école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

Bayesian Inference for Distributional Changes: The Effect of Western TV on Wage Inequality and Female Participation in Former East Germany

Edwin Fourrier-Nicolai
Michel Lubrano

WP 2020 - Nr 27

Aix*Marseille université Initiative d'excellence
A*Midex Initiative d'excellence Aix-Marseille
Investir l'avenir
Faculté d'Économie et de Gestion Aix*Marseille Université
cnrs
L'ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
CENTRALE MARSEILLE

Bayesian Inference for Distributional Changes: The Effect of Western TV on Wage Inequality and Female Participation in Former East Germany

Edwin Fourrier-Nicolai* Michel Lubrano†

July 2020

Abstract

This paper investigates the evolution of wage formation in a Mincer model with sample selection for which we develop Bayesian inference and growth incidence and poverty growth curves. We estimate the effect of an exogenous exposure to Western TV broadcasts on labour market participation and wage inequality in East Germany after the German reunification. Using the GSOEP, we find evidences that Western television had significantly increased wage inequality among males while it has significantly affected female labour participation and led the less productive females to drop out from the market, hiding thus a large increase in wage inequality among females.

Keywords: Bayesian inference, Labour market, Distributional changes, Sample selection, Wage inequality
JEL codes: C11, C24, D31, D91, J21

*Aix Marseille Univ, CNRS, AMSE, Marseille, France; Email: edwin.fourrier@univ-amu.fr; Address: 5-9 Bd Maurice Bourdet, 13001 Marseille, France.

†School of Economics, Jiangxi University of Finance and Economics and Aix Marseille Univ, CNRS, AMSE, Marseille, France. Email: michel.lubrano@univ-amu.fr. Amse-Greqam, 5-9 Bd Maurice Bourdet, 13001 Marseille, France.

[0]**Acknowledgments:** During the writing of this paper, we benefited from useful conversations and remarks by Stephen Bazen, Frank Cowell, Habiba Djebbari, Emmanuel Flachaire, Stephen Jenkins and Christian Robert, as well as participants of the 2019 ESEM held in Manchester and internal AMSE seminars. Of course, the usual disclaimers apply. This work was carried out thanks to the support of the A*MIDEX project (No ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR). This work was also supported by French National Research Agency Grants ANR-17-EURE-0020. Complementary financing was provided by the "National Natural Science Foundation of China" (No. 71764008).

[0]**Corresponding author:** Michel Lubrano, 5 Bd Maurice Bourdet, 13001 Marseille, France. Tel: 0033 413 55 25 57. Email:michel.lubrano@univ-amu.fr

1 Introduction

Income distributional changes are usually summarized by the variation of a descriptive statistics (e.g. the mean, the median income, the proportion of poor, or of some specific quantiles). The choice of these statistics is fundamentally arbitrary, leading to focus on some aspects of the distribution while ignoring others. The long-standing debate about the choice of the mean against the median to describe the centrality of an income distribution is particularly illuminating in that respect. What could be seen as a statistical debate has actually concrete policy implications. Defining the poverty threshold as a fraction of the median income instead of the mean can affect both the estimated level of poverty and indirectly the targeting of anti-poverty policies. Instead of focusing on one particular aspect of the distribution, Ravallion and Chen (2003) have proposed to consider the Growth Incidence Curve (GIC) for describing changes occurring over a complete income distribution. More precisely, the GIC measures the quantile-specific rate of income growth while guaranteeing that inequality and poverty comparisons can be done in an unambiguous way. In the same vein, Son (2004) has proposed the Poverty Growth Curve (PGC) for describing whether distributional changes have been favourable to the poor whatever the level of the poverty line. The advantage of this approach is that both curves are closely related to stochastic dominance as shown in Duclos (2009) and Araar et al. (2009). Consequently, distributional changes measured using these curves are consistent with a broad class of welfare functions as well as a large set of poverty and inequality indices. Hence, a growing literature makes use of these curves to describe distributional changes, see Lakner and Milanovic (2016) using the world income distribution, Chancel and Piketty (2017) for the Indian income distribution or Novokmet et al. (2017) for Russia among many possible examples. In addition, these curves have also featured in a long-standing literature that uses counterfactual income distributions to decompose and attribute differences in inequality and poverty to different factors, for example, changes in worker characteristics or in the returns to those characteristics. Early contributions to this literature include Juhn et al. (1993), Dinardo et al. (1996), and Donald et al. (2000). Each of these papers essentially search to account for difference across entire wage or income distributions - which could be formally expressed as GICs - using counterfactual distributions.

Despite their conceptual importance and widespread practical use, formal statistical inference for growth incidence curves and poverty growth curves have not yet been established firmly. Only very recently, Ferreira et al. (2019) have developed estimation and inference procedures in a parameter-free framework for the growth incidence curve. In this paper, we develop estimation and inference for both growth incidence curves and poverty growth curves in a Bayesian framework, and we propose a series of tests for dominance criteria in the context of wage distributions with sample selection corresponding to the decision of participating or not to the labour market. As a matter of fact, a Mincer equation explains formation of observed wages while when it is corrected for sample selection its explanation power is extended also to potential wages, which means also the wages of those who decided not to participate would have obtained if they had participated. In a distribution free framework, using quantile regressions, Arellano and Bonhomme (2017) found that correcting for selection strongly affect UK male wage distribution in the lower quantiles corresponding to the exclusion of low-skilled males out of the labour market. As the effect was much weaker for females the potential wage gap between males and females was much reduced over the period 1978-2000. In the present paper, we shall investigate how sample selection affect distributional changes by means of GIC and PGC in East Germany after reunification.

Although a parameter-free approach has the advantage of requiring no assumption on the structural relationship between income and its covariates, the advantages of a Bayesian approach are manifold. First, statistical tests can be performed directly once inference is obtained. Second, in a distribution-free approach like in Ferreira et al. (2019), estimated quantile functions - on which the GIC are based - experience a large local variability with usual sample sizes leading to an important uncertainty in both ends. This is a serious drawback when studying poverty and inequality. Parametric Bayesian inference overcomes these difficulties by imposing a structure on the data; parametric models compensate for the scarcity of observations and leads naturally to smooth GIC and PGC. While a distribution-free approach does not make any assumption about the shape of the income distribution, a parametric Bayesian approach relies on a particular representation of the income distribution. Therefore, by considering that the income distribution can be represented by a parametric distribution, we can take advantage of the parametric structure to increase accuracy and precision for inference and tests. Finally, because we have an underlying regression model, we can decompose changes in inequality and attribute these changes to specific factors. In this attempt, we augment the standard log-linear model of wage formation introduced by Mincer (1974) to allow for sample selection and decision to participate to the labour market. In doing so, we assume that the wage distribution follows a conditional log-normal process with conditional heteroskedasticity. This approach has the advantage, that using a well recognized model of wage formation, we obtain realistic parametric analytical expressions for the GIC and PGC. Then, we can test numerous hypotheses regarding distributional changes. For instance, we can test if distributional changes have been welfare improving in terms of stochastic dominance, if distributional changes have been relatively pro-poor, if every quantile has benefited equally from growth and, if distributional changes occurring in a specific group are preferred to those occurring in another group.

We illustrate our method to estimate the effect of an exogenous exposure to Western TV broadcasts in former German Democratic Republic (GDR) on wage inequality after the reunification. Indeed, by affecting social aspirations, Western TV might have affected behaviours on the labour market and, in fine, the evolution of wage inequality. Using the GSOEP, we find evidences that television had significantly increased wage inequality among males while it has significantly affected female labour participation and led the less productive females to drop out from the market, compensating thus a large increase in wage inequality among females.

The paper is organized as follows. In the next section, we present the growth incidence and poverty growth curves and their relationship with the quantile function, the Lorenz curve and stochastic dominance which allows to characterize distributional changes with respect to a broad class of ethical judgements. In the third section, we derive a parametric model for wage formation with sample selection. We derive the corresponding GIC and PGC and detail the impact of sample selection on distributional changes. In a fourth section, we provide Bayesian inference for the GIC and PGC using a MCMC algorithm for the selection equation and for the conditional wage equation with heterogeneity. We propose a set of tests for characterizing distributional changes. The empirical application is presented in the fifth section. The last section concludes. Some empirical details are regrouped in an Appendix.

2 Distributional changes and stochastic dominance

The analysis of distributional changes is closely related to the quantile function (Ravallion and Chen 2003) and to the Lorenz curve (Son 2004). In this section, we provide a simple derivation of the GIC and PGC. We show how these functions are related to stochastic dominance and we characterize distributional changes with respect to a broad class of ethical judgements.

2.1 Quantile functions and Lorenz curves

Let Y be a continuous random variable (e.g. income, consumption, wage) with cumulative distribution function (cdf) $F(y)$, probability density function (pdf) $f(y)$, with support contained on the non-negative real line. The quantile function is defined as the inverse of the cumulative distribution function $Q(p) = F^{-1}(y)$, a first step to introduce the Lorenz curve following Gastwirth (1971):

$$L(p) = \frac{1}{\bar{y}} \int_0^p Q(t)\, dt, \tag{1}$$

from which we have another relation between the quantile function and the Lorenz curve with:

$$Q(p) = \bar{y} L'(p), \tag{2}$$

In those expressions, $\bar{y}$ is the mean income expressed as:

$$\bar{y} = \int_0^\infty y f(y)\, dy = \int_0^1 Q(p)\, dp. \tag{3}$$

Because welfare depends both on the level of income and on its distribution, Shorrocks (1983) introduced the generalized Lorenz curve, which is simply the Lorenz curve multiplied by the mean income $\bar{y}$:

$$GL(p) = \int_0^p Q(t)\, dt = \bar{y} L(p).$$

Numerous inequality and poverty measures (among others, the Gini coefficient, the Foster et al. (1984)'s indices and the cumulative poverty gap curve) rely on the quantile function, and thus, can be derived from the Lorenz curve and the generalized Lorenz curve (see e.g. Foster and Shorrocks 1988 or Davidson and Duclos 2000).

2.2 Distributional changes

Let us now consider two dates $t-1$ and t and the corresponding income cumulative distributions $F_{t-1}(y)$ and $F_t(y)$. The growth incidence curve (GIC) of Ravallion and Chen (2003) measures the growth rate of the p-quantile for every p. It can be derived by taking logs and first differences in (2):

$$g_t(p) = \gamma_t + \Delta \log L_t'(p) = \Delta \log GL_t'(p), \tag{4}$$

where $\gamma_t = \Delta \log(\bar{y}_t)$ is the average growth rate. Thus the growth incidence curve corresponds either to the variation of the first derivative of the generalized Lorenz curve or to the average growth rate plus the variation of the first order derivative of the Lorenz curve. Graphically, the GIC associates the growth rate of income to a proportion p of individuals ordered by increasing base-year income. Thus, if inequality does not change then $g_t(p) = \gamma_t$ for all p, and we can say that the p-quantile increases (decreases) if $g_t(p) > 0$ ($g_t(p) < 0$). Observe that computing the GIC requires only cross-sectional data at two different dates instead of longitudinal data. By doing so, we consider only the shapes of the distribution and not individuals destinies per se. This is why the GIC curve is sometimes referred as being anonymous.

An alternative approach for assessing distributional changes has been proposed by Son (2004) who introduces the Poverty Growth Curve (PGC). The initial question of Son (2004) was to determine whether the mean income of the lower quantiles (corresponding to the poor) is growing faster than the mean income of the other quantiles. The poverty growth curve is thus defined as the variation in percentage of the average income of the bottom p% of the population and corresponds to $\Delta \log(\bar{y}_p)$, where $\bar{y}_p$ is the average income of the bottom p%. Using (1) and (3), the Lorenz curve can be written as:

$$L(p) = \frac{\int_0^p Q(t)\, dt}{\int_0^1 Q(t)\, dt} = \frac{p\bar{y}_p}{\bar{y}}.$$

Taking logs, first differences and rearranging the terms, the poverty growth curve of Son (2004) corresponds to:

$$G_t(p) = \Delta \log \bar{y}_p = \gamma_t + \Delta \log L(p) = \Delta \log GL(p). \tag{5}$$

Equation (5) shows that $G_t(p)$ results from the sum of the variation in overall income growth and the variation of inequality.

2.3 Relation to stochastic dominance

Because GIC and PGC are directly related to the quantile function and the generalized Lorenz curve, there is a direct link between GIC, PGC and stochastic dominance. Let us consider an income distribution observed between two periods y_{t-1}, y_t with a growth rate γ_t and a common poverty line z. First-order stochastic dominance of y_t over y_{t-1} up to poverty line z implies that $F(y_t) \leq F(y_{t-1})$ for all $y \leq z$. Since first-order stochastic dominance is essentially a comparison of cumulative distributions, quantile functions are related to stochastic dominance as well, this is the p-approach to dominance developed in Davidson and Duclos (2000). Thus, first-order stochastic dominance of y_t over y_{t-1} implies that $F^{-1}(y_t) \geq F^{-1}(y_{t-1})$ for all $y \leq z$, or in other terms that $Q_t(p) \geq Q_{t-1}(p)$ for all $p < q$ where $q = \max_p(p|z - Q_t(p) \geq 0)$. It follows directly from equation (4) that the condition of first-order stochastic dominance of y_t over y_{t-1} is equivalent to:

$$g_t(p) > 0, \quad \forall p \in [0, q].$$

Therefore, first-order stochastic dominance of the second period over the first period is verified if and only if the growth incidence curve is positive for every quantile up to q. In terms of poverty, this means that the proportion of individuals below the poverty line (the headcount ratio) is always greater in $F(y_{t-1})$ than in $F(y_t)$, for any poverty line lower than z. More generally, this can be regarded as an ordering for all indices from any non-decreasing social welfare function for which the headcount ratio is a special case (Davidson and Duclos 2000).

If the growth incidence curve is negative for some values of p, then we cannot conclude unambiguously about whether distributional changes have been welfare-improving or not. In such a case, one has to impose more normative conditions by considering stochastic dominance at the second order to obtain a possibly complete ordering. Because generalized Lorenz dominance is strictly equivalent to second-order stochastic dominance (Atkinson 1987, Foster and Shorrocks 1988), it follows directly from equation (5) that second-order stochastic dominance of y_t over y_{t-1} is equivalent to:

$$G_t(p) > 0, \quad \forall p \in [0, q[.$$

Therefore, second-order stochastic dominance of the second period over the first period is verified if and only if the poverty growth curve is positive for every quantile up to q. This can be regarded as an ordering for all indices from any non-decreasing concave social welfare function. In terms of inequality, this means that the mean income of the $p\%$ of the population is no smaller in y_t than in y_{t-1}, and for some p this mean income is greater. In terms of poverty, this is equivalent to say that all indices built on relative poverty gaps in y_{t-1} are always greater than in y_t, for any poverty line lower than z (Davidson and Duclos 2000).

Duclos (2009) and Araar et al. (2009) go a step further on and state that growth is relatively pro-poor if:

$$g_t(p) > \gamma_t, \quad \forall p \in [0, q].$$

This condition is verified if the quantiles of the poor increase at a pace greater than the average growth. Taking into account inequality among the poor, we can test whether:

$$G_t(p) > \gamma_t, \quad \forall p \in [0, q[.$$

As $G_t(p)$ can be decomposed between growth and variation in inequality, we can further qualify a distributional change. If $\Delta \log L(p) < 0$, which means an overall reduction of inequality, then $G_t(p) > \gamma_t$ for all p. In this case, growth is said to be *pro-poor* wherein both poverty and inequality are reduced. If $0 < G_t(p) < \gamma_t$ for all p, then growth reduces poverty but is accompanied by an increase in inequality. This is sometimes called a *trickle-down* growth wherein growth reduces poverty but the poor receive proportionally less benefits than the non-poor. Last, if $G_t(p) < 0$ for all p and $\gamma_t > 0$, then we have an *immiserizing* growth where growth increases poverty.

3 A parametric model for the labour market

Observed wages has been traditionally explained by a simple Mincer (1974) equation. However, the decision to participate to the labour market is generally not random leading to potential large differences between observed and potential wages, potential wages corresponding to the wages that all individuals (participants and non-participants) would have obtained if they had participated. The model we consider is essentially a Type 2 Tobit model of Amemiya (1985, p. 385) for which we derive the corresponding GIC and PGC. Equipped with this model, we have the necessary tool to examine the impact of Western television on the evolution of wages in the empirical application.

3.1 Reservation wage and sample selection

Economists usually view non-participation as a rational choice arising from a maximizing-utility framework. In this sense, individuals decide not to participate to the labour market when the utility provided by their reservation wage is higher than the utility gained by accepting their potential market wage. In this case, we observe a zero wage. We thus assume that the decision to participate follows the random utility model and that the underlying difference in utility between the offered wage and the reservation wage is determined by:

$$P_i^* = z_i'\zeta + v_i \qquad v_i \sim N(0,1). \tag{6}$$

Depending if the utility of participating P_i^* is positive, or negative, the observed wage is equal to:

$$y_i = \begin{cases} y_i^* & \text{if } P_i^* > 0, \\ 0 & \text{if } P_i^* \leq 0. \end{cases} \tag{7}$$

where the potential log wage is given by:

$$\log(y_i^*) = x_i'\beta + u_i. \tag{8}$$

We thus observe $P = 1$ when $y_i > 0$ and $P = 0$ when $y_i = 0$. We have two Gaussian random terms, u and v which are expected to be correlated. Hence, we can express this problem as a bivariate normal distribution with correlated error terms:

$$\begin{pmatrix} v_i \\ u_i \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}\right). \tag{9}$$

Therefore, for the subsample with a positive wage, this introduces a sample selection bias wherein the sample is not representative of the population, the conditional expectation of $\log(y^*)$ being given by:

$$\mathrm{E}(\log(y_i^*)|x_i, P_{it}^* > 0) = x_i'\beta + \mathrm{E}(u_i|v_i > -z_i'\zeta). \tag{10}$$

Knowing that v and u have a bivariate normal distribution, the conditional expectation of the error term is:

$$\mathrm{E}(u_i|v_i > -z_i'\zeta) = \rho\sigma \, \frac{\phi(-z_i'\zeta)}{1 - \Phi(-z_i'\zeta)}, \tag{11}$$

where $\phi(.)$ and $\Phi(.)$ denote, respectively, the pdf and cdf of the standard normal. By symmetry of the normal distribution, we recover the well-known inverse Mills ratio $\lambda(.)$:

$$\lambda_i = \frac{\phi(-z_i'\zeta)}{1 - \Phi(-z_i'\zeta)} = \frac{\phi(z_i'\zeta)}{\Phi(z_i'\zeta)}. \tag{12}$$

Then, the conditional expectation of $\log(y^*)$ can be written as:

$$\mathrm{E}(\log(y_i^*)|x_i, P_{it}^* > 0) = x_i'\beta + \rho\sigma \, \lambda_i, \tag{13}$$

leading to the regression equation:

$$\log(y_i) = x_i'\beta + \lambda_i\beta_\lambda + \epsilon_i. \tag{14}$$

In this regression the error term ϵ_i has a non-constant variance with:

$$\mathrm{Var}(\epsilon_i) = \sigma^2 - [z_i'\zeta \, \lambda_i + \lambda_i^2]\beta_\lambda^2. \tag{15}$$

As $\mathrm{Var}(\epsilon_i)$ is a function of z_i, we have a case of functional heteroskedasticity arising from sample selection which is fairly constrained. However, heteroskedasticity could occur even in the absence of sample selection (i.e. when $\rho = 0$) just because $\mathrm{Var}(\epsilon_i)$ could, for various practical reasons, depend on z_i. Thus, we consider a general form of functional heteroskedasticty:

$$\sigma_i^2 = \sigma^2 h(z_i'\delta + \lambda_i\tau + \lambda_i^2\psi). \tag{16}$$

where z_i represents a set of exogenous variables including x_i. So that in matrix notation including all the observations $Z = (X, B)$ and B are the excluded covariates. B may measure the opportunity cost of participating to the labour market. In the empirical application we shall use the household income and other replacement incomes. These variables are required for identification. Remark that we can easily test for selectivity bias by testing if β_λ is significantly different from zero in the regression equation (14).

3.2 Corresponding GIC and PGC

Because we are in a conditional log-normal framework, we can derive easily analytical expressions for the GIC and PGC. In the usual log-normal process, the two parameters μ and σ^2 are fixed, leading to $\exp(\mu + \sigma^2/2)$ for the mean, $\exp(\mu + \sigma\Phi^{-1})$ for the quantile function and $\Phi(\Phi^{-1}(p) - \sigma)$ for the Lorenz curve. In our conditional model, both μ and σ are function of covariates. Thus, the conditional mean of $y|y > 0$ becomes:

$$\mathrm{E}(y|z_i) = \exp(x_i'\beta + \lambda_i\beta_\lambda + \sigma^2 h(z_i'\delta + \lambda_i\tau + \lambda_i^2\psi)/2). \tag{17}$$

The conditional quantile function is found to be:

$$Q(p|z_i) = \exp(x_i'\beta + \lambda_i\beta_\lambda + \sigma\sqrt{h(z_i'\delta + \lambda_i\tau + \lambda_i^2\psi)}\Phi^{-1}(p)), \tag{18}$$

while the conditional Lorenz curve is:

$$L(p|z_i) = \Phi(\Phi^{-1}(p) - \sigma\sqrt{h(z_i'\delta + \lambda_i\tau + \lambda_i^2\psi)}).$$

In all these expressions, we have conditioned on z only because x is included in z. If we now consider two periods, $t-1$ and t, using equations (4) and (5), we can get the GIC and the PGC for two periods of time:

$$\begin{aligned} g_t(p|z_i) &= x_i'\beta_t + \lambda_i\beta_{\lambda t} + \sigma_t\sqrt{h(z_i'\delta_t + \lambda_i\tau_t + \lambda_i^2\psi_t)}\Phi^{-1}(p) \\ &\quad - \left(x_i'\beta_{t-1} + \lambda_i\beta_{\lambda t-1} + \sigma_{t-1}\sqrt{h(z_i'\delta_{t-1}) + \lambda_i\tau_{t-1} + \lambda_i^2\psi_{t-1}}\Phi^{-1}(p)\right), \end{aligned} \tag{19}$$

$$\begin{aligned} G_t(p|z_i) &= x_i'\beta_t + \lambda_i\beta_{\lambda t} + \sigma_t^2 h(z_i'\delta_t + \lambda_i\tau_t + \lambda_i^2\psi_t)/2 \\ &\quad -(x_i'\beta_{t-1} + \lambda_i\beta_{\lambda t-1} + \sigma_{t-1}^2 h(z_i'\delta_{t-1} + \lambda_i\tau_{t-1} + \lambda_i^2\psi_{t-1})/2) \\ &\quad + \log\left(\frac{\Phi(\Phi^{-1}(p) - \sigma_t\sqrt{h(z_i'\delta_t + \lambda_i\tau_t + \lambda_i^2\psi_t)})}{\Phi(\Phi^{-1}(p) - \sigma_{t-1}\sqrt{h(z_i'\delta_{t-1} + \lambda_i\tau_{t-1} + \lambda_i^2\psi_{t-1})})}\right). \end{aligned} \tag{20}$$

The height of these curves is determined by the average wage growth rate while their slope is essentially a function of the difference between the two skedastic functions. Particularly, the slope is positive if inequality has increased and negative if it has decreased.

3.3 The impact of sample selection on distributional changes

Because both curves given in (19) and (20) are conditional on the distributions of x and z, estimation of these curves requires to select particular values for x_i and z_i. In order to get posterior draws for a single curve, we have to integrate out z (knowing that x is included in z) with respect to its empirical distribution. Let us take the example of $g_t(p)$ (the approach is similar for $G_t(p)$). Then, we have:

$$\begin{aligned} g_t(p|\theta) &= \int \log Q_t(p|z_i, \theta) f_t(z) dz \\ &\quad - \int \log Q_{t-1}(p|z_i, \theta) f_{t-1}(z) dz. \end{aligned} \tag{21}$$

From (18), the log quantile is partly a linear function of z, so integration means that $g_t(p|\theta)$ can be evaluated at the sample mean of z. This is wrong for the non-linearities arising from heteroskedasticity and sample selection. However, we can think that the approximation is not severe. Then, we can estimate distributional changes among participants to the labour market by evaluating the GIC and PGC at the average value of z for the sample of participants for who $y_i > 0$ or alternatively for who $\Phi(-z_i'\bar{\zeta}) > 0.50$. In contrast, we can estimate distributional changes for potential wages of the whole population (including participants and non-participants) by evaluating the GIC and PGC at the average value of z for the whole sample. The difference between both estimates provides the impact of sample selection on distributional changes. In the same vein, we can easily estimate counterfactual distributional changes by predicting GIC and PGC when varying z, other things being equal.

4 Bayesian inference for GIC and PGC

Inference is conducted in two separate steps. We first conduct inference for the selection equation which relies essentially on a Gibbs sampler following Koop (2003, pages 214-216), but a Metropolis-Hastings algorithm would also be possible, see e.g. Marin and Robert (2007). Then, knowing the posterior draws of ζ, we compute the inverse Mills ratio $\lambda(z_i'\zeta)$. Conditionally on this value, we can conduct inference on the heteroskedastic model in a second step, which requires a Metropolis algorithm for the parameters of the skedastic function δ. Having obtained m posterior draws of the model parameters, we can easily transform these draws into draws of both GIC and PGC using (19) and (20) and then test several dominance hypotheses.

4.1 The marginal selection model

Bayesian inference for the probit model was first proposed by Albert and Chib (1993) using a Gibbs sampler. Following equation (6), if the latent variable P_i^* were known, this model would be a linear regression model with unit variance. Under a non-informative prior, the posterior density of ζ conditionally on the latent P^* and the observed P is a simple Gaussian density with:

$$\zeta|P^*, P \sim \mathrm{N}(\hat{\zeta}, (Z'Z)^{-1}),$$

with $\hat{\zeta} = (Z'Z)^{-1})Z'P^*$. The posterior distribution of P^*, conditionally on ζ and P is a truncated normal with:

$$P_i^* \sim \mathrm{TN}(z_i'\zeta, 1),$$

which is truncated at zero by the left if $P_i = 1$ or by the right if $P_i = 0$. A Gibbs sampler is devised by simulating alternatively P^* and ζ as detailed in Algorithm 1.

Algorithm 1 Gibbs sampler for the selection model

Initialize $\zeta^{(0)}$
for $j = 1$ to m **do**
  Generate $P^{(j)*} \sim TN(Z\zeta^{(j-1)}, 1)$
  Generate $\zeta^{(j)} \sim N((Z'Z)^{-1})Z'P^{(j)*}, (Z'Z)^{-1})$
  Compute $\lambda^{(j)} = \phi(Z\zeta^{(j)})/\Phi(Z\zeta^{(j)})$
  Store $\zeta^{(j)}$ and $\lambda^{(j)}$
end for

A Maximum likelihood estimator serves at initializing the chain with $\zeta^{(0)}$. Inference for the decision to participate is done on all the observations (both positive and zero wages). We then compute the inverse of the Mills ratio (IMR) for positive wages. Therefore, we obtain m draws of the IMR, $\lambda(\zeta^{(j)})$, that we note $\lambda^{(j)}$ for clarity. Inference on the wage equation is then conducted conditionally on $\lambda^{(j)}$, that will be integrated out within a MCMC loop.

4.2 The conditional heteroskedastic wage model

After sample selection, we have n positive wages stacked in the vector y, a matrix of k covariates including a constant term and m draws of the IMR, $\lambda^{(j)}$. A second matrix of covariates for modelling heteroskedasticity includes the covariates of the wage equation, the covariates of the selection equation plus $\lambda^{(j)}$ and $\lambda^{2(j)}$. We shall note these matrices of covariates respectively $X(\lambda^{(j)})$ and $Z(\lambda^{(j)})$. Integration of the IMR will be done in the MCMC loop. For the simplicity of notation, we note the matrices of covariates as $X(\lambda)$ and $Z(\lambda)$ and work conditionally on λ to detail the main steps of the procedure.

The wage equation is noted in a matrix form:

$$\log(y) = X(\lambda)'\beta + \epsilon, \qquad \epsilon \sim N(0, \sigma^2 H(\delta|\lambda)).$$

The $n \times n$ diagonal matrix H represents the variance-covariance of the error term ϵ:

$$H(\delta|\lambda) = \mathrm{diag}(h(z_1(\lambda)'\delta), \ldots, h(z_n(\lambda)'\delta)).$$

Because we explicitly model heteroskedasticity, the posterior standard error of β will be correctly evaluated, contrary to the usual two-step regression of Heckman which requires a specific correction (Heckman 1979). The likelihood function of y given $X(\lambda)$ and $Z(\lambda)$ is:

$$\begin{aligned} L(y; \beta, \sigma^2, \delta) &= \left(\prod_{i=1}^{n} (y_i)^{-1/2} \right) (2\pi)^{-n/2} \sigma^{-n} |H(\delta|\lambda)|^{-1/2} \\ &\quad \times \exp - \frac{1}{2\sigma^2} (\log(y) - X(\lambda)\beta)' H(\delta|\lambda)^{-1} (\log(y) - X(\lambda)\beta). \end{aligned}$$

Conditionally on λ, this likelihood function is identical to the one considered in Bauwens et al. (1999, Chap. 7), Griffiths (2001), and Koop (2003, Chap. 6), except for the Jacobian of the transform of y into $\log y$. We select a non-informative prior on all the parameters as done in Griffiths (2001):

$$\pi(\beta, \sigma^2, \delta) \propto 1/\sigma^2,$$

so that the posterior density of the parameters is proportional to:

$$\begin{aligned}\pi(\beta, \sigma^2, \delta|y, \lambda) \;\; &\propto \;\; \sigma^{-(n+1)}|H(\delta|\lambda)|^{-1/2} \\ &\quad\;\; \exp -\frac{1}{2\sigma^2}\left[s_*(\delta|\lambda) + (\beta - \beta_*(\delta|\lambda))' M_*(\delta|\lambda)(\beta - \beta_*(\delta|\lambda))\right], \quad (22)\end{aligned}$$

with:

$$\begin{aligned}M_*(\delta|\lambda) &= X'(\lambda)H^{-1}(\delta|\lambda)X(\lambda), & (23)\\ \beta_*(\delta|\lambda) &= M_*^{-1}(\delta|\lambda)X'(\lambda)H^{-1}(\delta|\lambda)\log(y), & (24)\\ s_*(\delta|\lambda) &= (\log(y) - X(\lambda)\beta_*(\delta))'H^{-1}(\delta|\lambda)(\log(y) - X(\lambda)\beta_*(\delta)). & (25)\end{aligned}$$

There are several ways of treating this posterior density in β, σ, δ (conditionally on λ). Koop (2003, Chap. 6) derives each conditional density and proposes a Gibbs sampler with a Metropolis step for the conditional distribution of δ. Bauwens et al. (1999, Chap. 7) and Griffiths (2001) prefer to note that conditionally on δ, we recover the conditional posterior distributions of β and σ^2 which are an inverted gamma 2 and a Student noted respectively:[1]

$$\begin{aligned}\pi(\sigma^2|\delta, \lambda, y) &= f_{i\gamma}(\sigma^2|n, s_*^2(\delta|\lambda)) & (26)\\ \pi(\beta|\delta, \lambda, y) &= f_t(\beta|\beta_*(\delta|\lambda), M_*(\delta|\lambda), s_*^2(\delta|\lambda), n). & (27)\end{aligned}$$

We can derive the marginal posterior density of δ by an analytical integration of (22) in β and σ^2. The result is immediately deduced as the constant of integration of the above Student times a term which comes from the likelihood function, so that:

$$\pi(\delta|y, \lambda) \propto |H(\delta|\lambda^{(j)})|^{-1/2} s_*(\delta|\lambda^{(j)})^{-(n-k)/2}|M_*(\delta|\lambda^{(j)})|^{-1/2}. \quad (28)$$

Because we should not forget that this density is conditional on λ and that we have already obtained m draws $\lambda^{(j)}$, we have conditioned this density on $\lambda^{(j)}$, which will be integrated out in the loop of Algorithm 2.

The conditional posterior density of δ does not belong to a known family. We can draw random numbers from $\pi(\delta|y, \lambda^{(j)})$ using a Metropolis algorithm and a Gaussian proposal density. The mean and the variance of the proposal density can be approximated by the posterior mode of δ and minus the inverse of the second order derivative of the log posterior density at this point, respectively, everything computed at the mean value $\bar{\lambda}$ to build the proposal. So we have a unique proposal.

Once we have drawn random numbers from $\pi(\delta|y, \lambda)$, it is easy to generate random numbers for β and σ. We have simply to replace $M_*(\delta|\lambda)$, $\beta_*(\delta|\lambda)$ and $s_*(\delta|\lambda)$ by $M_*(\delta^{(j)}|\lambda^{(j)})$,

[1] Notations for these two densities are provided in the appendix of Bauwens et al. (1999), together with procedures to draw random numbers from them.

$\beta_*(\delta^{(j)}|\lambda^{(j)})$ and $s_*(\delta^{(j)}|\lambda^{(j)})$ and then use the following densities for drawing values for σ^2 and β:

$$\pi(\sigma^2|\delta^{(j)}, \lambda, y) = f_{i\gamma}(\sigma^2|n, s_*^2(\delta^{(j)}|\lambda^{(j)})), \tag{29}$$

$$\pi(\beta|\sigma^2, \delta^{(j)}, \lambda, y) = f_N(\beta|\beta_*(\delta^{(j)}|\lambda^{(j)}), \sigma^2 M_*^{-1}(\delta^{(j)}|\lambda^{(j)})). \tag{30}$$

This approach by direct sampling reduces substantially the dependence in the draws which is inherent to any MCMC method.

We favour the independent Metropolis algorithm to the random walk Metropolis used in Griffiths (2001) because it implies a much lower rejection rate as underlined in Bauwens et al. (1999, Chap. 3). The usual argument in favour of the random walk Metropolis is that its proposal density is simpler to establish (only a variance-covariance matrix is needed). In our case, the proposal density is easy to calibrate as detailed above. Let us denote $\ell(\delta)$ the proposal density and $\pi(\delta|y, \lambda)$ the posterior density of δ given in (28), the independent Metropolis algorithm is implemented in Algorithm 2.

Algorithm 2 Independent Metropolis algorithm integrating out λ

Obtain m draws $\lambda^{(j)}$ from the selection model in a first step
Compute the posterior expectation of the IMR, $\bar{\lambda}$
Calibrate the proposal density $\ell(\delta|\bar{\lambda})$
for $j = 1$ to m **do**
  Build $X(\lambda^{(j)})$ and $Z(\lambda^{(j)})$
  Generate a proposal $\delta^p \sim \ell(\delta|\bar{\lambda})$
  Compute the probability of acceptance as

$$p_a = \min\left(\frac{\pi(\delta^p|y, X(\lambda^{(j)}), Z(\lambda^{(j)}))}{\pi(\delta^{(j-1)}|y, X(\lambda^{(j)}), Z(\lambda^{(j)}))}\frac{\ell(\delta^{(j-1)}|\bar{\lambda})}{\ell(\delta^p|\bar{\lambda})}, 1\right)$$

  Generate a uniform random number u.
  If
  if $u \leq p_a$ **then**
    accept $\delta^{(j)} = \delta^p$
  else
    keep $\delta^{(j)} = \delta^{(j-1)}$
  end if
  Generate $\sigma^{(j)2} \sim \pi(\sigma^2|\delta^{(j)}, y, X(\lambda^{(j)}), Z(\lambda^{(j)}))$
  Generate $\beta^{(j)} \sim \pi(\beta|\sigma^2, \delta^{(j)}, y, X(\lambda^{(j)}), Z(\lambda^{(j)}))$
  Store $\delta^{(j)}$, ${\sigma^{(j)}}^2$, ${\sigma^{(j)}}^2$
end for

This algorithm takes full account of the uncertainty contained in both equations (participation and wage equation) because λ is integrated out in a common loop with the other parameters, using the draws of λ obtained during the first step.

4.3 Bayesian testing

Once we have obtained m posterior draws of β, σ, δ and λ, we can easily transform them into draws of $g_t(p)$ and $G_t(p)$ on a predefined grid of np values for p, using formulae (19) and (20). These draws are stored in two matrices with m rows and np columns. The posterior mean curve is obtained by taking the mean of those np columns while taking the $\alpha/2$ and $1-\alpha/2$ quantiles over each column provides $(100 \times \alpha)\%$ confidence bounds. This method is straightforward in a Bayesian context contrarily to the classical approach as detailed in Araar et al. (2009) or more recently in Ferreira et al. (2019). Furthermore, we can use these m posterior draws of the growth incidence and of the poverty growth curves for testing several crucial hypotheses about the nature of growth and distributional changes, for instance:

- H_1: $g_t(p) > 0$ ($G_t(p) > 0$), $\forall p \in [0,1]$, meaning that distributional change has been welfare improving in terms of first-order (second-order) stochastic dominance,

- H_2: $g_t(p) > \gamma$ ($G_t(p) > \gamma$), $\forall p \in [0,q]$, meaning that distributional change has been relatively pro-poor (q being the head count poverty rate),

can be tested in a simple way. We only have to compute for each value p of the grid, the empirical probability that these conditions are satisfied. For instance, consider the probability that the growth incidence curve is (strictly) positive. A j^{th} Monte Carlo draw for this condition is noted $\mathbb{1}(g_t^{(j)}(p) > 0)$, where $\mathbb{1}(.)$ is the indicator function. We have simply to compute the following empirical mean:

$$\Pr(g_t(p) > 0) \simeq \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}(g_t^{(j)}(p) > 0).$$

5 Western television and wage inequality in East Germany

By affecting the aspirations of East Germans, exposure to Western TV broadcasts before reunification should have affected behaviours on the labour market once the two parts of Germany were re-unified. In order to estimate this causal effect, we use a natural experiment where, for exogenous reasons, only a part of former East Germany was exposed to Western television broadcasts. Then, we investigate how this exogenous exposure has affected the evolution of the wage distribution for males and females once the reunification has operated.

5.1 Aspirations in a natural experiment

Since the seminal paper of Veblen (1899) on conspicuous consumption, a growing literature has been investigating the formation of aspirations and how they can affect socio-economic outcomes (Stutzer 2004, Knight and Gunatilaka 2012). More recently, an empirical literature has been investigating how television can shape aspirations. In the context of the German reunification, Hyll and Schneider (2013) find consistent evidences that watching Western television in the former GDR increased material aspirations substantially while Bursztyn and Cantoni (2016) find that Western television exposure affects the structure of consumption. In doing so, television affects the perceived socio-economic context or, in other terms, the reference group of individuals; that is, income and consumption comparisons take place not only within individual's actual reference group (e.g. relatives, friends, neighbours, and colleagues as in Luttmer 2005 or Clark and Senik 2010), but also within a virtual reference group consisting of television characters. In this attempt, advertising and entertainment programs play a non-negligible role in shaping mental representations (e.g. Chong and Ferrara 2009 or Bursztyn and Cantoni 2016). More recently, Genicot and Ray (2017) demonstrate in a theoretical model that socially determined aspirations are sufficient to generate persistent income inequality by affecting a broad variety of investment decisions.

In this perspective, long-term exposure to Western television could have shifted aspirations of exposed East German households toward Western standards affecting their behaviours on the labour market once reunification occurred. More precisely, Western television could have affected their preferences and social aspirations which, in turn, could have affected incentives on the labour market and wage inequality through migration, education, fertility and marriage decisions. Therefore, we investigate the effect of aspirations on inequality with the understanding that it operates through these different channels. Finally, and perhaps most importantly, television could have affected female participation to the labour market, once the reunification operated. As noted in Rosenfeld et al. (2004), before reunification women's participation in East Germany was extremely high, roughly equal to that of men. In 1989, women participation was 89% compared to 92% for men. At the same date, women participation was only 56% in West Germany, well below that of men (83%). We can suppose that the women exposed to Western TV changed their mind concerning the utility they derived from working. Particularly in West Germany, there is an non-negligible stigma associated with working mothers, usually referred as the *Rabenmutter* stigma. This leads us to study separately distributional changes of males and females.

In order to identify the effect of Western TV, we exploit a natural experiment. Owing to exogenous variations in the signal strength, Bursztyn and Cantoni (2016) have identified two comparable groups with different access to television broadcasting. While one group (the control group) could receive only East German television channels, the other group (the treatment group) was also exposed to West-German television channels. The East German regions without access were located either in the North-East and in the South-East of the country, and were either too far away from the transmitters or were located on the other side of mountains that blocked the signals. Their Figures 1 and 3 display a map of the treatment and control areas. About 85% of the GDR population was exposed to Western television broadcasts, while the remaining 15% had only access to East German television broadcasts.

Previous studies using this natural experiment argue that both groups were initially homogeneous, so that it is sufficient to compare distributional changes of the treatment group to those of the control group for measuring the treatment effect.[2] Homogeneity is in fact a multidimensional concept and homogeneity is better controlled for when conditioning on covariates. Indeed, it is not guaranteed that between-group differences in distributional changes arise because of the treatment or because of differences in the composition of the groups. Thus, we rely on an explanatory model of wage formation in order to distinguish the effect of the treatment from that of potential differences in the composition of the groups.

Despite the fact that Bursztyn and Cantoni (2016) suggest that everybody was watching Western TV in the exposed regions, we cannot formally gauge whether each individual in the treatment group has actually been treated.[3] As a consequence, we are actually estimating an intention-to-treat (ITT) effect; that is, the effect of being offered the treatment. Since the process governing television exposure resembles random assignment, the ITT effect has a causal interpretation: it tells us the causal effect of offering the treatment. For this reason, the estimated ITT effect is generally smaller than the treatment effect on those who were treated i.e. those who actually watched Western television broadcasts daily.

Once we have estimated the causal effect of the treatment on individual wages, the second step consists in obtaining the associated conditional wage distribution for each group, and then computing distributional changes occurring in each group between the two periods using the growth incidence and the poverty growth curves. Particularly, the trajectory of both treated and controlled households is measured between 1992 (the first data for which we have income data) and 1995 (three years later). We do not consider a later period, because as time elapsed, the control group experiences the effect of Western television and many other effects could have inferred into the changes in the income distribution of the two groups.

[2]For instance, Bursztyn and Cantoni (2016) compare both regions in 1955 and 1990 and they do not observe any differential trend for the two groups between 1955 (before television became popular in East Germany) and 1990 (during the reunification). In the same direction, Hyll and Schneider (2013) demonstrate that economic, political, and social conditions in the control region did not systematically differ from conditions in the other regions of East Germany.

[3]Bursztyn and Cantoni (2016) show that there were no difference in TV set equipments between the treatment and control regions. Television was a basic good and the great majority of households had a television set in East Germany (98% of households in 1989). Moreover, they suggest that the great majority of those who were exposed to Western television actually consumed it even if it was officially forbidden. Figure 2 of Bursztyn and Cantoni (2016) indicates that 66% of respondents in districts with access to Western television declared they watched Western TV stations daily. In contrast, only 5% of the respondents in the district of Dresden declared so. Finally, there is no evidence of selective spatial sorting across groups before the reunification. The centralization of the economy and the chronic housing shortage inhibited selective spatial sorting across groups before the reunification.

5.2 Data

We use the German Socio-Economic Panel (GSOEP), a longitudinal representative survey of German households provided by the Institute for Economic Research (DIW). The survey started in 1984 in West-Germany and the first wave surveying East Germans started in 1990 for a limited set of variables including the regional policy regions (*ROR, Raumordnungsregionen*) wherein a household lived at a given year. We use the regional policy regions to determine whether a household belongs to the treatment or to the control group. Specifically, the control group was formed by the households located in one of the five following ROR in 1990: Stralsund-Greifswald, Rostock, Neubrandenburg, Oberlausitz and Dresden. We assign households living in the rest of East Germany to the treatment group. The final sample is constituted of the working-age individuals (between 16 and 65) living in East Germany (excluding Berlin since the city was part of both East and West parts of Germany) in 1990. Because individuals are followed-up over time by the GSOEP, this is a cohort study. Sample sizes are provided in Table 1. They naturally decrease over time because of attrition.

Numerous income variables and socio-economic characteristics have been reported in the GSOEP. Unfortunately, labour income, which is our primary interest, is observed in East Germany only after 1992, where changes due to the reunification had already operated. However, the income reported in 1992 is the income of the previous year, namely 1991. Still, the initial period should be 1990, taking 1991 as the initial period would potentially attenuate the treatment effect. Again, this only strengthens our case if we want to test the existence of a treatment effect. Labour income corresponds to wage earnings before taxes and redistribution. We use a consumer price index to make income comparable over time and between regions. A substantial proportion of individuals do not participate to the labour market as seen in the second part of Table 1, having in the data set either a zero wage or having worked zero hour during the last surveyed year.[4] We take this indicator as a measure for participation to the labour market.

5.3 Inference results

We have selected the usual explanatory variables for explaining the decision to participate to the labour market: age, age squared, number of years of education, number of children (between 0 and 14 years old) in the household, replacement incomes (market income from other household members, public and social transfers received by the household). Monetary quantities are scaled by the GSOEP consumer price index and are expressed in thousands real euros. To take into account household size effects, we have divided household resources by the square root of the number of household members. We added a dummy indicating

[4]The data set contains an employment status variable. Those who are registered as working must have a positive wage and a positive number of working hours in the previous year. Those registered as unemployed can have a positive wage recorded for the previous year. So it is difficult, if not impossible to differentiate unemployment from not-participating-to-the-labour-market on the sole observation of a zero wage.

Table 1: Sample sizes and participation rates

	1990	1991	1992	1993	1994	1995
			Sample sizes			
Female	1802	1673	1568	1493	1436	1380
Male	1767	1640	1524	1433	1373	1304
Treatment group	2 798	2 594	2 423	2 304	2 200	2 113
Control group	771	719	669	622	609	571
			Participation rates			
Female	-	-	0.747	0.658	0.627	0.620
Male	-	-	0.872	0.809	0.774	0.772
Treatment group	-	-	0.806	0.736	0.697	0.696
Control group	-	-	0.816	0.719	0.706	0.688

The sample is constituted of individuals between 16 and 65 years old in 1990. Participation rate is calculated as the ratio of the employed to the working-age population.

one for the treatment group and zero for the control group, and an interaction term with education.[5] Results are based on 10 000 draws while 1 000 additional draws were used for warming the chain. They are reported in Tables 6 and 7 for females and males in the Appendix. These two equations serve to make inference on the IMR that enters into the two wage equations. Our Mincer-type wage equations explain the log-wage by the number of years of education, age (as a proxy for experience) and age squared. Number of children and replacement incomes that enter the participation equations are excluded for identification. This equation is conditional on draws of inverse Mills ratio (IMR) and models risk by means of heteroskedasticity. We introduce the treatment effect with the same binary dummy variable as in the selection equation. We allow the treatment variable to affect both the mean wage and its variance. Inference results are reported in Tables 8 and 9 in the Appendix for 1992 and 1995. We used the previous 10 000 draws from the selection equation and generate for each of these draws a new draw from the posterior distribution of the wage equation. We used 1 000 extra draws for warming the chain. Wages are scaled by the consumer price index and are expressed in thousands of real euros. Age is divided by 10 for scaling.

The Mills ratio is always significant and has a strong correcting effect both on the mean and the variance, confirming a non-negligible selection bias for both genders. Coefficients cannot be interpreted directly, so we report marginal effects in Table 2 for the treatment and for education. In a Type 2 Tobit model the marginal effect of a variable is given by the difference between the regression coefficient and a selection correction term (see e.g. Cameron and Trivedi 2005, page 552). We have clearly a huge difference between males and females. The treatment has a strong effect on females wages, but this effect is entirely due to the selection mechanism. On the contrary, the treatment has no effect on male

[5]Some values for the variable education were missing. We imputed the missing values by taking the sample average while including a flag dummy in the regression (not reported in the results).

Table 2: Marginal effects on average wage

	1992			1995		
	Impact	Selection	Total	Impact	Selection	Total
			Females			
Treatment	-0.004	**-0.742**	**-0.746**	-0.048	**0.496**	**0.448**
	(0.043)	(0.303)	(0.300)	(0.058)	(0.142)	(0.145)
Education	**0.090**	**-0.019**	**0.109**	**0.074**	**-0.086**	**0.160**
	(0.009)	(0.009)	(0.010)	(0.012)	(0.016)	(0.015)
			Males			
Treatment	0.040	0.091	0.130	**-0.104**	0.260	0.156
	(0.039)	(0.081)	(0.089)	(0.048)	(0.161)	(0.161)
Education	**0.055**	**-0.021**	**0.076**	**0.057**	**-0.051**	**0.107**
	(0.006)	(0.006)	(0.007)	(0.009)	(0.012)	(0.010)

Figure in bold correspond to variables for which the regression coefficient is at least twice its standard deviation.

wages when we correct for sample selection. The return of education are higher for women than for males even if both marginal effects are increasing between 1992 and 1995. The marginal effect of education reduces when we take into account sample selection, suggesting that less educated people tend to participate less.

Marginal effects are not concerned by heteroskedasticity modelling as they are defined by the derivative of the expected wage with respect to a given variable. It would also be difficult to qualify the gender gap on the basis of these marginal effects. On the contrary, GIC and PGC are function of both the mean and the variance, so they will help to fully qualify the distributional effects of the treatment between 1992 and 1995.

5.4 Distributional effects of the treatment

In a first step, we compare the effect of selection on distributional changes for males and females. As explained in subsection 3.3, we compute growth incidence curves for both participating individuals and for the whole sample. Comparing these two curves reveals the discrepancy between potential wages and wages for the sample of participants. More precisely, the difference between both curves indicate the change in the selection effect and not the direct impact of selection for a given year as it is done in Arellano and Bonhomme (2017). Clearly, Figure 1 indicates that the selection effect has dramatically increased for females. When using the the selected sample which means the sample of persons participating to the labour market, there is a strong similarity between males and females. There has been a slight increase in inequality as the income of the top 80% has increased more than the mean and the reverse for the lower deciles. On average participating females got a wage rise of 23%, slightly greater than that of males with 18%. Table 3 shows that at the 90% level, male and females are not different when considering only the participating individuals. Thus, there does not seem to be important differences in distributional changes between males and females for those participating. However, we observe significant differences between the GIC of participating females and that of potential female wages (except for the upper 0.80 quantiles) suggesting that the

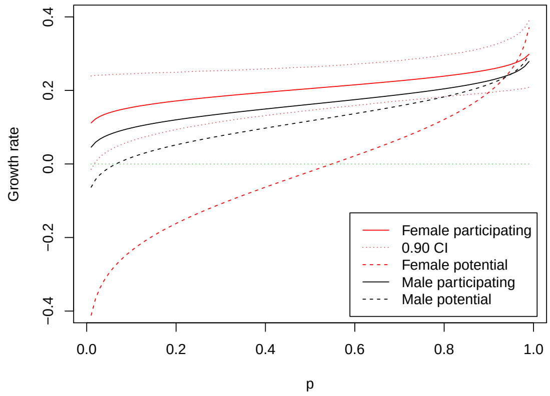

Figure 1: Potential and participating GIC for males and females (1992-1995)

gap between the potential distribution and that of participants has been increasing. The evolution of potential wages has been strongly detrimental for females, especially for lower quantiles. If all the females had decided to participate, given their personal characteristics, their 50% lower wages would have dramatically decreased while than that of males would have increased or, at least, being stable for lower wages. Therefore, focusing only on the evolution of the wages of participants would have hidden this fact as females were participating less than males (83% against 92% in 1992 and 69% against 83% in 1995).

Table 3: Testing for a selection effect: males and females

0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
Participating, $Females > Males$								
0.779	0.791	0.800	0.810	0.809	0.807	0.794	0.774	0.725
Potential, $Females < Males$								
0.990	0.989	0.989	0.984	0.977	0.953	0.904	0.806	0.630
Females, $Participating > Potential$								
1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.992	0.860
Males, $Participating < Potential$								
0.999	0.999	0.998	0.996	0.991	0.981	0.952	0.886	0.710

What is now the effect of exposure to Western television on the evolution of wages for males and females? Figure 2 compares separately for females and males the GIC of the treatment group and that of the control group both for participating individuals and for the whole sample. The left panel of Figure 2 suggests that the treatment had a strong effect on females participation. For those participating, the treatment led to a uniform wage growth around 20%, meaning that the fruits of growth were equally distributed among treated participating females. The potential situation of females was much less favourable as shown in Table 4. The female potential GIC is much below zero for the

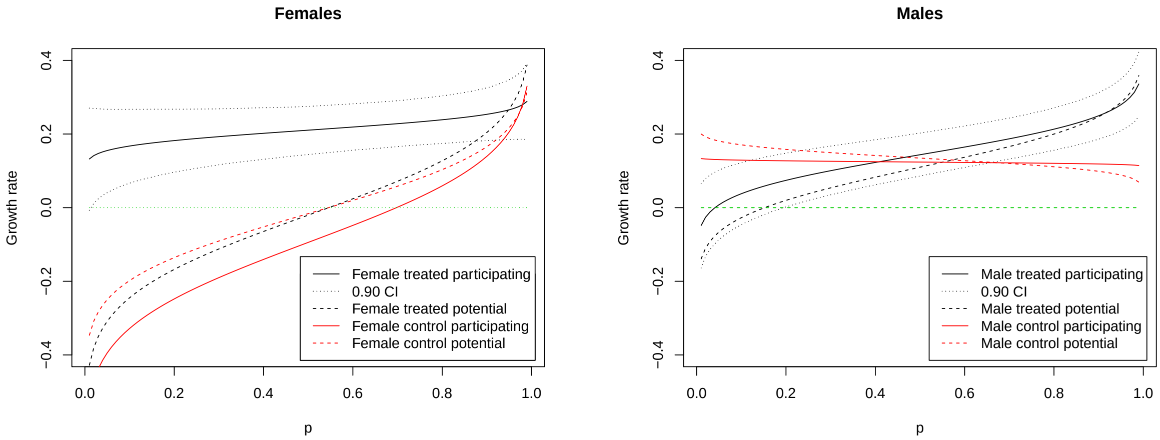

Figure 2: Treatment and control GIC (1992-1995)

lower quantiles, indicating that only females with the higher potential wages continued to participate to the labour market. Those with the lower potential wages were driven out of the labour market. In the UK, this happened for males according to Arellano and Bonhomme (2017). Interestingly, selection was not affected for the control group where the two curves are not statistically different, see Table 4 and participating females experienced a huge downgrade for the lower quantiles. Thus, the treatment effect on the evolution of female wages operates mainly through changes in participation. In term of stochastic

Table 4: Testing for a treatment effect on female wages

0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
Participating, $Treatment > Control$								
1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.971	0.866
Potential, $Control > Treatment$								
0.629	0.606	0.583	0.554	0.521	0.481	0.435	0.393	0.353
Treatment, $Participating > Potential$								
1.000	1.000	1.000	1.000	0.999	0.997	0.985	0.930	0.738
Control, $Potential > Participating$								
0.750	0.754	0.752	0.748	0.740	0.719	0.687	0.642	0.578

dominance, the wage distribution of 1995 dominates that of 1992 at the first order for treated and participating females (the probability for any quantile to be positive is always greater than 0.996). Thus, we could conclude that distributional changes occurring in the participating female treatment group has been welfare improving whereas this was the opposite for the control group where the lower wages of participating females have decreased.

For males, the story is totally different as seen when comparing the two panels of Figure 2. There is no impact of the treatment on participation because solid and dotted curves are similar (tested in Table 5). But the treatment had an impact on inequality which is increasing over the period. The lower quantiles experienced a zero growth while the highest quantile had a wage growth larger than that of females. On the contrary, the control group experienced an even wage increase of 10%, roughly equally distributed over the quantiles. In term of stochastic dominance, the distribution of 1995 dominates that of 1992 for the

Table 5: Testing for a treatment effect on male wages

0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
Participating, $Treatment > Control$								
0.219	0.297	0.385	0.480	0.590	0.698	0.797	0.879	0.939
Potential, $Control > Treatment$								
0.936	0.895	0.829	0.740	0.620	0.468	0.302	0.147	0.053
Treatment, $Participating > Potential$								
0.790	0.781	0.771	0.755	0.730	0.702	0.660	0.604	0.525
Control, $Potential > Participating$								
0.607	0.592	0.576	0.558	0.542	0.522	0.496	0.470	0.431

participating treated group at 95%, except for the first decile. This is also true for the participating control group, but at 90% (lower sample size), also except for the first decile.

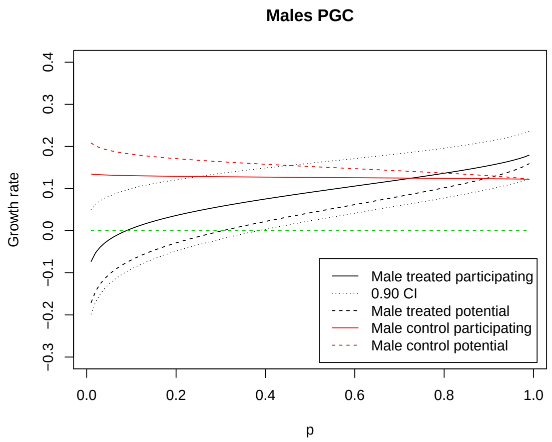

Figure 3: Treatment and control PGC for males (1992-1995)

Because treated and control curves intersect for male in the right panel of Figure 2, we cannot conclude that distributional changes occurring in the treatment group are preferred to those occurring in the control group. Therefore, we consider the PGC in Figure 3 which is related to second-order stochastic dominance. Although, the results are more striking with the PGC, we do not obtained a complete ordering.

6 Conclusion

Growth incidence and poverty growth curves have been proved to be very useful tools for the analysis of distributional changes. In this paper, we derive parametric growth incidence and poverty growth curves for a model of wage formation with sample selection. In the present paper, we provided Bayesian inference for these curves and provided test related to stochastic dominance.

We contribute to the literature on cultural economics by applying our method to estimate the distributional effect of Western television broadcasts on wages in East Germany after the reunification. We found first that selection played a very important role for female participation, the rate of participation falling down after the reunification. The phenomenon existed also for males, but to a much lesser extend. The impact of Western TV was prominent on females wages, and particularly on female labour participation. We found that it has significantly affected female labour participation and led the less productive females to drop out from the market, hiding thus a large increase in wage inequality among females. In contrast, Western television had significantly increased wage inequality among males.

Beyond these results, the present paper has some limits. Particularly, our methodology does not allow us to conclude about the persistence of this phenomenon, partly by lack of data (sample size and attrition), but mainly because the control group became treated after reunification as the Western television broadcasts progressively covered all the former East Germany. Indeed, the persistence of these effects could have feed the feeling of *ostalgie*, that is the nostalgia for the communist era in former German Democratic Republic. While our approach focus only on distributional changes limiting the data requirements to repeated cross-sections, investigating these issues will necessitate the use of longitudinal data to reveal income dynamics. We let these issues for future research.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.

Amemiya, T. (1985). *Advanced Econometrics*. Blackwell.

Araar, A., Duclos, J.-Y., Audet, M., and Makdissi, P. (2009). Testing for Pro-Poorness of Growth, With an Application to Mexico. *Review of Income and Wealth*, 55(4):853–881.

Arellano, M. and Bonhomme, S. (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica*, 85(1):1–28.

Atkinson, A. B. (1987). On the measurement of poverty. *Econometrica*, 55(4):749–764.

Bauwens, L., Lubrano, M., and Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

Bursztyn, L. and Cantoni, D. (2016). A Tear in the Iron Curtain: The Impact of Western Television on Consumption Behavior. *Review of Economics and Statistics*, 98(1):25–41.

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

Chancel, L. and Piketty, T. (2017). Indian income inequality, 1922-2014: From British Raj to billionaire Raj? Working Paper 2017/11, WID.world.

Chong, A. and Ferrara, E. L. (2009). Television and Divorce: Evidence from Brazilian Novelas. *Journal of the European Economic Association*, 7(2-3):458–468.

Clark, A. E. and Senik, C. (2010). Who Compares to Whom? The Anatomy of Income Comparisons in Europe. *The Economic Journal*, 120(544):573–594.

Davidson, R. and Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, 68(6):1435–1464.

Dinardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

Donald, S. G., Green, D. A., and Paarsch, H. J. (2000). Differences in Wage Distributions Between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates. *The Review of Economic Studies*, 67(4):609–633.

Duclos, J.-Y. (2009). What is Pro-Poor? *Social Choice and Welfare*, 32(1):37–58.

Ferreira, F. H., Firpo, S., and Galvao, A. F. (2019). Actual and counterfactual growth incidence and delta lorenz curves: Estimation and inference. *Journal of Applied Econometrics*, 34(3):385–402.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52(3):761–766.

Foster, J. and Shorrocks, A. (1988). Poverty orderings. *Econometrica*, 56(1):173–177.

Gastwirth, J. (1971). A general definition of the Lorenz curve. *Econometrica*, 39(6):1037–1039.

Genicot, G. and Ray, D. (2017). Aspirations and Inequality. *Econometrica*, 85(2):489–519.

Griffiths, W. (2001). Heteroskedasticity. In Baltagi, B. H., editor, *A Companion to Theoretical Econometrics*, chapter 4, pages 82–100. Blackwell Publishing Ltd, Oxford.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Hyll, W. and Schneider, L. (2013). The causal effect of watching TV on material aspirations: Evidence from the valley of the innocent. *Journal of Economic Behavior & Organization*, 86:37–51.

Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3):410–442.

Knight, J. and Gunatilaka, R. (2012). Income, aspirations and the hedonic treadmill in a poor society. *Journal of Economic Behavior & Organization*, 82(1):67 – 81.

Koop, G. (2003). *Bayesian Econometrics*. Wiley, Chichester, England.

Lakner, C. and Milanovic, B. (2016). Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession. *World Bank Economic Review*, 30(2):203–232.

Luttmer, E. F. P. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly Journal of Economics*, 120(3):963–1002.

Marin, J.-M. and Robert, C. P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer Texts in Statistics. Springer.

Mincer, J. (1974). Schooling, experience and earnings. Columbia university press, National Bureau of Economic Research, New York.

Novokmet, F., Piketty, T., and Zucman, G. (2017). From Soviets to Oligarchs: Inequality and Property in Russia, 1905-2016. Working Paper 23712, National Bureau of Economic Research.

Ravallion, M. and Chen, S. (2003). Measuring pro-poor growth. *Economics Letters*, 78:93–99.

Rosenfeld, R. A., Trappe, H., and Gornick, J. C. (2004). Gender and work in Germany: Before and after reunification. *Annual Review of Sociology*, 30:103–124.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica*, 50(197):3–17.

Son, H. H. (2004). A note on pro-poor growth. *Economics Letters*, 82(3):307 – 314.

Stutzer, A. (2004). The role of income aspirations in individual happiness. *Journal of Economic Behavior & Organization*, 54(1):89 – 109.

Veblen, T. (1899). *The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions*. The Macmillan Company.

Appendix

A Inference results

Table 6: Selection equation: Females

	1992			1995		
	Mean	S.D.	p-value	Mean	S.D.	p-value
Intercept	-4.383***	0.491	0.000	-2.754***	0.597	0.000
Age	3.205	0.191***	0.000	2.546***	0.217	0.000
Age^2	-0.437***	0.024	0.000	-0.348***	0.026	0.000
Education	0.112***	0.033	0.001	0.045	0.034	0.190
N.o. children 0-14	-0.247***	0.044	0.000	-0.444***	0.047	0.000
HH income	-0.020***	0.004	0.000	-0.025***	0.004	0.000
HH public transfers	-0.238***	0.017	0.000	-0.245***	0.014	0.000
HH social transfers	-0.137***	0.015	0.000	-0.138***	0.013	0.000
Treatment	0.955**	0.446	0.032	-1.474***	0.475	0.002
Treat x Education	-0.096**	0.037	0.010	0.117***	0.039	0.003

Treatment is a dummy variable which is 1 if the individual received western TV and 0 otherwise. Education represents the number of schooling years. HH income is the total income of the partner before redistribution and taxes. Public transfers are mainly social assistance and unemployment benefits. Social transfers are mainly pensions, child allowances and illness insurance. $p \in [0, 0.01]$ "***", $p \in [0.01, 0.05]$ "**", $p \in [0.05, 0.1]$ "*"

Table 7: Selection equation: Males

	1992			1995		
	Mean	S.D.	p-value	Mean	S.D.	p-value
Intercept	-1.855***	0.508	0.000	-4.499***	0.590	0.000
Age	2.227***	0.194	0.000	3.090***	0.224	0.000
Age^2	-0.297***	0.024	0.000	-0.402***	0.026	0.000
Education	0.036	0.030	0.228	0.064**	0.030	0.035
N.o. children 0-14	0.112**	0.047	0.016	0.135***	0.051	0.009
HH income	-0.020***	0.005	0.000	-0.018***	0.005	0.000
HH public transfers	-0.193***	0.017	0.000	-0.139***	0.014	0.000
HH social transfers	-0.158***	0.015	0.000	-0.104***	0.012	0.000
Treatment	-0.483	0.425	0.256	-0.201	0.447	0.653
Treat x Education	0.040	0.034	0.242	0.033	0.035	0.344

Treatment is a dummy variable which is 1 if the individual received western TV and 0 otherwise. Education represents the number of schooling years. HH income is the total income of the partner before redistribution and taxes. Public transfers are mainly social assistance and unemployment benefits. Social transfers are mainly pensions, child allowances and illness insurance. $p \in [0, 0.01]$ "***", $p \in [0.01, 0.05]$ "**", $p \in [0.05, 0.1]$ "*"

Table 8: Wage equation parameters: Females

	1992			1995		
	Mean	S.D.	p-value	Mean	S.D.	p-value
Intercept	1.142***	0.397	0.004	1.929***	0.557	0.001
Treatment	-0.004	0.044	0.925	-0.048	0.059	0.415
Age	0.090	0.204	0.659	-0.096	0.267	0.718
Age^2	0.008	0.026	0.750	0.033	0.033	0.317
Education	0.090***	0.009	0.000	0.074***	0.012	0.000
IMR	-0.936***	0.153	0.000	-1.024***	0.140	0.000
	Skedastic function					
	Mean	S.D.	p-value	Mean	S.D.	p-value
Sigma	0.521***	0.193	0.007	0.213**	0.090	0.018
Treatment	0.046	0.098	0.638	0.040	0.129	0.754
Age	-0.168***	0.041	0.000	-0.098**	0.049	0.046
Education	-0.013	0.021	0.550	0.044	0.032	0.168
IMR	2.212***	0.341	0.000	2.192***	0.309	0.000
IMR^2	-0.740***	0.191	0.000	-0.449***	0.132	0.001
Acceptation rate		0.657			0.386	

Treatment is a dummy variable which is 1 if the individual received western TV and 0 otherwise. IMR is the inverse Mills ratio. Age were divided by 10. $p \in [0, 0.01]$ "***", $p \in [0.01, 0.05]$ "**", $p \in [0.05, 0.1]$ "*"

Table 9: Wage equation parameters: Males

	1992			1995		
	Mean	S.D.	p-value	Mean	S.D.	p-value
Intercept	1.427***	0.303	0.000	3.555***	0.578	0.000
Treatment	0.040	0.038	0.299	-0.104**	0.048	0.031
Age	0.333**	0.146	0.022	-0.681**	0.271	0.012
Age^2	-0.028	0.018	0.119	0.102***	0.034	0.002
Education	0.055***	0.006	0.000	0.057***	0.009	0.000
IMR	-1.006***	0.167	0.000	-1.373***	0.215	0.000
	Skedastic function					
	Mean	S.D.	p-value	Mean	S.D.	p-value
Sigma	0.435***	0.109	0.000	0.259***	0.097	0.008
Treatment	-0.287***	0.097	0.003	0.426***	0.115	0.000
Age	-0.302***	0.033	0.000	-0.315***	0.048	0.000
Education	0.040**	0.017	0.016	0.064***	0.021	0.002
IMR	4.346***	0.414	0.000	2.503***	0.406	0.000
IMR^2	-1.782***	0.319	0.000	-0.349*	0.191	0.068
Acceptation rate		0.631			0.586	

Treatment is a dummy variable which is 1 if the individual received western TV and 0 otherwise. IMR is the inverse Mills ratio. Age were divided by 10. $p \in [0, 0.01]$ "***", $p \in [0.01, 0.05]$ "**", $p \in [0.05, 0.1]$ "*"