econometrics

MDPI

Article

Generalized Recentered Influence Function Regressions

Javier Alejo [1,†], Antonio Galvao [2,†], Julián Martínez-Iriarte [3,†] and Gabriel Montes-Rojas [4,*,†]

[1] Instituto de Economía, Universidad de la República, Montevideo 11200, Uruguay; javier.alejo@fcea.edu.uy
[2] Department of Economics, Michigan State University, East Lansing, MI 48824, USA; agalvao@msu.edu
[3] Department of Economics, University of California-Santa Cruz, Santa Cruz, CA 95064, USA; jmart425@ucsc.edu
[4] Instituto Interdisciplinario de Economía Política-CONICET, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires C1122, Argentina
* Correspondence: gabriel.montes@economicas.uba.ar
† These authors contributed equally to this work.

Abstract: This paper suggests a generalization of covariate shifts to study distributional impacts on inequality and distributional measures. It builds on the recentered influence function (RIF) regression method, originally designed for location shifts in covariates, and extends it to general policy interventions, such as location–scale or asymmetric interventions. Numerical simulations for the Gini, Theil, and Atkinson indexes demonstrate strong performance across a myriad of cases and distributional measures. An empirical application examining changes in Mincerian equations is presented to illustrate the method.

Keywords: influence functions; RIF regression; Gini; Atkinson; counterfactual distributions

1. Introduction

The recentered influence function (RIF) regression has emerged as a powerful tool for examining how changes in covariates influence the distributional properties of an outcome variable. This approach is particularly valuable in the analysis of income distribution, inequality, and poverty. First introduced in the seminal work of Firpo et al. (2009), RIF regression is based on a transformation of the outcome variable, leveraging the influence function of a specific distributional statistic of interest. The influence function, a fundamental concept in robust statistics, measures the sensitivity of an unconditional statistic to individual observations and is widely documented in the statistical literature.[1] This accessibility makes RIF regression both convenient and straightforward to implement. By regressing the RIF-transformed outcome on explanatory variables, researchers can estimate the marginal effects of these covariates on various distributional measures, offering a flexible and insightful framework for studying economic and social disparities.

RIF regression has been extensively applied in empirical research, particularly in the study of unconditional quantiles. In this context, the primary parameter of interest is the unconditional quantile effect (UQE), and the corresponding RIF regression is referred to as unconditional quantile regression (UQR). Beyond quantiles, RIF regression can be applied to a broad set of distributional statistics, including the Gini coefficient, the Theil index, and measures of polarization, provided that the influence function for the statistic is known. This flexibility has contributed to its widespread adoption across multiple disciplines, including labor economics, health economics, and public policy, where understanding the distributional effects of explanatory variables is crucial.

Building on the foundational work of Firpo et al. (2009), subsequent research has further developed RIF-based methodologies. Essama-Nssah and Lambert (2012), for instance, pro-

check for updates

Academic Editor: Willi Semmler

Received: 21 January 2025
Revised: 13 March 2025
Accepted: 7 April 2025
Published: 18 April 2025

Citation: Alejo, J., Galvao, A., Martínez-Iriarte, J., & Montes-Rojas, G. (2025). Generalized Recentered Influence Function Regressions. *Econometrics*, *13*(2), 19. https://doi.org/10.3390/econometrics13020019

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

vide a comprehensive derivation of influence functions for a range of distributional statistics, which serve as key components in RIF models. Moreover, Fortin et al. (2011) extend RIF regression to the Oaxaca–Blinder decomposition framework (see, e.g., Oaxaca, 1973; Oaxaca & Ransom, 1994), allowing for a decomposition of differences in distributional statistics across groups. Additional refinements have been proposed by Firpo and Pinto (2016) and Firpo et al. (2018), further enhancing the applicability and robustness of the method.

Despite its extensive use, the conventional RIF regression framework primarily captures location-shift effects, where a marginal change in a covariate affects the unconditional distribution of the outcome variable in a uniform manner. Specifically, the model of Firpo et al. (2009) considers shifts in the mean of a single covariate or marginal changes in the probability mass of a binary regressor. However, in many practical settings, policy interventions do not merely induce uniform changes in covariates but instead lead to broader transformations that may include shifts in both location and scale, as well as asymmetric changes. Addressing this limitation, we propose a more general and intuitive parametric representation of covariate shifts within the RIF framework.

Our contribution lies in extending RIF regression to accommodate general parametric covariate shifts, allowing for a more flexible evaluation of policy interventions. By formulating a parametric approach, we provide a simple yet effective method applicable to any distributional functional. This generalization enables researchers to assess policy effects beyond location shifts, capturing more complex transformations in covariates. Our proposed framework enhances the analytical power of RIF regression, making it a more versatile tool for studying the impact of structural changes in economic and social environments.

Recent research has explored various approaches to extending the location-only effects framework, particularly in the context of unconditional quantile regression (UQR). Chernozhukov et al. (2013) analyze this problem as a means of transforming the status quo population; however, their work does not explicitly address the limiting or marginal cases considered in Firpo et al. (2009).[2] While their study provides valuable insights, it does not propose a methodology for capturing general shifts in covariates or for extending the analysis to a broader set of distributional functionals.

A related contribution by Martínez-Iriarte et al. (2024) introduces a framework for modeling location and scale shift effects in UQR estimation, making it particularly useful for analyzing general policy interventions. However, their approach relies on maximum likelihood estimation rather than RIF regression.[3] Other notable extensions of UQR include Inoue et al. (2021), which examines the two-sample problem, Sasaki et al. (2022), which addresses high-dimensional settings, and Alejo et al. (2024), which investigates the relationship between conditional and unconditional quantiles.

To illustrate the effectiveness of our proposed method, we conduct Monte Carlo simulations using the Gini coefficient as the target distributional functional. The simulation results demonstrate that the method performs well across a range of parametric shift functions. Finally, we present an empirical application that examines the impact of different policy interventions on inequality under various shift scenarios. Specifically, we analyze how changes in education and work experience influence inequality within a Mincerian earnings equation framework.

The remainder of this paper is organized as follows. Section 2 reviews the RIF regression method and introduces our proposed estimands for capturing different types of covariate shifts. This section also discusses potential empirical applications in the existing literature. Section 3 presents the corresponding estimators and explores their asymptotic properties. Section 4 evaluates the finite sample performance of the estimators using Monte Carlo experiments. Section 5 applies the proposed estimators to analyze shifts in education and age within the Mincer equation framework. Finally, Section 6 concludes with a discussion of the key findings and potential avenues for future research.

2. RIF Regression

2.1. RIF Regression Framework for Pure Location-Shifts

Consider a functional $v(\cdot)$ defined on the distribution function F_Y of a random variable Y. Changes in v that measure the influence or impact of a given observation through F_Y are studied using the influence function (IF). More formally, the IF is the directional derivative of $v(F_Y)$ at F_Y and measures the effect of a small perturbation in F_Y. The IF can be calculated at each data point in the domain of Y. Let y be be such point, that contains a Dirac probability mass of ϵ_y, such that we define

$$IF(y; v; F_Y) = \lim_{t \to 0} \frac{v\left[t\epsilon_y + (1-t)F_Y\right] - v(F_Y)}{t}.$$

When $IF(\cdot; v; F_Y)$ is evaluated at Y, it becomes a random variable. Importantly, it has zero mean by construction, that is, $E_Y[IF(Y, v, F_Y)] = 0$. For example, if $v(F_Y) = F_Y^{-1}(\tau)$, the τ-quantile of F_Y, then

$$IF(y; v; F_Y) = \frac{\tau - 1\left\{y \leq F_Y^{-1}(\tau)\right\}}{f_Y(F_Y^{-1}(\tau))}.$$

Note that

$$E[IF(Y; v; F_Y)] = \frac{\tau - E\left[1\left\{Y \leq F_Y^{-1}(\tau)\right\}\right]}{f_Y(F_Y^{-1}(\tau))} = 0,$$

because $E\left[1\left\{Y \leq F_Y^{-1}(\tau)\right\}\right] = F_Y(F_Y^{-1}(\tau)) = \tau$. The Appendix A contains the influence function for the Gini, Atkinson, and Theil indices.

The recentered influence function (RIF) is simply defined as

$$RIF(Y, v, F_Y) = v(F_Y) + IF(Y, v, F_Y),$$

and by the previous result it satisfies $v(F_Y) = E_Y[RIF(Y, v, F_Y)]$. This is a key property as it allows to implement the Law of Iterated Expectations and to work with conditional models. In particular, if we consider a set of covariates of interest X, then we have that

$$v(F_Y) = E\{E[RIF(Y, v, F_Y) \mid X]\}. \tag{1}$$

Firpo et al. (2009) key result is based on modeling of the conditioning part, $m_v(x) \equiv E[RIF(Y, v, F_Y) \mid X = x]$ using standard regression tools, as this is referred to as RIF regression. Then, by the properties above, integrating over X links to the unconditional statistic of interest, i.e., $v(F_Y) = E_X[m_v(X)]$.

Firpo et al. (2009) focuses on marginal effects on covariates X, modeled as a location shift $X + \delta$ where δ is a small real valued perturbation. Changes in X have a corresponding effect on Y, the outcome of interest. To study this, consider a statistical tool built upon the IF. Define F_{Y_δ} as the distribution of Y after the perturbation in X. Then we consider a functional derivative given by

$$\lim_{\delta \to 0} \frac{v(F_{Y_\delta}) - v(F_Y)}{\delta} = \lim_{\delta \to 0} \frac{E[m_v(X+\delta)] - E[m_v(X)]}{\delta} = E\left[\lim_{\delta \to 0} \frac{m_v(X+\delta) - m_v(X)}{\delta}\right].$$

The last equality is valid if we assume that limit and expectation are interchangeable, which we assume throughout the paper. In fact, if we assume that $m_v(x)$ is differentiable at any point x, then we obtain the main result of RIF regression models, that is,

$$\Pi_v \equiv \lim_{\delta \to 0} \frac{v(F_{Y_\delta}) - v(F_Y)}{\delta} = E\left[\frac{\partial m_v(x)}{\partial x}\right] = E\left[\frac{\partial E[RIF(Y, v, F_Y) \mid X = x]}{\partial x}\right].$$

Π_v measures the impact of a given location shift in covariates on the unconditional functional statistic v of Y. Statistically, the expression for Π_v, is called an average derivative, and is an object whose estimation has been thoroughly studied. An excellent textbook reference is chapter 4 of Pagan and Ullah (1999).

2.2. General Unconditional Effects

While Firpo et al. (2009) focus mainly on a location shift, one can consider a general shift on a given covariate,

$$X_\delta = h(X, \delta), \tag{2}$$

where $\delta \geq 0$, $h(x, 0) = x$, and as a function of δ the function $h(X, \delta)$ is a continuously differentiable function. Then we consider general unconditional effects as

$$\Pi_{v,h} = \lim_{\delta \to 0} \frac{v(F_{Y_\delta}) - v(F_Y)}{\delta} \tag{3}$$

for general h functions.

Here we index the unconditional effect by h, which denotes the type of counterfactual policy analyzed. Naturally, depending on the particular form of $h(X, \delta)$ we obtain different effects. Here are few examples:

- Location shift This is the case developed above taken from Firpo et al. (2009) analysis that consider a location shift change in one covariate of the form

$$X_\delta = X + \delta. \tag{4}$$

- Location–scale shift

$$X_\delta = \frac{X - \mu_X}{s(\delta)} + \mu_X + \ell(\delta) \tag{5}$$

  where $\delta \geq 0$, $s(\delta) > 0$ and $\ell(\delta)$ are continuously differentiable functions with $s(0) = 1$ and $\ell(0) = 0$, respectively. Here, $s(\delta)$ acts as a shrinking parameter such that an increment in this parameter reduces the overall impact of the X variable. The pure location shift in Firpo et al. (2009) can be obtained by setting $s(\delta) = 1$ and $\ell(\delta) = \delta$. Martínez-Iriarte et al. (2024) have a similar shift written in a different manner. In fact, different alternatives can be developed based on how to combine the location and scale joint shifts.
- Asymmetric shift

$$X_\delta = X + a(\delta)(X_{max} - X)^\lambda \tag{6}$$

  where $\delta \geq 0$, and the map $\delta \mapsto a(\delta)$ satisfies: $a(\delta) > 0$, $a(0) = 0$, and is continuously differentiable. The factor X_{max} is maximum value in the support of X or an upper bound. The parameter λ determines the asymmetry of the shift effects: if $\lambda < 0$ then the shift is biased towards upper values, if $\lambda > 0$ the shift is biased towards lower values, if $\lambda = 0$ this would be a pure location-shift. Although this type of shift was applied as a numerical simulation exercise in Battiston et al. (2014), the novelty of our proposal is to include it analytically within the RIF regression strategy.

From the definition of RIF, one can compute the parameter of interest in equation for the three cases described previously. Let $m_v(x) = E[RIF(Y, v)|X = x]$ for a given functional $v = v(F_Y)$. Using the same derivation strategy as in Section 2.1 above, it is straightforward to compute that the estimand of interest is

$$\Pi_{v,h} = E\left[\left.\frac{\partial m_v(x)}{\partial x}\right|_{x=X} \left.\frac{\partial h(X,\delta)}{\partial \delta}\right|_{\delta=0}\right]. \tag{7}$$

The estimand is the product of the derivative of two functions. First the $m_v(x)$ function, which depends on the functional of interest, and second, of $h(x,\delta)$ which depends on the type of counterfactual policy. The formula in (7) can then be applied to each particular effect that the user is trying to analyze.

- Location shift

  $$\Pi_{v,h} = E\left[\left.\frac{\partial m_v(x)}{\partial x}\right|_{x=X}\right].$$

  This is indeed the estimand of Firpo et al. (2009) and the most popular amongst RIF regression empirical applications.
- Location–scale shift

  $$\Pi_{v,h} = E\left[\left.\frac{\partial m_v(x)}{\partial x}\right|_{x=X}\left(-(X-\mu_X)\frac{s'(0)}{s^2(0)} + \ell'(0)\right)\right].$$

  As in Martínez-Iriarte et al. (2024) we can define the location shift effect as

  $$\Pi_v^L := E\left[\left.\frac{\partial m_v(x)}{\partial x}\right|_{x=X} \ell'(0)\right],$$

  which corresponds to the Firpo et al. (2009) RIF regression coefficient representation, and the scale effect is

  $$\Pi_v^S := E\left[-\left.\frac{\partial m_v(x)}{\partial x}\right|_{x=X} (X-\mu_X)\frac{s'(0)}{s^2(0)}\right],$$

  such that $\Pi_{v,h} = \Pi_v^L + \Pi_v^S$.
- Asymmetric shift

  $$\Pi_{v,h} = E\left[\left.\frac{\partial m_v(x)}{\partial x}\right|_{x=X} a'(0)(X_{max} - X)^\lambda\right]. \tag{8}$$

2.3. Empirical Examples to Motivate the Estimands

Here we discuss some empirical examples where these models could be used.

Effect of increasing education on wage inequality. In a Mincer equation, log wages are modeled as a function of certain observable covariates such as years of education. Changes in education levels have different effects on income inequality and a positive shift may result in augmenting it, the so-called 'paradox of progress' (see Bourguignon et al., 2024). A study of the effect of a shift in education on wage inequality could be implemented using our proposed framework. We can accommodate a counterfactual policy experiment where there may be not only a general increase in the education level but also a change in its dispersion or an asymmetric shift towards the highest possible level of education.

Smoking and birth weight. Consider a cigarette consumption tax. Under some assumptions, the consumption X will be reduced to $X/(1+\delta)$ for some δ that depends on the elasticities of the supply and demand. One might wonder about the final impact of this tax policy on the distribution of birth weights. This is analyzed in Martínez-Iriarte et al. (2024).

Wage controls and earnings distribution Vickers and Ziebarth (2022) studies the effect of a more uniform (less dispersion) distribution of wage control brackets on the distributions of earnings, in the context of a policy that was implemented during World War II.

3. Generalized RIF Estimator

The proposed estimands can be estimated using the RIF regression method. This is implemented in three steps.

First, we need to specify a model for the $m_v(x)$ function, that is, the relationship between the RIF of the v functional together with an ad hoc model on how it relates to the covariates, i.e., $\frac{\partial m_v(x)}{\partial x}$. Let $\widehat{\frac{\partial m_v(x)}{\partial x}}$ be such estimator. Assume a parametric model

$$\frac{\partial m_v(x)}{\partial x} \equiv g_1(x, \eta_0)$$

for some parameter η with true value at η_0. The estimator is then $g_1(x_i, \hat{\eta})$, where $\hat{\eta}$ is a consistent estimator of η. Firpo et al. (2009) propose several alternatives, among which RIF OLS and RIF Logit are the most commonly used. Similar procedures can be applied to any functional, as is typically done in the applied literature. In turn, this requires that the proposed model is an appropriate representation of the true model. η can then be interpreted as the parameters from the RIF regression model and $\hat{\eta}$ as the corresponding estimator.

For example, a third-degree polynomial is a popular choice:

$$m_v(x) = \eta_0 + \eta_1 x + \eta_2 x^2 + \eta_3 x^3. \tag{9}$$

Second, depending on the proposed $h(X, \delta)$ model different alternatives arise as show in Section 2.2. This depends on the policy intervention of interest and the associated shift in covariates. Then, we can compute $h'_\delta(x, 0)$ for each case. This derivative may also include estimated parameters. Let $\widehat{h'_\delta(x, 0)}$ be such estimator. Assume a parametric model

$$h'_\delta(x, 0) \equiv g_2(x, \gamma_0)$$

for some parameter γ with true value at γ_0 and that $\hat{\gamma}$ is a consistent estimator of γ. Thus, the estimator is $g_2(x_i, \hat{\gamma})$.

Third, the unconditional effects can then be obtained by sample averages of the product formed by the two elements in the last paragraphs.

$$\widehat{\Pi}_{v,h} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\frac{\partial m_v(x_i)}{\partial x}} \widehat{h'_\delta(x_i, 0)} = \frac{1}{n}\sum_{i=1}^{n} g_1(x_i, \hat{\eta}) g_2(x_i, \hat{\gamma}). \tag{10}$$

Now define $g(x, \alpha) \equiv g_1(x, \eta) g_2(x, \gamma)$ for $\alpha = (\eta, \gamma)$. Consistency of $\widehat{\Pi}_{v,h}$ will follow from a uniform law of large numbers over α, and the correct specification of the parametric model for $m_v(x)$. For a uniform law of large numbers, sufficient conditions can be found in Lemma 4.3 in Newey and McFadden (1994).

Assumption 1. *Assume that* $\frac{\partial m_v(x)}{\partial x}\Big|_{x=X} h'_\delta(X, 0) = g(X, \alpha_0)$ *is continuous at* α_0 *with probability one, and there is a neighborhood* $\tilde{A}_0$ *of* α_0 *such that* $E\left[\sup_{\alpha \in \tilde{A}_0} \|g(X, \alpha)\|\right] < \infty$.

The correct specification of the model for $m_v(x)$ is quite challenging. Indeed, it is unlikely that a parametric specification is the correct one. However, one can view a specification like the one in (9) as a series approximation to the true model. A rigorous treatment is provided in Firpo et al. (2009) for the case $h'_\delta(X, 0) \equiv 1$.

Deriving the asymptotic normality of our estimators is more complex and requires a more intricate analysis. If the influence function were observed, it would exhibit the standard parametric asymptotic convergence rate, i.e., $\sqrt{n}$, as this follows from a standard OLS (or Logit) analysis. However, in most cases, the IF must be estimated and often

involves nonparametric components, such as densities. As noted by Firpo et al. (2009, p. 962), "because the density is nonparametrically estimated by kernel methods, the rate of convergence of the three estimators will be dominated by this slower term". Thus, we necessitate a case-by-case asymptotic analysis for different estimators. Firpo and Pinto (2016) derive the asymptotic distribution for the functionals used here (Gini, Theil, Atkinson), and the asymptotic distribution of our proposed estimators can, in principle, be obtained via the delta method. For the UQR case, the Supplemental Appendix in Firpo et al. (2009) provides a detailed example for the location–shift effect, which can be extended to any shift function of X. A specific extension is developed by Martínez-Iriarte et al. (2024), albeit without relying on RIF methods.

Thus, while the asymptotic normality of our proposed estimators can be derived, it involves additional technical challenges, and we leave this as an avenue for further research. In our implementation, we conduct inference using the wild bootstrap, a common approach in RIF regression methods.

4. Monte Carlo Experiments

Consider the Gini coefficient as our distributional target of interest v and the following data generating process (DGP):

$$Y_i = 3 + 2X_i + (1 + \theta X_i)u_i \tag{11}$$

with $i = 1, \ldots, n$ where $X_i \sim N(6, 1)$ and u_i is a random variable that is independent of X_i. The chosen value parameters will exclude $y < 0$. See the Appendix A for the details on the Gini index computation together with the RIF.

Then for each DGP we compute Π_v^δ by a simulation with $n = 10$ million sample size. The θ parameter allows for the presence of scale shifts in the effect of the X covariate in the linear model (11). In all cases, we replace X_δ in Equation (11) thus obtaining Y_δ and computing numerically $\Pi_v = [v(F_{Y_\delta}) - v(F_Y)]/\delta$ for $\delta = 0.0001$.

We consider the following cases:

- Pure location shift: $s(\delta) = 1$ and $\ell(\delta) = \delta$.
- Pure scale shift: $s(\delta) = 1 + \delta$ and $\ell(\delta) = 0$.
- Location–Scale shift: $s(\delta) = 1 + \delta$ and $\ell(\delta) = \delta$.
- Asymmetric shift: $a(\delta) = \delta$ and $\lambda = \{-0.5, 0, 0.5\}$. Here we set $X_{max} = 11.31$ such that no values in the simulations will exceed this.

For the RIF regression, we use a third-degree polynomial such that

$$m_v(x) = \eta_0 + \eta_1 x + \eta_2 x^2 + \eta_3 x^3$$

thus,

$$\frac{\partial m_v}{\partial x}(x_i) = \eta_1 + 2\eta_2 x_i + 3\eta_3 x_i^2$$

for $i = 1, \ldots, n$.

Then, for the location and scale shifts we have

$$\widehat{\Pi}_v^L = n^{-1} \sum_{i=1}^{n} \widehat{\frac{\partial m_v}{\partial x}}(x_i) = \hat{\eta}_1 + 2\hat{\eta}_2 \overline{x} + 3\hat{\eta}_3 \overline{x^2},$$

and

$$\widehat{\Pi}_v^S = n^{-1} \sum_{i=1}^{n} -(x_i - \overline{x}) \widehat{\frac{\partial m_v}{\partial x}}(x_i),$$

with $\overline{x} = n^{-1}\sum_{i=1}^{n} x_i$; $\overline{x^2} = n^{-1}\sum_{i=1}^{n} x_i^2$ and the η parameters are estimated by OLS as in the RIF regression method.

Then, for the asymmetric shift we use

$$\widehat{\Pi}_v = n^{-1}\sum_{i=1}^{n}(X_{max} - x_i)^{\lambda}\widehat{\frac{\partial m_v}{\partial x}}(x_i).$$

For the simulations, we consider the cases of $u_i \sim N(0,1)$ (Table 1) and $u_i \sim (\chi_1^2 - 1)/2$ (Table 2), using $\theta \in \{0, 0.3\}$, thus allowing for location-only and location–scale effects in (11).

Table 1. Estimators' finite sample performance (v = Gini index; $u_i \sim N(0,1)$).

Effect	n	$\theta = 0$			$\theta = 0.3$		
		Bias	**Var**	**MSE**	**Bias**	**Var**	**MSE**
Location	50	−0.0044	0.7563	0.7564	−0.1031	2.9639	2.9745
	100	0.0081	0.3482	0.3483	−0.0416	1.3842	1.3860
	500	−0.0086	0.0547	0.0548	−0.0260	0.2176	0.2183
	1000	−0.0066	0.0299	0.0299	−0.0184	0.1153	0.1157
	5000	−0.0005	0.0062	0.0062	0.0020	0.0252	0.0252
Scale	50	−0.1141	2.2196	2.2326	0.0156	6.9623	6.9625
	100	−0.1114	1.0436	1.0560	0.0144	3.3553	3.3555
	500	−0.0601	0.2052	0.2088	−0.0307	0.6320	0.6329
	1000	−0.0418	0.1078	0.1096	−0.0195	0.3407	0.3411
	5000	−0.0088	0.0205	0.0206	−0.0039	0.0606	0.0606
Both	50	−0.1185	3.5751	3.5891	−0.0876	10.4689	10.4766
	100	−0.1033	1.7181	1.7288	−0.0273	5.2379	5.2387
	500	−0.0688	0.2923	0.2970	−0.0568	0.8887	0.8920
	1000	−0.0485	0.1612	0.1635	−0.0379	0.4704	0.4718
	5000	−0.0094	0.0324	0.0325	−0.0019	0.0882	0.0883
Asymmetric ($\lambda = -0.5$)	50	0.0048	0.1453	0.1453	−0.0455	0.5998	0.6019
	100	0.0112	0.0658	0.0659	−0.0179	0.2760	0.2763
	500	0.0016	0.0107	0.0107	−0.0098	0.0439	0.0440
	1000	0.0019	0.0057	0.0057	−0.0066	0.0235	0.0236
	5000	0.0033	0.0012	0.0012	0.0018	0.0052	0.0052
Asymmetric ($\lambda = 0.0$)	50	−0.0044	0.7563	0.7564	−0.1031	2.9639	2.9745
	100	0.0081	0.3482	0.3483	−0.0416	1.3842	1.3860
	500	−0.0086	0.0547	0.0548	−0.0260	0.2176	0.2183
	1000	−0.0066	0.0299	0.0299	−0.0184	0.1153	0.1157
	5000	−0.0005	0.0062	0.0062	0.0020	0.0252	0.0252
Asymmetric ($\lambda = 0.5$)	50	−0.0365	4.3397	4.3410	−0.2329	16.0205	16.0748
	100	−0.0086	2.0279	2.0280	−0.0939	7.6135	7.6223
	500	−0.0350	0.3114	0.3127	−0.0665	1.1873	1.1917
	1000	−0.0270	0.1729	0.1737	−0.0473	0.6259	0.6281
	5000	−0.0059	0.0362	0.0363	0.0031	0.1357	0.1357

Note: calculations based on 1000 Monte Carlo experiments, Gini index is multiplied by 100.

All DGPs and simulation exercises show that the proposed implementation works. Bias and variance reduce monotonically as the sample size increase. Note that, as expected, the case of asymmetric shift with $\lambda = 0$ coincides with the location-shift effect. Moreover, the method works for all shifts models.

Table 2. Estimators' finite sample performance (v = Gini index; $u_i \sim (\chi^2_1 - 1)/\sqrt{2}$).

Effect	n	$\theta = 0$			$\theta = 0.3$		
		Bias	**Var**	**MSE**	**Bias**	**Var**	**MSE**
Location	50	0.0126	0.7075	0.7077	0.0170	3.8324	3.8327
	100	0.0100	0.2943	0.2944	0.0197	1.6137	1.6140
	500	0.0077	0.0566	0.0567	0.0074	0.3119	0.3119
	1000	0.0036	0.0266	0.0266	0.0085	0.1438	0.1439
	5000	0.0002	0.0057	0.0057	−0.0035	0.0298	0.0299
Scale	50	−0.1601	1.8447	1.8704	−0.0454	7.8484	7.8505
	100	−0.1201	0.9508	0.9652	−0.0340	3.9788	3.9800
	500	−0.0499	0.2079	0.2104	−0.0449	0.8343	0.8364
	1000	−0.0381	0.0933	0.0947	0.0000	0.3934	0.3934
	5000	−0.0040	0.0209	0.0209	0.0034	0.0845	0.0845
Both	50	−0.1477	2.2522	2.2741	−0.0286	5.7489	5.7497
	100	−0.1104	1.2147	1.2268	−0.0144	3.1012	3.1014
	500	−0.0424	0.2608	0.2626	−0.0375	0.6949	0.6963
	1000	−0.0347	0.1159	0.1171	0.0084	0.3021	0.3022
	5000	−0.0040	0.0265	0.0265	−0.0002	0.0645	0.0645
Asymmetric ($\lambda = -0.5$)	50	0.0137	0.1493	0.1495	0.0089	0.8879	0.8880
	100	0.0118	0.0617	0.0618	0.0109	0.3816	0.3817
	500	0.0084	0.0118	0.0119	0.0065	0.0728	0.0729
	1000	0.0063	0.0056	0.0056	0.0048	0.0342	0.0343
	5000	0.0036	0.0012	0.0012	0.0000	0.0071	0.0071
Asymmetric ($\lambda = 0.0$)	50	0.0126	0.7075	0.7077	0.0170	3.8324	3.8327
	100	0.0100	0.2943	0.2944	0.0197	1.6137	1.6140
	500	0.0077	0.0566	0.0567	0.0074	0.3119	0.3119
	1000	0.0036	0.0266	0.0266	0.0085	0.1438	0.1439
	5000	0.0002	0.0057	0.0057	−0.0035	0.0298	0.0299
Asymmetric ($\lambda = 0.5$)	50	−0.0075	3.6517	3.6517	0.0299	17.5340	17.5348
	100	−0.0053	1.5628	1.5629	0.0385	7.3357	7.3372
	500	0.0035	0.3033	0.3033	0.0060	1.4437	1.4437
	1000	−0.0031	0.1410	0.1410	0.0189	0.6523	0.6526
	5000	−0.0038	0.0307	0.0307	−0.0088	0.1349	0.1349

Note: calculations based on 1000 Monte Carlo experiments, Gini index is multiplied by 100.

5. Empirical Application

This section presents an empirical application. We use an extract from the Merged Outgoing Rotation Group of the Current Population Survey from 1983, 1984, and 1985, restricted to male respondents only. More details about the data can be found in Lemieux (2006). The variable of interest is Y, the hourly wage, and the covariates X are an indicator of whether the individual is unionized, years of education, marital status, race (non-white), and work experience. We use a cubic specification in all continuous covariates to give some flexibility to the RIF model.

The studied effects correspond to education (Table 3) and experience (Table 4) in order to simulate different interventions and their potential effects on measures of inequality. We consider the estimators of the location, scale, and asymmetric effects discussed above for the Gini, Theil, and Atkinson indices (see the Appendix A for more details on each index). All index values are multiplied by 100, and all effects correspond to a change in education/experience on the corresponding shift. Figures 1 and 2 show the modeled shift for each year of education/experience depending on the value chosen for λ.

Consider first the effect of education, as reported in Table 3. The location effect is positive, while the scale effect is negative for all cases. This determines that while a positive

shift in education levels has resulted in augmenting inequality (this is also called the 'paradox of progress' by different authors), a reduction in the overall impact of education has a decreasing effect on inequality. In fact, the location–scale joint effect is negative, thus reducing inequality. In the case of the asymmetric shift, note that hourly wages distribution is more egalitarian when the shift is biased towards lower educational levels ($\lambda > 0$), as long as inequality is measured using the Gini or Theil coefficient. However, if the index used is more inequality-averse (Atkinson(1) and Atkinson(2)), then all effects are unequalizing.

Table 3. Education effects.

Effect	Gini	Theil	Atkinson(1)	Atkinson(2)
Location	0.6086 * (0.0173)	0.5699 * (0.0218)	0.6282 * (0.0151)	1.2625 * (0.0233)
Scale	−4.8003 * (0.0824)	−4.7931 * (0.1074)	−4.1004 * (0.0714)	−6.4685 * (0.1047)
Both	−4.1918 * (0.0777)	−4.2232 * (0.1019)	−3.4722 * (0.0681)	−5.2060 * (0.1031)
Asymmetric ($\lambda = -0.5$)	0.3303 * (0.0086)	0.3111 * (0.0109)	0.3347 * (0.0075)	0.6600 * (0.0115)
Asymmetric ($\lambda = 0$)	0.6086 * (0.0173)	0.5699 * (0.0218)	0.6282 * (0.0151)	1.2625 * (0.0233)
Asymmetric ($\lambda = 0.5$)	−0.1681 * (0.0383)	−0.2530 * (0.0494)	0.0941 * (0.0343)	0.7458 * (0.0563)

Notes: the sample size is 266,956 observations; bootstrap standard errors (500 replications) in parentheses, * for $p < 0.01$.

Consider now the effect of experience in Table 4. In this case, the effects are clearly negative for both location and scale effects. In fact, in all cases there is a reducing effect on inequality measured by all indices. In the case of asymmetric change, the equalizing effect is greater when the increase in experience is biased towards lower levels ($\lambda > 0$).

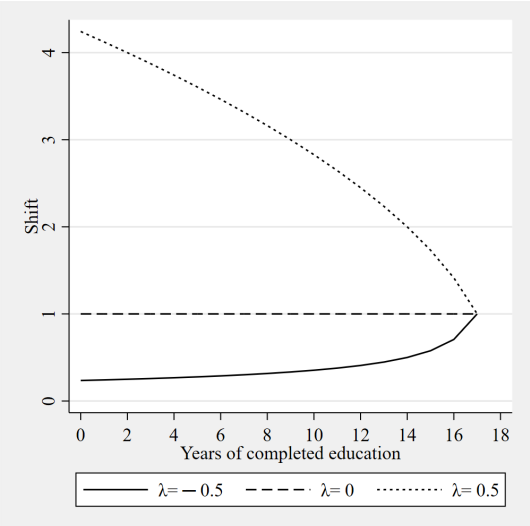


Figure 1. Asymmetric shift, education.

Table 4. Experience effects.

Effect	Gini	Theil	Atkinson(1)	Atkinson(2)
Location	−0.4025 * (0.0059)	−0.3963 * (0.0068)	−0.3291 * (0.0053)	−0.4686 * (0.0092)
Scale	−4.5253 * (0.0929)	−4.5701 * (0.1259)	−3.7068 * (0.0826)	−5.2172 * (0.1267)
Both	−4.9278 * (0.0959)	−4.9664 * (0.1292)	−4.0358 * (0.0853)	−5.6858 * (0.1314)
Asymmetric ($\lambda = -0.5$)	−0.0620 * (0.0010)	−0.0607 * (0.0011)	−0.0512 * (0.0009)	−0.0744 * (0.0015)
Asymmetric ($\lambda = 0$)	−0.4025 * (0.0059)	−0.3963 * (0.0068)	−0.3291 * (0.0053)	−0.4686 * (0.0092)
Asymmetric ($\lambda = 0.5$)	−2.8433 * (0.0410)	−2.8099 * (0.0482)	−2.3207 * (0.0370)	−3.2877 * (0.0626)

Notes: the sample size is 266,956 observations; bootstrap standard errors (500 replications) in parentheses, * for $p < 0.01$.

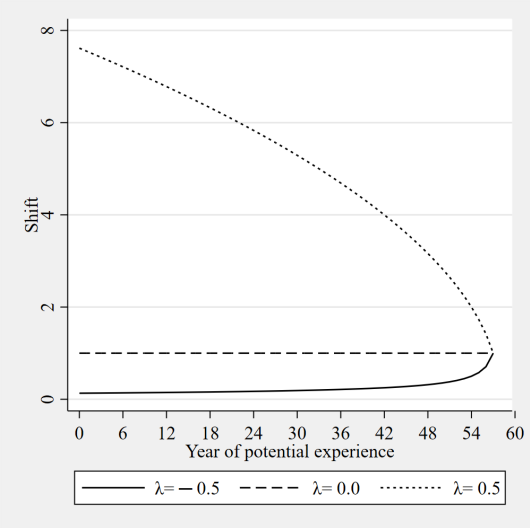


Figure 2. Asymmetric shift, experience.

6. Concluding Remarks

This paper introduces a generalized framework for analyzing the distributional impacts of covariate shifts on inequality and other distributional measures. Building on the RIF regression approach originally proposed by Firpo et al. (2009) for location shifts, our method extends its applicability by incorporating a broader class of shifts, following a framework similar to Martínez-Iriarte et al. (2024). Our Monte Carlo simulations demonstrate that the proposed approach performs well across a wide range of scenarios, highlighting its robustness and versatility.

Several avenues for future research emerge from this work. First, more precise asymptotic approximations could be developed by explicitly deriving the asymptotic distribution of functional estimators of influence functions. Additionally, alternative numerical and approximation methods could be explored for cases where closed-form solutions are unavailable. Second, the framework could be extended to accommodate calibrated covariate shifts that align with empirical policy interventions. For example, any targeted policy intervention could be translated into a counterfactual covariate distribution, to which our method could be applied. Finally, Oaxaca–Blinder-type decompositions could be formu-

lated for specific types of shifts, allowing for a broader range of decomposition analyses and further enhancing the interpretability of distributional changes.

Author Contributions: Conceptualization, J.A., A.G., J.M.-I. and G.M.-R.; methodology, J.A., A.G., J.M.-I. and G.M.-R.; software, J.A., A.G., J.M.-I. and G.M.-R.; formal analysis, J.A., A.G., J.M.-I. and G.M.-R.; investigation, J.A., A.G., J.M.-I. and G.M.-R.; resources, J.A., A.G., J.M.-I. and G.M.-R.; data curation, J.A., A.G., J.M.-I. and G.M.-R.; writing—original draft preparation, J.A., A.G., J.M.-I. and G.M.-R.; writing—review and editing, J.A., A.G., J.M.-I. and G.M.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Functionals and Their RIF

Appendix A.1. Gini Index

The Gini coefficient is defined as

$$v^{\text{Gini}} = 1 - 2\frac{R(F_Y)}{\mu},$$

where $\mu = E(Y)$ and $R(F_Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} z dF_Y(z) dF_Y(y)$. Note that if we define $p = F_Y(z)$ then $R(F_Y) = \int_0^1 GL(p, F_y)dp$, that is, the term $R(F_Y)/\mu$ acquires the classical interpretation of the area under the Lorenz Curve given by the expression $GL(p, F_Y) = \int_{-\infty}^{F_Y^{-1}(p)} z dF_Y(z)$.

Following Firpo et al. (2018), and after some algebra, the RIF for the Gini index v^{Gini} can be written as:

$$RIF(y, v^{\text{Gini}}, F_Y) = 1 + \left(\frac{1 - v^{\text{Gini}}}{\mu}\right)y + \frac{2}{\mu}[y(F_Y(y) - 1) - GL(F_Y(y), F_Y)]$$

Sample Estimator

Assume a random sample $y_1, \ldots, y_n$, we use a plug-in estimator of the previous equation. First, following Lambert (2001), consider the observations ordered from smallest to largest $y_1 \leq y_2 \leq \ldots \leq y_n$, then

$$\hat{v}^{\text{Gini}} = 1 + \frac{1}{n} - \frac{2}{\mu n^2}\sum_{i=1}^{n} y_i(n + 1 - i)$$

Finally, the rest of the components are estimated as follows

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\hat{F}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} 1(y_i \leq y)$$

$$\widehat{GL}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} y_i 1(y_i \leq y)$$

Appendix A.2. Theil Index

The Theil index is defined as

$$v^{\text{Theil}} = \frac{\eta}{\mu} - ln(\mu)$$

where $\eta = \int_{-\infty}^{\infty} y \ln(y) dF_Y(y)$ and $\mu = E(Y)$.

Using this notation, the RIF of the Theil is:

$$RIF(y, v^{\text{Theil}}, F_Y) = v^{\text{Theil}} + \frac{y \ln(y) - \eta}{\mu} - \left(\frac{\eta + \mu}{\mu^2}\right)(y - \mu)$$

Sample Estimator

Assume a random sample $y_1, \ldots, y_n$, the statistics to implement the RIF as a plug-in estimator are:

$$\hat{\eta} = \frac{1}{n} \sum_{i=1}^{n} y_i \ln(y_i)$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Appendix A.3. Atkinson Index

The Atkinson index is defined as

$$v^{\text{Atk}}(\epsilon) = \begin{cases} 1 - \frac{\kappa(\epsilon)^{\frac{1}{1-\epsilon}}}{\mu} & \text{if} \quad \epsilon \neq 1; \\ 1 - \frac{e^{\kappa(\epsilon)}}{\mu} & \text{if} \quad \epsilon = 1, \end{cases}$$

where $\mu = E(Y)$, $\kappa(\epsilon) = \int_{-\infty}^{\infty} y^{(1-\epsilon)\omega} \ln(y)^{(1-\omega)} dF_Y(y)$, $\omega = 1(\epsilon \neq 1)$ and ϵ is an inequality aversion parameter. The parameter ϵ defines the type of social preferences $W = W(y_1, \ldots, y_n)$ and is a value set by the researcher. The two extreme cases are $\epsilon = 0$, when the inequality index reflects a utilitarian welfare function ($W = \mu$) while if $\epsilon \to \infty$ it is a Rawlsian function (W is a Leontief type function).

The RIF for the Atkinson index is:

$$RIF(y, v^{\text{Atk}}(\epsilon), F_Y) = v^{\text{Atk}}(\epsilon) + A(\epsilon, y) + B(\epsilon)(y - \mu)$$

where

$$A(\epsilon, y) = \begin{cases} \frac{\kappa^{\frac{\epsilon}{1-\epsilon}}}{(\epsilon-1)\mu}(y^{1-\epsilon} - \kappa) & \text{if} \quad \epsilon \neq 1; \\ -\frac{e^{\kappa}}{\mu}(\ln(y) - \kappa) & \text{if} \quad \epsilon = 1, \end{cases}$$

and

$$B(\epsilon) = \begin{cases} \frac{\kappa^{\frac{1}{1-\epsilon}}}{\mu^2} & \text{if} \quad \epsilon \neq 1; \\ \frac{e^{\kappa}}{\mu^2} & \text{if} \quad \epsilon = 1, \end{cases}$$

Sample Estimator

Assume a random sample $y_1, \ldots, y_n$, the statistics to implement the RIF as a plug-in estimator are:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$\hat{k}(\epsilon) = \frac{1}{n} \sum_{i=1}^{n} y_i^{(1-\epsilon)\omega} \ln(y_i)^{(1-\omega)}$$

where $\omega = 1(\epsilon \neq 1)$.

Notes

[1] For an introduction to influence functions, see van der Vaart (1998).

[2] Martinez-Iriarte (2024) develops a sensitivity analysis procedure that accounts for both marginal and non-marginal (global) effects on unconditional quantiles, specifically when covariates are discrete.

[3] Various methods exist for estimating unconditional quantile effects (UQEs). Indeed, Firpo et al. (2009) rigorously derive three distinct estimation methods.

References

Alejo, J., Galvao, A. F., Martinez-Iriarte, J., & Montes-Rojas, G. (2024). Unconditional quantile partial effects via conditional quantile regression. *Journal of Econometrics*, 105678. [CrossRef]

Battiston, D., Garcia-Domench, C., & Gasparini, L. (2014). Could an increase in education raise income inequality? Evidence for Latin America. *Latin American Journal of Economics*, *51*(1), 1–39. [CrossRef]

Bourguignon, F., Lustig, N., & Ferreira, F. (2024). *The microeconomics of income distribution dynamics*. Oxford University Press.

Chernozhukov, V., Fernández-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*(6), 2205–2268. [CrossRef]

Essama-Nssah, B., & Lambert, P. (2012). Influence functions for policy impact analysis. In J. Bishop, & R. Salas (Eds.), *Inequality, mobility and segregation: Essays in honor of jacques silber research on economic inequality* (Vol. 20, pp. 135–159). Emerald Group Publishing.

Firpo, S., Fortin, N., & Lemieux, T. (2009). Unconditional quantile regression. *Econometrica*, *77*(3), 953–973.

Firpo, S., Fortin, N., & Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics*, *6*(2), 28. [CrossRef]

Firpo, S., & Pinto, C. (2016). Identification and estimation of distributional impacts of interventions using changes in inequality measures. *Journal of Applied Econometrics*, *31*(3), 457–486. [CrossRef]

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In O. Ashenfelter, & D. Card (Eds.), *Handbook of labor economics* (Vol. 4, pp. 1–12). Elsevier.

Inoue, A., Li, T., & Xu, Q. (2021). Two sample unconditional quantile effect. *arXiv*, arXiv:2105.09445.

Lambert, P. (2001). *The distribution and redistribution of income*. Manchester University Press.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review*, *96*(3), 461–498. [CrossRef]

Martinez-Iriarte, J. (2024). Sensitivity analysis in unconditional quantile effects. *arXiv*, arXiv:2105.09445.

Martínez-Iriarte, J., Montes-Rojas, G., & Sun, Y. (2024). Unconditional effects of general policy interventions. *Journal of Econometrics*, *238*(2), 105570. [CrossRef]

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle, & D. L. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp. 2111–2245). Elsevier.

Oaxaca, R. L. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693–709. [CrossRef]

Oaxaca, R. L., & Ransom, M. R. (1994). On discrimination and the decomposition of the wage differentials. *Journal of Econometrics*, *61*, 5–21. [CrossRef]

Pagan, A., & Ullah, A. (1999). *Nonparametric econometrics*. Cambridge University Press.

Sasaki, Y., Ura, T., & Zhang, Y. (2022). Unconditional quantile regression with high-dimensional data. *Quantitative Economics*, *13*(3), 955–978. [CrossRef]

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge University Press.

Vickers, C., & Ziebarth, N. L. (2022). *The effects of the national war labor board on labor income inequality.* Forthcoming American Economic Review. Available online: https://www.aeaweb.org/conference/2022/preliminary/paper/9B7GT7BY (accessed on 1 April 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.