Working Paper no. **80**

Shapley-Value Decompositions of Changes in Wage Distributions: A Note

Francesco Devicienti
University of Torino and LABORatorio Revelli - Collegio Carlo Alberto

Laboratorio R. Revelli, *Collegio Carlo Alberto*
Via Real Collegio, 30 - 10024 Moncalieri (TO)
Tel. +39 011 670.50.60 - Fax +39 011 670.50.61
www.laboratoriorevelli.it - labor@laboratoriorevelli.it

LABOR is an independent research centre of the Collegio Carlo Alberto

Shapley-Value Decompositions of Changes in Wage Distributions: A Note

by

Francesco Devicienti
University of Torino and LABORatorio Revelli - Collegio Carlo Alberto

Dipartimento di Scienze Economiche e Finanziarie "G. Prato",
Università di Torino - Facoltà di Economia,
Corso Unione Sovietica 218bis, 10134 Torino (Italy).
Tel: +39 0116706288. email: devicienti@econ.unito.it

Abstract

This note shows how the Shapley-value can be applied to the regression-based methods that are often used to decompose changes in wage distributions. The method remedies the path-dependency exhibited by existing approaches that compute the contributions due to (i) changes in sample observable characteristics, (ii) changes in the return of characteristics, (iii) changes in the distribution of unobservable characteristics.

JEL code: J31
Keywords: regression-based decomposition methods, wage distribution, Shapley-value.

I would like to thank Jacques Silber for helpful discussion and two anonymous referees for their comments on an earlier version of the paper. The usual disclaimer applies.

1. Introduction

Empirical analyses studying changes in wage distributions often rely upon simple, regression-based decomposition exercises, like the well-known Blinder-Oaxaca decomposition of mean earnings.[1] Juhn, Murphy and Pierce (1993) – JMP henceforth – proposed a more general methodology, aimed at decomposing the observed changes in a wage distribution in changes in three components of a regression function: the observed vector of characteristics in the sample (sample composition), the returns earned by those characteristics (regression coefficients), and the residual distribution (unobservables). More recently, Lemieux (2002) proposed a decomposition method that combines elements of the procedures of JMP and of DiNardo, Fortin and Lemieux (1996).

One problem with the present approaches is that they are sensitive to the order in which factors are analyzed when computing their contribution to the changes in the wage distribution. This paper shows how this "path dependency" can be easily remedied by applying the general Shapley-value approach proposed by Shorrocks (1999). While the Shapley-value has already found application in a number of diverse decomposition exercises[2], is has not yet been applied to regression-based decompositions of changes in wage distributions. This note intends to fill this gap, by showing how the Shapley-value can be used to decompose the observed changes in wage inequality in the contributions due to changes in sample composition, regression coefficients and unobservables. The main advantage of the proposed decomposition is that any (linear and non-linear) distributional statistics can be decomposed, and exact contributions obtained in each case. The resulting procedure encompasses existing methods as special cases. An application studying the changes in the Italian wage distribution, 1985-1999, illustrates the proposed Shapley-value decomposition and shows that the ordering of factor elimination matters for the results.

[1] The regression-based decomposition methods are reviewed in depth in a number of papers, among which Lemieux (2002) and Wan (2004).

[2] A non-exhaustive list of recent applications include Israeli (2007), Mussard and Terraza (2006), Sastre and Trannoy (2002).

2. Methodology

Most regression-based decomposition exercises start by writing (log) wages, Y_{it}, as:

$$Y_{it} = X_{it}\beta_t + u_{it} \qquad (1)$$

where X_{it} is a vector of characteristics, and β_t a coefficient vector representing the prices/returns of those characteristics. Residuals u_{it} capture unobserved factors and have cumulative distribution function $F_t(u_{it} \mid X_{it})$, conditional on observed characteristics. If θ_{it} denotes the rank of individual i in the cumulative residual distribution F_t, then we can write:

$$u_{it} = F_t^{-1}(\theta_{it} \mid X_{it}). \qquad (2)$$

With r representing a reference year (e.g., initial year), (1) can be re-written as:

$$Y_{it} = [X_{it}\beta_r + F_r^{-1}(\theta_{it} \mid X_{it})] + [X_{it}(\beta_t - \beta_r)] + [F_t^{-1}(\theta_{it} \mid X_{it}) - F_r^{-1}(\theta_{it} \mid X_{it})] \qquad (3)$$

showing that the wage distribution varies between t and r because of (*i*) changes in the covariate distribution, at fixed prices β_r (first term in the RHS), (*ii*) changes in prices at given quantities (second term), and (*iii*) changes in the residual distribution. Simple OLS estimates of (1) can be used to produce β_s and the empirical distribution $F_s(\cdot \mid X_{is})$, for $s=r,t$. JMP then propose a "residual imputation procedure" to find the counterfactual residual $F_r^{-1}(\theta_{it} \mid X_{it})$, i.e. the residual that a worker would get in year r if s/he were at the same rank θ_{it} as in the residual distribution of year t.

To obtain the factor contributions to the distributional changes between year r and year t, the idea behind (3) is to start by evaluating the distribution when the three factors have all been eliminated, sequentially re-introducing them one at a time and computing the distributional changes obtained at each stage. The process starts at the bottom of Figure 1, with a starting level of a chosen distributional indicator, $I(\cdot)$, at year t given by: $I(X_r\beta_r + F_r^{-1}(\theta_{it} \mid X_{it}))$. The underscore to X, β and F indicate that the covariate distribution, the price vector and the residual distribution are kept at their value in the reference year. Note that this starting (counterfactual) distribution is indistinguishable from the distribution in the reference year:

$$I(X_r\beta_r + F_r^{-1}(\theta_{it} \mid X_{it})) = I(X_r\beta_r + F_r^{-1}(\theta_{ir} \mid X_{ir})) \qquad (4)$$

as long as the distribution F is conditional on the same set of characteristics X in both t and r and the distributional statistics respect a minimal "anonymity property"[3].

Starting from the counterfactual distribution at the bottom of figure 1, factors contributing to the level of inequality in a given year are subsequently re-introduced. At each stage, a counterfactual distribution is first obtained by switching on/off the relevant factors; $I(\cdot)$ is then computed for each counterfactual distribution. For example, along sequence 3, moving from the bottom up, the counterfactual distribution $X_r\beta_t + F_r^{-1}(\theta_{it}| X_{it})$ is first computed where the price of observable characteristics is replaced by its level in t. Moving up along the same sequence, the covariate distribution is next set at the level in t, obtaining the counterfactual distribution $X_t\beta_t + F_r^{-1}(\theta_{it}| X_{it})$. Finally, the residual distribution of the reference year is replaced with the corresponding distribution of year t, obtaining the distribution $X_t\beta_t + F_t^{-1}(\theta_{it} | X_{it})$, which is nothing else than the observed distribution in year t.

However, the ordering in which factors are eliminated clearly matters for the calculation of each factor contribution, because of interactions between the factors. If $I(\cdot)$ depends on K factors, $I=I(z_1, z_2, \ldots, z_K)$, the marginal contribution of factor k may be obtained through a comparison of the values assumed by I when k is, respectively, operating (denoted z_k=1) and removed (denoted z_k=0). The resulting change in I will depend on the values the other factors have been kept fixed at. With K factors, there exist $K!$ sequences where the factors are eliminated in a different order. The value of ΔI corresponding to the turning on/off of factor k will in general depend on the particular path of eliminations followed up to that point.

How does one replace the distribution of characteristics of the reference year (X_r) with the distribution of characteristics in year t (X_t)? DiNardo et al. (1996) and Lemieux (2002) propose a re-weighting procedure that transforms the distribution of covariates of reference year r into the distribution of covariates of year t. Assume that the sample in each year consists of observations

[3] "Anonymity" (permutations of the observations do not alter the value of I) holds for standard distributional statistics.

(Y_{is}, X_{is}, w_{is}), where w_{is} are sample weights. To obtain the re-weighting factor the idea is to pool period r and period t samples and run a standard logit or probit model for the probability of being in year t, conditional on X:

$$P_{it} = \text{Prob}(\text{period} = t \mid X_{it}). \quad (5)$$

The re-weighting factor is:

$$\psi_i = [(1-P_{it})/P_{it}] \text{x} [P_t/(1-P_t)], \quad (6)$$

where P_t is the unconditional probability of being in year t. Finally, the counterfactual weight, $w_{ir}^{*} = w_{ir}/\psi_i$, is used to compute a distributional statistics that would have prevailed if the distribution of covariates X had been as it was in period t. Using this re-weighting procedure the distribution of covariates can be changed from its shape in r to its shape in t along any point of the sequences depicted in figure 1.

How does one compute the contribution of a given factor to the changes in wage distribution between year r and year t? The (first-round) marginal contribution of, say, factor β in year t, can be obtained as:

$$C_{\beta} = I(X_{ir}\beta_t + F_r^{-1}(\theta_{it} \mid X_{it})) - I(X_{ir}\beta_r + F_r^{-1}(\theta_{it} \mid X_{it}))$$

However, first-round contributions do not necessarily add-up to the total change to be decomposed. Moreover, there are other sequences where the factor β is eliminated at a different stage, creating a "path-dependency problem". As shown by Shorrochs (1999), the Shapley-value contribution of factor β is instead computed by averaging all possible changes in $I()$ obtained when substituting β_r with β_t at any stage over any of the $K!$ elimination sequences, where in our case K=3 ($K!$=6):

$$C^{shapley}{}_{\beta} = \sum_{j,h=t,r} \left[I(X_{ij}\beta_t + F_h^{-1}) - I(X_{ij}\beta_r + F_h^{-1}) \right] \cdot \pi_{jh} \quad (7)$$

Note that there are elimination sequences that deliver the same value of the marginal contribution of a given factor[4]; therefore marginal contributions are weighted by the probability π_{jh} that a given sequence is followed, with π_{jh}=1/6 if $h \neq j$ and π_{jh}=2/6 if $h=j$. As pointed out by Shorrocks (1999), the decomposition rule in (7) corresponds to the Shapley value for the cooperative games in which "output" I is shared amongst the set of "agents" K (Shapley, 1953).

It is easy to verify that Shapley-value decompositions satisfy the adding-up property. They also lend themselves to a natural interpretation: Shapley-value contributions are the expected marginal impact of each factor when the expectation is taken over all the possible elimination paths.

Figure 1 also clarifies the relationship between the Shapley-value decompositions and the decompositions of JMP and of Lemieux. The procedure of JMP is exactly the same as that obtained following sequence 1. In fact, JMP define the following counterfactual distributions, for $s=t,r$:

$$Y_{is}^{1} = X_{is}\beta_r + F_r^{-1}(\theta_{is} \mid X_{is}),$$

$$Y_{is}^{2} = X_{is}\beta_s + F_r^{-1}(\theta_{is} \mid X_{is})$$

$$Y_{is}^{3} = X_{is}\beta_s + F_s^{-1}(\theta_{is} \mid X_{is}) = Y_{is},$$

The change in the distributional statistics is then decomposed as:

$$\Delta I = I(Y_t)\text{-}I(Y_r) = [I(Y_t^1)\text{-}I(Y_r^1)] + [I(Y_t^2)\text{-}I(Y_r^2)\text{-}(I(Y_t^1)\text{-}I(Y_r^1))] + [I(Y_t^3)\text{-}I(Y_r^3)\text{-}((I(Y_t^2)\text{-}I(Y_r^2))]$$

$$= \Delta I^X + \Delta I^\beta + \Delta I^F$$

where ΔI^X is the covariate effect, ΔI^β the coefficient effect and the ΔI^F is the residual effect. After simplification, one gets:

$$\Delta I^X = I(X_{it}\beta_r + F_r^{-1}(\theta_{it}| X_{it}) \text{ - } I(X_{ir}\beta_r + F_r^{-1}(\theta_{is}| X_{is})),$$

$$\Delta I^\beta = I(X_{it}\beta_t + F_r^{-1}(\theta_{it}| X_{it}) \text{ - } I(X_{it}\beta_r + F_r^{-1}(\theta_{it}| X_{it})),$$

$$\Delta I^F = I(X_{it}\beta_t + F_t^{-1}(\theta_{it}| X_{it}) \text{ - } I(X_{it}\beta_t + F_r^{-1}(\theta_{it}| X_{it})),$$

which is precisely the decomposition followed in sequence 1 of Figure 1.

[4] For example, in figure 1 the marginal effect of β along sequences 3 and 4 and 2 and 5 is the same. Similarly, the marginal effect of X is the same along sequences 4 and 6.

Similarly, one can show that Lemieux's decomposition (what he calls JMP2, p. 681[5]) is identical to the one obtained following sequence 3. In practice, JMP first remove the effect of covariates, followed by coefficients and residuals; Lemieux, instead proceeds by first eliminating coefficients, next covariates and then residuals. When there are no compelling reasons to follow any particular sequences, the attractiveness of a Shapley-value approach naturally emerges.[6]

3. Discussion

Before turning to the empirical application of section 4, in this section I briefly discuss how the decomposition method proposed in the note relates to existing approaches and highlight directions for future research.

As is well known, the pioneering work of Oaxaca (1973) and Blinder (1973) assume a linear regression model for income, as in equation (1), and focus on decomposing the difference in mean income between two groups. The previous section has shown how JMP extend this approach, with a decomposition that depends on the difference in the entire earnings distribution rather than in the difference in the mean income only. In their approach any distributional statistics can be decomposed in a "price effect" (related to variation in β), a "quantity effect" (reflecting changes in X) and a "residual effect" (related to changes in the distribution of unobservables).

The decompositions of authors such as Fields (2002, 2004), Morduch and Sicular (2002), Cowell and Fiorio (2006) are also based on linear regression models, but they are not generally concerned with separating the "price effect" from the "quantity effect". In fact, following the inequality decomposition by income sources approach of Shorrocks (1982, 1983), they decompose total inequality in the contribution due to each source z_k, where $z_k = \beta_k X_k$.

[5] JMP2 is formally equivalent to the re-weighting procedure proposed by Lemieux in his section 3.
[6] Note that both JMP and Lemieux account for residuals (unobservables) at the end of the decomposition. However, it is hard to see why this component should always be eliminated lastly, rather then be treated symmetrically with respect to the other two components. In fact, in many cross-section Mincerian wage regressions – and certainly in the Italian case shown above – the unexplained component is higher than the explained component. Moreover, much of the empirical literature on wage inequality has been concerned with the rise in within-inequality (inequality in residuals) as much as it has been with the rise in between-inequality.

Other authors have relied on nonlinear regression models for their inequality decomposition. For example, the method of Wan (2004) allows for the decomposition of any inequality measure in the contribution due to each source z_k (but not of β_k and X_k separately) using both linear and non-linear regression models. Machado-Mata (2005) have based their decomposition on quantile-regressions (see Lemiuex, 2002, for a critical review). Bourguignon *et al.* (2001) use a framework similar to that of JMP, but relax the requirement that the income-generating process be linear.[7] Finally, a different strand of the literature uses semi-parametric and non- parametric methods as the basis for their decomposition exercises. A notable example is the kernel density estimation approach of DiNardo *et al.* (1996).[8]

What is important to our aims is to note that in virtually all the decomposition methods reviewed above there is a path-dependency problem, in that one can often think of a different order of factor elimination than the one followed by the authors in their decomposition.[9] This note has shown that that one cannot in general expect that the order of factor elimination is immaterial for the results of the decomposition. Within the special case of linear regression models, this note has illustrated how this path-dependency can be remedied using a Shapley-value approach. Future research may draw on the general ideas illustrated here, and in the other references provided, to fruitfully apply the Shapley-value approach to more general income-generating models (non-linear regression models as well as non-parametric models) and decompositions.

4. An application: Decomposing Changes in the Italian Wage Distribution, 1985-1999

[7] They postulate that the income of a household in a given year depend on four sets of arguments: its observable socio-demographic characteristics or those of its members (X), the set of prices and labor remuneration rates it faces (β), and a set of parameters describing the labor force participation and occupational choice behavior of its members (which they summarize in the vector λ). The model is clearly more complicated than the one originally used by JMP, in part reflecting their different focus: while JMP are concerned with the determinants of the observed changes in *wage* inequality, Bourguignon *et al.* are interested in the causes of changes in the distribution of *household income*. Therefore the labor participation decision of all household members is added to the analysis as an additional contributing factor to the observed distributional changes.

[8] Morduch and Sicular (2004, p. 93) note that, while these methods impose as little structure as possible, researchers often find it necessary to impose more structure in order to draw sharp conclusions, which may explain the co-existence of both strands of the literature.

[9] This problem is explicitly recognized by DiNardo et al., p. 1021, and by Bourguigon et al., p. 145.

The methodology of section 2 is now applied to decompose the change in wage inequality experienced by Italy between 1985 and 1999. I use the *Worker History Italian Panel*, an administrative dataset with information on the weekly wages of a large sample of Italian employees (about 100,000 observations yearly) in private firms, aged 15-64. The data are available from the LABORatorio R. Revelli, where detailed online documentation can be found (http://www.laboratoriorevelli.it). Reflecting the administrative nature of the data, earnings are likely to be accurately recorded from firms' records (Contini, 2002). Another advantage of the data is its large sample size and the availability of a number of workers' and firms' characteristics to be used as controls in Mincerian wage regressions. However, education is not observed, and has to be proxied by other worker's characteristics (e.g. a workers occupation). Borgarello and Devicienti (2006) further discuss data and sample details, institutional background and results.

While the procedure can be applied to decompose any distributional statistics, for simplicity I focus on subset of measures that have most commonly used in the literature on wage inequality. Table 1 shows that inequality increased between 1985 and 1999 according to each inequality measures, and that the raise is concentrated in the upper part of the distribution. The 90-50 gap increases by 0.134 log points and the 90-10 gap by 0.132, implying almost no change in the 50-10 differential. The variance of logarithms increased by 0.1 while the Gini coefficient raised by 0.009. A flexible specification is employeed for estimating (1) and (6), with X including a quartic in age and dummy indicators for a worker's occupation, gender, sector of activity, firm size and regional area, with all dummies fully interacted with the age polynomial. To compute the counterfactual residuals, the empirical F distribution is approximated by a step function (see Lemieux, 2002, footnote 21).

Table 1 considers the three procedures that have been discussed in section 2, namely the JMP, the JMP2 and the Shapley-value. The decomposition of each of the inequality measures shows that all three factors are responsible for the increase in wage inequality. The proportion explained by a given factor is however dependent on the decomposition procedure.[10] For the 90-50 gap, changes in the distribution of covariates explain 48% of the total increase, changes in the coefficients explain another 42% and changes in the residual explain the remaining 9% according to the JMP decomposition. However, along a different elimination sequence, the results obtained may be significantly different: compare for example the results obtained along sequence 3 of Figure 1 (namely the decomposition called JMP2) with the JMP results obtained along sequence 1. Not surprisingly, the Shapley-value decomposition – the average over the 6 sequences – is also significantly different: only 24% is now attributed to the effect of covariates, 55% is the coefficient effect and 21% the effect of residual. In this case, while the largest effect is played by the covariates in the JMP approach, the dominant factor becomes the coefficient vector in the Shapley-value decomposition.

The differences between the Shapley and the JMP method appear relevant also with respect to the other inequality measures displayed in the table. To formalize this statement, Table 1 also reports bootstrap standard errors (1000 replications) for each inequality decomposition, as well as a generalized Wald test statistics for the null hypothesis that the three elements of the Shapley-value decompositions are (jointly) indistinguishable for the corresponding elements in the JMP decomposition.[11] The test strongly rejects the null at standard levels of confidence for the Gini coefficient, the variance of logarithms and the p90-p50 gap; the null is rejected at the 90% level for the 75-25 gap, although it is not rejected for the 90-10 gap. Overall, we conclude that, while the results depends on the specific application at hand, one should in general not expect to find

[10] Note that the contribution of the residuals is the same in the and JMP2 by construction (see Figure 1).

[11] The Wald statistics for the hypothesis H_0: $(C^{shapley}_X, C^{shapley}_\beta, C^{shapley}_F) = (C^{JMP}_X, C^{JMP}_\beta, C^{JMP}_F)$, against the alternative H_A that the two decompositions are statistically different, is computed as follows (see Cameron and Trivedi (2005), p. 378): $W = (\theta^{shapley} - \theta^{JMP})' V_{boot}^{-1}\left[\theta^{shapley} - \theta^{JMP}\right](\theta^{shapley} - \theta^{JMP})$, where $\theta^{shapley} = (C_X^{shapley}, C_\beta^{shapley}, C_F^{shapley})'$ is a column vector of the estimated Shapley contributions and $\theta^{JMP} = (C_X^{JMP}, C_\beta^{JMP}, C_F^{JMP})'$ is the corresponding vector of JMP contributions. $V_{boot}\left[\theta^{shapley} - \theta^{JMP}\right]$ is a bootstrapped variance matrix, defined as:

$$V_{boot}\left[\theta^{shapley} - \theta^{JMP}\right] = \frac{1}{B-1}\sum_{b=1}^{B}\left[\left(\theta_b^{shapley} - \theta_b^{JMP}\right) - \left(\bar{\theta}^{shapley} - \bar{\theta}^{JMP}\right)\right]\left[\left(\theta_b^{shapley} - \theta_b^{JMP}\right) - \left(\bar{\theta}^{shapley} - \bar{\theta}^{JMP}\right)\right]'$$

where $\theta_b^{shapley}$ and θ_b^{JMP} are the estimated vectors in the b-th bootstrapped sample (b=1,.. B). The Wald statistics is then compared to the $\chi^2(J)$ critical values, where J is equal to the number of restrictions being tested.

immaterial differences between the Shapley-value and either the JMP or JMP2 decompositions. Therefore, the ordering in which factors are analyzed in empirical applications matters for the final decomposition and the Shapley-value approach offers a simple and sensible criterion to minimize the arbitrariness of the choices made.

Before concluding, I briefly discuss how the Shapley-value approach proposed in the note may be extended if one is interested in disentangling the contribution of the different characteristics included in X. For example, suppose that X is actually partitioned in $X=(X^e, X^o)$, where say X^e includes education-related variables and X^o contains all remaining variables. Correspondingly, the return-to-characteristics vector β is partitioned as $\beta=(\beta^e, \beta^o)$. At the stage in which the contribution of X is computed along any sequence in Figure 1, one may compute the contribution of, say, X^e by averaging the change in I obtained when (i) changing $X^e{}_r$ to $X^e{}_t$ while keeping X^o fixed at $X^o{}_{r,}$, and (ii) changing $X^e{}_r$ to $X^e{}_t$ after X^o has already been changes at $X^o{}_{t,}$. A similar nested procedure may be followed along the other sequences of Figure 1 at the stage where the contribution of X is computed. The corresponding contribution of the return β^e can be computed with an analogous procedure.[12] While possible in principle, one immediate problem with this extension is that if one attempts to recover the separate contribution of many of the regressors included in X, the number of factors (and sequences) to consider in Figure 1 grows rapidly and may entail a high computational burden. This burden is further increased if bootstrapped standard errors are required. Clearly devising efficient methods for handling such computations would be an interesting area of future research.

5. Conclusions

Existing approaches to decompose changes in the wage distribution into the effect of regression coefficients, covariate distribution and residuals, exhibit path-dependency. The proposed Shapley-value decomposition offers a natural solution to the problem and encompasses existing

[12] It should be noted, however, that an important shortcoming of the Shapley decomposition rule in this case is that it does not satisfy the principle of independence of the aggregation level (see Shorrocks, 1999). I thank an anonymous referee for pointing this out.

approaches as special cases. An application to the changes in the Italian wage inequality shows that the ordering of factor elimination matters for the results. Finally, it should be noted that, while the paper has focused on decompositions of wage distributions, the methodology is readily applicable to many other contexts in which regression methods are used to study the distribution of any variable of interest.

References

Blinder, A. S. (1973), "Wage discrimination: reduced form and structural estimates", *Journal of Human Resources*, 8(4):436-455.

Borgarello, A. and F. Devicienti (2006), "L'aumento della disuguaglianza dei salari in Italia: premi salariali per le "nuove" skill?", *Politica Economica*, n. 2. (English working paper version: http://www.labor-torino.it/workingpapers/wp18.htm).

Bourguignon, F., Fournier, M., and Gurgand, M. (2001), "Fast development with a stable income distribution: Taiwan, 1979-94", *Review of Income and Wealth*, 47(2):139-163.

Cameron A.C. and Trivedi P.K. (2005), Microeconometrics, Cambridge University Press.

Contini (2002), *Labor mobility and wage dynamics in Italy*, (eds.), Rosenberg & Sellier, Torino.

Di Nardo J., Fortin N., and T. Lemieux (1996), "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach", *Econometrica*, vol. 64, no, 5, 1001-1044.

Fields, G. S. (2003). "Accounting for income inequality and its change: a new method with application to distribution of earnings in the United States". In Polachek, S. W., editor, *Research in Labour Economics*, volume 22. Elsevier.

Israeli O. (2007) "A Shapley-based decomposition of the *R*-squared of a linear regression", *Journal of Economic Inequality*, 5, no. 2, 199-212.

Morduch, J. and Sicular, T. (2002), "Rethinking inequality decomposition, with evidence from rural china", *The Economic Journal*, 112(476):93.106(14).

Mussard S. and Terraza V. (2006), "The Shapley decomposition for portfolio risk", GREDI Working paper, no. 06-09.

Juhn, K., Murphy M. and B. Pierce (1993), "Wage inequality and the rise in returns to skill", *Journal of Political Economy*, vol.101, No. 3, 410-442.

Lemieux T. (2002), "Decomposing changes in wage distributions: a unified approach", *Canadian Journal of Economics*, 35, n. 2, 646-688.

Oaxaca, R. R. A. (1973), "Male-Female wage differentials in urban labor markets", *International Economic Review*, 14(3):693.709.

Sastre M. and A. Trannoy (2002), "Shapley inequality decomposition by factor components: Some methodological issues", in P. Moyes, C. Seidl and A.F. Shorrocks (eds.), Inequalities: Theory, Experiments and Applications, *Journal of Economics* 9, 51-90.

Shapley, L. (1953), "A value for n-person games," in Kuhn A.H. and Tucker A.W., eds., Contributions to the Theory of Games, Vol. 2 (Princeton University Press).

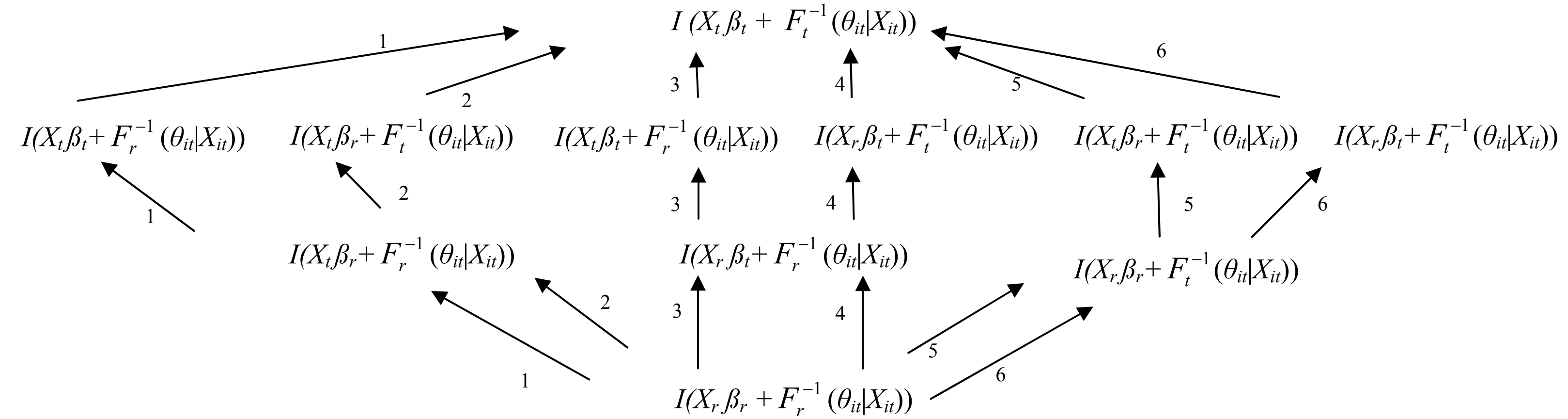
Shorrocks, A. F. (1982). "Inequality decomposition by factor components", *Econometrica*, 50:193-211.

Shorrocks, A. F. (1983). "The impact of income components on the distribution of family incomes", *The Quarterly Journal of Economics*, 98(2):311-326.

Shorrocks A. (1999), "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value", Essex University, mimeo.

Wan G. (2004), "Accounting for income inequality in rural China: a regression-based approach", *Journal of Comparative Economics*, 32, 348-363

Figure 1: Elimination sequences



.

Table 1: Wage inequality decompositions in Italy, 1999-1985

		ΔI	Factor contributions (% of total change)		
		Δ*I*	***X***	**Beta**	**Residual**
P90-P10		0.132			
	JMP		0.279	0.538	0.181
			(0.041)	(0.042)	(0.018)
	JMP2		0.325	0.493	0.181
			(0.027)	(0.032)	(0.018)
	Shapley		0.307	0.524	0.169
			(0.023)	(0.030)	(0.021)
	Wald tests	H_0: Shapley=JMP	H_0: Shapley= JMP2		
	(p-value)	1.77 (0.610)	1.76 (0.622)		
P90-p50		0.134			
	JMP		0.482	0.424	0.093
			(0.025)	(0.024)	(0.011)
	JMP2		0.124	0.782	0.093
			(0.016)	(0.020)	(0.012)
	Shapley		0.244	0.547	0.209
			(0.014)	(0.018)	(0.011)
	Wald tests	H_0: Shapley=JMP	H_0: Shapley= JMP2		
	(p-value)	190 (0.000)	181 (0.000)		
P75-25		0.057			
	JMP		0.182	0.598	0.219
			(0.053)	(0.052)	(0.025)
	JMP2		0.329	0.450	0.219
			(0.032)	(0.043)	(0.025)
	Shapley		0.276	0.547	0.176
			(0.023)	(0.035)	(0.026)
	Wald tests	H_0: Shapley=JMP	H_0: Shapley= JMP2		
	(p-value)	6.62 (0.085)	6.81 (0.078)		
Var-logs		0.098			
	JMP		0.076	0.322	0.601
			(0.016)	(0.022)	(0.027)
	JMP2		0.135	0.263	0.601
			(0.032)	(0.036)	(0.027)
	Shapley		0.131	0.302	0.566
			(0.030)	(0.036)	(0.027)
	Wald tests	H_0: Shapley=JMP	H_0: Shapley= JMP2		
	(p-value)	12.1 (0.007)	13.7 (0.003)		
Gini		0.009			
	JMP		0.143	0.537	0.318
			(0.032)	(0.037)	(0.029)
	JMP2		0.340	0.341	0.318
			(0.040)	(0.043)	(0.029)
	Shapley		0.279	0.472	0.248
			(0.035)	(0.034)	(0.031)
	Wald tests	Shapley=JMP	Shapley= JMP2		
	(p-value)	36.6 (0.000)	34.7 (0.000)		
	No. observations:	60548 in t=1985; 68728 in t=1999.			

Notes: bootstrap standard errors in brackets (1000 replications).