Hidden in Plain Sight:
Influential Sets in Linear Regression

Nikolas Kuschnig[1,*]

Gregor Zens[2,3]

Jesús Crespo Cuaresma[1,3,4,5,6]

[1] *Vienna University of Economics and Business*
[2] *Bocconi University*
[3] *International Institute for Applied Systems Analysis*
[4] *Wittgenstein Centre for Demography and Global Human Capital*
[5] *Austrian Institute of Economic Research*
[6] *CESifo*

Abstract

Influential sets are sets of observations that have considerable impact on econometric results. In this paper, we present a disciplined and insightful method for assessing the sensitivity of regression-based inference to influential sets. We explore algorithmic approaches to identify influential sets, discuss interpretation, and assess the sensitivity of earlier studies to these sets. We apply our method to established results in development economics, and show that results are driven by small influential sets. Identifying and analyzing these sets can reveal potential omitted variable bias, unobserved heterogeneity, a lack of external validity, and technical limitations of the methodological approach used.

**Keywords:** sensitivity, regression diagnostics, influence, development

1 Introduction

Econometric methods are an important instrument of scientific discovery, and are vital for the design of evidence-based policy. By approximating real-world phenomena, they provide us with empirical insights, allow us to test theories, and facilitate prediction. The sensitivity of these methods to modeling assumptions is a long-standing and active subject of research. The economic literature tends to focus on sensitivity along the *horizontal* dimension of the data, which is related to the functional form of model specifications. Examples go back to extreme bounds analysis (Leamer, 1983, 1985), and include model averaging (Steel, 2020), elaborate research designs (Angrist and Pischke, 2010), and randomization (Athey and Imbens, 2017). The sensitivity of inference to certain sets of observations, i.e., the *vertical* dimension of the data, however, has received less attention in the modern economic literature.

In this paper, we investigate the sensitivity of regression-based results to *influential sets* of observations. It is well-known that individual influential observations can drive the results of regression analysis (Cook, 1979). Influential sets may be substantially more impactful, but have not received much attention in the literature due to their inherent computational complexity. The empirical and theoretical analysis of these sets is intractable in most of the relevant cases, and approximate methods need to be used to identify them and assess their impact. Approximations, however, are prone to underestimate the impacts of influential sets due to two phenomena — *joint influence* and *masking*. Consider, for instance, the influence of sets of observations on the positive slope of the regression line presented in Figure 1. The three observations on the top right, marked 'a', are clearly influential, especially when considered as a set (see the top right panel). After removing them, the positive slope is driven by the observations on the center right, marked 'b'. This influential set, however, is masked — we only uncover its influence after accounting for set 'a' (see the bottom panels of Figure 1). These two phenomena affect the precision and accuracy of approximate methods for assessing influential sets of observations.

The main contribution of this paper is a disciplined and insightful method to identify and analyze influential sets in linear regression models, and to assess the sensitivity of inferential results to them. Our approach focuses on sets that are influential with respect to a particular quantity of interest, such as the size or significance of a coefficient. This way, we narrow down candidate sets to ones that (1) are immediately relevant to the question at hand, (2) allow for insightful interpretation, and (3) can be identified reliably. Our sensitivity check works by first identifying influential sets that overturn specific inferential results of interest, such as a positive sign or a significance threshold.

Figure 1: Influential sets in a univariate regression
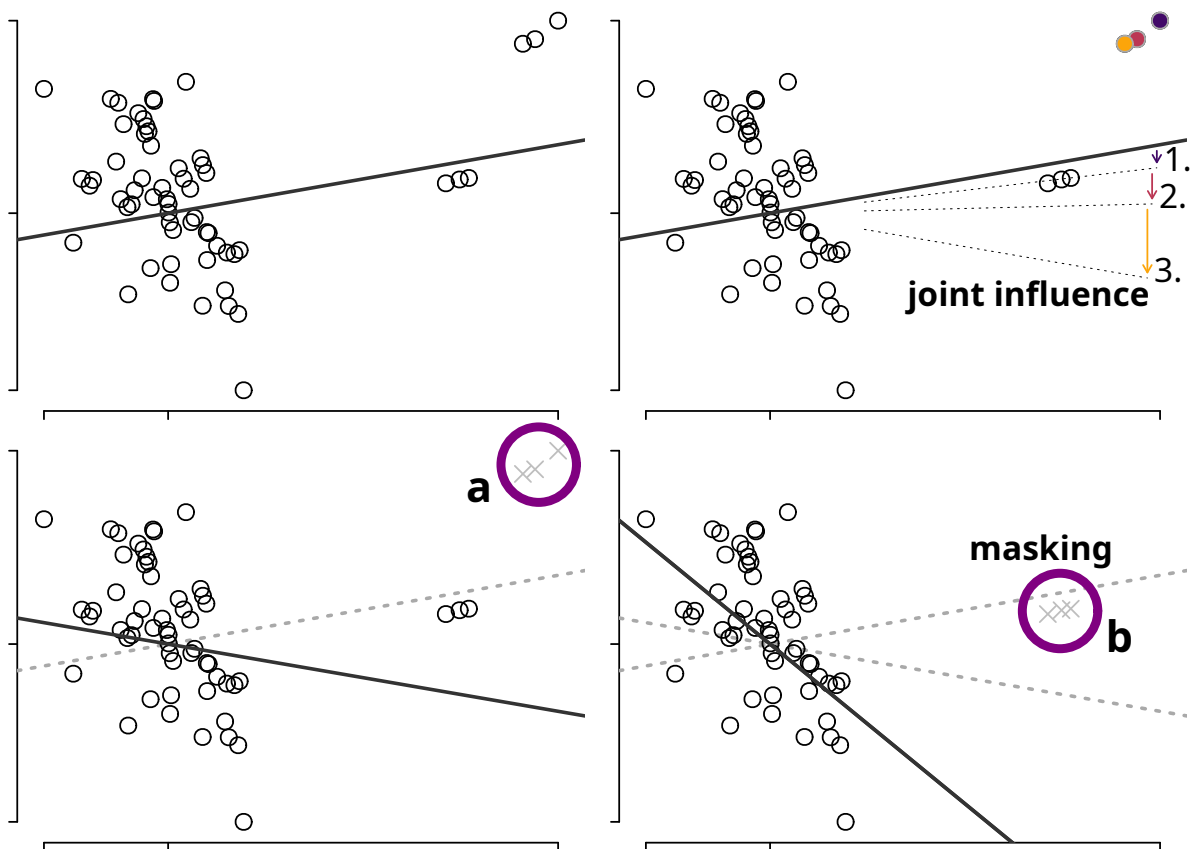
Univariate regression, with two sets of observations that are influential on the positive slope estimate, demonstrating joint influence and masking. In the top right panel, we can see that removing any member of a jointly influential set increases the influence of other members. In the bottom row, we can see the effects of masking. After removing set 'a' (left), the sign of the slope switches, and the influence of set 'b' is unmasked. After accounting for the influential set marked 'b' (bottom right), the slope is significantly negative.

Then, we analyze the identified influential sets, and investigate the characteristics of their members. We show that this analysis can provide insightful summary statistics, and can inform us about omitted variable bias, unobserved heterogeneity, a lack of external validity, and technical limitations of the methodological approach and data.

To implement our approach, we explore three computationally tractable algorithms to identify and account for influential sets. All three algorithms are straightforward to use, and differ in how they balance accuracy and precision against computational complexity. The first algorithm relies on the full-sample influence of individual observations for both the identification of sets and the estimation of their influence. This makes it cheap to compute, but comes at the price of accuracy and precision, as the algorithm cannot account for joint influence and masking. The second algorithm identifies influential sets in the same way, but computes the influence of a given set of observations exactly following a binary search pattern. This makes it extremely efficient computationally, and allows it to account for joint influence, but not masking. The third algorithm alleviates concerns with joint influence and masking by adaptively building an influential set. This way, joint influence is accounted for, and masked observations can be revealed, improving the accuracy and precision. We illustrate the performance of these algorithms using an illustrative example, and simulation exercises.

We revisit four established results in the field of development economics to demonstrate the virtue and wide applicability of our approach. We find that the mediating effect of rugged geography on income per capita in Africa (Nunn and Puga, 2012) and the development impact of the Tsetse fly (Alsan, 2015) hinge on small influential sets. Closer inspection of these sets and the underlying data suggest the existence of omitted confounders and issues related to limitations of the data employed. In a large-scale application to the origins of mistrust (Nunn and Wantchekon, 2011), influential sets are concentrated in terms of location, indicating heterogeneity in the effects considered. Technical sensitivity of instrumental variable estimates is illustrated in an application to the effects of skilled migration (Droller, 2018). Overall, we find that influential sets are prevalent in practice, and that they present an insightful sensitivity check in practice.

Our analysis builds on a long history of studies on the role of influential observations and outliers in the statistical and econometric literature.[1] In this paper, we specifically build upon the notion of influence (going back to at least Cook, 1979) and the influence function of Hampel et al. (2005), which we generalize to sets of observations. There are many approaches to identify and account for single influential observations (see Chatterjee and Hadi, 1986, for a review), and a number of useful results that allow for

---

[1]See, for example, Box (1953); Belsley et al. (1980); Hampel et al. (2005); Hansen and Sargent (2011); Huber (1964); Huber and Ronchetti (2009); Maronna et al. (2019); Rousseeuw and Leroy (1987).

efficient computation (e.g. Belsley et al., 1980; Phillips, 1977) that we build upon directly. In linear regression, single influential observations are generally easy to identify but limited in impact; most existing approaches to assess and account for sensitivity to these are of a holistic nature (this includes the statistics of Cook, 1979; Belsley et al., 1980, and approaches like winsorizing or trimming). The impact of influential sets of observations, however, is mostly sidestepped by the literature.

When sensitivity to subsets of the data is suspected, the most prominent statistical approach to correct for their influence is robust regression (see, e.g., Huber and Ronchetti, 2009; Lewis et al., 2021). Robust estimators, such as M-estimators and S-estimators, are (up to a breakdown point) resistant to a number of arbitrarily influential observation. Nonetheless, robust estimation is rarely used in practice. Important practical drawbacks (noted, for instance, in Stigler, 2010) include the elusive notion of robustness to arbitrary contamination, the loss of information, and the largely arbitrary choice of hyperparameters, which can have sizeable effects on the results.[2] By contrast, the approach proposed in this paper is directly relevant to a result of interest, involves limited choices beyond the influence function, and is both highly informative and readily interpretable. Moreover, we show that the types of sensitivity uncovered by robust estimation and influential sets differ considerably in practice.

An important and widely used approach that accounts for data sensitivity is based on resampling methods, such as the Bootstrap (Efron and Stein, 1981), the Jackknife (Efron and Tibshirani, 1994), or cross-validation (Picard and Cook, 1984). These methods build upon random samples of the data to produce estimates, and are regularly used for model selection and to quantify the overall uncertainty around estimates.[3] Our approach is related to these methods to the extent that we aim at uncovering the subsample that represents the worst-case scenario in terms of the robustness of the inferential result of interest. Another approach in the literature is concerned with detecting outliers. As opposed to influential observations, outliers are not well-defined (that is, they do not rely on an explicit influence function). This complicates the task of detection considerably, and makes (unsupervised) clustering methods a common choice for dealing with outliers in regression analysis (see e.g. Hyndman et al., 2015; Kaufman and Rousseeuw, 2009; Shotwell and Slate, 2011). Our preferred, adaptive algorithm is closely related to earlier sequential procedures for identifying multiple outliers (including ones by Belsley et al., 1980; Hadi, 1992; Hadi and Simonoff, 1993).

An important strand of the literature focuses on identifying or accounting for multiple influential observations (or outliers that are defined in more concrete terms). Going

---

[2]A 'monitoring' approach (Riani et al., 2014; Cerioli et al., 2018) can partly alleviate the last two issues.
[3]Model averaging methods can also be used to assess data sensitivity in a similar fashion (Steel, 2020).

back to at least Belsley et al. (1980), these approaches are often based on the influence of individual observations for the full-sample. A notable, recent example is the work of Broderick et al. (2021), who approximate the influence of a set of observations by accumulating individual influences that are computed using a linear approximation; this approach serves as an inspiration for our first algorithm. Another approach is taken by Peña and Yohai (1995), who employ the influence matrix as a heuristic to detect influential sets. We add to this literature in two ways. First, we formalize three algorithms for detecting influential sets and assess their susceptibility to the impacts of joint influence and masking. Our novel (second) algorithm considerably improves upon recent approaches by yielding precise results at negligible computational cost, and can readily be generalized beyond our linear setting. The second algorithm, which we formalize and implement in an efficient way, further improves accuracy and is rarely prohibitive to compute for linear models. Second, we show how analysis of the uncovered influential sets can yield deep insights into the issue at hand, and provide a guide for interpreting these sets.

By applying our approach to identify and assess influential sets, we contribute additional insights to the development and applied econometric literature literatures. First, we find that the differential effect of ruggedness in Africa as a determinant of economic development, which is considered a mediator for the impact of the slave trades (Nunn and Puga, 2012), may be a statistical artifact. By contrast, we find the impact of the slave trades on mistrust in modern-day Western Africa, which was subject to the Atlantic slave trades, to be considerably larger than previously thought (Nunn and Wantchekon, 2011). Next, our findings emphasize the longstanding need for more insightful summaries of the data (see Anscombe, 1973). We show that summaries based on influential sets could be an important addition to the regression statistics that are usually assessed. We also find that instrumental variable regression is particularly sensitive to influential sets. Simulation results (presented in the Appendix) show considerably more sensitivity than for ordinary least squares, even with strong instruments. This result complements the findings of Young (2022), who documents high susceptibility of instrumental variable regression to leverage. In addition, we document that poverty convergence across countries is not particularly sensitive to influential sets of observations (see the Appendix).

The remainder of this paper is structured as follows. In Section 2, we establish the theoretical framework. We present and illustrate the algorithmic approaches to identify and assess influential sets in Section 3. In Section 4, we demonstrate the analysis of influential sets by revisiting established empirical results on long-term development. We discuss the interpretation of sensitivity to influential sets, and illustrate with additional

applied examples in Section 5. In Section 6, we conclude. All codes and data used for this paper are available online,[4] and additional material is provided in the Appendix.

2   Theoretical framework

We start by providing the analytical framework upon which we build our influential set identification algorithms. We define influential sets, jointly influential sets and masking in this setting, and provide the link to popular measures of influence based on influential sets of single observations.

2.1   Influential sets

Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{1}$$

where $\mathbf{y}$ is an $N \times 1$ vector containing observations of the dependent variable, $\mathbf{X}$ is an $N \times P$ matrix with P explanatory variables, β is a $P \times 1$ vector of coefficients to be estimated, and ε is an $N \times 1$ vector of error terms with zero mean and unknown covariance matrix Σ. We denote the ith observation (row) of $\mathbf{y}$ and $\mathbf{X}$ as y_i and x_i. The deletion of an observation j is indicated with a subscript in parentheses, that is, $\mathbf{y}_{(j)}$ is the dependent vector without observation j.

A set of observations, $\mathcal{S}$, is defined as a non-empty subset of the set of all observations, i.e., $\mathcal{S} \subset \tilde{\mathcal{S}} = \{s | s \in \mathbb{Z} \cap [1, N]\}$. We use the shorthand $N_\alpha = \lceil N\alpha \rceil$ to denote a fraction $\alpha \in [0, 1]$ of the data, and indicate the cardinality of a set using a subscript, i.e., $|\mathcal{S}_\alpha| = N_\alpha$. The empty set is denoted by $\varnothing$ and the set of all sets of cardinality N_α by $[\mathcal{S}]^\alpha$.

We are interested in the impact of removing influential sets of observations on λ, some quantity of interest to the econometrician. A set of observations is an *influential set*, if its omission has a large impact on λ, when compared to most other sets of equal size. The impact of removing a set is measured with some (generalized) influence function Δ of λ and of the sets, $\mathcal{S}$ and $\mathcal{T}$ that are compared. In our example from Figure 1, for instance, we were interested in how the omission of sets of observations reduces the positive estimate of the slope coefficient. For this purpose, we can set

$$\lambda(\mathcal{S}) = \left(\mathbf{X}'_{(\mathcal{S})} \mathbf{X}_{(\mathcal{S})} \right)^{-1} \mathbf{X}'_{(\mathcal{S})} \mathbf{y}_{(\mathcal{S})}, \tag{2}$$

$$\Delta(\mathcal{S}, \varnothing, \lambda) = -\operatorname{sign}(\lambda(\varnothing)) \times \lambda(\mathcal{S}), \tag{3}$$

---

[4]The repository at `https://github.com/nk027/influential_sets` includes all scripts, data, and an R package with the approaches described in the paper (see Appendix A for more information).

where we will suppress the dependence on λ and the default value $\mathcal{T} = \varnothing$, and use the shorthand $\lambda_{(\mathcal{S})} = \lambda(\mathcal{S})$ for simplicity.

To summarize the sensitivity of λ, we focus on the *minimal influential set*, which we define as the smallest set whose omission achieves a select target impact on λ. We formalize this set by first defining the *maximally influential set*, $\mathcal{S}_\alpha^*$, which achieves the maximal influence for a given size of the omitted set, as

$$\mathcal{S}_\alpha^* = \arg\max_{\mathcal{S} \in [\mathcal{S}]^\alpha} \Delta(\mathcal{S}). \tag{4}$$

We can then formally define the minimal influential set, $\mathcal{S}^{**}$, with respect to the target impact of choice, Δ^*, as follows

$$\mathcal{S}^{**} = \mathcal{S}_{\arg\min_\alpha}^* \text{ s.t. } \Delta(\mathcal{S}_\alpha^*) \geq \Delta^*. \tag{5}$$

One relevant example is the minimal influential set that achieves sign switch of the slope in Figure 1, which can be achieved by setting $\Delta^* = 0$.

After obtaining a minimal influential set, we are interested in its size, both in absolute terms and relative to the full sample size, and the characteristics of its members. However, conclusively identifying such a set is computationally prohibitive — we would need to evaluate Δ for $\binom{N}{N_\alpha}$ potential sets. Instead, we rely on approximations for all but the most trivial settings. To assess the quality of these approximations, we introduce two concepts related to potential sources of error.

First, we define the notion of *joint influence*, which happens when a set of observations is jointly more influential than its members would be individually. Consider the influence of an observation i, given by the shorthand $\delta_i = \delta_{i|\varnothing} = \Delta(\{i\}, \varnothing)$. A *jointly influential set*, $\mathcal{U}$, satisfies $\delta_{i|\mathcal{J}} \gg \delta_{i|\varnothing}$ for all $i \in \mathcal{U}$ and non-empty $\mathcal{J} \subset \mathcal{U}$. The influence of any member i increases substantially after the removal of other members, as illustrated in the top right of Figure 1 for the jointly influential set 'a'. Joint influence is a major challenge for approximations, which rely on assessing a limited number of potential sets.

Second, we define *masking* as the phenomenon where members of a maximally influential set $\mathcal{S}_\alpha^*$ are not identified by the estimated set $\hat{\mathcal{S}}_\alpha^*$. This phenomenon can be observed in the bottom row of Figure 1, where set 'b' is masked by 'a' when considering sets of size four and above. Masking affects the accuracy of an approximation, and we can quantify its severity via the difference in influence between the true and estimated sets, or the cardinality of the set difference.

2.2 Computing influence estimates

The statistics literature is rich in results for identifying and dealing with single influential observations (see Chatterjee and Hadi, 1986, for a review). Effectively and efficiently measuring influence is central to this pursuit, and there is a wide variety of interrelated statistics serving this purpose. For most measures, the residuals ($\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$) and the leverage (the diagonal elements of the 'hat matrix', $\mathbf{H} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}$) are pivotal elements. One notable example directly measures the influence of an observation i on the OLS estimate of β, and can be expressed as

$$\delta_i = \lambda_{(\varnothing)} - \lambda_{(\{i\})} = \frac{\left(\mathbf{X'X}\right)^{-1} x_i' e_i}{1 - h_i}, \tag{6}$$

where e_i and h_i are the residual and leverage of observation i. This well-known quantity, termed DFBETA$_i$ by Belsley et al. (1980), facilitates the quick evaluation of the individual influence of all observations. Similar results are available for estimates of (spherical) variances, coefficient standard errors (Belsley et al., 1980), and two-stage least squares (2SLS) estimates (Phillips, 1977). Such convenient forms, together with efficient updating formulae (for example, to evaluate $(\mathbf{X}_{(i)}'\mathbf{X}_{(i)})^{-1}$) facilitate the computation of the influence, Δ. However, due to the combinatorial complexity, identifying a minimal influential set generally remains insurmountable, and approximate approaches are needed.[5]

3   Algorithms to assess influential sets

In this section, we formalize three simple algorithms for approximating minimal influential sets and their influence. Before we introduce the algorithms in detail, we illustrate their behavior for our example in Figure 1. The first algorithm, Algorithm 0, exclusively uses the full-sample influence of individual observations. As a result, it is extremely cheap to compute, but suffers in terms of accuracy and precision. Algorithm 1 identifies influential sets in the same way, but computes their influence exactly. This allows it to account for joint influence for little overhead. The last algorithm, Algorithm 2, uses a simple adaptive procedure for identifying influential sets, improving its precision and accuracy at slightly increased computational cost. Lastly, we discuss computational con-

---

[5]Consider a total number of observations of $N = 1,000$ and potential sets of size up to $N_\alpha = 10$. Assume that every calculation of λ needs one microsecond — very roughly the time needed to compute the cross-product of a four-by-four matrix. Enumeration would require about 8.35 billion years, or 1.8 times the age of the Earth, which is safely out of scope for non-tenured researchers.

cerns and the use of approximations for the identification and assessment of influential sets.[6]

3.1   An illustration

We illustrate the performance of the three algorithms via the example in Figure 1. Recall that we are interested in assessing how the omission of sets of observations reduces the slope of the OLS regression line, as expressed by Equations 2 and 3. The set marked 'a' (in the top right of the scatter plot) is clearly influential. This is generally reflected by standard diagnostics, and all three algorithms correctly identify this influential set. For the set marked 'b' (in the center right), which is initially masked by the first set, this is not the case. Neither Algorithm 0 nor Algorithm 1 identify any of its elements. The adaptive nature of Algorithm 2, however, allows it to identify the masked observations and account for their influence.
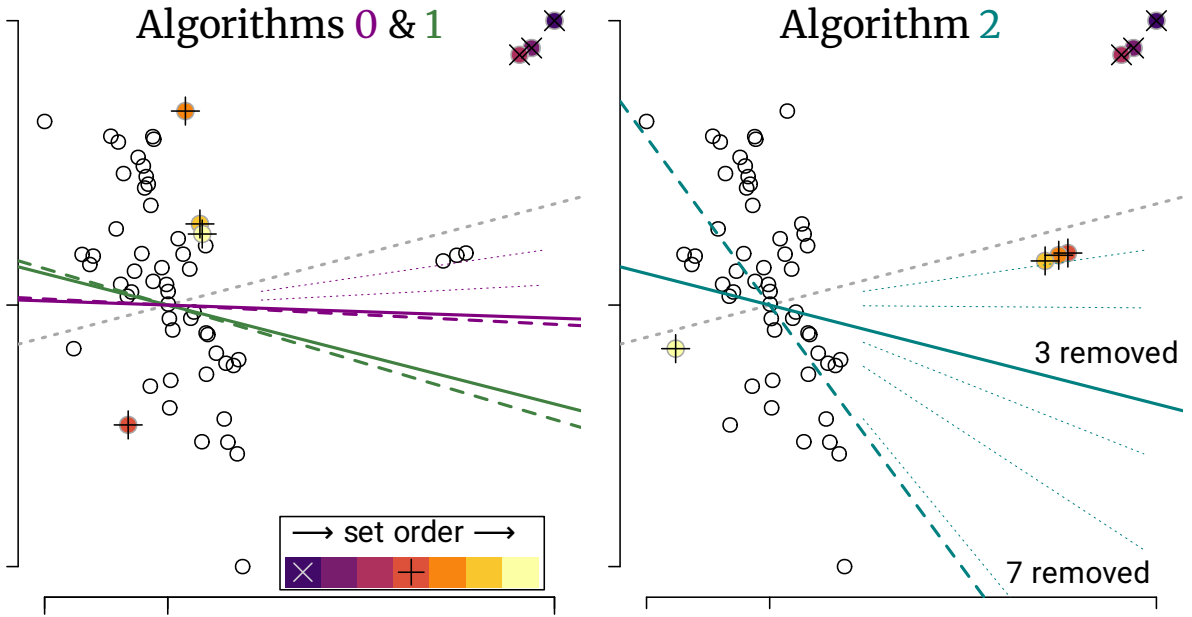

Figure 2: Influential sets with Algorithms 0, 1, and Algorithm 2. The order of the identified observations is color-coded and indicated by a cross (×, first three) and a crosshair (+, next four). The dashed gray line indicates the full-sample regression line. The solid and dashed lines are the implied regression lines after accounting for influential sets of sizes three and seven. Thin dotted lines show the influence of intermediate steps for Algorithms 0 and 2.

The results of applying the three algorithms to identify and account for influential sets of size three and seven are depicted in Figure 2. In the left panel, we can see that

---

[6]Additional results comparing the performance of the algorithms in simulated and applied settings are provided in Section B3 of the Appendix.

Algorithms 0 and 1 are affected by masking after set 'a' is identified. Algorithm 1 precisely accounts for this first set, but subsequent removals are relatively inconsequential. Meanwhile, Algorithm 0 cannot account for the joint influence of set 'a', underestimating the influence after the first removal. This is illustrated by the intermediate slopes (thin dotted lines), which decrease in impact after each removal. In the right panel, we can see that Algorithm 2 does not suffer from these problems. The implied slopes are accurate and precise, and intermediate steps clearly show how the joint influence of sets increases the impacts of individual removals. For these reasons, Algorithm 2 is our preferred choice for later empirical applications.

3.2 Algorithm 0: Initial approximation

Algorithm 0 builds on the full-sample influence of single observations. Maximally influential sets are proposed based on the order of individual, full-sample influences, and their influence is approximated by accumulating these individual influences. In our linear setting, it generalizes the approach of Broderick et al. (2021), who use a linear approximation to compute Δ and accumulate influences.[7]

Algorithm 0: Initial approximation.

set the function Δ, the target Δ^*, and the maximum size U, let $\mathcal{S} \leftarrow \varnothing$;
compute $\delta_i = \Delta(\{i\})$ for all $i \in \tilde{\mathcal{S}}$;
while $\Delta(\mathcal{S}) < \Delta^*$ **do**
    let $\mathcal{S} \leftarrow \mathcal{S} \cup \arg\max_j \delta_j$, for $j \notin \mathcal{S}$;
    let $\Delta(\mathcal{S}) \leftarrow \sum_{k \in \mathcal{S}} \delta_k$;
    if $|\mathcal{S}| \geq U$ **then return** *unsuccessful*;
end
return $\mathcal{S}, \Delta(\mathcal{S})$;

The algorithm works as follows. First, we set an influence function, target value, and a maximum size for the minimal influential set. Next, we compute the initial influence δ_i for each i. As discussed above, this computation is relatively cheap in many interesting cases; otherwise, approximations could be used. The first iterated step is the proposal of a maximally influential set, which is based on the union of the observations with the largest individual influences. The influence of this proposed set is then estimated by summing the individual influences of observations in the set. These two steps are repeated until the specified target or the maximum size is reached.

---

[7]See Section B2 in the Appendix for information on the approach used by Broderick et al. (2021).

The method embodied in Algorithm 0 can yield striking results, as demonstrated by the findings of Broderick et al. (2021). However, its low computational complexity comes at the price of precision and accuracy when influential observations are present. First, the algorithm is prone to underestimate joint influence due to summing up individual influences. Second, the estimated influences are not updated, and observations may remain masked behind the influence of already removed observations.

3.3 Algorithm 1: Initial binary search

Algorithm 1 rectifies the precision issues of Algorithm 0, while retaining high computational efficiency. Similar to Algorithm 0, the identification of maximally influential sets is based on the ordering of the individual, full-sample influences. The influence of these proposed sets, however, is calculated exactly. In order to guarantee low computational cost, the algorithm follows a binary search pattern when proposing sets.

Algorithm 1: Initial search.

set the function Δ, the target Δ^*, and the maximum size U, let $L \leftarrow 0$;

compute $\delta_i = \Delta(\{i\})$ for all $i \in \tilde{\mathcal{S}}$;

while $L \leq U$ **do**
  let $M \leftarrow \lfloor (L + U)/2 \rfloor$;
  let $\mathcal{S}$ be the union of the indices of the M largest δ_i;
  compute $\Delta(\mathcal{S})$;
  if $\Delta(\mathcal{S}) < \Delta^*$ **then** let $L \leftarrow M + 1$;
  else let $U \leftarrow M - 1$;
end

if $\Delta(\mathcal{S}) \geq \Delta^*$ **then return** $\mathcal{S}, \Delta(\mathcal{S})$;

else return *unsuccessful*;

Algorithm 1 works by iteratively setting the size of proposed sets, M, to halve a search interval $[L, U]$. In each step, the influence of the proposed set is computed exactly, and the bounds of the search interval are updated depending on the influence of the proposed set. If the target is reached, the upper bound is decreased to $M - 1$, otherwise the lower bound is increased to $M + 1$. If an approximate minimal influential set exists in the search interval, it is found after $\mathcal{O}(\log U)$ steps. As a result, Algorithm 1 adds negligible computational overhead over Algorithm 0, making this divide-and-conquer approach practical for computationally prohibitive problems.

By computing Δ explicitly, Algorithm 1 can account for joint influence, and yields precise influence estimates for a given set. However, its accuracy is not guaranteed, as

masking remains an issue. Because influence estimates are not updated, influential observations are likely to remain hidden behind already removed observations. The algorithm also relies on the (previously implicit) assumption that the influence of estimated maximally influential sets increases monotonically with the size of the set. The next algorithm abstracts from this assumption and is designed to explicitly address masking.

3.4 Algorithm 2: Adaptive approximation

The third algorithm (Algorithm 2) uses a simple adaptive procedure for identifying the minimal influential set. The maximally influential sets are constructed iteratively, using updated influence estimates. This facilitates the discovery of masked observations, improving accuracy. The adaptive nature of this algorithm allows for good accuracy and precision, while retaining computational tractability.

Algorithm 2: Adaptive approximation.

choose the function Δ, the target Δ^*, and the maximum size U, let $\mathcal{S} \leftarrow \varnothing$;

while $\Delta(\mathcal{S}) < \Delta^*$ **do**

    compute $\Delta(\mathcal{S} \cup j)$, for all $j \in \tilde{\mathcal{S}} - \mathcal{S}$;

    let $\mathcal{S} \leftarrow \mathcal{S} \cup \arg\max_j \Delta(\mathcal{S} \cup j)$;

    if $|\mathcal{S}| \geq U$ **then return** *unsuccessful*;

end

return $\mathcal{S}, \Delta(\mathcal{S})$;

Algorithm 2 starts by computing the influence of all individual observations. Then, maximally influential sets are proposed adaptively, by forming the union of the previous set (starting with the empty set) and the observation with the highest influence (at each step). This is repeated until a minimal influential set is found, or the maximal size is reached. The algorithmic complexity is linear in the cardinality of the set and falls well short of, e.g., a Jackknife approach. By computing individual influences after every removal, this approach reduces the risk of masking problems and allows us to more reliably investigate sensitivity to influential sets.

3.5 Approximations and computational concerns

The algorithms presented above are computationally straightforward, with complexities that are constant, logarithmic, and linear in the size of the minimal influential set. In our linear setting, most quantities of interest rely on matrix factorization, with a complexity

of $\mathcal{O}(NP^2)$. This means that the computation of the influence function, Δ dominates the overall complexity in practice. Fortunately, many quantities of interest, such as coefficients and standard errors, can be computed efficiently using algebraic results (for instance, Equation 6) and updating formulae (for cross products, inverses, and factorizations). Together with optimization of the underlying linear algebra, this allows us to quickly assess the sensitivity to influential sets in a wide range of applications.

Further speed gains can be realized with approximate methods. When the number of covariates is large, it can be helpful to marginalize out nuisance covariates before computing influence measures. Approximate methods are another option to speed up computation. There is, for instance, no convenient method to update clustered standard errors, and their computation becomes prohibitive with large N or P. By using classical standard errors to propose sets, we can facilitate sensitivity checks at large scales. Such simplifications, however, can be problematic if a large part of the influence is driven by the element which is simplified to achieve computational gains.

In our setting, one example is the linear approximation to the influence of individual observations on coefficient estimates (i.e. to Equation 6) that Broderick et al. (2021) propose. This approximation is given by $\delta_i \approx (\mathbf{X}'\mathbf{X})^{-1} x_i' e_i$, and disregards the role of leverage (see Section B2 of the Appendix for more information). As a result, the approximation underestimates the influence of all observations, and features a particularly large downward bias for observations with high leverage, and thus (ceteris paribus) high influence.

4   Influential sets and economic development

In this section, we investigate the sensitivity of empirical results related to the determinants of long-term economic development to influential sets. We concentrate on the influence on the t value of the main estimate of interest. For the thresholds, we follow Broderick et al. (2021) in reporting minimal influential sets that induce (1) a loss of significance at a given level, (2) an estimate of the opposite sign, and (3) a significant estimate of the opposite sign. We use Algorithm 2 unless mentioned otherwise.[8]

Identifying the driving forces of economic development has been a central endeavor in development economics over the last decades. Factors such as colonial experiences (Acemoglu et al., 2001), the slave trades (Nunn, 2008), precolonial institutions and centralization (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013), as well as migration (Tabellini, 2020) are known to play a fundamental role. The pathways through which these determinants affect development today are an important

---

[8]See Sections B3 and C4 of the Appendix for further material on the comparison of algorithms.

subject of research (Spolaore and Wacziarg, 2013). Three empirical studies on economic development in Africa find that the slave trades affect development via ruggedness (Nunn and Puga, 2012) and interpersonal trust (Nunn and Wantchekon, 2011), and that the Tsetse fly affects precolonial centralization (Alsan, 2015). Another important study on the Age of Mass Migration uses an instrumental variable approach to show the long-term impacts of skilled migration (Droller, 2018). Below, we assess the sensitivity of the empirical findings presented in these three studies to influential sets of observations.

4.1   Geography, development, and omitted variables

Exploiting cross-country differences, Nunn and Puga (2012) estimate linear regressions that relate GDP per capita to the ruggedness of terrain and other control variables. They allow for heterogeneity in the effect of ruggedness for African countries via an interaction term. Nunn and Puga (2012) find a significantly negative overall estimate of the effect of ruggedness, and a significantly positive estimate for the additional effect in African countries (see Table 1, Column 1). This differential effect is interpreted as rugged geography offering protection against the slave trades, and thus counteracting the negative effects of ruggedness on long-run economic development. The authors perform rigorous checks to address the sensitivity of these results to influential observations, and their findings appear to be robust. In particular, the results are unchanged when omitting observations that exceed a threshold of $|\text{DFBETA}_i| > 2/N$ (following Belsley et al., 1980), as well as the ten smallest and most rugged observations. In addition, we find that robust M-estimates of the effect support this conclusion (see Table 1, Column 2).

We investigate the sensitivity of this differential effect of ruggedness in Africa to influential sets. In the row labeled 'Thresholds' of Table 1, we report the sizes of influential sets that induce (1) an insignificant estimate, (2) an estimate of the opposite sign (in square brackets), and (3) a significant estimate of the opposite sign (in curly brackets) for the coefficient of interest. We find that an influential set of two observations overturns the significance of the differential effect. An influential set of five observations is enough to induce a sign-switch, and one of size eleven achieves a statistically significant sign-switch.

To provide more context for these observations, we visualize the five most influential ones in Figure 3, Panel A. Their influence is visualized in Panel B, which shows the effect of each subsequent observation (indexed by the ISO code of the corresponding country) on the coefficient's t value. Starting at a value of 2.53, the t-value decreases slightly with the removal of the most influential observation, which is of the Seychelles.

Table 1: The differential effect of ruggedness in Africa

	Baseline	Robust	Population	Land area
Ruggedness, Africa[†]	0.321	0.325	0.190	0.215
	(2.53)	(2.46)	(1.66)	(1.63)
Ruggedness	-0.231	-0.251	-0.231	-0.238
	(-2.99)	(-3.23)	(-2.94)	(-3.08)
Controls	Yes	Yes	Yes	Yes
Population in 1400	–	–	Yes	–
Land area	–	–	–	Yes
Thresholds[†]	2[5]11	—	–[3]6	–[4]8
N	170	170	168	170
R^2	0.537	0.533	0.571	0.554

The row labeled 'Thresholds' reports the sizes of influential sets that induce a loss of significance (at the 5% level), [a sign flip], and a significant sign flip of the coefficient that captures the differential effect of ruggedness in Africa (using Algorithm 2). The 'Baseline' column reproduces the results of Nunn and Puga (2012, see Table 1, Column 6) using OLS estimation, while 'Robust' uses robust M-estimation. The specifications labeled 'Population' and 'Land area' add the population level in the year 1400, and the land area of the country (both in logs) as covariates. Coefficient estimates are reported with t values based on HC1 robust standard errors in parentheses.

Its influence (as indicated by the height of its ISO code in the visualization) appears limited, but when using Algorithm 2 (on the left) we can clearly see that the influence of subsequent removals increases considerably. This pattern is indicative of a jointly influential set, which remains hidden from Algorithm 0 (on the right).

Figure 3: Two nations drive the blessing of bad geography
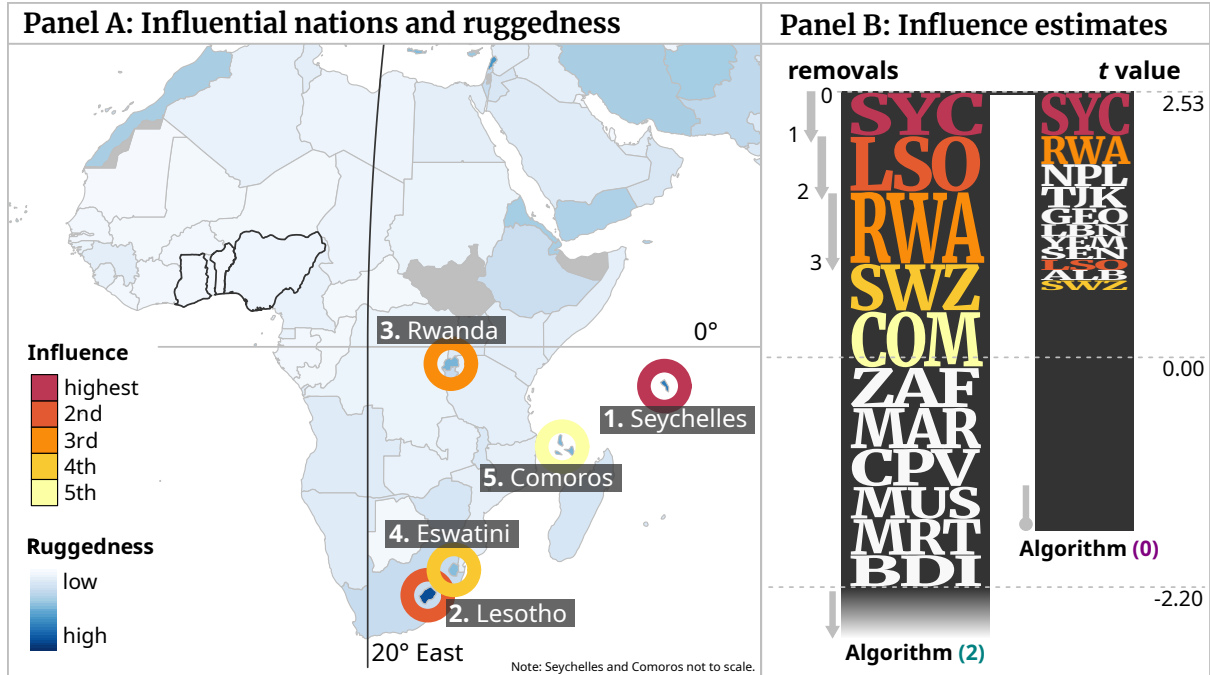
Panel A visualizes ruggedness, the explanatory variable of interest, and the five most influential observations. Panel B shows the cumulative reduction of the t value as observations (indicated with their ISO codes) are removed (from 2.53 at the top, to the bottom of the respective observation) using Algorithms 2 (left) and 0 (right).

Sensitivity checks to influential sets can provide valuable insights, allowing us to analyze the minimal influential set and the characteristics of its members. In this case, the influence of the Seychelles casts doubts on the interpretation presented in Nunn and Puga (2012). The island nation has only been inhabited permanently since the late 18th century (Fauvel, 1909), and its ruggedness played no role in mediating the effects of the slave trades. In addition, the five most influential nations are extraordinarily small in terms of land area and population. This may indicate a survivorship bias, where the inclusion of countries in the dataset is partly determined by land area, population, geography, or economic success. Past population sizes are likely related to confounding geographical features, and can also play an important role in mediating the impacts of the slave trades. In Columns 3 and 4 of Table 1, we present the results of specifications that include past population size and land area as additional controls. In these cases, no significant differential effect of ruggedness on economic development can be found for the African continent.

4.2 Slave trades and the heterogeneous origins of mistrust

Nunn and Wantchekon (2011) analyze the role of the Atlantic and East African slave trades as determinants of interpersonal trust, using individual-level survey data and historical information on the slave trades. They regress measures of the trust of relatives and neighbors (among others) on a measure of slave trade intensity, as well as a number of individual and district-level covariates and fixed effects at the country level. The design matrix of the regression model contains over 20,000 observations that span several countries in sub-Saharan Africa, and a total of 78 regressors. Uncertainty is quantified using standard errors that are clustered along ethnicity and district. The authors find statistically and economically significant negative effects of the slave trades on interpersonal trust. Their findings are robust to removing Kenya and Mali (which were also impacted by the trans-Saharan and Red Sea slave trades) from the sample.

Table 2: The origins of mistrust

	Trust of relatives		Trust of neighbors	
	Pooled	West | East	Pooled	West | East
Exports/area[†]	-0.133	-0.145	-0.159	-0.168
	(-3.68)	(-3.84)	(-4.67)	(-4.48)
Exports/area, East		0.053		0.023
		(0.96)		(0.32)
Individual controls	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Thresholds[†]	105[380]656	78[301]532	161[425]768	133[323]527
N	20,062	7,549 | 12,513	20,027	7,523 | 12,504
Ethnicity clusters	185	62 | 123	185	62 | 123
District clusters	1,257	628 | 651	1,257	628 | 651
R^2	0.133	0.199 | 0.097	0.156	0.228 | 0.117

The row labeled 'Thresholds' reports the sizes of influential sets that induce a loss of significance (at the 5% level), [a sign flip], and a significant sign flip of the coefficient that captures the effect of the slave trades (measured as exports of slaves per area) on trust (using Algorithm 2). The columns labeled 'Pooled' reproduce the results of Nunn and Wantchekon (2011, see Table 2, Columns 1 and 2). The columns labeled 'West | East' estimate separate models for observations to the west and east of the 20° Eastern meridian (thresholds refer to Western subset). Coefficient estimates are reported with t values based on two-way clustered standard errors in parentheses.

For this analysis, a sensitivity check based on the influential sets is computationally challenging, since no updating formula for two-way clustered standard errors exists.

To carry out our analysis, we only cluster standard errors after the removal of an observation, and not for proposed removals. We reproduce the empirical results of Nunn and Wantchekon (2011) in Table 2, and find that an influential set of size 105 (0.5% of the sample) induces a loss of significance of the estimated effect of the slave trades on trust of relatives, while 380 removals (1.9% of the sample) lead to a sign-flip that becomes significant after 656 removals (3.3% of the sample) for the trust of relatives. Results are similar for the trust of neighbors.

An analysis of these influential sets shows that 536 of the 600 most influential observations (89.3%) stem from three West African nations: Benin, Nigeria, and Ghana (marked with black borders in Figure 3, see also Table S2 in the Appendix for more details). These nations were major centers of the Atlantic slave trade, and their large influence may suggest differences in impacts between the Atlantic and East-African slave trades. When dividing the dataset into a Western and Eastern sample (with the 20° Eastern meridian as the dividing line), we find a significantly negative effect for the Western sample, and an insignificant effect for the Eastern sample (see Table 2, Columns 2 and 4). This result is consistent with the literature, which suggests larger impacts from the Atlantic slave trade on economic development (see Nunn, 2008).

4.3 The concentrated effects of the Tsetse fly

The Tsetse fly is an important vector of disease and considered a hindrance to early economic development. Alsan (2015) empirically investigates this detrimental effect, concentrating on its role as a determinant of agricultural practices, urbanization, and institutions. The empirical analysis is based on regressing various outcomes at the ethnic group level on a Tsetse suitability index (TSI) and a number of controls, clustering standard errors at the province level. Using ethnicity-level data, Alsan (2015) finds significant detrimental effects of the TSI across the board, with precolonial centralization as the major channel affecting present economic development. These results are robust to perturbations of the TSI and to corrections aimed at assessing the negative selection bias that may occur due to more developed ethnic groups displacing less developed ones.

We reproduce these results in Table 3, and investigate the sensitivity of the estimated effects to the sample. First, we focus on the sensitivity to influential sets. We find that the significance of results is induced by few observations, ranging from one (for the variable measuring whether a centralized state was present) to 33 (for the possession of large domesticated animals). Influential sets of sizes between 17 and 79 induce significant results of the opposite signs. The sensitivities of the empirical results in

Table 3: The effects of the Tsetse fly

	Animals	Intensive	Plow	Female	Density	Slavery	Central
TSI[†]	-0.231	-0.09	-0.057	0.206	-0.745	0.101	-0.075
	(-5.47)	(-3.29)	(-2.54)	(3.41)	(-3.25)	(2.51)	(-2.12)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Thresholds[†]	33[58]79	7[25]41	3[12]17	12[30]48	9[27]42	4[22]35	1[16]30
N	484	485	484	315	398	446	467
Clusters	44	44	44	43	43	44	44
R^2	0.296	0.268	0.462	0.285	0.254	0.178	0.139

The row labeled 'Thresholds' reports the sizes of influential sets that induce a loss of significance (at the 5% level), [a sign flip], and a significant sign flip of the coefficient that captures the effect of the TSI on the respective dependent variable. The dependent variables are binary and continuous, and measure whether a precolonial ethnic group (1) possessed large domesticated 'Animals', (2) adopted 'Intensive' agriculture, (3) adopted the 'Plow', (4) had 'Female' participation in agriculture, (5) log population 'Density', (6) practiced indigenous 'Slavery', and (7) had a 'Central'ized state. Coefficient estimates are reported with t values based on clustered standard errors in parentheses.

Alsan (2015) to excluding small sets of observations is pronounced, with a possible exception for the effects on the domestication of large animals. The findings of Alsan (2015) are supported by a relatively small amount of data, which may stem from a lack of variation in the outcome variables. In spite of the reasonable sample size, the six binary outcomes are relatively rare (the plow adoption, for example, only occurred for 37 of the 484 observations).

For comparison, we use robust M-estimation, and find that five of the seven specifications are effectively unchanged (see Table S3 in the Appendix for the estimation results). Two specifications (those referring to plow adoption and indigenous slavery as outcomes) yield a pathological fit. We also repeat the empirical analysis using a high-breakdown S-estimator. The effects on population density (which are not particularly robust to influential sets) remain unchanged. Meanwhile, the six other specifications yield a pathological fit, and between 106 and 286 observations (or 22 and 91 percent of the sample) are flagged as outliers. Thus, conclusions based on robust estimators depend on the type of estimator, and can differ strongly from ones based on influential sets. While robust estimators also flag sets of impactful observations, these tend to be larger than the influential sets obtained with the methods presented here, and are not as immediately relevant to a particular inferential result of interest.

4.4 Migration, instrumental variables, and influential sets

Migration reshapes global populations and, as a result, economic and cultural structures. In a recent study, Droller (2018) investigates the long-term impacts of European migration to Argentina in the late 19th and early 20th century. He uses a shift-share instrument (Borusyak et al., 2022; Goldsmith-Pinkham et al., 2020), to identify the causal effect of the share of European migration on GDP per capita in 2000. The sample includes 136 counties in the provinces of Buenos Aires, Santa Fe, Córdoba, and Entre Rios. The results in Droller (2018) show considerable impacts of skilled European migrants, who played an important role in industrialization, on GDP per capita, education outcomes, and skilled labor. The results are shown to be robust to a number of controls and to changing the construction of the instrument.

Table 4: The long-term migration impacts on development

log GDP/capita ∼	Baseline		Plain	
	1$^{\text{st}}$ stage	2SLS	1$^{\text{st}}$ stage	2SLS
European share		5.778		6.35
		(3.40)		(4.31)
Instrument	0.251		0.281	
	(6.01)		(4.84)	
Geographic controls	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	–	–
Thresholds†	10[18]26	5[11]27⟨18⟩	10[18]31	6[11]32⟨18⟩
N	136	136	136	136
R^2	0.816	0.622	0.688	0.591

The row named 'Thresholds' gives the numbers of observation necessary to remove significance (at the 5% level), [flip the sign], and significantly flip the sign of either the shift-share instrument (in columns labelled '1$^{\text{st}}$ stage'), or the effects of the share of European migrants (as reflected by 2SLS estimates); the angled brackets indicate the number of removals until ⟨2SLS estimates are numerically unstable⟩. The columns labelled 'Baseline' reproduce the 'Specification 2' in Table 4 of Droller (2018), the ones labelled 'Plain' reproduce 'Specification 1'. Coefficients are reported with t values based on HC0 robust standard errors in parentheses.

We reproduce the results of Droller (2018) in Table 4, and identify influential sets for both the first stage OLS, and the full 2SLS estimates. The first stage results are comparatively insensitive to influential sets. Ten observations need to be removed for the instrument to lose statistical significance in the first stage, suggesting a strong instrument. For the 2SLS estimates of the baseline specification, we find that an influential set of size five overturns significance, while one of size eleven induces a sign-switch. With

an influential set of size 27, a significant sign-switch is induced. It is notable that a mere 18 targeted removals (using $\Delta(\mathcal{S}) = -|t_{(\mathcal{S})}|$ as an influence function) can lead to pathological numerical stability (illustrated in the Appendix, Figure S8) and eliminate the relevance of the instrument.

This example illustrates how influential sets can arise from and unveil technical limitations of an estimator for a given dataset. In this case, the instrument is based on initial population shares that are shifted with total migrant flows. As a result, it features a strong geographical component that creates strong correlation structures between the instrument and the controls.[9] In addition, the HC0 standard errors used in the analysis are well-known to be susceptible to observations with high leverage (Cribari-Neto et al., 2007). Most importantly, 2SLS estimates themselves may be more susceptible to influential sets in general, due to the particular role of leverage (also see Young, 2022). This finding is supported by the simulation results presented in Section B4 of the Appendix.

5 Discussion

In the applications presented above, the main results appeared sensitive to influential sets of observations. Analysis of the nature of these influential sets helped reveal problems related to omitted variable bias, heterogeneity, or limited support from the data. However, the question of how to interpret minimal influential sets in general remains unanswered. In this section, we elaborate on this issue, provide some additional illustrations based on empirical examples, and give recommendations to applied researchers that are interested in assessing the sensitivity of empirical results to influential sets of observations.

There is no definite guide for the interpretation of influential sets. Expert knowledge is vital, and the expected degree of sensitivity provides important context for the interpretation. When searching for the proverbial needle in the haystack, we expect high sensitivity to little data. On the other hand, we expect to find a good amount of needles when investigating a sewing kit. Analyses of rare events such as disease outbreaks or economic crises, as well as policy interventions that only affect a small part of the population, are inherently sensitive to small subsets of the data. Phenomena that appear more universally, such as convergence in growth or income levels across economies, are expected to only be susceptible to the exclusion of larger fractions of

---

[9]The full-sample design matrix already displays a high degree of collinearity. Out of the fourteen variables, three have variance inflation factors over ten (including the variable whose effect is of interest), and eight have variance inflation factors over five.

the data. Having a good understanding of the distribution and intensity of the effects hypothesized (and of other underlying factors), allows us to gain valuable insights from analyzing influential sets.

First, we consider the role of influential sets when analyzing rare, concentrated phenomena. The effects of the slave trades on mistrust, for instance, are likely to manifest only in few impacted communities. Hence, the identified minimal influential set for a loss of statistical significance (whose size is of 105 observations, which corresponds to 0.5% of the sample) does not appear to be particularly worrying, and arguably turns out to be insightful, highlighting subsamples that are particularly affected. In the case of the effects of the Tsetse fly, we consider the sensitivities to be more problematic. The small cardinality of the influential sets identified imply that the results hinge on deceptively few observations. More generally, influential sets can bring potential heterogeneities in effects to the attention of the researcher, allowing for better summaries of the results and the use of more appropriate methods. One example concerns the impacts of microcredits on income (see Section B2 of the Appendix, and Broderick et al., 2021).

Second, we consider the context of more universal phenomena. One notable example whose robustness has been discussed extensively in the literature concerns the convergence of poverty rates across countries. Based on theoretical considerations regarding income convergence and the link between mean income growth and poverty dynamics (see Johnson and Papageorgiou, 2020, for a survey), global convergence in poverty rates is expected in the data. Ravallion (2012), however, finds statistically insignificant poverty convergence rates in a sample of 89 countries. As Crespo Cuaresma et al. (2016, 2022) point out, this result is likely to be driven by the idiosyncratic experience of Eastern European countries. An analysis of influential sets in their setting reveals exactly this result: an influential set formed by the observations of Belarus, Latvia, Ukraine, and Poland is behind the lack of significance of the convergence estimate in Ravallion (2012). An alternative specification (proposed by Crespo Cuaresma et al., 2016) leads to estimates of the convergence rate that are less sensitive to influential sets of observations (see Section C4 in the Appendix for more information on these results).

Finally, we consider the technical circumstances that give rise to an inferential result. Three tightly connected cornerstones are the composition of the data, the identification of parameters, and the estimation method employed. By their nature, OLS estimates are sensitive to observations with high leverage and large residuals, while shrinkage estimators will be less impacted by such data points. On the other hand, the compound nature of 2SLS estimators makes them more susceptible to small influential sets, and numerical instability of the estimators can quickly become a problem — as illustrated in the context of long-term impacts of migration (as studied by Droller, 2018), and

suggested by additional simulation results in Section B4 of the Appendix. Evidently, technical sensitivity is related to the identification of parameters and to the variability of the data. With 37 incidences, for instance, there is little variation to identify the effect of the Tsetse fly on plow adoption, and the type of sensitivity is unsurprising. Technical sensitivities are an inescapable feature of empirical analysis (see e.g. Young, 2022), and the identification of influential sets can bring potential issues to our attention.

To summarize and provide more concrete recommendations, we can conclude that sensitivity to influential sets of observations is highly context-dependent, and cannot be used to attest robustness of a given result in a definitive manner. However, the analysis of influential sets provides intuitive and insightful tools to shed light on a previously ignored dimension of sensitivity. Minimal influential sets that are small in relative terms may offer deeper insights, while ones that are small in cardinality, i.e. absolute terms, may convey deeper issues. Results that are driven by a single influential observation are rightly suspect; three or four thereof in an influential set likewise warrant a closer look. To address this hidden dimension of sensitivity, summaries based on the size of minimal influential sets behind the size and statistical significance of effects can help increase the transparency and quality of empirical research.

6   Conclusions

In this paper, we presented a method to assess the sensitivity of inferential quantities in linear regression to influential sets of observations. We showed how masking, where certain observations obscure the influence of others, and joint influence complicate sensitivity checks by inducing false negatives, and explored ways to address them. We investigated three empirical studies on the determinants of economic development in Africa for influential sets, and demonstrated the practical relevance and utility of our approach. Our analysis showed that sensitivity to influential sets is widespread among recent empirical studies in the field of development economics, and that their assessment can lead to important insights concerning phenomena under scrutiny.

In applied econometrics, the sensitivity to (sets of) observations is still rarely assessed in a systematic manner. Reinforced by our findings in this paper, we believe that influential sets have a role to play in the communication of results. They can deliver valuable insights, and summaries based on them, such as their absolute and relative size, can serve as intuitive indicators to inform econometricians and their audience, improving transparency in the communication of regression results. Two important advantages over other measures such as Cook's distance or robust estimates are their interpretability and salience. Sensitivity is directly tied to a statistic of interest, and influential sets can

easily be analyzed in detail. One disadvantage is that influential sets are necessarily approximate, and robustness cannot be attested conclusively. Nonetheless, many false negatives can be avoided at a reasonable computational cost, and approximations can unveil important insights that would otherwise remain hidden in plain sight.

There are several pathways for future work building upon this contribution. On the topic of identifying influential sets, there is certainly room for algorithms that yield more accurate results at reasonably higher computational complexity. Potentially fruitful approaches could use a probabilistic approach informed by heuristics that are well-suited for the specific influence at hand. In a wider sense, there is a lack of comprehensive indicators measuring the sensitivity of statistical inference to the data. The approach presented here opens the door for further replication studies and sensitivity checks of documented empirical phenomena, which may deliver valuable insights for researchers and policymakers alike.

References

Daron Acemoglu, Simon Johnson, and James A. Robinson. The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5):1369–1401, 2001. ISSN 0002-8282. doi:10.1257/aer.91.5.1369.

Marcella Alsan. The effect of the TseTse fly on African development. *American Economic Review*, 105(1):382–410, 2015. ISSN 0002-8282. doi:10.1257/aer.20130604.

Manuela Angelucci, Dean Karlan, and Jonathan Zinman. Microcredit impacts: Evidence from a randomized microcredit program placement experiment by Compartamos Banco. *American Economic Journal: Applied Economics*, 7(1):151–82, 2015. doi:10.1257/app.20130537.

Joshua D. Angrist and Jörn-Steffen Pischke. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30, 2010. doi:10.1257/jep.24.2.3.

F. J. Anscombe. Graphs in statistical analysis. *American Statistician*, 27(1):17–21, 1973. ISSN 0003-1305. doi:10.1080/00031305.1973.10478966.

Susan Athey and Guido W. Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32, 2017. doi:10.1257/jep.31.2.3.

Orazio Attanasio, Britta Augsburg, Ralph De Haas, Emla Fitzsimons, and Heike Harmgart. The impacts of microfinance: Evidence from joint-liability lending in Mongolia. *American Economic Journal: Applied Economics*, 7(1):90–122, 2015. doi:10.1257/app.20130489.

Britta Augsburg, Ralph De Haas, Heike Harmgart, and Costas Meghir. The impacts of microcredit: Evidence from Bosnia and Herzegovina. *American Economic Journal: Applied Economics*, 7 (1):183–203, 2015. doi:10.1257/app.20130272.

Abhijit Banerjee, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. The miracle of microfinance? evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, 7(1):22–53, 2015. doi:10.1257/app.20130533.

David A. Belsley, Edwin Kuh, and Roy E. Welsch. *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, 1980. doi:10.1002/0471725153.

Kirill Borusyak, Peter Hull, and Xavier Jaravel. Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1):181–213, 2022. ISSN 0034-6527. doi:10.1093/restud/rdab030.

George E. P. Box. Non-normality and tests on variances. *Biometrika*, 40(3/4):318–335, 1953. doi:10.2307/2333350.

Tamara Broderick, Ryan Giordano, and Rachael Meager. An automatic finite-sample robustness metric: Can dropping a little data make a big difference?, 2021.

Andrea Cerioli, Marco Riani, Anthony C. Atkinson, and Aldo Corbellini. The power of monitoring: How to make the most of a contaminated multivariate sample. *Statistical Methods & Applications*, 27(4):559–587, 2018. ISSN 1613-981X. doi:10.1007/s10260-017-0409-8.

Samprit Chatterjee and Ali S. Hadi. Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, 1(3):379–393, 1986. doi:10.1214/ss/1177013622.

Ralph Dennis Cook. Influential observations in linear regression. *Journal of the American Statistical Association*, 74(365):169–174, 1979. doi:10.2307/2286747.

Bruno Crépon, Florencia Devoto, Esther Duflo, and William Parienté. Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco. *American Economic Journal: Applied Economics*, 7(1):123–50, 2015. doi:10.1257/app.20130535.

Jesús Crespo Cuaresma, Stephan Klasen, and Konstantin M. Wacker. There is poverty convergence. *SSRN Electronic Journal*, 2016. doi:10.2139/ssrn.2718720.

Jesús Crespo Cuaresma, Stephan Klasen, and Konstantin M. Wacker. When do we see poverty convergence? *Oxford Bulletin of Economics and Statistics*, 2022. ISSN 0305-9049. doi:10.1111/obes.12492.

Francisco Cribari-Neto, Tatiene C. Souza, and Klaus L. P. Vasconcellos. Inference under heteroskedasticity and leveraged data. *Communications in Statistics - Theory and Methods*, 36(10): 1877–1888, 2007. doi:10.1080/03610920601126589.

Federico Droller. Migration, population composition and long run economic development: Evidence from settlements in the Pampas. *The Economic Journal*, 128(614):2321–2352, 2018. doi:10.1111/ecoj.12505.

Bradley Efron and Charles Stein. The Jackknife estimate of variance. *The Annals of Statistics*, 9 (3), 1981. doi:10.1214/aos/1176345462.

Bradley Efron and Robert J. Tibshirani. *An introduction to the Bootstrap*. CRC Press, 1994. doi:10.1201/9780429246593.

A. A. Fauvel. *Unpublished documents on the history of the Seychelles islands anterior to 1810*. Government Printing Office Mahe, Seychelles, 1909. URL https://www.loc.gov/item/unk83018617/.

Nicola Gennaioli and Ilia Rainer. The modern impact of precolonial centralization in Africa. *Journal of Economic Growth*, 12(3):185–234, 2007. ISSN 1573-7020. doi:10.1007/s10887-007-9017-z.

Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift. Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624, 2020. ISSN 0002-8282. doi:10.1257/aer.20181047.

Ali S. Hadi. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society: Series B* (*Methodological*), 54(3):761–771, 1992. doi:10.1111/j.2517-6161.1992.tb01449.x.

Ali S. Hadi and Jeffrey S. Simonoff. Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*, 88(424):1264–1272, 1993. ISSN 0162-1459. doi:10.1080/01621459.1993.10476407.

Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. *Robust statistics: The approach based on influence functions*, volume 196. John Wiley & Sons, 2005. doi:10.1002/9781118186435.

Lars Peter Hansen and Thomas J. Sargent. *Robustness*. Princeton University Press, Princeton, NJ, USA, 2011. ISBN 978-1-40082938-5. doi:10.1515/9781400829385.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. doi:10.1214/aoms/1177703732.

Peter J. Huber and Elvezio M. Ronchetti. *Robust statistics*. John Wiley & Sons, Ltd., Chichester, England, UK, January 2009. ISBN 978-0-47012990-6. doi:10.1002/9780470434697.

Rob J. Hyndman, Earo Wang, and Nikolay Laptev. Large-scale unusual time series detection. In *2015 IEEE International Conference on Data Mining Workshop* (*ICDMW*), pages 1616–1619. IEEE, 2015. doi:10.1109/ICDMW.2015.104.

Paul Johnson and Chris Papageorgiou. What remains of cross-country convergence? *Journal of Economic Literature*, 58(1):129–75, 2020. doi:10.1257/jel.20181207.

Dean Karlan and Jonathan Zinman. Microcredit in theory and practice: Using randomized credit scoring for impact evaluation. *Science*, 332(6035):1278–1284, 2011. doi:10.1126/science.1200138.

Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: An introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009. ISBN 978-0470316801. doi:10.1002/9780470316801.

Stephan Klasen and Mark Misselhorn. Determinants of the growth semi-elasticity of poverty reduction, 2008.

Edward E. Leamer. Let's take the con out of econometrics. *American Economic Review*, 73(1): 31–43, 1983. URL https://www.jstor.org/stable/1803924.

Edward E. Leamer. Sensitivity analyses would help. *American Economic Review*, 75(3):308–313, 1985. URL https://www.jstor.org/stable/1814801.

John R. Lewis, Steven N. MacEachern, and Yoonkyung Lee. Bayesian restricted likelihood methods: Conditioning on insufficient statistics in Bayesian regression. *Bayesian Analysis*, pages 1–38, 2021. doi:10.1214/21-BA1257.

Ricardo A. Maronna, R. Douglas Martin, Victor J. Yohai, and Matías Salibián-Barrera. *Robust statistics: Theory and methods (with R)*. John Wiley & Sons, 2019. doi:10.1002/9781119214656.

Stelios Michalopoulos and Elias Papaioannou. Pre-colonial ethnic institutions and contemporary African development. *Econometrica*, 81(1):113–152, 2013. ISSN 1468-0262. doi:10.3982/ECTA9613.

Nathan Nunn. The long-term effects of Africa's slave trades. *Quarterly Journal of Economics*, 123 (1):139–176, 2008. doi:10.1162/qjec.2008.123.1.139.

Nathan Nunn and Diego Puga. Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, 94(1):20–36, 2012. doi:10.1162/REST_a_00161.

Nathan Nunn and Leonard Wantchekon. The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7):3221–52, 2011. doi:10.1257/aer.101.7.3221.

Daniel Peña and Victor J Yohai. The detection of influential subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1): 145–156, 1995. doi:10.1111/j.2517-6161.1995.tb02020.x.

Garry D. A. Phillips. Recursions for the two-stage least-squares estimators. *Journal of Econometrics*, 6(1):65–77, 1977. doi:10.1016/0304-4076(77)90055-0.

Richard R. Picard and R. Dennis Cook. Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575–583, 1984. ISSN 0162-1459. doi:10.1080/01621459.1984.10478083.

Martin Ravallion. Why don't we see poverty convergence? *American Economic Review*, 102(1): 504–23, 2012. doi:10.1257/aer.102.1.504.

Marco Riani, Andrea Cerioli, Anthony C. Atkinson, and Domenico Perrotta. Monitoring robust regression. *Electronic Journal of Statistics*, 8(1):646–677, 2014. ISSN 1935-7524. doi:10.1214/14-EJS897.

Peter J. Rousseeuw and Annick M. Leroy. *Robust regression and outlier detection*. John Wiley & Sons, Ltd., Chichester, England, UK, 1987. ISBN 978-0-47185233-9. doi:10.1002/0471725382.

Matthew S. Shotwell and Elizabeth H. Slate. Bayesian outlier detection with Dirichlet process mixtures. *Bayesian Analysis*, 6(4):665–690, 2011. doi:10.1214/11-BA625.

Enrico Spolaore and Romain Wacziarg. How deep are the roots of economic development? *Journal of Economic Literature*, 51(2):325–69, 2013. ISSN 0022-0515. doi:10.1257/jel.51.2.325.

Mark F. J. Steel. Model averaging and its use in economics. *Journal of Economic Literature*, 58(3): 644–719, 2020. doi:10.1257/jel.20191385.

Stephen M. Stigler. The changing history of robustness. *American Statistician*, 64(4):277–281, 2010. ISSN 0003-1305. doi:10.1198/tast.2010.10159.

Marco Tabellini. Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *Review of Economic Studies*, 87(1):454–486, 2020. ISSN 0034-6527. doi:10.1093/restud/rdz027.

Alessandro Tarozzi, Jaikishan Desai, and Kristin Johnson. The impacts of microcredit: Evidence from Ethiopia. *American Economic Journal: Applied Economics*, 7(1):54–89, 2015. doi:10.1257/app.20130475.

Alwyn Young. Consistency without inference: Instrumental variables in practical application. *European Economic Review*, 147:104112, 2022. ISSN 0014-2921. doi:10.1016/j.euroecorev.2022.104112.

Appendix

Hidden in Plain Sight: Influential Sets in Linear Regression

This appendix features three sections. In Appendix A we provide replication material and interactive demonstrations of the algorithms proposed in the paper, in Appendix B, we provide additional technical details of the method, and in Appendix C we provide additional results for the empirical applications assessed in the paper, as well as for an example based on addressing poverty convergence across countries.

A   Software and reproducibility

All code used in this paper, and a preliminary R package that implements the proposed algorithms, are freely available at `https://github.com/nk027/influence`. In this section, we (1) provide an interactive illustration of influential sets in univariate settings for the browser, and (2) present ready-to-use code to install and load the package, and use the methods on simulated and real-world data from the main text.

Interactive demonstration

An interactive illustration that can be used in most browsers (desktop and mobile) is available at `https://n027.shinyapps.io/influence/`.

A1   Code

A script version of this code is available at `https://short.wu.ac.at/influence_demo`.

```r
# 0. Install and load the package -----
devtools::install_github("nk027/influence") # Install from GitHub
library("influence") # Load the package

# 1. Reproduce the toy example -----
set.seed(753) # Simulate some random (seeded) data ---
N <- 54 # We'll add 3 and 3 outliers to the end
x <- c(rnorm(N), rnorm(3, 6, 0.25), rnorm(3, 8, 0.25))
y <- c( # The outliers differ in effect
  x[seq(N)] * -0.5 + rnorm(N, 0, 1),
  x[seq(N + 1, N + 3)] * 0.1 + rnorm(3, 0, 0.1),
```

```r
  x[seq(N + 4, N + 6)] * 0.4 + rnorm(3, 0, 0.1)
)
mdl <- lm(y ~ x - 1) # Fit a linear model (w/o intercept) ---
plot(x, y); abline(mdl, col = "darkgray", lwd = 2)
# Find the influential sets ---
mdl_sens <- sens(mdl) # Turns the first coef negative by default
mdl_sens$influence$id[1:7] # Adaptive results
mdl_sens$initial$id[1:7] # Initial approximation

# 2. Reproduce the ruggedness application -----
data <- read.csv("https://short.wu.ac.at/rugged_data") # Read from GitHub
data$diamonds <- with(data, gemstones / (land_area / 100))
data <- data[!is.na(data$rgdppc_2000), ] # Remove NAs beforehand
# Fit the baseline model ---
mdl <- lm(log(rgdppc_2000) ~ rugged * cont_africa +
  diamonds * cont_africa + soil * cont_africa +
  tropical * cont_africa + dist_coast * cont_africa, data = data)
summary(mdl) # Summary and non-robust SE
# lmtest::coeftest(mdl, vcov = sandwich::vcovHC(mdl, "HC1")) # Robust SE
# Assess influential sets ---
mdl_sens <- sens(mdl, # Target t of rugged:cont_africa at position 8
  lambda = set_lambda("tstat", pos = 8, sign = sign(coef(mdl)[8])),
  options = set_options("all"), # No approximations
  cluster = seq_along(mdl$fitted) # For HC1 SE
)
plot(mdl_sens, threshold = qnorm(.975)) # Plot the path
summary(mdl_sens) # Get the summary of set sizes
data[mdl_sens$influence$id[seq(5)], ] # Check out the top 5 set
```

B Additional technical information

In this section, we provide additional technical details on the approximation of influential sets and their influence. First, we illustrate the underestimation caused by aggregating individual, full-sample influences, as performed in Algorithm 0. Then, we discuss approximations to the influence of sets of observations, using the approach of Broderick et al. (2021) as an example. Next, we conduct a simulation exercise to compare the three considered algorithms. Finally, we investigate technical aspects behind the sensitivity of OLS and 2SLS estimators in a simulation exercise.

B1 Aggregating individual influences

Algorithm 0 approximates the influence of a set of observations by aggregating individual, full-sample influences of its members. When assessing the influence on coefficient estimates, this method suffers from a downward bias. With every removal, the leverage and hence the influence of observations increases and differences in leverage are exacerbated. The other driver of influence, residuals, decreases on average, complicating the analysis. We illustrate the underestimation with a mean-only example.

Consider the model $\mathbf{y} = \mathbf{1}\theta + \varepsilon$, where we are interested in the influence on the estimate of θ, which we are (without loss of generality) trying to decrease ($\Delta(\mathcal{S}) = \hat{\theta}_{(\varnothing)} - \hat{\theta}_{(\mathcal{S})}$). We will show that the influence of a set of two influential observations y_1 and y_2 (i.e., they satisfy $y_1 \geq y_2 > \sum_i y_i/N$) is greater than the sum of influences of its members. We need to show $\delta_1 + \delta_2 < \Delta(\{1,2\})$, which is equivalent to

$$\hat{\theta} - \hat{\theta}_{(1)} + \hat{\theta} - \hat{\theta}_{(2)} < \hat{\theta} - \hat{\theta}_{\{(1,2)\}},$$

$$\frac{\sum_i y_i}{N} - \frac{\sum_{i\neq 1} y_i + \sum_{i\neq 2} y_i}{N-1} + \frac{\sum_{i\neq 1,2} y_i}{N-2} < 0.$$

Since $\sum_{i\neq 1} y_i + \sum_{i\neq 2} y_i = \sum_i y_i + \sum_{i\neq 1,2} y_i$, we can redistribute the second term for

$$\left(\frac{1}{N} - \frac{1}{N-1}\right)\sum_i y_i + \left(\frac{1}{N-2} - \frac{1}{N-1}\right)\sum_{i\neq 1,2} y_i < 0.$$

By assumption, we know that $\sum_{i\neq 1,2} y_i + \frac{2}{N-2}\sum_{i\neq 1,2} y_i < \sum_i y_i$, which implies that

$$\left(\frac{1}{N} - \frac{1}{N-1}\right)\sum_i y_i + \left(\frac{1}{N-2} - \frac{1}{N-1}\right)\sum_{i\neq 1,2} y_i <$$

$$\left(\frac{1}{N} - \frac{1}{N-1}\right)\left(\sum_{i\neq 1,2} y_i + \frac{2}{N-2}\sum_{i\neq 1,2} y_i\right) + \left(\frac{1}{N-2} - \frac{1}{N-1}\right)\sum_{i\neq 1,2} y_i = 0,$$

where the second term cancels out, completing the proof. □

If we use one or many regressors instead, the leverage of individual observations will vary, and we can no longer directly determine the impact of removals on the error. However, the overall problem is similar — any removal increases the leverage (and, hence, the influence) of non-removed observations. This effect is particularly strong when a high-leverage (and thus, influential) observation is removed.

B2   Assessing influence approximations

Estimates of the influence of sets of observations that ignore or approximate certain elements required to compute the exact value enable computation in otherwise unfeasible settings. However, such computational shortcuts can become problematic if the approximated elements are important determinants of the influence. One useful approximation is the Approximate Maximum Influence Perturbation (AMIP, by Broderick et al., 2021), which we investigate in our context.[10] We illustrate the limitations of this approximation using a simulated example, and then discuss performance in the context of seven studies of microcredits, which is featured in their contribution.[11]
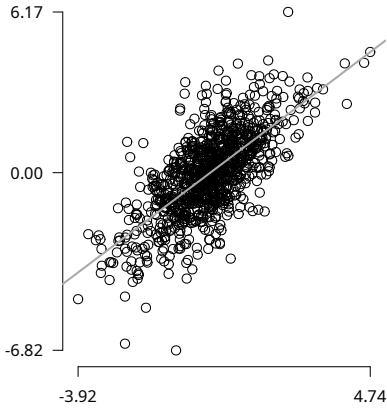

Figure S1: Simulated data and least squares regression line. The covariate (on the horizontal axis) and the innovations are drawn from a $t(8)$ distribution to allow for moderate leverage and residuals.

Consider a univariate regression model based on $N = 1,000$ observations, with a slope parameter $\beta = 1$ and a zero intercept, where innovations and observations of the covariate are drawn from $t(8)$ distributions. This setup allows for moderately high leverage and residuals, without inducing sensitivity to influential sets. The simulated dependent variable is visualized in Figure S1. First, we illustrate the behavior of AMIP when approximating influential sets and their influence with regard to the coefficient estimate of β. Influence is given by

$$\delta_i^{\text{AMIP}} = \left(\mathbf{x}'\mathbf{x}\right)^{-1} x_i' e_i \quad \approx \quad \text{DFBETA}_i = \frac{\left(\mathbf{x}'\mathbf{x}\right)^{-1} x_i' e_i}{1 - h_i},$$

---

[10]Our analysis is based on the results in Broderick et al. (2021) and their implementation (kindly made available at `https://github.com/rgiordan/zaminfluence`) as of commit 2d29fbb from 4/20/2022.

[11]We also consider AMIP in the simulation exercise in Section B3 and provide additional details.

where $\mathbf{x}$ is a vector containing the covariate and e is a vector containing the residual. The difference between δ_i^{AMIP} and DFBETA$_i$ is given by the suppression of the leverage component of the measure, h_i.

In Figure S2, we visualize the errors of approximation (when compared to the true influence). In the left panel, we relate the AMIP error to the leverage of the dataset. Since leverage and residuals are the only two factors driving influence in this simple case, the use of AMIP leads to a downward bias (of $(N-1)^{-1}$ percent) in the estimation of the influence. In the right panel, we relate the error to the true influence. As can be seen, the magnitude of the error is particularly pronounced for highly influential observations.
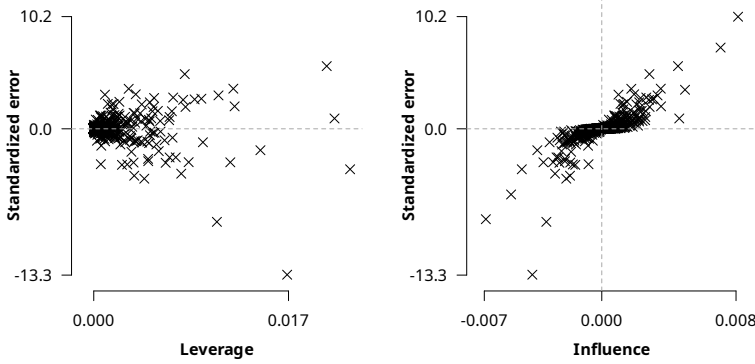

Figure S2: Errors from using AMIP to estimate influence on coefficients (compared to the exact value), plotted against the exact leverage (left) and the exact influence (right) of observations.

Next, we investigate the influence on the estimate of the standard error of the coefficient estimate. To showcase their relationship, we regress the exact influence on standard errors on its AMIP estimate. We find an R^2 of 0.86, and a coefficient value of 1.005 ($t = 78.8$), indicating good accuracy on average. However, the quality of the approximation deteriorates considerably with increasing influence. Figure S3 visualizes the relationship between the (standardized) regression residuals and the fitted values, as well as leverage.
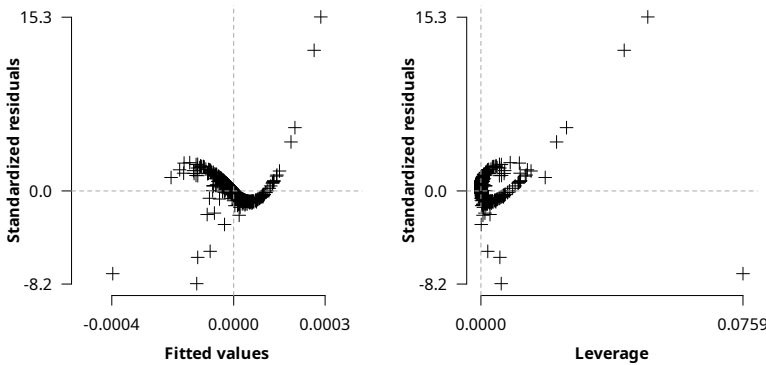

Figure S3: Diagnostic plots for the regression of exact observation influence on standard errors onto their AMIP estimates. Residuals are plotted against fitted values (left) and the leverage of the influence (on the right).

Finally, we consider the influence on the significance of the parameter estimate. Broderick et al. (2021) do not explicitly track influence on t values, but rather compute

the influence on significance for a given t value by accumulating

$$\delta_i^{\text{AMIP}} = \beta_{(i)}^{\text{AMIP}} + t\text{SE}_{(i)}^{\text{AMIP}}, \tag{S1}$$

where $\beta_{(i)}^{\text{AMIP}}$ and $\text{SE}_{(i)}^{\text{AMIP}}$ are the AMIP estimates of the influence of observation i on the estimate of β and its standard error. A minimal influential set is found by choosing a direction for Δ (i.e. significantly positive or negative), and setting the target value that needs to be exceeded to one of $\beta \pm t \times \text{SE}$. We investigate this approximation below using an empirical application.

Empirical application to microcredits

Many recent large-scale experimental studies evaluate the efficacy of microcredit as a tool for alleviating poverty and facilitating economic development. Seven of these studies are analyzed by Broderick et al. (2021), who assess the sensitivity of the average treatment effect to removing observations. These are randomized control trials in Bosnia and Herzegovina (Augsburg et al., 2015), Mongolia (Attanasio et al., 2015), Ethiopia (Tarozzi et al., 2015), Mexico (Angelucci et al., 2015), Morocco (Crépon et al., 2015), the Philippines (Karlan and Zinman, 2011), and India (Banerjee et al., 2015). The underlying model is a simple treatment effect model with a single (randomized) treatment dummy.

Table S1: Sensitivity of the average treatment effect of microcredits

		BIH	MON	ETH	MEX	MOR	PHI	IND
Estimate	β	37.53	-0.34	7.29	-4.55	17.54	66.56	16.72
	t	(1.90)	(-1.53)	(0.92)	(-0.77)	(1.54)	(0.85)	(1.41)
Sign-switch	A2	13	15	1	1	11	9	6
	A0	14	16	1	1	11	9	6
	B0	14	16	1	1	11	9	6
Significance	A2	35	34	10	9	29	38	28
	A0	68	58	387	41	42	89	76
	B0	40	38	66	15	30	58	32
Observations		1,195	961	3,113	16,560	5,498	1,113	6,863

The reported values are the sizes of influential sets that are needed to induce a sign-switch of the average treatment effect, and have this sign-flip become significant (at the 5% level). We use Algorithms 2 (labelled 'A2') and 0 using exact influences ('A0') and AMIP estimates ('B0'). For the sign-switch, we explicitly target the influence on β, and not on the t values.

In Table S1, we present the full-sample estimates and the sizes of minimal influential sets that induce a sign-switch, and a significant sign-switch. We compare three

approaches, based on (1) Algorithm 2, and Algorithm 0, using (2) exact influences ('A0') and (3) AMIP estimates ('B0', reproducing Broderick et al., 2021). The results for sign-switches are similar. By contrast, the sizes needed for a significant sign-switch differ strongly. When particularly influential observations are present, only Algorithm 2 properly accounts for the effects of subsequent removals. Notably, the AMIP estimates for inducing significance (following Equation S1) are considerably lower than one would expect.[12]
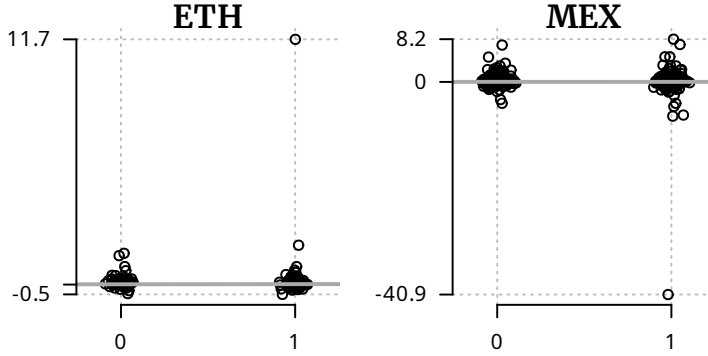

Figure S4: Household income in thousands (vertical axis) and the treatment dummy for the studies on Ethiopia and Mexico. The regression line is indicated in gray.
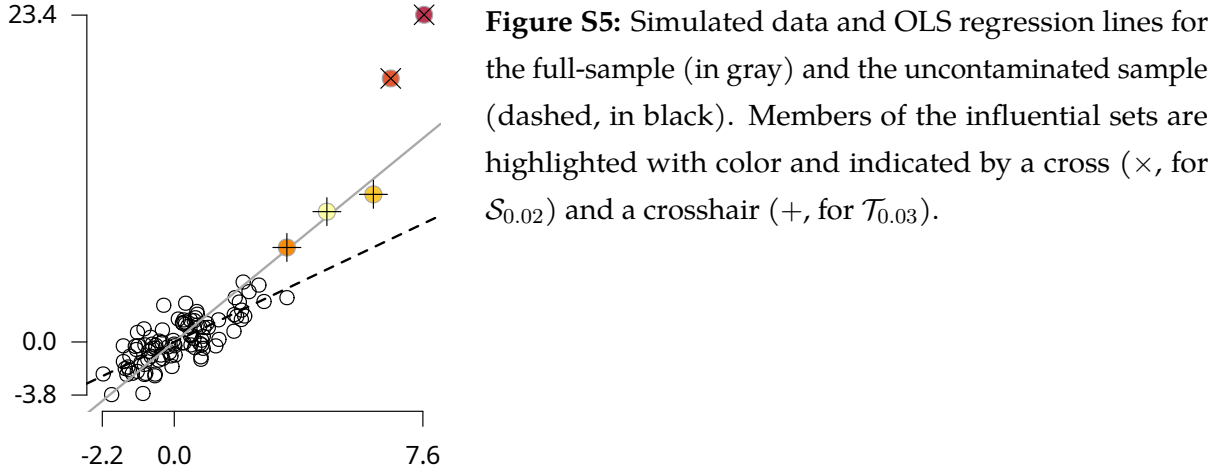
Thanks to randomization, the regression model underlying these results is remarkably simple, and Cook's distance (or other standard checks) already indicate sensitivity issues. The cases of Ethiopia and Mexico are particularly striking, (see Figure S4). Since leverage plays a limited role in this regression setting, it appears as a suitable application for Algorithm 1.

B3 Approximations in the presence of influential sets

Approaches for identifying influential sets that are based on approximations may suffer precisely in the presence of such influential sets. To showcase this issue, we consider the univariate regression from Section B2 (with $\beta = 1$, and innovations and covariates drawn from $t(8)$ distributions) with $N = 100$. We contaminate this setup with two influential sets. For the first set, $\mathcal{S}_{0.02}$ we use a coefficient value of three and increase covariate values by seven. For the second one, $\mathcal{T}_{0.03}$, the coefficient value is changed to two and covariate values are increased by five. One realization of this process is visualized in Figure S5.

We simulate 1,000 realizations of the process described above, and use four approaches to assess maximally influential sets (that lower coefficient estimates) of sizes

---

[12]This is peculiar, considering the downward biases of (1) accumulating influences, (2) the $\beta_{(i)}^{\mathrm{AMIP}}$ estimate in general, and (3) the estimates $\beta_{(i)}^{\mathrm{AMIP}}$ and $\mathrm{SE}_{(i)}^{\mathrm{AMIP}}$ for influential observations.

Figure S5: Simulated data and OLS regression lines for the full-sample (in gray) and the uncontaminated sample (dashed, in black). Members of the influential sets are highlighted with color and indicated by a cross ($\times$, for $\mathcal{S}_{0.02}$) and a crosshair ($+$, for $\mathcal{T}_{0.03}$).

one to ten.[13] These are Algorithms 2 (labelled 'A2') and 1 ('A1'), as well as Algorithm 0, with exact influences ('A0') and the AMIP estimate ('B0'). The results are presented in Figure S6. Both variants of Algorithm 0 perform relatively poorly, and do not recognize the influence of the first influential set. Algorithm 1 performs better, but does not reliably account for the influence of the second set. Estimates are considerably spread out, and the target value (after five removals) of unity in the slope coefficient is missed on average. The impact of removals levels off afterwards. By contrast, Algorithm 2 reliably finds both sets, and continues to identify impactful observations after their removal.
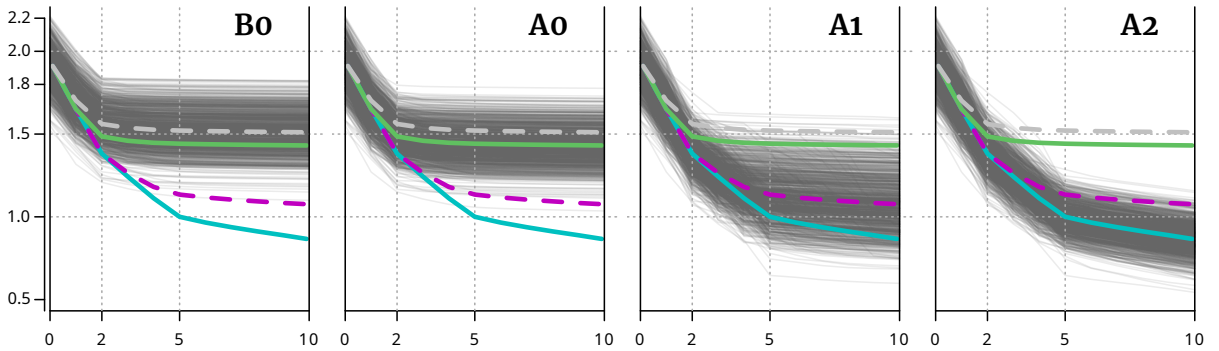

Figure S6: Transparent lines indicate individual runs, thick lines the average results of (from top to bottom) approach 'B0' (gray, dashed), 'A0' (green, solid), 'A1' (purple, dashed), and 'A2' (teal, solid). The vertical axis indicates estimates, the horizontal one the number of removals.

---

[13]The average estimates are $\beta_{LS} = 1.91$, $\hat{\beta}_{(\mathcal{S})} = 1.38$, and $\hat{\beta}_{(\{\mathcal{S},\mathcal{T}\})} = 1.00$. Interestingly, robust M-estimation slightly undershoots the influence of the first set, at $\beta_M = 1.43$.

B4　Simulations and method sensitivity for OLS and 2SLS

Sensitivity to influential sets depends on the measure of influence and the estimator used. We illustrate the role of the particular estimator by comparing influential sets for OLS and 2SLS coefficient estimates. Again, we use the univariate regression from Section B2 with $N = 100$,[14] and add a repeated layer of this setup to construct the instrumented variable. This gives us the following setup for 2SLS estimation

$$\mathbf{x} = \mathbf{z} + \mathbf{u}, \text{ with } \mathbf{u} \sim t(8) \text{ and } \mathbf{z} \sim t(8),$$

$$\mathbf{y} = \mathbf{x} + \mathbf{e}, \text{ with } \mathbf{e} \sim t(8).$$

We use Algorithm 2 to find and assess the influence of maximally influential sets from size one to 95 and replicate the exercise 1,000 times.

The results of our simulation exercise are visualized in Figure S7. On the vertical axes, we depict the OLS (top) and 2SLS (bottom) estimates after removing influential sets of increasing size (horizontal axis). OLS estimates remain stable until around twenty observations are left in the sample. After this point, we observe a limited divergence of estimates. The mean and median of the estimates stay close to each other throughout, indicating no serious numerical problems. This is not the case for the 2SLS estimates, where estimates start to diverge after about twenty removals. At 24 removals, we have the first drop-out due to a lack of numerical stability, which we indicate with a green cross on top. Before this, we can see a short spike in the mean estimate, which then drifts off again. Overall, about three quarters of the simulated runs for 2SLS estimation terminate prematurely due to a lack of numerical stability. At around 50 removals, more than half of the remaining 2SLS estimates are pathological (see median estimate), with vast fluctuations afterwards.

These results indicate that even with heavy-tailed errors and leverage, the OLS estimates of an otherwise correctly specified linear regression model do not appear to be particularly sensitive to influential sets. However, the 2SLS estimates appear to be susceptible to influential sets and run into numerical stability problems, even with this simple setup.

---

[14]That is, with $\beta = 1$, and innovations and covariates drawn from a $t(8)$ distribution to allow for moderately high errors and leverage. See Figure S5 for one realization.
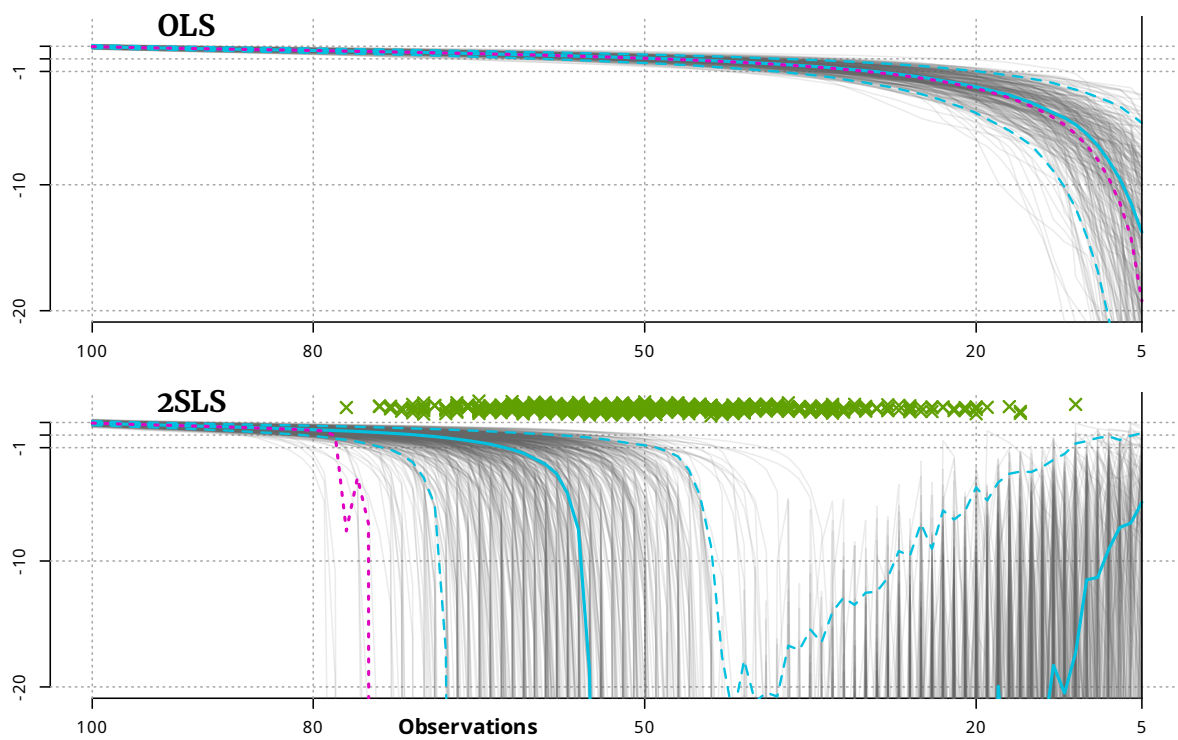
Figure S7: Transparent lines indicate (250 samples of) individual runs, thick lines indicate the median (solid, blue), the 95% and 5% quantile (dashed), and the average (dotted, pink) of the estimate. Crosses at the top of the 2SLS panel indicate drop-outs due to pathological numerical stability (within machine precision). The vertical axes indicate estimates, the horizontal ones the number of observations left.

C Additional results and applications

In this section, we provide supporting material for the applications discussed in the paper, and present one additional application to global cross-country poverty convergence.

C1 The origins of mistrust

Table S2: Distribution of the top 600 most influential observations on trust in relatives

	Benin	Nigeria	Ghana	Other
Top 100	54	12	19	15
Top 101–200	85	6	5	4
Top 201–300	80	12	7	1
Top 301–400	52	27	14	7
Top 401-500	3	67	6	24
Top 501–600	3	69	15	13
Top 600	277	193	66	64

Summary of the origin countries of the 600 most influential observations affecting the parameter estimate associated with the slave trades as a determinant of trust of relatives (see Column 1 of Table 2).

C2 The effects of the Tsetse fly

Table S3: The robust effects of the Tsetse fly

	Animals	Intensive	Plow	Female	Density	Slavery	Centralized
M-estimate	-0.259	-0.102	–	0.227	-0.761	–	-0.080
	(-5.53)	(-3.29)	–	(3.71)	(-3.00)	–	(-2.12)
S-estimate	–	–	–	–	-0.761	–	–
	–	–	–	–	(-4.63)	–	–
N	484	485	484	315	398	446	467

Robust M- and S-estimates for the effects of the Tsetse fly (see Table 3). Coefficient estimates are reported with t values based on clustered (for M-estimation) and classical (for S-estimation) standard errors in parentheses.

C3   The long-term migration impacts on development
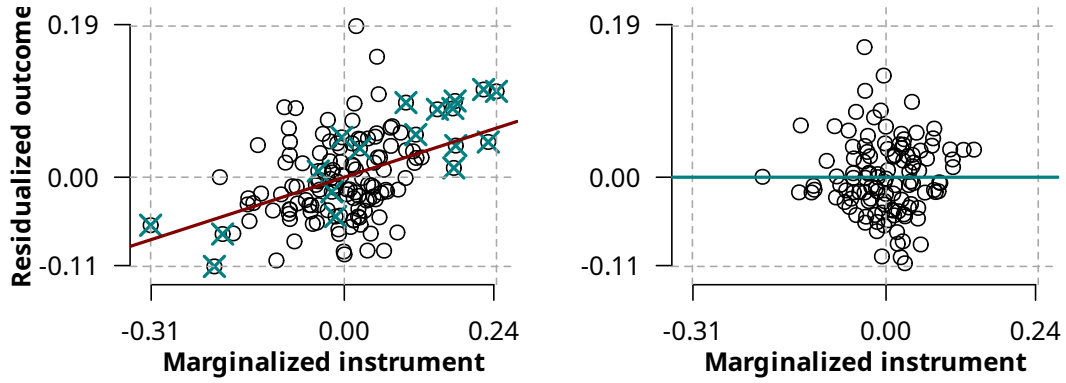

Figure S8: Residualized data (with other covariates marginalized out) and regression lines of the share of European migration against the shift-share instrument (see Table 4), before (left) and after (right) removing the 18 observations that induce numerical instability of 2SLS estimates.

C4   Poverty convergence and influential sets

Ravallion (2012) examines the convergence of poverty rates in a sample of 89 countries using the following linear regression model

$$T_i^{-1}(\ln H_{it} - \ln H_{it-1}) = \alpha + \beta \ln H_{it-1} + \varepsilon_{it}, \tag{S2}$$

where H_{it} denotes the poverty headcount ratio in country i and time period t, and T_i is the length of its observation period in years. In his study, Ravallion (2012) does not detect poverty convergence, instead obtaining a positive, statistically insignificant estimate of β, that we reproduce in Column 1 of Table S4.

This finding is likely due to idiosyncratic experiences in the former Eastern bloc. As Crespo Cuaresma et al. (2022, 2016) point out, these countries exhibit low initial poverty headcount ratios, implying that small absolute changes translate into large growth rates. This is due to the log-transformation applied, and makes observations of these countries particularly influential on OLS estimates. Crespo Cuaresma et al. (2016) consider two alternative specifications that yield empirical evidence of poverty convergence in the original dataset. These are (1) an extension of Equation S2 that controls for the experience of Eastern European countries, and (2) a model that is based on a semi-elastic relationship between poverty reduction and growth (see Klasen and Misselhorn, 2008), effectively dropping the log-transformation from Equation S2. See Table S4, Columns 2 and 3 for estimates.

Table S4: Sensitivity of poverty convergence evidence to influential sets

	Baseline	Eastern Europe	Alternative
Convergence[†]	0.006	-0.020	-0.019
	(0.84)	(-2.88)	(-9.68)
Convergence, Eastern Europe	–	0.044	–
	–	(2.12)	–
Thresholds[†]	–[1]{4}	3[10]{24}	26[32]{42}
N	89	89	124
R^2	0.375	0.375	0.607

The row labeled 'Thresholds' reports the sizes of influential sets that induce a loss of significance (at the 5% level), [a sign flip], and {a significant sign flip} of the coefficient that captures the poverty convergence effect (using Algorithm 2). The column labelled 'Baseline' reproduces the results of Ravallion (2012), the one labelled 'Eastern Europe' adds an interaction term for Eastern European countries. The column labelled 'Alternative' uses a different specification (proposed by Crespo Cuaresma et al., 2016) and an updated dataset. Coefficients are reported with t values in parentheses.

We assess the sensitivity of poverty convergence to influential sets, starting with the model in Equation S2. Figure S9 presents the data and regression line of Ravallion (2012), as well as the minimal influential set needed to attain significant poverty convergence (see 'Thresholds' in Table S4), as identified by Algorithm 2. The four members of this set are Belarus, Latvia, Ukraine, and Poland.[15] We investigate the two additional specifications next. When accounting for the experience of Eastern Europe, we find a significance threshold of three observations. For the alternative specification, we source an updated dataset from PovCalNet,[16] and find significant poverty convergence. Algorithm 2 indicates thresholds at 26 (insignificance, see Figure S10), 32 (sign-flip), and 42 (significant sign-flip) out of 124 observations. Algorithm 1 only indicate a loss of significance after 56 removals (see Figure S10).

---

[15]Subsequent removals would be the Russian Federation, Lithuania, Estonia, and Macedonia. Notably, Algorithm 0 indicates a threshold of sixteen removals, suggesting a relatively low degree of sensitivity of the result.

[16]We use poverty headcounts (at $2 a day) from household surveys, only include countries that are observed continuously for at least ten years, and use consumption- over income-based data when available.
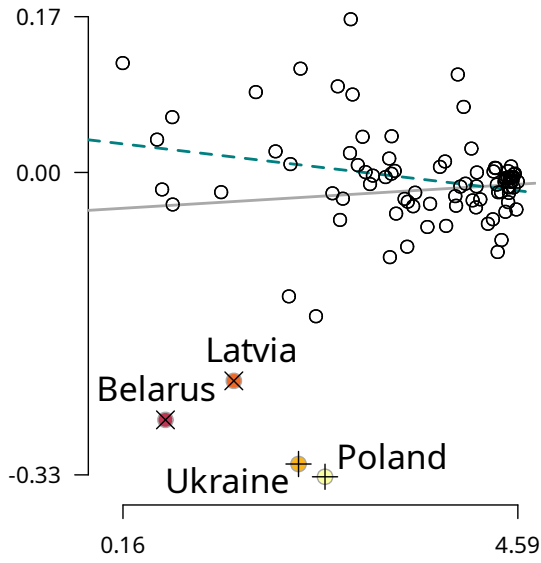
Figure S9: Data and regression line for Raval-lion (2012) before (solid line) and after (dashed line) removing the influential set $\hat{\mathcal{S}}_4^*$ (colored and marked with crosses, then crosshairs). The vertical axis holds annualized log differences of poverty headcount ratios; the horizontal axis the logarithm of the initial poverty headcount ratio.
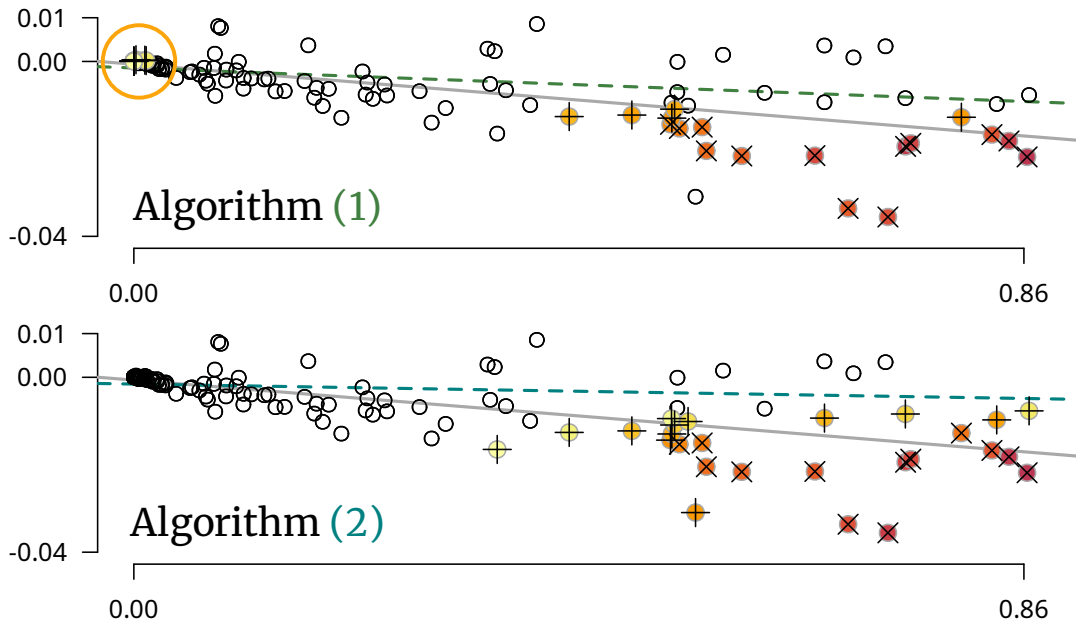

Figure S10: Data and regression line for the alternative specification before (solid line) and after (dashed line) accounting for an influential set of size 26 (colored and marked with crosses ×, then crosshairs +), using Algorithm 1 (top) and 2 (bottom).