Testing for Coefficient Distortion due to Outliers with an Application to the Economic Impacts of Climate Change

Xiyu Jiao[1,4], Felix Pretis[2,4]*, and Moritz Schwarz[3,4]

[1]Department of Economics, University of Oxford
[2]Department of Economics, University of Victoria
[3]School of Geography and the Environment, University of Oxford
[4]Nuffield College, University of Oxford

Abstract

Outlying observations can bias regression estimates, requiring the use of robust estimators. Comparing robust estimates to those obtained using OLS is a common robustness check, however, such comparisons have been predominantly informal due to the lack of available tests. Here we introduce a formal test for coefficient distortion due to outliers in regression models. Our proposed test is based on the difference between OLS and robust estimates obtained using a class of Huber-skip M-type estimators (such as Impulse Indicator Saturation or Robustified Least Squares). We show that our test has an asymptotic chi-squared distribution, and is valid for cross-sectional as well as panel and time series models. The argument requires establishing asymptotics of the corresponding Huber-skip M-estimators using an empirical process CLT recently developed by Berenguer-Rico et al. (2019). To improve finite sample performance and to alleviate concerns on distributional assumptions, we further introduce and explore three bootstrap testing schemes. Simulation studies lend support to the proposed testing methods and theory. Finally, we apply our outlier distortion test to estimates of the macro-economic impacts of climate change allowing for adaptation. We find evidence of income-driven adaptation, however, OLS estimates are significantly different to those obtained using a robust estimator. Controlling for outliers, the estimated impacts of climate on economic growth are dampened at all income levels compared to OLS estimates.

JEL Classification: C12, C52.
Keywords: outlier robustness checks, misspecification, outlier detection, robust estimation, iterated 1-step Huber-skip M-estimator, indicator saturation.

*Corresponding author: fpretis@uvic.ca.

1 Introduction

A common concern in empirical modelling centres around whether estimated regression coefficients are affected by a small set of outlying observations. Comparing OLS estimates to those obtained using a robust estimator is a common robustness check in the applied literature. However, such comparisons are mostly done heuristically, due to a lack of formal tests. Here we propose a formal test for outlier distortion assessing whether robust estimates obtained using popular Huber-skip estimators are different to those derived using OLS. We apply our test to estimates of the macro-economic impacts of climate change under adaptation.

Distorted estimates due to outlying observations are of particular concern in econometric research on the economic impacts of climate change. The standard approach in the empirical macro-economic climate-impact literature is to estimate country or region-level panel models, modelling GDP per capita growth as a function of population-weighted climate observations, see for example Dell et al. (2012); Burke et al. (2015); Pretis, Schwarz, et al. (2018); Kalkuhl & Wenz (2020); Newell et al. (2021). Control variables in these impact models are conventionally limited to unit and time fixed effects, together with unit-specific non-linear time trends. With such limited controls and the wide range of determinants of economic growth, there is the potential for many un-modelled idiosyncratic shocks to bias the estimated coefficients on climate variables. Existing estimates in the climate impact literature have relied on heuristic comparisons of OLS and outlier-robust estimates due to a lack of formal tests. For instance, comparing robust estimates to their main OLS specification, Dell et al. (2012) write in their seminal paper modelling GDP growth as a function of temperatures: "*When we use median regressions to reduce the impact of outliers, the estimated impact [of temperatures] for poor countries becomes slightly larger and substantially more statistically significant*". However, it is unclear whether 'slightly larger' implies that the estimates are statistically different from those obtained using OLS. Similarly, Pretis, Schwarz, et al. (2018) estimate an empirical impact model robust to outliers using Impulse Indicator Saturation (IIS - see Hendry et al. 2008), finding that "*controlling for these outlying observations reduces the slope of the estimated temperature curve [...] (suggesting reduced impact of temperatures on countries with low average annual temperatures)*". Nevertheless, this comparison also remains informal as no test was available to compare OLS to robust IIS estimates. Such potentially distorted coefficients may result in biases in the projections of the economic impacts of climate change and subsequently distort cost-benefit analyses of climate policy.

Robust estimators as in the above two cases pose a robustness-efficiency trade-off. When outliers are present, we should rely on robust estimators. When no outliers are present, OLS is more efficient. Unfortunately in practise the presence of outliers and any resulting distortion in coefficients is unknown. To assess whether the gain in robustness offsets the loss of efficiency when using robust estimators, we propose an outlier distortion test. Specifically, we introduce a test for outlier distortion comparing the difference between OLS and robust estimates derived from Huber-skip estimators. The robust procedure underlying Huber-skip estimators is to use least squares estimators on "clean" data obtained by removing observations with extreme residuals from an initial least squares regression fit. This two step procedure originates from the robust statistics literature, where it has been referred to as the Trimmed Least Squares (Ruppert & Carroll 1980); the Data-analytic Strategy (Welsh & Ronchetti 2002); and the 1-step

Huber-skip M-estimator (Johansen & Nielsen 2009). This procedure can start with any initial estimators, including the robust ones. Two special examples are Robustified Least Squares (RLS) and Impulse Indicator Saturation (IIS), which respectively use the full sample and split sample least squares as the initial estimators. Huber-skip type estimators have been widely applied in empirical studies so as to conduct the outlier robustness checks, with RLS being used to estimate the social returns to equipment investment (De Long & Summers 1991, 1994, Auerbach et al. 1994), the institutional impact on economic growth (Acemoglu et al. 2001, 2012, Albouy 2012, Acemoglu et al. 2019); with IIS being used to model wages (Castle & Hendry 2009), food expenditure and demand (Hendry & Mizon 2011), money demand (Dreger & Wolters 2014), housing markets (Anundsen 2015), unemployment (Nymoen & Sparrman 2015), exchange rates (Stillwagon 2016), debt forecasts (Ericsson 2017); with climate applications ranging from assessing climate model performances (Pretis et al. 2015) to detecting volcanic eruptions in temperature reconstructions (Schneider et al. 2017) to improving normalized hurricane damages (Martinez 2020). However, all of these studies have relied on heuristic comparisons of OLS and outlier robust estimators (RLS/IIS) and have not been able to formally assess whether outliers distort the coefficients on the variables of interest, since no statistical test was available to compare OLS to RLS/IIS.

The main contribution of this paper is to construct such a formal test of conducting outlier robustness checks and analysing whether or not conclusions in empirical studies are driven by a tiny set of outliers. When no outliers are present, we expect the difference between OLS and robust estimators to be small. In turn, when outliers are present and distort regression estimates, the difference between the two estimators will be large (see stylized example in Figure 1.1[1]). We show that our test, based on the difference between OLS and RLS/IIS similar to the Durbin (1954)-Hausman (1978)-Wu (1973) test statistics, has an asymptotic χ^2 distribution by analytically deriving the joint limiting distribution of the two estimators in cross-sectional and time series regressions, where regressors can be either stationary, deterministic trending, or unit root processes. It thus enables us to evaluate whether the gain in robustness is more valuable than the corresponding loss in efficiency when using RLS/IIS and to test whether the parameter of interest changes its value significantly before and after removing outliers, allowing formal outlier sensitivity analysis. Our simulations show that the test has a size close to nominal levels once sample sizes are moderately large ($n \geq 200$) and exhibits high power under a range of alternatives.

To establish the proposed outlier distortion test, we need to study the asymptotic behaviour of the RLS/IIS and two step robust procedure in general. Using the empirical process CLT recently developed by Berenguer-Rico et al. (2019), we build up an asymptotic theory of this class of robust estimators with an argument similar to Johansen & Nielsen (2009, 2013, 2019) which study M-estimators. The theory suggests two improvements to the naive two step procedure. First, the ordinary variance estimator needs to be bias-corrected due to the fact that some observations have certain chances to be wrongly classified as outliers and removed under the null of no outliers. Second, to further gain robustness the two step procedure can be iterated until converging to a fixed point which is shown to have the same first order asymptotics as the Huber-skip

[1] Also see the empirical study by Hendry & Mizon (2011) which showed extreme distortion of regression estimates from not modelling outliers when investigating the effect of relative prices on food expenditure, as the sign of the parameter changes from positive to negative after using the robust Huber-skip IIS estimator.

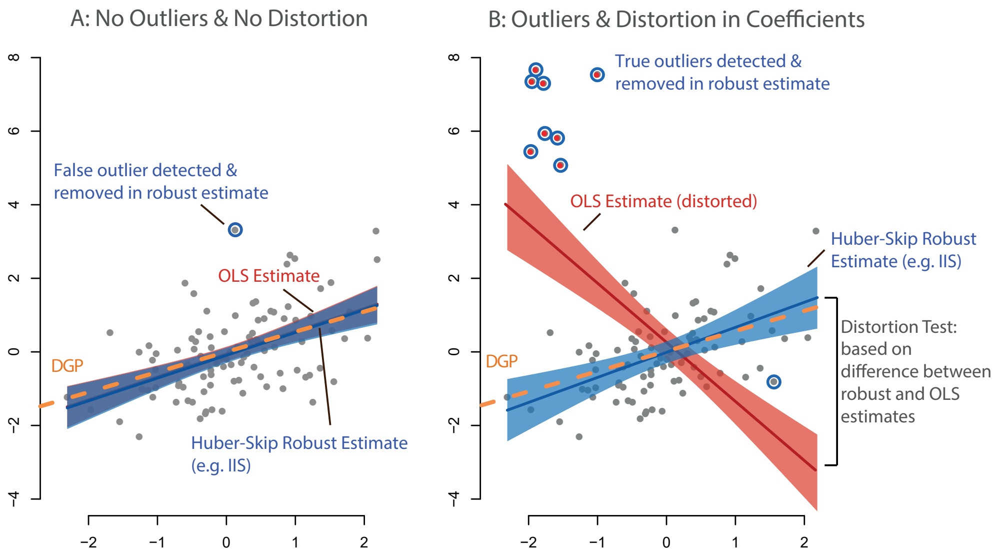

Figure 1.1: Stylized example of outlier distortion using artificial data. Left panel A shows OLS (red) and robust Huber-skip estimates (blue) when there are no distorting outliers. The DGP is shown as orange-dashed. Right panel B shows 8 outliers distorting the OLS estimates, where the difference between the OLS and robust coefficient estimates is used to construct our proposed distortion test. When no outliers are present, OLS is preferred. When outliers are present and distorting coefficient estimates, the robust estimator is preferred.

M-estimator. This paper then establishes tightness, stochastic expansion, and weak convergence of the robust estimators produced by the improved iterated procedure in cross-sectional or time series regressions with stationary, deterministically or stochastically trending regressors. Implementation of RLS/IIS requires selecting a cut-off value, the tuning parameter, to detect outlying observations based on their residuals. The cut-off value as robustification parameter determines the trade-off between robustness and efficiency for the robust estimator RLS/IIS. Jiao & Pretis (2020) provides a guidance on choosing this robustification parameter based on an asymptotic study of the false outlier detection rate. The theory derived in this paper holds uniformly for the robustification parameter.

To improve finite sample performance and to alleviate concerns on distributional assumptions on errors, we further introduce and explore three bootstrap testing schemes of our distortion test. First, we non-parametrically re-sample the raw data, which is conservative under the null hypothesis of no outliers and lacks power under alternatives. Second, we re-sample from the outlier-removed clean data. Here we face concerns about size as the cleaned data may appear rather different than the raw data. This is due to the fact that clean data was obtained by targeting a reference distribution. Therefore, we propose a third bootstrap scheme, in which we assess outlier distortion by re-sampling from the clean data but classify outliers at a looser cut-off threshold for bootstrapping samples. The idea being that outliers under a lower cut-off in the clean sample, mimic the relationship of the clean sample to the original raw data. This approach captures some of the characteristics of the data that are not forced to conform to the reference distribution. As expected, all schemes perform reasonably well under the null, however, re-sampling the raw data itself has close to zero power. The re-scaled clean data bootstrap performs best overall with good size and power properties in finite samples, even when the reference distribution is incorrect.

The literature on outlier robustness has not commonly focused on testing presence of distortion. Two papers, however, are notable exceptions and closely related to our approach. First, Kaji (2018) also asymptotically studies the outlier removal estimator RLS as well as other robust estimators, such as the winsorized estimator, all of which can be represented as L-statistics (integrals of transformations of empirical quantile functions with respect to corresponding random sample selection measures). To establish their asymptotic theory, he then develops a new empirical process theory and a functional delta method for its quantile process tailored for these L-statistics, with an important innovation to the choice of L_1 norm (instead of the standard L_∞ norm in the literature). His argument can be extended to instrumental variables regressions, but requires iid observations. Compared to Kaji (2018), we characterise the RLS/IIS and other Huber-skip type estimators as a different type and new class of weighted and marked empirical processes, whose theory has been established by Berenguer-Rico et al. (2019) via martingale decompositions. Our argument is constructed specially for the RLS/IIS, thus it allows us to dig deeper for these two algorithms. For example, our theory can explore the asymptotics of the variance estimator, iteration of the algorithms, the varying robustification parameter, and the algorithms starting with other initial estimators like least trimmed squares. In addition, our analysis does not require iid data and is also valid for time series regressions with stationary, deterministically trending, and unit root regressors.

To establish the outlier distortion test based on the difference between OLS and RLS estimators, Kaji (2018) applies the nonparametric bootstrap by randomly sampling the original data with replacement to draw the joint distribution of the outlier removal and the full sample estimators (RLS and OLS). The bootstrap scheme does not require a distributional assumption on the error term, however, it has lower power under a range of alternatives where there are influential outliers distorting regression parameters. This is because when resampling from the original data, the bootstrapping scheme has a certain chance of picking up outliers. Those bootstrapping samples which inlucde outliers will then distort the distribution of the testing statistics and drive it away from what it should be under the null of no outliers.

Second, Dehon et al. (2012) propose a similar Hausman type test for comparing OLS to the robust S-estimators. Asymptotic theory of the S-estimators has been well established in the robust statistics literature, thus they do not need to derive the asymptotics when constructing the outlier distortion test for the S-estimators. To study the difference between OLS and RLS/IIS, we need to fully establish the theory of RLS/IIS using the empirical process CLT recently developed by Berenguer-Rico et al. (2019), although there have been some related work to research M and L type estimators, see Hendry et al. (2008), Johansen & Nielsen (2009, 2013, 2016a,b, 2019), Jiao & Nielsen (2015). In addition, Dehon et al. (2012) restrict their study in the iid setup, whereas time series regressions are allowed in our framework.

In the wider literature on Huber-skip estimators, there is some other closely related work. Jiao & Pretis (2020) study whether the proportion (and number) of outliers is different to the expected proportion (and number) when no outliers are present. Jiao (2019) and Jiao & Kurle (2021) restrict the setup in iid data, and then extend asymptotic theory of the RLS/IIS algorithms and their false outlier detection rate to instrumental variables regressions. Berenguer-Rico & Nielsen (2018) and Berenguer-Rico & Wilms (2021) considers diagnostic testing on residuals for normality and heteroskedasticity after outlier removal by robust Huber-skip regressions.

We apply our proposed test of outlier distortion to estimate the macro-economic impacts of climate change using the robust IIS estimator. Specifically, we estimate a macro-economic climate impact model in line with the growing panel-econometric literature modelling GDP per capita growth as a function of (non-linear) climate variables. We make two contributions to the existing climate-econometric literature. First, beyond existing estimates, we consider income-based adaptation, allowing the impact of year-on-year changes in temperatures to vary by country-specific income levels. Second, to address un-modelled (and a-priori unknown) idiosyncratic shocks, we apply the robust IIS estimator. We then test for outlier distortion of the estimated coefficients. Our results show significant evidence of income-driven adaptation to temperatures, where climate effects are dampened as incomes increase. However, conventional OLS panel estimates are significantly different to those obtained using the robust estimator.

Once we control for these outliers (some of which coincide with the first Gulf War and collapse of the Soviet Union), the estimated impacts of temperatures on economic growth are attenuated at all income levels compared to conventional OLS estimates. Using these estimates, we then compare the effect of the robust estimation in projections of climate impacts to the end of the century under different temperature scenarios. Projection results consistently indicate that higher levels of warming could lead to higher impacts on GDP per capita, however, we show that the uncertainty reduces considerably when using robust estimation.

NOTE for MORITZ: One sentence here on what this means for the projections.

NOTE for MORITZ: Do we need to describe our application more so as to put more weight on climate side?

The paper proceeds as follows: §2 presents our main results with all proofs shown in Appendix §A. In particular, §2.1 and §2.2 introduces a regression model, a list of assumptions required for analysis, and a class of outlier-robust algorithms including RLS and IIS. §2.3 establishes asymptotic theory of RLS and IIS, whilst §2.4 proposes the outlier distortion test. Then, §3 conducts Monte Carlo studies with additional simulation results in Appendix §B. Lastly, §4 applies the outlier distortion test to the macro-economic impacts of climate change using the robust IIS estimator.

2 Outlier Distortion Test

We consider a linear regression model

$$y_i = x_i'\beta + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.1}$$

for the data $\{(y_i, x_i)\}_{i=1}^n$, where y_i is univariate and x_i is multivariate with the dimension d_x. This setting can represent both classical, time series, and panel regression models. Moreover, in our analysis, regressors x_i can be either stationary, deterministically trending, unit root, or explosive processes. Innovations ε_i/σ are independent of the filtration $\mathcal{F}_{i-1} = \sigma(x_1, \ldots, x_i, \varepsilon_1, \ldots, \varepsilon_{i-1})$ with the common density f and distribution function $\mathsf{F}(c) = \mathsf{P}(\varepsilon_i/\sigma \leq c)$. Denote g as the density of the absolute error $|\varepsilon_i|/\sigma$ and its distribution function by $\mathsf{G}(c) = \mathsf{P}(|\varepsilon_i|/\sigma \leq c)$ for $c > 0$. Assuming symmetry of f, $\mathsf{G}(c) = 2\mathsf{F}(c) - 1$ and $\mathsf{g}(c) = 2\mathsf{f}(c)$. Define $\psi_c = \mathsf{G}(c)$ so the probability of exceeding the cut-off c is $\gamma_c = 1 - \psi_c$. Suppose the k-th moment of the density f exists, we then define the kth moment and truncated moment as

$$\tau_k = \int_{-\infty}^{\infty} u^k \mathsf{f}(u)du, \qquad \tau_k^c = \int_{-c}^{c} u^k \mathsf{f}(u)du. \tag{2.2}$$

Thus, $\tau_0^c = \psi_c$, $\tau_2 = 1$ while $\tau_k = \tau_k^c = 0$ for odd k under symmetry. We define the conditional variance of ε_i/σ given $(|\varepsilon_i|/\sigma \leq c)$ as

$$\varsigma_c^2 = \frac{\tau_2^c}{\psi_c} = \frac{\int_{-c}^{c} u^2 \mathsf{f}(u)du}{\mathsf{P}(|\varepsilon_i| \leq \sigma c)}. \tag{2.3}$$

This is the bias correction factor for the variance estimate computed from the selected non-outlying sample. For the Normal reference $\varepsilon_i/\sigma \sim \mathsf{N}(0,1)$, then $\tau_2^c = \psi_c - 2c\mathsf{f}(c)$, $\tau_4^c = 3\psi_c - 2c(c^2+3)\mathsf{f}(c)$ and $\tau_4 = 3$.

Outliers are pairs of observations that do not conform with the model (2.1) or with the assumed density f. We are interested in the presence of outliers where the errors ε_i/σ are drawn from the reference distribution f but potentially contaminated by an arbitrary (possibly fatter tail) unknown distribution f^{c} under ϵ-contamination as in Huber (1964)

$$(1-\epsilon)\mathsf{f} + \epsilon \mathsf{f}^{\mathsf{c}}. \tag{2.4}$$

This framework allows the data generated by the reference distribution f to be contaminated by an ϵ (between 0 and 1) proportion of outliers. Compared to not imposing a parametric distribution on errors, the mixture model of f and f^{c} in (2.4) is an alternative way to relax the assumption on $\varepsilon_i/\sigma \sim \mathsf{f}$. The null hypothesis of no outliers in the model (2.1) can be formally defined under (2.4) as $H_0 : \epsilon = 0$ against the alternative $H_1 : \epsilon > 0$. The following section describes the iterated 1-step Huber-skip M-estimator where we subsequently derive its asymptotic properties and present a new Durbin-Hausman-Wu type test to formalize outlier robustness checks and to assess $H_0 : \epsilon = 0$.

2.1 Robust Algorithms to Outliers

There are two potential improvements to the trimmed least squares (the simple two step procedure) used in the empirical literature and the example of Acemoglu et al. (2019). First, Johansen & Nielsen (2009) suggest that the updated variance estimator should be corrected by the factor ς^{-2} introduced in (2.3), since the simple procedure underestimates σ^2 in the case where observations are identified by chance and falsely removed as outliers. Second, robustness of the estimator could be improved by iterating the 1-step procedure. Considering these two improvements, we introduce and study the so called *iterated 1-step Huber-skip M-estimators* in Algorithm 2.1:

Algorithm 2.1. *Choose a cut-off $c > 0$.*
1. Choose initial estimators $\widehat{\beta}_c^{(0)}$, $(\widehat{\sigma}_c^{(0)})^2$ and let $m = 0$.
2. Define indicator variables for selecting non-outlying observations

$$v_{i,c}^{(m)} = 1_{(|y_i - x_i'\widehat{\beta}_c^{(m)}| \leq \widehat{\sigma}_c^{(m)} c)}. \tag{2.5}$$

3. Compute least squares estimators

$$\widehat{\beta}_c^{(m+1)} = (\sum_{i=1}^{n} x_i x_i' v_{i,c}^{(m)})^{-1} (\sum_{i=1}^{n} x_i y_i v_{i,c}^{(m)}), \tag{2.6}$$

$$(\widehat{\sigma}_c^{(m+1)})^2 = \varsigma_c^{-2} (\sum_{i=1}^{n} v_{i,c}^{(m)})^{-1} \{\sum_{i=1}^{n} (y_i - x_i'\widehat{\beta}_c^{(m+1)})^2 v_{i,c}^{(m)}\}. \tag{2.7}$$

4. Let $m = m+1$ and repeat 2 and 3.

Having defined the robust algorithm 2.1, we need to specify initial estimators $\widehat{\beta}_c^{(0)}$, $(\widehat{\sigma}_c^{(0)})^2$. The first initial estimator in our analysis uses the full sample OLS estimates as in Acemoglu et al. (2019) and we refer to this as *Robustified Least Squares* (RLS). As a second approach, we use a robust initial estimator and refer to this approach as *Impulse Indicator Saturation* (IIS), described in Algorithm 2.2. In IIS we divide the full sample into two sub-samples and use regression estimates calculated from each sub-sample to detect outliers in the other sub-sample:

Algorithm 2.2. ***Stylized Impulse Indicator Saturation****. Choose a cut-off $c > 0$.*
1.1. Split full sample into two sets $\mathcal{I}_j$, $j = 1, 2$ of n_j observations where $\sum_{j=1}^2 n_j = n$.
1.2. Calculate least squares estimators based upon each sub-sample $\mathcal{I}_j$ for $j = 1, 2$

$$\widehat{\beta}_j = (\sum_{i\in\mathcal{I}_j} x_i x_i')^{-1}(\sum_{i\in\mathcal{I}_j} x_i y_i), \qquad \widehat{\sigma}_j^2 = \frac{1}{n_j}\sum_{i\in\mathcal{I}_j}(y_i - x_i'\widehat{\beta}_j)^2. \tag{2.8}$$

1.3. Define the initial indicator variables for selecting non-outlying observations

$$v_{i,c}^{(0)} = 1_{(i\in\mathcal{I}_1)} 1_{(|y_i - x_i'\widehat{\beta}_2| \leq \widehat{\sigma}_2 c)} + 1_{(i\in\mathcal{I}_2)} 1_{(|y_i - x_i'\widehat{\beta}_1| \leq \widehat{\sigma}_1 c)}. \tag{2.9}$$

1.4. Compute $\widehat{\beta}_c^{(1)}$, $(\widehat{\sigma}_c^{(1)})^2$ using (2.6), (2.7) with $m = 0$, and then let $m = 1$.
2. Follow the step 2,3,4 in Algorithm 2.1.

The iterated 1-step Huber-skip M-estimator mimics the Huber (1964) skip[2] estimator with the criterion function

$$\rho(t) = \begin{cases} \frac{t^2}{2}, & \text{if } |t| \leq c, \\ \frac{c^2}{2}, & \text{otherwise,} \end{cases} \tag{2.10}$$

which is immune to either outliers or a fatter tail distribution (defined relative to the reference f) under the ϵ-contamination $(1-\epsilon)\mathsf{f} + \epsilon\mathsf{f}^c$. Johansen & Nielsen (2013) argues that Algorithm 2.1 is an approximation to and thus an implementation of the Huber-skip regression. Jiao & Pretis (2020) provide the guidance for selecting the robustification parameter c and address the testing problem for overall presence of outliers by theoretically analyzing the false discovery rate (referred to as the gauge) of outliers detected by Algorithm 2.1. The main purpose of this paper is to formalize outlier robustness checks using Algorithm 2.1 as the Durbin-Hausman-Wu type test by establishing weak convergence of the iterated 1-step Huber-skip M-estimator with the drifting cut-off c.

2.2 Assumptions for Asymptotic Theory

Innovations ε_i and regressors x_i must satisfy moment conditions as outlined in Assumption 2.1 for our asymptotic analysis. Regressors x_i can be temporally dependent and deterministically or stochastically trending. We therefore require a normalisation matrix N that allows for different behaviour of the components of the regressor vector

[2] See Hampel et al. (1986) (p. 104) for the Huber-skip type estimator as opposed to the Huber estimator with the criterion function

$$\rho(t) = \begin{cases} \frac{t^2}{2}, & \text{if } |t| \leq c, \\ c|t| - c^2/2, & \text{otherwise.} \end{cases}$$

.

x_i. In the case of a stationary regressor we need a standard $n^{-1/2}$ normalisation so that N must be proportional to the identity matrix of the same dimension as x_i, that is $N = n^{-1/2} I_{d_x}$. Likewise, if x_i is a random walk we have $N = n^{-1} I_{d_x}$. Explosive processes are also allowed by our analysis, for example $x_i = 2^i$ such that the normalization becomes 2^{-n}. If the regressors are unbalanced as in $x_i = (1, i, 2^i)'$ we can choose $N = \mathrm{diag}(n^{-1/2}, n^{-3/2}, 2^{-n})$. Thus, denote the normalized regressors as $x_{in} = N' x_i$.

Assumption 2.1. *Let $\mathcal{F}_i$ be an increasing sequence of σ-fields so ε_{i-1} and x_i are $\mathcal{F}_{i-1}$ measurable and ε_i is independent of $\mathcal{F}_{i-1}$. Let ε_i/σ have a symmetric, continuously differentiable density f which is positive on the real line $\mathbb{R}$. For some values of η such that $0 < \eta \leq 1/4$, choose an integer $r \geq 2$ so*

$$2^{r-1} > 1 + (1/4 - \eta)(1 + d_x). \tag{2.11}$$

Let $q = 1 + 2^{r+1}$. Suppose
(i) the density f satisfies
(a) $|u|^q \mathsf{f}(u)$, $|u^{q+1} \dot{\mathsf{f}}(u)|$ are decreasing for large u;
(ii) the regressors x_i satisfy
(a) $\Sigma_n = \sum_{i=1}^n x_{in} x_{in}' \overset{\mathsf{D}}{\to} \Sigma \overset{a.s.}{>} 0$;
(b) $n^{-1} \mathsf{E} \sum_{i=1}^n |n^{1/2} x_{in}|^q = \mathrm{O}(1)$;
(iii) the initial estimator $(\widetilde{\beta}, \widetilde{\sigma}^2)$ satisfies
(a) $N^{-1}(\widetilde{\beta} - \beta) = \mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$;
(b) $n^{1/2}(\widetilde{\sigma}^2 - \sigma^2) = \mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$.

While these assumptions may appear abstract, conditions (i), (ii) are satisfied in a range of situations. In particular, the condition (i) is satisfied by the Normal and t distribution; see Remark 2.5 in Berenguer-Rico & Nielsen (2018), while the condition (ii) is satisfied by stationary, random walk and deterministically trending regressors; see Example 3.2 in Johansen & Nielsen (2016a), as well as by explosive processes; see Remark 4.2(c) in Berenguer-Rico et al. (2019). Condition (iii) allows the standardised estimation errors to diverge at a rate of $n^{1/4-\eta}$ rather than being bounded in probability. In particular, $\eta = 1/4$ can be chosen for estimators with standard convergence rates.

There is a trade-off in (2.11) between η, the divergence rate of initial estimators $\widetilde{\beta}$, $\widetilde{\sigma}^2$, and r, the required number of moments for innovations ε_i and regressors x_i. If we have standard initial estimators which are bounded in probability after normalization, such as Robustified Least Squares and Impulse Indicator Saturation, then η becomes $1/4$ so we can choose $r = 2$ regardless of dimension of regressors, implying we only require the lower number of moments. Whereas for non-standard diverging estimators, i.e. $0 < \eta < 1/4$ and $1/4 - \eta > 0$, then the required number r of moments grows linearly with the dimension of the regressor. This would be relevant for the $n^{1/3}$-consistent least median of squares regression estimator proposed by Rousseeuw (1984).

2.3 Weak Convergence

We now provide asymptotic theory for the iterated one-step Huber-skip M-estimator, such as tightness, stochastic expansions, fixed points of the iterated estimators calculated by Algorithm 2.1. The argument holds uniformly in cut-off values. Then, weak convergence is given for Robustified Least Squares and Impulse Indicator Saturation.

The first theorem shows that the iterated estimator produced by Algorithm 2.1 is tight in the iteration $m \in [0, \infty)$ and in the cut-off value $c \in [c_+, \infty)$. Note that $c_+ > 0$ is a small positive number.

Theorem 2.1. *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 2.1. Suppose Assumption 2.1 holds with $\eta = 1/4$. Then, as $n \to \infty$*

$$\sup_{0\leq m<\infty} \sup_{c_+\leq c<\infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \beta)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)| = \mathrm{O}_{\mathsf{P}}(1).$$

The proof involves empirical process theory recently developed by Berenguer-Rico et al. (2019). Firstly, Assumption 2.1(*iii*) with $\eta = 1/4$ corresponds to a standard convergence rate for the initial estimator. With the 1-step relationship between the updated and the original estimator provided by Lemma A.3, also see Corollary 2.1, the tightness can then be demonstrated by a geometric argument and mathematical induction.

Firstly, the above tightness theorem implies the uniform consistency of the iterated estimators computed by Algorithm 2.1, that is $\widehat{\beta}_c^{(m)} \overset{\mathsf{P}}{\to} \beta$ and $\widehat{\sigma}_c^{(m)} \overset{\mathsf{P}}{\to} \sigma$ uniformly in the cut-off value c and iteration step m as $n \to \infty$. Secondly, the tightness will also be used to establish fixed points $\widehat{\beta}_c^{(*)}$ and $\widehat{\sigma}_c^{(*)}$ of Algorithm 2.1 upon through infinite iterations when $m \to \infty$ and to demonstrate weak convergence theory of RLS and IIS.

The next theorem is to show stochastic expansions of any iterated step estimators of Algorithm 2.1 in terms of the initial estimators, kernels, and small remainder terms.

Theorem 2.2. *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 2.1. Suppose Assumption 2.1 holds with $\eta = 1/4$. Then, as $n \to \infty$ and uniformly in $c \in [c_+, \infty)$, we have for any $m \in [0, \infty)$*

$$N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = \varrho_{\beta,c}^{(m+1)} N^{-1}(\widehat{\beta}_c^{(0)} - \beta) + \varrho_{x\varepsilon,c}^{(m+1)} \Sigma_n^{-1} \sum_{i=1}^{n} x_{in}\varepsilon_i 1_{(|\varepsilon_i|\leq\sigma c)} + \mathrm{o}_{\mathsf{P}}(1),$$

$$n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = \varrho_{\sigma,c}^{(m+1)} n^{1/2}(\widehat{\sigma}_c^{(0)} - \sigma) + \varrho_{\varepsilon\varepsilon,c}^{(m+1)} n^{-1/2} \sum_{i=1}^{n} (\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2) 1_{(|\varepsilon_i|\leq\sigma c)} + \mathrm{o}_{\mathsf{P}}(1),$$

where coefficients have expressions

$$\varrho_{\beta,c}^{(m+1)} = \{\frac{2c\mathsf{f}(c)}{\psi_c}\}^{m+1}, \qquad \varrho_{x\varepsilon,c}^{(m+1)} = \frac{\psi_c^{m+1} - \{2c\mathsf{f}(c)\}^{m+1}}{\psi_c^{m+1}\{\psi_c - 2c\mathsf{f}(c)\}},$$

$$\varrho_{\sigma,c}^{(m+1)} = \{\frac{c(c^2-\varsigma_c^2)\mathsf{f}(c)}{\tau_2^c}\}^{m+1}, \qquad \varrho_{\varepsilon\varepsilon,c}^{(m+1)} = \sigma\frac{(\tau_2^c)^{m+1} - \{c(c^2-\varsigma_c^2)\mathsf{f}(c)\}^{m+1}}{2(\tau_2^c)^{m+1}\{\tau_2^c - c(c^2-\varsigma_c^2)\mathsf{f}(c)\}}.$$

The above theorem generalises Lemma A.3 in the sense that its proof is to recursively apply the one-step expansion of the updated estimator in terms of the original ones. Let $m = 0$ such that $\varrho_{\beta,c}^{(1)} = 2c\mathsf{f}(c)/\psi_c$, $\varrho_{x\varepsilon,c}^{(1)} = \psi_c^{-1}$, $\varrho_{\sigma,c}^{(1)} = c(c^2-\varsigma_c^2)\mathsf{f}(c)/\tau_2^c$, and $\varrho_{\varepsilon\varepsilon,c}^{(1)} = \sigma/(2\tau_2^c)$, then the expansion immediately reduces to the one-step case. Here, we provide the following corollary to re-express Lemma A.3 as a stochastic expansion of the fist step estimators in terms of the initial ones and the kernel terms.

Corollary 2.1. *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 2.1. Suppose Assumption 2.1 holds with $\eta = 1/4$. Then, as $n \to \infty$ and uniformly in*

$c \in [c_+, \infty)$, *we have*

$$N^{-1}(\widehat{\beta}_c^{(1)} - \beta) = \frac{2c\mathsf{f}(c)}{\psi_c} N^{-1}(\widehat{\beta}_c^{(0)} - \beta) + (\psi_c \Sigma_n)^{-1} \sum_{i=1}^n x_{in}\varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)} + \mathsf{o_P}(1),$$

$$n^{1/2}(\widehat{\sigma}_c^{(1)} - \sigma) = \frac{c(c^2 - \varsigma_c^2)\mathsf{f}(c)}{\tau_2^c} n^{1/2}(\widehat{\sigma}_c^{(0)} - \sigma) + \frac{\sigma}{2\tau_2^c} n^{-1/2} \sum_{i=1}^n (\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2) 1_{(|\varepsilon_i| \leq \sigma c)} + \mathsf{o_P}(1).$$

Initially the tight estimator is assumed to be available and it is subsequently iterated through the above one-step expansion. Assumption 2.1(i) implies that autoregressive coefficients $2c\mathsf{f}(c)/\psi_c$ and $c(c^2 - \varsigma_c^2)\mathsf{f}(c)/\tau_2^c$ are strictly bounded by one holding uniformly in c, suggesting that the above equation is a contraction mapping. Thus, Algorithm 2.1 will converge to a fixed point as the iteration step increases to be sufficiently large. Let $m \to \infty$ in Theorem 2.2 such that $\varrho_{\beta,c}^{(\infty)} = 0$, $\varrho_{x\varepsilon,c}^{(\infty)} = 1/\{\psi_c - 2c\mathsf{f}(c)\}$, $\varrho_{\sigma,c}^{(\infty)} = 0$, and $\varrho_{\varepsilon\varepsilon,c}^{(\infty)} = \sigma/[2\{\tau_2^c - c(c^2 - \varsigma_c^2)\mathsf{f}(c)\}]$, then the below theorem finds the fixed point $N^{-1}(\widehat{\beta}_c^{(*)} - \beta) = N^{-1}(\widehat{\beta}_c^{(\infty)} - \beta)$, $n^{1/2}(\widehat{\sigma}_c^{(*)} - \sigma) = n^{1/2}(\widehat{\sigma}_c^{(\infty)} - \sigma)$.

Theorem 2.3. *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 2.1. Suppose Assumption 2.1 holds with $\eta = 1/4$. Then, for all $\epsilon, \delta > 0$ a pair $m_0, n_0 > 0$ exists, so for all $m > m_0$ and $n > n_0$*

$$\mathsf{P}\{\sup_{c_+ \leq c < \infty} |N^{-1}(\widehat{\beta}_c^{(m)} - \widehat{\beta}_c^{(*)})| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \widehat{\sigma}_c^{(*)})| > \delta\} < \epsilon,$$

where

$$N^{-1}(\widehat{\beta}_c^{(*)} - \beta) = \frac{1}{\psi_c - 2c\mathsf{f}(c)} \Sigma_n^{-1} \sum_{i=1}^n x_{in}\varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)},$$

$$n^{1/2}(\widehat{\sigma}_c^{(*)} - \sigma) = \frac{\sigma}{2\{\tau_2^c - c(c^2 - \varsigma_c^2)\mathsf{f}(c)\}} n^{-1/2} \sum_{i=1}^n (\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2) 1_{(|\varepsilon_i| \leq \sigma c)}.$$

The proof is conducted as follows. According to Theorem 2.1, if the initial estimator is bounded in a large compact set with a large probability, then any iterated estimators of Algorithm 2.1 take values in the same compact set no matter what value of the cut-off c is chosen. Next, the argument is to further demonstrate that the deviation between the m-fold iterated estimator and the fixed point is the sum of two terms vanishing exponentially and in probability respectively as m and n go to infinity.

Theorem 2.3 shows the first order asymptotics of the fixed point of Algorithm 2.1 and it is in fact the same as that of the Huber (1964) skip estimator, thus the iterated one-step Huber-skip M-estimator mimics the Huber-skip estimator and Algorithm 2.1 can be understood as an implementation of the Huber-skip estimation as a non-linear optimisation problem.

The choice of the initial estimator does not affect Algorithm 2.1 in terms of finding the fixed point, as long as it has the tightness property. Thus, the fixed point theorem also applies for RLS and IIS as well as the theorems regarding to tightness and stochastic expansions, since the full sample or split sample least squares as their starting estimators are tight. Next concentrating on RLS and IIS, we first show their stochastic expansions in order to establish weak convergence theory.

Theorem 2.4. *Consider Robustified Least Squares (RLS) or split half Impulse Indicator Saturation where $n_1 = \text{int}[n/2]$ and $n_2 = n - n_1$ (IIS). Suppose Assumption 2.1(i, ii) holds for each sub-sample set $\mathcal{I}_1$, $\mathcal{I}_2$. Then, as $n \to \infty$ and uniformly in $c \in [c_+, \infty)$ we have for any $m \in [0, \infty)$*

$$N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = \varrho_{\beta,c}^{(m+1)} \Sigma_n^{-1} \sum_{i=1}^n x_{in}\varepsilon_i + \varrho_{x\varepsilon,c}^{(m+1)} \Sigma_n^{-1} \sum_{i=1}^n x_{in}\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} + \mathrm{o_P}(1),$$

$$n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = \varrho_{\sigma,c}^{(m+1)} \frac{\sigma}{2} n^{-1/2} \sum_{i=1}^n (\frac{\varepsilon_i^2}{\sigma^2} - 1)$$
$$+ \varrho_{\varepsilon\varepsilon,c}^{(m+1)} n^{-1/2} \sum_{i=1}^n (\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2) 1_{(|\varepsilon_i| \le \sigma c)} + \mathrm{o_P}(1),$$

where the coefficients $\varrho_{\beta,c}^{(m+1)}, \varrho_{x\varepsilon,c}^{(m+1)}, \varrho_{\sigma,c}^{(m+1)}, \varrho_{\varepsilon\varepsilon,c}^{(m+1)}$ are defined in Theorem 2.2.

Substituting the expansion of the full sample least squares as the initial estimator in Theorem 2.2, we immediately prove the above theorem for RLS. Then, we find that the split half version of IIS has the identical expansion as RLS. The first step updated estimator and the fixed point of RLS and IIS are two special cases of our main interest, so their expansions can be obtained by letting $m = 0$ and $m \to \infty$. With Theorem 2.4, we are now ready to establish the weak convergence theory for RLS and IIS in cases where the regressors are either stationary, deterministic trend, or unit root processes.

Theorem 2.1 demonstrates uniform consistency of the iterated estimators of β, σ^2. In this paper, we are concerned about whether the β coefficient is distorted due to outliers. Thus, the next step is to focus on distributional analysis of the iterated estimator of β. Thus, using the stochastic expansion of the β estimator in Theorem 2.4, it suffices to analyse the distribution of the kernel vector $\sum_{i=1}^n (x_{in}'\varepsilon_i, x_{in}'\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)})'$. With this purpose in mind, we need to first discuss in a few situations the choice of the normalisation matrix N in $x_{in} = N'x_i$ and the limiting behaviour of the covariance matrix of the normalised regressors $\Sigma_n = \sum_{i=1}^n x_{in}x_{in}'$.

Stationary case. Suppose the regressors are cross-sectional iid or arise from a stationary time series model. Thus, we choose $N = n^{-1/2} I_{d_x}$ for normalising the regressors such that $x_{in} = N'x_i = n^{-1/2}x_i$. Then, $\Sigma_n = \sum_{i=1}^n x_{in}x_{in}' = n^{-1}\sum_{i=1}^n x_i x_i'$ converges in probability to a deterministic term $\Sigma = \mathsf{E}x_i x_i'$ by LLN. To investigate the asymptotic behaviour of the kernel vector, we apply martingale CLT so that for any $c \in [c_+, \infty)$

$$\sum_{i=1}^n \begin{pmatrix} x_{in}\varepsilon_i \\ x_{in}\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} \end{pmatrix} = n^{-1/2} \sum_{i=1}^n \begin{pmatrix} x_i\varepsilon_i \\ x_i\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} \end{pmatrix} \xrightarrow{\mathsf{D}} \mathsf{N}\left\{ \begin{pmatrix} 0_{d_x} \\ 0_{d_x} \end{pmatrix}, \sigma^2 \tau_2^c \begin{pmatrix} \frac{1}{\tau_2^c}\Sigma & \Sigma \\ \Sigma & \Sigma \end{pmatrix} \right\}.$$

Drifting cut-off values c in the interval $[c_+, \infty)$, we obtain a sequence of processes $\mathbb{G}_n^{(m+1)}(c) = N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta)$ for any $m \in [0, \infty)$. We can now establish a weak convergence theory for $\mathbb{G}_n^{(m+1)}$, which follows from a finite dimensional convergence and tightness; see Billingsley (1968). The below theorem then shows that RLS and IIS are asymptotically approximated by a Gaussian process.

Theorem 2.5. *Consider RLS or IIS. Suppose Assumption 2.1(i, ii) holds. For any $m \in [0, \infty)$, denote the processes $\mathbb{G}_n^{(m+1)}(c) = n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta)$ with the argument*

$c \in [c_+, \infty)$. *Then as* $n \to \infty$, $\mathbb{G}_n^{(m+1)}$ *weakly converges to a zero mean Gaussian process* $\mathbb{G}^{(m+1)}$ *with variance*

$$\mathsf{Var}\{\mathbb{G}^{(m+1)}(c)\} = \{(\varrho_{\beta,c}^{(m+1)})^2 + 2\tau_2^c \varrho_{\beta,c}^{(m+1)} \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c (\varrho_{x\varepsilon,c}^{(m+1)})^2\}\sigma^2\Sigma^{-1},$$

where $\varrho_{\beta,c}^{(m+1)}$, $\varrho_{x\varepsilon,c}^{(m+1)}$ *are defined in Theorem 2.2.*

Again, the one-step updated estimator and fixed point are of particular interest in RLS or IIS. To explore their weak convergence, let $m = 0$ and $m \to \infty$ in Theorem 2.5 such that $\varrho_{\beta,c}^{(1)} = 2c\mathsf{f}(c)/\psi_c$, $\varrho_{x\varepsilon,c}^{(1)} = \psi_c^{-1}$ and $\varrho_{\beta,c}^{(\infty)} = 0$, $\varrho_{x\varepsilon,c}^{(\infty)} = 1/\{\psi_c - 2c\mathsf{f}(c)\}$, then we have the below corollary.

Corollary 2.2. *Consider RLS or IIS. Suppose Assumption 2.1(i, ii) holds. The cut-off* c *drifts in the interval* $[c_+, \infty)$. *Then as* $n \to \infty$, *for* $m = 0$ *a sequence of processes* $\mathbb{G}_n^{(1)}$ *of the initial estimator weakly converges to a zero mean Gaussian process* $\mathbb{G}^{(1)}$ *with variance*

$$\mathsf{Var}\{\mathbb{G}^{(1)}(c)\} = \frac{4c^2\mathsf{f}^2(c) + 4\tau_2^c c\mathsf{f}(c) + \tau_2^c}{\psi_c^2}\sigma^2\Sigma^{-1}.$$

In addition, for $m \to \infty$ *a sequence of processes* $\mathbb{G}_n^{(*)}$ *of the fixed point estimator weakly converges to a zero mean Gaussian process* $\mathbb{G}^{(*)}$ *with variance*

$$\mathsf{Var}\{\mathbb{G}^{(*)}(c)\} = \frac{\tau_2^c}{\{\psi_c - 2c\mathsf{f}(c)\}^2}\sigma^2\Sigma^{-1}.$$

Deterministic trends. To keep notations simple and analysis clear, here we consider a straightforward example which follows the regression

$$y_i = \beta_0 + \beta_1 i + \varepsilon_i, \quad i = 1, 2, \ldots, n.$$

For the regressors $x_i = (1, i)'$, we choose the normalisation matrix $N = \begin{pmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{pmatrix}$ so that $x_{in} = N'x_i = (n^{-1/2}, n^{-3/2}i)'$. Then, it follows

$$\Sigma_n = \sum_{i=1}^n x_{in}x_{in}' = \sum_{i=1}^n \begin{pmatrix} n^{-1} & n^{-2}i \\ n^{-2}i & n^{-3}i^2 \end{pmatrix} \to \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix} = \Sigma.$$

Notice that we use $\sum_{i=1}^n i = n(n+1)/2$ and $\sum_{i=1}^n i^2 = n(n+1)(2n+1)/6$ to obtain the above deterministic limit. The kernel vector $\sum_{i=1}^n (x_{in}'\varepsilon_i, x_{in}'\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)})'$ has a limiting normal distribution with the same form of mean and variance as given in the stationary case where instead $d_x = 2$ and Σ is derived immediately above. For any $m \in [0, \infty)$, denote a sequence of processes of the iterated estimators computed by RLS or IIS as $\mathbb{G}_n^{(m+1)}(c) = N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = \begin{Bmatrix} n^{1/2}(\widehat{\beta}_{0,c}^{(m+1)} - \beta_0) \\ n^{3/2}(\widehat{\beta}_{1,c}^{(m+1)} - \beta_1) \end{Bmatrix}$ with $c \in [c_+, \infty)$. Thus as $n \to \infty$, $\mathbb{G}_n^{(m+1)}$ weakly converges to a zero mean Gaussian process with the same form of variance as given in Theorem 2.5 and Corollary 2.2 where again Σ needs to be changed to the one shown above.

Other cases of deterministic trends can be studied using a similar analysis to the above example. For instance, the argument applies to trend stationary autoregressions but involves a notationally tedious detrending derivation; see Section 1.5.1 in Johansen & Nielsen (2009) for a related and more detailed description.

Unit roots. Consider the I(1) process, which follows the autoregression

$$y_i = \beta y_{i-1} + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\beta = 1$. Other cases of unit roots processes can be analysed similarly using the following argument. Firstly, we choose $N = n^{-1}$ for normalising the regressor $x_i = y_{i-1}$ so that $x_{in} = N'x_i = n^{-1}y_{i-1} = n^{-1}y_0 + n^{-1}\sum_{s=1}^{i-1}\varepsilon_s$. For any $c \in [c_+, \infty)$ and as $n \to \infty$, the functional CLT shows that a sequence of processes $n^{-1/2}\sum_{s=1}^{\mathrm{int}(nu)}\begin{pmatrix}\varepsilon_s \\ \varepsilon_s 1_{(|\varepsilon_s| \le \sigma c)}\end{pmatrix}$ weakly converges to a Brownian motion $\begin{pmatrix}B_{1,u} \\ B_{2,u}\end{pmatrix}$ with the argument $u \in [0,1]$ having the mean zero and variance $\sigma^2\begin{pmatrix}1 & \tau_2^c \\ \tau_2^c & \tau_2^c\end{pmatrix}$. Next, we find

$$\Sigma_n = \sum_{i=1}^n x_{in}x_{in}' = n^{-2}\sum_{i=1}^n y_{i-1}^2 = n^{-1}\sum_{i=1}^n (n^{-1/2}\sum_{s=1}^{i-1}\varepsilon_s + n^{-1/2}y_0)^2 \overset{\mathsf{D}}{\to} \int_0^1 B_{1,u}^2 du = \Sigma.$$

For any $c \in [c_+, \infty)$, the kernel vector then follows

$$\sum_{i=1}^n \begin{pmatrix} x_{in}\varepsilon_i \\ x_{in}\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)}\end{pmatrix} = n^{-1/2}\sum_{i=1}^n \begin{pmatrix} n^{-1/2}y_{i-1}\varepsilon_i \\ n^{-1/2}y_{i-1}\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)}\end{pmatrix} \overset{\mathsf{D}}{\to} \begin{pmatrix}\int_0^1 B_{1,u}dB_{1,u} \\ \int_0^1 B_{1,u}dB_{2,u}\end{pmatrix}.$$

Theorem 2.4 shows that

$$N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = \begin{pmatrix}\varrho_{\beta,c}^{(m+1)}\Sigma_n^{-1} \\ \varrho_{x\varepsilon,c}^{(m+1)}\Sigma_n^{-1}\end{pmatrix}'\sum_{i=1}^n \begin{pmatrix} x_{in}\varepsilon_i \\ x_{in}\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)}\end{pmatrix} + \mathrm{o}_{\mathsf{P}}(1).$$

With the above results, we can now establish the limiting distribution of the iterated estimator produced by RLS or IIS for any cut-off value so that for any $m \in [0, \infty)$ and $c \in [c_+, \infty)$ it follows as $n \to \infty$ that

$$n(\widehat{\beta}_c^{(m+1)} - \beta) \overset{\mathsf{D}}{\to} \frac{\varrho_{\beta,c}^{(m+1)}\int_0^1 B_{1,u}dB_{1,u} + \varrho_{x\varepsilon,c}^{(m+1)}\int_0^1 B_{1,u}dB_{2,u}}{\int_0^1 B_{1,u}^2 du}.$$

Remember from Theorem 2.2 that

$$\varrho_{\beta,c}^{(m+1)} = \{\frac{2c\mathsf{f}(c)}{\psi_c}\}^{m+1}, \qquad \varrho_{x\varepsilon,c}^{(m+1)} = \frac{\psi_c^{m+1} - \{2c\mathsf{f}(c)\}^{m+1}}{\psi_c^{m+1}\{\psi_c - 2c\mathsf{f}(c)\}}.$$

Once again, special interests lie in the first step and fixed point estimators of RLS and IIS. Thus, let $m = 0$ and $m \to \infty$ in the above limiting distribution such that $\varrho_{\beta,c}^{(1)} = 2c\mathsf{f}(c)/\psi_c$, $\varrho_{x\varepsilon,c}^{(1)} = \psi_c^{-1}$ and $\varrho_{\beta,c}^{(\infty)} = 0$, $\varrho_{x\varepsilon,c}^{(\infty)} = 1/\{\psi_c - 2c\mathsf{f}(c)\}$, then for any $c \in [c_+, \infty)$ it follows as $n \to \infty$ that

$$n(\widehat{\beta}_c^{(1)} - \beta) \overset{\mathsf{D}}{\to} \frac{2c\mathsf{f}(c)\int_0^1 B_{1,u}dB_{1,u} + \int_0^1 B_{1,u}dB_{2,u}}{\psi_c\int_0^1 B_{1,u}^2 du},$$

$$n(\widehat{\beta}_c^{(*)} - \beta) \overset{\mathsf{D}}{\to} \frac{\int_0^1 B_{1,u}dB_{2,u}}{\{\psi_c - 2c\mathsf{f}(c)\}\int_0^1 B_{1,u}^2 du}.$$

When $c \to \infty$ then $\psi_c \to 1$, $\tau_2^c \to 1$, and $c\mathsf{f}(c) \to 0$ such that $\varrho_{\beta,c}^{(m+1)} \to 0$ and $\varrho_{x\varepsilon,c}^{(m+1)} \to 1$ for any $m \in [0,\infty)$ due to Assumption 2.1(i). Furthermore, two Brownian motions B_1 and B_2 become identical, since $1_{(|\varepsilon_s| \leq \sigma c)} \to 1$ and their variances $\sigma^2 \begin{pmatrix} 1 & \tau_2^c \\ \tau_2^c & \tau_2^c \end{pmatrix} \to \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Thus, the limiting distribution of the iterated estimators

$$\frac{\varrho_{\beta,c}^{(m+1)} \int_0^1 B_{1,u} dB_{1,u} + \varrho_{x\varepsilon,c}^{(m+1)} \int_0^1 B_{1,u} dB_{2,u}}{\int_0^1 B_{1,u}^2 du} \to \frac{\int_0^1 B_{1,u} dB_{1,u}}{\int_0^1 B_{1,u}^2 du},$$

which is the usual Dicky-Fuller distribution. This is not surprising, since for any $m \in [0,\infty)$ and as $c \to \infty$ the iterated estimator $n(\widehat{\beta}_c^{(m+1)} - \beta)$ computed by RLS or IIS becomes the ordinary least squares $n(\widetilde{\beta} - \beta)$ which then converges to the Dicky-Fuller distribution as $n \to \infty$.

With the above weak convergence results of RLS and IIS, the next step is to construct tests for coefficients distortion due to outliers. The proposed tests formalise the common practice for outlier robustness checks by looking into the difference between RLS (or IIS) and OLS.

2.4 Testing Methods

A frequent concern in empirical economics is whether a tiny set of outliers may have invalidated empirical results. The common practice is to carry out robustness checks by redoing the analyses with the sample after trimming all outliers detected by Algorithm 2.1 and comparing results from the original ones with the full sample. Instead of heuristically checking the difference between the trimmed LS and OLS without knowing whether it is statistically significant, this paper formalizes the outlier robustness check as a new type of Durbin (1954)-Hausman (1978)-Wu (1973) test.

The test is based on trade-off between robustness and efficiency and enables to judge whether the least squares estimation is appropriate or the robust method should be preferred. The robust estimator produced by Algorithm 2.1 is consistent both under the null and alternatives (although less efficient under the null[3]), whereas OLS is efficient (and consistent) under the null, but inconsistent otherwise. The test statistics is looking on statistically significant difference between RLS/IIS and OLS. If the model is correctly specified, the Hausman type test statistics should be rather small under the null of no outliers, since two consistent methods should produce estimates that are very close to the population. On the contrary when outliers have large influence on least squares estimation, the robust method should be very different from the ordinary estimate, so OLS should be rejected and the RLS/IIS should be preferred. The proposed test thus will evaluate whether the gain in robustness is more valuable than the corresponding loss in efficiency.

In practice, empirical researchers run the full sample OLS $\widetilde{\beta}$ and compare to the trimmed LS $\widehat{\beta}_c^{(m+1)}$. Our proposed test can detect whether two estimates are significantly distinct by exploring on the L2 norm of the difference between $\widetilde{\beta}$ and $\widehat{\beta}_c^{(m+1)}$. Thus, for any $m \in [0,\infty)$ it is essential first to derive the stochastic expansion and

[3]The trimmed LS would throw away observations wrongly classified as outliers and thus has the higher asymptotic variance than OLS under the null; See the weak convergence result of RLS and IIS in §2.3; Also see the discussion in Johansen & Nielsen (2009) on the relative efficiency factor (efficiency loss) of IIS.

weak limit of a sequence of stochastic processes $\mathbb{H}_n^{(m+1)}(c) = N^{-1}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ with the argument $c \in [c_+, \infty)$. We concentrate on stationary regressors in this section, where $N = n^{-1/2} I_{d_x}$ such that $\mathbb{H}_n^{(m+1)}(c) = n^{1/2}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$, $x_{in} = n^{-1/2} x_i$, and $\Sigma = \mathsf{E} x_i x_i'$. With the weak convergence results shown in §2.3, other cases such as deterministic trends and unit roots can be analysed using the same argument.

Theorem 2.6. *Consider RLS or IIS. Suppose Assumption 2.1(i, ii) holds. For any $m \in [0, \infty)$, denote the processes $\mathbb{H}_n^{(m+1)}(c) = n^{1/2}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ with the argument $c \in [c_+, \infty)$. Then as $n \to \infty$, we have*

$$\mathbb{H}_n^{(m+1)}(c) = (\varrho_{\beta,c}^{(m+1)} - 1)\Sigma_n^{-1} \sum_{i=1}^{n} x_{in}\varepsilon_i + \varrho_{x\varepsilon,c}^{(m+1)} \Sigma_n^{-1} \sum_{i=1}^{n} x_{in}\varepsilon_i 1_{(|\varepsilon_i| \le \sigma c)} + \mathsf{o_P}(1),$$

where $\varrho_{\beta,c}^{(m+1)}$, $\varrho_{x\varepsilon,c}^{(m+1)}$ are defined in Theorem 2.2. Furthermore, $\mathbb{H}_n^{(m+1)}$ weakly converges to a zero mean Gaussian process $\mathbb{H}^{(m+1)}$ with variance given as

$$\mathsf{Var}\{\mathbb{H}^{(m+1)}(c)\} = \{(\varrho_{\beta,c}^{(m+1)} - 1)^2 + 2\tau_2^c(\varrho_{\beta,c}^{(m+1)} - 1)\varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c(\varrho_{x\varepsilon,c}^{(m+1)})^2\}\sigma^2\Sigma^{-1}.$$

The proof immediately follows from the stochastic expansion and weak convergence of $\widehat{\beta}_c^{(m+1)}$ in Theorem 2.4 and 2.5. Using the weak convergence result of $\mathbb{H}_n^{(m+1)}$, we next establish the outlier distortion test.

Corollary 2.3. *Consider RLS or IIS. Suppose Assumption 2.1(i, ii) holds. For any $m \in [0, \infty)$, $c \in [c_+, \infty)$ and as $n \to \infty$, we have*

$$n^{1/2}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) \overset{\mathsf{D}}{\to} \mathsf{N}\{0_{d_x}, \mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})\},$$

where $\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) = \{(\varrho_{\beta,c}^{(m+1)} - 1)^2 + 2\tau_2^c(\varrho_{\beta,c}^{(m+1)} - 1)\varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c(\varrho_{x\varepsilon,c}^{(m+1)})^2\}\sigma^2\Sigma^{-1}$. Then, the proposed test statistics has the weak limit

$$H_{n,c}^{(m+1)} = n(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})'\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})^{-1}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) \overset{\mathsf{D}}{\to} \chi^2_{d_x}.$$

When two estimators are correlated: one (RLS/IIS $\widehat{\beta}_c^{(m+1)}$) is always consistent but inefficient under the null (of no outliers), the other (OLS $\widetilde{\beta}$) efficient but not consistent under alternatives, then the asymptotic variance of their difference is given by the difference of their respective asymptotic variances under the null. Using the similar argument to Hausman (1978), Lemma A.4 in the appendix demonstrates the above statement in our context. In addition, we provide a different but more direct proof in Remark A.1 to show $\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) = \mathsf{avar}(\widehat{\beta}_c^{(m+1)}) - \mathsf{avar}(\widetilde{\beta})$ under the null of no outliers. It makes use of the asymptotics derived for $\widehat{\beta}_c^{(m+1)}$ in §2.3 and indicates that the regularity conditions required by Lemma A.4 hold for $\mathsf{f} \overset{\mathsf{D}}{=} \mathsf{N}(0, 1)$.

To guarantee the power performance of the outlier distortion test under alternatives, our first recommendation is to use $\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ as suggested in Corollary 2.3 rather than $\mathsf{avar}(\widehat{\beta}_c^{(m+1)}) - \mathsf{avar}(\widetilde{\beta})$ when constructing the test statistics. It is because although two are equal under the null of no outliers, this would not be the case under alternatives. Furthermore, we need to estimate $\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ in order to conduct the test. Given the chosen iteration step m, the cut-off value c, and the reference distribution f, the terms τ_2^c, $\varrho_{\beta,c}^{(m+1)}$, $\varrho_{x\varepsilon,c}^{(m+1)}$ appearing in $\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ are known, so parameters that need to be estimated are σ^2, Σ. Their estimators should be consistent under the null and robust under alternatives, thus our second recommendation is of estimating σ^2, Σ using the clean data with all outliers removed, since the full sample estimators are inconsistent under alternatives though efficient under the null. Given any chosen $m \in [0, \infty)$ and $c \in [c_+, \infty)$, we can consistently estimate σ^2 and $\Sigma = \mathsf{E} x_i x_i'$ under the null using the subsample of all non-outlying observations by

$$(\widehat{\sigma}_c^{(m+1)})^2 = \varsigma_c^{-2} (\sum_{i=1}^n v_{i,c}^{(m)})^{-1} \{\sum_{i=1}^n (y_i - x_i' \widehat{\beta}_c^{(m+1)})^2 v_{i,c}^{(m)}\},$$
$$\widehat{\Sigma}_c^{(m+1)} = (\sum_{i=1}^n v_{i,c}^{(m)})^{-1} (\sum_{i=1}^n x_i x_i' v_{i,c}^{(m)}).$$

Thus, our suggested estimator of $\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ is given by

$$\begin{aligned} &\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) \\ &= \{(\varrho_{\beta,c}^{(m+1)} - 1)^2 + 2\tau_2^c (\varrho_{\beta,c}^{(m+1)} - 1) \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c (\varrho_{x\varepsilon,c}^{(m+1)})^2\} (\widehat{\sigma}_c^{(m+1)})^2 (\widehat{\Sigma}_c^{(m+1)})^{-1}. \end{aligned} \quad (2.12)$$

Otherwise if we use $\mathsf{avar}(\widehat{\beta}_c^{(m+1)}) - \mathsf{avar}(\widetilde{\beta})$ to construct the testing statistics and estimate σ^2 and Σ in $\mathsf{avar}(\widetilde{\beta})$ using the full sample, then under alternatives the test would lose power and lead to incorrect results. More seriously, it is very likely under alternatives to have $\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(m+1)}) \leq \widehat{\mathsf{avar}}(\widetilde{\beta})$ such that $\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(m+1)}) - \widehat{\mathsf{avar}}(\widetilde{\beta}) \leq 0$, thus the Hausman type testing statistics is negative so that the test is meaningless in this case.

We can now fully establish the outlier distortion test for formalising the robustness checks. For any chosen $m \in [0, \infty)$ and $c \in [c_+, \infty)$, we have the testing statistics and its limiting distribution

$$n^{1/2} (\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) \overset{a}{\sim} \mathsf{N}\{0_{d_x}, \widehat{\mathsf{avar}}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})\},$$

and

$$\widehat{H}_{n,c}^{(m+1)} = n (\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})' \widehat{\mathsf{avar}}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})^{-1} (\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) \overset{a}{\sim} \chi_{d_x}^2.$$

Note that $\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ shown in (2.12) is invertible and its inverse is given by

$$\begin{aligned} &\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})^{-1} \\ &= \{(\varrho_{\beta,c}^{(m+1)} - 1)^2 + 2\tau_2^c (\varrho_{\beta,c}^{(m+1)} - 1) \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c (\varrho_{x\varepsilon,c}^{(m+1)})^2\}^{-1} (\widehat{\sigma}_c^{(m+1)})^{-2} \widehat{\Sigma}_c^{(m+1)}, \end{aligned}$$

since $(\varrho_{\beta,c}^{(m+1)} - 1)^2 + 2\tau_2^c (\varrho_{\beta,c}^{(m+1)} - 1) \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c (\varrho_{x\varepsilon,c}^{(m+1)})^2 \neq 0$. We therefore avoid the rank deficiency problem described and addressed by the generalised inverse in Hausman & Taylor (1981) and Holly (1982). The proposed test can either be performed as the two-sided test with the normal limit or as the one-sided test with the chi-squared limit. On implementing RLS or IIS, we are particularly interested in two specific estimators: the trimmed estimator just updated from the full sample OLS and the fixed point estimator iterated upon through infinite steps. The below corollary then provides the outlier distortion tests for these two special cases, checking whether $\widetilde{\beta}$ is distinct from $\widehat{\beta}_c^{(1)}$ when $m = 0$ or from $\widehat{\beta}_c^{(*)}$ when $m \to \infty$.

Corollary 2.4. *Consider RLS or IIS. Suppose Assumption 2.1(i, ii) holds. For any $c \in [c_+, \infty)$ and large n, then when $m = 0$ we have*

$$n^{1/2} (\widehat{\beta}_c^{(1)} - \widetilde{\beta}) \overset{a}{\sim} \mathsf{N}\{0_{d_x}, \widehat{\mathsf{avar}}(\widehat{\beta}_c^{(1)} - \widetilde{\beta})\},$$

and

$$\widehat{H}_{n,c}^{(1)} = n(\widehat{\beta}_c^{(1)} - \widetilde{\beta})' \widehat{\mathsf{avar}}(\widehat{\beta}_c^{(1)} - \widetilde{\beta})^{-1} (\widehat{\beta}_c^{(1)} - \widetilde{\beta}) \overset{a}{\sim} \chi^2_{d_x},$$

where

$$\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(1)} - \widetilde{\beta}) = \frac{\{2c\mathsf{f}(c) - \psi_c\}^2 + 2\tau_2^c \{2c\mathsf{f}(c) - \psi_c\} + \tau_2^c}{\psi_c^2} (\widehat{\sigma}_c^{(1)})^2 (\widehat{\Sigma}_c^{(1)})^{-1}.$$

In addition, when $m \to \infty$ we have

$$n^{1/2}(\widehat{\beta}_c^{(*)} - \widetilde{\beta}) \overset{a}{\sim} \mathsf{N}\{0_{d_x}, \widehat{\mathsf{avar}}(\widehat{\beta}_c^{(*)} - \widetilde{\beta})\},$$

and

$$\widehat{H}_{n,c}^{(*)} = n(\widehat{\beta}_c^{(*)} - \widetilde{\beta})' \widehat{\mathsf{avar}}(\widehat{\beta}_c^{(*)} - \widetilde{\beta})^{-1} (\widehat{\beta}_c^{(*)} - \widetilde{\beta}) \overset{a}{\sim} \chi^2_{d_x},$$

where

$$\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(*)} - \widetilde{\beta}) = \frac{\{2c\mathsf{f}(c) - \psi_c\}^2 + 2\tau_2^c \{2c\mathsf{f}(c) - \psi_c\} + \tau_2^c}{\{2c\mathsf{f}(c) - \psi_c\}^2} (\widehat{\sigma}_c^{(*)})^2 (\widehat{\Sigma}_c^{(*)})^{-1}.$$ [4]

Heuristic tests. Without the asymptotic theory, some empirical researchers simply use the similar form of the proposed testing statistics but incorrectly replace the standard error of the difference between two estimators $\widetilde{\beta}$ and $\widehat{\beta}_c^{(m+1)}$ by the standard error of the marginal distribution of the baseline estimator $\widetilde{\beta}$. They subsequently compare it with the critical value either drawn from the standard normal for the two-sided test or from the chi-square with d_x degree of freedom for the one-sided test. The above procedure is referred to as the heuristic method for the outlier robustness checks. Since the heuristic method is constructed in a mistaken way, its size does not converge to the nominal level under the null and it has a low statistical power under alternatives. Thus, even the heuristic method is informative for preliminary analysis, it is not consistent in theory and not reliable in practice.

Notice that the density f of the errors ε_i enters in the asymptotic variance of the difference between RLS/IIS and OLS, thus our proposed test needs to assume its distributional form. It is a regularity condition in the outlier detection literature and follows from the ϵ-contamination idea initially proposed by Huber (1964) and recently re-investigated by Johansen & Nielsen (2016a). Under the Huber's framework, data come from $(1-\epsilon)\mathsf{f} + \epsilon\mathsf{f}^c$, so ϵ proportion of outliers, generated by an arbitrary (possibly fatter tailed) contaminated distribution f^c, have to be defined relative to a reference distribution f for $1-\epsilon$ proportion of sample. Furthermore, many empirical studies implicitly assume a form of the density f to select the cut-off value c for conducting outlier robustness checks. For example, Acemoglu et al. (2019) choose $c = 1.96$ by implicitly imposing the standard normal for the density f to compute the RLS estimator when assessing the impact of democratisation on economic growth.[5] Thus, the distributional assumption on f is not necessarily restrictive in our context, and it can be testable by a specification test proposed by Berenguer-Rico & Nielsen (2018) and Jiao & Pretis (2020).

Unlike the proposed asymptotic test, the heuristic test does not depend on the form of f. This is because it wrongly replaces the asymptotic variance of the difference between

[4] If assume $\mathsf{f} \overset{\mathsf{D}}{=} \mathsf{N}(0,1)$, then $\tau_2^c = \psi_c - 2c\mathsf{f}(c)$ so that $\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(1)} - \widetilde{\beta})$ and $\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(*)} - \widetilde{\beta})$ can be further simplified.

[5] In addition, other IIS and climate applications mentioned in the introduction (from wages to hurricane damages) all implicitly assume a normal reference distribution of the error term.

RLS/IIS and OLS by the asymptotic variance of OLS. Thus, it is meaningless to apply this inconsistent test except for preliminary analysis just to obtain an informative result. For our testing problem, we next move down the bootstrap route to research whether it is possible to robustify the distributional assumption on f.

Bootstrap tests. Despite the parametric distribution of the error term being a common assumption, to alleviate concerns about performance under an incorrect reference distribution and to improve finite sample performance, we propose and investigate three bootstrap versions of the outlier distortion test. This paper focuses on a non-parametric bootstrap of the L_1 and L_2 norms of the difference between OLS and RLS/IIS.

First, the same as Kaji (2018), we investigate the standard non-parametric bootstrap by randomly sampling observations i from the data $\{(y_i, x_i)\}_{i=1}^n$ with replacement to draw the distribution of the L_1 norm of the difference between $\widehat{\beta}_c^{(m+1)}$ and $\widetilde{\beta}$. Unlike the asymptotic test proposed in this paper, the bootstrap does not require assuming the known form of the density f, but it likely suffers from low power under alternatives where the data is in fact contaminated. This is because under alternatives the bootstrap has a certain probability to sample outlying observations which would have large influence on β estimation. For these bootstrapping iterations having sampled large outliers from the original data, the testing statistics on difference between $\widehat{\beta}_c^{(m+1)}$ and $\widetilde{\beta}$ would become rather large, so its bootstrapping distribution would be highly distorted and significantly distinct from what it should be under the null.

Second, we consider re-sampling from the outlier-removed ('cleaned') data. While this approach likely exhibits better power properties, there is a concern that it may be over-sized, as outliers are determined relative to the specified reference distribution, thus forcing the 'clean' data to appear closer to the reference distribution even under the null of no outlier distortion.

We therefore propose a third bootstrap approach where we scale the cut-off used to determine outlying observations. We first identify outlying observations at a cut-off c in the raw data, and subsequently re-sample from the cleaned data and create a bootstrap sample of our test statistic (or RLS/IIS) by using c' where $c' < c$. In other words, while the target reference distribution forces the cleaned data to be closer to the assumed distribution, by then relaxing the cut-off, some of the properties of the underlying data will likely be regained. The idea being that outliers identified using a lower threshold c' on the clean data, mimics the removal of outliers from the raw data at c.

We study the performance of the three bootstrap schemes together with the proposed asymptotic type test using a range of simulations.

3 Finite Sample Performance using Simulations

Here we study the performance of the proposed outlier distortion test in a series of simulation experiments under the null hypothesis of no distortion, as well as under a range of alternatives. We simulate (2.1) varying the sample size n, the number of regressors tested d_x, the degree of persistence in the dependent variable (ranging from iid to stochastically-trending), as well as the underlying reference distribution, and the degree of outlier contamination under a range of alternatives (the proportion of outliers as well as outlier magnitude). Simulations are implemented using the R-package `getspanel`, which is based on the R-package `gets`, with $m = 10,000$ replications for asymptotic tests and $m = 1,000$ replications and bootstrap draws for the bootstrap versions of the test.

3.1 Simulation Performance under the Null of No Distortion

Simulation results under the null of no outliers are shown in Figure 3.1. The asymptotic test appears slightly over-sized for small samples ($n < 200$) but exhibits size close to the nominal level in larger samples ($n > 200$). Size appears unaffected by the threshold used to detect outliers (left panel in 3.1), the number of coefficients tested for distortion (middle panel in Figure 3.1), or the degree of persistence of the dependent variable even if y follows a stochastic trend (right panel in Figure 3.1).

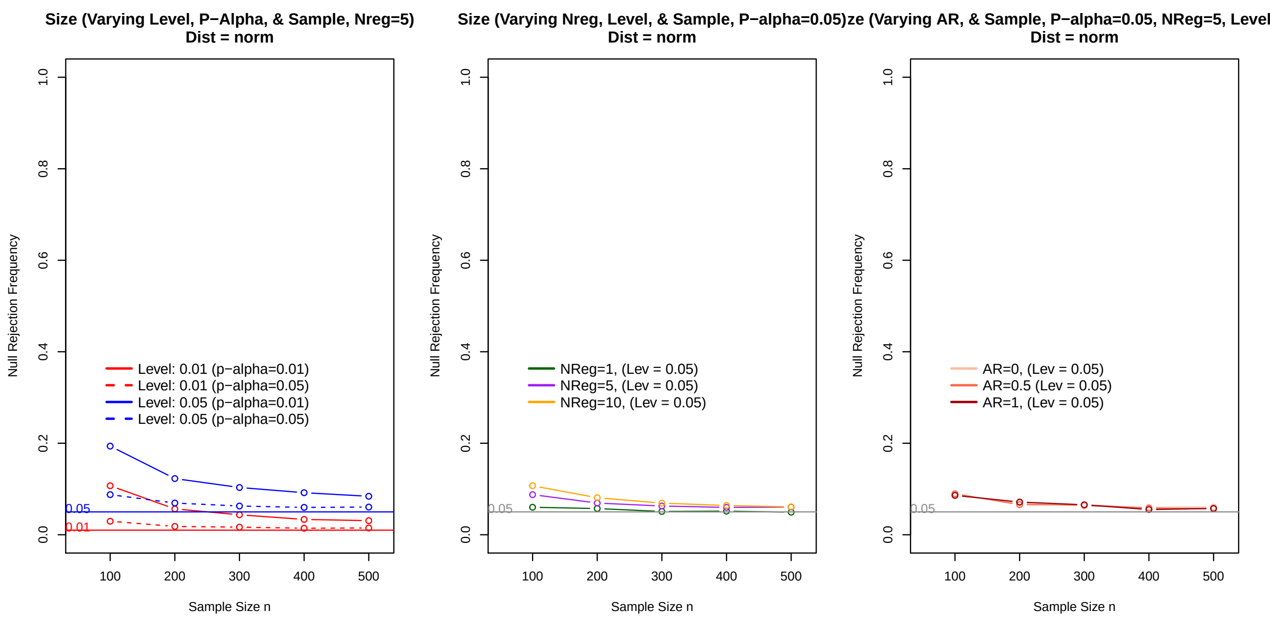

Figure 3.1: Simulation performance of the asymptotic test under the null of no distortion.

Bootstrap results are shown in Figure 3.2 using both the L_1 and L_2 norm of the difference between robust and OLS coefficient estimates. Here we also consider the performance of the bootstrap and asymptotic tests when the assumed reference distribution is incorrect. Specifically, we consider the case where we assume f to be a normal reference distribution, when the actual error distribution is fatter-tailed (a t-distribution with 3 degrees of freedom).

As expected, the bootstrap on the raw data is under-sized regardless of the reference distribution. In turn, the asymptotic test is over-sized when reference distribution is incorrect. Bootstrapping the clean (outlier-removed) data yields a size close to nominal levels for small samples (under a correct reference distribution), and also partly reduces the over-rejection when the reference distribution is incorrect. Our proposed bootstrap scaling approach improves the performance relative to the asymptotic test when the assumed reference distribution is to the true error distribution.

3.2 Simulation Performance in the Presence of Distortion

We assess the performance of the test in the presence of outlier distortion by introducing outliers of varying magnitudes to the DGP, varying the sample size and degree of

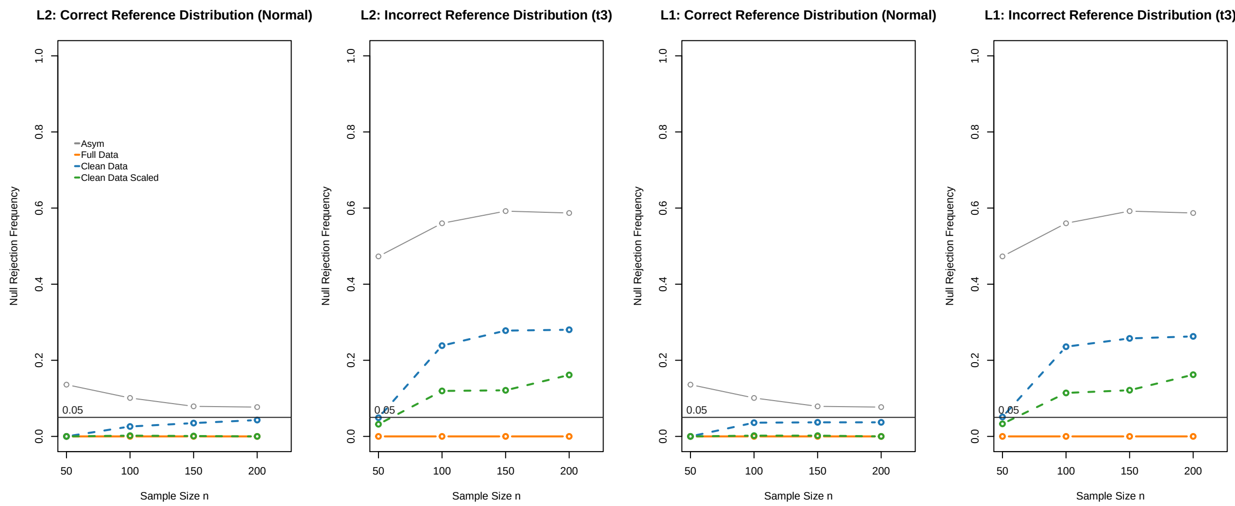

Figure 3.2: Simulation performance of the bootstrap tests under the null of no distortion when the reference distribution does and does not match the error distribution in the DGP.

autocorrelation in the dependent variable. We plot the null-rejection frequencies for different simulation specifications in Figures 3.3 and 3.4 for iid and stochastically trending dependent variables when the sample contains 10% of contaminated observations. To capture the degree of distortion we plot the null rejection frequency against the euclidian distance of all estimated (OLS) coefficients from their true underlying DGP counterparts, scaled by the maximum distance. As the results show, regardless of whether y is iid or stochastically-trending, the power of the test increases with the degree of outlier distortion. Consistent with our expectations, power also increases with sample size. Additional simulation results with varying proportions of outlier contamination and varying degrees of persistence in y are provided in the appendix.

Bootstrap results are shown in Figures 3.5 and 3.6 for different outlier magnitudes. As expected, bootstrapping using the raw data appears to have zero power as each bootstrap draw has a chance of including some outlying observations, thus rendering the resulting bootstrap distribution of the L1 or L2 norms uninformative when compared to the original test statistic. In turn, bootstrapping the clean (outlier-removed) data, or the clean data with re-scaled threshold shows desirable power properties: power increases with the sample size as well as with the degree of outlier distortion, even when the reference distribution is non-normal.

3.3 Summary of Simulation Results

The simulation results show well-controlled size and high power of our asymptotic test in large samples (n $>$ 200), with the tests being slightly oversized in small samples (n $<$ 200). In line with our theory results, the performance of the test is unaffected by the presence of stochastic trends. Bootstrapping the cleaned (outlier-removed) data, or bootstrapping with scaled significance levels on the cleaned data performs well under the null of no outliers, and reduces size-distortions of the asymptotic test in small samples and when the reference distribution is different to the assumed distribution. These bootstrap approaches also exhibit good power under a range of alternatives. We do not recommend bootstrapping the raw data, as expected, re-sampling the raw data has close

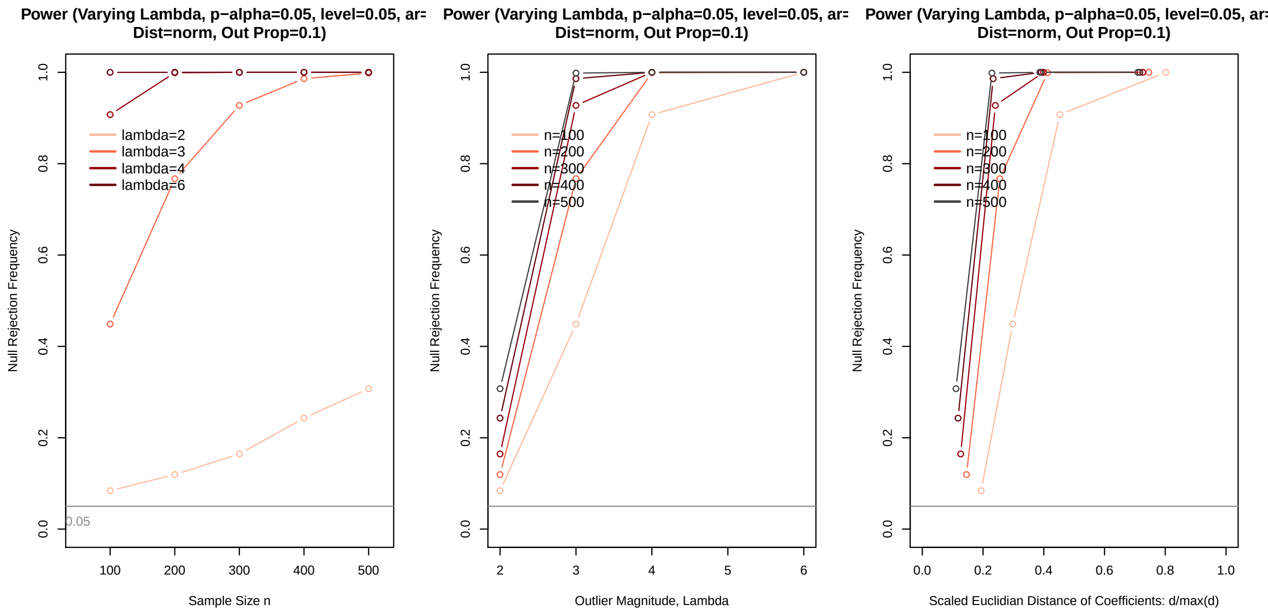

Figure 3.3: Simulation performance under the alternative for varying sample sizes and outlier magnitudes when the dependent variable exhibits no autocorrelation and 10% of the sample is outlier-contaminated.

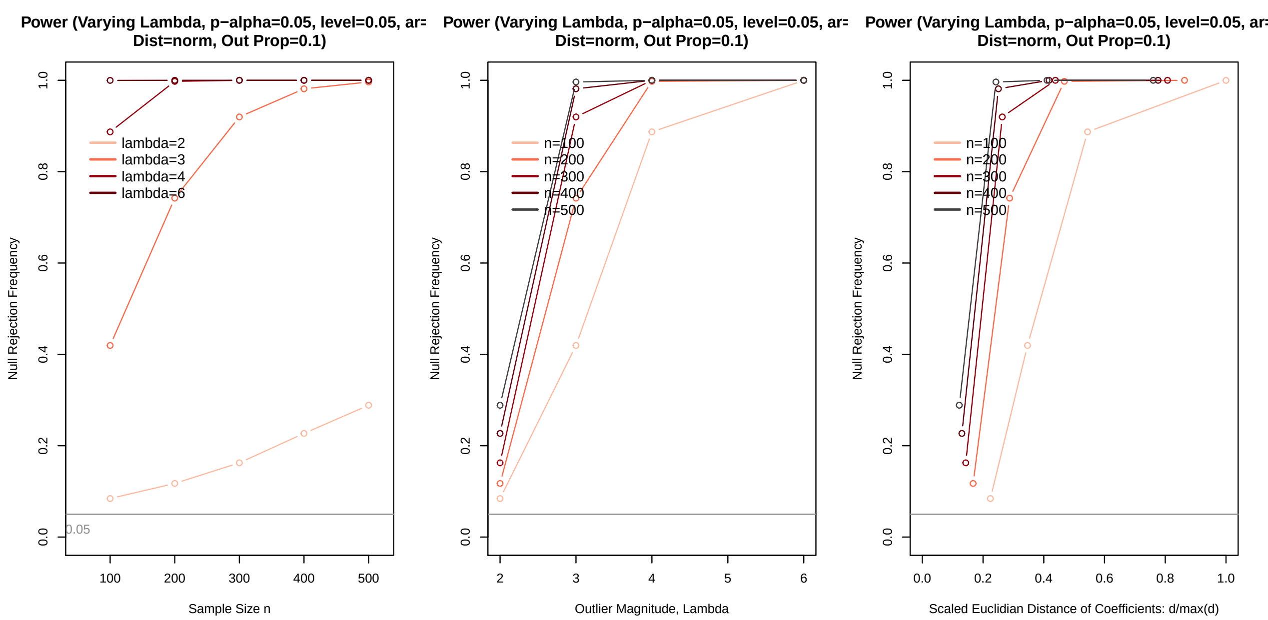

Figure 3.4: Simulation performance under the alternative for varying sample sizes and outlier magnitudes when the dependent variable is stochastically trending and 10% of the sample is outlier-contaminated.

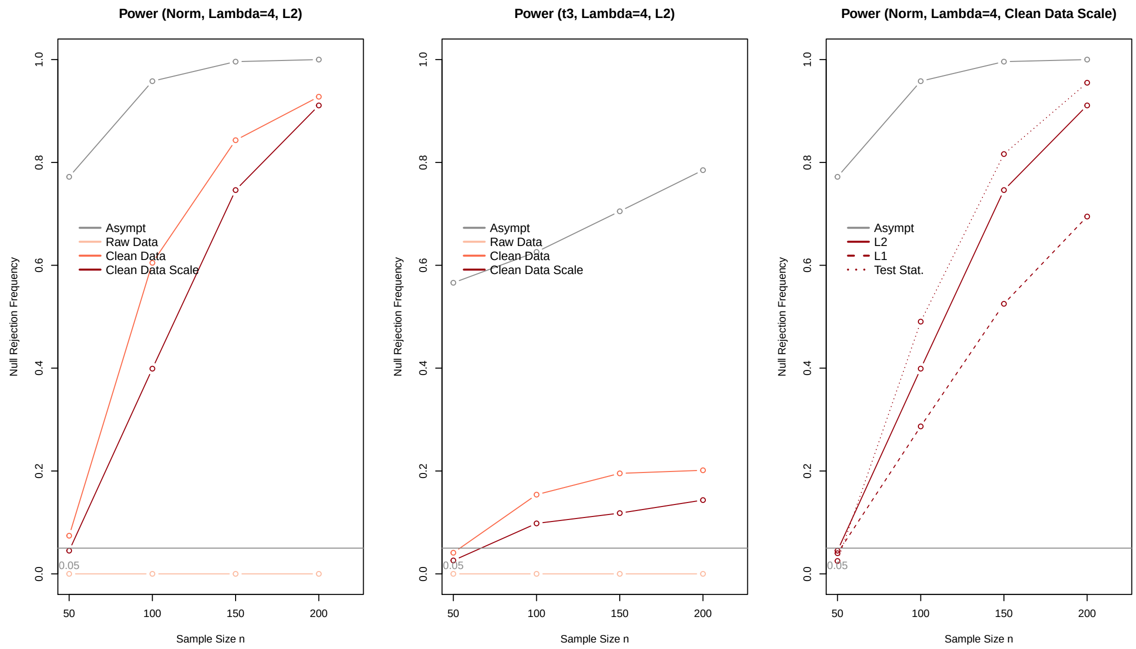

Figure 3.5: Simulation performance under the alternative for varying sample sizes when using the bootstrap implementations of our test (for outlier magnitude of 4SD of the error term).

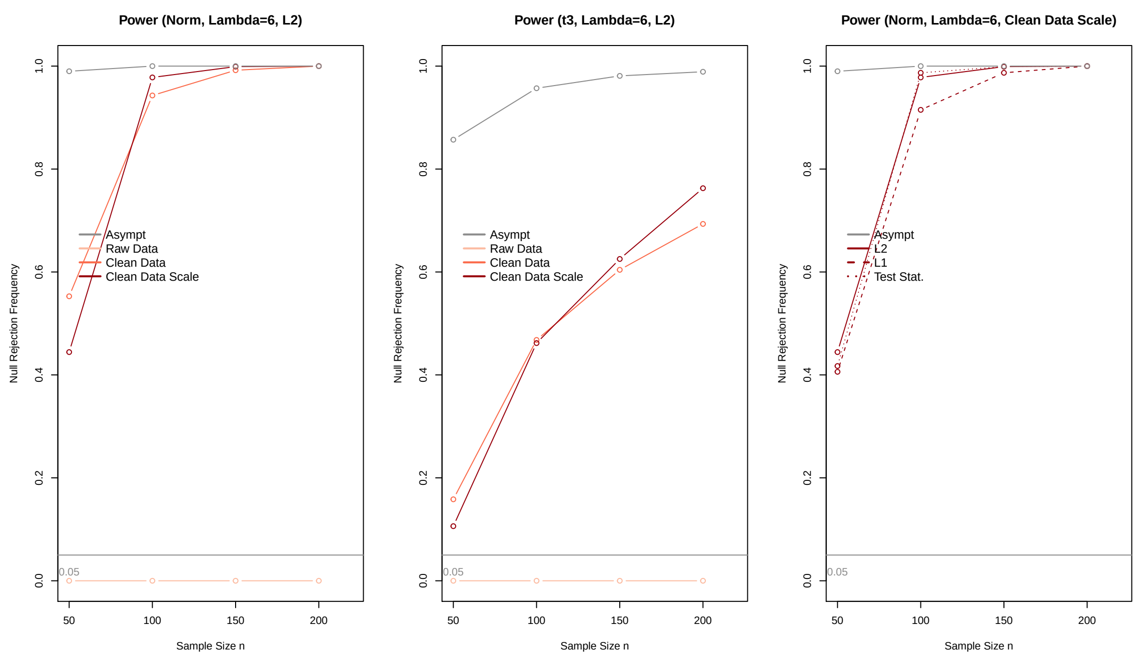

Figure 3.6: Simulation performance under the alternative for varying sample sizes when using the bootstrap implementations of our test (for outlier magnitude of 6SD of the error term).

to no power to detect distortion.

4 Application: Climate Adaptation & the Macro-Economic Impacts of Temperatures

We apply the outlier distortion test to estimates of the economic impacts of climate change allowing for income-driven adaptation. Specifically, we use the robust IIS estimator to estimate a panel model of the effects of temperature on GDP per capita growth using a global panel dataset in line with the existing climate-econometric impacts literature (e.g. Burke et al., 2015; Dell et al., 2012; Pretis, Schwarz, et al., 2018). While existing estimates strongly support a non-linear relationship between GDP per capita growth and temperatures, adaptation to climate change has been notably absent in empirical macro-economic cross-country analyses. Higher incomes might mitigate the impacts of climate change (see e.g. Acevedo et al., 2020; Schwarz & Pretis, 2020), and such adaptation might bias existing historical estimates and subsequent future projections of macro-economic climate impacts. We therefore estimate a panel model of climate impacts while allowing for adaptation driven by incomes, specified through interaction terms of temperatures and (lagged) log GDP per capita.

A concern with global panel-econometric models of climate impacts is the inability to account for all observed variation in economic growth. Even though conventional panel climate impact models control for time fixed effects, country fixed effects, and country-specific (non-linear) time trends, there are potentially numerous un-modelled idiosyncratic shocks. Such shocks – taking the form of outliers in the model – can erroneously be attributed to climate variation and thus distort the estimated coefficients on climate variables. To account for the presence of such un-modelled shocks, we estimate our panel climate impacts model with adaptation using both OLS and the robust IIS estimator and subsequently test whether there is evidence of coefficient distortion using our proposed outlier distortion test.

4.1 Data & Model

Following the macro-econometric literature studying the temperature effects on economic growth, we model the change in the log of real GDP per capita (World Development Indicators, World Bank, 2019) from 1961-2017 for 169 countries as a function of population-weighted climate variation. We obtain historical climate observations on temperatures and precipitation from (Matsuura & Willmott, 2018, version 5) covering global land areas at a $0.5°$ spatial resolution. To map gridded climate observations to individual countries in each year, we weight climate observations by gridded population data (CIESIN, 2016) at the same spatial resolution. Population-weighting (rather than area-weighting) is more likely to capture the effects of climate onto socio-economic activity (see Tol, 2017). To project climate impacts to the end of the century, we construct a dataset of country-level changes in temperature based on CMIP5 climate models (Taylor et al., 2012) and use long-term country-level forecasts of GDP per capita for our baseline (Müller et al., 2019).

Modelling GDP per capita growth in fixed effects panels as a non-linear (quadratic) function of annual-average temperatures has become common practise in the climate econometric literature. We expand on these existing models by allowing for climate adaptation at different income levels. Specifically, our model allows for adaptation by potentially attenuating (or amplifying) the coefficients on temperatures through interaction terms with (lagged) log of GDP per capita. Dell et al. (2012) estimate a simplified version of such a specification by interacting a linear temperature variable with a dummy variable for poor countries, and Burke et al. (2015) explore the interaction of a linear temperature variable with country-level incomes. Here we generalise this to non-linear temperature impacts interacted with the continuum of per capita incomes. A similar approach of estimating income-based adaptation has also been applied for heat-related mortality in Carleton et al. (2020) and Schwarz & Pretis (2020) for climate impacts of extremes. Our base model is given in equation (4.1):

$$\Delta log(y)_{i,t} = \alpha_i + \lambda_t + \beta_1 T_{i,t} + \beta_2 T_{i,t}^2 + \beta_3 \left[T_{i,t} \times log(y)_{i,t-1}\right] + \beta_4 \left[T_{i,t}^2 \times log(y)_{i,t-1}\right] + x'_{i,t}\gamma + u_{i,t} \tag{4.1}$$

where $\Delta log(y)_{i,t}$ is the year-on-year change in per capita income in country i in year t. The coefficients of interest are β_1 through β_4. specifically β_1 and β_2 capture the (potentially non-linear) impacts on GDP per capita growth and coefficients β_3, β_4 on interactions terms allow for the temperature-growth relationship to vary by country incomes. The vector $x_{i,t}$ denotes a set of control variables including population-weighted annual average precipitation (and its square), and country-specific linear and quadratic trends as in Burke et al. (2015) and Pretis, Schwarz, et al. (2018). To assess the impact of unknown idiosyncratic shocks, we estimate (4.1) using OLS (as is convention in the literature) and then compare the OLS estimates to those obtained when estimating (4.1) using the robust IIS estimator. Let $\tilde{\beta}$ denote the vector of coefficients estimated by OLS $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3, \tilde{\beta}_4)$, and $\hat{\beta}_c^{(1)}$ denote the vector of coefficients estimated using the robust IIS estimator at $\gamma_{c=2.57} = 0.01$ relative to a normal reference distribution f. The hypothesis of interest is $H_0 : \tilde{\beta} = \hat{\beta}_c^{(1)}$. In other words, we assess possible outlier distortion in the coefficient estimates.

4.2 Results

The estimation results (Figures 4.1, 4.2, and Table 4.1) show significant evidence of income-driven adaptation to temperatures, however, the estimates are distorted by the presence of outlying observations. Richer countries exhibit a different response function to temperatures compared to poor and middle-income countries, but the shape of the estimated effects overall shows a dampened temperature impact for the robust estimator compared to OLS. Figure 4.1 shows the difference in estimated coefficients between the robust IIS approach and OLS. Specifically, the robust coefficient estimates are attenuated relative to the OLS estimates. The difference between the OLS and robust IIS estimates is statistically different from zero when applying our proposed test for outlier distortion. Formally, we construct our test statistic for outlier distortion in the four coefficients of interest as:

$$H^1_{(n,c)} = n\left(\hat{\beta}_c^{(1)} - \tilde{\beta}\right)' \widehat{\text{avar}}\left(\hat{\beta}_c^{(1)} - \tilde{\beta}\right)^{-1}\left(\hat{\beta}_c^{(1)} - \tilde{\beta}\right) \tag{4.2}$$

where $c = 1.96$ for a normal reference distribution and a sample size of $n = 7007$. The difference in coefficients between the OLS and robust estimator shown in Table 4.1 (right column). We compare the resulting test statistic $H^1_{(n,c)} = 554.98$ against the critical values of the χ^2 distribution with $df = 4$, as we are testing for the distortion of four coefficients. The null hypothesis of no distortion is rejected at any conventional significance level, with $p < 0.0001$.

We show the estimated non-linear relationship between temperatures and growth for three different income levels (25^{th}, 50^{th}, and 75^{th} income percentiles) in Figure 4.2. Results across income levels show that controlling for outlying observations dampens the impacts of temperatures onto economic growth. Using robust estimates, the projected negative temperature impacts are less steep (but still negative) for low income countries, and dampened for middle income countries. For rich countries, the positive effects of higher temperatures are drastically weakened, suggesting attenuated positive effects.

Beyond distortion in coefficients, we show the distribution of the detected outlying observations in the panel model over time, countries, and both (Figures 4.3 and 4.4). The number of detected outliers greatly exceeds the number expected by chance. In total, 150 country-year observations identified as outlying compared to 70.7 expected under the null when there are no outliers. The observed proportion of outliers $\hat{\gamma}_{c=2.57} = 150/7007{=}0.021$ is statistically different from the expected proportion of $\gamma_{c=2.57} = 0.01$ ($p < 0.001$ using the Jiao & Pretis (2020) proportion test comparing 0.021 to 0.01). Further, the distribution of outliers does not appear to be random over time and space – outliers are concentrated in developing countries and there is a temporal cluster coinciding with the collapse of the Soviet Union and the onset of the Gulf War, which may well capture an idiosyncratic shock not modelled in the base panel model.

Overall, our results provide macro-econometric evidence of income-driven adaptation – temperatures have a non-linear effect on GDP per capita growth, but this relationship varies by income, where poor countries face negative impacts of warmer temperatures while rich countries face predominantly positive impacts, with middle-income countries seeing the familiar quadratic relationship. Moreover, our estimation results highlight the importance of controlling for outlying observations and testing for subsequent distortion. Idiosyncratic shocks in the model distort coefficient estimates, with robust results showing dampened climate impacts onto GDP per capita growth compared to conventional OLS estimates.

Note: Can we show that our residuals produced by IIS are iid normal or at least satisfy either normal, homoscedasticity, or serially independent? Any one of these will help us to defend our paper if referees criticise on our assumptions. In addition, in this case we have a very well specified climate impact model that can pass all the specification tests.

Table 4.1: OLS and IIS Panel Regression Results together with their difference in coefficients and the resulting outlier distortion test statistic. Coefficients on control variables are omitted.

	Base	Base IIS	Base Outlier Distortion Test	Adaptation	Adaptation IIS	Adaptation Outlier Distortion Test
Temperature	0.01734***	0.01032***	108.95397	-0.06224***	-0.03384***	186.25252
	(0.00348)	(0.00233)	[>0.001]	(0.01041)	(0.0072)	[>0.001]
Temperature^2	-0.00059***	-0.00039***	99.04793	0.00070	0.00001	88.57293
	(0.0001)	(0.00007)	[>0.001]	(0.00037)	(0.00026)	[>0.001]
Precipitation	0.00043	0.00073	1.97683	0.01018	0.01328***	7.86435
	(0.00111)	(0.00074)	[0.15973]	(0.00563)	(0.00383)	[0.00504]
Precipitation^2	-0.00004	-0.00004	0.46262	-0.00019	-0.00035**	17.55037
	(0.00003)	(0.00002)	[0.4964]	(0.00019)	(0.00013)	[0.00003]
Temperature x GDP_{pc}				0.00811***	0.00416***	317.58792
				(0.00111)	(0.00077)	[>0.001]
Temperature^2 x GDP_{pc}				-0.00012**	-0.00002	148.16829
				(0.00004)	(0.00003)	[>0.001]
Precipitation x GDP_{pc}				-0.00121	-0.00155***	6.86436
				(0.00066)	(0.00045)	[0.00879]
Precipitation^2 x GDP_{pc}				0.00002	0.00004*	17.52154
				(0.00002)	(0.00002)	[0.00003]
Num. Outliers			165			170
Outlier Distortion test statistic for Temp. Variables			$\chi^2_2 = 111.27$			$\chi^2_4 = 770.69$
Outlier Distortion p-value for Temp. Variables			>0.001			>0.001
Num.Obs.	7716	7716		7716	7716	
BIC	-18483.2	-23533.1		-18801.4	-23776.0	
Log.Lik.	11774.742	15038.149		11951.758	15199.897	
Fixed Effects	Country & Year	Country & Year		Country & Year	Country & Year	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
(Standard Errors) and [p-values]

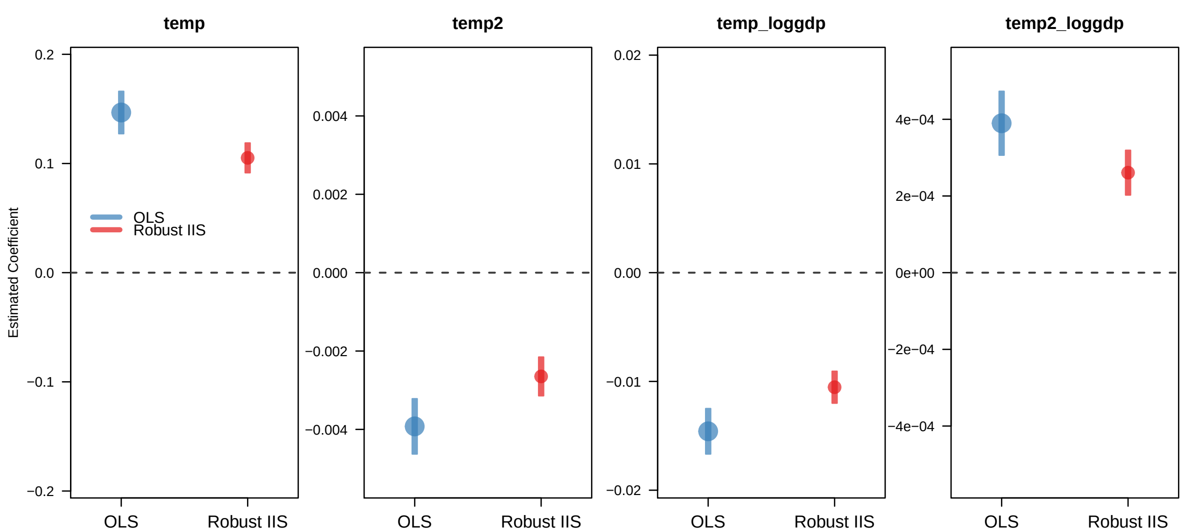

Figure 4.1: Estimated Impact of Temperatures on GDP Per Capita Growth Allowing for Adaptation and Controlling for Outliers. Coefficients on temperatures and the income interaction terms are shown using OLS and the robust IIS estimator. Note that scale of the y-axis differs across plots.

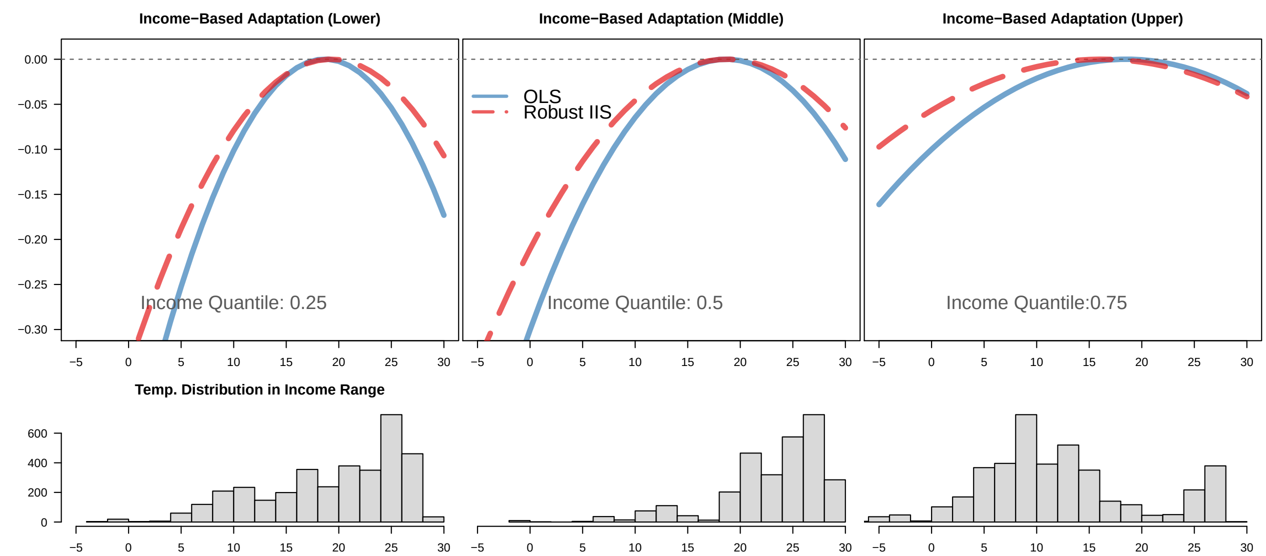

Figure 4.2: Estimated Impact of Temperatures on GDP Per Capita Growth Allowing for Adaptation and Controlling for Outliers. OLS-estimated relationship shown in blue, robust IIS estimated relationship shown in red. Estimated non-linear impact function for different three different income levels (middle panels) and the observed country level temperatures (bottom panels): countries in at the 25th income percentile, countries with median income, and countries at the 75th income percentile. Observed temperatures across income ranges are shown in the lower panel.

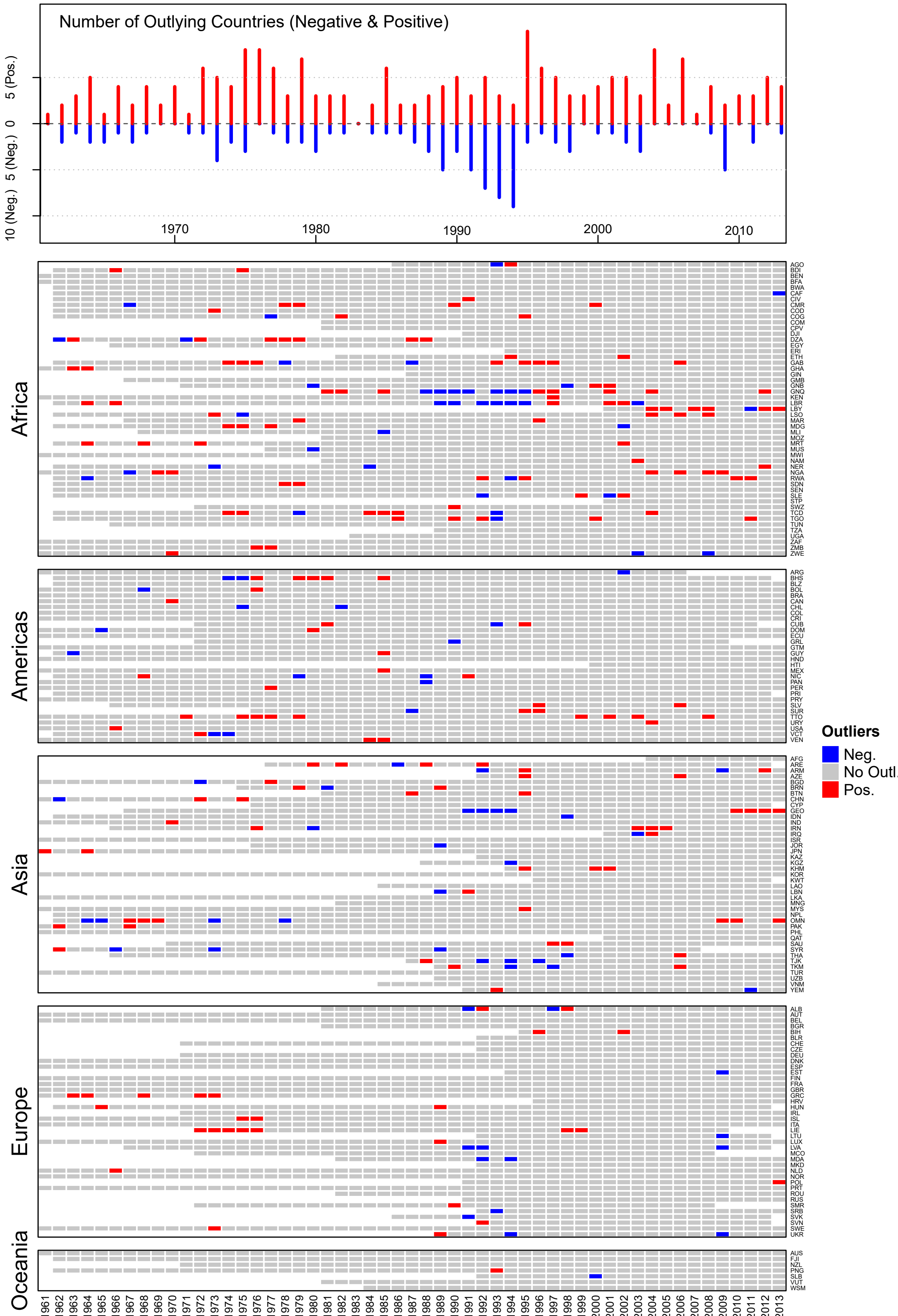

Figure 4.3: Detected Outliers using the robust IIS estimator across countries and time in the global cross-country panel from 1962-2013. Top Figure panel shows the number of positive (red) and negative (blue) outliers aggregated over countries in each year. Bottom panel shows country-year observations as gray when not outlying, blue when there is a negative outliers, and red for positive outliers.

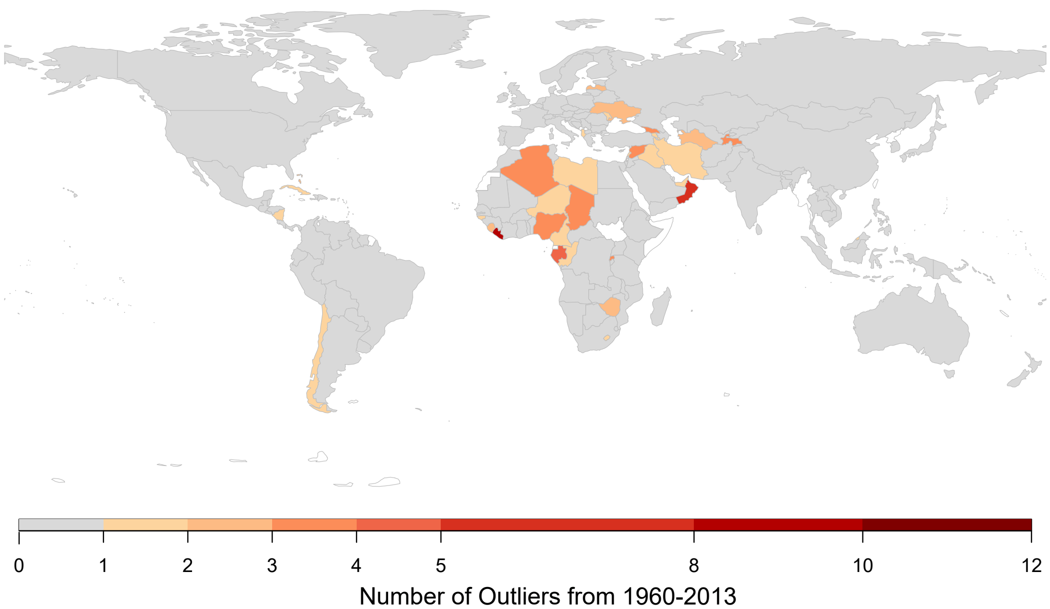

Figure 4.4: Detected outliers aggregated over the full sample by contry in the panel. Gray denotes no outliers detected.

5 Conclusion

We introduced a formal test to assess outlier distortion in regression models by comparing OLS estimates to those obtained using robust Huber-skip estimators (including Robustified Least Squares and Impulse Indicator Saturation). To develop the proposed test, we fully established asymptotic theory of RLS and IIS. Our analysis is valid in cross sectional and time series settings which include stationary, deterministically trending, and unit root processes. The test performs well in simulations with size close to nominal levels for large ($n > 200$) samples and high power under a range of alternatives. Our proposed bootstrap implementation of scaling the cut-off values used to determine outliers improves performance in finite samples and reduces size distortions for the asymptotic test when the reference distribution is different to the true underlying error distribution.

Our application of the outlier distortion test to the macro-economic impacts of climate change highlights the importance of assessing the influence of outliers in regression models. While we find evidence of income-driven adaptation to climate change, the estimates are sensitive to outlying observations. When using a robust estimator, the estimated impacts of temperatures onto GDP per capita growth are significantly different to those obtained when using OLS. Resulting projections of impacts under adaptation when using a robust estimator show that...

[**Note: some results here**].

More generally, our proposed test allows for the assessment of outlier-driven distortion in a wide set of regression models and can be readily implemented using the R-package ‘gets’ (Pretis, Reade, & Sucarrat 2018).

References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, *91*(5), 1369–1401.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2012). The colonial origins of comparative development: An empirical investigation: Reply. *American Economic Review*, *102*(6), 3077–3110.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, *127*(1), 47–100.

Acevedo, S., Mrkaic, M., Novta, N., Pugacheva, E., & Topalova, P. (2020). The effects of weather shocks on economic activity: what are the channels of impact? *Journal of Macroeconomics*, *65*, 103207.

Albouy, D. Y. (2012). The colonial origins of comparative development: an empirical investigation: comment. *American economic review*, *102*(6), 3059–76.

Anundsen, A. K. (2015). Econometric regime shifts and the us subprime bubble. *Journal of Applied Econometrics*, *30*(1), 145–169.

Auerbach, A. J., Hassett, K. A., & Oliner, S. D. (1994). Reassessing the social returns to equipment investment. *The Quarterly Journal of Economics*, *109*(3), 789–802.

Berenguer-Rico, V., Johansen, S., & Nielsen, B. (2019). The analysis of marked and weighted empirical processes of estimated residuals. *Working Paper*.

Berenguer-Rico, V., & Nielsen, B. (2018). Marked and weighted empirical processes of residuals with applications to robust regressions. *Working Paper*.

Berenguer-Rico, V., & Wilms, I. (2021). Heteroscedasticity testing after outlier removal. *Econometric Reviews*, *40*(1), 51–85.

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons.

Burke, M., Hsiang, S. M., & Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, *527*(7577), 235–239.

Carleton, T. A., Jina, A., Delgado, M. T., Greenstone, M., Houser, T., Hsiang, S. M., ... others (2020). *Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits* (Tech. Rep.). National Bureau of Economic Research.

Castle, J. L., & Hendry, D. F. (2009). The long-run determinants of uk wages, 1860–2004. *Journal of Macroeconomics*, *31*(1), 5–28.

CIESIN, F. (2016). Gridded population of the world, version 4 (gpwv4): population count grid. *Center for International Earth Science Information Network (CIESIN), Columbia University*.

Dehon, C., Gassner, M., & Verardi, V. (2012). Extending the hausman test to check for the presence of outliers. In *Essays in honor of jerry hausman.* Emerald Group Publishing Limited.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, *4*(3), 66–95.

De Long, J. B., & Summers, L. H. (1991). Equipment investment and economic growth. *The Quarterly Journal of Economics*, *106*(2), 445–502.

De Long, J. B., & Summers, L. H. (1994). Equipment investment and economic growth: reply. *The Quarterly Journal of Economics*, *109*(3), 803–807.

Dreger, C., & Wolters, J. (2014). Money demand and the role of monetary indicators in forecasting euro area inflation. *International Journal of Forecasting*, *30*(2), 303–312.

Durbin, J. (1954). Errors in variables. *Revue de l'institut International de Statistique*, 23–32.

Ericsson, N. R. (2017). How biased are us government forecasts of the federal debt? *International Journal of Forecasting*, *33*(2), 543–559.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: the approach based on influence functions.* New York: Wiley.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251–1271.

Hausman, J. A., & Taylor, W. E. (1981). A generalized specification test. *Economics Letters*, *8*(3), 239–245.

Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, *23*(2), 317–335.

Hendry, D. F., & Mizon, G. E. (2011). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, *3*(1).

Holly, A. (1982). A remark on hausman's specification test. *Econometrica: Journal of the Econometric Society*, 749–759.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Statistics*, *35*, 73–101.

Jiao, X. (2019). A simple robust procedure in instrumental variables regression. *Working Paper*.

Jiao, X., & Kurle, J. (2021). An asymptotic study of the false outlier detection rate in robust two stage least squares models. *Working Paper*.

Jiao, X., & Nielsen, B. (2015). Asymptotic analysis of iterated 1-step huber-skip m-estimators with varying cut-offs. In *Workshop on analytical methods in statistics* (pp. 23–52).

Jiao, X., & Pretis, F. (2020). Testing the presence of outliers in regression models. *Working Paper*.

Johansen, S., & Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. *Castle, and Shephard (2009)*, 1–36.

Johansen, S., & Nielsen, B. (2013). Outlier detection in regression using an iterated one-step approximation to the huber-skip estimator. *Econometrics*, *1*(1), 53–70.

Johansen, S., & Nielsen, B. (2016a). Analysis of the forward search using some new results for martingales and empirical processes. *Bernoulli*, *22*(2), 1131–1183.

Johansen, S., & Nielsen, B. (2016b). Asymptotic theory of outlier detection algorithms for linear time series regression models. *Scandinavian Journal of Statistics*, *43*(2), 321–348.

Johansen, S., & Nielsen, B. (2019). Boundedness of m-estimators for linear regression in time series. *Econometric Theory*, *35*(3), 653–683.

Kaji, T. (2018). Switching to the new norm: from heuristics to formal tests using integrable empirical processes. *Working Paper*.

Kalkuhl, M., & Wenz, L. (2020). The impact of climate conditions on economic production. evidence from a global panel of regions. *Journal of Environmental Economics and Management*, *103*, 102360.

Martinez, A. B. (2020). Improving normalized hurricane damages. *Nature Sustainability*, *3*(7), 517–518.

Matsuura, K., & Willmott, C. J. (2018). Terrestrial precipitation: 1900–2017 gridded monthly time series. *Electronic. Department of Geography, University of Delaware, Newark, DE*, *19716*.

Müller, U. K., Stock, J. H., & Watson, M. W. (2019). *An econometric model of international long-run growth dynamics* (Tech. Rep.). National Bureau of Economic Research.

Newell, R. G., Prest, B. C., & Sexton, S. E. (2021). The gdp-temperature relationship: implications for climate change damages. *Journal of Environmental Economics and Management*, *108*, 102445.

Nymoen, R., & Sparrman, V. (2015). Equilibrium unemployment dynamics in a panel of oecd countries. *Oxford Bulletin of Economics and Statistics*, *77*(2), 164–190.

Pretis, F., Mann, M. L., & Kaufmann, R. K. (2015). Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. *Climatic Change*, *131*(4), 705–718.

Pretis, F., Reade, J., & Sucarrat, G. (2018). Automated general-to-specific (gets) regression modeling and indicator saturation methods for the detection of outliers and structural breaks. *Journal of Statistical Software*, *86*(3).

Pretis, F., Schwarz, M., Tang, K., Haustein, K., & Allen, M. R. (2018). Uncertain impacts on economic growth when stabilizing global temperatures at 1.5 c or 2 c warming. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2119), 20160460.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association*, *79*(388), 871–880.

Ruppert, D., & Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, *75*(372), 828–838.

Schneider, L., Smerdon, J. E., Pretis, F., Hartl-Meier, C., & Esper, J. (2017). A new archive of large volcanic events over the past millennium derived from reconstructed summer temperatures. *Environmental Research Letters*, *12*(9), 094005.

Schwarz, M., & Pretis, F. (2020). Modelling historical adaptation rates to inform future adaptation pathways. In *Egu general assembly conference abstracts* (p. 21004).

Stillwagon, J. R. (2016). Non-linear exchange rate relationships: An automated model selection approach with indicator saturation. *The North American Journal of Economics and Finance*, *37*, 84–109.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of cmip5 and the experiment design. *Bulletin of the American meteorological Society*, *93*(4), 485–498.

Tol, R. S. (2017). Population and trends in the global mean temperature. *Atmósfera*, *30*(2), 121–135.

Varga, R. S. (2000). *Matrix iterative analysis*. Berlin: Springer.

Welsh, A. H., & Ronchetti, E. (2002). A journey in single steps: robust one-step m-estimation in linear regression. *Journal of Statistical Planning and Inference*, *103*(1-2), 287–310.

World Bank, W. (2019). World development indicators. *Washington, DC. http://wdi. worldbank. org/table/4.2*, *10*, 2001–2002.

Wu, D.-M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica: journal of the Econometric Society*, 733–750.

A Proofs of The Main Theorems

The proofs of the main results proceed as follows: initially the one-step stochastic expansion and tightness results are given for the iterated estimators of (β, σ^2) computed from Algorithm 2.1. Next, we show the expansion of the iterated estimators in any step in terms of the initial estimators and establish the fixed point of the iterated estimators upon through infinite iterations. Given the two different type of initial estimators by the Robustified Least Squares and Impulse Indicator Saturation algorithms, we then build up the stochastic expansions of their algorithmic estimators in any iterated step. All these arguments hold uniformly in the cut-off $c \in [c_+, \infty)$, so weak convergence theory can be established for the estimators of (β, σ^2) seen as stochastic processes in terms of the drifting cut-off c. The proof combines the tightness and finite dimensional convergence, showing that the weak limit varies with the stochastic properties of regressors. Finally, we derive the limiting distribution of a new Hausman type test statistics for checking outlier robustness in coefficients.

The updated estimators (2.6) and (2.7) of (β, σ^2) involve the weighted and marked empirical processes presented in the following lemma from Jiao & Pretis (2020), which is built on the empirical process theory recently developed by Berenguer-Rico et al. (2019) (Theorem 4.4). The lemma below is thus to give the first-order asymptotic expansions of the empirical processes appeared in the robust estimators (2.6) and (2.7).

Lemma A.1. *(Jiao & Pretis (2020), Lemma 8.2) Suppose Assumption 2.1(i, iib) holds. We have an expansion*

$$n^{-1/2}\sum_{i=1}^{n} w_{in}\varepsilon_i^p 1_{(|\varepsilon_i - x_{in}'b| \leq \sigma c + n^{-1/2}ac)} = n^{-1/2}\sum_{i=1}^{n} w_{in}\varepsilon_i^p 1_{(|\varepsilon_i| \leq \sigma c)} + \mathcal{B}_n(a,b,c) + R(a,b,c),$$

where the bias term is expressed as

$$\mathcal{B}_n(a,b,c) = 2\sigma^{p-1}c^p \mathsf{f}(c) n^{-1/2}\sum_{i=1}^{n} w_{in}(1_{(p\ even)} n^{-1/2}ac + 1_{(p\ odd)} x_{in}'b).$$

Notice w_{in} can be chosen as $1, n^{1/2}x_{in}, nx_{in}x_{in}'$ and p as either of $0, 1, 2$. For any $B > 0$ and as $n \to \infty$, the remainder term satisfies

$$\sup_{0<c<\infty}\ \sup_{|a|,|b| \leq n^{1/4-\eta}B} |R(a,b,c)| = \mathrm{o}_{\mathsf{P}}(1).$$

In addition, the normalized process $n^{-1/2}\sum_{i=1}^{n} w_{in}(\varepsilon_i^p 1_{(|\varepsilon_i| \leq \sigma c)} - \mathsf{E}_{i-1}\varepsilon_i^p 1_{(|\varepsilon_i| \leq \sigma c)})$ is tight in $c \in \mathbb{R}_+$, where $\mathsf{E}_{i-1}(\cdot) = \mathsf{E}(\cdot|\mathcal{F}_{i-1})$ and $\mathsf{E}_{i-1}\varepsilon_i^p 1_{(|\varepsilon_i| \leq \sigma c)} = \mathsf{E}\varepsilon_i^p 1_{(|\varepsilon_i| \leq \sigma c)} = \sigma^p \tau_p^c$.

Before showing the one-step stochastic expansion of the updated estimators, we first present the following lemma, which is a variation of the delta method required for attaining the expansion of $n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma)$ from $n^{1/2}\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\}$.

Lemma A.2. *Let $\{X_n\}$ be a sequence of random variables and θ be a deterministic parameter. Assume that a univariate function g has the first and second derivatives $\dot{g}, \ddot{g}$, then we have*

$$n^{1/2}\{g(X_n) - g(\theta)\} = \dot{g}(\theta)n^{1/2}(X_n - \theta) + n^{-1/2}\ddot{g}(\bar{\theta})\{n^{1/2}(X_n - \theta)\}^2,$$

where $|\bar{\theta} - \theta| \leq |X_n - \theta|$.

Proof of Lemma A.2. Approximate g around the point θ by the linear function using the Taylor expansion and particularly check the approximation at the point X_n, then

$$g(X_n) = g(\theta) + \dot{g}(\theta)(X_n - \theta) + \ddot{g}(\bar{\theta})(X_n - \theta)^2,$$

where $|\bar{\theta} - \theta| \leq |X_n - \theta|$. Rearranging the above immediately gives the expansion shown in the lemma. ■

Equipped with the empirical processes theory in Lemma A.1 and the delta method in Lemma A.2, we can now study the updated estimator (2.6) and (2.7) and build up its stochastic expansion in terms of the original estimator, a kernel and a small remainder term. Denote $c_+ > 0$ as a small positive number.

Lemma A.3. *Consider the iterated 1-step Huber-skip M-estimator in Algorithm 2.1. Suppose Assumption 2.1(i, ii) holds, and that $N^{-1}(\widehat{\beta}_c^{(m)} - \beta)$, $n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$ are $\mathrm{O}_{\mathsf{P}}(1)$. Then, uniformly in $c \in [c_+, \infty)$ and as $n \to \infty$*

$$N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = \frac{2c\mathsf{f}(c)}{\psi_c} N^{-1}(\widehat{\beta}_c^{(m)} - \beta) + (\psi_c \Sigma_n)^{-1} \sum_{i=1}^n x_{in}\varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)} + \mathrm{o}_{\mathsf{P}}(1),$$

$$n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = \frac{c(c^2 - \varsigma_c^2)\mathsf{f}(c)}{\tau_2^c} n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$$
$$+ \frac{\sigma}{2\tau_2^c} n^{-1/2} \sum_{i=1}^n (\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2) 1_{(|\varepsilon_i| \leq \sigma c)} + \mathrm{o}_{\mathsf{P}}(1).$$

Proof of Lemma A.3. The $m+1$ step estimators (2.6) and (2.7) of β, σ^2 are least squares estimators for the non-outlying observations and satisfy

$$N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = (\sum_{i=1}^n x_{in}x_{in}' v_{i,c}^{(m)})^{-1} (\sum_{i=1}^n x_{in}\varepsilon_i v_{i,c}^{(m)}), \tag{A.1}$$

$$n^{1/2}\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\} = \varsigma_c^{-2}(n^{-1}\sum_{i=1}^n v_{i,c}^{(m)})^{-1} n^{-1/2}\{\sigma^2 \sum_{i=1}^n (\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2) v_{i,c}^{(m)} \tag{A.2}$$
$$- (\sum_{i=1}^n \varepsilon_i x_{in}' v_{i,c}^{(m)})(\sum_{i=1}^n x_{in}x_{in}' v_{i,c}^{(m)})^{-1}(\sum_{i=1}^n x_{in}\varepsilon_i v_{i,c}^{(m)})\}.$$

We express the weight $v_{i,c}^{(m)}$ in (2.5) as

$$v_{i,c}^{(m)} = 1_{(|y_i - x_i'\widehat{\beta}_c^{(m)}| \leq \widehat{\sigma}_c^{(m)} c)} = 1_{(|\varepsilon_i - x_{in}'\widehat{b}_c^{(m)}| \leq \sigma c + n^{-1/2}\widehat{a}_c^{(m)} c)}, \tag{A.3}$$

where $\widehat{b}_c^{(m)} = N^{-1}(\widehat{\beta}_c^{(m)} - \beta)$ and $\widehat{a}_c^{(m)} = n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$ are estimation errors for β and σ in the m step of the algorithm.

By Assumption 2.1(i, iib) and $|\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathrm{O}_{\mathsf{P}}(1)$, we can apply the expansions of the empirical processes in Lemma A.1 to (A.1) and (A.2), so for $\widehat{\beta}_c^{(m+1)}$ we have

$$\widehat{b}_c^{(m+1)} = \frac{2c\mathsf{f}(c)}{\psi_c}\widehat{b}_c^{(m)} + (\psi_c \Sigma_n)^{-1} \sum_{i=1}^n x_{in}\varepsilon_i 1_{(|\varepsilon_i| \leq \sigma c)} + R_\beta(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

where the remainder $R_\beta(a,b,c)$ vanishes uniformly in $c_+ \leq c < \infty$ and $|a|,|b| \leq B$. A key to this is that c is bounded away from zero and that $\Sigma_n \overset{\mathsf{D}}{\to} \Sigma$ is almost surely positive definite by Assumption 2.1(iia) so that the denominator ψ_c, $\psi_c\Sigma_n$ is bounded away from zero.

For $\widehat{\sigma}_c^{(m+1)}$, first write $n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = n^{1/2}[\{(\widehat{\sigma}_c^{(m+1)})^2\}^{1/2} - (\sigma^2)^{1/2}]$ and let $g(x) = x^{1/2}$, $X_n = (\widehat{\sigma}_c^{(m+1)})^2$, $\theta = \sigma^2$, then apply Lemma A.2 to obtain

$$n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = \frac{1}{2\sigma} n^{1/2}\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\} + n^{-1/2}\mathrm{O}[n\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\}^2].$$

Notice $\dot{g}(x) = x^{-1/2}/2$ and $\ddot{g}(x) = -x^{-3/2}/4$. Next, apply the similar arguments as for $\widehat{\beta}_c^{(m+1)}$ to get

$$\widehat{a}_c^{(m+1)} = \frac{c(c^2 - \varsigma_c^2)\mathsf{f}(c)}{\tau_2^c}\widehat{a}_c^{(m)} + \frac{\sigma}{2\tau_2^c} n^{-1/2}\sum_{i=1}^n \left(\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2\right)1_{(|\varepsilon_i| \leq \sigma c)} + R_\sigma(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

where the remainder $R_\sigma(a,b,c)$ also vanishes uniformly. ■

The next step is to prove tightness of iterated estimators. Then, we show the contraction mapping for the one-step expansion of the updated estimator in terms of the original one. With $\eta = 1/4$ corresponding to a bounded initial estimators, this will be sufficient for tightness proof. Note that $|\cdot|$ refers to the usual Euclidean vector norm, while $\|M\| = \max\{\mathrm{eigen}(M'M)\}^{1/2}$ is the spectral norm for any matrix M. The norms are compatible so that $|Mx| \leq \|M\||x|$ for any vector x.

Proof of Theorem 2.1. To make the proof concise, write the one-step expansion as the compact form

$$\widehat{u}_c^{(m+1)} = \Gamma_c \widehat{u}_c^{(m)} + K_c + R_u(\widehat{u}_c^{(m)}, c), \tag{A.4}$$

where the remainder term satisfies $\sup_{c_+ \leq c < \infty} \sup_{|u| \leq B} |R_u(u,c)| = \mathrm{o}_{\mathsf{P}}(1)$ and

$$\widehat{u}_c^{(m)} = \begin{pmatrix} \widehat{b}_c^{(m)} \\ \widehat{a}_c^{(m)} \end{pmatrix}, \qquad \Gamma_c = \begin{Bmatrix} \frac{2c\mathsf{f}(c)}{\psi_c} I_{d_x} & 0_{d_x} \\ 0'_{d_x} & \frac{c(c^2 - \varsigma_c^2)\mathsf{f}(c)}{\tau_2^c} \end{Bmatrix}, \tag{A.5}$$

$$K_c = \begin{Bmatrix} (\psi_c \Sigma_n)^{-1} & 0_{d_x} \\ 0'_{d_x} & \frac{\sigma}{2\tau_2^c} \end{Bmatrix} \sum_{i=1}^n \begin{Bmatrix} x_{in}\varepsilon_i \\ n^{-1/2}(\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2) \end{Bmatrix} 1_{(|\varepsilon_i| \leq \sigma c)}. \tag{A.6}$$

Apply the autoregressive equation (A.4) recursively to obtain the representation

$$\widehat{u}_c^{(m+1)} = \Gamma_c^{m+1}\widehat{u}_c^{(0)} + \sum_{l=0}^m \Gamma_c^l \{K_c + R_u(\widehat{u}_c^{(m-l)}, c)\}. \tag{A.7}$$

Use the triangle inequality and $|Mx| \leq \|M\||x|$ to get

$$|\widehat{u}_c^{(m+1)}| \leq \|\Gamma_c^{m+1}\||\widehat{u}_c^{(0)}| + \{|K_c| + \max_{0 \leq l \leq m} |R_u(\widehat{u}_c^{(l)}, c)|\} \sum_{l=0}^m \|\Gamma_c^l\|.$$

Assumption 2.1(i) shows $\sup_{c_+ \leq c < \infty} \max\{|2c\mathsf{f}(c)/\psi_c|, |c(c^2 - \varsigma_c^2)\mathsf{f}(c)/\tau_2^c|\} < 1$; see Theorem 3.5 in Johansen & Nielsen (2013), so $\sup_{c_+ \leq c < \infty} \|\Gamma_c\| < 1$. Thus, by Gelfand's formula, see Theorem 3.4 in Varga (2000), $\lim_{m\to\infty}\|M^m\|^{1/m} = \max|\text{eigen}(M)|$, for some ω such that $\sup_{c_+\leq c<\infty}\|\Gamma_c\| < \omega < 1$ there exists $m_0 > 0$ so for all $m > m_0$ then

$$\sup_{c_+\leq c<\infty}\|\Gamma_c^m\| < \omega^m < 1. \tag{A.8}$$

Also note $(I_{d_x+1}-\Gamma_c)^{-1} = \sum_{l=0}^{\infty}\Gamma_c^l$. This in turn implies for some $1 < B_0 < \infty$

$$\sup_{0\leq m<\infty}\sup_{c_+\leq c<\infty}\|\Gamma_c^m\| < B_0, \qquad \sup_{c_+\leq c<\infty}\|(I_{d_x+1}-\Gamma_c)^{-1}\| \leq \sum_{l=0}^{\infty}\sup_{c_+\leq c<\infty}\|\Gamma_c^l\| < B_0. \tag{A.9}$$

Therefore, we have for all $m \in [0,\infty)$

$$|\widehat{u}_c^{(m+1)}| < B_0\{|\widehat{u}_c^{(0)}| + |K_c| + \max_{0\leq l\leq m}|R_u(\widehat{u}_c^{(l)}, c)|\}. \tag{A.10}$$

For any $c \in [c_+,\infty)$, Assumption 2.1(iii) with $\eta = 1/4$ guarantees tightness of $\widehat{u}_c^{(0)}$. Lemma A.1 shows that the kernel K_c process is tight using Assumption 2.1(i, iib). Thus, for all $\epsilon, \delta > 0$ there exist $n_0, U_0 > 0$ so that for all $n > n_0$ the set

$$\mathcal{A}_n = \{B_0\sup_{c_+\leq c<\infty}(|\widehat{u}_c^{(0)}| + |K_c|) \leq U_0/3, B_0\sup_{c_+\leq c<\infty}\sup_{|u|\leq U_0}|R_u(u,c)| < \delta/2\} \tag{A.11}$$

has probability larger than $1-\epsilon$.

Mathematical induction over m is used to show $\sup_{0\leq m<\infty}\sup_{c_+\leq c<\infty}|\widehat{u}_c^{(m)}| \leq U_0$ on the set $\mathcal{A}_n$. For $m = 0$ as induction starts, $\sup_{c_+\leq c<\infty}|\widehat{u}_c^{(0)}| \leq B_0^{-1}U_0/3 < U_0$ holds since $B_0 > 1$. The induction assumption is that $\sup_{0\leq l\leq m}\sup_{c_+\leq c<\infty}|\widehat{u}_c^{(l)}| \leq U_0$. This implies $B_0\max_{0\leq l\leq m}|R_u(\widehat{u}_c^{(l)}, c)| < \delta/2$, and then the bound in (A.10) becomes $\sup_{c_+\leq c<\infty}|\widehat{u}_c^{(m+1)}| < 2U_0/3 + \delta/2 < U_0$ so that $\sup_{0\leq l\leq m+1}\sup_{c_+\leq c<\infty}|\widehat{u}_c^{(l)}| \leq U_0$. ■

Next, we show the expansion of the iterated estimator from Algorithm 2.1 at any step in terms of its starting point.

Proof of Theorem 2.2. Directly apply the recursive representation (A.7) in the tightness proof, so uniformly in $c \in [c_+,\infty)$ and for any $m \in [0,\infty)$

$$\widehat{u}_c^{(m+1)} = \Gamma_c^{m+1}\widehat{u}_c^{(0)} + \sum_{l=0}^{m}\Gamma_c^l K_c + \sum_{l=0}^{m}\Gamma_c^l R_u(\widehat{u}_c^{(m-l)}, c),$$

where $\sup_{c_+\leq c<\infty}\sup_{|u|\leq B}|R_u(u,c)| = \mathrm{o_P}(1)$. Since the spectral radius of Γ_c is bounded by one, see (A.8), then (A.9) shows for $1 < B_0 < \infty$

$$\sup_{c_+\leq c<\infty}\|\sum_{l=0}^{m}\Gamma_c^l\| \leq \sup_{c_+\leq c<\infty}\sum_{l=0}^{m}\|\Gamma_c^l\| \leq \sum_{l=0}^{\infty}\sup_{c_+\leq c<\infty}\|\Gamma_c^l\| < B_0.$$

Further with tightness $\sup_{0\leq m<\infty}\sup_{c_+\leq c<\infty}|\widehat{u}_c^{(m)}| = \mathrm{O_P}(1)$ shown in Theorem 2.1 due to Assumption 2.1 with $\eta = 1/4$, the third term vanishes in the above recursive representation. Applying the equality

$$\sum_{l=0}^{m}\Gamma_c^l = (I_{d_x+1}-\Gamma_c)^{-1}(I_{d_x+1}-\Gamma_c^{m+1}) = (I_{d_x+1}-\Gamma_c^{m+1})(I_{d_x+1}-\Gamma_c)^{-1}, \tag{A.12}$$

we then rearrange the recursive representation to attain

$$\widehat{u}_c^{(m+1)} = \Gamma_c^{m+1}\widehat{u}_c^{(0)} + (I_{d_x+1} - \Gamma_c^{m+1})(I_{d_x+1} - \Gamma_c)^{-1}K_c + \mathsf{o}_\mathsf{P}(1), \tag{A.13}$$

uniformly in c, m. Recall the definition of Γ_c in (A.5), then

$$\Gamma_c^{m+1} = \begin{bmatrix} \{\frac{2c\mathsf{f}(c)}{\psi_c}\}^{m+1} I_{d_x} & 0_{d_x} \\ 0'_{d_x} & \{\frac{c(c^2-\varsigma_c^2)\mathsf{f}(c)}{\tau_2^c}\}^{m+1} \end{bmatrix},$$

$$(I_{d_x+1} - \Gamma_c^{m+1})(I_{d_x+1} - \Gamma_c)^{-1} = \begin{bmatrix} \frac{\psi_c^{m+1}-\{2c\mathsf{f}(c)\}^{m+1}}{\psi_c^m\{\psi_c-2c\mathsf{f}(c)\}} I_{d_x} & 0_{d_x} \\ 0'_{d_x} & \frac{(\tau_2^c)^{m+1}-\{c(c^2-\varsigma_c^2)\mathsf{f}(c)\}^{m+1}}{(\tau_2^c)^m\{\tau_2^c-c(c^2-\varsigma_c^2)\mathsf{f}(c)\}} \end{bmatrix}.$$

Finally, substitute these and $\widehat{u}_c^{(0)}$, K_c into (A.13) to establish the expansion of $\widehat{u}_c^{(m+1)}$; see $\widehat{u}_c^{(0)}$ in (A.5) and K_c in (A.6). ■

The next corollary re-expresses Lemma A.3 as a stochastic expansion of the fist step estimators in terms of the initial ones, which is a special case of Theorem 2.2 where $m = 0$.

Proof of Corollary 2.1. The proof follows by setting $m = 0$ in Theorem 2.2 where $\varrho_{\beta,c}^{(1)} = 2c\mathsf{f}(c)/\psi_c$, $\varrho_{x\varepsilon,c}^{(1)} = \psi_c^{-1}$, $\varrho_{\sigma,c}^{(1)} = c(c^2 - \varsigma_c^2)\mathsf{f}(c)/\tau_2^c$, and $\varrho_{\varepsilon\varepsilon,c}^{(1)} = \sigma/(2\tau_2^c)$. ■

We then establish the fixed point of the iterated one-step Huber-skip M-estimators defined in Algorithm 2.1.

Proof of Theorem 2.3. Since the spectral radius of Γ_c is strictly smaller than one, see (A.8), we have $\Gamma_c^{m+1} \to 0_{(d_x+1)\times(d_x+1)}$ uniformly in $c \in [c_+, \infty)$ as $m \to \infty$. Further with the boundedness of $\widehat{u}_c^{(0)}$ in probability as $n \to \infty$ by Assumption 2.1(iii) with $\eta = 1/4$, the first term in (A.13) vanishes. Notice that to attain the recursive representation (A.13) in the proof of Theorem 2.2, we also require Assumption 2.1(i, ii). Thus, let $n, m \to \infty$ in (A.13), we can then obtain the fixed point

$$\widehat{u}_c^{(*)} = \widehat{u}_c^{(\infty)} = (I_{d_x+1} - \Gamma_c)^{-1}K_c, \tag{A.14}$$

uniformly in c. Recall the definition of Γ_c in (A.5), we then have

$$(I_{d_x+1} - \Gamma_c)^{-1} = \begin{Bmatrix} \frac{\psi_c}{\psi_c-2c\mathsf{f}(c)} I_{d_x} & 0_{d_x} \\ 0'_{d_x} & \frac{\tau_2^c}{\tau_2^c-c(c^2-\varsigma_c^2)\mathsf{f}(c)} \end{Bmatrix}.$$

Substitute this and K_c into (A.14) to attain the expression of the fixed point $\widehat{u}_c^{(*)}$; see K_c in (A.6).

The next step is to formally prove that $\widehat{u}_c^{(*)}$ is indeed the fixed point. Replace (A.7) and (A.14) into the deviation $\widehat{\Delta}_c^{(m+1)} = \widehat{u}_c^{(m+1)} - \widehat{u}_c^{(*)}$ and apply (A.12) to attain

$$\widehat{\Delta}_c^{(m+1)} = \Gamma_c^{m+1}\{\widehat{u}_c^{(0)} - (I_{d_x+1} - \Gamma_c)^{-1}K_c\} + \sum_{l=0}^{m} \Gamma_c^l R_u(\widehat{u}_c^{(m-l)}, c).$$

To bound $\widehat{\Delta}_c^{(m+1)}$, use the triangle inequality and $|Mx| \leq \|M\||x|$ to get

$$|\widehat{\Delta}_c^{(m+1)}| \leq \|\Gamma_c^{m+1}\|\{|\widehat{u}_c^{(0)}| + \|(I_{d_x+1} - \Gamma_c)^{-1}\||K_c|\} + \max_{0\leq l\leq m} |R_u(\widehat{u}_c^{(l)}, c)| \sum_{l=0}^{m} \|\Gamma_c^l\|.$$

Further bound above using the inequalities (A.8) and (A.9), so for $m > m_0$

$$|\widehat{\Delta}_c^{(m+1)}| < \omega^{m+1}(|\widehat{u}_c^{(0)}| + B_0|K_c|) + B_0 \max_{0\le l\le m} |R_u(\widehat{u}_c^{(l)}, c)|.$$

On the set $\mathcal{A}_n$ as in (A.11), since $\sup_{0\le m<\infty} \sup_{c_+\le c<\infty} |\widehat{u}_c^{(m)}| \le U_0$ by Theorem 2.1, then $\sup_{c_+\le c<\infty} |\widehat{\Delta}_c^{(m+1)}| < \omega^{m+1}(B_0^{-1}U_0/3 + U_0/3) + \delta/2 < \omega^{m+1}U_0 + \delta/2$. As $0 < \omega < 1$, ω^{m+1} declines exponentially so m_0 can be chosen sufficiently large that for all $m > m_0$ then $\omega^{m+1}U_0 < \delta/2$. Thus, $\mathsf{P}(\sup_{c_+\le c<\infty} |\widehat{\Delta}_c^{(m+1)}| < \delta) > 1 - \epsilon$ for $m > m_0$, $n > n_0$. ■

Next considering RLS and IIS, we first build up their stochastic expansions.

Proof of Theorem 2.4. We first establish the expansion for the Robustified Least Squares. Then, we show that the Impulse Indicator Saturation has the identical expansion of the initial step updated estimator as the Robustified Least Squares so that they have the same general $m+1$ step expansion for any $m \in [0, \infty)$.

RLS starts with the full sample least squares $(\widetilde{\beta}, \widetilde{\sigma})$ which are tight and satisfy

$$N^{-1}(\widetilde{\beta} - \beta) = (\sum_{i=1}^n x_{in}x_{in}')^{-1}(\sum_{i=1}^n x_{in}\varepsilon_i),$$
$$n^{1/2}(\widetilde{\sigma} - \sigma) = \frac{\sigma}{2} n^{-1/2}\sum_{i=1}^n (\frac{\varepsilon_i^2}{\sigma^2} - 1) + \mathrm{O}_{\mathsf{P}}(n^{-1/2}).$$

Substitute the above expansion of the initial estimators $(\widehat{\beta}_c^{(0)}, \widehat{\sigma}_c^{(0)}) = (\widetilde{\beta}, \widetilde{\sigma})$ into Theorem 2.2 using Assumption 2.1(i, ii), then it holds for any $m \in [0, \infty)$,

$$N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta) = \varrho_{\beta,c}^{(m+1)}\Sigma_n^{-1}\sum_{i=1}^n x_{in}\varepsilon_i + \varrho_{x\varepsilon,c}^{(m+1)}\Sigma_n^{-1}\sum_{i=1}^n x_{in}\varepsilon_i 1_{(|\varepsilon_i|\le\sigma c)} + \mathrm{o}_{\mathsf{P}}(1),$$
$$n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = \varrho_{\sigma,c}^{(m+1)}\frac{\sigma}{2}n^{-1/2}\sum_{i=1}^n(\frac{\varepsilon_i^2}{\sigma^2} - 1)$$
$$+ \varrho_{\varepsilon\varepsilon,c}^{(m+1)} n^{-1/2}\sum_{i=1}^n (\frac{\varepsilon_i^2}{\sigma^2} - \varsigma_c^2)1_{(|\varepsilon_i|\le\sigma c)} + \mathrm{o}_{\mathsf{P}}(1),$$

uniformly in $c \in [c_+, \infty)$.

To demonstrate that two algorithms RLS and IIS have the identical expansion of $N^{-1}(\widehat{\beta}_c^{(m+1)} - \beta)$ for $m \in [0, \infty)$ even they start with different initial estimators when running Algorithm 2.1, it suffices to show that the expansion of $N^{-1}(\widehat{\beta}_c^{(1)} - \beta)$ for IIS is the same as RLS in the above. The updated estimator for β from an initial step in IIS is expressed as

$$N^{-1}(\widehat{\beta}_c^{(1)} - \beta) = \{\sum_{j=1,2}(N_{3-j}^{-1}N)'\sum_{i\in\mathcal{I}_{3-j}} x_{in_{3-j}}x_{in_{3-j}}' 1_{(|y_i - x_i'\widehat{\beta}_j|\le\widehat{\sigma}_j c)} N_{3-j}^{-1}N\}^{-1}$$
$$\times \{\sum_{j=1,2}(N_{3-j}^{-1}N)'\sum_{i\in\mathcal{I}_{3-j}} x_{in_{3-j}}\varepsilon_i 1_{(|y_i - x_i'\widehat{\beta}_j|\le\widehat{\sigma}_j c)}\}.$$

Notice $x_{in_j} = N_j' x_i$ denotes the normalized regressors for each subsample $i \in \mathcal{I}_j$, $j = 1, 2$, where N_j corresponds to the normalization matrix based on the subsample size n_j.

Argue along the lines of Lemma A.3 using the expansions in Lemma A.1, then it follows

$$
\begin{aligned}
N^{-1}(\widehat{\beta}_c^{(1)}-\beta) = (\psi_c\Sigma_n)^{-1}\{2c\mathsf{f}(c)&\sum_{j=1,2}(N_{3-j}^{-1}N)'\Sigma_{n_{3-j}}^{\mathcal{I}_{3-j}}(N_{3-j}^{-1}N_j)N_j^{-1}(\widehat{\beta}_j-\beta)\\
&+\sum_{i=1}^{n}x_{in}\varepsilon_i 1_{(|\varepsilon_i|\leq\sigma c)}\}+\mathsf{o}_{\mathsf{P}}(1),
\end{aligned}
$$

uniformly in $c\in[c_+,\infty)$ and where we denote $\Sigma_{n_j}^{\mathcal{I}_j}=\sum_{i\in\mathcal{I}_j}x_{in_j}x_{in_j}'$ for $j=1,2$. To apply the empirical processes argument in the above, we require Assumption 2.1(i,iib) holding for each subsample $\mathcal{I}_j$ and the fact that the initial estimators $N_j^{-1}(\widehat{\beta}_j-\beta)$, $n_j^{1/2}(\widehat{\sigma}_j-\sigma)$ are tight for $j=1,2$. Further substituting the expansion of least squares on each subsample $\mathcal{I}_j$, that is $N_j^{-1}(\widehat{\beta}_j-\beta)=(\Sigma_{n_j}^{\mathcal{I}_j})^{-1}\sum_{i\in\mathcal{I}_j}x_{in_j}\varepsilon_i$ for $j=1,2$, we have

$$
\begin{aligned}
N^{-1}(\widehat{\beta}_c^{(1)}-\beta) = \frac{2c\mathsf{f}(c)}{\psi_c}\Sigma_n^{-1}&\sum_{j=1,2}(N_{3-j}^{-1}N)'\Sigma_{n_{3-j}}^{\mathcal{I}_{3-j}}(N_{3-j}^{-1}N_j)(\Sigma_{n_j}^{\mathcal{I}_j})^{-1}\sum_{i\in\mathcal{I}_j}x_{in_j}\varepsilon_i\\
&+(\psi_c\Sigma_n)^{-1}\sum_{i=1}^{n}x_{in}\varepsilon_i 1_{(|\varepsilon_i|\leq\sigma c)}+\mathsf{o}_{\mathsf{P}}(1).
\end{aligned}
$$

Notice that we assume $n_1=\mathrm{int}[n/2]$ and $n_2=n-n_1$ so that $N_{3-j}^{-1}N_j\to I_{d_x}$ and $N_jN_{3-j}^{-1}\to I_{d_x}$ as $n\to\infty$ and $n_j\to\infty$ for $j=1,2$. Combine this fact and Assumption 2.1(iia) that $\Sigma_{n_j}^{\mathcal{I}_j}\overset{\mathsf{P}}{\to}\Sigma$ for $j=1,2$ to attain

$$
N^{-1}(\widehat{\beta}_c^{(1)}-\beta) = \frac{2c\mathsf{f}(c)}{\psi_c}\Sigma_n^{-1}\sum_{i=1}^{n}x_{in}\varepsilon_i+(\psi_c\Sigma_n)^{-1}\sum_{i=1}^{n}x_{in}\varepsilon_i 1_{(|\varepsilon_i|\leq\sigma c)}+\mathsf{o}_{\mathsf{P}}(1).
$$

Then, we find that RLS and IIS have the identical expansion of $N^{-1}(\widehat{\beta}_c^{(1)}-\beta)$ by noting $\varrho_{\beta,c}^{(1)}=2c\mathsf{f}(c)/\psi_c$ and $\varrho_{x\varepsilon,c}^{(1)}=\psi_c^{-1}$, so as for the general $m+1$ step beta estimators for $m\in[0,\infty)$. Using the similar reasoning as on beta, it follows that two algorithms also have the same expansion on sigma. ■

Drifting the cut-off value $c\in[c_+,\infty)$, we derive the weak convergence theory of the processes $\mathbb{G}_n^{(m+1)}(c)=N^{-1}(\widehat{\beta}_c^{(m+1)}-\beta)$ computed by RLS and IIS for any $m\in[0,\infty)$ in the stationary case, where $N=n^{-1/2}I_{d_x}$ such that $x_{in}=n^{-1/2}x_i$ and $\Sigma=\mathsf{E}x_ix_i'$.

Proof of Theorem 2.5. By Assumption 2.1(i,ii), Theorem 2.4 shows that for any $m\in[0,\infty)$ and uniformly in $c\in[c_+,\infty)$

$$
\mathbb{G}_n^{(m+1)}(c)=\begin{pmatrix}\varrho_{\beta,c}^{(m+1)}\Sigma_n^{-1}\\ \varrho_{x\varepsilon,c}^{(m+1)}\Sigma_n^{-1}\end{pmatrix}'\sum_{i=1}^{n}\begin{pmatrix}x_{in}\varepsilon_i\\ x_{in}\varepsilon_i 1_{(|\varepsilon_i|\leq\sigma c)}\end{pmatrix}+\mathsf{o}_{\mathsf{P}}(1).
$$

Then, by $\Sigma_n\overset{\mathsf{P}}{\to}\Sigma>0$, the limiting distribution of the kernel vector, and Slutsky's theorem, we have for any $c\in[c_+,\infty)$

$$
\mathbb{G}_n^{(m+1)}(c)\overset{\mathsf{D}}{\to}\begin{pmatrix}\varrho_{\beta,c}^{(m+1)}\Sigma^{-1}\\ \varrho_{x\varepsilon,c}^{(m+1)}\Sigma^{-1}\end{pmatrix}'\mathsf{N}\left\{\begin{pmatrix}0_{d_x}\\ 0_{d_x}\end{pmatrix},\sigma^2\tau_2^c\begin{pmatrix}\frac{1}{\tau_2^c}\Sigma & \Sigma\\ \Sigma & \Sigma\end{pmatrix}\right\}.
$$

Since a transformation of multivariate normal is still normal, it follows

$$\mathbb{G}_n^{(m+1)}(c) \xrightarrow{\mathsf{D}} \mathsf{N}[0_{d_x}, \{(\varrho_{\beta,c}^{(m+1)})^2 + 2\tau_2^c \varrho_{\beta,c}^{(m+1)} \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c (\varrho_{x\varepsilon,c}^{(m+1)})^2\} \sigma^2 \Sigma^{-1}].$$

Theorem 2.1 demonstrates the tightness of the processes $\mathbb{G}_n^{(m+1)}$ for any $m \in [0, \infty)$, so

$$\mathbb{G}_n^{(m+1)} \rightsquigarrow \mathbb{G}^{(m+1)},$$

where the weak limit $\mathbb{G}^{(m+1)}$ is a zero mean Gaussian process with the variance

$$\mathsf{Var}\{\mathbb{G}^{(m+1)}(c)\} = \{(\varrho_{\beta,c}^{(m+1)})^2 + 2\tau_2^c \varrho_{\beta,c}^{(m+1)} \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c (\varrho_{x\varepsilon,c}^{(m+1)})^2\} \sigma^2 \Sigma^{-1}. \qquad \blacksquare$$

Next, we give the weak convergence of the first step and fixed point estimator of RLS and IIS.

Proof of Corollary 2.2. These are two special cases of Theorem 2.5 where $m = 0$ such that $\varrho_{\beta,c}^{(1)} = 2c\mathsf{f}(c)/\psi_c$, $\varrho_{x\varepsilon,c}^{(1)} = \psi_c^{-1}$ and where $m \to \infty$ such that $\varrho_{\beta,c}^{(\infty)} = 0$, $\varrho_{x\varepsilon,c}^{(\infty)} = 1/\{\psi_c - 2c\mathsf{f}(c)\}$. $\blacksquare$

We now turn to proving the results for the outlier distortion test and first show the stochastic expansion and weak limit of the difference processes between the RLS/IIS and full sample OLS. The proof here mainly concentrates on stationary regressors, whereas for deterministic trends and unit roots the same argument can be applied using the weak convergence results shown in §2.3. Thus, we choose $N = n^{-1/2} I_{d_x}$ such that the RLS/IIS and OLS need to be normalised by $N^{-1} = n^{1/2} I_{d_x}$, $x_{in} = n^{-1/2} x_i$, and $\Sigma = \mathsf{E} x_i x_i'$.

Proof of Theorem 2.6. Rearrange $\mathbb{H}_n^{(m+1)}(c)$ as

$$\mathbb{H}_n^{(m+1)}(c) = n^{1/2}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) = n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta) - n^{1/2}(\widetilde{\beta} - \beta).$$

Incorporate the expansions of RLS/IIS $n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta)$ and OLS $n^{1/2}(\widetilde{\beta} - \beta)$ from Theorem 2.4 and its proof into the above term, then the expansion is immediately attained for $\mathbb{H}_n^{(m+1)}(c)$ for any $m \in [0, \infty)$ and $c \in [c_+, \infty)$. We can next obtain the weak Gaussian limit $\mathbb{H}^{(m+1)}$ of a sequence of processes $\mathbb{H}_n^{(m+1)}$ by arguing along the lines of Proof of Theorem 2.5 but replacing $\varrho_{\beta,c}^{(m+1)}$ by $\varrho_{\beta,c}^{(m+1)} - 1$. $\blacksquare$

Next, we establish the outlier distortion test and prove Corollary 2.3.

Proof of Corollary 2.3. Given any cut-off values, the Gaussian weak limit from Theorem 2.6 immediately implies that the difference between RLS/IIS and OLS converges pointwisely to a Normal distribution. Thus, the proposed Hausman type test statistics has a limiting chi-squared distribution. $\blacksquare$

Using the relative efficiency argument similar to Hausman (1978), we show in the below lemma that the asymptotic variance of the difference between RLS/IIS and OLS can be given by the difference of their respective asymptotic variances under the null of no outliers.

Lemma A.4. *Consider RLS or IIS. Suppose Assumption 2.1(i, ii) holds. For any $m \in [0, \infty)$, $c \in [c_+, \infty)$ and as $n \to \infty$, we have*

$$\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) = \mathsf{avar}(\widehat{\beta}_c^{(m+1)}) - \mathsf{avar}(\widetilde{\beta}).$$

Proof of Lemma A.4. Under the null of no outliers, for any $m \in [0, \infty)$, $c \in [c_+, \infty)$ RLS/IIS $\widehat{\beta}_c^{(m+1)}$ and OLS $\widetilde{\beta}$ are both consistent, although $\widehat{\beta}_c^{(m+1)}$ is less efficient than $\widetilde{\beta}$ in terms of having the higher asymptotic variance, see Theorem 2.5. We take a weighted average between RLS/IIS and OLS to construct a new estimator

$$\widehat{\theta}_c^{(m+1)}(\lambda) = \lambda \widehat{\beta}_c^{(m+1)} + (1-\lambda)\widetilde{\beta},$$

where $\lambda \in [0, 1]$. The class of estimators $\widehat{\theta}_c^{(m+1)}(\lambda)$ are consistent, and the choice of λ determines the trade-off between efficiency and robustness. The closer λ is to zero, the more efficient $\widehat{\theta}_c^{(m+1)}(\lambda)$ is. When $\lambda = 0$ the constructed estimator becomes OLS, so $\widehat{\theta}_c^{(m+1)}(0) = \widetilde{\beta}$ is the most efficient estimator in the class, that is to say $\mathsf{avar}\{\widehat{\theta}_c^{(m+1)}(\lambda)\}$ attains the minimum at zero. Notice

$$\mathsf{avar}\{\widehat{\theta}_c^{(m+1)}(\lambda)\} = \lambda^2 \mathsf{avar}(\widehat{\beta}_c^{(m+1)}) + (1-\lambda)^2 \mathsf{avar}(\widetilde{\beta}) + 2\lambda(1-\lambda)\mathsf{acov}(\widehat{\beta}_c^{(m+1)}, \widetilde{\beta}).$$

Then, we check its first and second derivatives

$$\begin{aligned}
\frac{d}{d\lambda}\mathsf{avar}\{\widehat{\theta}_c^{(m+1)}(\lambda)\} &= 2\{\lambda \mathsf{avar}(\widehat{\beta}_c^{(m+1)}) - (1-\lambda)\mathsf{avar}(\widetilde{\beta}) + (1-2\lambda)\mathsf{acov}(\widehat{\beta}_c^{(m+1)}, \widetilde{\beta})\}, \\
\frac{d^2}{d\lambda^2}\mathsf{avar}\{\widehat{\theta}_c^{(m+1)}(\lambda)\} &= 2\{\mathsf{avar}(\widehat{\beta}_c^{(m+1)}) + \mathsf{avar}(\widetilde{\beta}) - 2\mathsf{acov}(\widehat{\beta}_c^{(m+1)}, \widetilde{\beta})\} \\
&= 2\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) \geq 0.
\end{aligned}$$

Thus, the function $\mathsf{avar}\{\widehat{\theta}_c^{(m+1)}(\lambda)\}$ is convex and minimised at $\lambda = 0$, then it follows $\frac{d}{d\lambda}\mathsf{avar}\{\widehat{\theta}_c^{(m+1)}(\lambda)\}|_{\lambda=0} = 0$ subsequently implying $\mathsf{acov}(\widehat{\beta}_c^{(m+1)}, \widetilde{\beta}) = \mathsf{avar}(\widetilde{\beta})$. Finally,

$$\begin{aligned}
\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) &= \mathsf{avar}(\widehat{\beta}_c^{(m+1)}) + \mathsf{avar}(\widetilde{\beta}) - 2\mathsf{acov}(\widehat{\beta}_c^{(m+1)}, \widetilde{\beta}) \\
&= \mathsf{avar}(\widehat{\beta}_c^{(m+1)}) - \mathsf{avar}(\widetilde{\beta}).
\end{aligned}$$ ■

We provide a different but more direct proof in Remark A.1 to establish the equality $\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) = \mathsf{avar}(\widehat{\beta}_c^{(m+1)}) - \mathsf{avar}(\widetilde{\beta})$ under the null of no outliers. The proof relies on the asymptotics derived for RLS/IIS $\widehat{\beta}_c^{(m+1)}$ shown in Theorem 2.5. Furthermore, the remark indirectly indicates that the normal distribution for errors satisfies the regularity conditions required by Lemma A.4.

Remark A.1. *Rearrange the expression of* $\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ *from Corollary 2.3 to attain*

$$\{(\varrho_{\beta,c}^{(m+1)})^2 - 2\varrho_{\beta,c}^{(m+1)} + 1 + 2\tau_2^c \varrho_{\beta,c}^{(m+1)} \varrho_{x\varepsilon,c}^{(m+1)} - 2\tau_2^c \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c(\varrho_{x\varepsilon,c}^{(m+1)})^2\}\sigma^2\Sigma^{-1}.$$

Recall $\tau_2^c = \psi_c - 2c\mathsf{f}(c)$ *if* $\mathsf{f} \overset{\mathsf{D}}{=} \mathsf{N}(0,1)$ *and from Theorem 2.2 that*

$$\varrho_{\beta,c}^{(m+1)} = \{\frac{2c\mathsf{f}(c)}{\psi_c}\}^{m+1}, \qquad \varrho_{x\varepsilon,c}^{(m+1)} = \frac{\psi_c^{m+1} - \{2c\mathsf{f}(c)\}^{m+1}}{\psi_c^{m+1}\{\psi_c - 2c\mathsf{f}(c)\}}.$$

Apply these terms to have $-2\varrho_{\beta,c}^{(m+1)} + 1 - 2\tau_2^c \varrho_{x\varepsilon,c}^{(m+1)} = -1$*, and further notice from Theorem 2.5 that* $\mathsf{avar}(\widehat{\beta}_c^{(m+1)}) = \{(\varrho_{\beta,c}^{(m+1)})^2 + 2\tau_2^c \varrho_{\beta,c}^{(m+1)} \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c(\varrho_{x\varepsilon,c}^{(m+1)})^2\}\sigma^2\Sigma^{-1}$ *and* $\mathsf{avar}(\widetilde{\beta}) = \sigma^2\Sigma^{-1}$*. Thus, we finally shows*

$$\begin{aligned}
\mathsf{avar}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta}) &= \{(\varrho_{\beta,c}^{(m+1)} - 1)^2 + 2\tau_2^c(\varrho_{\beta,c}^{(m+1)} - 1)\varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c(\varrho_{x\varepsilon,c}^{(m+1)})^2\}\sigma^2\Sigma^{-1} \\
&= \{(\varrho_{\beta,c}^{(m+1)})^2 + 2\tau_2^c \varrho_{\beta,c}^{(m+1)} \varrho_{x\varepsilon,c}^{(m+1)} + \tau_2^c(\varrho_{x\varepsilon,c}^{(m+1)})^2\}\sigma^2\Sigma^{-1} - \sigma^2\Sigma^{-1} \\
&= \mathsf{avar}(\widehat{\beta}_c^{(m+1)}) - \mathsf{avar}(\widetilde{\beta}).
\end{aligned}$$

Finally, we prove Corollary 2.4, which provides two special cases of the outlier distortion tests comparing $\widetilde{\beta}$ with $\widehat{\beta}_c^{(1)}$ when $m = 0$ and with $\widehat{\beta}_c^{(*)}$ when $m \to \infty$.

Proof of Corollary 2.4. Set $m = 0$ and $m \to \infty$ such that $\varrho_{\beta,c}^{(1)} = 2c\mathsf{f}(c)/\psi_c$, $\varrho_{x\varepsilon,c}^{(1)} = \psi_c^{-1}$ and $\varrho_{\beta,c}^{(\infty)} = 0$, $\varrho_{x\varepsilon,c}^{(\infty)} = \{\psi_c - 2c\mathsf{f}(c)\}^{-1}$ for $\widehat{\mathsf{avar}}(\widehat{\beta}_c^{(m+1)} - \widetilde{\beta})$ in (2.12), then we immediately obtain our proposed tests. ∎

B Additional Simulation Results

Here we report additional simulation results in the presence of outliers. Figure B.1 shows the simulation results under a range of alternatives for 15% outlier contamination. Figure B.2 shows the simulation results when testing for distortion of a single coefficient under a range of alternatives. Figure B.3 shows the simulation results when testing for distortion when the dependent variable shows some persistence (autoregressive coefficient of 0.5) under a range of alternatives.

Figures B.4 and B.5 show the bootstrap simulation results for outlier magnitudes of 2 and 3 standard deviations of the error term respectively.

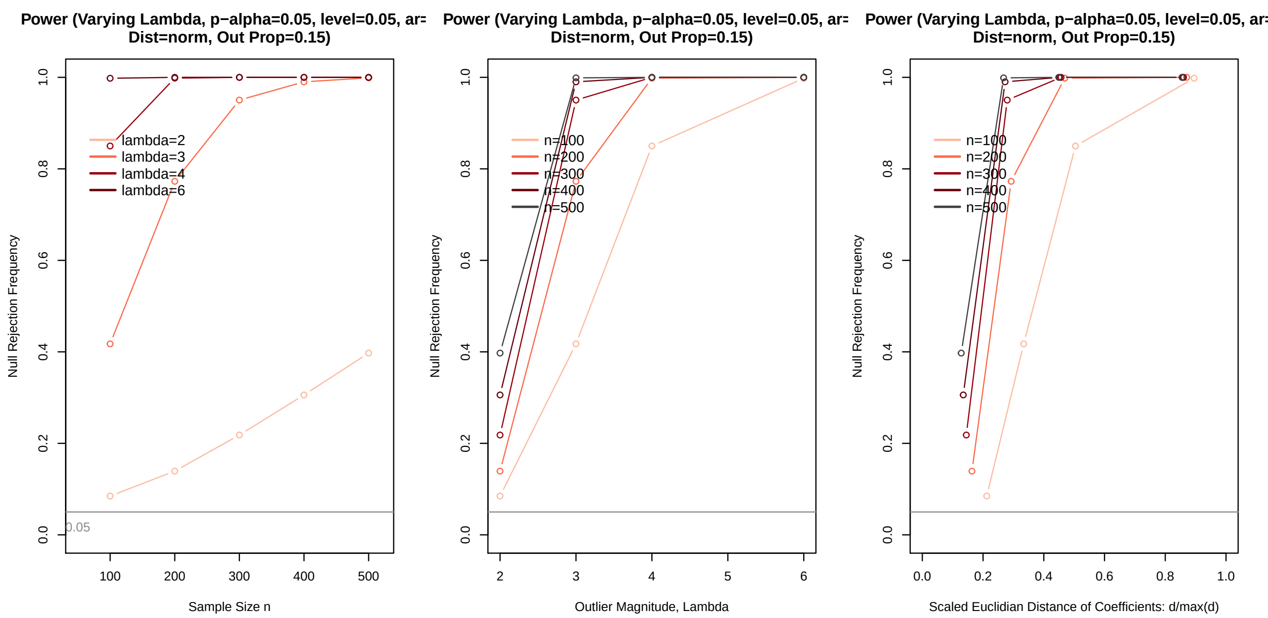

Figure B.1: Simulation performance under the alternative for varying sample sizes and outlier magnitudes when the dependent variable exhibits no autocorrelation and 15% of the sample is outlier-contaminated.

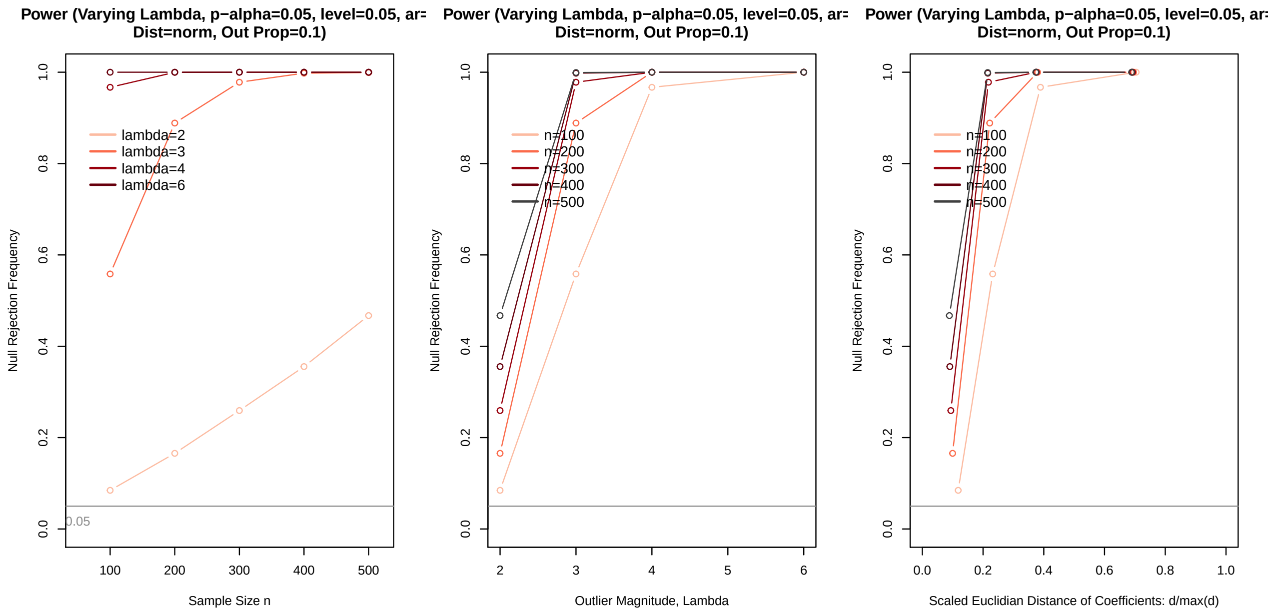

Figure B.2: Simulation performance under the alternative, testing a single regressor, for varying sample sizes and outlier magnitudes when the dependent variable exhibits no autocorrelation and 10% of the sample is outlier-contaminated.

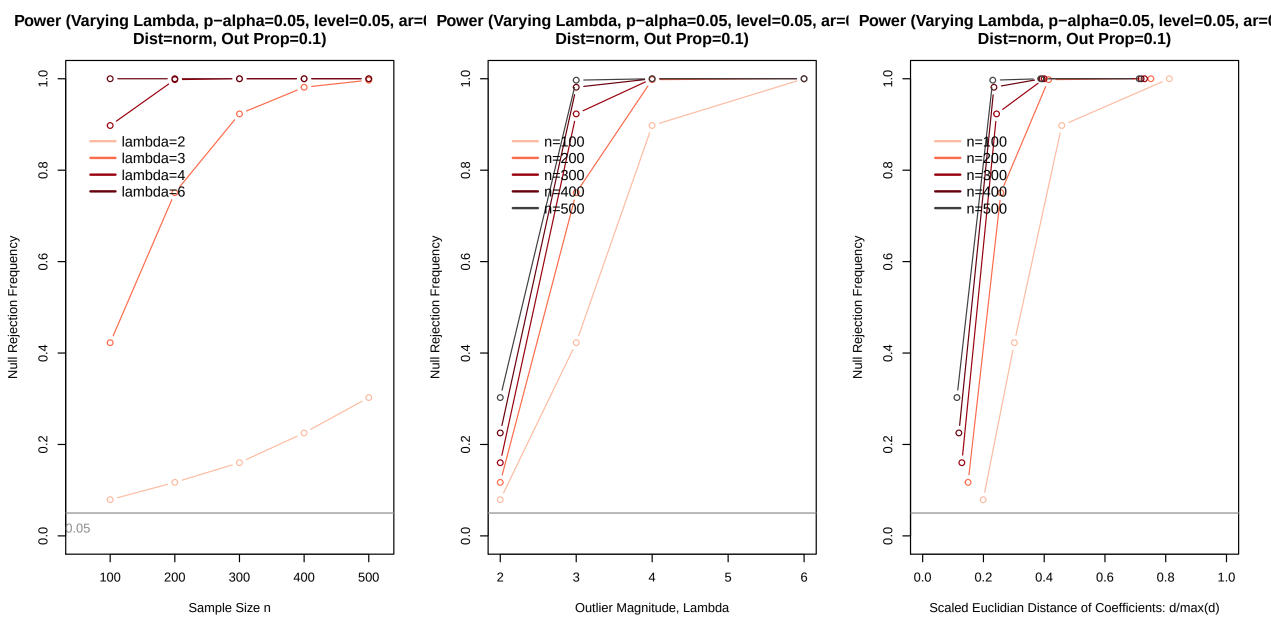

Figure B.3: Simulation performance under the alternative for varying sample sizes and outlier magnitudes when the dependent variable exhibits some autocorrelation (AR coefficient = 0.5) and 10% of the sample is outlier-contaminated.

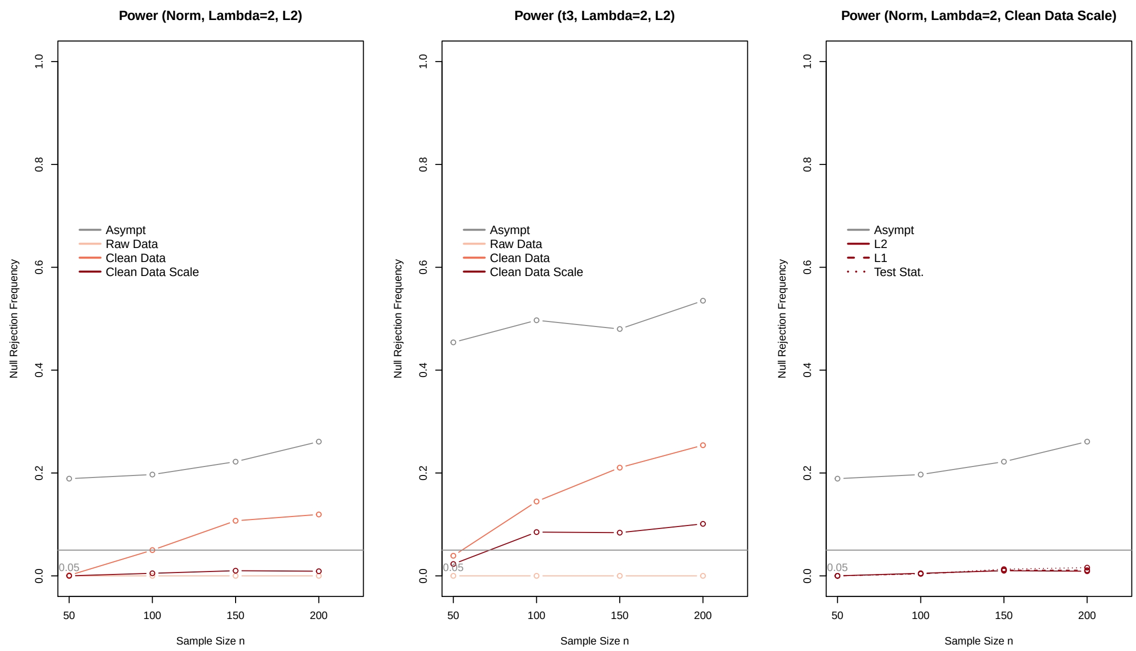

Figure B.4: Simulation performance under the alternative for varying sample sizes when using the bootstrap implementations of our test (for outlier magnitude of 2SD of the error term).

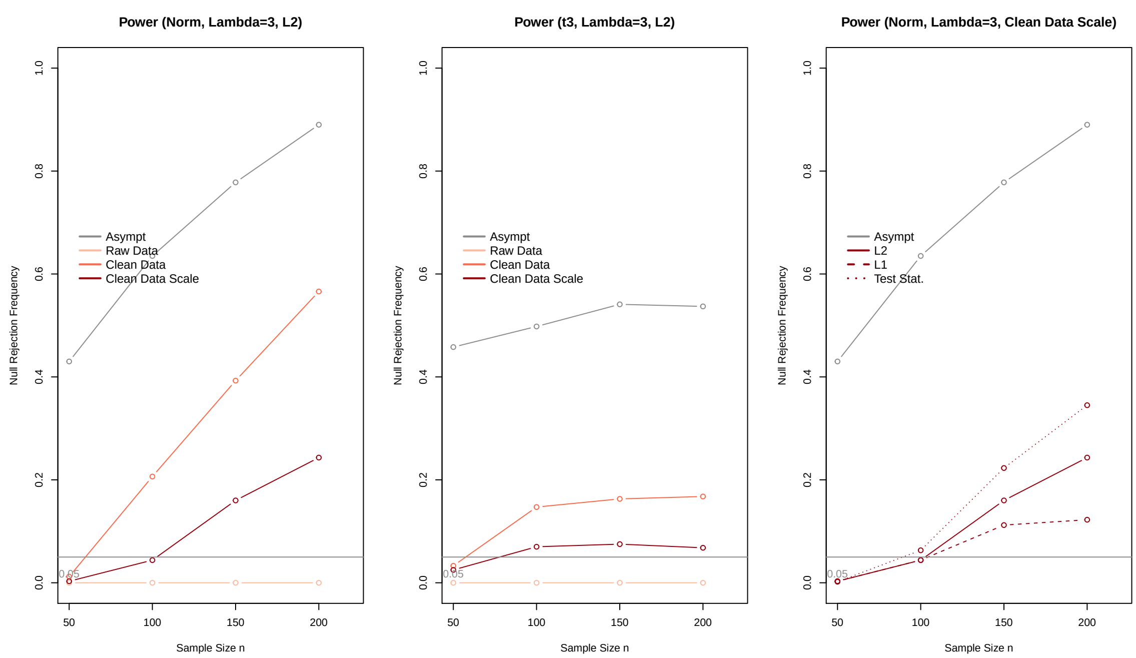

Figure B.5: Simulation performance under the alternative for varying sample sizes when using the bootstrap implementations of our test (for outlier magnitude of 3SD of the error term).