



HAL
open science

Bayesian inference for non-anonymous growth incidence curves using Bernstein polynomials: an application to academic wage dynamics

Edwin Fourier-Nicolaï, Michel Lubrano

► To cite this version:

Edwin Fourier-Nicolaï, Michel Lubrano. Bayesian inference for non-anonymous growth incidence curves using Bernstein polynomials: an application to academic wage dynamics. *Studies in Nonlinear Dynamics and Econometrics*, 2023, 10.1515/snde-2022-0109 . hal-04356211

HAL Id: hal-04356211

<https://hal.science/hal-04356211>

Submitted on 20 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Bayesian inference for non-anonymous Growth Incidence Curves using Bernstein polynomials: an application to academic wage dynamics*

Edwin Fourrier-Nicolai[†] Michel Lubrano[‡]

July 2023

*This paper was presented at the occasion of the 12th European Seminar on Bayesian Econometrics (ESOBE 2022) held in Salzburg, Austria in September 2022 in a session dedicated to Herman van Dijk to whom the Bayesian profession owes so much. Comments by the participants of ESOBE 2022, by Neil Shephard and by Herman van Dijk are gratefully acknowledged. During the writing of the first version this paper, we have benefited from very useful conversations with Mohammad Abu-Zaineh who provided a decisive help for identifying gender from names. The present version benefited from the remarks of two referees and of the editor. Discussions with Luc Bauwens are also gratefully acknowledged. Of course, remaining errors are solely ours.

The project leading to this publication has received funding from the French Government under the “France 2030” investment plan managed by the French National Research Agency (reference: ANR-17-EURE-0020) and from Excellence Initiative of Aix-Marseille University - A*MIDEX.

[†]School of International Studies and Department of Economics and Management, University of Trento, Italy., e-mail: fourrier.edwin@gmail.com.

[‡]Aix-Marseille Univ., CNRS, AMSE, 5 Bd Maurice Bourdet, 13001 Marseille, France, e-mail: michel.lubrano@univ-amu.fr.

Abstract

The paper examines the question of non-anonymous Growth Incidence Curves (na-GIC) from a Bayesian inferential point of view. Building on the notion of conditional quantiles of Barnett (1976), we show that removing the anonymity axiom leads to a complex and shaky curve that has to be smoothed, using a non-parametric approach. We opted for a Bayesian approach using Bernstein polynomials which provides confidence intervals, tests and a simple way to compare two na-GICs. The methodology is applied to examine wage dynamics in a US university with a particular attention devoted to unbundling and anti-discrimination policies. Our findings are the detection of wage scale compression for higher quantiles for all academics and an apparent pro-female wage increase compared to males. But this pro-female policy works only for academics and not for the para-academics categories created by the unbundling policy.

Keywords: Conditional quantiles, non-anonymous GIC, Bayesian inference, academic wage formation, gender policy.

1 Introduction

One of the main economic question is to understand how economic growth affects different segments of the population, i.e., whether everyone benefits from growth or only a small group of individuals and which ones. Growth incidence curves have been introduced in Ravallion and Chen (2003) to address this question. The purpose was to visualise who had benefited from growth and in particular if growth has been pro-poor or not. The basic ingredient of a GIC consists in comparing two quantile functions to explain changes in the income distribution. However, it soon appeared that this initial tool had a limit. When comparing quantiles, a GIC ignores the trajectory of individuals belonging to these quantiles, relying on individual anonymity. Only the global change in the income distribution was under scrutiny. If we want to take into account social mobility for evaluating welfare, our interest should rely in fact on a bivariate income distribution for building a complete growth incidence curve, instead of considering two independent quantile functions associated to two marginal income distributions.

The problem is however not trivial. A quantile function consists in proposing an ordering and as noted by Barnett (1976), ordering and quantiles are a one dimension property. There is no natural concept of a bivariate ordering for a bivariate distribution $F(x, y)$. Barnett (1976) has proposed

a series of four sub-ordering principles in order to circumvent the problem: *marginal ordering*, *reduced ordering*, *partial ordering* and finally *conditional ordering*. The economic literature about non-anonymous GIC (see e.g. Grimm 2007, Van Kerm 2009, Jenkins and Van Kerm 2016 or Palmisano and Peragine 2015 to quote just a few) has implicitly retained the notion of conditional ordering of Barnett (1976). In a conditional ordering, a first dimension is chosen, usually the one corresponding to $t - 1$ or the initial state. Conditionally on that ordering, the second component (corresponding to t) is ordered by blocks. This approach allows to point out income dynamics. So the literature on income dynamics considers individual transitions between income classes (which can correspond to marginal quantile intervals) when the non-anonymous GIC (na-GIC) relies on a conditional quantile function. Therefore, transition matrices explain the probability of going from one group of income to another group in the next period. The fact that these matrices are built around groups operate a kind of smoothing because individuals within one group are supposed to evolve similarly. In contrast, the conditional quantile function, when it does not operate on blocks, can follow the trajectory of each individual, and two individuals coming from the same initial quantile group do not have necessarily the same trajectory. So the graph of a conditional quantile function can be quite shaky and needs smoothing to become intelligible. Consequently, as we shall see later on, the estimation of a na-GIC becomes fundamentally a non-parametric econometric problem.

The aim of the present paper is threefold. First, we want to characterise the na-GIC as a growth rate computed between a marginal quantile function defined at $t - 1$ and a conditional quantile function defined at t . Second, we propose inference for na-GIC using a Bayesian semi-parametric framework based on Bernstein polynomials. Finally, we apply this methodology to examine wage dynamics in academia with a particular attention devoted to unbundling and anti-discrimination policies.

The paper is organised as follows. In section 2, we present two concepts for a na-GIC, show that both lead to a non-parametric regression and show how Bernstein regression can be a nice solution. We detail the properties of Bernstein polynomials and show how they are a nice alternative to cubic B-splines. In section 3, we present a Bayesian inference for this problem. We then detail how to compare models in order to select the degree k of the Bernstein polynomial and show which kind of prior is sufficient for using a marginal likelihood to select the degree of Bernstein polynomials. We end this section by providing statistical tests for growth incidence curves. Section 4 is devoted to analysing how wage growth is distributed over the quantiles in a large US public University, examining the gender pay gap in a different way. Section 5 concludes.

2 A non-parametric model for na-GIC based on Bernstein polynomials

The aim of the Growth Incidence curve of Ravallion and Chen (2003) was to measure the distributive impact of growth, and more specifically if growth was more favorable to certain segments of the distribution than others. Their answer is provided by the inspection of the growth rate of each quantile of two marginal income distributions. Let $F_{t-1}(x)$ and $F_t(y)$ represent these two marginal income distributions, a quantile function can be inferred from each distribution separately so that the GIC can be approximated by:

$$g_t(p) \simeq \log F_t^{-1}(p) - \log F_{t-1}^{-1}(p) = \log y(p) - \log x(p), \quad (1)$$

where $x(p)$ and $y(p)$ are the respective quantile functions (or the derivative of the respective generalised Lorenz curves). It is worth mentioning that there is no need to have the same individuals or households in x and y . From a statistical point of view, two cross-section samples are sufficient for estimating $g_t(p)$. The function $g_t(p)$ simply describes how growth has deformed the initial distribution. The procedure relies on an anonymity axiom, which means that this measure is independent of the initial ranking of the individuals.

There are however relevant reasons for removing the anonymity axiom. The statistician or the social planner may care about individual trajectories. To reach that goal, we should no longer consider two marginal distributions $F(x)$ and $F(y)$, but the entire joint distribution of incomes in $t - 1$ and t , that is $F(x, y)$. This is the starting point of all the literature on non-anonymous GIC (na-GIC), as well as that of income or poverty dynamics. However removing the anonymity axiom entails to solve much more complex statistical problems, one of the objectives of the present paper.

2.1 Two concepts for na-GIC

The term and concept of a na-GIC has been formally introduced in the literature by Grimm (2007). Let us start from the income quantile $x(p)$ and call p_x the order corresponding to this quantile function, so that $x(p_x)$ are the order statistics. Let assume that we observe the same individuals in the next period and that $y(p)$ is the quantile function of the second period. The na-GIC of Grimm (2007) is simply defined as:

$$g(p_x) = \log y(p_x) - \log x(p_x). \quad (2)$$

It corresponds thus to considering a quantile function computed from the initial income distribution $F(x)$ which is compared to a conditional quantile

function extracted from $F(y|X = x(p_x))$. So the initial order is given by $F(x)$ and maintained over the second period. This is a conditional ordering in the sense of Barnett (1976).

Van Kerm (2009) goes a step further (but without using the term na-GIC) as he relates what he calls an income mobility profile (in fact a na-GIC) to the income mobility literature, saying that “*An income mobility profile is a graphical tool to portray income mobility and identify the association between individual movements and initial status*”. Starting from the bivariate income distribution $F(x, y)$, he defines a distance function $\delta(x, y)$ that measures the income growth for an individual between $t - 1$ and t . A mobility index M is then defined as:

$$M = \int_0^\infty \int_0^\infty \delta(x, y) dF(x, y). \quad (3)$$

Because we are not interested in mobility for itself, but only in a certain type of mobility, for instance upward mobility or progressive mobility (see e.g. Benabou and Ok 2001), the distance function $\delta(x, y)$ has to be directional as detailed in Jenkins and Van Kerm (2016), verifying the following properties:

$$\delta(x, x) = 0, \quad \delta(x, y) = -\delta(y, x), \quad \delta(x, \rho x) > 0 \text{ for } \rho > 1. \quad (4)$$

Scale or translation invariance can also be added. Possible choices for the distance function are $\delta(x, y) = \log y - \log x$ for proportional income growth or $\delta(x, y) = y - x$ for absolute income growth.

Let us define $p_x = F_X(x)$ as the rank of income x at period $t - 1$ and $x(p_x)$ the corresponding quantile function. The mobility profile, or in other terms the na-GIC, is the function $m(p)$ defined as:

$$m(p) = m(X, Y|X = x(p_x)) = \delta(x(p_x), y(p_x)). \quad (5)$$

The simple GIC would correspond to the distance function $\delta(x(p_x), y(p_y))$ as indicated in Jenkins and Van Kerm (2016).

Dominance conditions to compare two growth situations were developed with different results in Bourguignon (2011), Jenkins and Van Kerm (2016) and Palmisano and Peragine (2015). Starting with a S-Gini, Jenkins and Van Kerm (2006) decompose inequality changes between progressivity and re-ranking components. All these papers have considered the economic status in the first period p_x as the reference. Lo-Bue and Palmisano (2020) adopt another point of view and consider complete poverty trajectories and their associated dominance conditions, independently of the initial status.

2.2 Non-anonymous GIC as a non-parametric problem

The natural estimator for the quantile function $x(p)$ is based on the definition of order statistics with:

$$x(p = i/n) = (x_{[1]}, \dots, x_{[i]}, \dots, x_{[n]}), \quad (6)$$

where $x_{[i]}$ are the order statistics coming from the n ordered observations such that $x_{[1]} < x_{[2]} < \dots < x_{[n]}$. This natural estimator leads to a fairly smooth curve with respect to p . There is no imperious need for smoothing even if Yang (1985) has proposed a kernel estimate for the quantile function which is equivalent to a Nadaraya-Watson non-parametric regression of the order statistics $x_{[i]}$ over its rank i/n .

The na-GIC in equation (2) is based on the difference between the log of a conditional quantile and the log of a marginal quantile. What would be a natural estimator for a conditional quantile? By analogy with the natural estimator of the marginal quantile $x(p = i/n)$, a natural estimator for the conditional quantile of $y|x$ is obtained by first defining the order of x that we call $(i_{[1]}, i_{[2]}, \dots, i_{[n]})$ so that $x_{[j]} = x_{i_{[j]}}$. This is equivalent to ordering the bivariate variable (x, y) according to its first component x . The natural estimator for a conditional quantile is then given by:

$$y(p = i_{[j]}/n) = (y_{i_{[1]}}, y_{i_{[2]}}, \dots, y_{i_{[n]}}). \quad (7)$$

However, this natural estimator provides a very shaky function of p , so that a generalisation of the non-parametric estimator of Yang (1985) proposed for the usual quantile function becomes very useful in this case.

The na-GIC of Grimm (2007) provided in equation (2) considers the difference between two quantile functions and thus could justify the smoothing of these two quantile functions separately. The mobility profile (5) of Van Kerm (2009) on the contrary does not proceed by considering the difference of two smoothed quantiles, but by defining the desired function as a conditional expectation in itself, starting directly from the $\delta(\cdot)$ function. More precisely, the problem is defined as a regression of $\delta(x(p_x), y(p_x))$ over p_x . Because the conditional expectation is a non-linear function of p for which a parametric form is not evident, a non-parametric approach is needed. Van Kerm (2006) has proposed to use the local weighted regression of Cleveland (1979) to regress $\delta(x(p_x), y(p_x))$ over p_x , a method also used in Jenkins and Van Kerm (2016).¹ It has better properties at boundaries than the Nadaraya-Watson kernel regression. In this paper, we propose to use Bernstein polynomials as

¹The method is available in R with the function `loess`, implying the choice of a degree of smoothing.

they are more suitable for depicting changes at boundaries while Bayesian inference can be derived straightforwardly.

2.3 Smoothing using Bernstein polynomials

Bernstein polynomials have been popular among engineers for approximating a complicated function $f(x)$ for $x \in [0, 1]$. The approximation $g_k(x)$ of $f(x)$ is given by the following polynomial of order k :

$$g_k(x) = \sum_{j=0}^k f(j/k) C_k^j x^j (1-x)^{k-j} = \sum_{j=0}^k f(j/k) b_k(x, j), \quad (8)$$

where C_k^j is the binomial coefficient and $b_k(x, j) = C_k^j x^j (1-x)^{k-j}$. The approximation can be made as precise as desired by increasing the order k of the polynomial. It was used to prove the Weierstrass approximation theorem. It is worth mentioning that if the support of x is outside the segment $[0, 1]$, say at values on the segment $[a, b]$, it can be transformed so as to lie in the required $[0, 1]$ interval, using $y = (x - a)/(b - a)$.

Stadtmuller (1986), Tenbusch (1997) were the first to propose Bernstein polynomials for curve estimation in a problem where the observations are n observed couples (y_i, x_i) :

$$y_i = m(x_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2). \quad (9)$$

The approximation of the unknown function $m(x_i)$ is given by running the following regression:

$$y_i = \sum_{j=0}^k \beta_j b_k(x, j) + \epsilon_i. \quad (10)$$

In this model, the β_j are the estimates of $f(j/k)$ so $g_k(x)$ is now defined by:

$$g_k(x) = \sum_{j=0}^k \beta_j C_k^j x^j (1-x)^{k-j}. \quad (11)$$

The regressors are deterministic functions of x and k forming a basis Z defined as:

$$Z = [z_i] = [b_k(x, 0), \dots, b_k(x, k)], \quad (12)$$

with $k + 1$ columns and n rows. So the Bernstein coefficients $f(j/k)$ are estimated by a linear regression model without an intercept:

$$y_i = z_i \beta + \epsilon_i. \quad (13)$$

The degree k of the polynomial is directly related to the degree of smoothing introducing a common bias-variance trade-off. A higher k will reduce the bias at the cost of increasing the number of parameters, estimated with less and less precision.

Remark 1 *The matrix Z has some interesting properties. Whatever k , its rows sum to 1.0, which explains the fact that regression (13) has to be run without adding an intercept. When $k = 1$, the predicted values of y form a straight line. So $(\hat{\beta}_0 + \hat{\beta}_1)/2$ is exactly equal to the mean of y . For $k > 1$, the predicted values of y usually depart from a straight line, so the average of the regression coefficients are only close to the mean of y . For $k \geq 3$, the vector formed by the row sum of the outer columns ($Z[, 1] + Z[, (k+1)]$) is equivalent to 1 minus the vector formed by the row sum of all the inner columns, i.e. $\text{rowSums}(Z[, 2:k])$, up to a rounding error, using the syntax of R.*

2.4 Comparison with other methods

Bernstein polynomials have a certain number of advantages for non-parametric regression. Brown and Chen (1999) underline that the competing method of kernel regression has problems at the boundaries of x . Solving this problem would mean considering different kernels inside the same regression problem. This is naturally done when using Bernstein polynomials as can be seen from the left panel of Figure 1 where we have displayed the elements of the basis Z for $k = 10$. This property is particularly relevant since we are generally concerned by the very poor and very rich in inequality studies. Moreover in our particular case the explanatory variable is naturally in the segment $[0,1]$, with no transformation needed. Bernstein polynomials behave

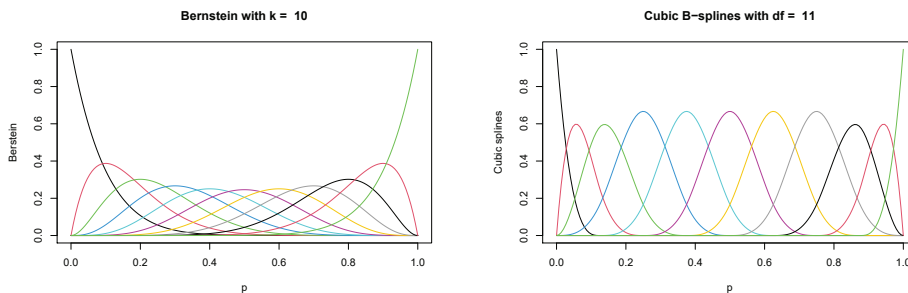


Figure 1: Comparing Bernstein polynomials with Cubic B-splines

like a variable asymmetric kernel, with different shapes depending on the value of the explanatory variable p .

A more serious competitor to Bernstein polynomials are splines which can be seen as another way of building the basis Z when estimating regression (13). Splines are defined by minimising a specific loss function in order to approximate the regression function $m(x)$:

$$\sum_{i=1}^n (y_i - m(x_i))^2 + \lambda \int (m''(x))^2 dx, \quad (14)$$

where λ is a smoothing parameter monitoring the smoothing penalty expressed as the second order derivative of the curve $m(x_i)$ to be found. Following Hardle (1990, pp. 56-65), this minimisation problem has a unique solution $m_\lambda(x)$ over the space of twice differentiable functions which is provided by cubic splines.² We have plotted on the right panel of Figure 1 cubic B-splines with $df = 11$ (including a constant term).³ It results that Bernstein polynomials and cubic B-splines reproduce the same general pattern.⁴

This can be shown through a small Monte Carlo experiment. We run $m = 1,000$ simulations with n draws from a bivariate lognormal distribution with parameters

$$\mu = (1.0, 1.2), \quad \Sigma = \begin{pmatrix} 0.50 & -0.20 \\ -0.20 & 0.10 \end{pmatrix}, \quad (15)$$

and computed the empirical estimate of the na-GIC with $\delta(x(p_x), y(p_x))$. We then adjusted a linear regression model without intercept using first Bernstein polynomials, then B-splines and finally natural splines. We have chosen various values of k for Bernstein polynomials and $df=k+1$ for splines, leading the to the same number of regressors in each case.⁵ In Table 1, we report the RMSE of each regression and its evaluation for the lower and upper 5% tails. We consider $k = 3, 5, 10$ and $n = 100, 1000$. In all cases, Bernstein polynomials should be preferred to splines, and particularly to natural splines. Estimates based on Bernstein polynomials have a greater or equal accuracy and precision whatever is the sample size or the degree of complexity of the model. The advantage of Bernstein polynomials is especially visible for lower and higher quantiles, provided the value of k is not too small. As underlined before, this property is particularly interesting when one is concerned by

²See Hardle (1990, pp. 56-65) for more detailed and historical references.

³We have used the `bs` function of the R package `splines`.

⁴The similarity is due to the use of B-splines. If we had used natural splines, (function `ns` of the R package `splines`), we would have obtained totally different shapes, in particular at the lower boundary.

⁵The R commands are `bs(x, df=k+1, intercept = TRUE)` for B-splines and `ns(x, df=k+1, intercept = TRUE)` for natural splines.

Table 1: The relative performance of Bernstein polynomials versus splines

	$n = 100$			$n = 1000$		
	Bernstein	B-splines	Natural splines	Bernstein	B-splines	Natural splines
	$k = 3$					
$\hat{\sigma}$	0.184 (0.025)	0.184 (0.025)	0.200 (0.030)	0.178 (0.009)	0.178 (0.009)	0.196 (0.010)
$\sigma_{\hat{0.05}}$	0.268 (0.126)	0.268 (0.126)	0.326 (0.141)	0.309 (0.054)	0.309 (0.054)	0.372 (0.056)
$\sigma_{\hat{0.95}}$	0.277 (0.141)	0.277 (0.141)	0.334 (0.158)	0.309 (0.054)	0.309 (0.054)	0.372 (0.056)
	$k = 5$					
$\hat{\sigma}$	0.166 (0.019)	0.169 (0.021)	0.182 (0.026)	0.158 (0.006)	0.162 (0.007)	0.177 (0.008)
$\sigma_{\hat{0.05}}$	0.214 (0.114)	0.226 (0.121)	0.274 (0.140)	0.237 (0.047)	0.254 (0.048)	0.308 (0.053)
$\sigma_{\hat{0.95}}$	0.205 (0.103)	0.217 (0.110)	0.266 (0.129)	0.236 (0.044)	0.253 (0.046)	0.307 (0.050)
	$k = 10$					
$\hat{\sigma}$	0.151 (0.013)	0.153 (0.014)	0.161 (0.019)	0.147 (0.004)	0.149 (0.005)	0.159 (0.006)
$\sigma_{\hat{0.05}}$	0.147 (0.064)	0.161 (0.074)	0.203 (0.102)	0.178 (0.030)	0.192 (0.034)	0.244 (0.045)
$\sigma_{\hat{0.95}}$	0.147 (0.066)	0.164 (0.078)	0.205 (0.108)	0.179 (0.031)	0.193 (0.036)	0.246 (0.045)

Standard errors are indicated between parenthesis.

the very poor and very rich in inequality studies. Note nevertheless that B-splines are not very far, compared to the poor result of natural splines (which are known to have problems at the boundaries). The experiment relies on a automatic choice for the knots, based on the quantiles. All the regressors were significant in the experiment.

A final note on splines. One can think that with an optimal choice of the knots, splines could lead to a slightly better fit than Bernstein polynomials where the only possible choice is the value of k . This value is usually selected using an information criteria such as the BIC or comparing marginal likelihoods in a Bayesian framework as detailed below. There is no precise method for choosing the knots in splines. Usually one starts from a great number of knots around the quantiles of the explanatory variable and then selects the significant knots by a variable selection method which can be quite tedious. In a Bayesian framework, Dimatteo et al. (2001) have provided an automatic Reversible-Jump MCMC algorithm to select the knots. However, the procedure is time consuming, compared to the simple information criteria used to select the degree of a Bernstein polynomial.

2.5 Parametric restrictions

Choosing a high value for k increases the accuracy of the approximation to $m(x)$, but can lead to over-fitting. There is thus an incentive to be able to introduce restrictions on the sequence of the β parameters in order to constraint the shape of $\hat{m}(x)$. Those restrictions are introduced by different sets of inequality constraints which have their origin from considering the first or the second order derivatives of $g_k(x)$. Let us give the first derivative as an example:

$$g'_k(x) = k \sum_{j=0}^{k-1} k(\beta_{j+1} - \beta_j) C_{k-1}^j x^j (1-x)^{k-j-1}. \quad (16)$$

The sign of $g'_k(x)$ determines if $m(x)$ is monotone increasing or decreasing. The sign of each of the k elements of this sum depends only on the sign of $\beta_{j+1} - \beta_j$ so that the monotonicity can be described by a restriction of the form $A\beta > 0$ with A being a $k \times (k+1)$ matrix defined by:

$$A = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & & 1 & -1 \end{pmatrix}. \quad (17)$$

The restriction $A\beta > 0$ means that the k differences $\beta_{j+1} - \beta_j$ are positive, which implies that $m(x)$ is monotone increasing (monotone decreasing for $A\beta < 0$). See Wang and Ghosh (2012) for other types of A matrices implying for instance concave or convex restrictions. Wang and Ghosh (2012) propose a classical estimator for β satisfying this type of restrictions. Using a Bayesian approach, Ghosal and Ghosh (2022) propose a general inference method for linear regression models with linear inequality constraints. An over-parameterised regression model can lead to ambiguity about the general shape of a na-GIC as seen in the next example.

Example 1 *We took the same parameters as in the previous example, but inflated the variance parameter of the second period by 2. For smoothing the na-GIC obtained from this bivariate distribution, we used a Bernstein regression with $k = 5$. The graph of the first order derivative as reported in the right panel of Figure 2 indicates that for $k = 5$ the derivative curve is always negative, implying a strict negative slope for the smoothed na-GIC. However, with $k = 20$ which implies a large over-parametrisation, the derivative can be positive on some parts of the segment $[0,1]$. This illustrates the danger of over-fitting and the necessity to choose carefully the order of the polynomial k .*

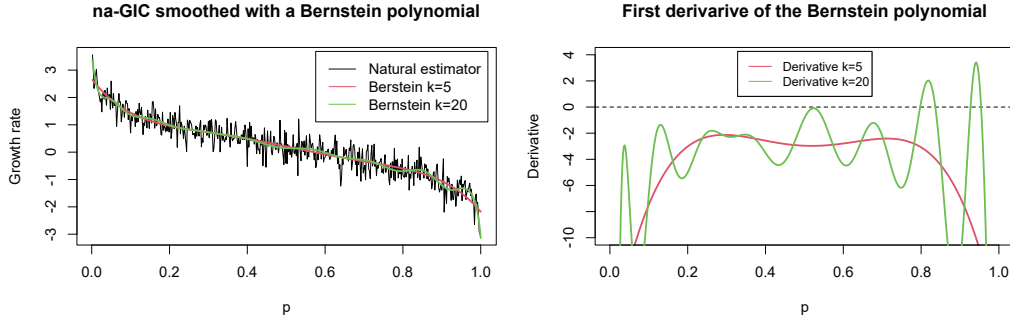


Figure 2: Impact of smoothing and under-smoothing in a Bernstein regression

3 Bayesian inference for na-GIC

In this section, we develop Bayesian inference for na-GIC. We derive the posterior na-GIC. We discuss the choice of the polynomial's order and we demonstrate how the choice of the prior can help in model selection. Last, we provide statistical tests for growth incidence curves.

3.1 Posterior densities

Under normality of the error term, the likelihood function is:

$$l(y|\theta, k) \propto (\sigma^2)^{-n/2} |Z(k)'Z(k)|^{-1/2} \exp -\frac{1}{2\sigma^2} (y - Z(k)\beta)'(y - Z(k)\beta). \quad (18)$$

In this notation, $Z(k)$ is the Bernstein basis and its size depends on k , which explains the presence of the factor $|Z(k)'Z(k)|^{-1/2}$. When the analysis is conducted conditionally on k , this factor can be neglected.

As we are in the framework of the linear regression model, there exist a natural conjugate prior on (β, σ^2) with:

$$\varphi(\sigma^2) = f_{IG}(\sigma^2|s_0, \nu_0), \quad (19)$$

$$\varphi(\beta|\sigma^2) = f_N(\beta|\beta_0, \sigma^2 M_0), \quad (20)$$

while a non-informative prior would mean:

$$\varphi(\beta) \propto 1, \quad \varphi(\sigma^2) \propto 1/\sigma^2. \quad (21)$$

Following standard textbook results (see e.g. Bauwens et al. 1999, pages 56-64), the posterior density of σ^2 and β are respectively an inverted gamma2 and a Student:

$$\varphi(\sigma^2|y) = f_{IG}(\sigma^2|s_*, \nu_*), \quad \varphi(\beta|y) = f_t(\beta|\beta_*, M_*, s_*, \nu_*), \quad (22)$$

where the hyper-parameters are given by:

$$M_* = Z(k)'Z(k) + M_0, \quad (23)$$

$$\beta_* = M_*^{-1}(Z(k)'y + M_0\beta_0), \quad (24)$$

$$s_* = y'y - \beta_*'M_*\beta_* + s_0 + \beta_0'M_0\beta_0, \quad (25)$$

$$\nu_* = n + \nu_0. \quad (26)$$

We are interested in non-linear transformations of the parameters in order to get the posterior density of the na-GIC. From the student posterior density of β , we can obtain m draws, so that we can build a $m \times np$ matrix M of posterior draws from the posterior density of the na-GIC on a grid of np predetermined points of p . These draws are obtained using the transformation $\beta_i^{(j)}b_k(x, i)$, where $\beta_i^{(j)}$ is the j^{th} draw of β_i leading to:

$$M[j, \cdot] = \sum_{i=0}^k \beta_i^{(j)} C_k^i p^i (1-p)^{k-i}, \quad (27)$$

From this matrix, we can determine a posterior confidence interval for the na-GIC, by selecting specific row quantiles of M , (e.g. 0.05 and 0.95 for a 10% confidence interval). In the same vein, we can construct a $m \times np$ matrix Ms of the derivative of the na-GIC. Elements of this second matrix are based on the following transformation:

$$Ms[j, \cdot] = k \sum_{i=0}^{k-1} (\beta_{i+1}^{(j)} - \beta_i^{(j)}) C_{k-1}^i p^i (1-p)^{k-i-1}. \quad (28)$$

Constraints on the sign of the first derivative can be imposed with the constraint $A\beta > 0$, following Koop (2003, pp.77-80). The prior density (20) is transformed so as to obtain:

$$\varphi(\beta|\sigma^2) \times \mathbb{1}(A\beta > 0), \quad (29)$$

where $\mathbb{1}(\cdot)$ is the indicator function. There is no longer an analytical formula for the posterior density of β , but the later can easily be evaluated using a simulation method. We simply draw random numbers from the unconstrained posterior Student posterior $\varphi(\beta|y) = f_t(\beta|\beta_*, M_*, s_*, \nu_*)$ and reject the draws that do not verify the constraint $A\beta > 0$. Even if constraining the shape of the smoothed $m(x)$ can be appealing, it does not solve the issue of selecting the order k of the polynomial.

3.2 Model comparison

The choice of k can be seen as a variable selection problem like in Curtis and Ghosh (2009) and Choi et al. (2016). They both use the approach of Geweke (1996) while they are in a context where they want to impose restrictions on the shape of the Bernstein polynomial. However, even if they are confronted to the same type of restrictions, Ding and Zhang (2016) prefer to use the usual information criteria, i.e. the BIC, the AIC or the deviance information criteria (DIC) of Spiegelhalter et al. (2002), in order to select the degree k of the Bernstein basis. These quantities are:

$$BIC(k) = -2 \log p(y|\hat{\theta}, k) + (k + 1) \log(n), \quad (30)$$

$$AIC(k) = -2 \log p(y|\hat{\theta}, k) + 2(k + 1), \quad (31)$$

$$DIC(k) = -2E_{\theta}[\log(p(y|\theta))] + p_D. \quad (32)$$

In this writing, $p(y|\hat{\theta}, k)$ is the posterior density or its approximation by the likelihood function and $\hat{\theta}$ is the point where maximum of the posterior density is reached. The expectation needed for the DIC is obtained in the same way. Finally, p_D is the effective dimension of the model as suggested in Spiegelhalter et al. (2002). Because we have no hidden parameters, the penalty p_D can be simplified to $2(k + 1)$.

These selection methods work even when the prior density is non-informative. In the case of an informative prior density, one can also rely on the marginal likelihood to compare models as it is a direct ingredient for posterior odds.⁶ Following Koop (2003, page 41), we can write the marginal likelihood of our linear model as:

$$p(y|M) = (|M_0|/|M_*|)^{1/2} \times s_0^{\nu_0/2} / s_*^{\nu_*/2} \times \Gamma(\nu_*/2) / \Gamma(\nu_0/2) \times \pi^{-n/2}. \quad (33)$$

The model with the higher marginal likelihood is selected. Knowing the structure of the Bernstein bases, the choice of appropriate prior information can help to determine k . The minimum possible model is with $k = 1$ so that $Z(p, 1)\beta$ represents the equation of a straight line and the average of β equals the empirical mean of y . When $k > 1$, the two extreme parameters β_0 and β_k continue to play a crucial role in representing the general shape of the curve when imposing the constraint $\beta_1 = \dots = \beta_{k-1} = 0$. So the idea is to be non-informative on β_0 and β_k while imposing the prior information

⁶We are grateful to a referee for this suggestion. Note that the BIC is an asymptotic approximation of the marginal likelihood $p(y|M)$ where M is the model to be evaluated as $p(y|M) = \exp(-BIC/2 + O(1))$. See e.g. Konishi and Kitagawa (2008, page 215, section 9.1.3) for a derivation using Laplace method. This might be useful for explaining some of our results.

that $\beta_1 = \dots = \beta_{k-1} = 0$ in a probabilistic way. This is obtained for a prior expectation equal to zero ($\beta_0 = 0$) and devising a specific prior precision matrix M_0 . Usually prior precision matrices are not easy to specify and the G-prior of Zellner (1986) is a convenient solution. It results that our prior precision matrix has the following form:

$$M_0 = \gamma_0 \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \tilde{Z}'\tilde{Z} & \vdots \\ 0 & \dots & 0 \end{pmatrix} \quad (34)$$

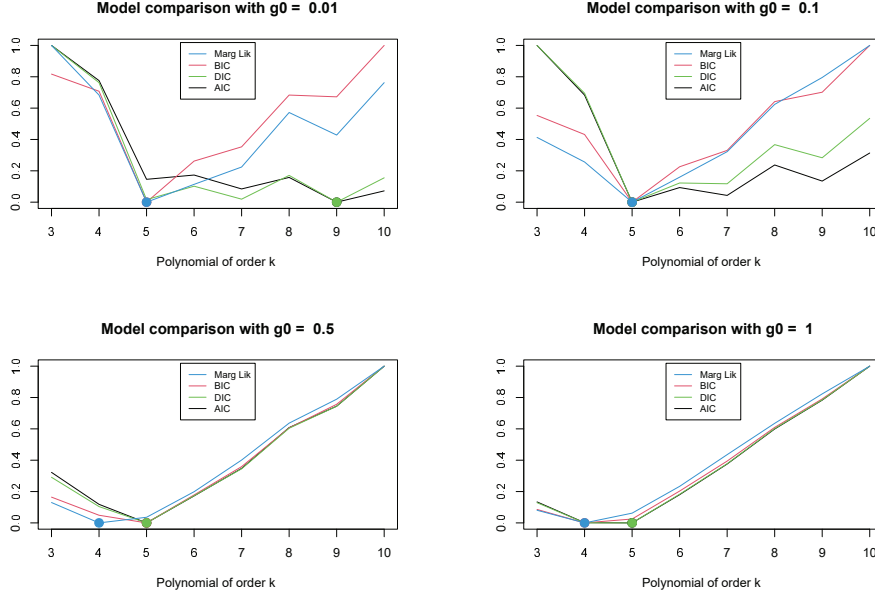
where $\tilde{Z} = Z[2 : k]$ is a submatrix of Z obtained by removing its first and last columns. γ_0 is a tuning parameter that monitors the strength of the prior information. The prior on σ^2 is of less importance and we can choose for instance $s_0 = 1$ and $\nu_0 = 3$.⁷ Note that the proposed solution implied that the precision matrix is singular, so that (33) cannot be evaluated. We propose to replace the determinant $|M_0|$ in (33) by $\gamma_0^{k-1}|\tilde{Z}'\tilde{Z}|$ as by definition $|\tilde{Z}'\tilde{Z}|$ is strictly positive. For numerical reasons, we shall compute the log of the marginal likelihood.

Example 2 *We continue using our previous simulated data to illustrate the role of γ_0 for selecting k and visualize our results in Figure 3. With a very small $\gamma_0 = 0.01$, marginal likelihood and BIC provide the same answer of $k = 5$ (remember that BIC is an asymptotic approximation of the marginal likelihood), while DIC and AIC select a less parsimonious model with $k = 9$ as expected. When increasing $\gamma = 0.1$, all the criteria provide the same answer $k = 5$. When increasing again $\gamma = 0.5$, the marginal likelihood select a more parsimonious model with $k = 4$ while the other criteria still prefer $k = 5$. With the very strong $\gamma = 1.0$, the DIC selects $k = 5$, while the other criteria select $k = 4$.*

3.3 Statistical tests

The Bayesian approach is not only particularly relevant for model selection but also for statistical tests. Most of hypotheses can be tested in a straightforward manner. Let us start with the simple growth incidence curve under the anonymity axiom. We simply need to have a convenient modelling of the two quantile functions indexed respectively by θ_1 and θ_2 . Using m draws from the posterior density of these parameters and build a $m \times np$ matrix of

⁷For an alternative way of building a prior density that incorporates restrictions, see Bauwens et al. (2022) in a context of time series and long memory.



The highest value of each criterion is normalised to 1.0 and for the marginal likelihood we have plotted 1.0 minus the normalised vector of the log marginal likelihood.

Figure 3: Bayesian model choice

draws from the adequate transformation of these draws:

$$M_{GIC}[j, \cdot] = \log y(p|\theta_2^{(j)}) - \log x(p|\theta_1^{(j)}). \quad (35)$$

For modelling the quantile functions $y(p|\theta_2)$ and $x(p|\theta_2)$, we can use the simple model of Kakwani (1980) detailed in Fourier-Nicolai and Lubrano (2021). For each quantile p_i of matrix M_{GIC} , we can compute the empirical probability that $g(p_i) > 0$ to know which quantiles have significantly grown:

$$\Pr g(p_i) > 0 \simeq \frac{1}{m} \sum_{j=1}^m \mathbb{1}(M_{GIC}[j, i] > 0), \quad (36)$$

where $\mathbb{1}(\cdot)$ is the indicator function. We could also test if certain quantiles are growing more rapidly than the rest of the distribution (e.g. if we want to test for pro-poor growth) by computing the empirical probability that $g(p_t) > \gamma_t$ where γ_t is the average growth rate. In doing so, we compare two marginal distributions using their quantile functions.

With the na-GIC, the statistical problem is quite similar once we have obtained m draws from the parameters and stored their transformation defined in (27). The interpretation is of course different as we look at the

individual trajectories. Let us call the matrix (27) M_{naGIC} . The probability that $m(p) = \delta(x(p_x), y(p_x)) > 0$ is computed as:

$$\Pr(m(p_i) > 0) \simeq \frac{1}{m} \sum_{j=1}^m \mathbb{1}(M_{naGIC}[j, i] > 0). \quad (37)$$

We can also compute the probability that the slope of the na-GIC is negative for all p_i using the matrix Ms of draws (28) corresponding to the the first order derivative of the na-GIC curve:

$$\Pr\left(\frac{\partial m(p_i)}{\partial p_i}\right) \simeq \frac{1}{m} \sum_{j=1}^m \mathbb{1}(Ms_{naGIC}[j, i] < 0). \quad (38)$$

Numerous other hypotheses can be easily tested in a Bayesian approach provided they can be expressed in terms of the posterior parameters.

4 An application to wage academic dynamics

Understanding wage dynamics in academia is critical for society as it is at the core of producing knowledge, innovation and intellectual capital. Stephan (1996), in her survey paper, underlines the importance of the life cycle model as a major mechanism for earning functions. But there is no dynamics in these functions as they are estimated on cross sections. Her numerous followers tried to find complementary mechanisms, introducing various measures of productivity which could, for instance, revert the negative relation between wages and seniority (e.g. Moore et al. 1998). Gender issues came next as an important concern. The recent paper of Blackaby et al. (2005) is considering a small survey of 349 UK economists for estimating a probit equation to explain rank achievement differences between males and females, underlying the importance of outside offers, more important for males than for females. For the Canadian University of Manitoba, Brown et al. (2011) attempt at explaining the evolution of the gender pay gap between 1993 and 2003, following the introduction of an anti-discrimination policy in that university. They derive the Wellington (1993) variant of the Oaxaca-Blinder decomposition to explain the nature of changes. The major change is that for full professors, more females were promoted but the yield of that rank became lower for females. In this study, we aim to analyse the dynamics of academic wages in light of the recent policy trends and their relation with the gender pay gap.

4.1 Institutional background and data

Michigan State University (MSU) is a very large US public university with more than 50,000 students. As all US public universities, it has the legal obligation of publishing the wages of its members. It was possible to find on its web site detailed pdf files containing names, wages, grade, type of contract and seniority for two years 2006 and 2012, covering a total of 8,964 observations from which we build a panel of 4,855 individuals belonging to five different broad colleges with Social Sciences, Science, Medicine, Agriculture and Humanities.⁸ Building a panel means that we have kept all the individuals present in 2006, keeping for the while those who have left in 2012, but excluding those who enter only in 2012. The balanced panel needed for computing a na-GIC requires further exclusion of all those who have left MSU in 2012. With the complete panel, we have a clear idea of individual trajectories which cover both wage increase and promotion. Because we have a clear description of their status, we can try to investigate the impact on those trajectories that a major phenomenon could have, the unbundling as described in e.g. Macfarlane (2011).

As in many UK and US universities, MSU is practicing an unbundling policy which intends to separate the traditional academic tasks of teaching, research and service into distinct jobs by introducing new positions like *Educator*, *Instructor*, *Lecturer*, *Specialist* and *Librarian-Archivist*. For Macfarlane (2011), the objective was to “*lower the costs of a university education and obtain better productivity from those working to deliver it*”. As a matter of fact, these new jobs, which require only a Bachelor or a Master degree, have a much lower mean wage: \$51,907 against \$96,494 for regular academics (Assistant, Associate, Full, Endowed Professors) in 2006.⁹ These members are most of the time recruited on a fixed term contract and the proportion of females is more important. However, Macfarlane (2011) notes also that this system leads universities “*to restructure their reward systems around para-academic or specialist professional career paths*”. As a matter of fact, the wage increase we observe at MSU is 24% for academics and 28% for para-academics. Consequently, there is an interest to assess to what extent their wage dynamics differ.

An important institutional feature at MSU is the anti-discrimination policy. We can find the following announcement on MSU web site in 2022:

⁸The college affiliation is detailed in the original data base. After those dates, the information content of the published files diminished, names disappeared and 2012 is the last useful available date.

⁹These figures were computed with the fully balanced panel, excluding thus those who were present in 2006, but have left in 2012.

Thus, even if not illegal, acts are prohibited under this policy if they discriminate against any University community member(s) through inappropriate limitation of employment opportunity, access to University residential facilities, or participation in education, athletic, social, cultural, or other University activities on the basis of age, color, gender, gender identity, disability status, height, marital status, national origin, political persuasion, race, religion, sexual orientation, veteran status, or weight. [...] For purpose of this Policy, “employment opportunity” is defined as job access and placement, retention, promotion, professional development, and salary.

So we can question gender pay gaps, keeping in mind previous studies like Brown et al. (2011) who were examining the effectiveness of a policy introduced in 1993 at the University of Manitoba. Note also that external factors can explain the gender pay gap; for instance Monroe and Chiu (2010) who were trying to check if gender discrimination can be due to an insufficient number of female candidates in the hiring pool.

4.2 Descriptive statistics

Let consider the balanced panel of individuals who were present both in 2006 and 2012.¹⁰ The characteristics of this sample are displayed in Tables 2 and 3, indicating the distinction between males and females. Unfortunately, the gender was not reported in the original dataset, then it was determined manually using the first names. There were 47 unresolved cases for which only first name’s initials were provided.

Table 2: Sample characteristics for academics stayers in 2006

Title	Total		Females		Males		Fem. freq ratio	Fem. wages ratio
	Freq	Salary	Freq	Salary	Freq	Salary		
Assistant Prof	462	69,538	206	66,289	252	72,379	0.446	0.953
Associate Prof	469	88,227	152	83,172	314	90,650	0.324	0.943
Full Prof	580	115,832	119	113,401	449	116,682	0.205	0.979
Endowed Prof	73	166,557	14	152,211	58	170,256	0.192	0.914
Total	1584	96,494	491	85,384	1,073	101,555	0.310	0.885

The last two columns indicate the ratio between females values and total average values.

There is a clear decreasing proportion of females when climbing the ladder of promotions. Female wages are lower than the average from 2% to 8%, and

¹⁰Note that this leads to exclude all individuals leaving the MSU between 2006 and 2012.

the difference is maximum for Endowed Professors (8%). This is in coherence with the results found in the literature.

Table 3: Sample characteristics for para-academics stayers in 2006

Title	Total		Females		Males		Fem. freq ratio	Fem. wages ratio
	Freq	Salary	Freq	Salary	Freq	Salary		
Instructor	125	41,579	75	41,472	49	41,978	0.600	0.997
Lecturer	10	47,088	3	48,069	7	46,668	0.300	1.021
Educator	132	46,073	73	41,006	59	52,342	0.553	0.890
Research Assoc	16	52,357	12	52,662	4	51,445	0.750	1.006
Specialist	239	58,286	122	53,845	117	62,916	0.510	0.924
Libra-Archiv	51	63,235	28	62,781	23	63,789	0.549	0.993
Total	573	51,907	313	48,585	259	56,007	0.564	0.888

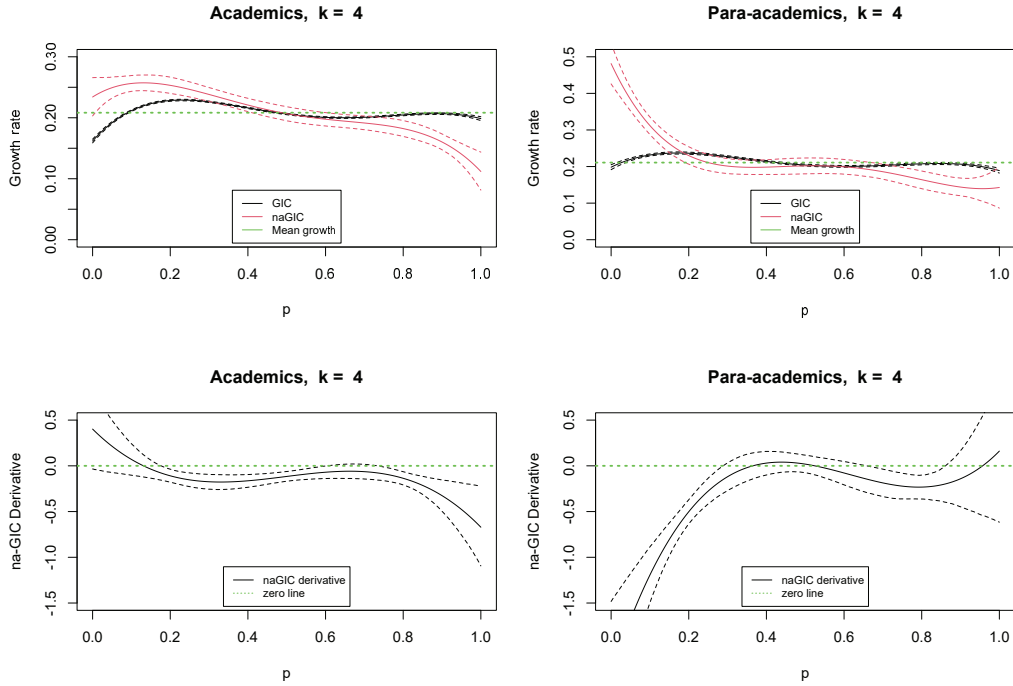
The last two columns indicate the ratio between females values and total average values.

The configuration is quite different with para-academic in Table 3. The proportion of females is 0.564, and this majority is found in all the categories, except for the marginal category Lecturer. On average, female earn 89% of the average wage. But this difference is not at all homogeneous. For two grades, females earn slightly more than the average. But for the highest categories of specialist and librarian, they earn less than average. So finally, wage differential between males and females is more complex than what we found in the literature, once we take into account the unbundling.

4.3 na-GICs for academics and para-academics

Now we consider wage dynamics by computing the na-GIC. While the marginal likelihood suggests $k = 3$, for the academics and $k = 4$ for the para-academics, we select $k = 4$ in both cases for the sake of consistency. Figure 4 displays slightly different wage dynamics between academics and para-academics. In the top panels, we provide GIC and na-GIC for academics on the left and the ones for para-academics are provided on the right. While the simple GIC indicates that all quantiles have benefited homogeneously from growth, the na-GIC instead indicates that individuals at the lower quantiles tend to experience greater wage growth than those on the upper part of the distribution.

The lower part of the panel provides the derivatives of the na-GIC for academics and para-academics. For academics, the curve is increasing till the second decile and decreasing after. Indeed, the na-GIC is over the average growth rate up to the fourth decile, and lower after. Its slope is regularly negative after the second decile. The na-GIC of the para-academics is quite different. The first decile experienced a very high wage increase, up to 40%.



The bottom panel displays the corresponding derivative of the na-GIC, which provides an indication if the na-GIC is increasing or decreasing.

Figure 4: Who benefited from wage growth among Academics and para-academics

The slope of the na-GIC is roughly zero between the third and seventh deciles and for those deciles the wage increase is equal to the mean. Starting from the eighth decile, the wage increase is lower than the mean, but the slope of the curve is uncertain.

4.4 Gender issues

It is clear from Table 2 and 3 that females were earning less than males, especially for academic positions in 2006. Brown et al. (2011) question the role of gender policy in diminishing the gender gap at the University of Manitoba. Their findings suggest that the gap did not disappear, but its nature had changed. na-GICs can be used so as to show how the general wage increase was distributed over the quantiles and that for males and females separately. So by comparing the two na-GIC, we can look if a pro-active policy was effective to decrease the gender pay gap.

Using the same degree $k = 4$, the left panel of Figure 5 shows that for roughly all quantiles, the wage for female academic was slightly larger than

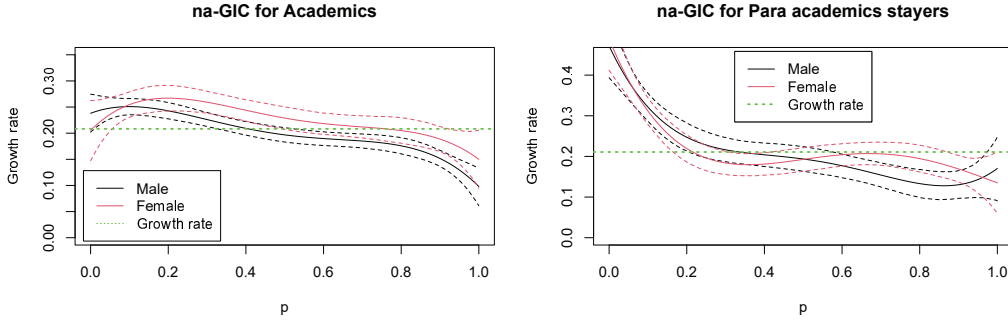


Figure 5: Gender issues for wage dynamics

for male. However, we do not find the same configuration for para-academics in the right panel. So if there was a pro-active compensation policy, it was effective only for the most visible part of the academic staff, as it was earlier suggested in Monroe and Chiu (2010). In their view, anti-discrimination statements should rather be considered as an insurance against the risk of lawsuits than a goal in itself. Indeed, Title IX (of the Education Amendments of 1972) prohibits discrimination in education programs receiving federal funds. Consequently administrations fear recognising gender discrimination because of possible lawsuits. It results that anti-discrimination policies are essentially a mean to “*ultimately protects the institution against claims of discrimination*”, leading to an heterogeneity of responses according to the level of risk.

We can complement the interpretation of these plots by computing the probability that female got a higher pay rise than male for each point of the curves. The results are provided in Table 4. The overall probability

Table 4: Statistical tests

p	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
	Academic males										
$naGIC > \gamma$	0.91	1.00	1.00	0.99	0.54	0.08	0.01	0.00	0.00	0.00	0.00
	Academic females										
$naGIC > \gamma$	0.49	1.00	1.00	1.00	1.00	0.94	0.80	0.60	0.40	0.09	0.05
	Comparing academic males and females										
$\Pr(Female > Male)$	0.21	0.57	0.91	0.99	0.99	0.98	0.98	0.96	0.95	0.98	0.90
Average	0.86										
	Comparing para-academic males and females										
$\Pr(Female > Male)$	0.59	0.35	0.18	0.10	0.17	0.48	0.87	0.98	0.99	0.91	0.29
Average	0.54										

that females got a higher pay rise is 0.86 for academics, which confirms our

Table 5: Promotion or administration?

	Males		Females	
	Promotion	Admi	Promotion	Admi
Assist-Assoc	0.371	0.019	0.324	0.026
Assoc-Full	0.255	0.061	0.201	0.065
Full-Endow	0.038	0.072	0.023	0.070
Endow		0.064		0.200

These transition probabilities were computed using the unbalanced panel.

visual interpretation of the na-GIC plots. But this probability reduces to 0.54 for para-academics, thus raising doubts on the existence of an effective anti-discrimination policy aiming to reduce the initial gender pay gap.

Taking an administrative position is said to be a secure way of getting a wage increase as noted in Hamermesh et al. 1982. Table 5 indicates that females have a higher tendency to take an administrative position, especially when they are Endowed Professors. This is certainly part of the wage increase we found with the na-GIC in Table 4. Para-academics do not have this possibility, which might explain partly the fact that the gender pay gap persists in this category.

5 Conclusion

In this paper, we have defined a na-GIC based on conditional ordering of a bivariate distribution, following one of the options depicted in Barnett (1976). This is coherent with the usual way of defining a na-GIC in the literature, taking the initial ranking of the first period p_x as the reference. For instance, when analysing if growth was pro-poor, the measure of Grimm (2007) describes what has happened to those who were initially poor. But, as underlined in Lo-Bue and Palmisano (2020), this approach ignores what happens to those who become poor during the second period. Taking the first period as the reference is justified on the ground of life trajectories, careers in our empirical application, as initial conditions determine greatly what is happening next. However, the reference ordering can be of particular importance, specifically when assessing welfare. This motivates Lo-Bue and Palmisano (2020) to prefer promoting more robust welfare criteria that take into account all individual trajectories and not only those determined by the initial conditions. To come back to Barnett (1976) paper, this would lead to consider another ordering than conditional ordering, such as for instance what he calls *reduced ordering*. However, by adopting such an ordering, we

would depart from the quantile transition matrices of Formby et al. (2004) and from all the related literature on the meaning of income mobility.

We have demonstrated that estimation of na-GIC is essentially a non-parametric problem for which we have proposed a Bernstein regression in order to model and smooth a conditional quantile function. In particular, Bernstein polynomials seem to be more efficient for depicting changes in extreme quantiles than other traditional methods. Bayesian inference turns out to be particularly useful in this approach as it allows to develop straightforwardly statistical tests associated with growth incidence curves but also to help in model selection. In particular, we have proposed priors which can help to determine the degree of the polynomial and we have derived four criteria (marginal likelihood, BIC, AIC and DIC) for model selection in the context of na-GICs.

We have applied our methodology to examine wage dynamics at the Michigan State University between 2006 and 2012. Our results indicate that the unbundling policy leads to two different wage mobility patterns while this policy may worsen the efforts to alleviate gender discrimination. Furthermore, our results suggest that this may partly due to the unequal access to administrative promotions between academics and para-academics.

This paper contributes to the literature on income mobility by proposing a new tool aiming to capture extreme variations among the lower and upper parts of the distribution. Its main practical limitation relies on its inability to take into account sample attrition which is an important empirical issue. Being beyond the scope of this paper, we let this to further research.

References

- Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, 139(3):318–355.
- Bauwens, L., Chevillon, G., and Laurent, S. (2022). We modeled long memory with just one lag! LIDAM Discussion Papers CORE 2022016, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).
- Bauwens, L., Lubrano, M., and Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.
- Benabou, R. and Ok, E. A. (2001). Mobility as progressivity: Ranking income processes according to equality of opportunity. Working Paper 8431, NBER.

- Blackaby, D., Booth, A. L., and Frank, J. (2005). Outside offers and the gender pay gap: Empirical evidence from the UK academic labour market. *The Economic Journal*, 115(501):F81–F107.
- Bourguignon, F. (2011). Non-anonymous growth incidence curves, income mobility and social welfare dominance. *The Journal of Economic Inequality*, 9(4):605–627.
- Brown, B. M. and Chen, S. X. (1999). Beta-Bernstein smoothing for regression curves with compact support. *Scandinavian Journal of Statistics*, 26(1):47–59.
- Brown, L. K., Troutt, E., and Prentice, S. (2011). Ten years after: Sex and salaries at a Canadian university. *Canadian Public Policy*, 37(2):239–255.
- Choi, T., Kim, H.-J., and Jo, S. (2016). Bayesian variable selection approach to Bernstein polynomial regression model with stochastic constraints. *Journal of Applied Statistics*, 43(15):2751–2771.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.
- Curtis, S. M. and Ghosh, S. K. (2009). A variable selection approach to Bayesian monotonic regression with Bernstein polynomials. Technical report, University of Washington.
- Dimatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, 88(4):1055–1071.
- Ding, J. and Zhang, Z. (2016). Bayesian regression on non-parametric mixed-effect models with shape-restricted Bernstein polynomials. *Journal of Applied Statistics*, 43(14):2524–2537.
- Formby, J. P., Smith, W. J., and Zheng, B. (2004). Mobility measurement, transition matrices and statistical inference. *Journal of Econometrics*, 120(1):181–205.
- Fourrier-Nicolai, E. and Lubrano, M. (2021). Bayesian inference for parametric growth incidence curves. In Bandyopadhyay, S., editor, *Research on Economic Inequality: Poverty, Inequality and Shocks, Volume 29*, chapter 2, pages 31–55. Emerald Publishing.

- Geweke, J. (1996). Variable selection and model comparison in regression. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics, 5*, volume 5, pages 609–620. Oxford University Press.
- Ghosal, R. and Ghosh, S. K. (2022). Bayesian inference for generalized linear model with linear inequality constraints. *Computational Statistics and Data Analysis*, 166:107335.
- Grimm, M. (2007). Removing the anonymity axiom in assessing pro-poor growth. *The Journal of Economic Inequality*, 5(2):179–197.
- Hamermesh, D. S., Johnson, G. E., and Weisbrod, B. A. (1982). Scholarship, citations and salaries: Economic rewards in economics. *Southern Economic Journal*, 49(2):472–481.
- Hardle, W. (1990). *Applied Nonparametric Regression*. Econometric Society Monographs. Cambridge University Press.
- Jenkins, S. P. and Van Kerm, P. (2006). Trends in income inequality, pro-poor income growth, and income mobility. *Oxford Economic Papers*, 58(3):531–548.
- Jenkins, S. P. and Van Kerm, P. (2016). Trends in individual income growth: Measurement methods and British evidence. *Economica*, 83(332):679–703.
- Kakwani, N. (1980). On a class of poverty measures. *Econometrica*, 48(2):437–446.
- Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer.
- Koop, G. (2003). *Bayesian Econometrics*. Wiley, New-York.
- Lo-Bue, M. C. and Palmisano, F. (2020). The individual poverty incidence of growth. *Oxford Bulletin of Economics and Statistics*, 82(6):1295–1321.
- Macfarlane, B. (2011). The morphing of academic practice: Unbundling and the rise of the para-academic. *Higher Education Quarterly*, 65(1):59–73.
- Monroe, K. R. and Chiu, W. F. (2010). Gender equality in the academy: The pipeline problem. *PS: Political Science and Politics*, 43(2):303–308.
- Moore, W. J., Newman, R. J., and Turnbull, G. K. (1998). Do academic salaries decline with seniority? *Journal of Labor Economics*, 16(2):352–366.

- Palmisano, F. and Peragine, V. (2015). The distributional incidence of growth: A social welfare approach. *Review of Income and Wealth*, 61(3):440–464.
- Ravallion, M. and Chen, S. (2003). Measuring pro-poor growth. *Economics Letters*, 78:93–99.
- Spiegelhalter, D. J., Best, Nicola G. and Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64(4):583–639.
- Stadtmüller, U. (1986). Asymptotic properties of nonparametric curve estimates. *Periodica Mathematica Hungarica*, 17(2):83–108.
- Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature*, 34(3):1199–1235.
- Tenbusch, A. (1997). Nonparametric curve estimation with Bernstein estimates. *Metrika*, 45:1–30.
- Van Kerm, P. (2006). Comparisons of income mobility profiles. ISER Working paper 2006-36, ISER, University of Essex.
- Van Kerm, P. (2009). Income mobility profiles. *Economics Letters*, 102:93–95.
- Wang, J. and Ghosh, S. (2012). Shape restricted nonparametric regression with Bernstein polynomials. *Computational Statistics and Data Analysis*, 56(9):2729–2741.
- Wellington, A. J. (1993). Changes in the male/female wage gap, 1976-85. *The Journal of Human Resources*, 28(2):383–411.
- Yang, S.-S. (1985). A smooth nonparametric estimator of a quantile function. *Journal of the American Statistical Association*, 80(392):1004–1011.
- Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Goel, P. and Zellner, A., editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, volume 6 of *Studies in Bayesian Econometrics and Statistics*, pages 233–243. Elsevier, New York.