Decomposition Based on Expectile RIF-Regression

Abdoul Aziz Ndoye* Moctar Laghlal†

Abstract

In this study, we develop a reliable decomposition method based on the expectile RIF-regression, which is more efficient under the normality assumption than the quantile RIF-regression. The proposed approach constitutes a natural extension of the conventional Oaxaca-Blinder decomposition of means over the entire distribution. It is illustrated to document the factors that contributed to the level of income inequality in Egypt during the ten-year period preceding the 2011 revolution. The empirical results show that young heads of middle-income households in urban areas were more affected by the Egyptian economic crisis, which may explain their leading position in Tahrir Square on January 25, 2011.

Keyword: Expectile, RIF-regression, Oaxaca-Blinder decomposition.
JEL classification: C18, C51, J31.

1 Introduction

In recent years, there has been a resurgence of interest in utilizing econometrics methods to fill knowledge gaps related to sources of the recent rapid rise in inequality. The seminal Oaxaca-Blinder (OB) (Oaxaca 1973; Blinder 1973) decomposition method decomposes differences in the mean outcomes between two groups into an explained part associated with differences in observed characteristics and an unexplained part attributable to the differences in the estimated regression coefficients. This latter effect is generally considered as a measure of discrimination under certain conditions in some empirical studies on inequality. However, despite its popularity, the mean OB decomposition only provides a partial framework to describe inequalities in the

*Corresponding author: University of Orleans - LEO, email: abdoul-aziz.ndoye@univ-orleans.fr
†LEO, email: el-moctar.laghlal@univ-orleans.fr

different parts of the distribution of the outcome variable. For example, it fails when the gap between groups varies differently along the entire distribution. Several alternatives based on the conditional quantile regression models have been developed to overcome some of these limitations. However, in regression, conditional quantiles do not average up to their unconditional population counterparts as conditional means do in a classical linear regression model. Furthermore, all proposed resampling extensions involve, either several restrictive hypotheses or computational difficulties that can be avoided to perform an aggregate quantile decomposition (Oaxaca and Ransom 1994; Machado and Mata 2005; Melly 2005). A detailed quantile decomposition in the same spirit of the conventional mean OB decomposition is based on the RIF-regression (Firpo et al. 2009; Firpo et al. 2011). However, the RIF for quantile has a bounded dichotomous structure, therefore, regressing it on a set of covariates in a linear gaussian regression model is not appropriate to evaluate its conditional expectation. Moreover, the reweighing extension of Firpo et al. (2018) in a probability RIF-regression model is restrictive because the two bounds are also decomposable. Hence, going beyond the mean remains an econometric challenge to extending the conventional mean OB decomposition to the whole distribution.

In this paper, we develop a decomposition method based on the expectile RIF-regression. Introduced by Newey and Powell (1987), expectiles are reliable alternatives to quantiles related to the mean in the same way as quantiles are related to the median. Similarly to quantiles, a sequence of expectiles can describe the distribution of a variable with the mean and few expectiles above and below it. In "expectiles are quantiles", Jones (1994) shows that an expectile for a given distribution corresponds to a quantile for a related distribution. Similarly, Breckling and Chambers (1988) argued that expectiles can be understood as a particular form of M-quantiles. In regression, they could generalize the classical mean regression model. By employing the RIF for expectiles, we may benefit from the property that the unconditional expectation of the expectile RIF is the expectile itself. Furthermore, contrariwise to quantiles, the RIF for expectile is continuous and does not depend on the distribution of the outcome variable. Therefore, under appropriate regularity conditions, the one-step OLS-estimation of the αth linear expectile RIF-regression coefficients is consistent and more efficient than the two-steps estimation of the quantile RIF-regression coefficients.

We illustrate the proposed approach to document the factors that contributed to the level of income distribution inequality in Egypt during the ten-year period preceding the 2011 revolution. The empirical results demonstrate that the differences in income from 1999 to 2010 are mainly due to the return effects at each level of expectiles. The returns to higher education decreased in the upper part of the distribution and was accompanied by a positive difference in the marginal effects of age and an increasing return effect of rural. These findings demonstrate that the young educated heads of household living in urban areas being the most heavily affected by the Egyptian's economic crisis, explaining their frustration and justifying their leading position in Tahrir Square on January, 2011.

The paper is organized as follows. Section 2 provides a remind of the different decomposition methods based on RIF-regression. Section 3 describes the proposed decomposition approach. Section 4 illustrates our approach to analyze income inequality in Egypt between 1999 and 2010. Section 5 concludes the paper.

2 Notations and Background

Let y be a vector of n real-valued dependent variables characterized by a c.d.f F_y. We denote by $\nu(F_y)$[1] the distributional statistics of F_y, X the (n × K) matrix of independent regressors and $(\mathbf{y}, \mathbf{X})$ the sample of the independent realizations of (y, X). The Influence Function (IF) of $\nu(F_y)$ is a widely used concept in robust statistics introduced by Hampel (1974) to describe the influence of an infinitesimal change in the distribution of X on $\nu(F_y)$. It is defined as :

$$\lim_{\epsilon\to 0}\frac{\nu(F_{y,\epsilon.G_y})-\nu(F_y)}{\epsilon}=\frac{\partial\nu(F_{y,\epsilon.\Delta_y})}{\partial\epsilon}|_{\epsilon=0}=\int_{\mathbb{R}}\mathrm{IF}_{\nu}\,.\,\mathrm{d}\left(G_y-F_y\right) \tag{1}$$

where $F_{y,\epsilon.G_y}=(1-\epsilon)F_y+\epsilon G_y$ (see also Firpo et al. 2009). The RIF for ν is defined as $\mathrm{RIF}_\nu=\nu+\mathrm{IF}_\nu$ with the property $\mathbf{E}\left(\mathrm{RIF}_{\nu(F_y)}\right)=\nu$. The linear RIF-regression consists of regressing RIF_ν on $\mathbf{X}$, $\mathrm{RIF}_\nu=\mathbf{X}\beta+u$ where u is the error term with a null conditional expectation given $\mathbf{X}$ and β the vector of parameters which can be consistently estimated by OLS. By the law of iterated expectation, we may derive the convenient property that the unconditional distributional statistics of y average up to the unconditional mean of X, $\nu=\sum_{k=1}^{K}\overline{\mathbf{X}}_k\widehat{\beta}_k$. From this latter, the differences of the given distributional statistics between two groups A and B can be decomposed into two components as:

$$\nu_A-\nu_B=\sum_{k=1}^{K}\left(\overline{\mathbf{X}}_{Ak}-\overline{\mathbf{X}}_{Bk}\right)\widehat{\beta}_{Bk}+\sum_{k=1}^{K}\left(\widehat{\beta}_{Ak}-\widehat{\beta}_{Bk}\right)\overline{\mathbf{X}}_{Ak}. \tag{2}$$

[1] $\nu(F_y)=\mathbf{E}(y)$ for the expectation of y, $\nu(F_y)=q_\tau(F_y)$ for the τth quantile and $\nu(F_y)=e_\alpha(F_y)$ for the αth expectile of $F(y)$. We denote by ν the realization of $\nu(F_y)$: $\nu=\overline{\mathbf{y}}$ for the unconditional mean of y, $\nu=q_\tau$ for the empirical τth quantile and $\nu=e_\alpha$ for the empirical αth expectile.

The first right-hand component is attributed to the explained part associated with the difference in characteristics between the two groups given the structure of returns in the second group. The second right-hand term corresponds to the difference in returns to covariates given individual characteristics of the group of reference.

Since the the unconditional mean-RIF is $\mathbf{y}$, $\text{RIF}_{\overline{\mathbf{y}}} = \mathbf{y}$ then the decomposition based on the mean RIF-regression corresponds to the conventional mean OB decomposition. The RIF for the τth quantile developed by Firpo et al. (2009) is expressed as $\text{RIF}_{q_\tau} = c_{1,\tau}\mathbb{1}(\mathbf{y} > q_\tau) + c_{2,\tau}$ where $c_{1\tau} = 1/f_y(q_\tau)$ and $c_{2\tau} = q_\tau - (1-\tau)c_{1\tau}$. A linear τth quantile RIF-regression model would specify the conditional expectation of $\text{RIF}_{q_\tau(F_y)}$ given X as $\mathbf{X}\beta_\tau$. However, such a model fails to impose the condition that $\mathbf{E}\left(\text{RIF}_{q_\tau(F_y)}|X\right)$ belongs to $[c_{1\tau}, c_{1\tau} + c_{2\tau}]$ which must hold only if $\mathbf{X}$ is constrained. Therefore, regressing the RIF_{q_τ} on $\mathbf{X}$ in a linear gaussian regression model is not appropriate to model its conditional expectation. Moreover, the reweighing extension of Firpo et al. (2018) in a probability regression model is restrictive because $c_{1\tau}$ and $c_{2\tau}$ are also decomposable because they both depend on the probability density function of y, f_y.

In this paper, we extend the RIF-regression model for expectiles and perform a decomposition method from which all expectiles are decomposed in the same way as the mean is decomposed in the conventional OB decomposition method. Moreover, since the mean is the .5th expectile, we directly derive from our approach the conventional mean OB decomposition.

3 Decomposition based on Expectile RIF-regression

Following the description of quantiles[2], Newey and Powell (1987) characterized the αth expectile e_α as the solution of the minimization of the asymmetrically weighted mean squared deviations criterion using a continuously differentiable weighted $\mathbb{L}_2$ loss function:

$$e_\alpha = \arg\min_{\zeta\in\mathbb{R}} \mathbf{E}\left[\rho_\alpha\left(y-\zeta\right)\right] \tag{3}$$

where $\rho_\alpha(u) = u^2\left[\alpha - \mathbb{1}(u<0)\right]$. From the first-order condition of (3), the αth expectile e_α can be expressed as the weighted average of random weights varying according to the observations below it

$$e_\alpha = \mathbf{E}\left[\frac{|\alpha - \mathbb{1}(y<e_\alpha)|}{\mathbf{E}\left(|\alpha - \mathbb{1}(y<e_\alpha)|\right)}y\right].$$

[2]The τth quantile q_τ of Koenker and Bassett (1978) is the minimizer of the least absolute deviation $q_\tau = \arg\min_{\zeta\in\mathbb{R}} \mathbf{E}\left[\phi_\tau\left(\mathbf{y}-\zeta\right)\right]$ using a weighted L_1 convex loss function ρ_τ defined as $\rho_\tau(u) = u\left[\tau - 1(u<0)\right]$.

3.1 Expectile RIF-regression

The αth expectile RIF-regression consists of regressing the RIF_{e_α} on the set of covariates $\mathbf{X}$. It can be described under the normality assumption as $\text{RIF}_{e_\alpha} = \mathbf{E}\left(\text{RIF}_{e_\alpha}|X\right) + u$ where $\mathbf{E}\left(u|X\right) = 0$. It allows to evaluate the effect[3] of changes in the distribution of $\mathbf{X}$ on the αth unconditional expectile, e_α of the distribution of the outcome variable y.

Theorem 1. *The Influence Function for the αth expectile of the distribution of the outcome variable, IF_{e_α} can be expressed as*

$$IF_{e_\alpha} = \psi\left(\mathbf{y}\right) - \int_{\mathbb{R}} \psi\left(\mathbf{y}\right) dF_y \tag{4}$$

where $\psi\left(.\right)$ is a bounded linear functional defined as

$$\psi\left(\mathbf{y}\right) = \frac{1-2\alpha}{\alpha}(\mathbf{y} - e_\alpha)\mathbb{1}(\mathbf{y} < e_\alpha) + \mathbf{y}. \tag{5}$$

Lemma 1.1. *The RIF for the αth expectile is given by $RIF_{e_\alpha} = \psi\left(\mathbf{y}\right)$ in (5).*

Contrariwise to the RIF_{q_τ}, the RIF_{e_α} is a linear functional independent of the distribution of y. For the sake of decomposition, we consider a linear expectile RIF-regression specified as $\text{RIF}_{e_\alpha} = \mathbf{X}\beta_\alpha + \mathbf{u}$, where each element of the error term u is assumed to be normally, independently, and identically distributed, with zero mean and a fixed variance. Under these assumptions, the one-step OLS-estimation of the αth expectile RIF-regression coefficients given by $\widehat{\beta}_\alpha = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{RIF}_{\widehat{e}_\alpha}$ is consistent and more efficient than the two-steps estimation of the quantile RIF-regression coefficients.

3.2 Decomposition based on Expectile RIF-regression

Let the unconditional αth expectile of the outcome variable for a given group g be the distributional statistics of F_{y_g} and $\widehat{\beta}_{\alpha,g}$ the OLS coefficients of the αth expectile-RIF regression. From (2), the difference of the αth unconditional expectile of F_{y_g} between two groups A and B can be decomposed as:

$$e_{A,\alpha} - e_{B,\alpha} = \sum_{k=1}^{K}\left(\overline{\mathbf{X}}_{Ak} - \overline{\mathbf{X}}_{Bk}\right)\widehat{\beta}_{\alpha,Bk} + \sum_{k=1}^{K}\left(\widehat{\beta}_{\alpha,Ak} - \widehat{\beta}_{\alpha,Bk}\right)\overline{\mathbf{X}}_{Ak}. \tag{6}$$

[3] The marginal effect of a continuous covariate $\mathbf{x}$ can be evaluated as : $\frac{\partial \mathbf{E}(\text{RIF}_{e_\alpha}|X=\mathbf{x})}{\partial \mathbf{x}}$.

where the first right hand component denoted by $\Delta_{\alpha,\mathbf{X}}$ is attributed to the "covariate effect" while the second right hand term, $\Delta_{\alpha,\beta}$ is associated to the "return effect". Since $\text{RIF}_{e_{0.5}} = \overline{\mathbf{y}}$ the .5th expectile RIF-regression approach corresponds to the conventional mean Oaxaca-Blinder decomposition method.

4 Income Inequality in Egypt (1999–2010)

The 2011 Egyptian revolution is one of the most widely discussed revolutions in recent history. Although several social, political, and religious reasons have emerged as causes of this revolution, factors such as lack of democracy, widespread government corruption, and religious tensions were not the main triggers of the revolution. The Egyptian revolution was a demonstration against the deterioration of living standards and the state of the economy, which resulted in increased poverty, unemployment, and increasingly unequal income distribution and disparity. Our empirical application focuses on analyzing the temporal evolution of income over the 10-year period preceding the 2011 revolution to explain a part of the income disparities between different income groups that contributed to the Egyptian crisis.

4.1 Baseline Model

We consider the typical income equation:

$$\begin{aligned} \text{RIF}(e_\alpha, \ln \mathbf{y}) \quad = \quad & \beta_{0\alpha} + \beta_{1\alpha}\mathbf{Rural} + \beta_{2\alpha}\mathbf{Fem} + \beta_{3\alpha}\mathbf{Married} + \beta_{4\alpha}\mathbf{Hsize} + \\ & \beta_{5\alpha}\mathbf{Age} + \beta_{6\alpha}\left(\mathbf{Age}\right)^2 + \beta_{7\alpha}\mathbf{Leduc} + u, \end{aligned} \tag{7}$$

where $\mathbf{y}$ is the vector of equivalized[4] real disposal income of households, **Rural** is the place of residence (rural/urban), **Fem** is the gender, **Age** is the age and **Leduc** corresponds to the level of education considered as dummies (i.e., primary, secondary, post-secondary, university, postgraduate) and $(\beta_{1\alpha}, \beta_{2\alpha}, \beta_{3\alpha}, \beta_{4\alpha}, \beta_{4\alpha}, \beta_{5\alpha}, \beta_{6\alpha}, \beta_{7\alpha})'$ is a vector of the RIF-regression coefficients. The analysis covers the 10-years period preceding the 2011 revolution (1999–2010). Bootstrap procedure is used to estimate standard errors with 1000 replications.

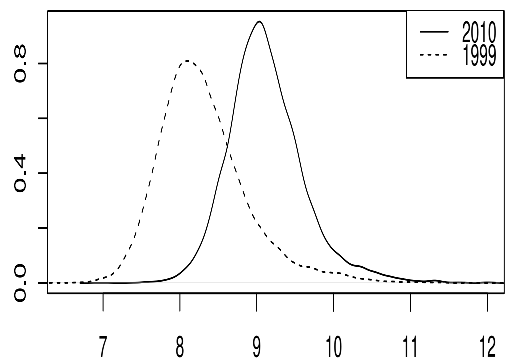
[4]We use the OECD-modified scale that assigns a value of 1 to household head, of 0.5 to each additional adult member and of 0.3 to each child. A correction for inflation was also implemented so as to be able to compare living standards across the years.

4.2 Data and Descriptive Statistics

The empirical analysis made use of data from the Egypt Household Income, Expenditure, and Consumption Survey (HIECS) for the years 1999 to 2010. The HIECS provides information on the living standards of households and individuals as well as the usual socio-demographic characteristics of individuals. These surveys, which covered all governorates throughout Egypt, were conducted every two years to accurately track changes in the living standards of Egyptian households and to define average household and per-capita income from different sources.

The empirical expectiles of equivalized income have continuously increased over the years preceding the 2011 revolution.

Descriptive Statistics	1999	2010
N. of observations	23975	7718
$e_{.10}$	7.87	8.74
$e_{.25}$	8.08	8.92
Mean	8.30	9.12
$e_{.75}$	8.57	9.34
$e_{.90}$	8.87	9.59
Median	8.24	9.07



The shapes of estimated log equivalized income densities remained relatively stable over the years 1999 to 2010.

Table 1 describes the evolution of certain characteristics of households. From 1999 to 2010, household characteristics such as family size, sex and marital status of the head member did not vary greatly. However, the educational level of household heads have slightly evolved over this period along with a general improvement in the primary and secondary education levels. Moreover, the illiteracy rate has fallen while the unemployment rate has risen monotonically.

Table 1: Descriptive of the different explanatory variables

Characteristics of households	Years		Level of education	Years	
	1999	2010		1999	2010
Rural	40.03	53.54	Primary	8.41	12.31
Hsize	4.72	4.41	Secondary	18.41	25.34
Female	14.94	16.70	Psecondary	3.92	4.20
Married	79.81	79.41	University	14.56	11.71
Uempl	21.20	21.83	Pgraduate	0.88	0.80

5 Empirical Results

The decomposition results based on the expectile RIF-regression reported in Table 2 indicate a positive overall income gap from 1999 to 2010 in each part of the distribution. The covariate effects have a weak contribution on the temporal evolution of the income except rural as also shown by the descriptive statistics presented in Table 1. Therefore, the income inequality in Egypt during the ten-year period preceding the 2011 revolution was mainly due to the return effect[5] in each part of the distribution. The estimated regression coefficients presented in Table 3 display the same pattern across expectiles for both years. The household heads incomes were higher in urban areas. However, the positive and monotonously increasing return effects of rural across all expectiles show that there was an improvement in living conditions in rural areas between 1999 and 2010 compared to urban areas. Consequently, the rural population increased sharply from 1999 to 2010 as indicated by the covariate effects.

The monotonous increase in the marginal effect of age between 1999 and 2010 across expectiles explains why young heads of household have more difficulty reaching a certain level of income than older ones. As for the negative differences in the marginal effects of household size, they reflect a decline in the standard of living of large households above the 10th expectile.

The return to education increased on average from 1999 to 2010 for each level. The difference in primary school returns was significant only in the lower part of the distribution, which is in line with our expectations because policies to improve primary education aim to increase the income of the lower part of the distribution.

[5]These return effects correspond to the difference in the estimated regression coefficients given the characteristics in 2010.

Table 2: Decomposition based on the Expectile-RIF regression

Level α	Below the Mean		Mean	Above the Mean	
	0.1	0.25	0.5	0.75	0.90
Return Effects					
Total	0.8648^{***} (0.0020)	0.8534^{***} (0.0032)	0.8361^{***} (0.0055)	0.8173^{***} (0.0114)	0.7970^{***} (0.0247)
Intercept	0.8407^{***} (0.0127)	0.8194^{***} (0.0238)	0.8216^{***} (0.0454)	0.9108^{***} (0.1050)	1.2236^{***} (0.2532)
Rural	0.0133^{***} (0.0022)	0.0258^{***} (0.0036)	0.0397^{***} (0.0062)	0.0411^{***} (0.0116)	0.0043^{ns} (0.0233)
Female	0.0005^{ns} (0.0017)	-0.0004^{ns} (0.0031)	-0.0055^{ns} (0.0063)	-0.0230^{ns} (0.0152)	-0.0665^{*} (0.0368)
Age	0.0250^{***} (0.0074)	0.0474^{***} (0.0126)	0.0800^{***} (0.0221)	0.1120^{**} (0.0445)	0.0894^{ns} (0.1018)
Hsize	-0.0085^{ns} (0.0056)	-0.0212^{**} (0.0085)	-0.0419^{***} (0.0134)	-0.0675^{***} (0.0245)	-0.0797^{*} (0.0469)
Maried	-0.0023^{ns} (0.0068)	-0.0064^{ns} (0.0136)	-0.0322^{ns} (0.0277)	-0.1103^{ns} (0.0682)	-0.3059^{*} (0.1670)
Primary	-0.0016^{**} (0.0008)	-0.0027^{**} (0.0013)	-0.0042^{**} (0.0022)	-0.0064^{ns} (0.0041)	-0.0081^{ns} (0.0074)
Secondary	-0.0014^{ns} (0.0013)	-0.0039^{*} (0.0021)	-0.0071^{*} (0.0036)	-0.0082^{ns} (0.0076)	-0.0036^{ns} (0.0148)
Psecondary	0.0001^{ns} (0.0003)	-0.00002^{ns} (0.0006)	-0.0002^{ns} (0.0012)	0.0005^{ns} (0.0024)	0.0011^{ns} (0.0052)
University	-0.0009^{*} (0.0006)	-0.0042^{***} (0.0011)	-0.0129^{***} (0.0026)	-0.0294^{***} (0.0067)	-0.0551^{***} (0.0163)
Pgraduate	-0.0001^{ns} (0.0001)	-0.0003^{ns} (0.0002)	-0.0010^{ns} (0.0007)	-0.0021^{ns} (0.0020)	-0.0023^{ns} (0.0060)
Covariate Effects					
Total	-0.0014^{ns} (0.0010)	-0.0061^{***} (0.0023)	-0.0180^{***} (0.0046)	-0.0410^{***} (0.0090)	-0.0762^{***} (0.0159)
Rural	-0.0084^{***} (0.0005)	-0.0185^{***} (0.0010)	-0.0356^{***} (0.0019)	-0.0583^{***} (0.0032)	-0.0815^{***} (0.0045)
Female	-0.0006^{***} (0.0003)	-0.0013^{***} (0.0003)	-0.0019^{***} (0.0005)	-0.0031^{***} (0.0011)	-0.0046^{**} (0.0019)
Age	-0.0001^{ns} (0.0001)	-0.0005^{ns} (0.0002)	-0.0010^{ns} (0.0007)	-0.0022^{ns} (0.0014)	-0.0043^{ns} (0.0029)
Hsize	0.0040^{***} (0.0005)	0.0079^{***} (0.0007)	0.0138^{***} (0.0013)	0.0216^{***} (0.0020)	0.0318^{***} (0.0032)
Maried	0.0000^{ns} (0.0000)	0.0001^{ns} (0.0001)	0.0003^{ns} (0.0003)	0.0006^{ns} (0.0008)	0.0011^{ns} (0.0016)
Primary	0.0023^{***} (0.0004)	0.0044^{***} (0.0005)	0.0075^{***} (0.0009)	0.0113^{***} (0.0014)	0.0149^{***} (0.0021)
Secondary	0.0048^{***} (0.0004)	0.0101^{***} (0.0009)	0.0187^{***} (0.0016)	0.0308^{***} (0.0026)	0.0449^{***} (0.0042)
Psecondary	0.0002^{ns} (0.0002)	0.0005^{ns} (0.0004)	0.0009^{ns} (0.0009)	0.0015^{ns} (0.0015)	0.0022^{ns} (0.0021)
University	-0.0036^{***} (0.0005)	-0.0087^{***} (0.0013)	-0.0198^{***} (0.0030)	-0.0408^{***} (0.0063)	-0.0759^{***} (0.0117)
Pgraduate	-0.0001^{ns} (0.0002)	-0.0004^{ns} (0.0006)	-0.0009^{ns} (0.0014)	-0.0023^{ns} (0.0035)	-0.0049^{ns} (0.0074)

Bootstrapped standard errors (1000 replications) are indicated in parentheses.

Beyond the mean, only differences in college performance were significant. The monotonic decline in its return effects shows a decrease in the earnings of higher education in the upper part of the distribution between 1999 and 2010.

6 Conclusion

In this study, we develop a reliable decomposition method based on the expectile RIF-regression, which constitutes a natural extension of the conventional Oaxaca-Blinder decomposition of means to the entire distribution. The proposed expectile RIF-regression is more efficient in close-to-Gaussian situations than the quantile RIF-regression. It is illustrated to document the factors that contributed to the level of income inequality in Egypt during the ten-year period preceding the 2011 revolution (1999–2010). The empirical results demonstrate that the overall gap in income across expectiles was mainly due to the differences in returns to covariates, while the differences in characteristics did not vary much expect rural. The returns to university education decreased in the upper part of the distribution and was accompanied by a positive difference in the marginal effects of age and an increasing return effect in rural areas. These findings demonstrate that young heads of middle-income households in urban areas were more affected by the Egyptian economic crisis, which may explain their leading position in Tahrir Square on January 25, 2011.

References

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4):436–455.

Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika*, 75(4):761–772.

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

Firpo, S., Fortin, N. M., and Lemieux, T. (2011). Decomposition methods in economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, Part A of *Handbook of Labor Economics*, pages 1–102. Elsevier Science Publishers, Amsterdam.

Firpo, S., Fortin, N. M., and Lemieux, T. (2018). Decomposing Wage Distributions Using Recentered Influence Function Regressions. *Econometrics*, 6.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393.

Jones, M. (1994). Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters*, 20(2):149–153.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Machado, J. A. F. and Mata, J. (2005). Counterfactual decompositions of changes in wage distributions using quantile regressions. *Journal of Applied Econometrics*, 20(4):445–465.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12(4):577–590.

Newey, W. and Powell, J. L. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica*, 55(4):819–47.

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

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

Von Mises, R. (1947). On the Asymptotic Distribution of Differentiable Statistical Functions. *The Annals of Mathematical Statistics*, 18:309 – 348.

Appendix

Proof of Theorem 1 and Lemma 1.1

Proof of Theorem 1. Following Newey and Powell (1987), the αth expectile can be determined by:

$$\begin{aligned} e_\alpha &= \frac{1-2\alpha}{\alpha}\int_{\infty}^{e_\alpha}(\mathbf{y}-e_\alpha)\mathrm{d}F_y + \mathbb{E}(y) \\ &= \int_{\mathbb{R}}\left[\frac{1-2\alpha}{\alpha}(\mathbf{y}-e_\alpha)\mathbb{1}(\mathbf{y}<e_\alpha)+\mathbf{y}\right]\mathrm{d}F_y. \end{aligned} \tag{8}$$

If we denote by $\nu(F_y) = e_\alpha$ and $\psi(\mathbf{y}) = \frac{1-2\alpha}{\alpha}(\mathbf{y} - e_\alpha)\mathbb{1}(\mathbf{y} < e_\alpha) + \mathbf{y}$ then the αth expectile in (8) is a linear functional of F expressed as :

$$\nu(F_y) = \int_{\mathbb{R}} \psi(\mathbf{y})\, \mathrm{d}F_y.$$

Hence,

$$\nu(F_{y,\epsilon.G_y}) - \nu(F_y) = \int_{\mathbb{R}} \psi(\mathbf{y})\, \mathrm{d}F_{y,\epsilon.G_y} - \int_{\mathbb{R}} \psi(\mathbf{y})\, \mathrm{d}F_y.$$

The Von Mises (1947) linear expansion of the functional $\nu(F_{y,\epsilon.G_y})$ is given by

$$\nu(F_{y,\epsilon.G_y}) = \nu(F_y) + \int_{\mathbb{R}} \left[\psi(\mathbf{y}) - \int_{\mathbb{R}} \psi(\mathbf{y})\, \mathrm{d}F_y\right] \mathrm{d}\left(F_{y,\epsilon.G_y} - F_y\right) + o\left[\delta(F_{y,\epsilon.G_y}, F_y)\right]$$

where $o\left[\delta(F_{y,\epsilon.G_y}, F_y)\right]$ is the rest that tends to zero and $\delta(F_{y,\epsilon.G_y}, F_y)$ is the distance between $F_{y,\epsilon.G_y}$ and F_y.

$$\begin{aligned}
\nu(F_{y,\epsilon.G_y}) - \nu(F_y) &= \int_{\mathbb{R}} \left[\psi(\mathbf{y}) - \int_{\mathbb{R}} \psi(\mathbf{y})\, \mathrm{d}F_y\right] \mathrm{d}\left(\epsilon G_y - \epsilon F_y\right) + o\left[\delta(F_{y,\epsilon.G_y}, F_y)\right] \\
&= \int_{\mathbb{R}} \left[\psi(\mathbf{y}) - \int_{\mathbb{R}} \psi(\mathbf{y})\, \mathrm{d}F_y\right] \epsilon \mathrm{d}(G_y - F_y) + o\left[\delta(F_{y,\epsilon.G_y}, F_y)\right].
\end{aligned}$$

Therefore,

$$\lim_{\epsilon \to 0} \frac{\nu(F_{y,\epsilon.G_y}) - \nu(F_y)}{\epsilon} = \int_{\mathbb{R}} \left[\underbrace{\psi(\mathbf{y}) - \int_{\mathbb{R}} \psi(\mathbf{y})\, \mathrm{d}F_y}_{IF_\nu}\right] \mathrm{d}(G_y - F_y).$$

Hence,

$$IF_\nu = \psi(\mathbf{y}) - \int_{\mathbb{R}} \psi(\mathbf{y})\, \mathrm{d}F_y.$$

□

Proof of Lemma 1.1.

$$\begin{aligned}
\mathrm{RIF}_\nu &= \nu + IF_\nu \\
&= \int_{\mathbb{R}} \psi(\mathbf{y})\, dF_y + \psi(\mathbf{y}) - \int_{\mathbb{R}} \psi(\mathbf{y})\, dF_y(y) \\
&= \psi(\mathbf{y})
\end{aligned}$$

□

Table 3: Expectile-RIF regression estimates, 1999-2010

Level α	**Below the Mean**		**Mean**	**Above the Mean**	
	0.1	0.25	0.5	0.75	0.90
Expectile-RIF regression estimates, 1999					
Intercept	8.2076^{***} (0.0556)	8.2082^{***} (0.0313)	8.3186^{***} (0.01931)	8.5290^{***} (0.0134)	8.8203^{***} (0.0097)
Rural	-0.5619^{***} (0.0191)	-0.4117^{***} (0.0107)	-0.2634^{***} (0.00664)	-0.1440^{***} (0.0046)	-0.0670^{***} (0.0033)
Female	-0.3085^{***} (0.0439)	-0.1839^{***} (0.0247)	-0.1065^{***} (0.01524)	-0.0590^{***} (0.0106)	-0.0289^{***} (0.0076)
Age	0.0034^{***} (0.0007)	0.0042^{***} (0.0004)	0.0038^{***} (0.00024)	0.0028^{***} (0.0002)	0.0018^{***} (0.0001)
Hsize	-0.1172^{***} (0.0043)	-0.0766^{***} (0.0024)	-0.0444^{***} (0.00149)	-0.0232^{***} (0.0010)	-0.0114^{***} (0.0007)
Maried	-0.0472^{ns} (0.0395)	-0.0644^{***} (0.0222)	-0.0656^{***} (0.01371)	-0.0517^{***} (0.0095)	-0.0303^{***} (0.0069)
Primary	0.5260^{***} (0.0329)	0.3366^{***} (0.0185)	0.1920^{***} (0.01141)	0.0966^{***} (0.0079)	0.0424^{***} (0.0057)
Secondary	0.6293^{***} (0.0252)	0.4349^{***} (0.0142)	0.2695^{***} (0.00875)	0.1480^{***} (0.0060)	0.0720^{***} (0.0044)
Psecondary	0.7419^{***} (0.0465)	0.5250^{***} (0.0261)	0.3324^{***} (0.01615)	0.1839^{***} (0.0112)	0.0902^{***} (0.0081)
University	1.1343^{***} (0.0277)	0.9182^{***} (0.0156)	0.6949^{***} (0.00960)	0.4786^{***} (0.0066)	0.2967^{***} (0.0048)
Pgraduate	1.6558^{***} (0.0940)	1.4631^{***} (0.0529)	1.2432^{***} (0.03264)	0.9879^{***} (0.0226)	0.7084^{***} (0.0164)
Expectile-RIF regression estimates, 2010					
Intercept	8.8665^{***} (0.1053)	8.9717^{***} (0.0569)	9.1402^{***} (0.0334)	9.3501^{***} (0.0221)	9.5969^{***} (0.0158)
Rural	-0.3384^{***} (0.0315)	-0.2670^{***} (0.0170)	-0.1893^{***} (0.0100)	-0.1184^{***} (0.0066)	-0.0661^{***} (0.0047)
Female	-0.2834^{***} (0.0826)	-0.1917^{***} (0.0446)	-0.1392^{***} (0.0262)	-0.1049^{***} (0.0174)	-0.0732^{***} (0.0124)
Age	0.0081^{***} (0.0012)	0.0072^{***} (0.0006)	0.0054^{***} (0.0004)	0.0036^{***} (0.0002)	0.0020^{***} (0.0002)
Hsize	-0.1346^{***} (0.0080)	-0.0911^{***} (0.0043)	-0.0539^{***} (0.0025)	-0.0283^{***} (0.0017)	-0.0134^{***} (0.0012)
Maried	-0.0732^{ns} (0.0777)	-0.0886^{***} (0.0420)	-0.1062^{***} (0.0247)	-0.0980^{***} (0.0163)	-0.0731^{***} (0.0117)
Primary	0.4107^{***} (0.0493)	0.2716^{***} (0.0266)	0.1581^{***} (0.0156)	0.0793^{***} (0.0104)	0.0351^{***} (0.0074)
Secondary	0.5799^{***} (0.0403)	0.3893^{***} (0.0218)	0.2412^{***} (0.0128)	0.1372^{***} (0.0085)	0.0704^{***} (0.0061)
Psecondary	0.7709^{***} (0.0777)	0.5236^{***} (0.0420)	0.3276^{***} (0.0247)	0.1877^{***} (0.0163)	0.0930^{***} (0.0117)
University	1.0676^{***} (0.0515)	0.8102^{***} (0.0278)	0.5848^{***} (0.0164)	0.3947^{***} (0.0108)	0.2444^{***} (0.0078)
Pgraduate	1.5571^{***} (0.1691)	1.3331^{***} (0.0914)	1.1174^{***} (0.0537)	0.8989^{***} (0.0355)	0.6760^{***} (0.0254)