(x'')| \leq L|x' - x''| \leq L\delta$$

for all $x'' \in B_\delta(x') \subset \mathcal{X}$. Now we pick $\delta > 0$ small enough such that the set $Z := B_\delta(x') \times B_{L\delta}(f(x'))$ has diameter η . The diameter of Z is $\text{diam}(Z) = \delta\sqrt{2+L^2}$, so that if we define $\delta < \frac{\eta}{\sqrt{2+L^2}}$, then $\text{diam}(Z) < \eta$. We can do this under Assumption 2. This implies in particular that $Z \cap J_v = \emptyset$.

Now the fact that $\mathcal{H}^d(Z \cap \Gamma_f) = \mathcal{H}^d((Z \cap \Gamma_f) \setminus J_v) > 0$ follows from an application of the area formula since Γ_f is a Lipschitz graph. In fact, we get (e.g. Ambrosio et al., 2000, p. 88)

$$\mathcal{H}^d(Z \cap \Gamma_f) = \int_{B_\delta(x')} \sqrt{1 + |\nabla f|^2} \, d\mathcal{L}^d > 0.$$

Denote

$$W := \{(x, t) \in J_v : x \in B_\delta(x')\}.$$

We have just shown that J_v does not contain $Z \cap \Gamma_f$ on a set $B_\delta(x')$ of positive d -dimensional Hausdorff measure, which implies that for \mathcal{H}^d -almost every $(x, t) \in W$, $t \neq f(x)$. Moreover, by definition $|v^+(x, t) - v^-(x, t)| > 0$ for \mathcal{H}^d -almost all $(x, t) \in J_v$. Since $f \in SBV(\mathcal{X})$ it holds by definition that J_v is measurable with respect to \mathcal{H}^d . Therefore,

$$(11) \quad \lim_{\lambda \rightarrow +\infty} \sup_{p \in K} \int_{J_v} (v^+ - v^-) p \cdot \rho_v \, d\mathcal{H}^d \\ = \lim_{\lambda \rightarrow +\infty} \sup_{p \in K} \left[\int_W (v^+ - v^-) p \cdot \rho_v \, d\mathcal{H}^d + \int_{J_v \setminus W} (v^+ - v^-) p \cdot \rho_v \, d\mathcal{H}^d \right].$$

Now note that it must hold that $\mathcal{H}^d(W) > 0$. In fact, since $v \in SBV(\mathcal{X} \times \mathbb{R})$, it holds that J_v is countably \mathcal{H}^d -rectifiable (e.g. Ambrosio et al., 2000, chapter 4). Since $B_\delta(x')$ is compact, W is also compact. This directly implies by Proposition 2.66 in Ambrosio et al. (2000) that $\mathcal{H}^d(W) \geq \mathcal{L}^d(\pi_{\mathcal{X}}(W))$, where $\pi_{\mathcal{X}}$ is the projection onto \mathcal{X} . But $\pi_{\mathcal{X}}(W) = B_\delta(x')$, which is of positive \mathcal{L}^d measure. The first term on the right hand side of (11) diverges to $+\infty$ because the range of p^t increases to \mathbb{R} as $\lambda \rightarrow +\infty$ and $t'' \neq f(x'')$ for \mathcal{L}^d -almost every $x'' \in B_\delta(x')$. This implies that such a v cannot be a solution to (4) in the limit as $\lambda \rightarrow +\infty$.

Second subcase. We now suppose $x \in S_f$. Since $x \in S_f$, we now have to work with the points $(x, f^+(x))$ and $(x, f^-(x))$. Without loss of generality focus on the former and suppose that $\text{dist}((x, f^+(x)), J_v(\lambda)) > \eta$ for all $\lambda > 0$ and as $\lambda \rightarrow +\infty$.

By Assumption 1, we know that $\mathcal{H}^{d-1}(S_f) < +\infty$, so that (e.g. Mattila, 1999, Theorem 4.7) $\mathcal{L}^d(S_f) = 0$ on \mathcal{X} . The first part of Assumption 2 implies that $\mathcal{L}^d(\bar{S}_f) = 0$ since it implies $\mathcal{H}^{d-1}(\bar{S}_f \setminus S_f) = 0$. This further implies that the projection $\pi_{\mathcal{X}}(B_\eta((x, f^+(x))))$ is such that it intersects $\mathcal{X} \setminus \bar{S}_f$ in such a way that

$$\mathcal{L}^d(\pi_{\mathcal{X}}(B_\delta((x, f^+(x)))) \cap (\mathcal{X} \setminus \bar{S}_f)) > 0.$$

Moreover, by Assumption 2 we know that $f(x)$ is Lipschitz away from \bar{S}_f , which directly implies that there must be points $(x', f(x'))$ in $B_\delta((x, f^+(x)))$ with $\text{dist}((x', f(x')), J_v) > 0$. This brings us back to the first subcase above and we can derive a contradiction this way and proves that (9) cannot be a solution in the limit as $\lambda \rightarrow +\infty$.

Showing (10) leads to a contradiction.

Suppose (10) holds and denote $Z^\eta := J_v \setminus \Gamma_f^\eta$ for some fixed $\eta > 0$. By our assumption, we know that Z^η is not empty, so there is some $(x, t) \in (\mathcal{X} \times \mathbb{R}) \cap Z^\eta$. As before J_v must be countably \mathcal{H}^d -rectifiable since $v \in SBV(\mathcal{X} \times \mathbb{R})$. This implies (e.g. Mattila, 1999, Lemma 15.5 (2)) that Z^η is countably \mathcal{H}^d -rectifiable. One direction of a theorem based on results by Besicovitch-Marstrand-Mattila (Ambrosio et al., 2000, Theorem 2.63) implies that

$$\theta_d(Z^\eta, (x, t)) = 1 \quad \mathcal{H}^d\text{-almost every } (x, t) \in \mathcal{X} \times \mathbb{R},$$

where

$$\theta_d(A, y) := \liminf_{r \downarrow 0} \frac{\mathcal{H}^d(B_r(y))}{\omega_d r^d}$$

is the lower density of the Hausdorff measure with respect to the volume of the d -dimensional unit ball ω_d . We may hence assume that there exist some $r_0 > 0$ and $c > 0$ such that

$$\mathcal{H}^d(Z^\eta \cap B_r((x, t))) \geq cr^d, \quad (x, t) \in Z, \quad 0 < r < r_0.$$

Recall that for any $(x, t) \in Z^\eta \cap B_r((x, t))$ it must hold that $t \neq f(x)$. Since Z^η is

by definition \mathcal{H}^d -measurable, it holds by the same argument as in (11) that

$$(12) \quad \lim_{\lambda \rightarrow +\infty} \sup_{p \in K} \int_{J_v} (v^+ - v^-) p \cdot \rho_v \, d\mathcal{H}^d \\ = \lim_{\lambda \rightarrow +\infty} \sup_{p \in K} \left[\int_{Z^\eta \cap B_r((x,t))} (v^+ - v^-) p \cdot \rho_v \, d\mathcal{H}^d + \int_{J_v \setminus (Z^\eta \cap B_r((x,t)))} (v^+ - v^-) p \cdot \rho_v \, d\mathcal{H}^d \right].$$

Now focus on the first term on the right and recall that we only consider the limit as $\lambda \rightarrow +\infty$. Hence, for every constant $M > 0$ and every $r > 0$ there exists $\lambda > 0$ such that the first term on the right hand side of (12) is greater than M . This means that the first term on the right hand side of (12) diverges to $+\infty$ as $\lambda \rightarrow +\infty$ because we can always pick a large enough λ for every r and because the range of p^t increases to \mathbb{R} as $\lambda \rightarrow +\infty$, and $t \neq f(x)$ for every $(x,t) \in Z$. This shows the contradiction to (10) for every $\eta > 0$ and proves that $\lim_{\lambda \rightarrow +\infty} d_H(J_v, \Gamma_f) = 0$ for λ that diverges fast enough. \square

B.3 Proof of Theorem 2

We split the proof of Theorem 2 into two parts. In the first part, we prove convergence for a *deterministic* analogue of the problem, for which the estimators \hat{f}_{N_n} and \hat{f}_{X,N_n} are known functions whose values are given in the center of each pixel. We do this for two reasons. First, it simplifies the proof for the statistical setting. Second, it provides a novel convergence result in the mathematical literature on image recognition, complementing recent convergence results (Carocchia et al., 2020; Chambolle and Pock, 2021; Ruf, 2019). Throughout, we assume that $v \in [0, 1]^{d+1}$ without loss of generality since all functions are defined on a compact subset of \mathbb{R}^{d+1} by Assumption 4. The deterministic part is captured in the following lemma.

Lemma 1. *Let Assumption 4 hold, let $\hat{f}_{N_n} \equiv f$ be a fixed function whose values are given in the center \bar{x}_k of each cube Q_k , and set $f_X(x) \equiv 1$ everywhere on \mathcal{X} . Then $E_N(v)$ Γ -converges in the weak*-topology to*

$$E(v) := \begin{cases} \sup_{p \in K} \int_{[0,1]^{d+1}} p \cdot Dv & \text{if } v \in C \\ +\infty & \text{else} \end{cases}.$$

To prove Lemma 1, we need another lemma, which relates the discrete problem to the approximation of the continuous problem via cubes. In the following, and to lighten the notational burden in proofs, we define

$$v_{k_1, \dots, k_j, \dots, k_{d+1}}^\uparrow := v_{k_1, \dots, k_j+1, \dots, k_{d+1}}$$

the forward value of $v_{k_1, \dots, k_j, \dots, k_{d+1}}$ for a given dimension k_j . As in the main text, we write $k := k_1, \dots, k_{d+1}$, and hence $v_k := v_{k_1, \dots, k_{d+1}}$. We write v_k^j if we want to emphasize one specific dimension $j = 1, \dots, d+1$, otherwise the dimension is defined by the context.

Lemma 2. *For the empirical analogues Dv_N of Dv and p_N of p as defined in the main text, it holds that*

$$\int_{[0,1]^{d+1}} p \cdot Dv_N = \frac{1}{N^{d+1}} \langle p_N, D_N v_N \rangle_N,$$

with

$$\langle p_N, D_N v_N \rangle_N = \sum_{0 \leq k_1, \dots, k_{d+1} \leq N} N \left(v_{k_1, \dots, k_{d+1}}^\uparrow - v_{k_1, \dots, k_{d+1}} \right) p_{k_1, \dots, k_{d+1}}^\uparrow.$$

Proof of Lemma 2. Note that $v_N \in SBV([0, 1]^{d+1})$ for all $N \in \mathbb{N}$, since it is a piecewise constant function and the partition \mathcal{Q}_N is a Caccioppoli partition by Theorem 4.16 in Ambrosio et al. (2000) in combination with Theorem 4.5.11 in Federer (2014). We can therefore write

$$\int_{[0,1]^{d+1}} p \cdot Dv_N = \sum_{0 \leq k \leq N} \int_{Q_k} p \cdot \nabla v_k \, dx + \sum_{0 \leq k \leq N} \int_{\partial Q_k^\uparrow \cap \partial Q_k} \left(v_k^\uparrow - v_k \right) p \cdot s \, d\mathcal{H}^d,$$

where the orientation $s \in \mathbb{S}^{d+1}$ is chosen in the “forward direction”, i.e. from j to $j+1$, which means that it takes the form of unit vectors $e_j \in \mathbb{R}^{d+1}$ with zeros everywhere except a 1 in one of the j positions. Since v_k is constant on Q_k it holds that $\nabla v_k = 0$ on Q_k , so that the first term vanishes.

For the second term, we have

$$\sum_{0 \leq k \leq N} \int_{\partial Q_k^\uparrow \cap \partial Q_k} \left(v_k^\uparrow - v_k \right) p_k \cdot s \, d\mathcal{H}^d = \sum_{0 \leq k \leq N} \left(v_k^\uparrow - v_k \right) p_k^\uparrow \frac{1}{N^d},$$

where we define $p_k^\uparrow = p_k \cdot e^\uparrow$, where e^\uparrow is the unit vector in the forward direction for respective j . The inequalities follow because p_k is constant on the boundary of the respective cube where it is defined and since the Hausdorff measure of a face of the hypercube in $d + 1$ dimensions with sidelength $\frac{1}{N}$ is $\frac{1}{N^d}$. Also recall that the orientation s is in terms of forward differences, so that p_k in the above expression is the one that lies on the boundary of Q_k and the corresponding Q_k^\uparrow .

We therefore have

$$\begin{aligned} \int_{[0,1]^{d+1}} p \cdot Dv_N &= \frac{1}{N^d} \sum_{0 \leq k \leq N} (v_k^\uparrow - v_k) p_k^\uparrow \\ &= \frac{1}{N^{d+1}} \sum_{0 \leq k \leq N} N (v_k^\uparrow - v_k) p_k^\uparrow \\ &\equiv \frac{1}{N^{d+1}} \langle p_N, D_N v_N \rangle_N. \end{aligned}$$

□

With this result, we are ready to prove Lemma 1.

Proof of Lemma 1. In the following, we denote weak*-convergence by $\overset{*}{\rightharpoonup}$. To prove Γ -convergence we need to show (Dal Maso, 2012, Proposition 8.1)

(i) for every $v \in C$ and every sequence $v_N \in \tilde{C}_N$ with $v_N \overset{*}{\rightharpoonup} v$ it holds that

$$E(v) \leq \liminf_{N \rightarrow \infty} E_N(v_N)$$

and

(ii) for every $v \in C$ there exists a sequence $v_N \in \tilde{C}_N$ with $v_N \overset{*}{\rightharpoonup} v$ such that

$$E(v) \geq \limsup_{N \rightarrow \infty} E_N(v_N).$$

Part (i): Let $v_N \overset{*}{\rightharpoonup} v$ and assume $\liminf_{N \rightarrow +\infty} E_N(v_N) < +\infty$. Let $p \in C_c^\infty([0, 1]^{d+1}, \mathbb{R}^{d+1}) \cap K$.

As in the main text, the discrete approximation of p is achieved by decomposing $[0, 1]^{d+1}$ into hypercubes of sidelength $\frac{1}{N}$. We define $p_k^{j+\frac{1}{2}}$ as the average flux through

the boundary of two adjacent hypercubes Q_k^{j+1} and Q_k^j , i.e.

$$p_k^{j+\frac{1}{2}} = N^d \int_{\partial Q_k^{j+1} \cap \partial Q_k^j} p \cdot s \, d\mathcal{H}^d,$$

where $s \in \mathbb{S}^d$ is the orientation chosen in the direction from j towards $j+1$. We then define

$$p_N = \sum_{0 \leq k \leq N} p_k.$$

Now we need to analyze the above constructed p_N in terms of \tilde{K} , which consists of two constraints. Let us start with the first. Each p_k is a $(d+1)$ -dimensional vector, and we have to distinguish between the first d entries of this vector and the last entry. To do this, we will write p_k^x as the vector consisting of the first d entries and p_k^t as the last entry of p_k . Furthermore, for the cube Q_k we denote its centerpoint by x_k .

We can now analyze

$$(13) \quad \left| \frac{1}{N} \sum_{\kappa_1 \leq k_{d+1} \leq \kappa_2} p_k^x \right| = \left| \frac{1}{N} \sum_{\kappa_1 \leq k_{d+1} \leq \kappa_2} p_k^x + p^x(x_k) - p^x(x_k) \right| \\ \leq \left| \frac{1}{N} \sum_{\kappa_1 \leq k_{d+1} \leq \kappa_2} p_k^x - p^x(x_k) \right| + \left| \frac{1}{N} \sum_{\kappa_1 \leq k_{d+1} \leq \kappa_2} p^x(x_k) \right|$$

for $\kappa_1, \kappa_2 \in \{1, \dots, N\}$. Since p is smooth and in particular Lipschitz, we can bound the first term in terms of the Lipschitz constant and the edge length. Let $L < +\infty$ be the Lipschitz constant and note that the Euclidean distance between x_k and the center of one of the boundaries of Q_k is $\frac{1}{2N}$. We can hence bound the first term as

$$\left| \frac{1}{N} \sum_{\kappa_1 \leq k_{d+1} \leq \kappa_2} p_k^x - p^x(x_k) \right| \leq \frac{1}{N} \sum_{\kappa_1 \leq k_{d+1} \leq \kappa_2} |p_k^x - p^x(x_k)| \\ \leq \frac{1}{N} \sum_{\kappa_1 \leq k_{d+1} \leq \kappa_2} \frac{L}{2N} \\ \leq \frac{L}{2N}.$$

The second term in (13) is bounded above by $\nu + o(1)$ since $p \in K$ and a Riemann

sum argument. So overall, we have

$$\frac{1}{N} \left| \sum_{\kappa_1 \leq k_{d+1} \leq \kappa_2} p_k^x \right| \leq \nu + o(1) + \frac{L}{2N}.$$

So for $p_N \in \tilde{K}$ we would have to normalize each p_k^x as

$$\tilde{p}_k^x = p_k^x \frac{\nu}{\nu + o(1) + \frac{L}{2N}}.$$

Let us now analyze the second part of the constraint \tilde{K} . Recall that the constraint for K reads

$$|p^x(x, t)|^2 \leq 4 (p^t(x, t) + \lambda(t - f(x))^2).$$

By Assumption 4 the term in square brackets is bounded for all t and x . We now analyze

$$|p_k^x|^2 \leq (|p_k^x - p^x(x_k)| + |p^x(x_k)|)^2,$$

where x_k is again the center point of the cube Q_k . The first term can be bounded in the same way as above:

$$|p_k^x - p^x(x_k)| \leq \frac{L}{2N}.$$

By the fact that $p \in K$ the second term satisfies

$$\begin{aligned} |p^x(x_k)| &\leq 2\sqrt{p^t(x_k) + \lambda(t_k - f(\tilde{x}_k))^2} \\ &\leq 2\sqrt{|p^t(x_k) - p_k^t| + p_k^t + \lambda(t_k - f(\tilde{x}_k))^2} \\ &\leq 2\sqrt{\frac{L}{2N} + p_k^t + \lambda(t_k - f(\tilde{x}_k))^2}, \end{aligned}$$

where we define $x_k \equiv (\tilde{x}_k, t_k) \in \mathbb{R}^{d+1}$ and the second line follows from $(a - b) \leq |a - b|$. Note that f is a bounded function by Assumption 4, which in our discretization scheme is approximated in the center of each cube. Putting both terms together gives

$$|p_k^x|^2 \leq \left(\frac{L}{2N} + \sqrt{\frac{2L}{N} + 4(p_k^t + \lambda(t_k - f(\tilde{x}_k))^2)} \right)^2$$

$$\begin{aligned}
&\leq \left(\frac{L}{2N} + \sqrt{\frac{2L}{N}} + \sqrt{4(p_k^t + \lambda(t_k - f(\tilde{x}_k))^2)} \right)^2 \\
&= \left(\frac{L}{2N} + \sqrt{\frac{2L}{N}} \right)^2 + 2 \left(\frac{L}{2N} + \sqrt{\frac{2L}{N}} \right) \sqrt{4(p_k^t + \lambda(t_k - f(\tilde{x}_k))^2)} + \\
&\quad 4(p_k^t + \lambda(t_k - f(\tilde{x}_k))^2) \\
&:= C_p,
\end{aligned}$$

where we used the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ in the second line.

So in order to make $\tilde{p}_N \in \tilde{K}_N$, we would need to normalize

$$\tilde{p}_k^x = p_k^x \left(\frac{\nu}{\nu + o(1) + \frac{L}{2N}} \right) \quad \text{and} \quad \tilde{p}_k^x = p_k^x \sqrt{\frac{4(p_k^t + \lambda(t_k - f(\tilde{x}_k))^2)}{C_p}}.$$

Taking the minimum of these two expressions for \tilde{p}_k^x , and calling this minimum $0 \leq m(N) \leq 1$ we have $\tilde{p}_k^x = m(N)p_k^x \leq p_k^x$ with $m(N) \rightarrow 1$ as $N \rightarrow \infty$. It also follows that

$$\frac{1}{N^{d+1}} \langle p_N, D_N v_N \rangle_N \leq m(N) E_N(v_N).$$

We now show the convergence of the liminf using the information we have just derived. First, since $v_N \xrightarrow{*} v$ holds by definition of weak*-convergence that

$$\int_{[0,1]^{d+1}} p \cdot Dv = \lim_{N \rightarrow \infty} \int_{[0,1]^{d+1}} p \cdot Dv_N = \lim_{N \rightarrow \infty} \frac{1}{N^{d+1}} \langle p_N, D_N v_N \rangle_N,$$

where the second inequality follows from Lemma 2. From the above argument, letting $N \rightarrow +\infty$ and then taking the supremum over p we get

$$E(v) \equiv \sup_{p \in K} \int_{[0,1]^{d+1}} p \cdot Dv \leq \liminf_{N \rightarrow \infty} E_N(v_N).$$

Note that the Neumann boundary conditions of the population problem are preserved in the finite sample version and consistency for those follows immediately by construction. This shows the first part of Γ -convergence.

Part (ii): For the second part we have to construct a recovering sequence (Dal Maso, 2012; Chambolle and Pock, 2021), which is a sequence v_N of discrete functions of the

form

$$\tilde{v}_N(x) := \sum_{0 \leq k \leq N} \tilde{v}_k \mathbb{1}\{x \in Q_k\}$$

with $\tilde{v}_N \xrightarrow{*} v$ and

$$\limsup_{N \rightarrow \infty} E_N(\tilde{v}_N) \leq E(v).$$

By [Ambrosio et al. \(2000, Theorem 3.9\)](#) we can approximate $v \in BV([0, 1]^{d+1})$ by a sequence $v_N \in C_c^\infty([0, 1]^{d+1})$ of mollifiers in L^1 with

$$\lim_{N \rightarrow \infty} \int_{[0, 1]^{d+1}} |\nabla v_N| \, dx = |Dv|.$$

We then construct the \tilde{v}_N by

$$\tilde{v}_N(x) = \sum_{0 \leq k \leq N} v_N(x_k) \mathbb{1}\{x \in Q_k\},$$

where x_k is the center point of the corresponding cube Q_k . But note that

$$\begin{aligned} & \lim_{N \rightarrow \infty} |\tilde{v}_N - v_N|_{L^1([0, 1]^{d+1})} \\ &= \lim_{N \rightarrow \infty} \int_{[0, 1]^{d+1}} \left| \sum_{0 \leq k \leq N} v_N(x_k) \mathbb{1}\{x \in Q_k\} - v(x) \right| dx \\ &\leq \lim_{N \rightarrow \infty} \int_{[0, 1]^{d+1}} \left| \sum_{0 \leq k \leq N} v_N(x_k) \mathbb{1}\{x \in Q_k\} - v_N(x) \right| dx + o(1) \\ &= \lim_{N \rightarrow \infty} \int_{[0, 1]^{d+1}} \left| \sum_{0 \leq k \leq N} v_N(x_k) \mathbb{1}\{x \in Q_k\} - \sum_{0 \leq k \leq N} v_N(x) \mathbb{1}\{x \in Q_k\} \right| dx + o(1) \\ &\leq \lim_{N \rightarrow \infty} \sum_{0 \leq k \leq N} \int_{Q_k} |v_N(x_k) - v_N(x)| \, dx + o(1) \\ &\leq \lim_{N \rightarrow \infty} \sum_{0 \leq k \leq N} \frac{L_N + o(1)}{2N} \frac{1}{N^{d+1}} + o(1) \\ &\leq \lim_{N \rightarrow \infty} \sum_{0 \leq k \leq N} \frac{L_N + o(1)}{2N^{d+2}} + o(1) \\ &\leq \lim_{N \rightarrow \infty} \frac{L_N + o(1)}{2N} + o(1) = 0, \end{aligned}$$

where we used the mean-value inequality and

$$L_N := \int_{[0,1]^{d+1}} |\nabla v_N| \, dx,$$

which we know converges to a bounded quantity since $v \in BV([0,1]^{d+1})$. This shows that $\tilde{v}_N \rightarrow v$ in $L^1([0,1]^{d+1})$.

Now we need to show that $\tilde{v}_N \xrightarrow{*} v$, for which we have to show that $\sup_{N \in \mathbb{N}} |D\tilde{v}_N| < +\infty$. Since $v_N \rightarrow v$ in $L^1([0,1]^{d+1})$, it holds that

$$|Dv_N| = \int_{[0,1]^{d+1}} |\nabla v_N| \, dx < +\infty$$

for N large enough (Ambrosio et al., 2000, Theorem 3.9). By construction, for any $N \in \mathbb{N}$ it holds that

$$\begin{aligned} \int_{[0,1]^{d+1}} p \cdot D\tilde{v}_N &= \sum_{0 \leq k \leq N} \int_{Q_k} p \cdot D\tilde{v}_N \\ &= \sum_{0 \leq k \leq N} \int_{Q_k} p \cdot \nabla \tilde{v}_N + \sum_{0 \leq k \leq N} \int_{\partial Q_k^\uparrow \cap \partial Q_k} (\tilde{v}_k^\uparrow - \tilde{v}_k) p \cdot s \, d\mathcal{H}^d \\ &= \sum_{0 \leq k \leq N} (\tilde{v}_k^\uparrow - \tilde{v}_k) p^\uparrow \frac{1}{N^d} \\ &\leq \frac{1}{N^d} \sum_{0 \leq k \leq N} \frac{L}{N} p^\uparrow, \end{aligned}$$

where p^\uparrow is the value of p on the boundary of one of the forward directions. Since p is continuous and $[0,1]^{d+1}$ is compact, it follows

$$\sup_{N \in \mathbb{N}} \sup_{p \in C_0([0,1]^{d+1})} \int_{[0,1]^{d+1}} p \cdot D\tilde{v}_N \leq \sup_{N \in \mathbb{N}} \sup_{p \in C_0([0,1]^{d+1})} \frac{1}{N^d} \sum_{0 \leq k \leq N} \frac{L}{N} p^\uparrow < +\infty,$$

which shows that $\sup_{N \in \mathbb{N}} |D\tilde{v}_N| < +\infty$. This in turn implies that that $\tilde{v}_N \xrightarrow{*} v$, which shows that \tilde{v}_N is the required recovering sequence, since the Neumann boundary conditions are also preserved as before. \square

With these lemmas, we can now prove Theorem 2.

Proof of Theorem 2. The proof follows along the same lines as the proof of Lemma

1, except that we have to account for the randomness of the estimators $\hat{f}_{Nn}(\tilde{x}_k)$ and $\hat{f}_{X,Nn}(\tilde{x}_k)$. The only random element in the optimization problem is one constraint in \tilde{K} , and this is what we analyze. All the other arguments are the same as in the deterministic case and are omitted.

Recall that the constraint for \tilde{K}_{Nn} reads

$$|p^x(x, t)|^2 \leq 4\hat{f}_{X,Nn}(\tilde{x}_k) \left(p^t(x, t) + \lambda\hat{f}_{X,Nn}(\tilde{x}_k) \left(t - \hat{f}_{Nn}(\tilde{x}_k) \right)^2 \right).$$

As before we analyze

$$|p_k^x|^2 \leq (|p_k^x - p^x(x_k)| + |p^x(x_k)|)^2,$$

where x_k is again the center point of the cube Q_k . The first term can be bounded in the same way as above:

$$|p_k^x - p^x(x_k)| \leq \frac{L}{2N}.$$

By the fact that $p \in K$ the second term satisfies

$$\begin{aligned} |p^x(x_k)| &\leq 2\sqrt{\hat{f}_{X,Nn}(\tilde{x}_k)p^t(x_k) + \lambda\hat{f}_{X,Nn}^2(\tilde{x}_k) \left(t_k - \hat{f}_{Nn}(\tilde{x}_k) \right)^2} \\ &\leq 2\sqrt{\hat{f}_{X,Nn}(\tilde{x}_k) |p^t(x_k) - p_k^t| + \hat{f}_{X,Nn}(\tilde{x}_k)p_k^t + \lambda\hat{f}_{X,Nn}^2(\tilde{x}_k) \left(t_k - \hat{f}_{Nn}(\tilde{x}_k) \right)^2} \\ &\leq 2\sqrt{\hat{f}_{X,Nn}(\tilde{x}_k)\frac{L}{2N} + \hat{f}_{X,Nn}(\tilde{x}_k)p_k^t + \lambda\hat{f}_{X,Nn}^2(\tilde{x}_k) \left(t_k - \hat{f}_{Nn}(\tilde{x}_k) \right)^2}, \end{aligned}$$

where we define $x_k \equiv (\tilde{x}_k, t_k) \in \mathbb{R}^{d+1}$ and the second line follows from $(a-b) \leq |a-b|$. The difference to the deterministic case is that \hat{f}_{Nn} and $\hat{f}_{X,Nn}$ are random estimators of $f(x)$ and $f_X(x)$, which in our discretization scheme are imputed in the center of each cube. Putting both terms together gives

$$|p_k^x|^2 \leq \left(\frac{L}{2N} + \sqrt{\hat{f}_{X,Nn}(\tilde{x}_k)\frac{2L}{N} + \hat{f}_{X,Nn}(\tilde{x}_k)p_k^t + \lambda\hat{f}_{X,Nn}^2(\tilde{x}_k) \left(t_k - \hat{f}_{Nn}(\tilde{x}_k) \right)^2} \right)^2.$$

Just as in the discrete case, we bound this further by

$$\begin{aligned}
|p_k^x|^2 &\leq \left(\frac{L}{2N} + \sqrt{\hat{f}_{X,Nn}(\tilde{x}_k) \frac{2L}{N}} \right)^2 \\
&+ 2 \left(\frac{L}{2N} + \sqrt{\hat{f}_{X,Nn}(\tilde{x}_k) \frac{2L}{N}} \right) \sqrt{4\hat{f}_{X,Nn}(\tilde{x}_k) \left(p_k^t + \lambda \hat{f}_{X,Nn}(\tilde{x}_k) (t_k - \hat{f}_{Nn}(\tilde{x}_k))^2 \right)} \\
&\quad + 4\hat{f}_{X,Nn}(\tilde{x}_k) \left(p_k^t + \lambda \hat{f}_{X,Nn}(\tilde{x}_k) (t_k - \hat{f}_{Nn}(\tilde{x}_k))^2 \right)
\end{aligned}$$

Let us consider the two terms $\hat{f}_{X,Nn}(\tilde{x}_k)$ and $\hat{f}_{Nn}(\tilde{x}_k)$ one by one. By definition, $\hat{f}_{X,Nn}(\tilde{x}_k)$ is a standard kernel smoothing (i.e. Nadaraya-Watson) or histogram estimator and by assumption, we have $\frac{n}{N} \rightarrow \infty$. Therefore, since the X_i are iid random draws from X which has a density f_X which is bounded away from zero on its compact support, we can use a standard law of large numbers in the sense that $\hat{f}_{X,Nn}(\tilde{x}_k) = f_X(\tilde{x}_k) + o_P(1)$, where $o_P(1)$ denotes a vanishing term in probability as $n \rightarrow \infty$. The Continuous Mapping Theorem (e.g. [van der Vaart and Wellner, 1996](#), Theorem 1.3.6) implies that $\hat{f}_{X,Nn}^2(\tilde{x}_k)$ also converges.

A similar argument holds for \hat{f}_{Nn} by assumption that $\mathbb{E}[\varepsilon_i | X_i] = 0$ and $\text{Var}(\varepsilon_i) = \sigma^2 < +\infty$. The law of large numbers implies that $\hat{f}_{Nn}(x) \rightarrow f(x)$ in probability. The continuous mapping theorem and Slutsky's theorem (e.g. [van der Vaart and Wellner, 1996](#), Example 1.4.7) implies that the entire expression converges in probability as $n \rightarrow \infty$, which is what we wanted to show. Now the argument is exactly the same as in the proof of Lemma 1, only that the bounds are now

$$\tilde{p}_k^x = p_k^x \left(\frac{\nu}{\nu + o(1) + \frac{L}{2N}} \right) \quad \text{and} \quad \tilde{p}_k^x = p_k^x \sqrt{\frac{S^2}{R^2 + 2RS + S^2}},$$

with

$$R := \frac{L}{2N} + \sqrt{\frac{2L}{N}} \quad \text{and} \quad S := \sqrt{4\hat{f}_{X,Nn}(\tilde{x}_k) \left(p_k^t + \lambda \hat{f}_{X,Nn}(\tilde{x}_k) (t_k - f(\tilde{x}_k))^2 \right)},$$

and one needs to account for the convergence in probability throughout. \square

C Uncertainty Quantification

C.1 Subsampling

Algorithm 1 Subsampling

-
- 1: **Input:** Data (X_i, Y_i) , $\mathcal{I} := \{1, \dots, n\}$, miscoverage level $\alpha \in (0, 1)$, regression algorithm \mathcal{A}
 - 2: **Output:** Confidence band, over $x \in \mathcal{Q}_{N_-}$
 - 3: $\hat{u} = \mathcal{A}(\{(X_i, Y_i) : i \in \mathcal{I}\})$
 - 4: **for** $j = 1$ to J **do**
 - 5: Randomly sample \mathcal{I} into K subsets $\mathcal{I}_1, \dots, \mathcal{I}_k$ of size b_1, \dots, b_K
 - 6: $u_{j,k}^* := \mathcal{A}(\{(Y_i, X_i) : i \in \mathcal{I}_k\})$
 - 7: $Z[j, k] = \max\{|u_{j,k}^* - \hat{u}|\}$
 - 8: **end for**
 - 9: $\bar{y}_k := \frac{1}{L} \sum_{\ell=1}^L \log [G_{b_k}^{*-1}(t_\ell) - G_{b_k}^{*-1}(s_\ell)]$, the mean of L log differences of the empirical quantile functions $G_b^{*-1}(t)$ at quantiles s_ℓ, t_ℓ
 - 10: $\hat{\beta} := -\frac{\text{cov}\{\bar{y}_k, \log(b_k)\}}{\text{var}\{\log(b_k)\}}$ the rate of convergence
 - 11: For some $k \in 1, \dots, K$:
 - 12: $Z^* = b_k^{\hat{\beta}} \cdot \max\{|\mathbf{u}_k^* - \hat{u}|\}$ with $\mathbf{u}_k^* = (u_{1,k}^*, \dots, u_{J,k}^*)$
 - 13: $z_\alpha := \text{sort}(Z^*)[(J+1) \cdot (\alpha)]$ the critical value
 - 14: **return** $C_{\text{sub}}(x, \alpha) = [\hat{u}(x) - z_\alpha/n^{b_k}, \hat{u}(x) + z_\alpha/n^{b_k}]$, for all $x \in \mathcal{Q}_{N_-}$
-

We construct uniform confidence bands by way of a subsampling approach with an estimated rate of convergence (Politis et al., 1999, Ch.8). The only assumption required for this approach to be consistent is that the limiting distribution exists and is non-degenerate for the rate of convergence we estimate. Under this assumption, we construct uniform confidence bounds using Algorithm 1, where \mathcal{Q}_{N_-} is the grid on the domain only, that is, without the lifted dimension. Note that, since we aim to construct confidence bands around a non-parametric estimate \hat{u} on a grid of size N , we keep the grid size fixed when calculating the subsampled estimates $u_{j,k}^*$. With this algorithm, we obtain the confidence bands and 95% significant jump sets depicted in Figures 1 and 5. To obtain confidence bands on the jump sizes, we simply repeat lines 8–14 in Algorithm 1 for the forward difference of \hat{u} and of the subsampled u^* s and conclude that a jump at point $x \in S_u$ is significant at the $\alpha\%$ level if either $\max\{y : y \in C_{\text{sub, lower}}^D(x, \alpha)\} > 0$, or $\min\{y : y \in C_{\text{sub, upper}}^D(x, \alpha)\} < 0$, where C^D indicates the confidence bands for the forward differences and “lower”, “upper”

indicate whether we consider the upper or lower bound. We take the max and min since in practice we set the jump size equal to the largest forward difference along any of the axes.

C.2 Conformal Inference

As a computationally efficient, though more conservative, alternative to quantifying the uncertainty of the function \hat{u} and the jump set S_u , we rely on distribution-free conformal prediction methods developed for the non-parametric regression setting (Lei et al., 2018). Conformal prediction, originally, is a method for constructing bands around predictions produced by machine learning methods. Its appeal lies in the fact that it can produce prediction bands for any general estimator without requiring assumptions about its properties or about the distribution of the data-generating process. In particular, we construct confidence bands around u and $|\nabla u|$ using the Split Conformal Prediction Algorithm 2 proposed in Lei et al. (2018, Algorithm 2). The authors show that, for $(X_i, Y_i)_{i=1, \dots, n}$ i.i.d. and assuming that the residuals $R_i, i \in$

Algorithm 2 Split Conformal Prediction (Lei et al., 2018)

- 1: **Input:** Data $(X_i, Y_i), i = 1, \dots, n$, miscoverage level $\alpha \in (0, 1)$, regression algorithm \mathcal{A}
 - 2: **Output:** Prediction band, over $x \in \mathbb{R}^d$
 - 3: Randomly split $\{1, \dots, n\}$ into two equal-sized subsets $\mathcal{I}_1, \mathcal{I}_2$
 - 4: $\hat{\mu} = \mathcal{A}(\{(X_i, Y_i) : i \in \mathcal{I}_1\})$
 - 5: $R_i = |Y_i - \hat{\mu}(X_i)|, i \in \mathcal{I}_2$
 - 6: $d(\alpha) =$ the k th smallest value in $\{R_i : i \in \mathcal{I}_2\}$, where $k = \lceil (n/2 + 1)(1 - \alpha) \rceil$
 - 7: **return** $C_{\text{split}}(x, \alpha) = [\hat{\mu}(x) - d(\alpha), \hat{\mu}(x) + d(\alpha)]$, for all $x \in \mathbb{R}^d$
-

\mathcal{I}_2 have a continuous joint distribution, which is guaranteed in the limit under the Gaussian assumption for the SURE,

$$\mathbb{P}(Y_{n+1} \in C_{\text{split}}(X_{n+1}, \alpha)) \leq 1 - \alpha + \frac{2}{n+2}.$$

Thus, C_{split} delivers prediction bands for every grid point $x \in \mathcal{Q}_{N_-}$ on which we estimate $\hat{u}(x)$.

In practice, we adapt Algorithm 2 to our grid-based setting as follows. Keeping the hyperparameters λ, ν and the grid \mathcal{Q}_N fixed, randomly split $\{1, \dots, n\}$ into equal

subsets $\mathcal{I}_1, \mathcal{I}_2$ and cast $(X_i, Y_i), i \in \mathcal{I}_1$ onto \mathcal{Q}_N as in Section 3.2. Solve problem (5) on the grid, and then calculate R_i as the difference between Y_i and the prediction $\hat{u}(X'_i)$ for the grid point $X'_i \in \mathcal{Q}_{N_-}$ closest to X_i . This gives prediction bands for the outcome variable Y_i . To obtain prediction bands for the jump set J_f , we also calculate the residual vector R_i^J for the implied predictions $D\hat{u}$ by computing DY'_i where Y'_i is the outcome value corresponding to the data point X_i closest to the grid point X'_i . Inference on the jump set then proceeds identically to the subsampling case. Finally, we stress that these *prediction* bands allow us to do inference on Y , but not on the conditional mean $f(X)$. Confidence bands for the latter will always be smaller than prediction bands for the former, so this approach provides a computationally efficient but very conservative hypothesis test regarding $f(X)$ and the corresponding jump set J_f . If the goal is to do inference on $f(X)$, we recommend using the subsampling approach described in the previous section.

D Implementation

We solve (5) using a primal-dual algorithm (Chambolle and Pock, 2011). At each step, we need to project the iterand v^n , which denotes the estimated primal function v at the n -th iteration of the algorithm, onto the sets C and K . We now discuss how these projections are implemented, before presenting the algorithm. The implementation follows those of Strekalovskiy et al. (2012) and Bauer (2016).

D.1 Projection Onto Constraint Sets

Projection Onto C The projection onto C can be done using a straightforward clipping,

$$(14) \quad v^{n+1} = \min\{1, \max\{0, v^n\}\},$$

where v^n is the n -th iteration of the discretized function v . We also need to impose the discretized limits from C , by setting $v^{n+1}(k_1, \dots, k_d, 1) = 1$ and $v^{n+1}(k_1, \dots, k_d, N) = 0$.

Projection Onto K The projection onto K is more involved. The first constraint,

$$(15) \quad K_1 = \left\{ p^t(k) \geq \frac{|p^x(k)|_2^2}{4} - \lambda \left(\frac{k}{S} - f(k_1, \dots, k_d) \right)^2 \right\},$$

constitutes a pointwise projection onto a parabola, which can be rewritten as an optimization program. To see this, let $\alpha > 0, p^x \in \mathbb{R}^d, p^t \in \mathbb{R}$ and $p = (p^x, p^t)^T \in \mathbb{R}^d \times \mathbb{R}$. Assume that $p_0^t < \alpha |p_0^x|_2^2$ holds for a point $p_0 \in \mathbb{R}^d \times \mathbb{R}$. The projection of p_0 onto the parabola $\alpha |p^x|_2^2$ can then be written as the following optimization program:

$$\begin{aligned} \min_{p \in \mathbb{R}^d \times \mathbb{R}} \quad & \frac{1}{2} |p - p_0|_2^2 \\ \text{subject to} \quad & p^t \geq \alpha |p^x|_2^2, \end{aligned}$$

where $\alpha = \frac{1}{4\hat{f}_{X, N_n}(k_1, \dots, k_d)}$ and we leave out the constant $\lambda \hat{f}_{X, N_n}(\dots) \left(\frac{k}{N} - f(\dots) \right)^2$ for ease of notation. The first-order conditions of the corresponding Lagrangian are,

$$(16) \quad \begin{pmatrix} p^x - p_0^x + \mu 2\alpha p^x \\ p^t - p_0^t - \mu \\ \alpha |p^x|_2^2 - p^t \end{pmatrix} = 0.$$

Solving for μ and recombining gives the cubic equation,

$$t^3 + 3bt - 2a = 0$$

with $a = 2\alpha |p_0^x|_2^2, b = \frac{2}{3}(1 - 2\alpha p_0^t)$ and $t = 2\alpha |p^x|_2^2$, which can be solved analytically as (McKelvey, 1984),

$$p^x = \begin{cases} p_0^x & \text{if } p_0^t \geq \alpha |p_0^x|_2^2 \\ \frac{w}{2\alpha} \frac{p_0^x}{|p_0^x|_2^2} & \text{if } p_0^t < \alpha |p_0^x|_2^2 \text{ and } p_0^x \neq 0 \\ 0 & \text{else} \end{cases}$$

and

$$p^t = \begin{cases} p_0^t & \text{if } p_0^t \geq \alpha |p_0^x|_2^2 \\ \alpha |p^x|_2^2 & \text{else} \end{cases},$$

where

$$d = \begin{cases} a^2 + b^3 & \text{if } b \geq 0 \\ (a - \sqrt{-b^3})(a + \sqrt{-b^3}) & \text{else} \end{cases}$$

and

$$w = \begin{cases} 0 & \text{if } c = 0 \\ c - \frac{b}{c} & \text{if } d \geq 0 \text{ and } c > 0, \\ 2\sqrt{-b} \cos\left(\frac{1}{3} \arccos \frac{a}{\sqrt{-b^3}}\right) & \text{else} \end{cases}$$

with $c = \sqrt[3]{a + \sqrt{d}}$.

The second constraint, we decouple from the first by way of Lagrange multipliers (Strekalovskiy et al., 2012). This lets us avoid the Dykstra projection originally used in Pock et al. (2009), which requires nested iterations (outer loop: primal-dual algorithm, inner loop: Dykstra's algorithm) and is thus computationally costly. In particular, we introduce a set of auxiliary variables $s_{s_1, s_2} := \sum_{s_1 \leq k_{d+1} \leq s_2} p^x(k)$ and of Lagrange multipliers $\mu_{s_1, s_2} \in \mathbb{R}^d$ and write the second constraint set as,

$$(17) \quad K_2 = \left\{ |s_{s_1, s_2}| \leq \nu \text{ s.t. } s_{s_1, s_2} = \sum_{s_1 \leq k_{d+1} \leq s_2} p^x(k) \right\},$$

with corresponding Lagrangian,

$$(18) \quad \mathcal{L}(v, \mu, p, s) = \langle p, D_N v \rangle_N + \sum_{s_1=1}^S \sum_{s_2=k_1}^S \left\langle \mu_{s_1, s_2}, \sum_{s_1 \leq k_{d+1} \leq s_2} p^x(k) - s_{s_1, s_2} \right\rangle.$$

Then, let (v^*, p^*) be the solution to (5) and $(v^*, p^*, \mu_{s_1, s_2}^*, s_{s_1, s_2}^*)$ the solution to,

$$(19) \quad \min_{v \in C} \max_{\substack{p \in K_1 \\ \mu_{s_1, s_2} \mid |s_{s_1, s_2}| \leq \nu}} \mathcal{L}(v, \mu, p, s).$$

We have that,

$$(20) \quad \langle p^*, D_N v^* \rangle_N = \mathcal{L}(v^*, \mu^*, p^*, s^*),$$

because $p \in K$ implies $p \in K_2$ and hence the Lagrange constraint $s_{s_1, s_2} = \sum_{s_1 \leq k_{d+1} \leq s_2} p^x(k)$ always holds with equality at the optimum.

D.2 Algorithm

Putting everything together, we solve the discretized problem using a primal-dual gradient descent-ascent algorithm (Chambolle and Pock, 2011). To calculate the gradient updates, we just need the derivatives with respect to the Lagrangian above, which are,

$$\begin{aligned}
 (21) \quad \frac{\partial \mathcal{L}(u, \mu, p, s)}{\partial u} &= D_N^T p \\
 \frac{\partial \mathcal{L}(u, \mu, p, s)}{\partial s_{s_1, s_2}} &= -\mu_{s_1, s_2} \\
 \frac{\partial \mathcal{L}(u, \mu, p, s)}{\partial \mu_{s_1, s_2}} &= p^x(k) - s_{s_1, s_2} \\
 \frac{\partial \mathcal{L}(u, \mu, p, s)}{\partial p} &= D_N u + \tilde{p},
 \end{aligned}$$

where

$$(22) \quad \tilde{p} = \begin{pmatrix} \sum_{s_1=1}^l \sum_{s_2=l}^S \mu_{s_1, s_2}^1 \\ \sum_{s_1=1}^l \sum_{s_2=l}^S \mu_{s_1, s_2}^2 \\ 0 \end{pmatrix}.$$

We get the following algorithm, It was proved in Chambolle and Pock (2011, Thm.1) that $v \rightarrow v^*$ for this algorithm as the number of iterations n goes to infinity, under the condition that the step sizes satisfy $\tau\sigma L^2 < 1$, where $L = |K| := \max\{|Kx| : x \in X \text{ with } |x| \leq 1\}$ with K the continuous linear operator on the primal variable.

D.3 Computational Details and Code

We implement the described algorithm in [PyTorch](#), a modern deep learning framework that offers native support for GPU acceleration. Running the algorithm on a GPU, as opposed to a CPU, provides several advantages. GPUs are inherently designed for high-throughput parallelism, enabling the simultaneous processing of thousands of

Algorithm 3 Primal-Dual Algorithm

- 1: Choose $(v^0, p^0, \mu^0, s^0) \in C \times K_p \times \mathbb{R}^{d \times N_1 \times \dots \times N_d \times I} \times \mathbb{R}^{d \times N_1 \times \dots \times N_d \times I}$ and let $\bar{v}^0 = u^0, \bar{\mu}^0 = \mu^0 = 0, p^0 = 0$.
- 2: Set $\tau_v = \sigma_p = \frac{1}{4(d+1)}$ and $\tau_\mu = 1/I, \sigma_s = 1$.
- 3: For each $n \geq 0$:
- 4:

$$\left\{ \begin{array}{l} p^{n+1} = \Pi_{K_p} (p^n + \sigma_p (D_N \bar{v}^n + \tilde{p}^n)) \\ s_{s_1, s_2}^{n+1} = \Pi_{|\cdot| \leq \nu} (s_{s_1, s_2}^n - \sigma_s \bar{\mu}_{s_1, s_2}^n) \\ v^{n+1} = \Pi_C (v^n - \tau_v D_N^* p^{n+1}) \\ \mu_{s_1, s_2}^{n+1} = \mu_{s_1, s_2}^n + \tau_\mu \left(s_{s_1, s_2}^{n+1} - \sum_{s_1 \leq k_{d+1} \leq s_2} p^{x, n+1}(k) \right) \\ \bar{v}^{n+1} = 2v^{n+1} - v^n \\ \bar{\mu}_{s_1, s_2}^{n+1} = 2\mu_{s_1, s_2}^{n+1} - \mu_{s_1, s_2}^n, \end{array} \right.$$

where $\Pi_D(x)$ denotes the projection of x onto the set D , $I := \frac{S^2+S}{2}$ the number of K constraints, and $D_N^* = D_N^T := N \cdot D^T := -N \cdot \text{div}$ the adjoint of the discrete gradient operator.^a

- 5: **return** v^*

^aImplemented in practice as backward differences with a Neumann boundary condition.

matrix operations. PyTorch offers native support for the whole spectrum of GPU architectures, including NVIDIA, Apple Silicon (for late 2020 Apple computers onward), and Intel GPUs. This ensures that researchers and practitioners with varied hardware setups can readily apply our estimator. For efficient computation of the SURE, which requires solving the algorithm a large number of times, we rely on the hyperparameter tuning suite in [Ray](#), an open-source distributed computing framework that supports parallel and fractional GPU processing. This can drastically speed up the hyperparameter selection by efficiently allocating jobs across multiple partitions of a single GPU on a local machine or multiple GPUs on a high-performance computing cluster.

The accompanying Python library can be found at <https://github.com/Davidvandijcke/fdr>. A fully developed Python package as well as extensions to R and STATA are in progress. Though the package is compatible with all major GPU architectures, researchers without GPUs on their local machines can refer to our Google Colab notebooks, which provide free access to cloud-based machines with GPU support. The current implementation of the algorithm takes 69.37 seconds on a Nvidia Tesla A100 Tensor Core GPU to converge on a 2D dataset with the number of raw data

points $n = 90,000$, the number of grid points $N = 12,750$ (25%), and the discretization of the lifted dimension $S = 32$. This computation time can be further improved by approximating the objective function using “sublabel-accurate” relaxations (Mollenhoff and Cremers, 2017). As this requires further extensions of our statistical convergence arguments, we leave this to future work.

E Additional Results

E.1 Figures

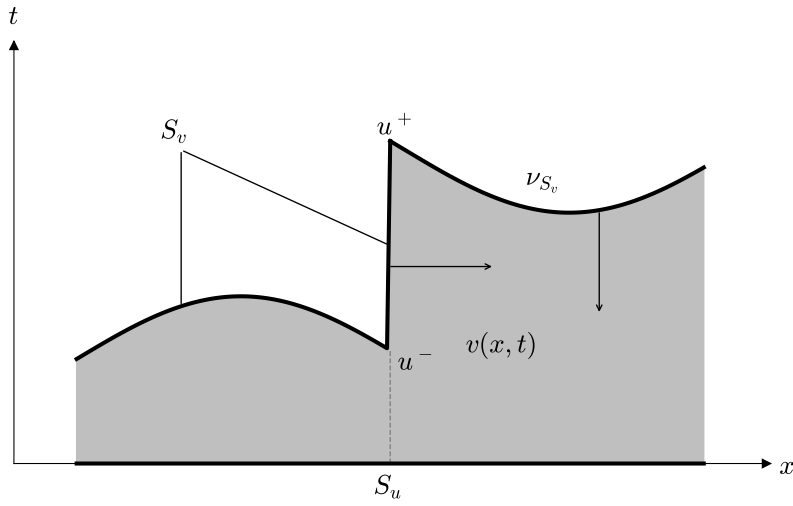


Figure 6: Convex Relaxation Through Functional Lifting

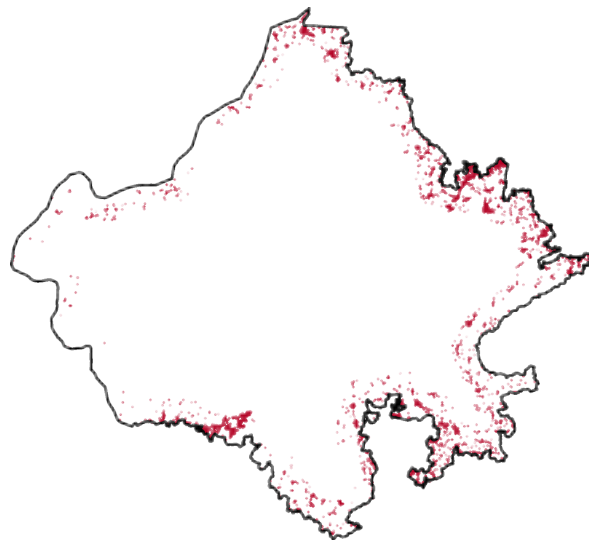
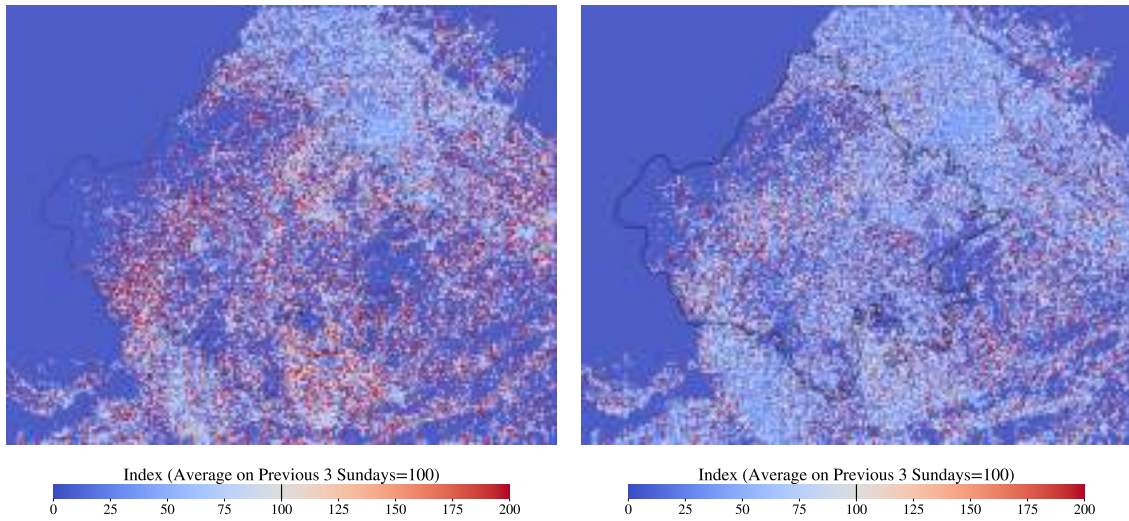


Figure 7: Catchment Area for Estimating Degree of Selection

Note: Figure plots pings inside 40km band within Rajasthan around the state boundary during the shutdown period. The degree of self-selection is estimated by comparing the share of devices associated with these pings that cross into neighboring states with its average in the prior month.



(a) Mobile: Raw Data, Day Of, 6 pm – Mid- (b) Mobile: Raw Data, Day After, 6 am – 6
night pm

Figure 8: Post-Shutdown Activity

Note: Plot shows the raw mobile device data on a 5x5km grid with the fill color of each cell indicating the value of \bar{Pings} , for the hours between 6 pm and midnight on the day of the shutdown in (a) and for the time spanning the shutdown window the day after the shutdown in (b). The outline of Rajasthan state is indicated by black lines.



(a) Esri Satellite Imagery

(b) OpenStreetMap

Figure 9: Rajasthan: Terrain View

Note: **9a** shows the satellite view of Rajasthan, obtained from Esri; **9b** shows the street map of Rajasthan obtained from OpenStreetMap. The outline of Rajasthan is depicted in black.