RedNIE Red Nacional de Investigadores en Economía

RIF regression via sensitivity curves

Javier Alejo (IECON/Universidad de la República)
Gabriel Montes-Rojas (IIEP-BAIRES/Universidad de Buenos Aires/CONICET)
Walter Sosa Escudero (Universidad de San Andrés/CONICET)

DOCUMENTO DE TRABAJO 2021-41

Marzo de 2021

Citar como:
Alejo, Javier, Gabriel Montes-Rojas y Walter Sosa Escudero (2021). RIF regression via sensitivity curves. *Documento de trabajo RedNIE, 2021-41*.

RIF Regression via Sensitivity Curves*

Javier Alejo†
Gabriel Montes-Rojas‡
Walter Sosa-Escudero§

January 18, 2021

Abstract

This paper proposes an empirical method to implement the recentered influence function (RIF) regression of Firpo, Fortin and Lemieux (2009), a relevant method to study the effect of covariates on many statistics beyond the mean. In practically relevant situations where the influence function is not available or difficult to compute, we suggest to use the *sensitivity curve* (Tukey, 1977). We illustrate the proposal with an application to the polarization index of Duclos, Esteban and Ray (2004).

JEL classification: J01, J31
Keywords: recentered influence function, sensitivity, polarization

*We thank Ezequiel Smucler for useful references. Errors and omissions are our responsibility.

†IECON-Universidad de la República, Gonzalo Ramírez 1926, C.P. 11200, Montevideo, Uruguay. E-mail: javier.alejo@ccee.edu.uy

‡IIEP-BAIRES - Universidad de Buenos Aires and CONICET. Email: gabriel.montes@fce.uba.ar

§Universidad de San Andrés and CONICET, Buenos Aires, Argentina. Email: wsosa@udesa.edu.ar

1 Introduction

The *recentered influence function* (RIF) regression, as proposed by Firpo, Fortin and Lemieux (2009), is a powerful tool to study the impact of changes in covariates on the unconditional distribution of a given outcome variable. Concretely, let Y be a random variable with cumulative distribution function F, and $\nu(F)$ any 'functional' of interest related to F. For example, if Y is income, $\nu(F)$ can be the mean, the Gini index, a quantile, or the poverty rate. The RIF is defined as $\mathsf{RIF}(y, \nu, F) = \nu(F) + \mathsf{IF}(y, \nu, F)$, where $\mathsf{IF}(y, \nu, F)$ is the *influence function* (IF) (Hampel, 1974) that measures the marginal impact of a particular data point in the support of F in the value of $\nu(F)$. Influence funtions play a key role in the robust statistics literature.

Firpo et al. (2009, 2018) note that since $\mathsf{E}[\mathsf{RIF}(Y, \nu, F)] = \nu(F)$, by the law of iterated expectations $\mathsf{E}_Y\left[\mathsf{E}_{Y|X}\mathsf{RIF}(Y, \nu, F)\right] = \nu(F)$, and show that the effect on $\nu(F)$ that arises from shifting a scalar covariate from X to $X + t$, where $t \downarrow 0$, is given by:

$$\int \frac{d\mathsf{E}[\mathsf{RIF}(Y, \nu, F)|X = x]}{dx} dF(x).$$

Hence, by properly modelling $\mathsf{E}[\mathsf{RIF}(Y, \nu, F)|X = x]$ in a regression fashion, the effect of X on ν can be recovered as an 'average derivative' of regressing $\mathsf{RIF}(Y, \nu, F)$ on X. The implementation of the method requires to construct $\mathsf{RIF}(Y, \nu, F)$ analytically for the functional of interest ν and then

regress it on X. In many relevant cases the influence function required to obtain $\mathsf{RIF}(\mathsf{Y}, \nu, \mathsf{F})$ is immediately available; Fortin, Lemieux and Firpo (2011) present a useful 'catalog' that includes the mean, the quantiles, the variance and the Gini index (see also Essama-Nssah and Lambert (2015) and Cowell and Flachaire (2015)).

In this paper we propose a practical computation method based on the *sensitivity curve* (Tukey, 1977). This procedure consists in comparing the full sample functional ν with that computed when the $\mathsf{j}-$th observation is left out; this is the influence of this particular observation on the empirical version of ν. The relevance of the proposed strategy derives from the fact that, under general conditions, the sensitivity curve (SC) converges in probability to the IF (see Nasser and Alam (2006) for a discussion). We provide an intuitive proof of this result.

The SC has some practical advantages over the IF. First, even when analytically available, in many cases the estimation of the IF involves dealing with the problem of selection of the meta-parameters, like bandwidths, which may add further complications. Second, in some relevant cases the IF may be difficult when not impossible to derive analytically. Finally, many relevant examples where the IF can be easily derived involve additive or quasi-additive measures that do not apply to many important situations.

This paper is organized as follows. Section 2 presents the main statistical derivations. Section 3 discusses an empirical exercise that shows that the

performance of the SC is close to that of the analytical IF. Then we implement the method to the case of the *polarization index* of Duclos, Esteban and Ray (2004), where the analytic IF is not available.

2 Influence via sensitivity curves

Let $\nu(F)$ a real-valued functional, where $\nu : \mathcal{F}_\nu \to \mathbb{R}$ and $\mathcal{F}_\nu$ is a class of distribution functions such that $F \in \mathcal{F}_\nu$ if $|\nu(F)| < \infty$. Consider two cummulative distribution functions, F and G, and let $H_{t,F,G} = tG + (1-t)F$, $t \in [0,1]$. Then, using the Von Mises (1947) expansion:

$$\nu(H) = \nu(F) + t\partial\nu\left(H_{t,F,G}\right)/\partial t\,|_{t=0} + r(t,F,G) \tag{1}$$

with

$$\begin{aligned}\partial\nu\left(H_{t,F,G}\right)/\partial\, t|_{t=0} &= \lim_{t\downarrow 0}\frac{\nu\left(H_{t,F,G}\right)-\nu(F)}{t}\\ &= \int \psi(y)d(G-F)(y).\end{aligned} \tag{2}$$

When $G = \Delta_y$ and Δ_y is the CDF of a random variable with probability mass of 1 at y, $\psi(y) = \partial\nu\left(H_{t,F,\Delta_y}\right)/\partial\, t|_{t=0}$ is the *influence function (IF)* of the functional ν, labeled as $IF(y,\nu,F)$ (see Huber and Ronchetti (2009) for a general discussion; here we are following the derivation in Firpo et al. (2009, p.956)).

Consider now the last term in (1). Following Von Mises (1947):

$$r(t,F,G) = \frac{\tilde{t}^2}{2}\partial^2\nu\left(H_{t,F,\Delta_y}\right)/\partial\, t^2\big|_{t=0}, \tag{3}$$

for some $\tilde{t} \in [0, t]$, where

$$\partial^2 \nu \left(H_{t,F,G} \right) / \partial t^2 \mid_{t=0} = \iint \phi(y, z) d(G - F)(y) d(G - F)(z) \tag{4}$$

with $\phi(y, z)$ a symmetric function; again, see Von Mises (1947, p. 325) for details. Note that if $\nu(F) = \nu(cF)$ for all $c > 0$ (scale invariance) then:

$$\begin{array}{ll} \text{(i)} & \int \psi(y) dF(y) = 0 \\ \text{(ii)} & \iint \phi(y, z) dF(y) dF(z) = 0 \end{array}$$

The proof of (i) and (ii) follows from Jaeckel (1972).[1]

The *recentered influence function (RIF)*, is defined as $RIF(y, \nu, F) \equiv \nu(F) + IF(y, \nu, F)$, where, trivially, $E[IF(y, \nu, F)] = \nu(F)$, from property (i) above. Firpo et al. (2009) develop a RIF-regression framework that is similar to a standard regression except that the dependent variable, Y, is replaced by the IF of the statistic of interest, which allows to estimate the effects of covariates X on $\nu(F)$.

Unfortunately, not all indicators have an IF with a specific analytical form and thus the RIF-regression may not be practically feasible. Our proposal consists of replacing the IF by the SC.

Let $\{y_i\}_{i=1}^n$ be an iid sample and define $\nu_n = \nu(F_n)$ as the sample coun-

[1] Let $G = 2F$, then $H_{(t,F,G)} = (1 + t)F = cF$ and then by the invariance to scale

$$\partial \nu \left(H_{t,F,G} \right) / \partial \, t|_{t=0} = \lim_{t \downarrow 0} \frac{\nu(cF) - \nu(F)}{t} = \lim_{t \downarrow 0} \frac{0}{t} = 0.$$

Moreover,

$$\partial^2 \nu \left(H_{t,F,\Delta_y} \right) / \partial \, t^2 \big|_{t=0} = 0.$$

terpart of $\nu(F)$, and let $\nu_n^{(j)} = \nu(F_n^{(j)})$ denote the case where $j-$th observation is left out, then:

$$\begin{aligned} F_n(y) &= \frac{1}{n}\sum_{i=1}^{n} 1\left(y_i \leqslant y\right) \\ F_n^{(j)}(y) &= \frac{1}{n-1}\sum_{i\neq j} 1\left(y_i \leqslant y\right). \end{aligned}$$

The *sensitivity curve (SC)* is defined as

$$SC\left(y_j, \nu_n, F_n\right) \equiv n \cdot \left[\nu\left(F_n\right) - \nu\left(F_n^{(j)}\right)\right]. \tag{5}$$

The key property that links the IF to the SC is the following:

Proposition 1. *Assume that* $\nu(F)$ *is twice continuously differentiable with respect to* F *and* $\psi(y)$ *and* $\phi(y,z)$ *exist, and that* ν *is invariant to scale (i.e.,* $\nu(F) = \nu(cF)$ *for* $c > 0$*). Then,* $SC\left(y_j, \nu_n, F_n\right) \xrightarrow{p} IF\left(y_j, \nu, F\right)$ *as* $n \to \infty$*.*

Proof. See the Appendix. □

Consequently, if the functional ν is smooth enough, the SC can be used instead of the analytical IF. Nasser and Alam (2006) show that Fréchet differentiability is sufficient for consistency. Of course smoothness is a strong requirement. For example, for the case of quantiles $IF(y, Q_\tau, F) = (y - 1[y \leqslant Q_\tau(F)])/f_y(Q_\tau(F))$, where $1[.]$ is an indicator function, $f_y(Q_\tau(F))$ is the density of the marginal distribution of y evaluated at the τ-quantile, and $Q_\tau(F)$ is the population τ-quantile of the unconditional distribution of y. The indicator function makes it non twice differentiable.

The *recentered sensitivity curve (RSC)* is defined as:

$$\mathsf{RSC}(\mathsf{y}_j, \nu_n, \mathsf{F}_n) \equiv \nu_n + \mathsf{SC}(\mathsf{y}_j, \nu_n, \mathsf{F}_n)$$

Trivially $\mathsf{RSC} \xrightarrow{p} \mathsf{RIF}$. Hence, our proposal is to replace RIF with RSC.

3 Empirical illustration

This section presents empirical applications. We first compare the empirical performance of RIF and RSC for the variance and the Gini index, for which the IF can be obtained analytically. Then we add the DER polarization index (Duclos, Esteban, and Ray, 2004) where an explicit analytical closed-form solution for IF is not available.

We use an extract from the Merged Outgoing Rotation Group of the Current Population Survey of 1983, 1984 and 1985 for males 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, whether he is married, non-white, his experience and its square.

Obtaining the RSC for each observation using the leave-one-out method can be computationally intensive if $\mathfrak{n}$ is too large since it requires a separate calculation for each observation. Therefore, we also consider computing the RSC by intrapolating an estimated spline using 1000 random points in the distribution of Y (this is denoted as RSC(sp)).

Table 1 shows results for the variance and the Gini index. Remarkably, the differences between the RIF and RSC regressions are negligible. Interestingly, the approximation obtained through the spline intrapolation seems to be accurate, suggesting that it is a convenient computational strategy relative to the leave-one-out method.

[INSERT TABLE 1 HERE]

Polarization is an important welfare concept in economics and political science. Intuitively, it measures the tension between individuals in a society, that depends positively on how distant individuals are between groups (alienation) and how close they are within a group (identification). From this perspective, a standard measure of inequality like the Gini index focuses on just the first component. Duclos et al. (2004) provide a full axiomatic framework that leads to a logically coherent measure of polarization. For a detailed empirical study on polarization for the case of Latin America and the Caribbean, see Gasparini et al. (2008).

Let $y_1, y_2, \ldots, y_n$ be and iid sample of incomes, ordered from lowest to highest. Duclos et al. (2004) propose the following empirical measure of polarization:

$$P_\alpha = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(y_i)^\alpha \hat{a}(y_i)$$

where $\hat{a}(y_i) = \hat{\mu} + y_i \left(n^{-1}(2i-1) - 1\right) - n^{-1} \left(2 \sum_{j=1}^{i-1} y_j + y_i\right)$, $\hat{\mu}$ is the

sample mean and $\hat{f}(y_i)$ is an estimate of the density of incomes. The parameter α is set exogenously and plays a key role in characterizing polarization. As a matter of fact, when $\alpha = 0$ polarization reduces to the Gini index. Larger values of α result in the index giving relatively more importance to identification, that is, to how close individuals are 'surrounded' by others of similar income. The axiomatic approach of Duclos et al. (1994) imposes lower and upper bounds to the values α may take in practice.

Table 1 also shows results for the DER polarization indexes, for which the Gini columns correspond to a particular cse ($\alpha = 0$), and for proper polarization we set $\alpha = 0.8$ following Duclos et al. (2004). We stress the fact that the IF function is not available for this case, hence we obtain results based on the RSC solely. Again, the computationally convenient spline approximation produces similar results than when RSC is computed directly. Even though a detailed study of the effects on inequality and polarization exceeds the scope of this note, we remakr that all factors reduce both measures (i.e., higher levels education predict less *unconditionally* inequality and polarization), and that effects are stronger for inequality.

References

Cowell, F.A., Flachaire, E. 2015. Statistical Methods for Distributional Analysis. In Anthony B. Atkinson and Francois Bourguignon (eds.), Handbook of Income Distribution. Amsterdam: Elsevier.

Davies, J.B., Fortin, N.M., Lemieux, T. 2017. Wealth inequality: Theory, Measurement and Decomposition. *Canadian Journal of Economics/Revue Canadienne d'Économique* 50(5): 1224-1261.

DiNardo, J., Fortin, N.M., Lemieux, T. 2017. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica* 64(5): 1001-1044.

Duclos, J.-Y., Esteban, J., Ray, D. 2004. Polarization: Concepts, Measurement, Estimation. *Econometrica* 72(6): 1737-1772.

Essama-Nssah, B., Lambert, P.J. 2015. Chapter 6: Influence Functions for Policy Impact Analysis. In John A. Bishop and Rafael Salas (eds.), Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber, pp.135-159. Bigley, UK: Emerald Group Publishing Limited.

Firpo, S.P., Fortin, N.M., Lemieux, T. 2009. Unconditional Quantile Regressions. *Econometrica* 77(3): 953-973.

Firpo, S.P., Fortin, N.M., Lemieux, T. 2018. Decomposing Wage Distribu-

tions Using Recentered Influence Function Regressions. *Econometrics* 6(3): 41.

Fortin, N.M., Lemieux, T., Firpo, S.P. 2011. Decomposition Methods in Economics. In Orley Ashenfelter and David Card (eds.), Handbook of Labor Economics. Amsterdam: Elsevier.

Gasparini, L., Horenstein, M., Molina, E., Olivieri, S. 2008. Income Polarization in Latin America: Patterns and Links with Institutions and Conflict, *Oxford Development Studies*, 36: 461-484.

Hampel, F. 1974. The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association* 69(346): 383-393.

Huber, P., Ronchetti, E.M. 2009. *Robust Statistics* (2nd edition). Wiley.

Jaeckel, L.A. 1972. Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics* 43, 1449-1458.

Lemieux, T. 2006. Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? *American Economic Review* 96(3): 461-498.

Nasser, M., Alam, M. 2006. Estimators of Influence Function. *Communications in Statistics - Theory and Methods*, 35(1), 21-32.

Tukey, J.W. 1977. Exploratory Data Analysis, Addison-Wesley, Reading, MA.

von Mises, R. 1947. On the Asymptotic Distribution of Differentiable Statistical Functions. *Annals of Mathematical Statistics* 18(3): 309-348.

Appendix

Proof of Proposition 1.

Using eq. (1) with F_n and $F_n^{(j)}$ for the case of $t = 1$:

$$\nu(F_n) = \nu\left(F_n^{(j)}\right) + \int \psi_n(y) d\left(F_n - F_n^{(j)}\right)(y) + r\left(\tilde{t}, F_n, F_n^{(j)}\right), \qquad (6)$$

for some $\tilde{t} \in [0,1]$. Note that $\psi_n(y) = IF(y, \nu, F_n) \xrightarrow{p} \psi(y)$ by continuity of the probability limit.

Now note that $n\left[F_n - F_n^{(j)}\right] = 1(y_j < y) + O_p(1)$ because

$$\begin{aligned} F_n(y) &= \tfrac{1}{n} 1(y_j \leqslant y) + \tfrac{n-1}{n} F_n^{(j)}(y), \\ F_n(y) - F_n^{(j)}(y) &= \tfrac{1}{n} 1(y_j \leqslant y) + \tfrac{n-1}{n} F_n^{(j)}(y) - F_n^{(j)}(y), \\ F_n(y) - F_n^{(j)}(y) &= \tfrac{1}{n} 1(y_j \leqslant y) - \tfrac{1}{n} F_n^{(j)}(y). \end{aligned}$$

That is,

$$n\left[F_n(y) - F_n^{(j)}\right] = 1(y_j \leqslant y) - a_n, \qquad (7)$$

with $a_n = F_n^{(j)}(y) \xrightarrow{p} F(y)$ by the Law of Large Numbers.

Then,

$$
\begin{aligned}
n \cdot \left[v\left(F_n\right) - v\left(F_n^{(j)}\right) \right] &= \int \psi_n(y) d\left(1\left(y_j \leqslant y\right) - a_n\right)(y) + n \cdot r\left(\tilde{t}, F_n, F_n^{(j)}\right) \\
n \cdot \left[v\left(F_n\right) - v\left(F_n^{(j)}\right) \right] &= \int \psi_n(y) d\left(1\left(y_j \leqslant y\right)\right)(y) - \int \psi_n(y) d\left(a_n\right)(y) + n \cdot r\left(\tilde{t}, F_n, F_n^{(j)}\right)
\end{aligned} \tag{8}
$$

Using the fact that $1\left(y_j \leqslant y\right)$ is the Dirac function, the first term of eq. (8) is

$$
\int \psi_n(y) d\left(1\left(y_j \leqslant y\right)\right)(y) = \psi_n\left(y_j\right) \xrightarrow{p} \psi\left(y_j\right).
$$

Noting that $a_n \xrightarrow{p} F(y)$ and $\psi_n(y) \xrightarrow{p} \psi(y)$, by continuity of the probability limit, the second term of (8) becomes

$$
\operatorname{plim} \int \psi_n(y) d\left(a_n\right)(y) = \int \psi(y) dF(y) = 0,
$$

because of property (i). Then,

$$
\operatorname{plim} \int \psi(y) d\left(1\left(y_j \leqslant y\right) + a_n\right)(y) = \int \psi(y) d\left(1\left(y_j \leqslant y\right)\right)(y) = \psi(y).
$$

It remains to study the third term in (8). From (3),

$$
\begin{aligned}
n \cdot r\left(\tilde{t}, F_n, F_n^{(j)}\right) &= n \cdot \tfrac{\tilde{t}^2}{2} \iint \psi(y, z) d\left(F_n - F_n^{(j)}\right)(y) d\left(F_n - F_n^{(j)}\right)(z) \\
n \cdot r\left(\tilde{t}, F_n, F_n^{(j)}\right) &= \tfrac{1}{n} \cdot \tfrac{\tilde{t}^2}{2} \iint \psi(y, z) d\left[n\left(F_n - F_n^{(j)}\right)\right](y) d\left[n\left(F_n - F_n^{(j)}\right)\right](z)
\end{aligned}
$$

for some $\tilde{t} \in [0, 1]$. Then using (7) and property (ii),

$$
\operatorname{plim} \iint \phi(y, z) d\left[n\left(F_n - F_n^{(j)}\right)\right](y) d\left[n\left(F_n - F_n^{(j)}\right)\right](z) = \phi\left(y_j, y_j\right).
$$

Then it follows that

$$\operatorname{plim}\left\{n \cdot r\left(\tilde{t}, F_n, F_n^{(j)}\right)\right\} = \left(\operatorname{plim}\frac{1}{n}\right) \cdot \frac{\tilde{t}^2}{2}\phi\left(y_j, y_j\right) = 0.$$

Then, the result follows,

$$\operatorname{plim}\frac{\nu\left(F_n\right) - \nu\left(F_n^{(j)}\right)}{1/n} = \psi\left(y_j\right) = IF\left(y_j, \nu, F\right). \text{ QED}$$

Table 1: Wage inequality and polarization

	Variance			**Gini DER($\alpha = 0$)**			**DER($\alpha = 0.80$)**	
	RIF	**RSC**	**RSC(sp)**	**RIF**	**RSC**	**RSC(sp)**	**RSC**	**RSC(sp)**
Union	-14.95***	-14.95***	-15.76***	-6.22***	-6.22***	-6.45***	-1.43***	-1.47***
	(0.190)	(0.163)	(0.158)	(0.055)	(0.046)	(0.046)	(0.011)	(0.011)
Education	1.67***	1.67***	1.45***	-0.35***	-0.34***	-0.41***	-0.12***	-0.16***
	(0.031)	(0.037)	(0.036)	(0.009)	(0.010)	(0.010)	(0.002)	(0.002)
Experience	-0.56***	-0.56***	-0.75***	-0.62***	-0.62***	-0.66***	-0.15***	-0.17***
	(0.025)	(0.024)	(0.024)	(0.007)	(0.008)	(0.008)	(0.002)	(0.002)
Experience2	0.02***	0.02***	0.02***	0.01***	0.01***	0.01***	0.00***	0.00***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	-8.41***	-8.41***	-8.96***	-4.00***	-3.99***	-4.12***	-0.89***	-0.94***
	(0.198)	(0.194)	(0.189)	(0.058)	(0.059)	(0.059)	(0.013)	(0.013)
Non-white	1.23***	1.23***	1.70***	1.75***	1.75***	1.88***	0.46***	0.51***
	(0.262)	(0.248)	(0.245)	(0.076)	(0.079)	(0.079)	(0.017)	(0.017)
Constant	21.90***	21.90***	27.06***	31.71***	30.83***	32.98***	17.95***	18.65***
	(0.434)	(0.529)	(0.514)	(0.126)	(0.152)	(0.150)	(0.028)	(0.028)
Observations	266,956	266,956	266,956	266,956	266,953	266,956	266,953	266,956

Source: Extract from the Merged Outgoing Rotation Group of the Current Population Survey of 1983, 1984 and 1985. Notes: Standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; (sp) indicates that the RSC was estimated using a Spline with a random subsample of 1000 points; all estimates are multiplied by 100.