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

Edwin Fourrier-Nicolai, Michel Lubrano

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

Working Papers / Documents de travail

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

Edwin Fourrier-Nicolai
Michel Lubrano

WP 2022- Nr 27

Aix*Marseille université Socialement engagée
A*Midex Initiative d'excellence Aix-Marseille
FRANCE 2030
Faculté d'Économie et de Gestion Aix*Marseille Université
cnrs
L'ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
Centrale Marseille

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

Edwin Fourrier-Nicolai† Michel Lubrano‡

November 30, 2022

**Abstract**

This 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 non-parametric inference problem. From a Bayesian point of view, an approach using Bernstein polynomials provides a simple solution and immediate confidence intervals, tests and a way to compare two na-GIC. The paper illustrates the approach to the question of academic wage formation and tries to shed some light on wether academic recruitment leads to a super stars phenomenon, that is a large increase of top wages, or not. Equipped with Bayesian na-GIC's, we show that wages at Michigan State University experienced a top compression leading to a shrinking of the wage scale. We finally analyse gender and ethnic questions in order to detect if the implemented pro-active policies were efficient.
**Keywords:** Conditional quantiles, non-anonymous GIC, Bayesian inference, wage formation, gender policy, ethnic discrimination.
**JEL classification:** C11, C22, I23.

*During the writing of this paper, we have benefited from very useful conversations with Mohammad Abu-Zaineh who provided a decisive help for identifying gender and ethnicity from names. Previous versions of this paper were presented at the LAGV June 2022 in Marseille, ESOBE September 2022 in Salzboug and at AMSE in October 2022. We thank the participants for their useful comments which led to substantial revisions. 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.

# 1 Introduction

Growth incidence curves (GIC) have now a long history starting with Ravallion and Chen (2003). 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 *conditional ordering*. The economic literature about non-anonymous GIC (see e.g. Grimm 2007, Van Kerm 2009, Jenkins and Van Kerm 2011 or Lo-Bue and Palmisano 2020 to quote just a few) has implicitly retained the notion of conditional ordering of Barnett (1976). In the conditional ordering, a first dimension is chosen, 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 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 follows 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 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 at $t$. Second, we propose inference for na-GIC using a Bayesian semi-parametric framework. Finally, we examine an empirical question concerning academic wage formation using the tool of na-GIC in order to take fully into account wage dynamics and examine gender and ethnic questions.

The paper is organised as follows. In section 2, we review a part of the existing literature on the non-anonymous growth incidence curve. We then show that this particular curve can be seen as a non-parametric problem. In section 3, we show how Bernstein polynomials could be a convenient tool for deriving the shape of this curve. We propose a method to select the optimal degree of the polynomial. We finally propose Bayesian inference for this problem. In section 4, we illustrate the method by analysing wage dynamics at Michigan State University. Gender and ethnic questions are examined in section 5. Section 6 concludes.

# 2 GIC and Non-anonymous GIC

We first review the definition of a GIC, before exploring how the literature has developed the notion of na-GIC. We finally explain why the na-GIC results in a non-parametric estimation problem.

## 2.1 Anonymous GIC

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 pro-poor or not. 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:

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

where $x(p)$ and $y(p)$ are the respective quantile functions or the derivative of the respective generalised Lorenz curves. 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. This measure simply provides an indication on how growth has impacted the shape of the income distribution. It is based on an anonymity axiom, which means that this measure is independent of the initial ranking of the individuals. Using that framework, a vast literature compares and ranks two income distributions. In particular, there is a correspondence between the shape of a GIC and stochastic dominance at the order one. Duclos (2009) and Araar et al. (2009) have shown how to build classical stochastic dominance tests from a GIC and also how to test for pro-poor growth when lower quantiles increase more than average.

All these questions have been investigated from a Bayesian viewpoint in Fourrier-Nicolai and Lubrano (2021). Essentially the question was to find a parametric representation for a quantile function. A first solution consists in modelling the underlying income distribution using simple densities for which the quantile function has a closed analytical form. If we assume, for example, that the income distribution at each of the two periods can be modelled by a log-normal distribution, we have a direct analytical form for the quantile function and consequently for the GIC. However, in this case the shape of the GIC is very much constrained as its slope depends solely on the difference between inequality in the first and in the second period. Flexibility can be reached provided the income distribution is modelled according to a mixture model for the underlying income distribution. However, in this case, the quantile function is semi-explicit and has to be evaluated numerically. Another way consists in adjusting directly a functional form for the Lorenz curve and deriving its first-order derivative to find the corresponding quantile function. Fourrier-Nicolai and Lubrano (2021) finds particularly suitable the Kakwani (1980)'s functional form based on the Beta distribution to fit a GIC.

There are however relevant reasons for removing the anonymity axiom. The statistician or the social planner may care about the dynamics of income. To reach that goal, we should no longer consider the 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 much more complex problems, especially if we want to adopt a parametric Bayesian approach. The solutions explored in Fourrier-Nicolai and Lubrano (2021) for the simple GIC cannot be easily generalised. For instance Bayesian inference for mixtures of bivariate log-normal distributions is not an easy task. So a new approach has to be found and this is one of the objectives of the present paper.

## 2.2 Non-anonymous GIC

The term and concept of 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, in fact the increasing rank of the observations so that $x(p_x)$ are the order statistics. We assume that in the next period we observe the same individuals. The quantile function of the second period corresponds to $y(p)$. The na-GIC defined in Grimm (2007) corresponds to:

$$g(p_x) = \log y(p_x) - \log x(p_x). \tag{2}$$

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). Following the approach of Ravallion and Chen (2003), Grimm (2007) computes the rate of pro-poor growth as:

$$\frac{1}{q}\int_0^q g_t(p_x)\mathrm{d}p_x, \tag{3}$$

where $q$ is the poverty head-count for a poverty line $z$ and the income distribution $F(x)$. This measures the variation over time of the Watts poverty index for those who were poor during the first period.

Van Kerm (2009) goes 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 is then defined as:

$$M = \int_0^\infty \int_0^\infty \delta(x, y)\mathrm{d}F(x, y).$$

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 (2011), verifying the following properties:

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

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 and $\delta(x, y) = y - x$ for absolute income growth.

Let us decompose the joint distribution $F(x, y)$ into the product of marginal and conditional distributions:

$$F(x, y) = F_{Y|x}(y)\, F_X(x),$$

so that the previous mobility index becomes:

$$\begin{aligned} M &= \int\int \delta(x, y) \mathrm{d}F_{Y|x}(y)\, \mathrm{d}F_X(x) && (4) \\ &= \int m(X, Y)|X = x) \mathrm{d}F_X(x) && (5) \\ &= \int_0^1 m(X, Y|X = x(p))\, \mathrm{d}p, && (6) \end{aligned}$$

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

$$m(p) = m(X, Y|X = x(p_x)) = \delta(x(p_x), y(p_x)), \tag{7}$$

when seen as a function of $p_x$. The simple GIC would correspond to the distance function $\delta(x(p_x), y(p_y))$ as indicated in Jenkins and Van Kerm (2011).

Dominance conditions to compare two growth situations were developed with different results in Bourguignon (2011), Jenkins and Van Kerm (2011) 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.

## 2.3 Non-anonymous GIC as a non-parametric problem

The natural estimator for the quantile function $x(p)$ is quite simple. It is based on the definition of order statistics. This is the simple empirical quantile function corresponding to:

$$x(p = i/n) = (x_{[1]}, \cdots, x_{[i]}, \cdots, x_{[n]}), \tag{8}$$

where $x_{[i]}$ are the order statistics coming from the $n$ ordered observations. This natural estimator leads to a pretty smooth aspect when we plot it against $p$. 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-parameter regression of the order statistics $x_{[i]}$ over $i/n$.

The na-GIC in (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 $(i_{[1]}, i_{[2]}, \ldots, i_{[n]})$ so that $x_{[j]} = x_{i_{[j]}}$. This is equivalent to ordering the bivariate variable $(x, y)$ according to $x$. The natural estimator for a conditional quantile is then given by:

$$y(p = i_{[j]}/n) = (y_{i_{[1]}}, y_{i_{[2]}}, \ldots, y_{i_{[n]}}). \tag{9}$$

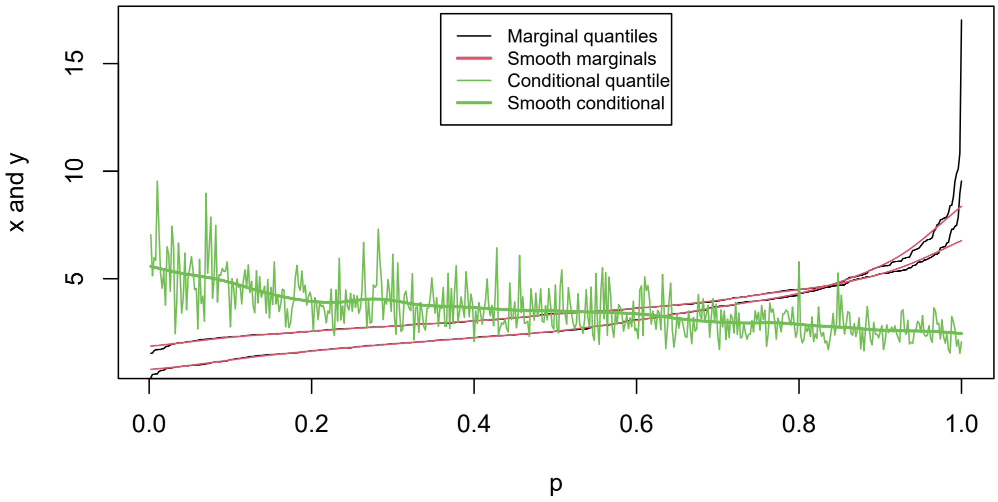
But this natural estimator provides a very shaky function of $p$, so that a generalisation of the non-parametric estimator of Yang (1985) proposed for usual quantile function becomes very useful in this case. Let us detail an example, using simulated data.

**Example 1** *We have generated 500 observations from a bivariate log-normal with a negative correlation, obtained using the following parameters:*

$$\mu = (1.0, 1.2), \quad \Sigma = \begin{pmatrix} 0.35 & -0.12 \\ -0.12 & 0.10 \end{pmatrix},$$

*leading to an empirical negative correlation of -0.55, a growth rate of 15% and Gini coefficients of respectively 0.31 and 0.18 for the two periods. This simulation could correspond to what has happened in China during the fifties with the great equalisation experience. Figure 1 presents the two marginal quantile functions, together with the conditional quantile function. The estimator of Yang (1985) has problems for extreme marginal quantiles as it is based on a symmetric kernel. But it manages to provide a clear interpretation of the conditional quantile function which is otherwise too shaky.*

The na-GIC (2) of Grimm (2007) consider the difference between two quantile functions and thus could justify the smoothing of these two quantile functions separately, or even just the smoothing of the conditional quantile function. The mobility profile (7) 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.




The data have been simulated from a bivariate lognormal distribution with 500 observations. The window size for smoothing has been determined by Silverman's rule.

Figure 1: Marginal and conditional quantiles estimated for a simulated bivariate lognormal distribution

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 (2011). The method is available in `R` with the function `loess`, implying the choice of a degree of smoothing. It has better properties at boundaries than the Nadaraya-Watson kernel regression. We illustrate the difference between (2) and (7) in the next example.

**Example 2** *Using the same simulated data as in the previous example, Figure 2 shows that the differences between (2) and (7) depends solely on what we smooth. For (2) we have smoothed only the conditional quantile function, so there are more details in the extreme quantiles. But otherwise, there are no significant differences between the two approaches.*

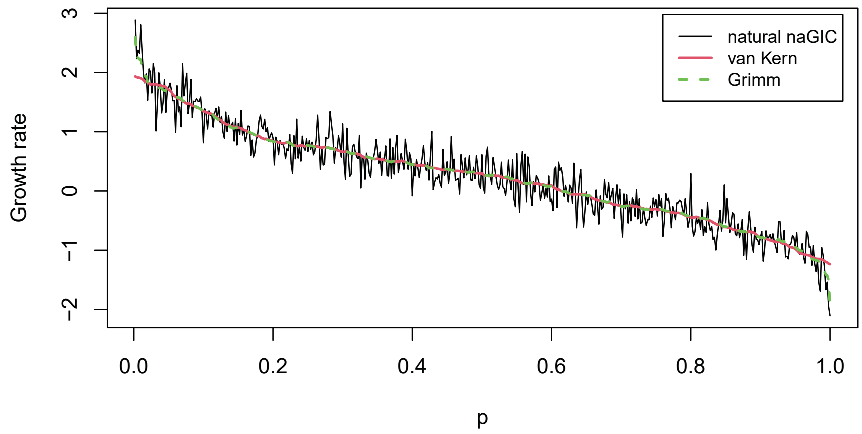



Figure 2: GIC and na-GIC with np smoothing

# 3 A Bayesian approach using Bernstein polynomials

We are looking for a Bayesian way of performing the single non-parametric regression implied by (7) which takes into account the fact that our explanatory variable is restricted to the range [0,1] and check if a Bayesian approach could make better justice to the extreme quantiles. Bayesian regression using Bernstein polynomials is a nice solution as it fulfill both requirements. The literature on Bernstein Bayesian regressions is not abundant and mainly devoted to cases where prior restrictions are imposed on the shape of the Bernstein function.

## 3.1 Bernstein polynomials

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

$$g(x,k) = \sum_{j=0}^{k} f(j/k) C_k^j x^j (1-x)^{k-j} = \sum_{j=0}^{k} f(j/k) b_k(x,j), \qquad (10)$$

where $C_k^j$ is the binomial coefficient and $b_k(x,j) = C_k^j x^j (1-x)^{k-j}$. If $x$ is outside the segment $[0,1]$, it can be transformed so as to lie in the required interval.[1] 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.

## 3.2 Smoothing using Bernstein polynomials

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

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

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

$$y_i = \sum_{j=0}^{k} \beta_j C_k^j x_i^j (1 - x_i)^{k-j} + \epsilon_i = \sum_{j=0}^{k} \beta_j b_k(x, j) + \epsilon_i. \qquad (11)$$

For a given basis $Z = [b_k(x, 0), \cdots, b_k(x, k)]$ with $k + 1$ columns and $n$ rows, the Bernstein coefficients $f(j/k)$ can be estimated by the following linear regression model without an intercept:

$$y_i = z_i \beta + \epsilon_i.$$

The degree $k$ of the polynomial is directly related to the degree of smoothing introducing a common bias-variance tradeoff. A higher $k$ will reduce the bias at the cost of increasing the number of parameters. The elements of the basis have the following form as displayed in Figure 3. They behave like a variable asymmetric kernel, with different shapes depending on the value of the explanatory variable $p$.

## 3.3 Properties

Bernstein polynomials have a certain number of advantages for nonparametric regression. Brown and Chen (1999) underline that the competing method of kernel regression has problems at the boundaries of the sample. Solving this problem would mean considering different kernels inside the same regression problem. This is automatically done when using Bernstein polynomials as can be guessed from Figure 3. This property is particularly interesting since we are generally concerned by the very poor and very rich in inequality studies. Moreover our explanatory variable is naturally in the segment [0,1], with no transformation needed.

[1] If $x$ is at values on the segment $[a, b]$, then $y = (x - a)/(b - a)$ is at value on [0,1].

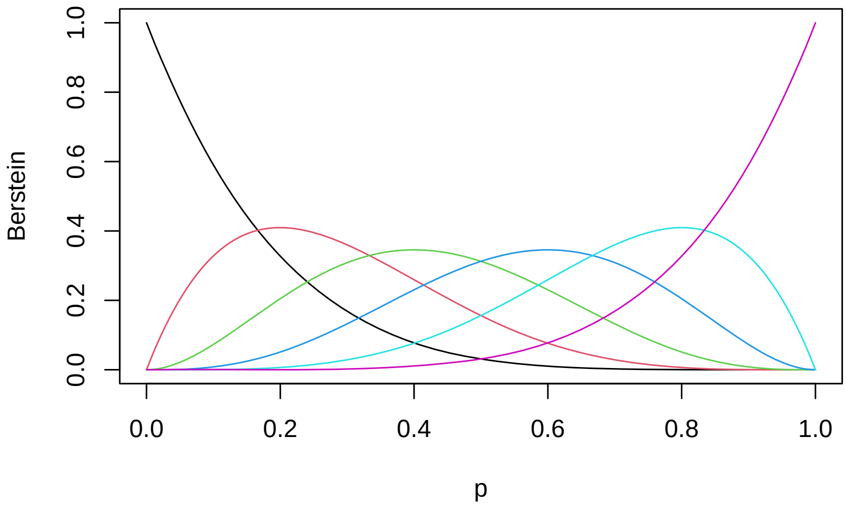



Figure 3: Bernstein polynomial of degree $k = 5$

One competing method would be splines. There is a huge literature about how to build splines which are interpolation methods (linear, cubic, etc,...) defined around knots. There is no precise method for choosing the knots. 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. With Bernstein polynomials, there is only one value for $k$ to be chosen, so the situation is in a way more simple and can rely on an usual information criteria.

Several authors (Curtis and Ghosh 2009, Wang and Ghosh 2012, Kim et al. 2019 to quote a few) have considered inference with constraints on the parameters so as to impose monotonicity or convexity. These conditions are derived by inspecting the first or second derivative and then imposing adequate sign constraints. The first order derivative of $g(x)$ with respect to $p$ is:

$$g'(x) = k \sum_{j=0}^{k-1} (\beta_{j+1} - \beta_j) C_{k-1}^{j} x^j (1-x)^{k-j-1}. \tag{12}$$

The sign of each of the $k$ elements of the sum depends only on the sign of $\beta_{j+1} - \beta_j$ so that the monotonicity is obtained for:

**Increasing** $\beta_j \leq \beta_{j+1}, \quad \forall j = 0, \cdots, k-1$

**Decreasing** $\beta_j \geq \beta_{j+1}, \quad \forall j = 0, \cdots, k-1$

An over-parameterised regression model can lead to ambiguity about the general shape of a na-GIC as seen in the next example.

**Example 3** *For smoothing the na-GIC obtained from the previous example, we used a Bernstein regression with $k = 5$, a value selected by application of a BIC. The graph of the first order derivative as reported in the right panel of Figure 4 indicates that for $k = 5$ the derivative*

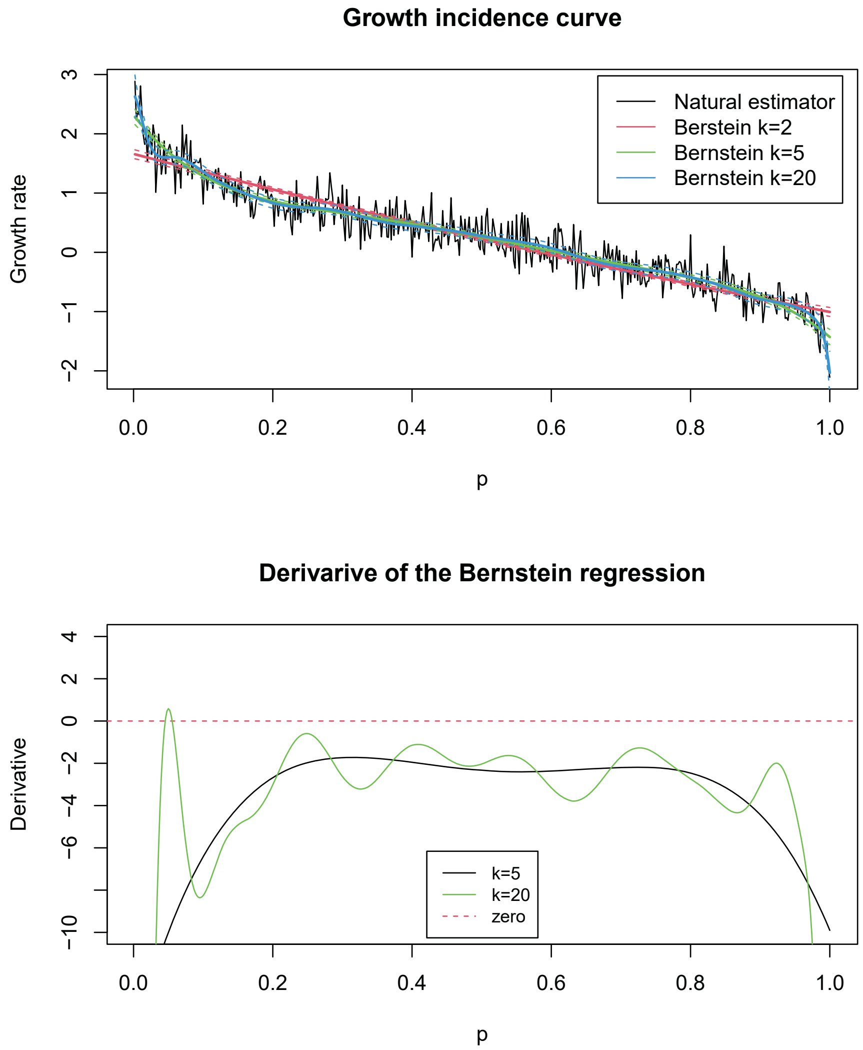



Figure 4: Impact of smoothing and under-smoothing on a Bernstein regression

*curve is always negative, implying a strict negative slope for the 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.*

## 3.4 Bayesian inference

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

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

In this notation, $Z(k)$ is the Bernstein basis and its size depends on the parameter $k$, 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.

Because they were concerned with specific problems, Curtis and Ghosh (2009) (a shape restricted Bernstein regression) and Kim et al. (2019) (sample selection) made use of particular prior distributions. Our case is much simpler. We have two parameters, $\beta$ and $\sigma^2$ while $k$ is part of the model specification. Because we are not interested in specifying a priori a particular restriction on the shape of the Bernstein polynomials, we can consider a usual non-informative prior with:

$$\varphi(\beta) \propto 1, \qquad \varphi(\sigma^2) \propto \frac{1}{\sigma^2}. \tag{13}$$

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

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

where the hyper-parameters are given by:

$$M_* = Z(k)'Z(k), \tag{14}$$
$$\beta_* = M_*^{-1}Z(k)'y, \tag{15}$$
$$s_* = y'y - y'Z(k)M_*^{-1}Z(k)'y, \tag{16}$$
$$\nu_* = n. \tag{17}$$

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 $\beta$, 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 $\beta_i$ leading to:

$$M[j,.] = \sum_{i=0}^{k} \beta_i^{(j)} C_k^i p^i (1-p)^{k-i}, \tag{18}$$

where the $np$ columns corresponds to the $np$ values of the grid. From this grid and this matrix, we can determine a posterior confidence interval for the na-GIC, just by determining empirically the row quantiles of $M$. 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:

$$Ms[j,.] = k \sum_{i=0}^{k-1} (\beta_{i+1}^{(j)} - \beta_i^{(j)}) C_{k-1}^i p^i (1-p)^{k-i-1}. \tag{19}$$

## 3.5 Model specification

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:

$$\begin{aligned} BIC(k) &= -2\log p(y|\hat{\theta}, k) + (k+1)\log(n), & (20)\\ AIC(k) &= -2\log p(y|\hat{\theta}, k) + 2(k+1), & (21)\\ DIC(k) &= -2\mathrm{E}_\theta[\log(p(y|\theta))] + p_D. & (22) \end{aligned}$$

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 $2(k+1)$.

Using an information criteria implies that the most parsimonious model is looked for. Hence, there is a tradeoff between efficiency and complexity of the Bernstein polynomial. The idea is to select $k$ so that local details will be smoothed out while not constraining too much the general shape, as we shall see in the next example.

**Example 4** *We continue using our previous simulated data. When looking for an optimal $k$, the first panel in Figure 5 indicates that $k = 5$ is the optimal value using a BIC and 500 simulations of the posterior density. With this specification, we draw $m = 5,000$ values for $\beta$. The*

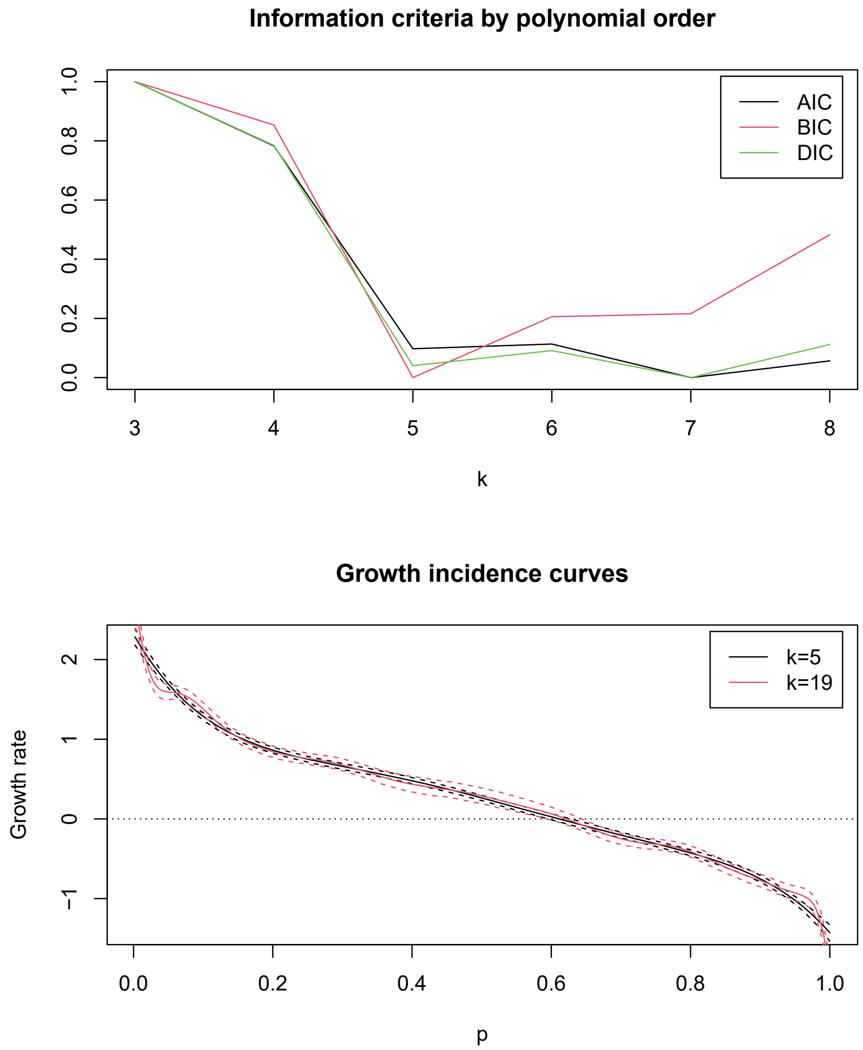



Figure 5: Bayesian inference for a Bernstein regression

*next panel displays the posterior graph for the na-GIC curve together with its 90% confidence bands for two cases, $k = 5$ and $k = 19$. From this example, it is clear that the gain in efficiency with $k = 19$ does not compensate the increasing complexity of the model.*

### 3.6 Testing for pro-poor growth

In a Bayesian framework, it is straightforward to test for pro-poor growth under anonymity, or in other words to test if the income of lower quantiles has grown more than that of the rest of the population. We simply need to have a convenient modelling of the two quantile functions indexed respectively by $\theta_1$ and $\theta_2$. Using $m$ draws from the posterior density of these parameters and build a $m \times np$ matrix of draws from the adequate transformation of these draws:

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

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 Fourrier-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:

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

where $\mathbb{1}(\cdot)$ is the indicator function. To test for pro-poor growth, we can also compute and the empirical probability that $g(p_i) > \gamma_t$ where $\gamma_t$ is the average growth rate over the period. 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 (18). The interpretation is of course different as we look at the trajectories of those who were initially poor, if their income has grown enough so that they managed to get out of poverty. Let us call the matrix (18) $M_{naGIC}$. The probability that $m(p) = \delta(x(p_x), y(p_x)) > 0$ is computed as:

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

We can also compute the probability that the slope of the of the na-GIC is negative for all $p_i$ using the matrix $Ms$ of draws (19) corresponding to the the first order derivative of the na-GIC curve:

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

**Example 5** *Using the same simulated data, Table 1, reports the empirical probability that the na-GIC curve is lower than the average growth rate $\gamma$; then the probability that it is positive for each decile; and finally the probability that the derivative is positive.*

Table 1: Posterior probabilities for quantiles

| $p$ | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
|---|---|---|---|---|---|---|---|---|---|
| na-GIC$>\gamma$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.01 | 0.00 | 0.00 | 0.00 |
| na-GIC$>0$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.94 | 0.00 | 0.00 | 0.00 |
| Derivative$>0$ | 0.00 | 0.03 | 0.01 | 0.01 | 0.02 | 0.00 | 0.01 | 0.01 | 0.07 |

# 4 Wage dynamics at MSU

Michigan State University (MSU) used to publish detailed pdf files documenting the wages paid to its academic staff. It is thus possible to construct a panel of individuals together with their grade, wage, type of contract and seniority for two years 2006 and 2012. After those dates, the information content of the published files diminished, names disappeared and 2012 is the last useful available date. The 2006 and 2012 samples have already been studied by Benzidia and Lubrano (2020) to explore academic wage formation in a static case. Benzidia and Lubrano (2020) were looking for a superstar effect, which means the presence of some top academics paid at a much higher wage than their followers. That type of effect was found in private firms for Chief Executive Officers that led a signification increase in wage inequality (Gabaix and Landier 2008). No super star was detected among MSU academics, except slightly when recruiting some Assistant Professors in some fields such as Medicine and Economics. On the contrary, there seemed to be a phenomenon of wage compression when climbing up the wage ladder, suggesting a glass ceiling effect. We have now new Bayesian tools, the na-GIC curve which allows to follow individual trajectories. This makes a great difference as we can study wage dynamics, pointing our attention also to gender and ethnic issues from this point of view.

## 4.1 A case study in university's recruitment policy

As in many American universities, MSU was tempted by an unbundling policy which intends to separate the traditional academic tasks (teaching, research and service) into distinct jobs by introducing new positions like *Educator*, *Instructor*, *Lecturer* or *Specialist* (Macfarlane 2011). These new jobs have a much lower mean wage: $46,941 against $91,820 for regular academics (Assistant, Associate, Full, Endowed Professors) in 2006, together with a lower dispersion. And these people are most of the time recruited on a fixed term contract. The other characteristics of the recruiting policy is the overall rise of fixed term contracts. From 2006 to 2012, the number of fixed term contracts has increased by 35% among regular academics at MSU, while the number of individuals in the tenure system remained stable. This increase concerned mainly Assistant professors. As a consequence, the unbundling policy can increase wage inequality and even foster polarisation.

One of the aims of this application is to investigate if this policy has managed to reallocate funds in order to be able to increase academic wages and in particular if top wages in academia have grown more than the average wage as in other sectors of the economy (Gabaix and Landier 2008). The simple GIC will help us to depict the evolution of the overall academic wage distribution and illustrating which of its parts have increased more rapidly. But the na-GIC allows us to follow individual trajectories as it explicitly permit possible re-ranking. With a na-GIC, we can depict which type of career an academic can expect. The second aim of this application is to investigate gender and ethnic issues and in particular what was the effect of the unbundling policy on the growth rate of female wages, and also if the pro-active anti-discrimination policies promoted by MSU were effective.

## 4.2 Wage dynamics and mobility

The estimation of a transition matrix between 2006 and 2012 describes the mobility of academics at MSU. Status transition can be staying in the same position, being promoted or quitting MSU. There were 2,271 academics in 2006 and 34% of them were females. Among those academics, 687 have left in 2012. The probability for a female academic to leave is 33%, but this rate falls down to 29% for males. Assistant Professors are those with the highest quitting rate, due to institutional reasons, while Associate, Full and Endowed Professors are by far mostly staying in the same position with an increasing probability, as indicated in Table 2. It is more difficult for females to be

Table 2: Mobility of academics between 2006 and 2012

| | Assist. Prof | Assoc. Prof | Full Prof | Endowed Prof | Quit | Othe |
|---|---|---|---|---|---|---|
| Assistant Prof | 0.23 (0.25) | 0.34 (0.32) | 0.01 (0.00) | 0.00 | 0.40 | 0.02 (0. |
| Associate Prof | 0.00 | 0.48 (0.49) | 0.23 (0.20) | 0.01 (0.00) | 0.21 (0.24) | 0.06 (0. |
| Full Prof | 0.00 | 0.00 | 0.60 (0.59) | 0.03 (0.02) | 0.27 (0.30) | 0.09 |
| Endowed Prof | 0.00 | 0.00 | 0.04 | 0.56 (0.45) | 0.27 (0.32) | 0.13 (0. |

Rows sum to one. The column Other corresponds to Emeritus or administrative tasks. Figures for females are indicated in parentheses when they are different from those of males.

promoted from Assistant to Associate and from Associate to Full professor. When having been promoted, they quit more often MSU. At the top of the ladder, they stay a shorter time Endowed, preferring either to quit or to take an administrative position, which is a secure way of getting a wage increase (Hamermesh et al. 1982). There is thus a clear gender issue for females at MSU.

Let us now estimate a Mincer equation explaining the log wage as a function of experience and other characteristics for the two groups, stayers and movers. The comparison between these two wage equations is quite illuminating as presented in Table 3. The movers have a yield of

Table 3: Wage equations in 2006 for academics staying at MSU or leaving before 2012

| | Stayers | | Movers | |
|---|---|---|---|---|
| | Estimate | t value | Estimate | t value |
| (Intercept) | 11.119 | 707.532 | 11.128 | 380.837 |
| Exp | 0.005 | 2.498 | 0.007 | 1.553 |
| $\text{Exp}^2$ | -0.000 | -1.338 | -0.000 | -2.041 |
| Associate Prof | 0.226 | 12.816 | 0.148 | 4.154 |
| Full Prof | 0.489 | 25.810 | 0.446 | 12.017 |
| Endowed Prof | 0.847 | 25.335 | 0.821 | 12.527 |
| Contract | 0.020 | 1.063 | -0.175 | -6.111 |
| Female | -0.064 | -4.712 | -0.028 | -1.125 |
| $R^2$ | 0.510 | | 0.459 | |
| $\sigma$ | 0.246 | | 0.306 | |
| $N$ | 1,584 | | 687 | |

experience which is decreasing after 12 years while it starts decreasing after 30 years for stayers. This makes a strong incentive to justify staying or leaving. The second difference is in term of type of contract. There are two types of contracts, the tenure system and the fixed term contract system where contracts can be renewed, but not necessarily. For the movers, the fact of holding a fixed term contract implies a wage loss of 18% when this does not play any role for the stayers. Compared to the Assistant Professors, Associate Professors have a gain of 15% for the movers, but 23% for the stayers, Full Professors have a gain of 45% for the movers, but 49% for the stayers. Finally, the Endowed Professors have a gain of 82% for the movers, but 85% for the stayers. These results suggest that those quitting MSU expect a better yield of experience and a larger probability of being promoted elsewhere allowing them to escape from fixed term contracts. The situation of the females is slightly different from the whole lot. Those who decided to stay experience a wage differential of 6%, when this penalty is not significant for the movers.

### 4.3 Bayesian na-GIC for academic stayers

We can observe only the wage of the stayers, so we have to limit our analysis of wage dynamics to those staying at MSU. When comparing the na-GIC and the GIC, we shall see that these two curves imply quite different wage dynamics. We have to choose first the degree $k$ for the Bernstein polynomial. For the na-GIC, we get $k = 1$ with BIC and DIC, which are not very plausible values and $k = 4$ with AIC. We used $m = 5,000$ draws. In Figure 6, we show the impact of choosing different values for $k$. In order to have fine results, we must

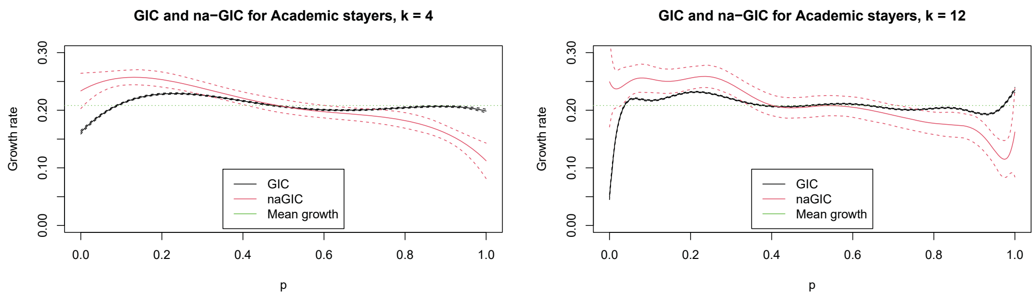



Figure 6: na-GIC versus GIC and the impact of under-smoothing for academic stayers

use a rather over-parameterised model with $k = 12$, taking advantage of the large number of observations ($n = 1,584$). Even with $k = 12$, the smoothed GIC has a very narrow 90% confidence interval. It is roughly horizontal, but says that very top wages have increased more than the average, while the 10% lower wages have increased much less than $\gamma$, the average wage growth. This result would be consistent with the hypothesis that MSU has tried to keep superstars by increasing top wages more than the average, but not with the story that it has tried to retain top Assistant Professors.

The na-GIC tells a totally different story. It is downward sloping. Wages below the median have increased more than $\gamma$ maintaining an attractive wage policy for Assistant professors. For wages above the median, they all increased less than average. So removing anonymity reveal a phenomenon of wage compression at the top of the scale. A potential explanation is that the academic market is less competitive for full and endowed professors who are more reluctant to move.

Table 4: Quantile probabilities for academic stayers with $k = 4$

| $p$ | 0.00 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GIC | | | | | |
| Pr(GIC> 0) | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Pr(GIC> $\gamma_t$) | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 |
| Pr(Deriv> 0) | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.10 | 0.00 |
| | | | | | | na-GIC | | | | | |
| Pr(na-GIC> 0) | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Pr(na-GIC> $\gamma_t$) | 0.91 | 1.00 | 1.00 | 1.00 | 0.97 | 0.41 | 0.05 | 0.01 | 0.00 | 0.00 | 0.00 |
| Pr(Deriv> 0) | 0.93 | 0.74 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.10 | 0.00 | 0.00 | 0.01 |

Table 5: Quantile probabilities for academic stayers with $k = 12$

| $p$ | 0.00 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GIC | | | | | |
| Pr(GIC> 0) | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Pr(GIC> $\gamma_t$) | 0.00 | 1.00 | 1.00 | 1.00 | 0.01 | 0.67 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Pr(Deriv> 0) | 1.00 | 0.17 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| | | | | | | na-GIC | | | | | |
| Pr(na-GIC> 0) | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Pr(na-GIC> $\gamma_t$) | 0.81 | 1.00 | 1.00 | 1.00 | 0.52 | 0.43 | 0.38 | 0.07 | 0.00 | 0.00 | 0.16 |
| Pr(Deriv> 0) | 0.35 | 0.36 | 0.72 | 0.02 | 0.10 | 0.65 | 0.27 | 0.21 | 0.35 | 0.13 | 0.84 |

Let us now confirm this judgement by examining the dominance properties for each quantile in Tables 4 and 5. First of all, we have the confirmation that the GIC has a varying slope while the na-GIC has a slope which is always negative with at most some ambiguity for the 10% quantile for $k = 4$. The na-GIC is always positive, which means that all wages have increased. However, the na-GIC becomes lower than the average growth rate right above the median without ambiguity. Taking into account mobility along the wage distribution can reveal very different patterns from a simple GIC.

# 5 Gender and ethnic issues at MSU

MSU displays an anti-discrimination policy on its web site where we can find the following announcement in 2022:

*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.*

What is the distance between cups and lips?

## 5.1 Gender issues

We have seen in Table 3 that in 2006 female academics were paid less by 6% compared to males academics. For those who decided to stay till 2012, what was their wage rise? Do we see any type of compensation in terms of a different wage increase for lower wages or is there an everlasting wage discrimination against females?

For consistency with the previous part, we keep k=4 for estimating GIC for males and females separately. We have 1,073 males and 511 females. Male and female academics follow the same tendency, but we

Table 6: Male and female academics

| $p$ | 0.00 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Academic males | | | | | |
| $naGIC > 0$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $naGIC > \gamma$ | 0.91 | 1.00 | 1.00 | 0.99 | 0.55 | 0.08 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | | | | | Academic females | | | | | |
| $naGIC > 0$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $naGIC > \gamma$ | 0.57 | 1.00 | 1.00 | 1.00 | 1.00 | 0.93 | 0.74 | 0.54 | 0.35 | 0.08 | 0.05 |
| | | | | | | Comparing academic males and females | | | | | |
| $\Pr(Male > Female)$ | 0.72 | 0.33 | 0.07 | 0.01 | 0.01 | 0.03 | 0.04 | 0.05 | 0.06 | 0.02 | 0.08 |
| | | | | | | Comparing para-academic males and females | | | | | |
| $\Pr(Male > Female)$ | 0.34 | 0.55 | 0.82 | 0.92 | 0.89 | 0.64 | 0.25 | 0.05 | 0.04 | 0.15 | 0.70 |

notice that the female curve in Figure 7 is quite systematically over the male curve, except for the minimum quantile. The overall probability that male wages increases more than female wages is only of 0.13. So there is a kind of compensation policy for females. They have a slightly lower wage at the beginning, but their prospect for a wage increase is greater than for males.

Let us now discover what is happening for para-academics. They are 522 in number, 236 males and 286 females, a totaly different proportion than for academics. The the wage equation reported in Table 7 indicates a penalty of 11% for females, so much more important than for female academics.

The overall probability that the increase of male wages is greater than that for females is now 0.49. So there is no general policy of compensation despite the fact that the gender gap is much greater than for academics. The two panels of Figure 7 illustrate gender differences for

Table 7: Wage equation for para-academics stayers in 2006

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 10.638 | 0.038 | 282.422 | 0.000 |
| Exp | 0.019 | 0.005 | 3.893 | 0.000 |
| $Exp^2$ | -0.000 | 0.000 | -0.845 | 0.399 |
| Instructor | 0.094 | 0.039 | 2.407 | 0.016 |
| Lecturer | 0.078 | 0.088 | 0.884 | 0.377 |
| Research Assoc | 0.322 | 0.072 | 4.475 | 0.000 |
| Specialist | 0.293 | 0.030 | 9.670 | 0.000 |
| Contract | -0.113 | 0.032 | -3.548 | 0.000 |
| Female | -0.107 | 0.023 | -4.602 | 0.000 |
| $R^2$ | 0.363 | | | |
| $\sigma$ | 0.260 | | | |
| $N$ | 522 | | | |

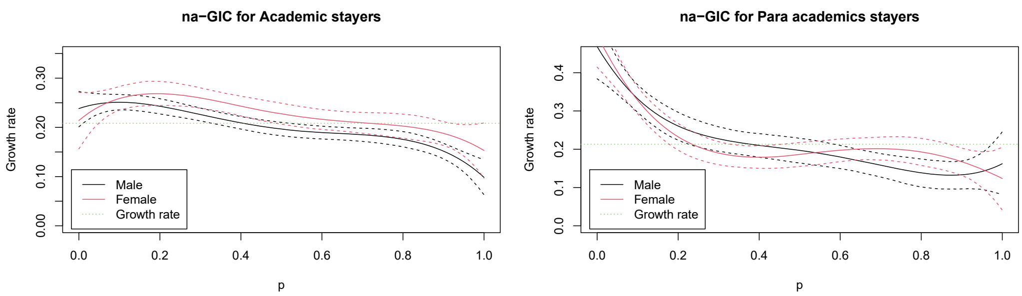



Figure 7: Gender issues for wage dynamics

academics and para-academics. In dynamics, there is a very selective gender policy at MSU in favour of academic females, even if Figure 7 shows that an important effort has been made for increasing low wages for both genders.

## 5.2 Ethinc issues

The US Bureau of Census distinguishes essentially four ethnic groups in the US: White, Black, Asian, Natives, the later including American Indians, Alaska and Hawaii Natives. To this list is made the distinction between Hispanic and non-Hispanic. Tzioumis (2018) gives the probability for a series of first names and last names to belong to a given ethnic group, exploiting exploiting data from the US Census. We used this methodology to classify names at MSU by ethnicity, using the `R` package `predictrace`. There is however a large proportion, 17% of cases that this methodology fails to predict. We decided to solve these 795 cases by hand and it resulted that 88% of the unresolved cases were likely to be classified as white according to the census classification originating generally from Eastern Europe or from the MENA region. We present in Table 8 wage characteristics of all the members of MSU, academics, para-academics and administration for 2006 and 2012. Hispanic are quite few in number and seem to be the most

Table 8: Wages and ethnicity for stayers at MSU

| Ethnicity | Stayers | Leavers | $q_{0.05}$ 2006 | Mean 2006 | $q_{0.95}$ 2006 | Mean 2012 |
|---|---|---|---|---|---|---|
| hispanic | 65 | 63 | $34,635 | $71,236 | $125,000 | $87,917 |
| asian | 221 | 232 | $36,000 | $83,211 | $154,500 | $104,708 |
| white | 2,505 | 1,640 | $37,038 | $90,982 | $171,649 | $111,268 |
| black | 86 | 42 | $32,060 | $90,105 | $190,530 | $112,918 |

disadvantaged category. Blacks, also quite few in number, have both the lowest $q_{0.05}$ and the highest $q_{0.95}$ in 2006. The exercise is now to compare the wage growth rate of Hispanic, Asian and Black employees to that of White ones. If we estimate a Mincer equation for explaining the 2006 log wages for those who will stay till 2012, we only find a significant negative wage different of 10% for Hispanic compared to white employees. The question is now to know if as a compensation the

Table 9: Wage equation for MSU 2006 stayers

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 11.065 | 0.048 | 232.931 | 0.000 |
| Exp | 0.013 | 0.002 | 7.283 | 0.000 |
| $\text{Exp}^2$ | -0.000 | 0.000 | -3.744 | 0.000 |
| Asian | 0.001 | 0.021 | 0.048 | 0.962 |
| Black | 0.038 | 0.032 | 1.195 | 0.232 |
| Hispanic | -0.100 | 0.036 | -2.770 | 0.006 |
| Female | -0.086 | 0.011 | -7.488 | 0.000 |
| $R^2$ | 0.624 | | | |
| $\sigma$ | 0.287 | | | |
| $N$ | 2,877 | | | |

Dummy variables were introduced for statuses, but their coefficient were not reported here. They explain roughly half of the variance.

wages of the Hispanic group has grown faster than that of the white group. Looking for a degree of the Bernstein polynomial for white (the most important group), we find $k = 6$ for the BIC and $k = 7$ for the AIC or DIC. For a fair comparison between groups, we adopted $k = 6$ for all the groups. When we compute the probabilities that the white

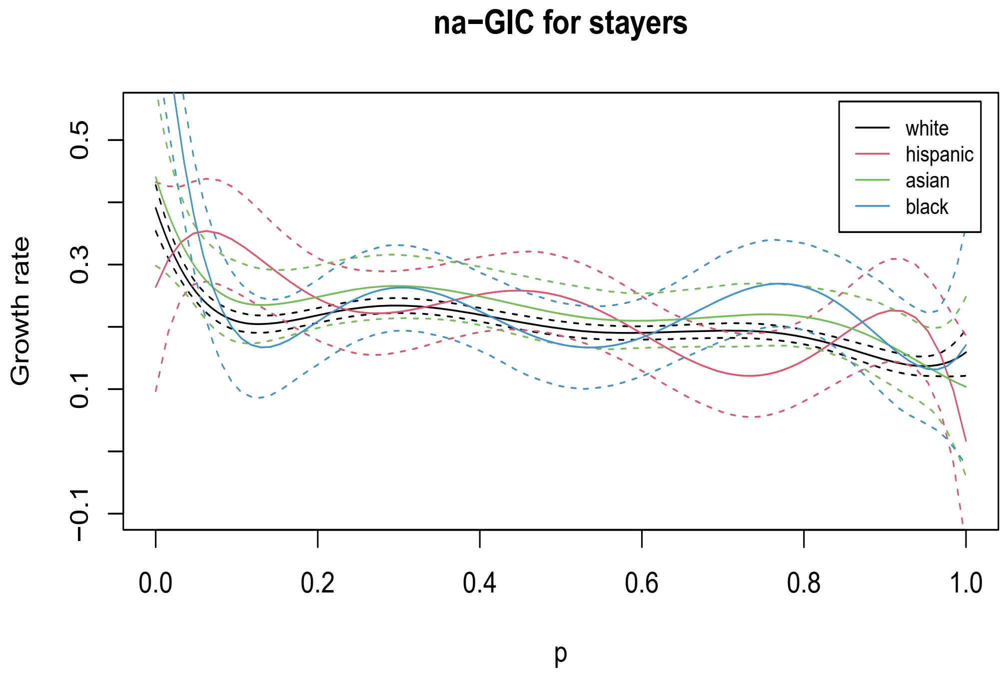



Figure 8: Ethnic issues for wage dynamics

group has larger wage rise than a given group, we find 0.48 for the Hispanic, 0.25 for Asian and 0.35 for Black. This means that Asian have 0.75 chances of a higher pay rise than White, Black 0.65 and Hispanic 0.52 when this group received on average a wage 10% lower than White. Let us now try to detail wage growth by quantile. Table

Table 10: Who gets more than average?

| Ethnicity | 0.00 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| White | 1.00 | 0.71 | 0.97 | 1.00 | 0.99 | 0.16 | 0.02 | 0.07 | 0.00 | 0.00 | 0.02 |
| Hispanic | 0.72 | 1.00 | 0.84 | 0.69 | 0.91 | 0.88 | 0.26 | 0.02 | 0.11 | 0.66 | 0.04 |
| Asian | 1.00 | 0.82 | 0.93 | 0.98 | 0.95 | 0.72 | 0.57 | 0.68 | 0.66 | 0.21 | 0.13 |
| Black | 1.00 | 0.46 | 0.54 | 0.92 | 0.71 | 0.21 | 0.29 | 0.86 | 0.92 | 0.26 | 0.38 |

10 tells us that above the median, the white group has 5% chances to get a pay rise greater than the measured $\gamma = 20\%$. For the Hispanic

group, this probability goes up to 33%, but for the Asian group and the Black group, this probability goes up respectively to 50% and 49%.

From all these numbers, combining the initial wage differentials and the results from estimating a na-GIC, we can conclude to some kind of affirmative action for the Asian and Black groups, while the Hispanic group seems to be left behind. As seen in Appendix A, the rate of exit for the Hispanic Assistant Professors is twice that of the other groups. And when they are Associate, their chance of being promoted full professor is a mere 5%, compared to 36% for Black, 32% for Asian and 23% for White.

# 6 Conclusion

In this paper, we have defined a na-GIC based on conditional ordering in a bivariate distribution, following one of the options depicted in Barnett (1976). This was coherent with the usual way of defining a na-GIC, taking the initial ranking of the first period $p_x$ as the reference. When analysing if growth was pro-poor, the measure of Grimm (2007) detailed in (3) describes what has happened to those who were initially poor. But, as underlined in Lo-Bue and Palmisano (2020), it ignores what happens to those who become poor at the second period. Taking the first period as the reference is justified on the ground of life trajectories, career 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 reduced ordering. However, by adopting such an ordering, we depart from the quantile transition matrix of Formby et al. (2004) and all the related literature on the meaning of income mobility.

We have adopted a Bernstein regression in order to model and smooth a conditional quantile function. This is a very convenient way for introducing Bayesian inference and it proved to be very convenient for computing probabilities and comparing trajectories. However, there is the pitfall of selecting the degree of the Bernstein polynomial. We have proposed to use three well-known information criteria (BIC, AIC and DIC) for model selection. Generally speaking BIC favours a much more parsimonious model whereas AIC and DIC usually agree for a larger parametrisation, which may be preferred when there are many observations. But we should also warn against the danger of over-fitting when using a Bernstein polynomial. It has a tendency to pick up unobserved heterogeneity, for instance if we had decided to use a Bernstein polynomial for modelling the non-linear effect of experience in our Mincer wage equations. But this is of course a different topic, which needs further investigation.

# References


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.

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, 139(3):318–355.

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. Technical report, NBER.

Benzidia, M. and Lubrano, M. (2020). A Bayesian look at American academic wages: From wage dispersion to wage compression. *Journal of Econonomic Inequality*, 18:213–238.

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.

Choi, T., Kim, H.-J., and Jo, S. (2016). Bayesian variable selection approach to a 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.

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.

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

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.

Gabaix, X. and Landier, A. (2008). Why has CEO Pay Increased So Much?*. *The Quarterly Journal of Economics*, 123(1):49–100.

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.

Grimm, M. (2007). Removing the anonymity axiom in assessing propoor 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.

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. (2011). Trends in individual income growth: Measurement methods and British evidence. Technical report, IZA.

Kakwani, N. (1980). On a class of poverty measures. *Econometrica*, 48(2):437–446.

Kim, H.-J., Roh, T., and Choi, T. (2019). Bayesian analysis of semiparametric Bernstein polynomial regression models for data with sample selection. *Statistics*, 53(5):1082–1111.

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.

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.

Stadtmuller, U. (1986). Asymptotic properties of nonparametric curve estimates. *Periodica Methematrica Hungarica*, 17(2):83–108.

Tenbusch, A. (1997). Nonparametric curve estimation with bernstein estimates. *Metrika*, 45:1–30.

Tzioumis, K. (2018). Demographic aspects of first names. *Scientific Data*, 5(1).

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.

Yang, S.-S. (1985). A smooth nonparametric estimator of a quantile function. *Journal of the American Statistical Association*, 80(392):1004–1011.

# Appendix

## A Transition matrices for academics by ethnicity

Table 11: Transition matrices for Academics by ethnicity

| | Assist | Assoc | Full | Endowed | Exit | Other |
|---|---|---|---|---|---|---|
| | | | Hispanic | | | |
| Assistant Prof | 0.219 | 0.156 | 0.000 | 0.000 | 0.625 | 0.000 |
| Associate Prof | 0.000 | 0.789 | 0.053 | 0.000 | 0.158 | 0.000 |
| Full Prof | 0.000 | 0.000 | 0.667 | 0.000 | 0.083 | 0.250 |
| Endowed Prof | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | | Black | | | |
| Assistant Prof | 0.261 | 0.348 | 0.000 | 0.000 | 0.348 | 0.043 |
| Associate Prof | 0.000 | 0.429 | 0.357 | 0.000 | 0.143 | 0.071 |
| Full Prof | 0.000 | 0.000 | 0.667 | 0.000 | 0.125 | 0.208 |
| Endowed Prof | 0.000 | 0.000 | 0.000 | 0.500 | 0.500 | 0.000 |
| | | | Asian | | | |
| Assistant Prof | 0.174 | 0.404 | 0.000 | 0.000 | 0.422 | 0.000 |
| Associate Prof | 0.000 | 0.464 | 0.321 | 0.018 | 0.161 | 0.036 |
| Full Prof | 0.000 | 0.000 | 0.679 | 0.054 | 0.179 | 0.089 |
| Endowed Prof | 0.000 | 0.000 | 0.000 | 0.500 | 0.250 | 0.250 |
| | | | White | | | |
| Assistant Prof | 0.236 | 0.337 | 0.010 | 0.000 | 0.390 | 0.028 |
| Associate Prof | 0.002 | 0.477 | 0.225 | 0.004 | 0.223 | 0.069 |
| Full Prof | 0.000 | 0.000 | 0.590 | 0.034 | 0.290 | 0.086 |
| Endowed Prof | 0.000 | 0.000 | 0.044 | 0.567 | 0.267 | 0.122 |

The column Other represents other cases such as emeritus, administrative or para-academic positions. The total number of academics staying or quitting is 1,916 for White, 229 for Asian, 63 for Hispanic and 63 for Black. So the transition matrices are to be interpreted with caution for Black and Hispanic.