

Dimensão, magnitude e localização das populações pobres no Brasil

Bernardo Campolina*
Fernando Gaiger Silveira†
Alexandre Carvalho Ywata‡
Carlos Roberto Azzoni§
Antônio Ibarra**

Resumo

A ênfase recente dos trabalhos sobre pobreza tem se preocupado em determinar o tamanho das populações pobres que servem de subsídio para os programas de combate a fome e a pobreza. São diversas as perguntas que tem sido feita, Qual a linha de pobreza? Qual o número de pobres? Quanto de dinheiro deve ser dado a uma família pobre para que ela tenha condições de sair da linha de indigência? A construção de linhas de pobreza baseadas em renda nacional em geral produz grandes distorções regionais. Usando informações da Pesquisa de Orçamento Familiar (POF) do IBGE que cobre regiões urbanas, metropolitanas, e rurais do país este artigo desenvolve e aplica uma nova metodologia para definir linhas de pobreza para diferentes regiões. Estas linhas são definidas a partir de requerimentos calóricos, de forma a evitar o problema da paridade do poder de compra. Além disso, a POF pesquisou a renda não monetária e produção para o próprio consumo, aspectos que são de extrema importância principalmente na área rural. Por ultimo, é feito um *matching* com dos dados do Censo, permitindo que seja feita a espacialização da pobreza nos municípios.

Código JEL: I 32; J18; R 29

Palavras chaves: Pobreza regional; nutrição; linhas de pobreza; regressão semi-paramétrica; modelo logito

Abstract

The recent emphasis on fighting poverty in Brazil makes the determination of the size of the targeted population an important issue (What is the right poverty line? What is the real size of the poor population? How much money should be given to each poor family?). The application of poverty lines based on national income levels tends to produce important distortions at the regional level. Using data from a Household Expenditure Survey (HES) that covered some regions in Brazil, the paper develops and applies a methodology to define poverty lines for all regions and urban areas. These lines are based on nutritional requirements, thus avoiding the purchasing power parity problem, and take into account non-monetary income and in-kind consumption, aspects that are very important at the rural level. The HES results are matched with Census data, allowing for the estimation of rural and urban poverty lines for Brazilian regions.

JEL Classification: I 32; J18; R 29

Key Words: Regional Poverty; nutrition; poverty lines; semi-parametric models; logit

* Professor Adjunto da Universidade Federal do Tocantins. Bolsista DCR/CNPq/SECT.

† Técnico de Pesquisa do IPEA (DISOC) e Doutor em Economia pela IE/Unicamp.

‡ Técnico de Pesquisa do IPEA (DIRUR).

§ Professor Titular da FEA/USP.

** Consultor do IPEA.

Dimensão, magnitude e localização das populações pobres no Brasil

1 Introdução

A importância do tema da pobreza no Brasil e, mais especificamente, da identificação das populações em situação de risco alimentar pode ser medido pela repercussão que teve a implantação do Programa Fome Zero e os debates que ensejou. Processo que levou o Governo Federal a alterar a abordagem da questão, com a unificação dos programas de transferência de renda e de combate a fome no Bolsa Família. Em junho último, o Bolsa Família atendeu a 11,2 milhões de famílias, em 5560 municípios, totalizando R\$ 682 milhões. Soma-se a isso os benefícios ainda concedidos pelo Bolsa Escola, pelo Bolsa Alimentação, pelo Cartão Alimentação e pelo Auxílio Gás, que, juntos, em junho último, atingiram a 1,5 milhões de famílias, com um desembolso total da ordem de R\$ 25 milhões.

Uma forma de avaliar a pobreza é através de medidas antropométricas, isso é feito de duas formas, uma consiste em determinar a participação de crianças com baixa estatura supera ao que é considerado normal (5%); a segunda forma consiste em estimar a população com deficiência energética crônica com base na proporção de indivíduos magros (índice de massa corporal inferior a 18 kg/m^2) ^{††}. Outra forma de mensurar indigência e pobreza consiste em determinar um valor de renda que permita a família atender as necessidades básicas das famílias, particularmente as alimentares.

Nesse caso, observam-se duas abordagens: de um lado, aquelas que empregam como rendas mínimas valores fixos, sejam parcelas do salário mínimo – atual ou valor real de outro período – sejam valores “internacionalmente” definidos – como o US\$ 1 ajustado ao poder de compra local do Banco Mundial; e, de outro, as que estimam esses valores limites com base nos dados do consumo alimentar observado. Abordagem essa que é realizada no presente estudo. Ambas abordagens determinam as chamadas linhas de indigência ou de pobreza que, em um segundo momento, serão cotejadas com os recebimentos apurados nas pesquisas domiciliares, especialmente àquelas de maior amplitude territorial e temporal.

Assim, no presente estudo, considera-se a extrema pobreza e a pobreza como, respectivamente, as seguintes situações: famílias que apresentam renda inferior ao necessário para cobrir gastos alimentares, cuja disponibilidade calórica atende ao **mínimo** requerido; famílias com renda insuficiente para adquirir os bens **básicos** em moradia, transporte, vestuário e, em alguns casos, saúde e educação^{‡‡} da família.

Dimensionar e apontar a concentração espacial da indigência e da pobreza são os principais objetivos do presente trabalho, empregando, para tanto, a relação entre a disponibilidade calórica e a renda para definir as populações com insuficiência de renda, para, de um lado, cobrir os gastos alimentares que fornecem tal disponibilidade e, por outro, para cobrir as necessidades básicas. Tendo em conta que a definição tanto das linhas de indigência e pobreza como das populações nessas situações é realizada com base na POF, cuja subenumeração das rendas é bem menor daquela que as outras investigações domiciliares incorrem, decidiu-se mudar a forma de espacializar a pobreza, empregando, para tanto, de técnicas de associação entre pesquisas com base em um conjunto mais amplo de indicadores socioeconômicos e

^{††} Em junho último, foram divulgados os dados da POF 2002-03 referentes a antropometria das crianças e adolescentes, sendo que a prevalência de déficit peso para idade das crianças menores de 5 anos ficou em 4,6%, com a região Norte, onde este indicador atingiu a 8%, destoando do restante do país.

^{‡‡} Não considerar os gastos em saúde e educação se deve ao fato que cabe ao estado fornecer aos cidadãos, educação básica e saúde gratuita - é direito do cidadão e obrigação do Estado.

demográficos. Fica, ainda, pendente métodos para acompanhar o comportamento dos indicadores de pobreza e indigência, que não se restrinjam ao cotejamento entre as renda e valores atualizados das linhas de indigência e pobreza.

O artigo está organizado em cinco tópicos, além dessa introdução: o segundo trata da base de dados, isto é, apresenta sumariamente as principais características da POF. No terceiro item são apresentadas as escolhas metodológicas para a estimação das populações indigentes e pobres, comparando-se a metodologia proposta pela Comissão Mista IPEA-Cepal-IBGE (ou Comissão de Pobreza) com a aqui proposta. Cabe destacar que as sugestões da Comissão foram, em grande medida, incorporadas no presente estudo. No tópico 4, encontram-se os resultados das estimativas de famílias indigentes e pobres na POF. Nos dois itens seguintes discutem-se a metodologia de espacialização da indigência e pobreza e, por fim, os resultados desse esforço de discriminação geográfica.

2 Base de Dados

A investigação domiciliar empregada no presente trabalho é a Pesquisa de Orçamentos Familiares, comumente conhecida por POF, realizada pelo IBGE, entre julho de 2002 e junho de 2003. A POF permite aferir a qualidade de vida da família brasileira, ou melhor, o perfil socioeconômico da população por meio de suas estruturas orçamentária e de recebimentos (IBGE, 2004).

A coleta de dados da POF 2002-2003 foi feita por meio de seis questionários, que cobrem, respectivamente: a) informações sobre as condições do domicílio - abastecimento de água, infra-estrutura sanitária, número de cômodos, condição de ocupação etc. - e as características dos indivíduos - sexo, nível de instrução, idade, frequência à escola, peso, altura -; b) o questionário com despesas coletivas, ou seja, com serviços públicos, com aluguéis, taxas, decoração, melhorias e reformas do domicílio, com bens de consumo duráveis e com serviços domésticos; c) caderneta de despesa coletiva, que engloba alimentação e higiene e limpeza; d) gastos individuais - vestuário, saúde, educação, higiene pessoal, transporte etc.; e) os recebimentos (salariais e não-salariais) reportados individualmente; e f) avaliação subjetiva das condições de vida das famílias. Nas informações sobre as despesas coletivas e individuais, há diferentes períodos de referência, como sete, 30 e 90 dias e seis meses, a depender do tipo de gasto - sazonalidade, frequência, facilidade de recordação etc.. A data de referência da pesquisa é 15 de janeiro de 2003, com todos os valores corrigidos para essa data.

Destaca-se que a nova POF também levantou as despesas e recebimentos não-monetários, fato este que enriquece a análise e reabre novos rumos no debate sobre a insuficiência de renda para mensurar o total de indigentes e pobres no Brasil. Esta informação é de grande relevância posto que atinge todas as camadas sociais. A produção própria e trabalho da família, no meio rural, e a ação das redes de proteção social (merenda escolar, alimentação na empresa, etc.), especialmente nas médias e grandes cidades, são exemplo da importância da investigação dessas despesas.

A amostra da POF abrangeu 48.470 domicílios, onde residiam 182.333 pessoas, representando 0,1% do total de domicílios do país. Seu desenho amostral permite análises para 70 domínios geográficos, a saber: para as populações totais e urbanas das 27 unidades da federação, as áreas rurais das cinco grandes regiões, as nove regiões metropolitanas - Porto Alegre, Curitiba, São Paulo, Rio de Janeiro, Belo Horizonte, Salvador, Recife, Fortaleza e Belém -, a cidade de Goiânia e o Distrito Federal.

3 Metodologias de estimação das populações rurais indigentes (extrema pobreza) e pobres

A metodologia aqui proposta visa estimar a dimensão da indigência (extrema pobreza) e da pobreza no Brasil, com base na Pesquisa de Orçamento Familiar (POF), para, posteriormente, associar os resultados aos dados da Amostra do Censo 2000, com o objetivo de discriminar espacialmente os resultados. Assim, a idéia básica é utilizar, em uma primeira etapa, modelos de regressão não-paramétrica, para obter uma estimativa dos valores das linhas de indigência e de pobreza para domínios geográficos selecionados. Em seguida as famílias que, na POF, se encontram em situação de indigência e pobreza, são caracterizadas demográfica, social e economicamente, visando verificar, em que medida, tais características determinam a indigência e/ou a pobreza. Tendo sido selecionadas as variáveis determinantes, por meio de um modelo lógico, realiza-se, em seguida, o pareamento entre a POF e o Censo.

3.1 Nova metodologia proposta de cálculo das linhas de indigência e pobreza

A metodologia atualmente aplicada para cálculo da linha de pobreza Comissão de Pobreza, apesar de ser intuitiva e de simples utilização tem alguns problemas, abaixo enumerados:

1. A escolha de 100 grupos de renda (100 centésimos) é arbitrária, assim como o número 20 para a média móvel de consumo. De fato, é provável que alterações nesses números resultem em valores bem distintos para o número de indigentes e para as conclusões gerais.
2. Mesmo com 20 grupos de renda na média móvel, isso ainda não garante monotonicidade na curva de consumo por equivalente adulto versus renda *per capita*, o que indica a fragilidade das estimativas obtidas. Dada a grande variabilidade nas medições de consumo calórico, pode acontecer que a curva estimada, de acordo com a média móvel de 20 centésimos, apresente trechos crescentes e trechos decrescentes.
3. Essa metodologia não se enquadra como um estimador estatístico a partir de técnicas regressão padrões.
4. Para amostras menores, o estimador pode não funcionar e a não monotonicidade da relação consumo calórico por equivalente adulto versus renda *per capita* pode ficar ainda mais evidente.
5. Por não se tratar de uma metodologia nitidamente via modelos de regressão, ela não permite a utilização de testes de hipóteses, intervalos de confiança e erros padrões.
6. Finalmente, utilizam-se médias móveis com 20 centésimos para todos os contextos geográficos indistintamente. Com isso, não se leva em conta a diferente disponibilidade de informação específica de cada contexto, nem a variabilidade de medição do consumo calórico. De fato, no contexto geográfico do Nordeste Rural, por exemplo, o número de famílias amostradas é bem maior do que o número de famílias em cada região metropolitana. Portanto, o número de observações para estimação dos modelos estatísticos varia de contexto para contexto.

Diante das restrições descritas acima na abordagem atual para cálculo da linha de indigência, propõe-se, a seguir, uma metodologia, baseada em modelos de regressão não-paramétrica, ou modelos de regressão semi-paramétrica, para a função do consumo calórico versus a renda *per capita*.

- a. estimação não-paramétrica da curva consumo-renda

Nesta seção apresentam-se os modelos de estimação não-paramétrica, ou semi-paramétrica, para estimar a curva de consumo calórico por equivalente adulto versus a renda *per capita*. O termo regressão semi-paramétrica deve-se ao fato de se utilizar expansões aparentemente

paramétricas para aproximar formas funcionais desconhecidas^{§§}. Apesar de as estimações realizadas neste trabalho terem sido realizadas por contexto geográfico, na exposição apresentada a seguir, suprime-se o subscrito k , correspondente ao contexto k , de forma a simplificar a notação. A idéia da estimação não-paramétrica neste projeto consiste em estimar uma função do tipo:

$$c_i / n_i = g(r_i / m_i) + \varepsilon_i, \quad (1)$$

onde, conforme apresentado acima, c_i é o consumo total da família, o valor r_i é a renda total da família, n_i é o número de pessoas, em equivalente adultos, e m_i é o número total de indivíduos na família. O quociente r_i/m_i corresponde à renda familiar *per capita*, enquanto c_i/n_i corresponde ao consumo calórico por equivalente adulto. O índice i corresponde à família na base de dados, e o termo ε_i é uma variável aleatória, com média zero e variância desconhecida.

A função $g(r_i/m_i)$ possui forma funcional desconhecida e deverá ser estimada a partir dos dados. Para estimá-la, emprega-se uma estimação semi-paramétrica, utilizando uma expansão de funções base (*basis functions*). Essa expansão baseia-se na aproximação da curva desconhecida $g(r_i/m_i)$ utilizando a forma paramétrica flexível:

$$g(r_i / m_i) \cong \sum_{l=1}^L b_l \times u_l(r_i / m_i), \quad (2)$$

onde as funções $u_l(r_i / m_i)$ são as funções base, com forma funcional conhecida. A forma funcional em (2) engloba uma série de modelos comumente encontrados na literatura de *data mining*. Entre esses modelos encontram-se, por exemplo, regressões de redes neurais e regressões de *wavelets* (vide Hastie, Tibshirani e Friedman, 2001).

Neste trabalho, utiliza-se uma expansão do tipo *B-splines* de ordem q . Devido ao fato de as funções de base $u_l(x)$ nas expansões de *B-splines* dependerem de q , escreve-se explicitamente $u_{l,q}(x)$. Assume-se que a variável explicativa (independente) x varia no intervalo $[x_{min}, x_{max}]$ ^{***}. Considere-se um vetor de w pontos (x_1, x_2, \dots, x_w) dividindo o intervalo $[x_{min}, x_{max}]$, onde $x_{min} < x_1 < x_2 < \dots < x_w < x_{max}$. A idéia da expansão de *B-splines* é ajustar um polinômio de grau $(q-1)$ em cada intervalo definido por pontos consecutivos (pontos no conjunto $x_{min}, x_1, x_2, \dots, x_w, x_{max}$). Em geral, utiliza-se $q = 3$ ou 4 , de forma que os polinômios utilizados possuem grau 2 ou 3 respectivamente.

Considere agora o vetor de nós $x_{min}, \dots, x_{min}, x_1, x_2, \dots, x_w, x_{max}, \dots, x_{max}$, onde os valores x_{min} e x_{max} nos extremos são repetidos um número q de vezes. Para facilitar a discussão a seguir, escreve-se o vetor $(x_{min}, \dots, x_{min}, x_1, x_2, \dots, x_w, x_{max}, \dots, x_{max})$ na forma $(t_1, t_2, \dots, t_w + 2q)$. Portanto, $t_1 = x_{min}, \dots, t_q = x_{min}$. A partir do vetor de nós $(t_1, t_2, \dots, t_w + 2q)$ e da ordem q , as funções base $u_{l,q}(x)$ podem ser construídas recursivamente, como segue

$$u_{l,1}(x) = \begin{cases} 1, & t_l \leq x < t_{l+1} \\ 0, & \text{caso contrário,} \end{cases} \quad (3)$$

^{§§} Modelos de regressão não-paramétrica são aqueles nos quais a forma funcional da função resposta não é conhecida e tem de ser estimada a partir dos dados. Em muitos casos, essa estimação é feita por meio de expansões polinomiais de funções base (*basis functions*) e o problema não paramétrico transforma-se em um problema de estimar um modelo paramétrico, onde as transformações nas variáveis explicativas são apropriadamente construídas. Nesse caso, o modelo de regressão não-paramétrica pode ser denominado modelo de regressão semi-paramétrica. Para mais detalhes, ver Hastie, Tibshirani e Friedman (2001).

^{***} Neste trabalho, a variável explicativa é a renda *per capita* r_i/m_i .

$$u_{l,p}(x) = \frac{x-t_l}{t_{l+p-1}-t_l} u_{l,p-1}(x) + \frac{t_{l+p}-x}{t_{l+p}-t_{l+1}} u_{l+1,p-1}(x), \text{ para } p = 2, \dots, q. \quad (4)$$

Observe-se que no denominador do segundo termo na expressão (4), aparece o valor t_{l+p} . Portanto, em *B-splines* de ordem q , o número total de funções base será igual ao número de nós menos o valor q . Ou seja, o número L de funções base é dado por $L = w + q$, e as funções base serão $u_{1,q}(x), \dots, u_{L,q}(x)$. Pode-se então reescrever a expressão (2), especificamente para a expansão de *B-splines*, como

$$g(r_i / m_i) \cong \sum_{l=1}^L b_l \times u_{l,q}(r_i / m_i). \quad (5)$$

O grau de flexibilidade da forma funcional em (5) é regulado pelo número L de funções base. Esse número está diretamente relacionado ao número de pontos divisórios w , bem como à ordem q . Quanto maior o número L , maior a flexibilidade da expansão semi-paramétrica. Uma vez fixado o número de funções base L para estimar a função desconhecida $g(r_i/m_i)$, a estimação dos parâmetros $b_l, l = 0, 1, 2, \dots, L$ pode ser efetuada via estimadores paramétricos tradicionais. Pode-se utilizar, por exemplo, estimadores de mínimos quadrados ordinários, ou algum estimador do tipo mínimos quadrados ponderados, com correção para heteroscedasticidade, ou para a presença de observações discrepantes (*outliers*). Observe-se agora que o problema de ajuste da curva $g(r_i/m_i)$ consiste na estimação do modelo de regressão (linear nos parâmetros):

$$(c_i / n_i) = \sum_{l=1}^L b_l \times u_{l,q}(r_i / m_i) + \varepsilon_i. \quad (6)$$

A escolha do número L de funções base pode ser efetuada utilizando-se algum critério de seleção de modelos, como por exemplo, o AIC ou o BIC^{†††} (regras de escolha *in-sample*), ou utilizando-se alguma regra de validação cruzada (*cross-validation*), que correspondem a regras de escolha *out-of-sample*. Esses procedimentos têm por objetivo evitar um sobre ajuste (*overfitting*) do modelo semi-paramétrico. De fato, quando se aumenta o número de funções base indefinidamente, obtém-se um modelo com ajuste perfeito dentro da amostra, mas com um poder preditivo questionável para observações fora da massa de dados. Por outro lado, para um número L muito baixo, obtém-se uma baixa flexibilidade da expansão de funções base, o que também incorre na perda de poder preditivo do modelo. Os diversos critérios de seleção de modelos auxiliam na escolha do número L de forma a maximizar a relação de compromisso entre flexibilidade do modelo e número excessivo de parâmetros desconhecidos^{†††}. Para maiores detalhes em escolhas do grau de flexibilização em modelos semi-paramétricos e não-paramétricos, vide Hastie, Tibshirani e Friedman (2001).

Observe-se que a estimação da função $g(r_i, n_i)$ pode ser efetuada diretamente nos micro dados, a partir da POF. Portanto, não há necessidade de um agrupamento prévio desses dados em centésimos, por exemplo, conforme efetuado na metodologia atualmente vigente para cálculo da linha de pobreza (vide seção anterior). Por outro lado, a escolha do grau de flexibilidade dos estimadores da função $g(\cdot)$ utilizando os critérios mais robustos descritos acima possibilita uma diminuição no grau de arbitrariedade empregado na escolha do número de centésimos nas médias móveis da metodologia atual. Além disso, escolhendo-se formas e valores L diferentes para diferentes contextos geográficos, podem-se obter expansões semi-paramétricas mais apropriadas para as particularidades do conjunto de informações disponível em cada região.

^{†††} Para maiores detalhes sobre o AIC e o BIC, ver Burnhan e Anderson (1998).

^{†††} Na literatura em estatística, esse problema é comumente conhecido como *trade-off* viés-variância (*bias-variance*).

b. imposição de monotonicidade na relação consumo-renda

Para a relação entre consumo calórico por equivalente adulto e renda *per capita*, uma hipótese plausível é que a função $g(r_i/m_i)$ seja monotonicamente crescente em r_i/m_i . Nesse caso, a expansão de funções base utilizada deve ser escolhida de forma a garantir essa monotonicidade, conforme discutido em Chen e Conley (2001) e em Leitenstorfer e Tutz (2005a e 2005b). Para garantir monotonicidade na expansão de B-splines em (6), basta impor nos parâmetros a serem estimados a restrição $b_1 \leq b_2 \leq \Lambda \leq b_L$.

Os métodos de estimação tradicionais, do tipo mínimos quadrados ordinários, não incorporam diretamente as restrições nos parâmetros b_1, b_2, \dots, b_L para garantir monotonicidade. Para contornar esse problema, utilizou-se um mecanismo de estimação via método de mínimos quadrados ponderados restrito. Utilizando essa abordagem, foi possível acomodar também, de maneira simples, os pesos dados a diferentes observações na amostra da POF, devido ao desenho amostral específico. Além disso, escolhendo-se iterativamente os pesos, é possível incorporar também o tratamento de observações discrepantes (*outliers*). Um passo importante nesse procedimento de estimação para garantir monotonicidade (imposta via restrição $b_1 \leq b_2 \leq \Lambda \leq b_L$) foi a utilização de métodos de maximização quadrática, com restrições lineares, utilizando os algoritmos descritos em Winston (2003) ou Hillier e Lieberman (2002).

O estimador de mínimos quadrados ponderado restrito, conforme discutido acima, fornece estimativas pontuais para os parâmetros b_1, b_2, \dots, b_L . A partir desses valores e das formas funcionais conhecidas para as funções base, é possível construir estimadores pontuais para o valor de consumo calórico médio c_i/n_i , por equivalente adulto, correspondente a um determinado valor de renda *per capita* r_i/m_i . Variando-se continuamente os valores de r_i/m_i , é possível obter-se a estimativa $\hat{g}(r_i/m_i)$ da curva $g(r_i/m_i)$ de consumo calórico por equivalente adulto versus renda *per capita*.

Uma vez escolhida e estimada a forma funcional de $g(r_i/m_i)$, o próximo passo é determinar os valores de renda *per capita* r_i/m_i de corte para os diferentes contextos geográficos. Essa etapa é feita de forma análoga ao que está discutido no passo (7) da metodologia atualmente empregada para a estimação da linha de pobreza com a metodologia atual (Seção 3.1). A informação principal aqui é o número de calorias de corte, por equivalente adulto. Seja C esse valor mínimo de calorias por equivalente adulto^{§§§}. O valor de renda *per capita* de corte é dado pelo ponto de interseção entre a curva estimada $\hat{g}(r_i/m_i)$ e o a reta horizontal passando pelo valor $c_i/n_i = C$. Portanto, a renda *per capita* de corte \hat{r}_{corte} é simplesmente a solução da expressão $\hat{g}(\hat{r}_{corte}) = C$. A partir da estimativa \hat{r}_{corte} , pode-se utilizar o procedimento descrito na subseção (c) mais adiante, onde replicaram-se as regressões semi-paramétricas para estudar a relação entre gasto alimentar *per capita* e renda *per capita*. Chega-se, então, com base nessas curvas ao gasto de corte \hat{g}_{corte} , correspondente a \hat{r}_{corte} . Gasto de corte que definirá as linhas de indigência e de pobreza para as diversas áreas geográficas no Brasil, determinando, conseqüentemente, o número de famílias indigentes e pobres, ou seja, as famílias com renda *per capita* inferior ao gasto de corte \hat{g}_{corte} . Além disso, pode-se determinar especificamente quais famílias na amostra disponível (essa amostra será na verdade extraída a partir de filtros nos micro dados da POF) estão abaixo dessas linhas. Essa informação será particularmente

§§§ Conforme discutido no início desta seção, apesar de todas as estimações serem feitas com amostras específicas por contexto geográfico, para facilitar a exposição suprimem-se os subscritos k , referentes aos contextos.

importante ao trabalharmos com os algoritmos de *matching* para localizar os pobres nos municípios brasileiros.

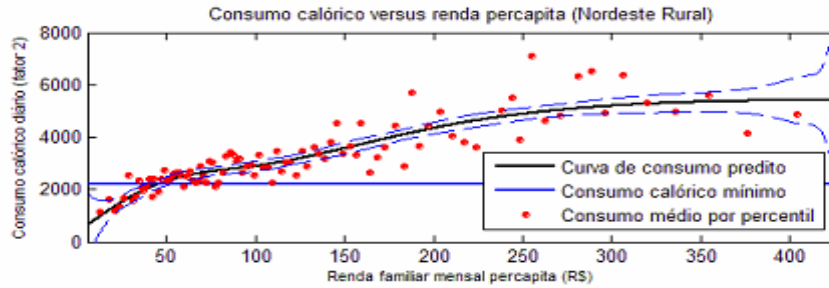
Uma vantagem na utilização de uma estimação não-paramétrica, ou semi-paramétrica, para a função $g(r_i/m_i)$ é a possibilidade de estabelecer intervalos de confiança, utilizando-se, por exemplo, estimadores *bootstrap* (vide, por exemplo, Davison e Hinkley, 1997, Hall, 1992), ou aproximações de primeira ordem (vide, por exemplo, Lehmann, 1999). Com isso, também é possível estabelecer intervalos de confiança para a estimativa da renda *per capita* de corte \hat{r}_{corte} e estabelecer também intervalos para a estimativa do número de indigentes. Uma alternativa computacionalmente menos demandante do que o *bootstrap* é utilizar intervalos de confiança para a média $\hat{g}(r_i/m_i)$ de consumo calórico por equivalente adulto versus renda *per capita*, utilizando estimadores baseados em métodos de estimação via mínimos quadrados tradicionais (vide Draper e Smith, 1998).

Com base nessa metodologia foram estimadas as curvas de consumo calórico por equivalente adulto versus renda *per capita* para os vinte e dois contextos geográficos, assumindo-se que o valor da caloria fora de casa é igual ou é duas vezes o valor da caloria dentro da residência. Como ilustração são apresentados, na Figura 1, os gráficos para o Nordeste Rural, considerando as duas hipóteses de valoração da caloria extra-domicílio. Nesses gráficos, o eixo vertical corresponde ao consumo calórico diário por equivalente adulto, enquanto o eixo horizontal corresponde à renda *per capita*. A curva em preto corresponde à estimativa $\hat{g}(r_i/m_i)$ da curva $g(r_i/m_i)$ de consumo calórico por equivalente adulto versus a renda *per capita*, enquanto as curvas tracejadas em azul em torno da curva estimada $\hat{g}(r_i/m_i)$ referem-se às bandas de estimação da curva $g(r_i/m_i)$. Essas curvas correspondem a um nível de significância de 90%. A curva horizontal sólida em azul corresponde ao corte consumo calórico mínimo do contexto geográfico.

Para dar uma idéia do ajuste das curvas estimadas, apresentam-se os pontos em vermelho, referentes às médias de consumo calórico por centésimo de renda. Para evitar problemas causados por valores de renda muito altos na amostra, excluíram-se das estimações as observações referentes aos centésimos de renda acima de 95%. Portanto, nas estimativas, as 5% famílias com maiores rendas *per capita* foram excluídas. Nota-se uma grande dispersão nos valores, o que fornece uma idéia do grau de dispersão dos micro dados disponíveis na POF. Apesar de os gráficos mostrarem as médias dos centésimos de renda, é importante ressaltar que todas as estimativas foram diretamente realizadas com todos os microdados da POF, utilizando-se toda a informação disponível, uma das vantagens do método apresentado neste trabalho.

A interseção entre a curva estimada $\hat{g}(r_i/m_i)$ com a curva horizontal correspondente ao consumo calórico de corte fornece a renda *per capita* de corte, que define as famílias pobres. Além disso, a interseção das curvas tracejadas em azul com a curva horizontal de consumo calórico mínimo fornece os limites inferiores e superiores para a renda *per capita* de corte. Em alguns casos, a banda superior não intercepta a curva horizontal de consumo calórico mínimo, impossibilitando a inferência em relação ao limite inferior da renda *per capita* de corte.

Figura 1
Curvas de consumo calórico diário equivalente adulto versus renda familiar *per capita*,
Nordeste Rural.



c. estimação das linhas de indigência e pobreza

Com a definição do nível de renda familiar *per capita* \hat{r}_{corte} compatível com o consumo calórico adequado, realizou-se, então, a estimação do gasto alimentar *per capita* que está associado àquela renda. Essa etapa é equivalente ao passo (8) na metodologia atualmente empregada (Seção 3.1). O valor deste gasto será definido, então, como sendo a linha de indigência, tendo sido estimadas linhas de indigência para cada um dos contextos. São consideradas indigentes aquelas famílias cuja renda familiar *per capita* é inferior ao valor da linha de indigência. Já as linhas de pobreza são valoradas a partir das linhas de indigência, empregando para tanto multiplicadores distintos para os meios urbanos e rurais, com base em estudos da Cepal.

Neste item, discorre-se sobre a etapa final de estimação das linhas de indigência e pobreza, ou seja, a definição dos gastos alimentares *per capita* associados as renda de corte. Empregou-se basicamente a mesma metodologia utilizada para estimar a curva de consumo versus renda, ou seja, aplicou-se regressões semi-paramétricas para a associação renda versus gasto.

$$g_i / m_i = h(r_i / m_i) + u_i, \quad (7)$$

onde g_i é o gasto mensal com alimentação da unidade de consumo i , m_i é o número de moradores na unidade de consumo, r_i é a renda familiar mensal e u_i é o termo idiossincrático, representando os fatores não incluídos na regressão. Portanto, o modelo acima corresponde a curva desconhecida $h(\cdot)$, que fornece a relação média entre gasto alimentar *per capita* e renda familiar *per capita*. Para obter os estimadores $\hat{h}(\cdot)$ da função $h(\cdot)$, nós utilizamos os mesmos procedimentos via expansão de B-splines, descritos anteriormente.

Para cada contexto geográfico, seja \hat{r}_{corte} o valor da renda *per capita* de corte obtido na seção anterior. A partir desse valor \hat{r}_{corte} , podemos encontrar o valor de gasto *per capita* correspondente $\hat{g}_{corte} = \hat{h}(\hat{r}_{corte})$, a partir da curva estimada $\hat{h}(\cdot)$. Esse valor de gasto *per capita* de corte \hat{g}_{corte} será então o valor indicativo da linha de indigência para o contexto correspondente. Multiplicando esse valor de corte \hat{g}_{corte} por dois, obtém-se o valor de corte para a pobreza nas áreas urbanas e multiplicando \hat{g}_{corte} por 1,75 obtém-se o valor de corte para a pobreza nas áreas rurais. Lembramos que, devido ao problema de imputação para o valor da caloria consumida fora de casa, todos os resultados foram calculados assumindo-se, de uma parte, que o valor da caloria fora de casa é duas vezes o valor dela no domicílio e, de outra, que tais valores são iguais.

4 – Resultados

Os resultados mostraram-se bem consistentes, tendo, inclusive, certo paralelismo com a auto-avaliação da população sobre sua situação alimentar. Concretamente, o IBGE, na Pnad de 2004, investigou, em suplemento específico o quadro de “*Segurança Alimentar*” do país, tendo apontado que cerca de 14 milhões de pessoas se encontra em grave insegurança alimentar. Observa-se ainda que todas as linhas de indigência são inferiores a R\$ 100, de 15 de janeiro de 2003, salvo na RM de São Paulo, que atinge a R\$ 103, quando se considera o fator 2 de valoração do preço da caloria fora. Isto é, 1/2 salário mínimo da época, para a indigência. Além disso, a razão entre as maiores e as menores linhas de indigência e de pobreza encontra-se, ao redor, de 3 vezes. O que pareça razoável como reflexo das diferenças de custo de vida, de hábitos de consumo, bem como das características dos mercados alimentares, entre os contextos geográficos (Tabela 3).

Tabela 3

Valores das linhas de indigência e pobreza, segundo contextos geográficos e hipóteses de valoração do preço da caloria fora do domicílio. - (R\$ de janeiro de 2003)

Contexto Geográfico	Indigência		Pobreza	
	fator 1	fator 2	fator 1	fator 2
<i>Belém</i>	51,01	55,11	102,01	110,21
<i>Norte Urbano</i>	39,48	40,77	78,95	81,54
<i>Norte Rural</i>	33,91	34,62	59,34	60,59
<i>Fortaleza</i>	52,84	55,00	105,68	109,99
<i>Recife</i>	45,19	61,57	90,37	123,14
<i>Salvador</i>	51,90	61,67	103,79	123,34
<i>Nordeste Urbano</i>	43,39	53,40	86,79	106,79
<i>Nordeste Rural</i>	33,16	34,68	58,02	60,70
<i>Belo Horizonte</i>	57,19	59,56	114,38	119,12
<i>MG+ES+RJ Urbano</i>	55,61	57,35	111,21	114,70
<i>MG+ES+RJ Rural</i>	44,23	44,23	77,40	7,40
<i>Rio de Janeiro</i>	58,34	68,92	116,67	137,83
<i>São Paulo</i>	81,98	102,55	163,97	205,10
<i>São Paulo Urbano</i>	52,33	53,27	104,65	106,53
<i>São Paulo Rural</i>	40,47	40,79	70,82	71,38
<i>Curitiba</i>	52,48	57,28	104,96	114,56
<i>Porto Alegre</i>	72,38	75,70	144,76	151,39
<i>Sul Urbano</i>	47,18	50,48	94,36	100,97
<i>Sul Rural</i>	44,71	45,81	78,25	80,17
<i>Distrito Federal</i>	75,22	84,92	150,45	169,83
<i>Centro-Oeste Urbano</i>	44,69	50,61	89,38	101,22
<i>Centro-Oeste Rural</i>	40,79	42,37	71,38	74,15

Fonte: microdados da POF 2002-03

Parece evidente que as pressões no orçamento dos gastos em habitação e transporte nas maiores cidades e a importância da alimentação fora se refletem no fato do estrato populacional de referência na Região Metropolitana de São Paulo – aquele que atinge a disponibilidade calórica –apresentar renda domiciliar *per capita* superior a R\$ 700 para um gasto alimentar – linha de indigência - de R\$ 103 com a caloria fora custando o dobro da domiciliar. Mesmo quando se consideram iguais os preços da caloria, a relação entre gastos alimentares e renda não ultrapassa a 1/3 no Rio de Janeiro, no Distrito Federal e em São Paulo, onde é, inclusive, menor que 20%. Não resta dúvida que estas proporções não se coadunam com a situação de indigência; pode-se dizer que pobres gastam no mínimo 40% com alimentação, participação essa sendo um dos divisores de águas na definição de pobreza.

Tabela 4

População, incidência e distribuição regional dos indigentes e pobres, segundo domínio geográfico e fator de preço de caloria fora

Domínio Geográfico	População				Incidência (% sobre população total)				Distribuição Regional (%)			
	Indigentes (Fator 1)	Indigentes (Fator 2)	Pobres (Fator 1)	Pobres (Fator 2)	Indigentes (Fator 1)	Indigentes (Fator 2)	Pobres (Fator 1)	Pobres (Fator 2)	Indigentes (Fator 1)	Indigentes (Fator 2)	Pobres (Fator 1)	Pobres (Fator 2)
RM de Belém	51.670	57.129	303.296	363.530	2,8	3,1	16,4	19,6	0,7	0,6	1,1	1,1
Norte Urbano Não Metropolitano	472.813	505.012	1.716.171	1.807.738	5,7	6,1	20,8	21,9	6,3	5,1	6,2	5,5
Norte Rural	217.234	229.178	781.437	809.291	6,1	6,5	22,1	22,9	2,9	2,3	2,8	2,5
Norte	741.716	791.319	2.800.905	2.980.559	5,4	5,8	20,5	21,8	9,9	8,0	10,2	9,1
RM de Fortaleza	224.761	231.487	710.761	757.693	7,4	7,6	23,5	25,0	3,0	2,3	2,6	2,3
RM de Recife	97.736	152.984	395.553	729.857	2,9	4,6	11,8	21,9	1,3	1,5	1,4	2,2
RM de Salvador	80.129	141.668	416.638	552.464	2,6	4,6	13,4	17,8	1,1	1,4	1,5	1,7
Nordeste Urbano Não Metropolitano	1.944.042	3.115.450	7.130.971	9.372.681	7,6	12,1	27,8	36,5	25,9	31,4	25,9	28,5
Nordeste Rural	1.600.324	1.803.625	4.446.906	4.728.344	11,5	12,9	31,9	33,9	21,3	18,2	16,1	14,4
Nordeste	3.946.992	5.445.213	13.100.829	16.141.038	8,0	11,1	26,7	32,9	52,5	54,9	47,5	49,0
RM de Belo Horizonte	49.556	49.556	257.075	306.109	1,1	1,1	5,8	6,9	0,7	0,5	0,9	0,9
MG+ES+RJ Urbano Não Metropolitano	812.649	887.172	2.409.054	2.570.805	4,8	5,3	14,3	15,2	10,8	8,9	8,7	7,8
MG+ES+RJ Rural	142.758	142.758	679.701	679.701	3,5	3,5	16,7	16,7	1,9	1,4	2,5	2,1
RM do Rio de Janeiro	357.941	532.924	1.341.419	1.636.057	3,2	4,8	12,1	14,7	4,8	5,4	4,9	5,0
RM de São Paulo	372.807	824.902	1.989.567	2.889.703	2,1	4,7	11,2	16,3	5,0	8,3	7,2	8,8
São Paulo Urbano Não Metropolitano	352.847	352.847	1.300.472	1.401.387	1,9	1,9	7,1	7,7	4,7	3,6	4,7	4,3
São Paulo Rural	5.916	5.916	69.297	69.297	0,2	0,2	2,8	2,8	0,1	0,1	0,3	0,2
Sudeste	2.094.473	2.796.075	8.046.585	9.553.059	2,8	3,7	10,7	12,7	27,9	28,2	29,2	29,0
Rm de Curitiba	33.640	33.640	103.889	123.979	1,3	1,3	3,9	4,7	0,4	0,3	0,4	0,4
RM de Porto Alegre	53.501	66.294	415.487	419.960	1,5	1,8	11,3	11,4	0,7	0,7	1,5	1,3
Sul Urbano Não Metropolitano	229.058	267.884	1.234.132	1.438.563	1,5	1,8	8,2	9,5	3,0	2,7	4,5	4,4
Sul Rural	25.633	25.633	281.788	293.142	0,6	0,6	6,3	6,6	0,3	0,3	1,0	0,9
Sul	341.832	393.451	2.035.296	2.275.644	1,3	1,5	7,9	8,8	4,6	4,0	7,4	6,9
Distrito Federal	87.705	113.531	362.057	432.043	4,0	5,2	16,7	19,9	1,2	1,1	1,3	1,3
Centro-Oeste Urbano, exclusive DF	242.286	314.504	1.073.468	1.377.082	2,8	3,6	12,4	15,9	3,2	3,2	3,9	4,2
Centro-Oeste Rural	56.735	59.821	139.516	150.587	4,1	4,3	10,1	10,9	0,8	0,6	0,5	0,5
Centro-Oeste	386.726	487.856	1.575.041	1.959.713	3,2	4,0	12,9	16,0	5,1	4,9	5,7	6,0
Brasil	7.511.739	9.913.914	27.558.656	32.910.014	4,3	5,6	15,7	18,7	100,0	100,0	100,0	100,0

Fonte: Elaboração Própria.

Verdade que, antes de se afastar da definição da indigência por meio da relação consumo alimentar e renda *per capita*, sustenta-se que parte disso se deve as dificuldades na captação do consumo alimentar, especialmente, nas regiões metropolitanas. Duas hipóteses para esse vazamento da disponibilidade calórica: a dificuldade de captar e valorar os consumos institucionais e das redes de proteção social, mesmo depois da investigação dos gastos não monetários; e o custo da caloria ser menor que o domiciliar, já que parcela expressiva dos serviços de alimentação extra-domiciliar para os estratos pobres e intermediários das regiões metropolitanas se caracterizam pela abundância calórica. Exemplar disso são os carrinhos de cachorro quente e sanduíches, as grandes redes de "fast-food", a alimentação a peso e os variados "self serv". E outro importante "item" da alimentação fora, a institucional na escola e no trabalho, não parece apresentar um custo calórico superior ao domiciliar. Isso sem afastar a hipótese de que ocorre uma emulação pelos estratos de menor renda do consumo dos mais ricos, o que encarece o consumo domiciliar – em termos de disponibilidade calórica.

Assim acredita-se que os valores estimados são condizentes com as situações que se deseja caracterizar, apesar dessas dificuldades, ou seja, os valores para delimitar a pobreza e a indigência, a incidência por região e situação dos domicílios e a distribuição regional são bastante próximos daquilo que maior parte dos estudos e indicadores aponta (Tabela 4). Metade das populações pobres se encontram no Nordeste, onde as proporções de indigentes e pobres atingem cerca de 10% e 30%, respectivamente. Os contextos geográficos que se destacam negativamente, com expressivas participações de populações pobres são, em primeiro plano, o Nordeste Urbano não metropolitano e o Nordeste Rural. Com valores ao redor ou maiores que 20% encontram-se todos os do Norte, as RMs Nordestinas e o Distrito Federal. A distribuição regional não se altera entre as populações indigentes e pobres e para os diferentes fatores de cálculo do preço caloria fora. Alguma mudança se observa nas regiões Sul e Nordeste, preferencialmente, no meio urbano, com a participação no primeiro crescendo e no segundo caindo (Tabela 4).

5 Matching POF x Censo

Estimados os valores das linhas de indigência e de pobreza, com base nos dados de consumo alimentar da POF, o próximo passo consiste em obter estimativas discriminadas geograficamente e ao longo do tempo das populações indigentes e pobres. Tendo em vista que a POF não conta com uma discriminação espacial e tem sua realização restrita temporalmente, emprega-se, comumente, a Pnad, para o acompanhamento da pobreza ao longo do tempo, e do Censo, no caso da espacialização da indigência e da pobreza.

Como a captação da renda é diferente em cada uma das distintas pesquisas, torna-se necessário compatibilizar e ajustar as rendas. Efetivamente, de um lado, na Pnad e no Censo não são considerados os recebimentos não monetários, dentre os quais sobressai-se o valor da habitação dos imóveis próprios e/ou cedidos e a produção para autoconsumo. De outro lado, sabe-se que as investigações domiciliares incorrem em subdeclaração dos rendimentos, particularmente nos estratos superiores de renda. Faz-se necessário então tratar as rendas apuradas nessas investigações, tanto no sentido de considerar o aluguel estimado, quanto as rendas não monetárias e as subdeclarações de renda.

No caso do aluguel estimado, existem duas alternativas de tratamento no caso das Pnads: a primeira consiste em descontar o aluguel da renda daquelas famílias que efetivamente pagam aluguel, ou alternativamente acrescer o aluguel estimado às rendas das famílias proprietárias ou cedentes. O argumento subjacente do primeiro procedimento é de se aproximar, ao

máximo, da renda disponível, ou seja, dos rendimentos que as famílias dispõem para consumir. Tal procedimento é, entretanto, pouco recomendável pelas dificuldades que as Pnads tem em captar outros rendimentos provenientes do não-trabalho, além das sub-declarações¹³, que tornam difícil a tarefa de atingir a renda disponível.

Para a espacialização das estimativas das populações pobres e indigentes aplicamos técnicas de associação estatística – *matching* – entre a POF e a Amostra do Censo Demográfico 2000. A partir da identificação das famílias pobres na POF, é realizado, em um primeiro momento, o mapeamento das características demográficas e socioeconômicas dessas famílias, associando-as as situações de indigência e pobreza. Parte-se do pressuposto de que essas características determinam as situações de indigência e pobreza, o que permitirá, em um segundo momento, identificar, com base nesses indicadores, tais populações no Censo. Evitar-se-á, assim, a realização de ajustes, em grande medida, arbitrários, às rendas no Censo para cotejá-las com as linhas de indigência e pobreza.

5.1 A metodologia do *matching*

O modelo empregado foi do tipo logit, de variáveis dependentes binárias (no caso, indigente ou não, pobre ou não), para caracterizar as famílias indigentes (vide, por exemplo, Greene, 1993, Wooldridge, 2002). O modelo a ser estimado será então

$$\text{Pr}[\text{indivíduo ser pobre}] = h(x_1, x_2, x_3, \dots, x_p), \quad (8)$$

onde a probabilidade $\text{Pr}[\text{indivíduo ser pobre}]$ do indivíduo estar no grupo dos pobres (ou indigentes) dependerá das covariáveis x_1, x_2, \dots, x_p . Esse modelo será utilizado em uma segunda etapa para encontrar os indivíduos indigentes nos dados do Censo 2000. Portanto, na escolha das covariáveis x_1, x_2, \dots, x_p , devemos nos restringir ao grupo de variáveis que estão disponíveis também nos micro-dados do Censo 2000. O objetivo de encontrar os indivíduos pobres no Censo 2000 é que a partir do modelo de probabilidade de ser pobre, de acordo com a equação acima, encontraremos quantos indivíduos pobres, em média, encontram-se em cada município brasileiro.

O algoritmo para mapear o número de pobres (indigentes) por município pode ser descrito como se segue:

1. Considere um município específico j , nos dados do Censo 2000;
2. Considere o indivíduo i no município j , com covariáveis (também pelos micro dados do Censo) x_1, x_2, \dots, x_p ;
3. Utilizando o modelo estatístico representado equação (2), nós estimamos a probabilidade $prob_i$ do indivíduo i ser indigente;
4. Repetimos os passos 2 e 3 acima para todos os indivíduos do município j , e encontramos todos os valores $prob_i$;
5. Somamos todos os valores $prob_i$ para todos os indivíduos do município j de forma a termos uma primeira estimativa num_j do número de indigentes no município j ;
6. Repetimos os passos 1 a 5 para todos os municípios j do Brasil, e obtemos as diversas estimativas num_j de cada um deles;

¹³ A CEPAL realiza um procedimento de “ajuste”, para os dois décimos de renda mais elevada, das rendas das PNADs com as Contas Nacionais, tomando como pressuposto que a sub-declaração ocorre apenas nos estratos mais altos. Este procedimento foi adotado quando da divulgação das estimativas da Comissão IPEA-Cepal-IBGE, mas não muita repercussão por considerar apenas os 20% mais ricos, entre os quais o grupo de referência para estimação das linhas de indigência e de pobreza nunca se encontrariam.

7. É provável que ao somarmos todos os valores num_j de todos os municípios, o cômputo geral de indigentes no Brasil, ou por sub-regiões (como, por exemplo, por macro-região brasileira ou por unidade da federação), seja diferente dos números estimados diretamente da POF. Portanto, nessa etapa, precisamos fazer um ajuste dos valores de forma a compatibilizar as estimativas ao nível de municípios. Para isso, reescalamos linearmente o número as estimativas num_j para cada município de forma a termos o total de indigentes por região igual ao total estimado diretamente a partir da POF. Essas novas estimativas serão representadas como num_j^* ;

8. Finalmente, os números estimados num_j^* poderão então ser mapeados para obtermos representações geo-referenciadas do número de indigentes nos municípios brasileiros.

É aí que reside o que consideramos maiores problemas na discriminação da pobreza espacial e temporalmente, pois se fazem necessários ajustes nas outras pesquisas – PNAD e Censo. Para não incorrer nas arbitrariedades da estimação das rendas não monetárias e em ajustes nas rendas monetárias, decidiu-se aplicar uma metodologia de associação – *matching* – entre POF e Censo. Assim, empregaram-se as rendas monetárias juntamente com outras variáveis socioeconômicas e demográficas.

Transformaram-se todas as variáveis consideradas de modo a serem indicadores domiciliares – ou familiares¹⁴. Assim, variáveis relacionadas às pessoas do domicílio – chefe, crianças e idosos, renda – foram agrupadas no mesmo arquivo com as características dos domicílios - famílias. É interessante observar que para a maioria das variáveis tanto a média como o desvio padrão são bastante semelhantes, o que mostra que as variáveis aqui utilizadas tem um aderência. As diferenças podem ser explicadas em função do tamanho da amostra do Censo Demográfico, muito maior, cujo resultado é uma variância maior. No caso das rendas, ficam patente as diferenças na renda do trabalho e nas aposentadorias, com a renda média monetária mensal familiar per capita, na POF superando em 25% a do Censo, o que pode ser creditado a inflação do período.

Na tabela 5 encontram-se as principais estatísticas relativas ao *matching* para a pobreza. Como se pode observar os resultados são bastante robustos, com os percentuais de emparelhamento situando-se ao redor de 90% para todos os domínios geográficos considerados. As principais estatísticas relacionadas ao processo, quais sejam o Somer'D, Gamma e Tau-a mostram-se, em todos os casos, significativas, demonstrando a robustez do processo de associação entre a POF e o Censo.

14 A POF considera como família o grupo de indivíduos que compartilham as despesas alimentares e de higiene domiciliar, nomeando-a, também, como unidade de consumo. Assim, difere do tratamento dado pela PNAD e Censo, onde família tem por base as relações de parentesco. Essa conceituação da POF se assemelha a de domicílio no Censo e na PNAD. Cabe, contudo, ter claro que na POF domicílio e unidade de consumo são conceitos diferentes, pois pode haver uma ou mais unidades de consumo no mesmo domicílio. Para maiores detalhes, ver IBGE – citar doc. POF onde estão os conceitos.

Tabela 5
Principais estatísticas do *matching* para pobres

Domínio Geográfico	Pares concordantes (%)	Somers'D	Gamma	Tau-a
<i>RM de Belém</i>	99,7	0,994	0,994	0,200
<i>Norte Urbano Não Metropolitano</i>	98,4	0,968	0,969	0,242
<i>Norte Rural</i>	95,1	0,902	0,903	0,232
<i>RM de Fortaleza</i>	99,1	0,983	0,984	0,282
<i>RM de Recife</i>	98,8	0,976	0,977	0,255
<i>RM de Salvador</i>	98,8	0,976	0,976	0,202
<i>Nordeste Urbano Não Metropolitano</i>	98,4	0,969	0,970	0,390
<i>Nordeste Rural</i>	96,5	0,931	0,932	0,349
<i>RM de Belo Horizonte</i>	99,3	0,987	0,987	0,090
<i>MG+ES+RJ Urbano Não Metropolitano</i>	98,0	0,961	0,962	0,175
<i>MG+ES+RJ Rural</i>	95,2	0,905	0,906	0,175
<i>RM do Rio de Janeiro</i>	98,8	0,975	0,976	0,158
<i>RM de São Paulo</i>	99,3	0,986	0,986	0,197
<i>São Paulo Urbano Não Metropolitano</i>	99,2	0,984	0,984	0,083
<i>São Paulo Rural</i>	99,4	0,988	0,988	0,045
<i>Rm de Curitiba</i>	99,8	0,996	0,996	0,062
<i>RM de Porto Alegre</i>	99,0	0,980	0,980	0,127
<i>Sul Urbano Não Metropolitano</i>	99,0	0,981	0,981	0,106
<i>Sul Rural</i>	97,5	0,951	0,951	0,075
<i>Distrito Federal</i>	98,6	0,973	0,973	0,225
<i>Centro-Oeste Urbano, exclusive DF</i>	98,5	0,969	0,970	0,190
<i>Centro-Oeste Rural\$</i>	98,2	0,964	0,965	0,147

Fonte: Elaboração Própria.

6 Espacialização das populações indigentes e pobres

Os dados espacializados da incidência da indigência e da pobreza encontram-se ilustrados, respectivamente, nos mapas da figura 2, evidenciando-se a concentração da indigência na região Nordeste, tendo, com uma relativa presença no Norte de Minas e nos estados do Norte, especialmente no Amazonas e no Acre. Já no Centro-Sul observam-se os menores níveis de incidência da indigência, ainda que, no Mato Grosso, no Mato Grosso do Sul e Norte de Goiás observem-se municípios com a proporção de indigentes na faixa de 5 a 10%. Vale ter presente que os indicadores relativos de indigência, ou seja, de proporção de indigentes sobre a população total obscurecem o fato de que os números absolutos são de maior envergadura nas regiões mais populosas. Nesse sentido, como se pode observa nos dados da tabela 4, 1/3 dos indigentes se encontram nas regiões Sul e Sudeste.

Figura 2
Mapas da proporção de indigentes na população, por municípios

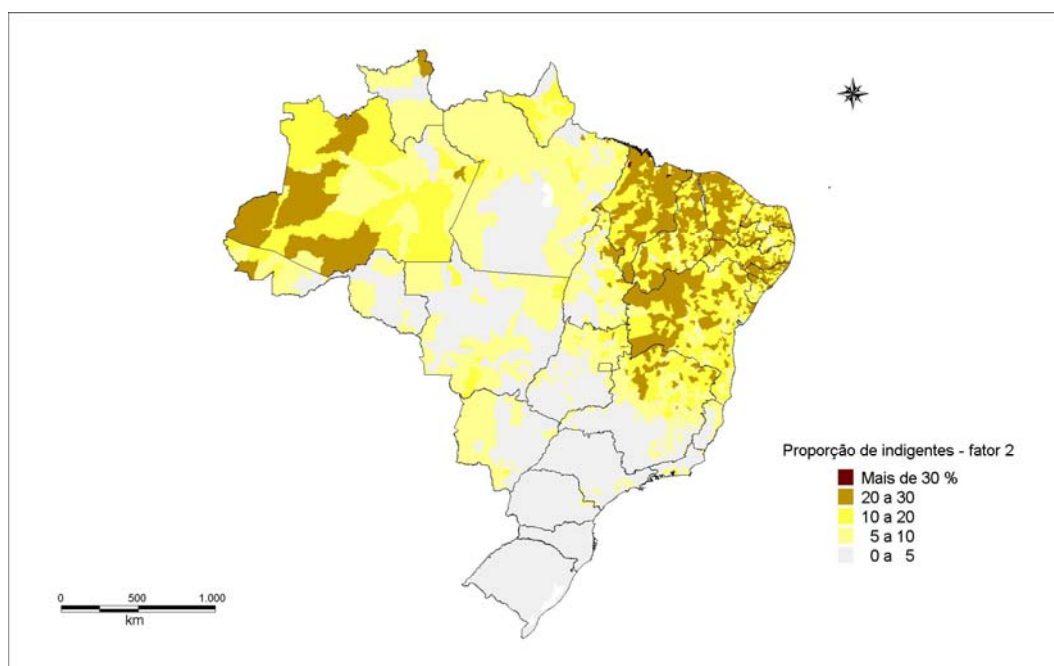
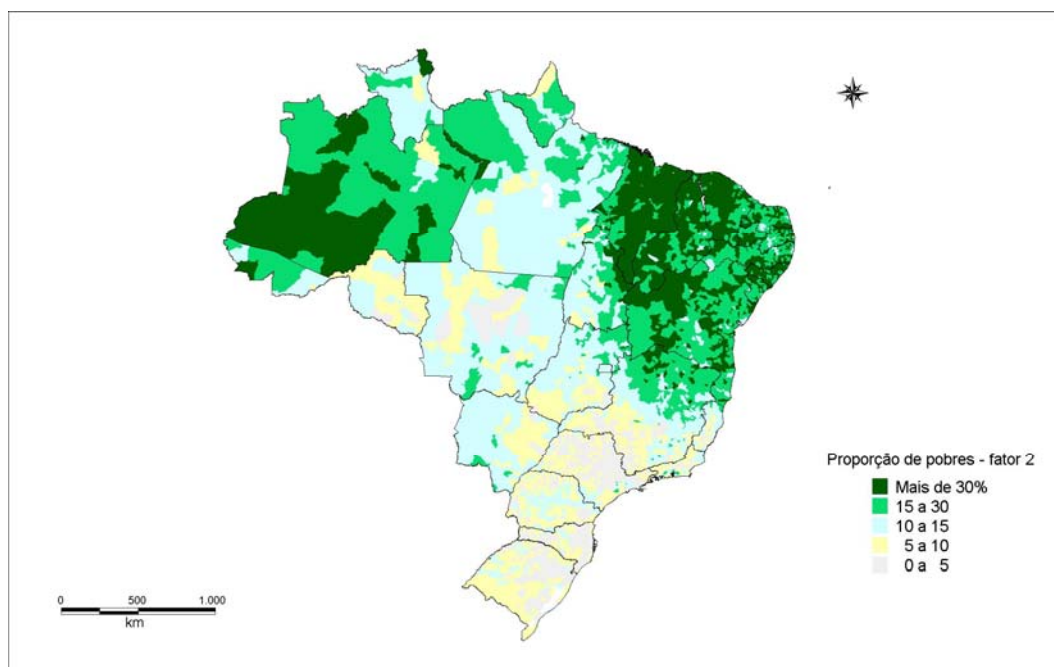


Figura 3
Mapas da proporção de pobres na população, por municípios



Os resultados da espacialização da pobreza são bem consistentes com aquilo que a maior parte dos estudos e análises apontam. Há uma concentração da pobreza nos municípios das regiões Nordeste e Norte, abarcando também a parcela norte do estado de Minas Gerais, região que faz parte da área de atuação da Sudene. Fica evidente que se pode traçar uma linha imaginária cortando o país no sentido de leste para oeste, que dividiria o país em duas

metades, uma, ao norte, com maior concentração de pobreza e outra, ao sul, com menor concentração de pobreza.

Analisando o mapa 2, além de reforçar os elevados percentuais de pobres nas regiões Nordeste e Norte, com destaque para os casos do Maranhão, do Piauí e do Amazonas, verifica-se que nos estados do Rio Grande do Sul, de Santa Catarina e de São Paulo apresentam o maior número de municípios com baixa incidência de pobres, chamando a atenção, particularmente, as áreas de concentração da pequena propriedade familiar – Vale do Itajaí e Serra Gaúcha. De outra parte, chama a atenção a proporção de pobres no Vale do Ribeira e no Pontal do Paranapanema, em São Paulo, no Centro Paranaense e na Fronteira Oeste do Rio Grande do Sul. Esses territórios se caracterizam por apresentar menor dinamismo econômico e baixos indicadores sociais. No caso do Amazonas, deve-se ter cuidado em função do tamanho dos municípios, que por serem bastante extensos fornecem uma falsa impressão de concentração de populações pobres. Vale dizer que os mapas podem esconder o fato que o número absoluto de pobres no estado de São Paulo e, de modo geral, na região Sudeste, ultrapassam aos 3 e 8 milhões de pessoas, respectivamente. Isso implica em uma participação da região sudeste no total de pobres do país da ordem de 1/3 (Tabela 4).

7 Referências bibliográficas

- ARIAS, Alfonso. *Proposta sobre a utilização do método da renda na preparação das medições de indigência e pobreza baseadas em linhas preparadas através da POF 2002-2003*. Brasília, IPEA, dezembro 2003. mimeo (publicação restrita).
- BARRRETTO, S. A. J. & CYRILLO, D. C. Análise da composição dos gastos com alimentação no município de São Paulo (Brasil) na década de 1990. *Revista de Saúde Pública*, São Paulo, v. 35, n. 1, 2001.
- BURNHAN, K. e ANDERSON, D. *Model Selection and Inference. A Practical Information-Theoretic Approach*. Springer, 1998.
- CEPAL. *Medición de la pobreza en Brasil: una estimación de las necesidades de energía y proteínas de la población*. Santiago, CEPAL, 1996.
- CHEN, X. e CONLEY, T. A New Semiparametric Spatial Model for Panel Time Series. *Journal of Econometrics*. No. 105, pp 59—83, 2001.
- COCHRAN, W. *Sampling Techniques*. Wiley, 1977.
- DAVISON, A. e HINKLEY, D. *Bootstrap Methods and their Applications*. Cambridge University Press, 1997.
- FAO/WHO/UNU. *Necessidades de Energia e Proteínas*. Genebra, FAO/WHO, 1985. *Série Informes Técnicos n° 724*.
- GREENE, W. *Econometric Analysis*. Prentice Hall, 1993.
- HALL, P. *The Bootstrap and the Edgeworth Expansion*. Springer, 1992.
- HASTIE, T., TIBSHIRANI, R. e FRIEDMAN, J. *The Elements of Statistical Learning. Data Mining Inference and Prediction*. Springer, 2001.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Pesquisa de Orçamentos Familiares 2002-2003: primeiros resultados – Brasil e Grandes Regiões*. 2. edição, Rio de Janeiro, 2004.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Pesquisa de Orçamentos Familiares 2002-2003: análise da disponibilidade domiciliar de alimentos e do estado nutricional no Brasil*. Rio de Janeiro, 2004.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Segurança alimentar: 2004*. Rio de Janeiro, 2006.

- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Pesquisa de Orçamentos Familiares 2002-2003: antropometria e análise do estado nutricional de crianças e adolescentes no Brasil*. Rio de Janeiro, 2006.
- IPEA. *Contribuições para a Construção de Linhas de Indigência para o Brasil*. Brasília, 06 de outubro de 1997, mimeo da Comissão Mista IBGE/CEPAL/IPEA
- LEHMANN, E. *Elements of Large-Sample Theory*. Springer, 1999.
- LEITENSTORFER, F. e TUTZ, G. Generalized Monotonic Regression Base don B-Splines with an Application to Air Pollution Data. *Ludwig-Maximilians-Universitat*. Discussion Paper n. 444. Maio 2005.
- LUSTOSA, Tânia. *Cálculo de las necesidades energéticas de la población brasileña para la construcción de una línea de pobreza*. Santiago, CEPAL, 1999 (apresentado no 4º Taller Regional del MECOVI).
- PAES DE BARROS, R. Et al. A estabilidade inaceitável: desigualdade e pobreza no Brasil. In: HENRIQUES, R. (org.) *Desigualdade e pobreza no Brasil*. Rio de Janeiro: IPEA, 2000.
- ROCHA, Sônia. “Medindo a Pobreza no Brasil: evolução metodológica e requisitos de informação básica” in Lisboa, Marcos & Menezes Filho, Naércio (Orgs.). *Microeconomia e Sociedade no Brasil*. Rio de Janeiro, Contracapa, 2001.
- ROCHA, Sônia. *Pobreza no Brasil: afinal do que se trata?* Rio de Janeiro: Editora FGV, 2003.
- TUTZ, G. e LEITENSTORFER, F. Generalized Smooth Monotonic Regression. *Ludwig-Maximilians-Universitat*. Discussion Paper n. 417. Março 2005.
- WOOLDRIDGE, J. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 2002.