

A Proporção de Pobres do Ceará Está Subestimada? Uma Aplicação da Teoria do Valor Extremo

Área Temática: Proteção Social e Geração de Oportunidades: Programas Sociais, Crime, Educação, Saúde, Pobreza;

Autores: Carlos Wagner Rios Pinto (CAEN/UFC); Ronaldo A Arraes (CAEN/UFC)

Emails: cwrios@ig.com.br; ronald@ufc.br

Telefone: 85 8607 2246 / 85 32814783

A Proporção de Pobres do Ceará Está Subestimada? Uma Aplicação da Teoria do Valor Extremo

Resumo

As pesquisas de estimação de distribuição da pobreza no Brasil têm se concentrado com o uso de ferramentas de inferência estatística contestáveis ou com o uso *ad hoc* de determinadas distribuições, ou ainda através de estudos de convergência. Este trabalho contribuiu com uma discussão de diferentes metodologias de inferência estatística não paramétrica, com o intuito de se estimar a evolução da densidade dos pobres. Inicialmente, através de suavização por núcleo estocástico (Kernel Density), com base nos dados coletados pela Pesquisa Nacional por Amostra de Domicílios – PNAD (2001, 2003, 2005 e 2007), conclui-se que a renda média familiar per capita vem aumentando no período da análise. Tomando o Ceará como um estado representativo para aplicação empírica, contrastou-se duas metodologias na estimação da proporção de pobres nesse período, sendo elas: A metodologia tradicional (uso de formulação Discreta Uniforme) e uma aplicação com base na Teoria do Valor Extremo (TVE). Para prover maior sustentação ao que se propõe, é feito um cenário contrapondo as taxas de pobreza extraídas do Censo-2000 com as obtidas através do IPEADATA e as avaliadas pela metodologia aqui empregada, ambas com base nos dados da PNAD-2001. Do contraste dos resultados pode-se concluir que há forte indicação de haver subestimação das taxas de pobreza ao se utilizar metodologias tradicionais.

Palavras-Chave: Núcleo Estocástico, Teoria do Valor Extremo, Distribuição de Pobres, Ceará.

Abstract

The research of estimating the distribution of poverty in Brazil have been concentrated with the use of tools of statistical inference or the inefficient use of some *ad hoc* distributions, or through studies of convergence. This work contributes to a discussion of different methods of non-parametric statistical inference in order to estimate the evolution of the density of poor individuals by smoothing a stochastic Kernel density, based on data from the National Household Survey - PNAD (2001, 2003, 2005 and 2007), and it concluded that the average family income per capita has been increasing during the analysis. By taking the state of Ceará as representative for empirical analysis, it was contrasted two methodologies for estimating the proportion of poor in the period, which are: The traditional method (use of Discrete Uniform formulation) and an application based on the theory of extreme values (TVE). To provide more support to what is proposed, it made a scene contrasting the poverty rates from the Census-2000 with those obtained by IPEADATA and evaluated by the methodology employed here, the latter two based on PNAD-2001 dataset. By contrasting the results it is concluded that there is strong evidence of underestimation of the poverty rates as the traditional methodology is applied.

Keywords: Kernel Density, Theory of Extreme Values, Poverty Distribution, Ceará.

1 Introdução

O objetivo da Inferência Estatística é produzir informações sobre dada característica da população a partir de informações colhidas de uma de suas partes. Se houvesse informações completas sobre uma população, como o seu comportamento, não haveria necessidade de se colher uma amostra, pois, toda a informação desejada seria obtida por meio de sua distribuição. Isso raramente acontece, pois, ou não se conhecem os parâmetros relativos à variável, a qual identifica a população, ou não se conhece a curva da distribuição, ou ainda, o que é mais comum, não se tem idéia de coisa alguma sobre ela. Daí a necessidade do uso de uma amostra que extraia o máximo possível de informações a seu respeito, que possibilite estimar, por técnicas de inferência estatística, a função densidade de probabilidade de uma variável aleatória x , denotada por $f(x)$, que descreve o real comportamento da distribuição dos dados. O conhecimento dessa função possibilitaria, por exemplo, calcular probabilidades de eventos associados a essa variável, ou, em estudos aplicados de distribuição de renda, conhecer-se a proporção de indivíduos situados em determinados intervalos de renda. Tais cálculos seriam procedidos da seguinte maneira:

$$p(a < x < b) = \int_a^b f(x) dx \quad \text{Para todo } a < b$$

Uma das aplicações dessa escolha metodológica na ciência econômica será utilizada neste trabalho cujo objetivo concentra-se em realizar técnicas de inferência estatística para se avaliar a evolução de renda das famílias pobres do Ceará no período compreendido entre 2001 e 2007. Para tanto, se utilizou amostras da PNAD nos anos de 2001, 2003, 2005 e 2007. Aplicaram-se à variável denotada pelo rendimento domiciliar *per capita* (RDPC) técnicas de inferência estatística paramétrica e não paramétrica, com o intuito de realizar estimativas dos seus parâmetros (Média, moda e desvio padrão), da curva da densidade bem como sua evolução durante o período supra mencionado e a estimativa da real proporção de pobres no Ceará no período adotado. As técnicas de inferência estatística não paramétrica aplicadas nessa análise tiveram o objetivo de se estimar a curva da densidade da renda dos pobres no Ceará bem como sua evolução no período adotado. Neste trabalho adotou-se a estimação de suavização por núcleo estocástico Sala-i-Martin (2002) se utilizou da técnica de suavização por Núcleo Estocástico, porém, aplicando-se à estimação da proporção de pobres. A contribuição desta formatação metodológica é a verificação da evolução da densidade da renda dos pobres do Ceará, estimando, via estatística não paramétrica que possui, de uma maneira geral, suposições bem menos rígidas.

Vale observar que no método de estimação de suavização por núcleo o que se obtém é um esboço do comportamento da real distribuição da renda nesta unidade da federação, com ênfase nos valores que compõem a cauda da distribuição, portanto, não se deve esperar como resultado que se consiga revelar uma fórmula da função densidade de probabilidade dessa variável, por exemplo, uma distribuição dentre as mais conhecidas. Estende-se ainda da análise, portanto, a verificação de se estar havendo convergência para um determinado nível de renda ou até mesmo uma confirmação de simetria ou uni modalidade.

O uso de técnicas de inferência estatística paramétrica neste trabalho possibilitou a estimação da proporção de pobres, tendo como referencial um estado representativo de uma região menos desenvolvida, no caso, o estado do Ceará. Muitos são os trabalhos a respeito de estimação da proporção de pobres com o uso *ad hoc* de determinadas distribuições, por exemplo, Barros e Mendonça (1997), Hoffman (2005), Foster et al (1984) e Sala-i-Martin (2002). Por outro lado, Arraes (2008), utilizou testes estatísticos não paramétricos para estimar a densidade de renda que melhor se ajustasse aos dados, tendo por base as unidades da federação. No trabalho ora apresentado a estimação das taxas de pobreza e indigência se

procedeu seguindo duas metodologias distintas: Na primeira, estimou-se a proporção de pobres pelo método mais tradicional que corresponde simplesmente ao quociente entre o número de pobres observados na amostra e a quantidade total de observações. O segundo método, já utilizados por Sen (1976) e posteriormente por Foster et al (1984), consiste em se calcular a integral, definida nos limites pelos quais se definem pobreza, da densidade da variável renda. A metodologia aqui empregada se diferencia das já apresentadas, no sentido de se estimar a densidade de renda via Teoria do Valor Extremo (TVE), que diferentemente de outros trabalhos publicados, por exemplo, em Manfred Gilli e Evis Kaellezi (2006), que se aplicou tal técnica para estimação do VAR (Valor em risco) de ativos financeiros. Aqui, utilizar-se-á TVE para se estimar a densidade da calda inferior da densidade da renda dos cearenses, resultando numa estimativa bastante consistente da proporção de pobres por motivos que serão apesados neste trabalho.

Inicia-se esse trabalho pela discussão sobre diferentes metodologias em diferentes artigos que tratam da estimação da densidade e proporção de pobres independente da região e o período da análise. Na seqüência são apresentadas as metodologias empregadas nas estimativas de densidade e proporção de pobres seguindo dos resultados e os possíveis contrastes encontrados na literatura. Na última seção são feitas as considerações finais.

2. Estimação Núcleo Estocástica da distribuição de renda dos pobres

Essa seção é dedicada à apresentação de metodologias de inferência estatística não paramétrica para estimação de densidade de uma variável aleatória. Nela, contrastar-se-ão metodologias como Histograma, por exemplo, com o método de estimação por suavização por núcleo estocástico. Para verificação empírica será tomado como referência o estado do Ceará.

Histograma

Para se realizar a estimação de uma densidade de probabilidade, é muito comum iniciá-la com uma investigação informal das propriedades dos dados observados. Um simples gráfico de dispersão dos dados pode mostrar evidências ou fortes indicações de simetria ou bi modalidade, por exemplo. O mais antigo e amplamente usado método de estimação de densidade trata-se do Histograma. Devido a sua simplicidade, ou seja, escolhe-se, a partir de experiências adquiridas das características das variáveis envolvidas no processo de estimação, ou até mesmo bom senso, o ponto de partida dos valores que irão compor o primeiro intervalo de dados na distribuição de freqüência, que por sua vez será a origem x_0 do gráfico, e adicionalmente a largura da caixa h (*bin width*) que corresponde à amplitude dos intervalos da distribuição de freqüência. A partir daí, formar-se-ão as outras caixas de mesma largura h , mas com altura correspondente a freqüência absoluta das observações que pertencem ao respectivo intervalo que geralmente são escolhidos serem fechados no lado esquerdo e abertos no lado direito. Outra maneira de construção do Histograma corresponde em fixar a quantidade de caixas e, conseqüentemente, tornando o parâmetro h em função disso. O estimador de densidade Histograma é então definido por:

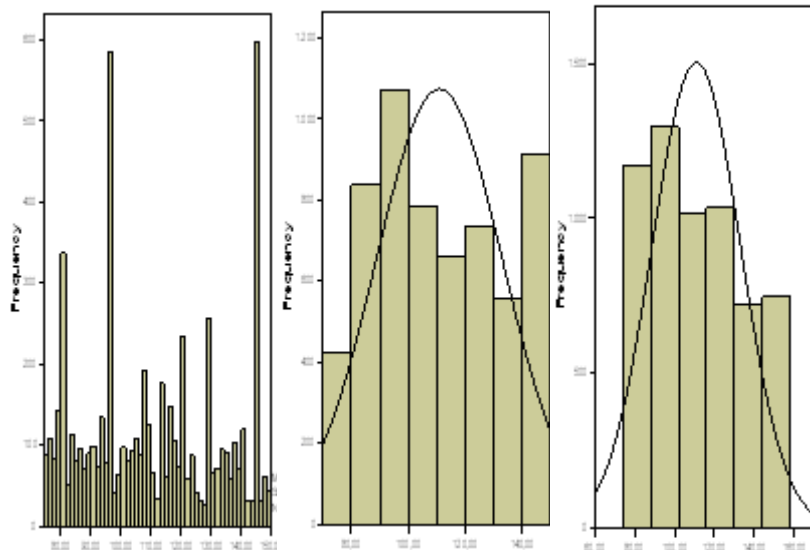
$$\hat{f}(x) = \frac{1}{nh} (\text{N}^\circ \text{ de } x \text{ no mesmo bin}) \quad (1)$$

Ou então uma generalização do Histograma, que permite que h varie. Portanto o estimador torna-se:

$$\hat{f}(x) = \frac{1}{n} \frac{(\text{N}^\circ \text{ de } x \text{ no mesmo bin})}{(\text{Largura do bin contendo } x)} \quad (2)$$

Nesta última versão apresentada, a largura do bin, como já comentado, pode variar de acordo com o que o pesquisador considerar conveniente. Por exemplo, em contas de energias é interessante classificar as classes de consumo como, primeiramente, uma faixa que determina o nível máximo de consumo que garante um desconto na conta daqueles usuários que não ultrapassem tal faixa. Também seria interessante determinar outra amplitude h do intervalo que determina níveis de consumo acima do que seria desejado às companhias de energia em certo período de crise energética, a fim de que houvesse punições, como uma multa, por exemplo, para aqueles consumidores que consumissem o nível de energia que pertencessem a tal intervalo. Além de um prévio procedimento que é determinar a largura dos *bins* no histograma, deve-se atribuir o ponto x_0 de partida, se não vejamos: Note que ao se determinar a origem, todas as larguras das “caixas”, inclusive a da primeira, serão determinadas a partir desta seguindo a seguinte lei $[x_0 + mh, x_0 + (m + 1)h]$ para todo inteiro m . Conclui-se que, ao se variar h ou x_0 ou ambos, ter-se-iam estimativas diferentes da função densidade de probabilidade. A figura 1 ilustra exemplos de estimativas da renda dos pobres do Ceará usando histograma como estimador.

Figura 1: Histogramas da renda dos pobres no Ceará em 2001



Na figura acima, o histograma (A) mostra a distribuição do rendimento domiciliar *per-capita* obtida através do *software* SPSS 12.0 onde não se determinou nenhum dos parâmetros do estimador da densidade, isto é, a largura h dos *bins*, a quantidade das caixas ou o ponto de partida; Neste caso o *software* atribui automaticamente, por métodos já programados (*default*), os valores dos respectivos parâmetros. Note que, com essa configuração, a disposição do arranjo das frequências não permite inferir sobre a real forma da densidade da variável investigada.

O histograma (B) foi desenhado com a largura das caixas pré-fixadas em R\$ 12.00 e a quantidade das caixas ficaram em função disso. O ponto de partida permaneceu constante. Com essa nova configuração visualiza-se uma tendência de bimodalidade na estimativa da densidade. No histograma (C), por sua vez, foram atribuídos novos valores para o Histograma (C): Como o ponto de partida que antes era de R\$ 70.00 e agora foi alterado para R\$ 60.00 e a quantidade das caixas ficou em 6 unidades, ficando, portanto, como função disso a largura h

das mesmas. Visualiza-se nesta nova configuração uma estimativa de curva completamente diferente das anteriores, apresentando agora uma assimetria à direita e une-modalidade.

Apesar de ser um dos estimadores de densidade mais usados por muitos pesquisadores pela sua simplicidade, mostrou-se um estimador de pouca eficiência, pois, constata-se uma grande variância nas estimativas apresentadas. Notou-se que basta que se varie pelo menos um dos três parâmetros que o define, para que a estimativa se torne completamente diferente, mesmo que se utilize a mesma série de dados, que neste caso se tratou da PNAD com corte nos valores entre R\$ 95,00 e R\$ 190,00 para o ano de 2001.

Estimador Naive

Define-se uma função densidade de probabilidade de uma variável aleatória x como sendo:

$$f(x) = \lim_{h \rightarrow 0} \frac{1}{2h} p(x-h < X < x+h) \quad (3)$$

Para qualquer dado h , pode-se estimar $p(x-h < X < x+h)$ pela proporção amostral que pertence ao intervalo $(x-h < X < x+h)$. Portanto um estimador natural (*Naive Estimator*) de $f(x)$ é dado por:

$$\hat{f}(x) = \frac{1}{2h} (\text{numero de } X_1, X_2, \dots, X_n \text{ que pertence ao intervalo } (x-h < X < x+h)) \quad (4)$$

Defina uma função peso dada por:

$$w(x) = \frac{1}{2} I_{|x| < 1} \quad (5)$$

Dada (5) pode-se reescrever (4) da seguinte maneira:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} w\left(\frac{x - X_i}{h}\right) \quad (6)$$

Segue então que a estimativa construída pelo estimador Naive consiste em pôr caixas de largura $2h$ e peso $(2nh)^{-1}$ em cada observação e então as somando. O estimador Naive leva vantagem sobre o Histograma, pois, ele pode ser visto como uma tentativa de construir um Histograma em que todos os pontos amostrais se situam no centro de cada smooth (intervalo ou caixa), livrando-se então da dependência do ponto de partida x_0 , logo suas estimativas são mais eficientes. A dependência da largura do *bin* (parâmetro *smooth*) continua e vale salientar que apesar do estimador *Naive* apresentar vantagens sobre o Histograma, por ser mais eficiente, o mesmo apresenta nas suas estimativas, uma forma bastante enrugada (áspera) da densidade, que pode muitas vezes também distorcer ou maquiagem o que seria a verdadeira forma da densidade além do mais, as derivadas em qualquer ponto pertencente ao range de variável que o define é zero.¹

¹ Para ver exemplo e mais detalhes consultar *Silverman* (1998).

O Estimador Kernel

O método de estimação de densidade proposto neste trabalho e que minimiza problemas encontrados pelo método do Histograma e do estimador *Naive*, é o método de suavização por núcleo estocástico.

Trata-se de uma generalização do estimador *Naive* que consiste basicamente por redefinir a função peso por:

$$\int_{-\infty}^{\infty} w(x) dx = 1 \quad (7)$$

Note que a função peso, definida por $w(x)$, atende a um quesito básico para definição de uma função densidade de probabilidade, ou seja, a integral definida nos Reais é igual a um. Note adicionalmente que a escolha dessa função deve seguir um comportamento razoável dos dados que pertencem a sua amostra. Geralmente deve-se escolher uma função simétrica como a distribuição Normal para a função peso $w(x)$.

O estimador *Kernel*, por analogia ao estimador *Naive*, é definido por:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} w\left(\frac{x - X_i}{h}\right) \quad (8)$$

Onde h é a largura do bin, também chamado de parâmetro alisador (smoothing); n é o tamanho da amostra; $w(x)$ é função peso já definida e x é a variável aleatória assumida em (7). A estimativa gerada pelo processo de Kernel é definida como sendo uma soma de *bumps* (protuberâncias) postas nas observações, isto é, a função $w(x)$ determina a forma dos *bumps* enquanto o parâmetro h determina a largura dos mesmos. Outra vantagem sobre as outras técnicas é que a estimativa se dá por uma soma de *bumps*, tornando suave a forma resultante e desde que $w(x)$ é escolhida ser uma função contínua, conclui-se que a estimativa será contínua e as derivadas de qualquer ordem e em qualquer ponto existem, resultando numa aproximação bastante razoável da verdadeira densidade.

Ressalte-se ainda que se tem uma dependência da escolha da largura dos *bumps*. No trabalho ora apresentado se seguiu a idéia apresentada e exemplificada em Silverman (1977, p.15), que a escolha de grandes larguras mascaram a real natureza da distribuição. Em contrapartida, uma pequena amplitude na largura torna a estimativa bastante enrugada, assemelhando-se à estimativa apresentada pelo estimador *Naive*. Qual então será o critério ótimo de escolha para h ? Optou-se por um método de tentativa, de se escolher h variando-a em seguidas vezes até que se encontre uma largura dada como conveniente. A base que se tomou para o critério de avaliar se determinado valor para h é considerado grande ou pequeno, foi utilizar-se da opção do default do *software E-Views* versão 6.0 que traz as duas possibilidades para largura do *bump* na caixa de diálogo *Silverman* ou *User specified*. A primeira opção se refere a um método que trata de um procedimento de escolha de h como sendo uma minimização do erro quadrático integral médio². Na seção 4.2 apresentam-se as estimativas de densidade, pelo método Kernel, da densidade dos pobres do Ceará nos anos de 2001 a 2007.

² Ver Silverman (1998)

3. Teoria do Valor Extremo

Essa seção, principal foco do trabalho, é dedicada à aplicação da Teoria do Valor Extremo à estimação da proporção de pobres no Ceará no período analisado.

Conhecer a magnitude da pobreza em determinada região é interesse de muitos economistas e governantes. Muitas tentativas de estimação da proporção de pobres já foram realizadas, contudo, muitas das quais se diferem substancialmente, seja por adotar diferentes linhas de pobreza ou por se utilizar de diferentes metodologias de estimação. Tais divergências podem gerar um nível considerável de desconfiança, por parte dos pesquisadores ou mesmo dos governantes que, porventura, necessitem se utilizar dessa informação, seja por não se saber ao certo qual é o verdadeiro valor (ou que mais se aproxima) ou o que é mais importante, se as estimativas subestimam ou superestimam a proporção de pobres.

Uma função densidade de probabilidade capaz de modelar o comportamento de uma população é definida por todos os possíveis valores da variável que a identifica. Tais valores, em geral, pertencem em sua maioria ao corpo da densidade. Para uma densidade Normal, por exemplo, o corpo se localiza no centro da distribuição, concentrando mais de 90% dos valores. As caudas concentram os valores mais raros de serem observados. No caso da distribuição de renda do Brasil, em particular do Ceará, espera-se observar uma curva assimétrica à direita, pois, trata-se de uma das unidades com maior desigualdade de renda da federação, portanto, concentra-se, em maior parte, por valores baixos de renda.

Independente dos métodos já apresentados por outros autores, já mencionados na introdução deste trabalho, a estimação da proporção de pobres de uma determinada região é realizada utilizando-se de uma função que modele a distribuição de renda, seja por considerar que essa variável segue um tipo de comportamento específico (uso *ad hoc* de certas distribuições), seja por realizar testes para escolha ótima da distribuição por métodos estatísticos. A pergunta a ser realizada é: Será que os valores que compõem a cauda da distribuição também têm o mesmo comportamento que os demais? Observe que a renda disponível por pobres, por mais assimétrica que seja a curva, pertence à cauda inferior da distribuição de renda.

A Teoria dos Valores Extremos foi aplicada neste trabalho por prover um sólido fundamento teórico necessário para construção de uma modelagem estatística das observações extremas de renda, pois, considera-se uma metodologia mais apropriada pela sua importante característica de concentrar-se no ajuste da distribuição apenas sobre os valores extremos inferiores da variável aleatória, diminuindo, portanto, a influência dos demais valores.

Ao se estimar a proporção de elementos que contêm certa característica de uma população, constrói-se, a partir de uma amostra, um modelo que aproxime o seu real comportamento, utilizando-se de resultados assintóticos, de modelos *ad hoc* específicos ou ainda por se utilizar de testes paramétricos e não paramétricos para escolha ótima do modelo. A partir daí, independentemente do método adotado, calcula-se a área sob a qual concentra os elementos que portam a característica de interesse. A população de que trata o presente trabalho é o rendimento familiar *per-capita* dos cearenses, mais especificamente do RFPC pelo qual se definiu pobreza. Note que a variável que representa a renda deve estar definida nos Reais não negativos, pois, admite-se que não há renda menor que zero. Os cearenses pobres detêm um rendimento mensal baixo, espera-se, portanto, que os rendimentos deles pertençam aos valores iniciais da variável que os descrevem. O uso da TVE se destaca e leva vantagem na utilização das outras metodologias, sobretudo, na capacidade de construir um modelo estatístico, de maneira teórica sólida, que aproxime o comportamento dos valores que compõem as caudas da distribuição, ou seja, dar maiores pesos aos valores que descrevem o RDPC dos cearenses pobres.

Ao se modelar mínimos ou máximos de uma variável aleatória pela TVE, na realidade, estar se usando resultados baseados no fundamento do Teorema do Limite Central, isto é, consiste na convergência da distribuição assintótica de uma série de mínimos ou máximos padronizados.

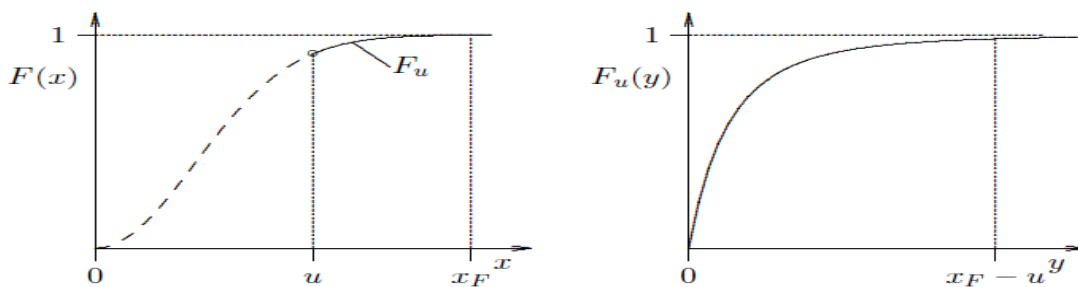
Há dois caminhos para se identificar valores extremos de uma variável aleatória: O primeiro considera máximos ou mínimos da amostra tomados em sucessivos períodos como dias ou semanas, ou em blocos, quando se tratar de dados “undated”. O segundo, conhecido pelo método do limite, é arbitrar um valor limite “u” no qual se considerará valor extremo da amostra o valor da observação que ultrapassar tal limite. Aos valores da amostra que foram coletados seguindo quaisquer dos dois caminhos, considerar-se-ão os valores extremos dessa variável.

O método de máximos em blocos é indicado principalmente em casos onde há presença de sazonalidade na série. Como a variável a ser tratada corresponde aos valores de renda pelos quais ultrapassam certo limite (linha de pobreza), optou-se pela estimação da densidade através da distribuição dos excedentes, realizada pelo método do limite.

3.1. Distribuição dos Excedentes

O método do limite, conhecido na literatura por POT (*peak over threshold method*), trata-se de um método de estimação da distribuição de uma variável aleatória que considera, na amostra, apenas os valores que ultrapassem um limite pré-determinado. Tal situação é ilustrada na figura 2 onde se considera uma função distribuição F (desconhecida) de uma variável aleatória X. Estamos interessados em estimar a distribuição F_u dos valores de X que ultrapassam o limite pré-determinado u.

Figura 2: Função distribuição F e função distribuição condicional F_u



A função distribuição F_u é chamada de função distribuição excesso condicional e é definida por:

$$F_u(Y) = P(X - u \leq y / X > u), \quad 0 \leq y \leq x_F - u;$$

Onde X é uma v.a., u é o limite pré-estabelecido, $Y = x - u$ é o excesso e $x_F \leq \infty$ é o limite superior de F. Perceba que os valores da variável aleatória X devem pertencer, em sua maioria, ao intervalo $(0; u]$, portanto não há grandes dificuldades na estimação de F; Porém, dependendo do valor de u, deve-se esperar que poucas observações pertençam ao intervalo complementar $[u; x_F)$, dificultando a estimação de F_u .

Pickands (1975), Balkema e de Haan (1974) mostraram que para uma grande classe de famílias de distribuição F a função distribuição excesso condicional F_u , para u grande, pode ser bem aproximada por:

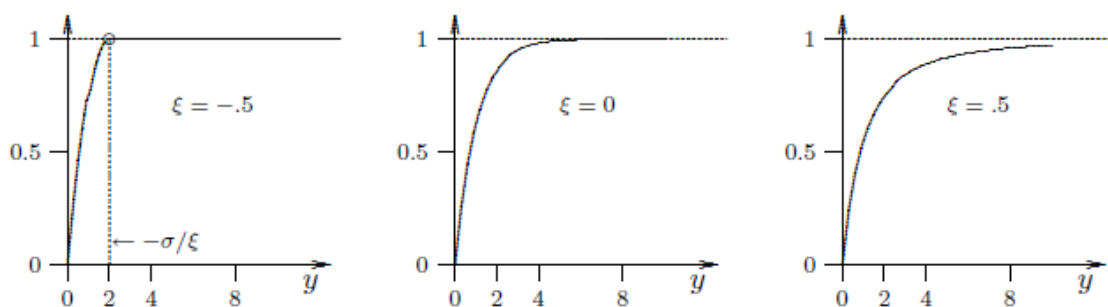
$$F_u(y) \approx G_{\xi, \sigma}(y), \quad u \rightarrow \infty,$$

$$\text{Onde } G_{\xi, \sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}} & \text{se } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}} & \text{se } \xi = 0 \end{cases}$$

Para $y \in [0, (x_F - u)]$ se $\xi \geq 0$ e $y \in \left[0, -\frac{\sigma}{\xi}\right]$ se $\xi < 0$. $G_{\xi, \sigma}(y)$ é então chamada de Distribuição Generalizada de Pareto (DGP). Provendo, portanto, à TVE um poderoso resultado sobre a função distribuição excesso condicional.

Se X é definido como sendo $x = u + y$, a GPD também pode ser escrita como uma função de X , isto é, $G_{\xi, \sigma}(y) = 1 - \left(1 + \xi \left(\frac{x - u}{\sigma}\right)\right)^{-\frac{1}{\xi}}$,

Figura 3: Forma da Distribuição Generalizada de Pareto $G_{\xi, \sigma}$ para $\sigma = 1$



O índice de cauda ξ dá uma indicação do peso da cauda; Para um ξ grande, tem-se uma cauda pesada.

4. Resultados

No trabalho ora apresentado foi considerado como unidade de observação o rendimento domiciliar *per capita* (RDPC), definido como o quociente entre o rendimento domiciliar e o número de pessoas residentes, e considerando apenas os domicílios particulares permanentes com declaração de rendimento. Note que a necessidade de renda não cresce linearmente com o tamanho da família, e que idosos, adultos e crianças precisam de volumes distintos de recursos pra viver [Barros, Carvalho, Franco e Mendonça (2007)]. Como não há informações sobre a importância dos ganhos de escala, e tampouco sobre as necessidades específicas de cada faixa etária, tal como na vasta literatura sobre distribuição de renda no Brasil, optou-se por essa modalidade de renda. Apesar de sua simplicidade, acredita-se que para que um indivíduo pertença a uma ou a outra classe social, deve-se levar em conta todos os familiares e também os agregados que moram no mesmo domicílio. A razão para tal

decorre do fato de o bem-estar de um indivíduo depender não apenas de seus próprios recursos, mas também (e talvez em grande medida) dos recursos da família a que ele pertence.

Ressalta-se, entretanto, que para medir a taxa e a evolução da pobreza no Ceará com base nos dados amostrais da distribuição da renda colhida pelas PNADs 2001, 2003, 2005 e 2007 (Anos adotados para medição), foi necessário se fixar um limite pelo qual se definiu pobreza (linha de pobreza). Neste trabalho, a abordagem para delinear a pobreza seguiu a mesma metodologia adotada pelo IPEA, ou seja, serão considerados pobres os componentes das famílias que detiveram rendimento médio de até R\$ 190,00, correspondendo à metade de salário mínimo, que em 2007 correspondia em R\$ 380,00, e considerados indigentes os componentes das famílias que detiveram rendimento médio de até um quarto de salário mínimo, o qual corresponde a R\$ 95,00. No Brasil, o salário mínimo (ou mesmo seus múltiplos) também tem sido recorrentemente usado como linha de pobreza (Pfefferman e Webb, 1983; Hoffman, 1984; Albuquerque, 1993; Paes de Barros, Henriques e Mendonça, 2000).

Sabe-se que existem várias metodologias a respeito da definição para linha de pobreza. Sabe-se, inclusive, que ao se considerar uma outra definição para pobres, por exemplo, levando em consideração outro nível de renda ou até mesmo outra metodologia que define através de outro fator que não a renda (número de calorias necessárias para sobrevivência (convertidas em valores monetários, por exemplo), poder-se-iam obter resultados divergentes dos alcançados no trabalho ora apresentado. Mas o objetivo aqui não é o de discutir qual a definição de pobreza a ser empregada, muito menos se é correto definir uma linha oficial de pobreza, mas tentar contribuir com resultados de estimativas de taxas de pobreza bem como a evolução da distribuição de renda dos cearenses pobres, através de metodologias até então não utilizadas, possibilitando, inclusive, uma contribuição na formulação de políticas adequadas de combate à pobreza, pois, acredita-se num maior vigor teórico e adequação da metodologia aqui empregada para extração dos resultados comparados aos outros trabalhos da literatura.

Os dados provêm da PNAD, que é realizada anualmente pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e constitui a principal fonte de informação sobre concentração de renda no país.

O indicador econômico analisado (RDPC) foi obtido através de uma transformação a partir das variáveis:

V5030 = Código da unidade da federação (Anos de 2001, 2003, 2005 e 2007).

V4722 = Valor do rendimento mensal familiar incluindo os agregados;

V4724 = Número de componentes da família inclusive os agregados;

Portanto a variável Rendimento familiar *per-cápita* dos pobres do estado do Ceará, definida por Rendpobre, foi criada a partir dos seguintes procedimentos:

Filtraram-se os valores da variável V4722 pareados aos da variável V5030 que apresentavam valor 23 (código da UF correspondente ao Ceará); Isso retornou somente valores com respeito ao estado do Ceará;

O próximo passo foi dividir os valores filtrados da V4722 pela V4724, assim teremos os valores dos rendimentos mensais médios familiares somente do estado do Ceará em cada ano de interesse;

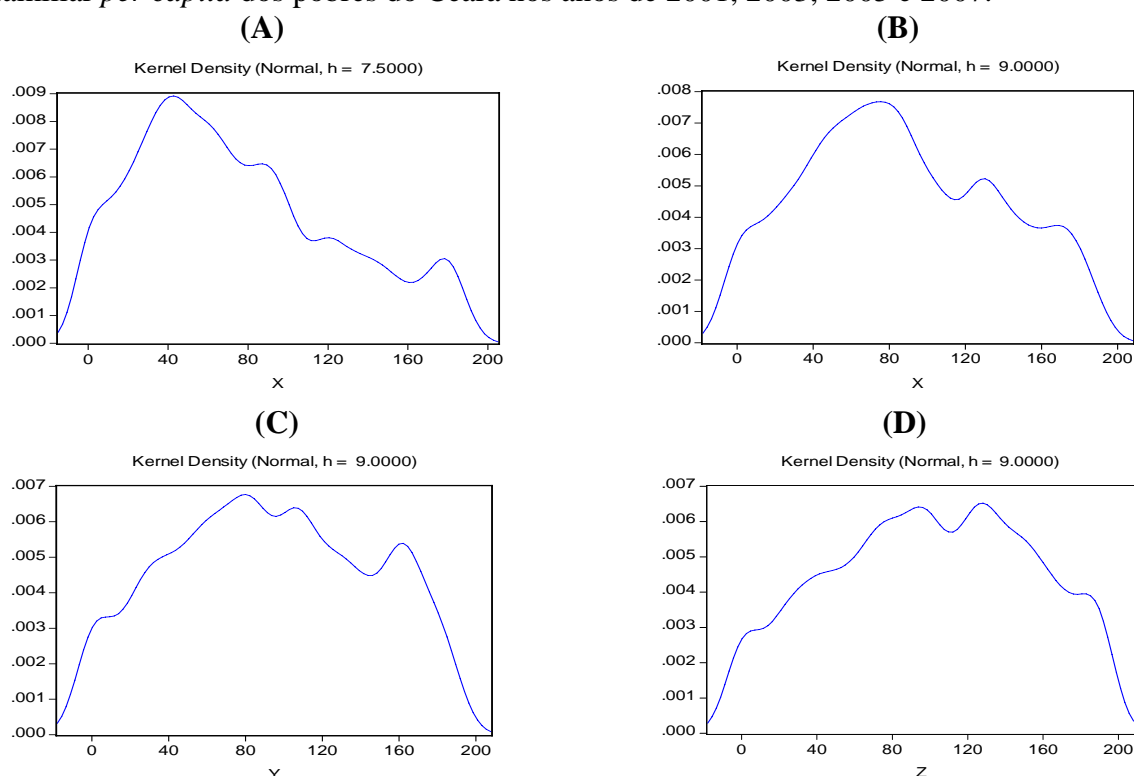
Posteriormente foi realizado um corte nesses dados no sentido de considerar apenas os níveis de renda que pertençam à faixa de zero à metade de um salário mínimo que, em 2007, esse intervalo correspondia a R\$0.00 a R\$190.00.(faixa adotada pelo IPEA que determina pobreza), e de zero a um quarto de salário mínimo, isto é, de R\$0,00 a R\$95,00 (faixa que determina a indigência) de acordo com a metodologia adotada pelo IPEA.

Finalmente, pra se comparar os rendimentos nos diferentes anos adotados neste trabalho, eles serão expressos em reais de Setembro de 2007, mês de realização da PNAD,

utilizando o INPC como deflator. Portanto a variável *Rendpobre* definida por rendimento domiciliar *per-capita* do estado do Ceará é dada pelo seguinte processo:

4.1. Evolução da densidade do rendimento familiar per-cápita dos pobres referente aos anos de 2001, 2003, 2005 e 2007.

Figura 4: Estimativas de Suavização por Núcleo das densidades do rendimento médio familiar *per capita* dos pobres do Ceará nos anos de 2001, 2003, 2005 e 2007.



A figura 4 refere-se às estimativas das densidades através de suavização por núcleo estocástico da renda domiciliar *per-capita* (considerou-se apenas os rendimentos médios mensais estejam entre zero à metade de um salário mínimo referente ao ano de 2007); O procedimento foi realizado com o uso do software *E-Views* 6.0 adotando como função peso a densidade Normal padronizada. A largura dos “*bumps*” foi de R\$ 7,50 para o ano de 2001 e R\$9,00 para os demais, determinados seguindo o procedimento descrito na seção anterior.

A respeito da estimativa da densidade para o ano de 2001, apresentada no gráfico (A), infere-se o seguinte: Aparentemente, a curva é uni modal e apresenta assimetria à direita. Sendo o corpo da distribuição composto por famílias que dispõem de R\$ 0,00 a um pouco mais de R\$ 150,00 médios mensais. Há uma discreta formação de um segundo grupo, do qual é composto pelas famílias cuja renda média mensal pertence ao intervalo de R\$ 160,00 a R\$ 190,00 aproximadamente. Ademais, há uma grande concentração de indivíduos no intervalo de renda de R\$10,00 a um pouco menos de R\$80, 00, revelando, portanto, que em 2001 a grande maioria dos cearenses pobres encontrava-se em situação de indigência.

Com relação à estimativa referente a 2003 (gráfico (B)), percebe-se, um achatamento (maior dispersão) e um considerável deslocamento à direita do cume do corpo da densidade com relação à situação encontrada no ano de 2001. Note que, em 2001, o cume se localizava em torno dos R\$ 40,00, e em 2003, esse se localiza próximo aos R\$ 80,00. Encontra-se, também na estimativa de densidade para os pobres do Ceará em 2003, uma maior densidade nos valores que compreendem o intervalo de R\$ 110,00 a R\$ 200,00, com relação ao ano de

2001. Fatos que se levam a concluir que houve um aumento na renda *per-capita* dos pobres em 2003 com relação à 2001.

O comportamento da curva da densidade no ano de 2005 (gráfico (C)) se distribuiu mais uniformemente que em 2003, pois, a assimetria à direita não se apresenta nesta situação. Isso mostra uma queda na desigualdade de renda entre os pobres, isto é, a proporção cearense em situação de indigência (com rendimento domiciliar *per-capita* inferior à R\$ 95,00) diminuiu, aumentando, portanto, a proporção dos indivíduos que ganham acima de R\$ 95,00 e abaixo de R\$ 190,00. Conclui-se que, em média, o rendimento *per-capita* dos pobres em 2005 subiu com relação aos anos de 2001 e 2003. Barreto, *et al* (2009) mostrou que o rendimento médio da distribuição da renda domiciliar *per capita* no Brasil aumentou de 15,2% no segundo quinto da população brasileira mais pobre (Os que dispunham de RMPC de R\$ 128,00 em 2001 e R\$ 147,50 em 2005). Observou-se também um aumento de 26,6% no RDPC do quinto mais pobre do Brasil de 2005 em relação a 2001 e de 11,4% no terceiro quinto. Barreto, *et al* (2007) verificou, ainda, o número e a proporção de pobres ($RPDC \leq R\$100,00$) diminuíram de 2001 a 2004. Fatos que corroboram o deslocamento para direita da densidade da renda dos pobres mesmo em se tratando da distribuição de renda dos pobres apenas no estado do Ceará.

Note que, na estimativa para o ano de 2007 (gráfico (D)), há a presença de dois cumes (bimodalidade), sendo o primeiro formado pelos indivíduos com menores rendimentos ($RDPC \leq R\$110,00$ aproximadamente). O segundo é formado por indivíduos que se encontram acima da faixa que determina a indigência ($R\$110,00 \leq RDPC \leq R\$190,00$). Ao se confrontar a estimativa de 2007 com a de 2005 percebe-se, um deslocamento à direita de toda a curva. Observa-se, inclusive, a assimetria à esquerda nesta estimativa.

A forma da densidade referente a 2007 se diferencia substancialmente com relação às dos outros anos considerados na análise. Os fatos constatados sobre esta curva reforçam, por mais forte razão, o aumento na média do rendimento mensal das famílias cearenses pobres em 2007. Barreto, *et al* (2009) foi realizada uma análise da renda média familiar *per capita* dos pobres no período de 1995 a 2007 mostrando a evolução desse indicador econômico, tendo como base de dados as PNADs referentes a cada ano pertencente ao período analisado, portanto, a mesma amostra utilizada neste trabalho. Foi mostrada uma expressiva expansão da renda dos pobres no período de 2002 a 2006. Em 2007 essa situação continuou, representando, ainda, uma ligeira melhora no âmbito nacional. Ademais, Soares, *et al* (2006) mostrou que os programas brasileiros de transferência direta de renda à população de baixa renda foram bastante importantes no aumento do rendimento familiar e na redução da desigualdade social no Brasil em 2004. Note que esses fatos corroboram com as deduções realizadas a cerca da evolução da renda dos pobres do Ceará no período analisado, tendo como ferramenta estimativa das densidades dessa unidade experimental através de suavização por núcleo estocástico.

4.2. Estimação de taxas de pobreza do Ceará

4.2.1 Método tradicional

Define-se X a variável aleatória tal que:

$$X = \begin{cases} 1, & \text{se no domicílio: Rendmédiam} \in] 0.00 ; 190.00 [\\ 0 & \text{caso contrário.} \end{cases}$$

Considera-se “sucesso” ($x=1$) o rendimento médio familiar mensal pertencente ao intervalo acima; Logo, o “fracasso”, é qualquer outro valor de rendimento. Portanto X assim definido segue *Bernoulli*(p) e sua fdp é dada pela fórmula:

$$f(x) = p^x(1-p)^{1-x}; X \in \{0;1\} \text{ e } p \text{ é a probabilidade de sucesso.}$$

Define-se $Y = \sum_{i=1}^n X_i$, a quantidade de sucessos em uma amostra de tamanho “ n ”, ou seja, a quantidade de cearenses considerados pobres (pela metodologia aplicada pelo IPEA, isto é, as que detêm rendimento médio mensal entre R\$ 0,00 e R\$ 190,00) pertencentes à amostra. Como a amostra é extraída de forma aleatória pela PNAD, isto é, x_1, x_2, \dots, x_n é uma amostra aleatória independente, é possível mostrar que Y segue Binomial($n; p$) com $E(Y) = np$ e $Var(Y) = np(1-p)$; Onde n é o tamanho da amostra; A fdp de Y é dada pela fórmula:

$f(y) = C_{p;n} p^y(1-p)^{n-y}; y \in \{0,1, \dots, n\}$; Onde $C_{p;n}$ é a combinação de n elementos tomados de ‘ p a p ’ maneiras.

Define-se por $\hat{P} = \frac{Y}{n}$ a proporção amostral de pobres do Ceará. Note ainda que:

$\hat{P} = \frac{\sum_{i=1}^n X_i}{n}$ Corresponde à \bar{x} , a média amostral de pobres; Pelo Teorema Central do Limite, para amostras de tamanho significativo, $\hat{P} = \bar{x}$ segue aproximadamente *Normal* $\left(\mu = p; \sigma^2 = \frac{p(1-p)}{n}\right)$. Logo $Z = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}}$ segue *Normal* $(\mu = 0; \sigma^2 = 1)$

Tabela 1: Quantidade e proporção de pobres no Ceará (Método Tradicional)

Anos	Casos					
	Rendmédiam $\in] 0,00; 190,00 [$		Rendmédiam $\notin] 0,00 ; 190,00 [$		Todos os casos	
	N	%	N	%	N	%
2001	13.687	57,89	9.958	42,11	23.645	100,0
2003	14.732	61,25	9.319	38,75	24.051	100,0
2005	14.303	57,12	10.697	42,88	25.000	100,0
2007	13.021	51,95	12.044	49,6	25.066	100,0

Fonte: Elaboração própria com base nos dados das PNADs 2001, 2003, 2005 e 2007

Tabela 2: Quantidade e taxa de indigentes no Ceará (Método Tradicional)

Anos	Casos					
	Rendmédiam $\in] 0,00; 95,00 [$		Rendmédiam $\notin] 0,00; 95,00 [$		Todos os casos	
	N	%	N	%	N	%
2001	7.562	32,0	16.083	68,0	23.645	100,0
2003	8.118	33,8	15.933	66,2	24.051	100,0
2005	7.466	29,9	17.534	70,1	25.000	100,0
2007	5.918	23,6	19.148	76,4	25.066	100,0

Fonte: Elaboração própria com base nos dados das PNADs 2001, 2003, 2005 e 2007

A tabela 1 apresenta, considerando a linha de pobreza adotada por este trabalho, as estimativas das taxas de pobres, que são de 57,89% em 2001, crescendo 3,36 pontos percentuais em 2003 (61,25%), decrescendo 4,13 pontos percentuais em 2005 (57,12%), e

manteve-se decrescente até atingir a proporção de 50,4% da população em 2007. Na tabela 2 apresentou-se a estimativa de indigentes, considerando-se como tal, aquele indivíduo que tem rendimentos inferiores a um quarto de salário mínimo, que em 2007 representava R\$ 95,00.

Como visto nesta seção, trata-se de estimativas pontuais de P dadas pelos quocientes entre a quantidade de amostras que pertencem à R\$0,00 a R\$95,00 e a quantidade total de observações que em 2007, por exemplo, corresponde a $\hat{P} = \frac{5.918}{25.066} = 0,236$, ou seja, estima-se, pelo método tradicional, que havia 23,6% de indigentes no Ceará no ano de 2007.

4.2.2 Método TVE

O procedimento de estimação deste método juntamente com simulações pré-liminares, por exemplo, da forma da cauda da curva (parâmetro “forma” da GPD), limite “u” utilizado, entre outros foram realizados com o uso de funções pertencentes ao pacote EVIM do *software* MATLAB 6.5.

A metodologia utilizada neste artigo parte da idéia proposta inicialmente por Sen (1976) e Foster et al (1984), também implementada por de Sala-i-Martin (2002a, 2002b), que calcula o número de pobres e a taxa de pobreza a partir da integral da função densidade de probabilidade gerada pela distribuição de renda das unidades observadas. Isto é feito para se obter a área sob a função densidade à esquerda da renda que define a linha de pobreza, ou seja, o valor da função densidade acumulada até o nível de renda que define tal linha.

$$R_{\alpha}(y_j; p) = \int_0^p [(p - y_j)/p]^{\alpha} f(y_j) dy_j \quad (9)$$

Onde; p é a linha de pobreza, y_j é a renda até p , $f(y_j)$ é a função densidade de probabilidade da renda, e α é o parâmetro que expressa aversão à pobreza. Para o cálculo da estimativa da proporção de pobres atribuir-se-á $\alpha = 0$ tal como realizado por Sala-i-Martin (2002) e Quah (2003).

Da expressão acima se pode derivar diversas medidas de pobreza, sendo as mais utilizadas: O índice de proporção de pobres (p^0), o hiato médio de pobreza (p^1) e o hiato quadrático de pobreza (p^2), para os valores de alfa = 0, 1 e 2, respectivamente. No primeiro caso, alfa igual à zero, tem-se a medida de incidência da pobreza que é simplesmente o percentual de pobres numa determinada economia. Quando se faz alfa igual a um, tem-se a medida de insuficiência média de renda. Quanto maior esta medida, menor é a renda média dos pobres em relação à linha de pobreza. Para alfa igual a dois, tem-se a medida de desigualdade entre os indivíduos que vivem na condição de pobreza.

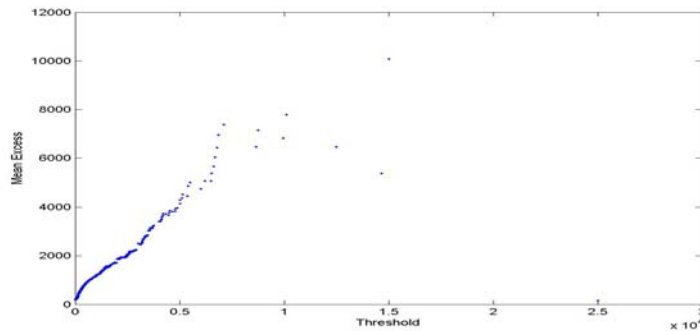
O hiato médio ou p^1 constitui um indicador mais interessante que o p^0 por diferenciar o muito pobre do pouco pobre. A vantagem do p^0 é obviamente a sua simplicidade. Por outro lado, p^1 nos dá diretamente o custo de um programa mais eficiente de combate à pobreza que pode ser implementado. A medida de pobreza p^1 confere maior peso aos mais pobres, mas o impacto de uma dada transferência de renda sobre o índice independe do nível de renda daqueles que recebem a transferência. A medida p^2 resolve este problema atribuindo maior peso aos mais pobres, pois, trata-se de uma de desvios ao quadrado em torno da linha de pobreza. Resumindo, à medida que subimos de p^2 , aumentam-se os pesos dos indivíduos mais pobres.

4.2.2.1 Resultados via TVE e comparação Amostral versus Populacional

Define-se a função excesso média amostral (MEF), por:
$$e_n(u) = \frac{\sum_{i=1}^n (x_i - u)}{\sum_{i=1}^n I_{\{x_i > u\}}(x)}$$

Onde u é o valor limite definido no método *peack over threshold* (POT). A MEF corresponde ao quociente entre a soma dos desvios tomados em relação a u , e o total de observações extremas da amostra. A estatística $e_n(u)$ retorna a estimativa da função excesso média para a parcela da população pertencente à uma de suas caudas. Se a MEF empírica é uma linha reta positivamente inclinada, isso é uma indicação que os dados seguem uma GPD com uma forma positiva do parâmetro *shape*.

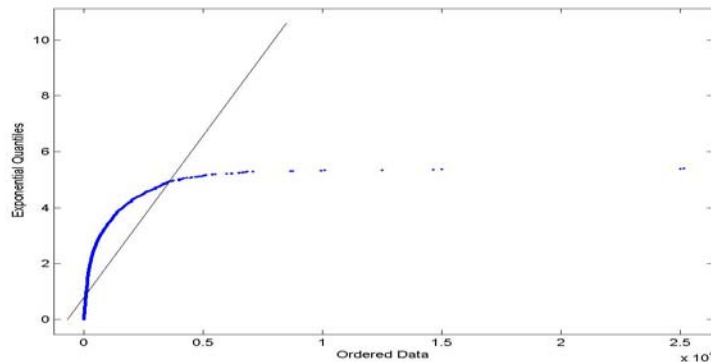
Figura 5: MEF da PNAD 2007



A figura 5 ilustra a MEF da PNAD 2007. Note que há, claramente, uma forte tendência positiva na estimativa da curva.

Em estatística, um QQ-plot (quantil-quantil plot) é uma conveniente ferramenta visual para analisar se uma amostra provém de uma distribuição teórica específica. Especificamente, os quantis de uma distribuição empírica são plotados contra os quantis de uma distribuição hipotética. Se a amostra provém da distribuição hipotética, o QQ-plot é linear. Na teoria dos valores extremos e aplicações, o QQ-plot é normalmente plotado contra a distribuição exponencial (isto é, uma distribuição com uma cauda de tamanho médio) para medir o peso da cauda da distribuição dos dados da amostra. Se os dados são de uma distribuição exponencial, os pontos no gráfico se encontram ao longo de uma reta linha. Se existe uma presença de concavidade, isto pode indicar uma distribuição com cauda pesada, ou seja, o parâmetro *shape* $\xi > 0$. Por outro lado, a presença de convexidade é uma indicação de cauda leve na distribuição.

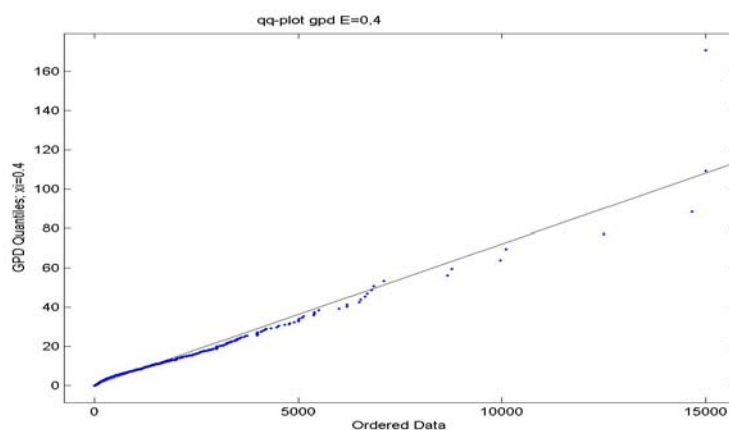
Figura 6: QQ-plot da PNAD 2007



A figura 6 apresenta o QQ-plot da Pnad 2007 contra uma distribuição exponencial. Repare que a presença de concavidade no QQ-plot é uma indicação de que essa distribuição possui cauda pesada.

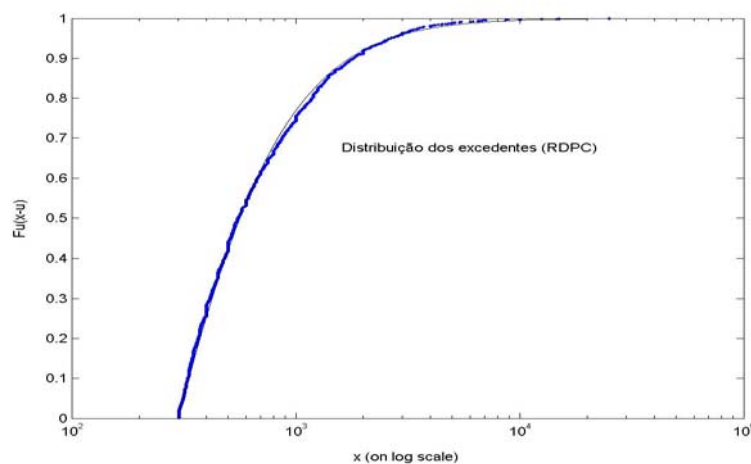
Deve-se, portanto, conduzir algumas tentativas no intuito de se encontrar uma estimativa para o parâmetro *shape* que torne o QQ-plot linear.

Figura 7: QQ-plot ajustado da PNAD 2007



A figura 7 mostra o QQ-plot da curva GPD ajustada contra uma GPD teórica de parâmetro *shape* igual a 0.4. Cabe notar que com essa configuração o gráfico QQ-plot torna-se linear.

Figura 8: Distribuição dos Excedentes da Pnad 2007



A figura 8 delinea a curva estimada da Distribuição dos Excedentes contra a curva teórica. Note que há um ajustamento bastante satisfatório, tendo em vista as duas curvas quase que coincidem. Os parâmetros relativos à curva ajustada para Pnad 2007 são: ($\xi = 0.3720$; $\sigma = 195.95$) e. Logo a GPD ajustada é:

$$G_{\xi,\sigma}(y) = 1 - \left(1 + \frac{0.3720}{195.95} y \right)^{-\frac{1}{0.3720}}$$

Há realmente um sério problema com relação às diferentes estimativas sobre a proporção de pobres, dada a elevada variabilidade dos valores existentes. Em vista disto, procurou-se tomar como referencial o censo-2000, extrair daí uma taxa, e compara-la com aquelas advindas dos dados amostrais do IPEADATA e estimada através do ajustamento da GPD via TVE, ambas baseadas nos dados da PNAD-2001. Dos resultados advindos desses dois procedimentos amostrais, torna irrefutável em favor daquele que mais se aproxime dos valores censitários.

Tabela 03: Comparação da proporção de pobres (p^0)

Estado	CENSO (a)	IPEA (b)	Variação [(b/a) - 1]x100	ESTIMADA (c)	Variação [(c/a) - 1]x100
Ceará	38,86	49,32	+ 26,92	40,12	3,24

Fonte: Elaboração própria com base nos dados do Censo e PNAD.

De acordo com a tabela 3, pode-se observar que a taxa de pobreza calculada através do procedimento aqui adotado é a que mais se aproxima, em termos relativos, daquela obtida com dados censitários. Há forte indicação pelo resultado que as taxas de pobreza obtidas através da TVE possam embutir erros de estimação menores do que as obtidas através de outros métodos. Em vista dessa sustentação metodológica, a tabela 4 expõe estimativas das taxas de pobreza para os anos considerados na análise.

Tabela 04: Valores de p^0 , p^1 e p^2 para a população cearense nos anos analisados

Anos	p^0	p^1	p^2
2001	0,5933	0,3634	0,2686
2003	0,5939	0,3639	0,2690
2005	0,6150	0,3786	0,2805
2007	0,5631	0,3341	0,2428

Fonte: Elaboração própria com base nas PNADs 2001, 2003, 2005 e 2007

Com base numa metodologia portadora de sólido fundamento teórico necessário para construção de uma modelagem estatística das observações extremas da renda dos cearenses, salienta-se, portanto, a grande relevância dos resultados deste trabalho, pois, as estimativas de pobreza aqui reveladas, mostram-se bastante discrepantes das divulgadas em outras análises.

Barreto, *et al* (2007) apresentou, dentre outros resultados, as estimativas de taxas de pobres e indigentes no estado do Ceará no período de 2002 a 2005. Em 2005, por exemplo, estimou-se que 56,38% dos cearenses eram pobres. Com o uso da TVE essa estimativa, usando a mesma base de dados, é de 61,5%, isto é, são 5.12 pontos percentuais a menos na estimativa usando o método tradicional.

Barreto, *et al* (2009) apresentou, dentre muitos outros resultados, a proporção de pobres no estado do Ceará em 2007. A amostra tratou-se da PNAD 2007 utilizando-se da mesma metodologia da seção 4.2.1 empregada neste trabalho, portanto, resultando em estimativas equivalentes às do trabalho ora apresentado. São 51,95% da população cearense pobres em 2007. A estimativa da proporção de pobres resultante do método TVE é 4.35 pontos percentuais mais alta do que pelo método tradicional. Isso mostra a subestimação dessa e de outras taxas que se utilizam do método tradicional, quando não se leva em consideração a real natureza populacional e, o que é mais importante, quando essa taxa envolve valores extremos da unidade experimental.

5. CONCLUSÕES

Existe uma extensa literatura a respeito de investigação sobre densidade da distribuição de renda nos estados brasileiros usando critérios de β -convergência e σ -convergência. Este trabalho traz uma contribuição no que diz respeito ao critério utilizado para se estimar a densidade e focando o estado do Ceará, que na opinião de muitos autores se trata de um dos estados brasileiros mais pobres e com maior desigualdade social.

Pretendeu-se, no trabalho ora apresentado, evidenciar características e, principalmente, inferir sobre a população dos considerados pobre no Ceará durante os anos de 2001 a 2007 usando uma técnica de inferência Estatística não paramétrica de suavização por Núcleo Estocástico. Apesar de muitos trabalhos que tratam de estimação de densidade de qualquer que seja a variável aleatória usarem como ferramenta o Histograma, este trabalho mostrou, na seção 3.1, que esse estimador não é eficiente. O uso da Suavização por Núcleo Estocástico para estimar a densidade do rendimento médio mensal dos pobres do Ceará, além de ser comprovadamente mais eficiente, possibilitou vislumbrar a evolução do comportamento dessa variável e, portanto, inferir em alguns aspectos: Há um constante movimento na curva da densidade de renda dos pobres do Ceará, ano a ano. Vislumbraram-se formações de grupos de famílias pobres em seus respectivos rendimentos, isto é, a partir desse trabalho será possível programar políticas governamentais apropriadas a cada grupo de famílias pobres. Ademais, concluiu-se que, em média, as famílias pobres estão dispondo de mais recursos a cada ano.

Com respeito à estimação da proporção de pobres no Ceará, notou-se uma disparidade grande em relação às estimativas encontradas na literatura e através das duas técnicas utilizadas neste trabalho. A estimativa da proporção de pobres apresentada se utilizando do método tradicional (seção 4.3.1) é inferior á apresentada pela aplicação da teoria do valor extremo. Trata-se de uma diferença de 4,35 pontos percentuais em 2007, por exemplo, que em se tratando de uma população superior aos sete milhões de habitantes, isso se torna bastante significativo. Portanto mostrou-se uma subestimação na proporção de pobres do Ceará. Deve-se lembrar que a estimação via TVE dá um peso bastante maior às observações extremas (observações da cauda inferior da distribuição), portanto, tal estimativa deve representar melhor a verdadeira proporção de habitantes pobres no estado do Ceará.

Considera-se que a principal contribuição desse trabalho foi à aplicação da TVE na estimação da proporção de pobres e indigentes, até então desconhecida existir na literatura. Vale salientar que, independente da linha de pobreza que se atribua, essa é facilmente aplicada à TVE, bastando, para tanto, redefinir os limites de integração.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ARRAES, R. A. *Convergência e Crescimento Econômico do Nordeste*, Revista Econômica do Nordeste, Banco do Nordeste. V. 28, n. Especial, p. 31 – 40, 1997.
- _____. *Há Superestimações das Taxas de Pobreza nas Regiões do Brasil?* Revista Econômica do Nordeste, Banco do Nordeste. V. 39, n. Especial, p. 31 – 40, 2008.
- BARRETO, Flávio Ataliba; MANSO, Carlos Alberto; SIQUEIRA, Marcelo Lettieri; TEOPHILO, Beatriz; PARANGUÁ, Marcelo. *Uma breve análise da evolução dos indicadores de pobreza e desigualdade no Ceará: Período 2002 a 2005*. Laboratório de Estudos da Pobreza CAEN-UFC, 2007.
- BARRETO, Flávio Ataliba; MANSO, Carlos Alberto; MATOS, Paulo Faustino; COSTA, Pedro Andrade da. *O Estado do Ceará de Tasso Jereissati a Cid Gomes: Período 2002 a 2005*. Laboratório de Estudos da Pobreza CAEN-UFC, Relatório de Pesquisa n° 3, 2009.

- BARROS, R.P.; HENRIQUES, R.; MENDONÇA, R. *A estabilidade inaceitável: desigualdade e pobreza no Brasil*. In: HENRIQUES, R.(org.). *Desigualdade e Pobreza no Brasil*. Rio de Janeiro: IPEA, 2000.
- BARROS, R. P.; CARVALHO, M.; FRANCO, S.; MEDONÇA, R. **A Queda Recente da Desigualdade de Renda no Brasil**. Rio de Janeiro: IPEA, jan. 2007. (Texto para Discussão, 1258).
- BARROS, R. P.; FRANCO, S.; MEDONÇA, R. **A Recente Queda da Desigualdade de Renda e o Acelerado Progresso Educacional Brasileiro da Última Década**. Rio de Janeiro: IPEA, set. 2007. (Texto para Discussão, 1304).
- BARROS R.P. de, MENDONÇA, R.S.P. de, ROCHA, S. *Welfare, inequality poverty, social indicators and social programs in Brazil in the 1980s. mimeo*, 1993.
- BARROS, R.P; CARVALHO, Mirela de.; FRANCO, Samuel. E MENDONÇA, Rosane. Uam análise das principais causas da queda recente na desigualdade de renda brasileira. Revista Econômica, Junho de 2006.
- BLACKWELL,D. *Estatística básica*. São Paulo: Editora McGraaw-Hill do Brasil Ltda., 1973.
- CASTILLO, E. and HADI, A. (1997). *Fitting the Generalized Pareto Distribution to Data*. Journal of the American Statistical Association, 92(440): 1609-1620.
- COLES, S. *An Introduction to Statistical Modeling of Extreme Values*. Springer. 2001.
- FERNANDES, R.; MENEZES-FILHO, N. *A Evolução da Desigualdade no Brasil Metropolitano entre 1983 e 1997*. Estudos Econômicos, São Paulo, v.30, n.4, p.549-569, out-dez., 2000
- FERREIRA, F. H. G.; LEITE, P. G.; LITCHFIELD, J.; ULYSSEA, G. A. *Ascensão e Queda da Desigualdade de Renda no Brasil*. *Econômica*, Rio de Janeiro, v.8, n.1, p.147-169, jun., 2006.
- GILLI, Manfred; KELLEZI, Evis. *An Application of Extreme Value Theory for Measuring Financial Risk*. Computational Economics 27(1), 2006, 1-23.
- GRIMSHAW, S. (1993). *Computing the Maximum Likelihood Estimates for the Generalized Pareto Distribution to Data*. *Technometrics*, 35(2): 185-191.
- HOFFMANN, Rodolfo. *Considerações sobre a evolução recente da distribuição da renda no Brasil*. Revista de Administração de Empresas, 13 (4): 7-17, out/dez de 1973.
- HOFFMANN, Rodolfo. *Elasticidade e pobreza em relação à renda média e à desigualdade no Brasil e nas unidades da federação*. Revista Econômica, Julho de 2005.
- HOFFMANN, Rodolfo. *Transferências de renda e a redução da desigualdade no Brasil*. Revista Econômica, Junho de 2006.
- HOFFMANN, R. *Transferências de Renda e a Redução da Desigualdade no Brasil e cinco Regiões entre 1997 e 2004*. *Econômica*, Rio de Janeiro, v.8, n.1, p.55-81, jun., 2006.
- HOGG, ROBERT V. e CRAIG, ALLEN T. CRAIG. *Introduction to Mathematical Statistics*. 5 ed. Prattice Hall, Upper Saddle River, New Jersey, 1995.
- SILVERMAN, B. W. *Density Estimation for Statistics and Data Analysis*. 1 ed., New York, Wasshington D.C: Chapman & Hall / CRC, 1998.
- SOARES, Sergei; MEDEIROS, Marcelo; OSÓRIO, Rafael G. *Programas de transferência de renda no Brasil: impactos sobre a desigualdade*. [td] Instituto de Pesquisa Econômica Aplicada - IPEA - Dominio Publico, 2006.