

Texto para Discussão

Série Economia

TD-E 14 / 2009

**DIMENSÃO, MAGNITUDE E LOCALIZAÇÃO DAS POPULAÇÕES POBRES
NO BRASIL**

Fernando Gaiger Silveira
Alexandre Xavier Ywata Carvalho
Carlos Roberto Azzoni
Bernardo Campolina
Antonio Ibarra

Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade
de Ribeirão Preto

Reitora da Universidade de São Paulo
Suely Vilela

Diretor da FEA-RP/USP
Rudinei Toneto Junior

Chefe do Departamento de Administração
André Lucirton Costa

Chefe do Departamento de Contabilidade
Adriana Maria Procópio de Araújo

Chefe do Departamento de Economia
Walter Belluzzo Junior

CONSELHO EDITORIAL

Comissão de Pesquisa da FEA-RP/USP

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Avenida dos Bandeirantes, 3900
14049-900 Ribeirão Preto - SP

A série TEXTO PARA DISCUSSÃO tem como objetivo divulgar: i) resultados de trabalhos em desenvolvimento na FEA-RP/USP; ii) trabalhos de pesquisadores de outras instituições considerados de relevância dadas as linhas de pesquisa da instituição. A série foi subdividida em função das principais áreas de atuação da FEA-RP/USP: Economia, Administração e Contabilidade. Veja o site da CPq na Home Page da FEA-RP: www.fearp.usp.br. Informações: e-mail: cpq@fearp.usp.br

TEXTO PARA DISCUSSÃO Nº 1278

DIMENSÃO, MAGNITUDE E LOCALIZAÇÃO DAS POPULAÇÕES POBRES NO BRASIL*

Fernando Gaiger Silveira^{}**
Alexandre Xavier Ywata Carvalho^{}**
Carlos Roberto Azzoni^{*}**
Bernardo Campolina^{**}**
Antonio Ibarra^{***}**

Brasília, maio de 2007

* Este trabalho foi desenvolvido na Fundação Instituto de Pesquisas Econômicas (Fipe), com o apoio do Núcleo de Estudos Agrários e Desenvolvimento Rural (Neado) e do Instituto Interamericano de Cooperação para a Agricultura (Iica).

** Alexandre Xavier Ywata de Carvalho é Coordenador de Estudos Espaciais da Diretoria de Estudos Regionais, Urbanos e Fiscais do Ipea. Fernando Gaiger Silveira é Pesquisador da Diretoria de Estudos Setoriais do Ipea.

*** Professor da FEA – USP.

**** Economista, Doutor em Geografia Humana – USP e Pesquisador da Fipe.

***** Sociólogo, Consultor da Fipe.

SUMÁRIO

SINOPSE

1	INTRODUÇÃO	7
2	BASE DE DADOS	8
3	METODOLOGIAS DE ESTIMAÇÃO DAS POPULAÇÕES RURAIS INDIGENTES (EXTREMAS POBREZA) E POBRES	9
4	RESULTADOS	23
5	<i>MATCHING</i> POF <i>VERSUS</i> CENSO	27
6	ESPACIALIZAÇÃO DAS POPULAÇÕES INDIGENTES E POBRES	30
	REFERÊNCIAS	39
	ANEXO	41

SINOPSE

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 à fome e à pobreza. São diversas as perguntas que têm sido feitas: 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 produz, em geral, grandes distorções regionais. Usando informações da Pesquisa de Orçamento Familiar (POF) do IBGE, pesquisa que cobre regiões urbanas, metropolitanas e rurais do país, este artigo desenvolve e aplica 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 último, é feito um *matching* com dados do Censo, de modo a permitir efetuar a espacialização da pobreza nos municípios.

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.

1 INTRODUÇÃO

A importância do tema da pobreza no Brasil e, mais especificamente, da identificação das populações em situação de risco alimentar, pode ser medida pela repercussão que teve a implantação do Programa Fome Zero e os debates que este ensejou. Tal processo levou o governo federal a alterar a abordagem da questão, unificando os programas de transferência de renda e de combate à fome no Bolsa Família. Em junho último, o Bolsa Família atendeu a 11,2 milhões de famílias, em 5.560 municípios, totalizando R\$ 682 milhões. Somam-se a isso os benefícios concedidos pelo Bolsa Escola, pelo Bolsa Alimentação, pelo Cartão Alimentação e pelo Auxílio-gás que, juntos, atingiram em junho último 1,5 milhão de famílias, com desembolso total da ordem de R\$ 25 milhões.

Pode-se considerar que existem duas maneiras de mensurar as populações em situação de pobreza extrema, ou seja, aquelas pessoas que se encontram em situação de insegurança alimentar. Em uma delas associa-se a pobreza diretamente à fome e à desnutrição, utilizando-se os indicadores de prevalência de desnutrição infantil e de desnutrição crônica. Tais indicadores são calculados com base nas medidas antropométricas das populações de crianças menores de cinco anos e de adultos maiores de 18 anos. Efetivamente, verifica-se, no primeiro caso, o quanto a participação de crianças com baixa estatura supera o que é considerado normal. No segundo, de modo similar, estima-se a população com deficiência energética crônica com base na proporção de indivíduos magros (índice de massa corporal inferior a 18 kg/m^2), que em populações em que sabidamente inexistente o problema da fome esta se situa ao redor de 5%. Segundo dados de Paes de Barros *et al.* (2000), enquanto a população indigente no Brasil era, em 1999, da ordem de 14,5%, as estimativas de subnutrição infantil situavam-se ao redor de 6%. Os dados antropométricos da população adulta, coletados pela Pesquisa de Orçamentos Familiares (POF) de 2002-03, não apontaram a prevalência de *déficits* de peso – participações superiores a 5% de indivíduos magros – no país, salvo para as mulheres, especialmente aquelas residentes no meio rural nordestino (7,2%). Em junho último foram divulgados os dados da POF 2002-03 referentes à antropometria das crianças e adolescentes, com a prevalência de *déficit* de peso segundo a idade das crianças menores de cinco anos ficando em 4,6%; a região Norte, na qual este indicador atingiu 8%, destoa do restante do país.

Outra forma de dimensionar as populações indigentes, ou seja, de determinar o montante de famílias com dificuldades de garantir a alimentação cotidiana, é estabelecer o valor da renda que permita atender às necessidades básicas das famílias, particularmente as alimentares. Neste caso, podem ser utilizadas 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 US\$ 1 ajustado ao poder de compra local do Banco Mundial; de outro, as que estimam esses valores limites com base nos dados do consumo alimentar observado. Esta é a abordagem utilizada no presente estudo. Ambas as 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 aquelas de maior amplitude territorial e temporal.

Assim, no presente estudo consideram-se como de extrema pobreza e pobreza as famílias que se encontram nas 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 educação¹ da família.²

Dimensionar e apontar a concentração espacial da indigência e da pobreza são os principais objetivos do presente trabalho. Para tanto, emprega-se 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, cobrir as necessidades básicas. Tendo em conta que a definição das linhas de indigência e pobreza bem como das populações nessas situações é realizada com base na POF, cuja subenumeração das rendas é bem menor do que aquela incorrida em outras investigações domiciliares, decidiu-se mudar a forma de espacializar a pobreza, empregando, para tanto, técnicas de associação entre pesquisas com base em um conjunto mais amplo de indicadores socioeconômicos e demográficos. Ficam, ainda, pendentes os métodos para acompanhar o comportamento dos indicadores de pobreza e indigência, que não se restrinjam ao cotejamento entre as rendas e valores atualizados das linhas de indigência e pobreza.

O trabalho está organizado em cinco tópicos, além dessa introdução. O segundo trata da base de dados, apresentando sumariamente as principais características da POF. No terceiro tópico 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 apresentada. Cabe destacar que as sugestões da Comissão foram, em grande medida, incorporadas ao presente estudo. No quarto tópico encontram-se os resultados das estimativas de famílias indigentes e pobres na POF. Nos dois itens seguintes discute-se a metodologia de espacialização da indigência e pobreza e, por fim, apresentam-se os resultados do esforço de discriminação segundo domínio geográfico.

2 BASE DE DADOS

A fonte de dados domiciliares utilizada 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 procura 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, 2004b).

A coleta de dados da POF 2002-03 foi feita por meio de seis questionários, cobrindo, 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,

1. Não considerar os gastos em saúde e educação se deve ao fato de que cabe ao Estado fornecer, aos cidadãos, educação básica e saúde gratuita – este é um direito do cidadão e obrigação do Estado.

2. Como se pode notar, os conceitos de suficiência, de mínimo requerido e necessidades básicas são comumente utilizados para determinar as situações de indigência – extrema pobreza – e pobreza. Ora, tais conceitos são, muitas vezes, subjetivos, arbitrários, histórico-culturais e políticos, indicando que tais medidas não são definitivas e tampouco absolutas.

frequência à escola, peso, altura; b) dados referentes às despesas coletivas, ou seja, com serviços públicos, aluguéis, taxas, decoração, melhorias e reformas do domicílio, bens de consumo duráveis e serviços domésticos; c) caderneta de despesa coletiva, que engloba alimentação e higiene e limpeza; d) gastos individuais com 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, tais 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.

Vale destacar que a nova POF também levantou as despesas e recebimentos não-monetários, fato 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 a todas as camadas sociais. A produção própria e o trabalho da família no meio rural, e a ação das redes de proteção social, especialmente nas médias e grandes cidades, são exemplos da importância de se investigar essas 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 a 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 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, com vistas a verificar em que medida tais características estão correlacionadas a indigência e/ou a pobreza. Tendo sido selecionadas as variáveis determinantes por meio de um modelo logit, realiza-se, em seguida, o emparelhamento entre a POF e o Censo.

3.1 ESTÁGIO ATUAL DO CÁLCULO DA LINHA DE EXTREMA POBREZA – INDIGÊNCIA

Os requerimentos calóricos mínimos

Na metodologia de estimação de populações indigentes e pobres realiza-se, em um primeiro momento, a seleção de estratos populacionais que apresentem gastos alimentares cujo conteúdo calórico atenda às recomendações mínimas. Entre os

principais estudos sobre requerimentos calóricos necessários a uma alimentação adequada executados nas últimas décadas encontram-se os realizados pela FAO/WHO/UNU em 1974, e revistos em 1985, nos quais são *recomendadas* 2.236 quilocalorias diárias por pessoa. O documento não aborda a questão da diferença entre residentes em áreas rurais, urbanas não-metropolitanas e metropolitanas, em que atividades laborais específicas poderiam determinar distintas necessidades calóricas. Além disso, a alimentação tem um componente etário, uma vez que são bastante diversas as necessidades de energia de uma criança, de um adolescente e de um idoso. Infelizmente, o estudo supracitado recomenda a adoção de único valor para todas as faixas etárias.

Assim sendo, a adoção de medidas de equivalência no processo de estimação de linhas de indigência e pobreza é uma maneira de tratar essas diferenças alimentares segundo o sexo e a idade. Ademais, é relativamente consensual a utilização de escala de equivalência, haja vista o enorme benefício que tal procedimento traz para a realização de estimativas sobre a extrema pobreza. Seria errôneo pensar que cada indivíduo, independentemente da idade e sexo, tem as mesmas necessidades alimentares e calóricas. Neste sentido, decidiu-se utilizar a escala de diferenciação de necessidades energéticas proposta pela Cepal, na qual são discriminadas oito faixas etárias, bem como o sexo, com o homem adulto servindo de parâmetro para a faixa etária de 18 a 30 anos.

TABELA 1

Tabela de consumo energético – Cepal

Faixa etária	Masculino	Feminino
Abaixo de 1 ano	0.269	0.249
de 1 a 3 anos	0.494	0.461
de 4 a 6 anos	0.640	0.577
de 7 a 9 anos	0.736	0.650
de 10 a 13 anos	0.812	0.717
de 14 a 17 anos	0.974	0.762
de 18 a 30 anos	1.000	0.721
de 31 a 60 anos	1.001	0.739
Mais de 60 anos	0.843	0.678

Fonte: Rocha (2001).

A tabela acima tem servido como referência para alguns estudos. Contudo, cabe salientar que não se trata de escala de equivalência *stricto sensu*, mas da forma como as calorias consumidas se diferenciaram por faixa etária e sexo. A utilização de escala de equivalência, qualquer que ela seja, tem como efeito principal o aumento do total de calorias *per capita* e a conseqüente diminuição do grupo de referência, o que resulta em linhas de indigência mais baixas. Outro ponto que merece destaque no uso de escalas de equivalência é que elas se referem apenas aos componentes alimentares, ou seja, apenas às linhas de indigência. No caso das linhas de pobreza, tal escala não deve ser utilizada porque não existem razões para se acreditar que a diferenciação na distribuição das outras despesas segue este mesmo raciocínio.

O principal determinante na construção de linhas de indigência baseada em requerimentos energéticos está na composição da cesta de alimentos para atingir tais necessidades. Pode-se optar por construir linhas utilizando todos os alimentos adquiridos pelas famílias no grupo de referência ou pela construção de cestas de alimentos. O primeiro procedimento parece o mais recomendável por registrar o que

realmente foi adquirido, enquanto o segundo envolveria critérios a serem adotados para compor uma cesta qualquer. O argumento contrário à utilização de todos os produtos registrados nas POFs deriva da dificuldade de atualização de seus valores ao longo do tempo. A verdade é que a atualização dos valores por meio do INPC-Alimentação para as linhas de indigência e do INPC-Geral para as de pobreza parece não provocar grandes distorções, a despeito da mudança de hábitos de consumo da população brasileira.

A pesquisadora Sônia Rocha tem sugerido, em vários estudos, o emprego dos requerimentos quilocalóricos *mínimos necessários* (algo em torno de 1.700 quilocalorias/dia por pessoa), ajustando as quantidades dos produtos de maior representatividade na POF. Concretamente, em um primeiro momento, selecionam-se os produtos com maior representatividade no gasto e na disponibilidade energética, e que atendam a essa necessidade energética mínima. Em uma segunda etapa, excluem-se os alimentos “que representam ingestão inferior a uma caloria por dia”, ajustando-se, por fim, as quantidades de modo a garantir a ingestão calórica recomendada em consonância com o domínio geográfico. Este procedimento para compor uma “cesta ajustada” tem como principal fundamento “obter a cesta alimentar de menor custo e que permita o atendimento dessas necessidades” (energéticas médias de cada área de análise) (ROCHA, 2003, p. 54).³ Logo, este “ajustamento” faz que, partindo de número menor de produtos, sejam alteradas as quantidades para que as famílias atinjam o total de calorias recomendadas, tendo, no entanto, como propósito central construir uma cesta mais barata, conforme sugere a própria pesquisadora. Vale sublinhar que, ao escolher os produtos que irá consumir, a família está maximizando uma função-objetivo que expressa o seu conjunto de preferências de consumo, considerando-se a restrição orçamentária. No entanto, o ajuste para uma cesta de valor mais baixo, de forma a atingir a necessidade calórica, somente faria sentido caso a função-objetivo “ingestão calórica” coincidissem com aquela baseada no sistema de preferências. Uma consequência imediata é que se estaria assumindo, implicitamente, que todas as famílias (ou pelo menos as mais pobres) deveriam ter o mesmo sistema de preferências. Portanto, a metodologia baseada no ajuste da cesta de valor mínimo não parece fazer sentido do ponto de vista de maximização de utilidade.

A tabela 2 apresenta as recomendações calóricas mínimas (em Kcal/dia *per capita*) apontadas pela Cepal (1996) e as elaboradas por Tânia Lustosa (1999). Percebe-se que os valores estimados por Lustosa são sempre inferiores aos da Cepal, salvo no caso do Sul Rural. Porém, as diferenças são de pequena monta. Não obstante isto, decidiu-se empregar as recomendações da Cepal, por sua tradição nessa área de estudo e pelo fato de a adoção da proposta de Lustosa implicar o uso de menores populações indigentes e pobres.

3. Há de sublinhar os estudos desenvolvidos por pesquisadores da USP, referentes à adequação dietética do consumo/gasto alimentar com a aplicação de modelos de programação linear, e que objetivam verificar em que medida os recursos gastos em alimentação são suficientes para a aquisição de cestas alimentares nutricionalmente balanceadas. Sinteticamente, procuram construir cestas alimentares que atendam, simultaneamente, às necessidades nutricionais básicas, à restrição orçamentária e aos hábitos de consumo da população. Ver, para mais detalhes, Barretto e Cyrillo (2001).

TABELA 2

Recomendações calóricas por domínio geográfico (em Kcals/dia *per capita*)

Domínio geográfico	Cepal (1996)	Lustosa (1999)
Reg. Met. Belém	2.191	2.160
Norte Urb. Não-met.	2.191	2.125
Norte Rural	2.258	2.125
Reg. Met. Fortaleza	2.200	2.098
Reg. Met. Recife	2.200	2.126
Reg. Met. Salvador	2.200	2.127
Nordeste Urb. Não-met.	2.200	2.169
Nordeste Rural	2.207	2.142
Reg. Met. Rio de Janeiro	2.288	2.233
Rio de Janeiro Urb. Não-met.	2.288	2.246
Rio de Janeiro Rural	2.318	2.309
Reg. Met. São Paulo	2.288	2.233
São Paulo Urb. Não-met.	2.288	2.246
São Paulo Rural	2.318	2.309
Reg. Met. Belo Horizonte	2.288	2.233
MG+ES Urb. Não-met.	2.288	2.246
MG+ES Rural	2.318	2.309
Reg. Met. Curitiba	2.313	2.282
Reg. Met. Porto Alegre	2.313	2.284
Sul Urb. Não-met.	2.313	2.287
Sul Rural	2.400	2.408
Distrito Federal	2.259	2.198
Centro-Oeste Urb. Não-met.	2.259	2.220
Centro-Oeste Rural	2.328	2.229

Fonte: Elaborado a partir de Arias (2003).

O emprego de quintos móveis

Qual o grupo populacional que atinge o consumo calórico adequado com o menor gasto alimentar? A resposta a essa pergunta define o valor da linha de indigência. Uma das maneiras de se estimar os grupos populacionais para cada contexto é utilizar a metodologia de quintos móveis. A adoção desta metodologia procura equacionar algumas dificuldades decorrentes do formato da POF. Esta pesquisa investiga os gastos alimentares no e fora do domicílio por meio do acompanhamento *pari passu* das despesas durante uma semana. Isso é feito para evitar que ocorram vazamentos de informações, ou seja, dos gastos alimentares que, por serem recorrentes, são de difícil coleta quando se empregam métodos recordatórios. Ao se coletar os gastos alimentares no período de uma semana, os dados individuais – por família/domicílio – não serão representativos do consumo e da despesa alimentar, dado que não são contabilizados os estoques, nem são consideradas as particularidades do gasto efetuado naquela semana em função da época (mês e ano) em que a despesa é realizada. O desenho amostral da pesquisa é efetuado de maneira a fazer que sejam contrabalançadas as diferenças de aquisição alimentar entre famílias semelhantes, isto é, situadas nos mesmos estratos geográfico e de renda. Dito de outro modo, são coletadas informações tanto de famílias que nada compraram no período de referência – o que não significa que seus componentes não se alimentaram – como de outras famílias que realizaram compras para mais de uma semana. Cabe sublinhar, ademais, que como a pesquisa se desenvolve ao longo de um ano, a sazonalidade do consumo de determinados itens alimentares é também contemplada. Assim, fica evidente que não se pode tratar cada família separadamente, pois a investigação das aquisições alimentares está

relacionada a um *cluster* de famílias semelhantes tanto geograficamente como por níveis de renda.

São também adotados alguns pressupostos relacionados a essas particularidades da pesquisa. O primeiro é o de que as recomendações calóricas são atingidas via alimentos adquiridos, monetariamente ou não, na semana de referência, pois não se investiga quais produtos foram efetivamente consumidos. O segundo pressuposto é que as famílias comprem o que vão consumir, mesmo que não seja naquela semana em especial. Poder-se-ia argumentar que o consumo efetivo poderia ser pesquisado, mas isso implicaria utilizar outro questionário específico, onerando ainda mais a pesquisa e sujeitando os entrevistados a processos extremamente exaustivos, o que poderia elevar o risco de informações inconsistentes.

Tendo em vista as limitações para se trabalhar os dados alimentares de modo individual, ou melhor, por famílias, uma das soluções empregadas é a metodologia de quintos móveis, com a ordenação das famílias segundo sugestão da Comissão Mista Ipea/Cepal/IBGE. Efetivamente, o primeiro quinto móvel é composto das famílias entre os 0% e os 20% mais pobres na distribuição de renda, enquanto o segundo quinto engloba as famílias de 1% a 21%, e assim sucessivamente até perfazer total de 81 quintos móveis. Este procedimento metodológico garante que, em determinado grupo, existam, simultaneamente, famílias que não adquiriram alimentos na semana de referência e famílias que compraram alimentos para além da referida semana. Os passos atualmente adotados para o cálculo da linha de pobreza podem ser assim resumidos:

- 1) Construção de uma base de dados com todas as famílias na POF. Nessa base, constam as seguintes colunas, para cada família i : número de pessoas na família, em equivalentes adultos (n_i); total da renda familiar (r_i); total do consumo calórico (c_i) da família; dispêndio total da família com alimentação (d_i); número total de pessoas na família (m_i).
- 2) A partir da base de dados acima, calcula-se a renda familiar *per capita* (r_i/m_i) e o consumo calórico por equivalente adulto (c_i/n_i).
- 3) Ordenam-se as famílias pela renda *per capita* e dividem-se todas as famílias em 100 grupos, com idêntico número de famílias em cada grupo.
- 4) Para cada um desses 100 grupos (100 centésimos), calcula-se a média do consumo calórico por equivalente adulto para as famílias dentro do grupo, assim como a renda familiar média do grupo.
- 5) Monta-se um gráfico com essas 100 médias de consumo calórico por equivalente adulto, contra as 100 rendas médias de cada grupo de rendas crescentes.
- 6) É de se esperar que esse gráfico do consumo *per capita* médio não seja crescente com as 100 rendas, o que não garante a monotonicidade da curva. Uma solução para isso é fazer uma média móvel com 20 grupos (dos 100) em cada janela. Com isso, obtém-se suavização da linha de consumo calórico por equivalente adulto *versus* renda *per capita*. Esta suavização é, então, a estimativa do gráfico (ou curva) consumo calórico *versus* renda.

- 7) Com base em um valor de corte C_k para o consumo calórico por equivalente adulto (vide tabela 2), calcula-se o valor da renda *per capita* R_k de corte das famílias que conseguem consumir o valor mínimo de calorias, com base nos preços dos alimentos e dos demais itens de despesa familiar. Utiliza-se o subscrito k para o consumo calórico e para a renda de corte C_k e R_k , de forma a explicitar a dependência desses valores com o contexto geográfico específico k .
- 8) Finalmente, como último estágio na aferição de indigência e pobreza, calcula-se o valor da média D_k de despesa d_i familiar *per capita* com alimentação, correspondente às famílias com renda *per capita* em torno da renda de corte R_k . Este valor pode ser, então, considerado como valor de corte de renda *per capita* a ser usado na definição de pobres ou indigentes.
- 9) Resta, por fim, ajustar as rendas das diferentes pesquisas para contemplar os ganhos não-monetários, o custo da moradia própria e a subdeclaração dos rendimentos a fim de cotejar com os valores definidos como linhas de indigência e pobreza (D_k).

Vale observar que se dispõe de dois valores de corte para a renda família *per capita*, e que podem ser utilizados como linhas de renda para definir indigência e pobreza. A primeira definição, mais afeita ao conceito de indigência, é utilizar a linha de corte D_k , específica do contexto geográfico k (vide passo (8), acima). Portanto, de acordo com essa definição, todas as famílias que possuírem renda *per capita* abaixo de D_k estariam no grupo de indigentes. Esta é a definição proposta pela Comissão de Pobreza para a linha de indigência no Brasil, cabendo, então, para mensurar as populações em situação de indigência, cotejá-las com a renda das investigações domiciliares de maior amplitude temporal – Pnad – e espacial – Amostra do Censo. Neste caso, são feitos diversos ajustes nas rendas apuradas nessas pesquisas, a fim de contemplar os ganhos não-monetários, o custo de moradia própria e as subdeclarações de rendimentos. Em termos conceituais, essa definição de pobreza utilizando o valor D_k para a renda *per capita* de corte corresponde às famílias que, mesmo gastando toda a sua renda na aquisição de bens alimentares, ainda assim não conseguiriam consumir a quantidade mínima de calorias diárias.

A segunda definição, mais próxima do conceito de pobreza, considera o valor R_k como valor de corte para a renda *per capita* familiar, para o contexto geográfico k . Em geral, espera-se que esse valor seja maior do que o valor D_k . Portanto, se se utilizar essa segunda definição, tem-se número maior de pobres do que o calculado segundo a definição discutida no parágrafo anterior. Em termos conceituais, as famílias com renda *per capita* abaixo de R_k podem ser vistas como as famílias que não possuem rendimento familiar mínimo para acomodar, no seu orçamento, despesas alimentares que lhes permitam consumir os valores mínimos necessários de calorias. Pode-se argumentar que ao se adotar R_k como o valor de corte para renda familiar *per capita* automaticamente se está levando em conta todos os demais itens de despesas. Verificou-se que os resultados, com base nessa definição, não se mostram sustentáveis, uma vez que os diferenciais dos valores de corte entre contextos foram superiores a dez vezes, e nas regiões metropolitanas as rendas de corte foram excessivamente elevadas. Supõe-se que isso ocorra em razão de a maior segmentação do consumo alimentar, a maior participação da alimentação extradomiciliar e a expressiva

importância das redes públicas e privadas de assistência social dificultarem a captação da efetiva disponibilidade calórica.

O consumo fora do domicílio

A POF consegue registrar razoavelmente bem as categorias de alimentos que são compradas para consumo no domicílio, para os quais se atribui, com o emprego das tabelas de conversão, o valor calórico. No caso da parcela da alimentação realizada fora do domicílio, não se investiga a composição alimentar das refeições, ou seja, que produtos e em que quantidades são consumidos. Assim, uma das dificuldades encontradas na estimação das curvas de consumo calórico por equivalente adulto *versus* renda familiar *per capita* é não se ter idéia precisa do total de calorias consumidas fora da residência. Adicionalmente, como bem aponta a tabela 3, a alimentação fora de casa tem importância significativa, chegando a representar, em média, 1/3 dos gastos alimentares nos maiores centros urbanos do país. Como a POF registra o valor total em dinheiro gasto pela família nas refeições consumidas fora de casa, é possível estimar o consumo calórico fora do domicílio com base no valor da caloria consumida domesticamente. Empregaram-se, no presente trabalho, duas hipóteses de imputação da disponibilidade calórica com a alimentação fora de casa: a primeira considera que a caloria adquirida fora do domicílio representa o dobro do valor monetário da caloria dentro do domicílio, na suposição de que os custos e as margens de lucro do setor de alimentação fora de casa implicam preços por calorias superiores àqueles observados no domicílio; a segunda trata como idênticos os custos monetários das calorias dentro e fora de casa. Acredita-se que o valor real dessa relação (caloria fora *versus* caloria na residência) deve variar de acordo com o contexto geográfico, mas não deve ficar fora do intervalo entre 1 e 2.

TABELA 3

Gasto alimentar fora da residência por domínio geográfico

Domínio geográfico	Participação dos gastos alimentares na renda familiar	Participação dos gastos com alimentação fora do domicílio no total dos gastos alimentares
Brasil	17%	24%
Belém	22%	23%
Norte Urbano	22%	22%
Norte Rural	33%	9%
Fortaleza	17%	22%
Recife	20%	22%
Salvador	18%	31%
Nordeste Urbano	23%	20%
Nordeste Rural	41%	12%
Belo Horizonte	13%	30%
MG+ES+RJ Urbano	17%	25%
MG+ES+RJ Rural	28%	15%
Rio de Janeiro	14%	33%
São Paulo	12%	31%
São Paulo Urbano	15%	22%
São Paulo Rural	18%	19%
Curitiba	13%	32%
Porto Alegre	13%	28%
Sul Urbano	15%	23%
Sul Rural	26%	14%
Distrito Federal	10%	38%
Centro-Oeste Urbano	15%	22%
Centro-Oeste Rural	26%	12%

Fonte: Microdados da POF 2002-03.

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 discriminação espacial ampla, 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. De um lado, na Pnad e no Censo não são considerados, efetivamente, os recebimentos não-monetários, entre os quais merecem destaque o valor da habitação dos imóveis próprios e/ou cedidos e a produção para autoconsumo. De outro lado, tem-se consciência que as investigações domiciliares incorrem em subdeclaração dos rendimentos, particularmente nos estratos superiores de renda. Faz-se, então, necessário 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 referente às Pnads: a primeira consiste em descontar o aluguel da renda daquelas famílias que efetivamente pagam aluguel, ou, alternativamente, acrescentar o aluguel estimado às rendas das famílias proprietárias ou cedentes. O argumento subjacente ao primeiro procedimento é o de se aproximar, ao máximo, da renda disponível, ou seja, dos rendimentos de que as famílias dispõem para consumir. Tal procedimento é, entretanto, pouco recomendável pelas dificuldades que têm as Pnads de captar outros rendimentos provenientes do não-trabalho, além das subdeclarações,⁴ o que constitui obstáculo para a tarefa de se obter a renda disponível.⁵

3.2 NOVA METODOLOGIA PROPOSTA PARA O CÁLCULO DAS LINHAS DE INDIGÊNCIA E POBREZA

Apesar da metodologia atualmente aplicada, conforme descrito acima, ser intuitiva e de simples utilização, algumas críticas podem ser apontadas:

- 1) Não permite o cálculo de intervalos de confiança para as estimativas.
- 2) 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.
- 3) Mesmo com 20 grupos de renda na média móvel, isso ainda não garante monotonicidade na curva de consumo por equivalente adulto *versus* renda *per*

4. A Cepal realiza um procedimento de "ajuste", para os dois décimos de renda mais elevada, das rendas das Pnads com as Contas Nacionais, usando como pressuposto o fato de que a subdeclaração ocorre apenas nos estratos mais altos. Este procedimento foi adotado quando da divulgação das estimativas da Comissão Ipea/Cepal/IBGE, mas sua repercussão não foi muito grande pelo fato de considerar apenas os 20% mais ricos, entre os quais o grupo de referência para estimação das linhas de indigência e de pobreza, em que estes nunca se encontrariam.

5. Vale sublinhar que a incorporação da propriedade em ou cedência da moradia via desconto da renda do aluguel efetivamente pago tem muito pouco impacto na renda das famílias rurais, desconsiderando uma de suas retaguardas em termos de renda.

capita, o que aponta para a fragilidade das estimativas obtidas. Dada a grande variabilidade nas medições de consumo calórico, pode ocorrer que a curva estimada de acordo com a média móvel de 20 centésimos apresente trechos crescentes e trechos decrescentes, o que contrariaria a suposição de que maior renda implica maior consumo calórico.

- 4) Essa metodologia não se enquadra como estimador estatístico a partir de técnicas de regressão-padrão.
- 5) 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.
- 6) Por não se tratar, nitidamente, de metodologia desenvolvida via modelos de regressão, ela não permite a utilização de testes de hipóteses, intervalos de confiança e erros-padrão que levam em conta o desenho amostral da POF.
- 7) Finalmente, utilizam-se médias móveis com 20 centésimos para todos os contextos geográficos, indistintamente. Com isso, não se leva em conta a disponibilidade distinta 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 a 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, metodologia baseada em modelos de regressão não-paramétrica, ou modelos de regressão semiparamétrica para a função do consumo calórico *versus* a renda *per capita*.

Estimação não-paramétrica da curva consumo/renda

Nesta seção apresentam-se os modelos de estimação não-paramétrica, ou semiparamétrica, para estimar a curva de consumo calórico por equivalente adulto *versus* a renda *per capita*. O termo regressão semiparamétrica deve-se ao fato de se utilizar expansões aparentemente paramétricas para aproximar formas funcionais desconhecidas.⁶ Apesar de as estimações feitas neste trabalho terem sido realizadas por contexto geográfico, na exposição apresentada a seguir suprime-se o subscrito *k*, correspondente ao contexto *k*, para simplificar a notação.

A idéia da estimação não-paramétrica neste projeto consiste em estimar uma função do tipo:

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

em que, conforme apresentado acima, c_i é o consumo total da família, r_i é a renda total da família, n_i é o número de pessoas, em equivalente adulto, e m_i é o número

6. 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 se transforma em um problema de estimar um modelo paramétrico, em que as transformações nas variáveis explicativas são apropriadamente construídas. Neste caso, o modelo de regressão não-paramétrica pode ser denominado modelo de regressão semiparamétrica. Para mais detalhes, ver Hastie, Tibshirani e Friedman (2001).

total de indivíduos na família. O quociente r/m_i corresponde à renda familiar *per capita*, enquanto c/n_i representa o 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 $g(r/m_i)$ possui forma funcional desconhecida e deverá ser estimada a partir dos dados. Para estimá-la, emprega-se estimação semiparamétrica, utilizando expansão de funções-base (*basis functions*). Esta expansão baseia-se na aproximação da curva desconhecida $g(r/m_i)$, utilizando a forma paramétrica flexível:

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

em que $u_l(r_i / m_i)$ 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 os quais encontram-se, por exemplo, regressões de redes neurais e regressões de *wavelets* (vide Hastie, Tibshirani e Friedman, 2001). Neste trabalho, utiliza-se expansão do tipo *B-splines* de ordem q . Devido ao fato de as funções-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 $[x_{\min}, x_{\max}]$.⁷ Considere-se um vetor de w pontos $(x_1, x_2, \dots, x_w)$ dividindo o intervalo $[x_{\min}, x_{\max}]$, em que $x_{\min} < x_1 < x_2 < \dots < x_w < x_{\max}$. 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 modo a fazer que os polinômios utilizados possuam grau 2 ou 3, respectivamente.

Considere agora o vetor de nós $x_{\min}, \dots, x_{\min}, x_1, x_2, \dots, x_w, x_{\max}, \dots, x_{\max}$, em que 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:

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

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

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. Este número está diretamente relacionado ao número de pontos

7. Neste trabalho, a variável explicativa é a renda *per capita* r/m_i .

divisórios w , bem como à ordem q . Quanto maior o número L , maior a flexibilidade da expansão semiparamétrica. Uma vez fixado o número de funções-base L para estimar a função desconhecida $g(r/m_i)$, a estimação dos parâmetros b_l , $l = 0, 1, 2, \dots, L$ pode ser efetuada via estimadores paramétricos tradicionais. Podem ser utilizados, 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 $g(r/m_i)$ 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 BIC⁸ (regras de escolha *in-sample*), ou se utilizando alguma regra de validação cruzada (*cross-validation*), que corresponde a regras de escolha *out-of-sample*. Tais procedimentos têm por objetivo evitar sobreajuste (*overfitting*) do modelo semiparamétrico. De fato, quando se aumenta o número de funções-base indefinidamente, obtém-se modelo com ajuste perfeito dentro da amostra, mas com poder preditivo questionável para observações fora da massa de dados. Por outro lado, para um número L muito baixo, obtém-se baixa flexibilidade da expansão de funções-base, incorrendo-se também 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.⁹ Para mais detalhes em escolhas do grau de flexibilização em modelos semiparamétricos e não-paramétricos, ver Hastie, Tibshirani e Friedman (2001).

Observe-se que a estimação da função $g(r, n)$ pode ser efetuada diretamente nos microdados, a partir da POF. Portanto, não há necessidade de agrupamento prévio desses dados em centésimos, por exemplo, conforme efetuado na metodologia atualmente vigente para o 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 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 diversos para diferentes contextos geográficos, podem-se obter expansões semiparamétricas mais apropriadas para as particularidades do conjunto de informações disponível em cada região.

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 $g(r/m_i)$ seja monotonicamente crescente em r/m_i . Neste 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 $b_1 \leq b_2 \leq \dots \leq b_L$.

8. Para mais detalhes sobre o AIC e o BIC, ver Burnhan e Anderson (1998).

9. Na literatura em estatística esse problema é comumente conhecido como *trade-off* viés-variância (*bias-variance*).

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 este problema utilizou-se 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 o tratamento de observações discrepantes (*outliers*).

Seja Y um vetor contendo todas as observações empilhadas para a variável dependente no modelo de regressão linear. Neste trabalho, a variável dependente corresponde ao consumo calórico por equivalente adulto c/n_i . Portanto, o vetor Y corresponde a uma lista dos valores c/n_i para todas as N famílias i na amostra. Seja X a matriz com dimensão $N \times L$, em que a coluna j corresponde ao empilhamento de todos os N valores $u_{j,q}(r_i/m_i)$ para a função-base j em (6). Finalmente, seja β o vetor de dimensão $L \times 1$, composto pelos parâmetros desconhecidos $b_1, b_2, \dots, b_L$. Note-se que tanto Y quanto X são completamente observados, dado que se conhece a forma funcional das funções-base na expansão de *B-splines*.

O problema de estimação de β via mínimos quadrados ordinários pode ser escrito como problema de maximização quadrática. De fato, considere-se a soma dos quadrados dos erros

$$\begin{aligned} S(\beta) &= [Y - X\beta]^T \times [Y - X\beta] \\ &= Y^T Y - 2[Y^T X]\beta + \beta^T [X^T X]\beta, \end{aligned}$$

em que A^T corresponde à transposta da matriz (vetor) A . Portanto, a minimização da soma dos erros quadrados $S(\beta)$ corresponde à maximização da forma quadrática $2[Y^T X]\beta - \beta^T [X^T X]\beta$. Pode-se, então, incorporar a restrição de monotonicidade da expansão de *B-splines*, $b_1 \leq b_2 \leq \dots \leq b_L$, o que leva ao problema de maximização quadrática restrita

$$\begin{aligned} \max_{\beta} \quad & [Y^T X]\beta - \frac{1}{2} \beta^T [X^T X]\beta \\ \text{sujeito a} \quad & M\beta \geq z, \end{aligned}$$

em que a matriz M e o vetor z são dados por

$$M = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}_{(L-1) \times L}, \quad z = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{bmatrix}_{L-1}$$

O problema de maximização acima corresponde a um problema clássico de programação quadrática e pode ser resolvido com os algoritmos descritos em Winston (2003) ou Hillier e Lieberman (2004). O método de estimação baseado na maximização acima corresponde à estimação via método de mínimos quadrados restrito.

A sistemática de estimação via método de mínimos quadrados restrito pode ser modificada para acomodar observações com diferentes pesos. Nesse caso, tem-se a estimação via método de mínimos quadrados ponderados restrito, em que cada observação i é ponderada pelo peso w_i . Considere-se, então, a matriz diagonal W , em que os elementos da diagonal principal correspondem a todos os N valores w_i ao quadrado w_i^2 . A estimação via método de mínimos quadrados ponderados restrito corresponde à maximização

$$\max_{\beta} [Y^T W X] \beta - \frac{1}{2} \beta^T [X^T W X] \beta$$

sujeito a $M\beta \geq z$.

Os estimadores de mínimos quadrados restrito e mínimos quadrados ponderado restrito permitem obter estimadores 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/n_i , por equivalente adulto, correspondente a determinado valor de renda *per capita* r/m_i . Variando-se continuamente os valores de r/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 corte de renda *per capita* r_i/m_i para os diferentes contextos geográficos. Esta 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.¹⁰ 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 a reta horizontal passando pelo valor $c/n_i = C$. 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 “Estimação das linhas de indigência e pobreza” mais adiante, em que se replicaram as regressões semiparamétricas para estudar a relação entre gasto alimentar *per capita* e renda *per capita*. Chega-se, assim, com base nessas curvas, ao gasto de corte $\hat{g}_{corte}$, correspondente a $\hat{r}_{corte}$, gasto este 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, é possível determinar especificamente quais famílias, na amostra disponível (essa amostra será, na verdade, extraída a partir de filtros nos microdados da POF), encontram-se abaixo dessas linhas. Esta informação será particularmente importante ao se trabalhar com os algoritmos de *matching* para localizar os pobres nos municípios brasileiros.

Uma vantagem na utilização de uma estimação não-paramétrica, ou semiparamétrica, para a função $g(r_i/m_i)$ é a possibilidade de se estabelecer intervalos de confiança, lançando mão, por exemplo, de estimadores *bootstrap* (vide, por exemplo, Davison e Hinkley, 1997; Hall, 1992), ou aproximações de primeira ordem (vide, por exemplo, Lehmann, 1999). Com isso, é possível estabelecer intervalos

10. 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.

de confiança para a estimativa da renda *per capita* de corte $\hat{r}_{\text{corte}}$, bem como estabelecer 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*, usando 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 22 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. As figuras A1 a A11 no apêndice A apresentam os gráficos para os 22 contextos geográficos, assumindo-se o valor da caloria consumida fora do domicílio como igual ao valor da caloria consumida em casa. As figuras B1 a B11 apresentam os gráficos para a caloria fora de casa tendo um preço duas vezes maior do que a caloria consumida na residência. Nesses gráficos, o eixo vertical corresponde ao consumo calórico diário por equivalente adulto, enquanto o eixo horizontal representa a renda *per capita*. A curva em linha cheia 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 em linhas tracejadas 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 linha horizontal pontilhada corresponde ao corte de consumo calórico mínimo em cada contexto geográfico considerado.

Para dar uma idéia do ajuste das curvas estimadas, apresentam-se os pontos, referentes às médias de consumo calórico por centésimo de renda. Para evitar problemas originados de 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 grande dispersão nos valores, o que dá idéia do grau de dispersão dos microdados 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, o que constitui 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 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.

Estimação das linhas de indigência e pobreza

Com a definição do nível de renda familiar *per capita* $\hat{r}_{\text{corte}}$ compatível com o consumo calórico adequado, realizou-se a estimação do gasto alimentar *per capita* que está associado àquela renda. Esta etapa é equivalente ao passo (8) na metodologia atualmente empregada (Seção 3.1). O valor deste gasto será então definido como a linha de indigência, tendo sido estimadas tais linhas 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-se, 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 à renda de corte. Empregou-se basicamente a mesma metodologia utilizada para estimar a curva de consumo *versus* renda, isto é, aplicaram-se regressões semiparamétricas para a associação renda *versus* gasto. O modelo tem a seguinte expressão:

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

em que 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 à 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 $h(\cdot)$ utilizaram-se 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. Partindo desse valor $\hat{r}_{corte}$, pode-se 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. Cabe lembrar 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 seu valor 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 encontram-se 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 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, valor que parece 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 4).

Parecem evidentes as pressões sobre o orçamento, resultantes dos gastos com habitação e transporte nas maiores cidades. Juntamente com a importância da

alimentação extradomicílio, tais pressões se refletem no fato de o 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 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 1/3 no Rio de Janeiro, no Distrito Federal e em São Paulo, onde é, inclusive, menor que 20%. Não restam dúvidas que estas proporções não se coadunam com a situação de indigência; pode-se dizer que os pobres gastam no mínimo 40% de sua renda com alimentação, participação esta que é um dos divisores de água na definição de pobreza. A verdade é que, sem se afastar da definição da indigência que leva em conta a relação entre consumo alimentar e renda *per capita*, sustenta-se que parte disso se deve às dificuldades de captação do consumo alimentar, especialmente nas regiões metropolitanas. Duas são as hipóteses para esse vazamento da disponibilidade calórica: a dificuldade de captar e valorar os consumos institucionais e os das redes de proteção social, mesmo depois da investigação dos gastos não-monetários; e o custo da caloria fora ser menor que o domiciliar, já que parcela expressiva dos serviços de alimentação extradomicílio para os estratos pobres e intermediários das regiões metropolitanas se caracteriza pela abundância calórica. Exemplos disso são os carrinhos de cachorro quente e sanduíches, as grandes redes de *fast food*, a alimentação a peso e os variados *self-services*. Outro importante item da alimentação fora, a institucional na escola e no trabalho, não parece apresentar custo calórico superior ao domiciliar. Isso sem desconsiderar a hipótese de que existe emulação que leva as famílias dos estratos de menor renda a tentarem igualar o consumo dos mais ricos, encarecendo, assim, o consumo domiciliar – em termos de disponibilidade calórica.

TABELA 4

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

Fonte: Microdados da POF 2002-03.

Apesar das dificuldades antes descritas, acredita-se que os valores estimados são condizentes com as situações que se deseja caracterizar. Os valores para delimitar a pobreza e a indigência, a incidência por região, a situação dos domicílios e a distribuição regional são bastante próximos daquilo que a maior parte dos estudos e indicadores apontam (tabela 4). A tabela 5 apresenta o número de indigentes e pobres por contexto, bem como a incidência e a distribuição espacial destes.

TABELA 5

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

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

Elaboração dos autores.

Observa-se que metade das populações pobres encontram-se no Nordeste, onde as proporções de indigentes e pobres atingem, respectivamente, cerca de 10% e 30%. Os contextos geográficos que se destacam negativamente, com expressivas participações de populações pobres, são, em primeiro lugar, o Nordeste Urbano Não-metropolitano e o Nordeste Rural. Com valores ao redor ou maiores que 20% encontram-se todas as populações do Norte, das RMs Nordestinas e do Distrito Federal. A distribuição regional não se altera entre as populações indigentes e pobres e para os diferentes fatores de cálculo do preço da caloria fora. Alguma mudança se observa nas regiões Sul e Nordeste, preferencialmente no meio urbano, com a participação no primeiro grupo crescendo e no segundo caindo.

Abordar-se-á, agora, a incidência e a distribuição da indigência e da pobreza no país segundo os locais de moradia. Estas informações, que constam na tabela 6, apontam, de um lado, para o fato de que a participação dos indigentes e pobres é superior no meio rural que nas áreas urbanas e metropolitanas, o que implica maior participação do meio rural na indigência e pobreza ante a sua contribuição

populacional. De outro, dado o alto grau de urbanidade da população brasileira, a participação das populações indigentes e pobres urbanas é de grande envergadura, respondendo, quando se considera o fator 2 para o preço da caloria extradomiciliar, por 77% e 80% da indigência e da pobreza, respectivamente. O descompasso entre as participações na população global e nas de indigentes e pobres é bem pouco expressivo no caso do meio urbano, porém, verificam-se algumas diferenças em se tratando dos meios rural e metropolitano.

TABELA 6

Incidência e distribuição da indigência e da pobreza, segundo a situação de domicílio – Brasil, 1995-96

População – Situação de domicílio	Metropolitano	Urbano	Rural
	Incidência		
Indigentes 1	3%	4%	7%
Indigentes 2	4%	6%	8%
Pobres 1	12%	16%	21%
Pobres 2	15%	19%	23%
	Distribuição		
Indigentes 1	19%	54%	27%
Indigentes 2	22%	55%	23%
Pobres 1	23%	54%	23%
Pobres 2	25%	55%	20%
População	30%	53%	17%

Elaboração dos autores.

A distribuição das populações indigentes e pobres entre os meios rural, urbano e metropolitano por grandes regiões é apresentada na tabela 7, permitindo avançar um pouco mais na caracterização da indigência e pobreza do que a já apontada concentração em termos de maior incidência nas regiões Nordeste e Norte. Fica patente, de um lado, que a indigência e a pobreza são predominantemente rurais nas regiões Nordeste e Centro-Oeste, sendo que nesta última, em se tratando da pobreza, o meio rural responde por parcela menor do que no referente ao caso da população como um todo. Esta situação é, por outro lado, a observada na região Sul, onde a população domiciliada no meio rural representa 17% da população total, ao passo que considerando o fator 2 as populações de indigentes e pobres rurais respondem por, respectivamente, 7% e 13% da indigência e pobreza regional. Já nas regiões Sudeste e Sul não se observam descompassos entre as contribuições populacionais dos diferentes locais de moradia tanto em termos globais quanto para as situações de indigência e pobreza. Na verdade, no caso da indigência no Sudeste, há um ligeiro descompasso, tendo o meio rural participação inferior à observada na população total, mas este descompasso é bem menos expressivo do que o verificado no Sul. Nesse particular, ou seja, nesta região, a indigência é predominantemente urbana, o que indica o quão efetivas, em termos de qualidade de vida, são a distribuição mais equânime da terra e a ascensão econômica da pequena propriedade familiar.

TABELA 7

Distribuição da indigência e da pobreza, segundo a situação de domicílio, por Grande Região – 1995-96

População – Situação de domicílio	Metropolitano	Urbano	Rural
Norte			
Indigentes 1	7%	64%	29%
Indigentes 2	7%	64%	29%
Pobres 1	11%	61%	28%
Pobres 2	12%	61%	27%
População	14%	61%	26%
Nordeste			
Indigentes 1	10%	49%	41%
Indigentes 2	10%	57%	33%
Pobres 1	12%	54%	34%
Pobres 2	13%	58%	29%
População	19%	52%	28%
Sudeste			
Indigentes 1	37%	56%	7%
Indigentes 2	50%	44%	5%
Pobres 1	45%	46%	9%
Pobres 2	51%	42%	8%
População	44%	47%	9%
Sul			
Indigentes 1	25%	67%	7%
Indigentes 2	25%	68%	7%
Pobres 1	26%	61%	14%
Pobres 2	24%	63%	13%
População	24%	58%	17%
Centro-Oeste			
Indigentes 1	23%	63%	15%
Indigentes 2	23%	64%	12%
Pobres 1	23%	68%	9%
Pobres 2	22%	70%	8%
População	18%	71%	11%

Elaboração dos autores.

5 MATCHING POF VERSUS CENSO

5.1 A METODOLOGIA DO MATCHING

Para a espacialização das estimativas das populações pobres e indigentes aplicaram-se 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 às 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.

Foi empregado modelo 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:

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

em que a probabilidade $\text{Pr}[\text{indivíduo ser pobre}]$ do indivíduo estar no grupo dos pobres (ou indigentes) dependerá das co-variáveis $x_1, x_2, \dots, x_p$. Esse modelo será utilizado em uma segunda etapa para encontrar os indivíduos indigentes nos dados

do Censo 2000. Portanto, na escolha das co-variáveis $x_1, x_2, \dots, x_p$, é necessário restringir-se ao grupo de variáveis que estão também disponíveis nos microdados 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, poder-se-á encontrar quantos indivíduos pobres, em média, estão 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 co-variáveis (também pelos microdados do Censo) $x_1, x_2, \dots, x_p$.
- 3) Utilizando o modelo estatístico representado na equação (2), estima-se a probabilidade $prob_i$ do indivíduo i ser indigente.
- 4) Repetem-se os passos 2 e 3 acima para todos os indivíduos do município j e encontram-se todos os valores $prob_i$.
- 5) Somam-se todos os valores $prob_i$ para todos os indivíduos do município j de forma a se obter uma primeira estimativa num_j do número de indigentes no município j .
- 6) Repetem-se os passos 1 a 5 para todos os municípios j do Brasil e obtêm-se as diversas estimativas num_j de cada um deles.
- 7) É provável que ao se somar todos os valores num_j de todos os municípios, o cômputo geral de indigentes no Brasil, ou por sub-regiões (como, por exemplo, por macrorregião brasileira ou por unidade da federação), seja diferente dos números estimados diretamente da POF. Portanto, nessa etapa é preciso fazer ajuste dos valores de forma a compatibilizar as estimativas em nível de municípios. Para isso, reescalona-se linearmente o número de estimativas num_j para cada município de forma a se obter 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 ser, então, mapeados para a obtenção de representações georeferenciadas do número de indigentes nos municípios brasileiros.

5.2 A CONSTRUÇÃO DA BASE DE DADOS

Existe um núcleo comum entre as investigações domiciliares do IBGE, ou seja, entre as Pnads, amostras dos Censos, POFs e PPV, destacando-se o gênero, a raça, a idade e a relação com o chefe da família, para as pessoas, e no caso dos domicílios, o tipo, a existência de energia elétrica, água, esgotamento, levantamento dos bens, número de cômodos, número de banheiros, entre outras informações. No anexo consta a tabela com as variáveis selecionadas e as estatísticas descritivas básicas.¹¹

No caso da renda, as diferenças são um pouco maiores, uma vez que a POF, ao investigar detalhadamente os orçamentos familiares, apura de forma mais precisa os

11. Enquanto a amostra do Censo Demográfico de 2000 é composta de 5 milhões de registros, no caso das informações por domicílios, a POF conta com amostra de pouco menos que 50 mil unidades de consumo. Concretamente, a amostra do Censo é da ordem de 10% ante os 0,1% da POF.

rendimentos, chegando inclusive a pesquisar as rendas não-monetárias, entre as quais se sobressaem o valor do aluguel dos imóveis próprios ou cedidos, a produção própria, as retiradas em bens dos empreendimentos e as doações privadas e públicas. Já o Censo apura de forma menos detalhada as rendas monetárias, particularmente as oriundas de aluguéis e das aplicações de capital, não investigando as rendas não-monetárias (doação, produção para o próprio consumo etc.). Assim, é necessário ter cuidado ao se comparar os rendimentos das pesquisas. É neste aspecto que reside o que se consideram os maiores problemas na discriminação da pobreza espacial e temporalmente, pois se faz necessário efetuar ajustes nas outras pesquisas – Pnad e Censo. Para não incorrer em arbitrariedades na estimação das rendas não-monetárias e em ajustes nas rendas monetárias, decidiu-se aplicar 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 se obter indicadores domiciliares – ou familiares.¹² Assim, as 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 têm uma aderência. As diferenças podem ser advindas do tamanho (muito maior) da amostra do Censo Demográfico, cujo resultado é uma maior variabilidade. No caso das rendas, ficam patentes 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 à inflação do período. No caso do inventário dos bens domésticos (rádio, televisão, máquina de lavar roupa etc.), as pesquisas possuem perguntas distintas, com o Censo Demográfico investigando a presença ou não do item, enquanto que a POF coleta informações sobre o número de itens existentes, tornando a variável binária – ter ou não. Como já se apontou, para efeito do *matching* consideraram-se apenas as rendas monetárias, uma vez que o Censo Demográfico não apura os rendimentos não-monetários. Assim, no caso das variáveis de renda são apresentadas apenas aquelas relativas à parte monetária, pois, como colocado anteriormente, há dificuldade em efetuar a comparação da renda investigada entre as pesquisas.

Como forma de sintetizar a análise da espacialização, passa-se a apresentar os resultados para o caso de a caloria extradomiciliar ter o dobro do custo da intradomiciliar. Assim, na tabela 8 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.

12. 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, em que família tem por base as relações de parentesco. Esta conceituação da POF se assemelha à 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.

TABELA 8

Principais estatísticas do *matching* para pobres

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

Elaboração dos autores.

6 ESPACIALIZAÇÃO DAS POPULAÇÕES INDIGENTES E POBRES

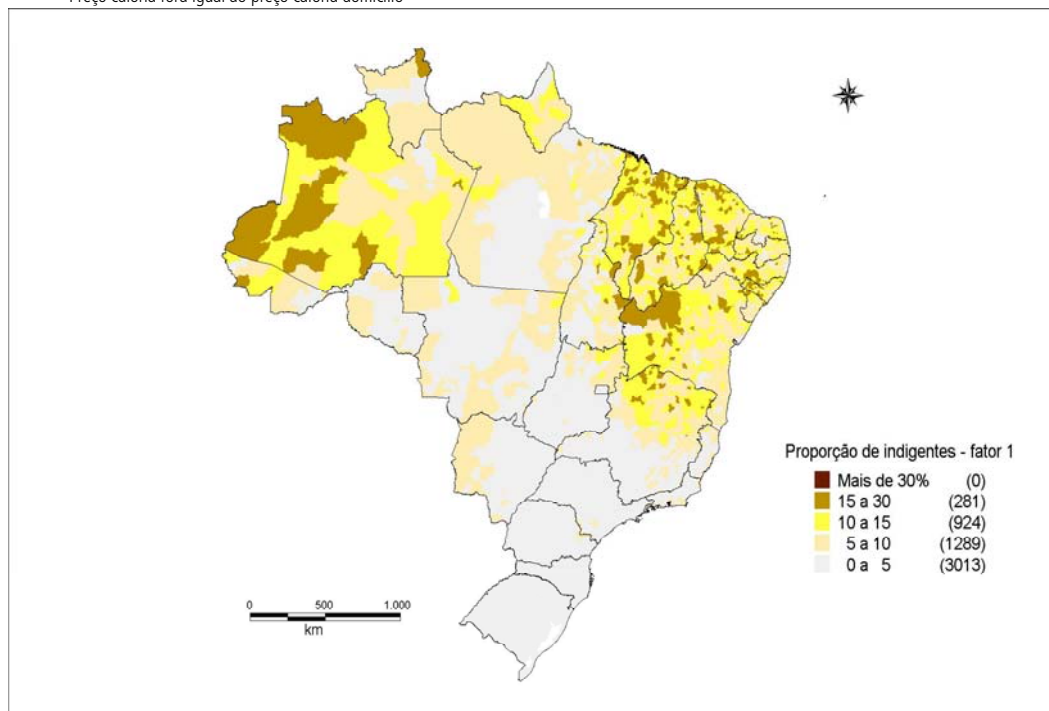
Os dados espacializados da incidência da indigência encontram-se ilustrados nos mapas 1 e 2, nos quais se consideraram, respectivamente, os fatores 1 e 2 para a razão preço calórico fora *versus* dentro do domicílio. Fica patente que a indigência está concentrada na região Nordeste, tendo também relativa presença nos estados do Norte, especialmente no Amazonas e no Acre. Já o Centro-Sul apresenta baixos níveis de incidência de populações indigentes, 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%.

De modo mais detalhado, e considerando particularmente o fator 2, a presença de populações indigentes se mostra mais expressiva nos Estados do Maranhão, Piauí, Ceará e Alagoas, onde está mais disseminada. Somam-se a esses, as regiões oeste do Amazonas e noroeste da Bahia, com manchas no norte mineiro, no Acre e nos sertões de Pernambuco e Paraíba. Nestes casos, cerca de 1/4 da população encontra-se em situação de indigência, apontando, portanto, para o fato de a indigência situar-se no nordeste setentrional e no extremo oeste da Amazônia. Como se pode verificar, isso se deve à ruralidade dessas áreas, haja vista que o comportamento da incidência da indigência no meio rural – mapa 8 – é muitíssimo similar. Tais resultados detalham espacialmente os dados obtidos na POF – tabela 5. Por outro lado, vale ter presente que os indicadores relativos de indigência, ou seja, da 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 observar nos dados da tabela 5, 1/3 dos indigentes encontram-se nas regiões Sul e Sudeste.

MAPA 1

Proporção de indigentes por município

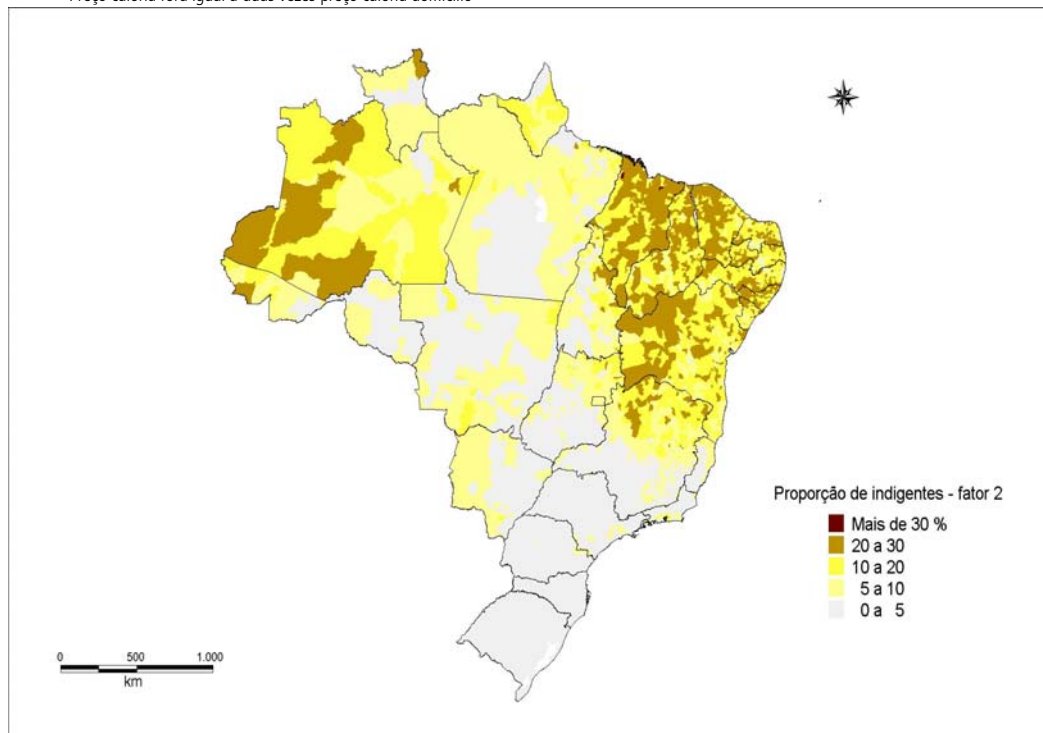
Preço caloria fora igual ao preço caloria domicílio



MAPA 2

Proporção de indigentes por município

Preço caloria fora igual a duas vezes preço caloria domicílio



Os resultados da espacialização da pobreza são bem consistentes com aquilo que apontam a maior parte dos estudos e análises. Em ambos os casos – fatores 1 e 2 –, há 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 (mapas 3 e 4). É notório que se pode traçar uma linha imaginária cortando o país no sentido leste para oeste, dividindo-o em duas metades, uma ao norte, em que é maior a proporção de pessoas pobres, e outra ao sul, em que a proporção de pobres na população não ultrapassa 5%.

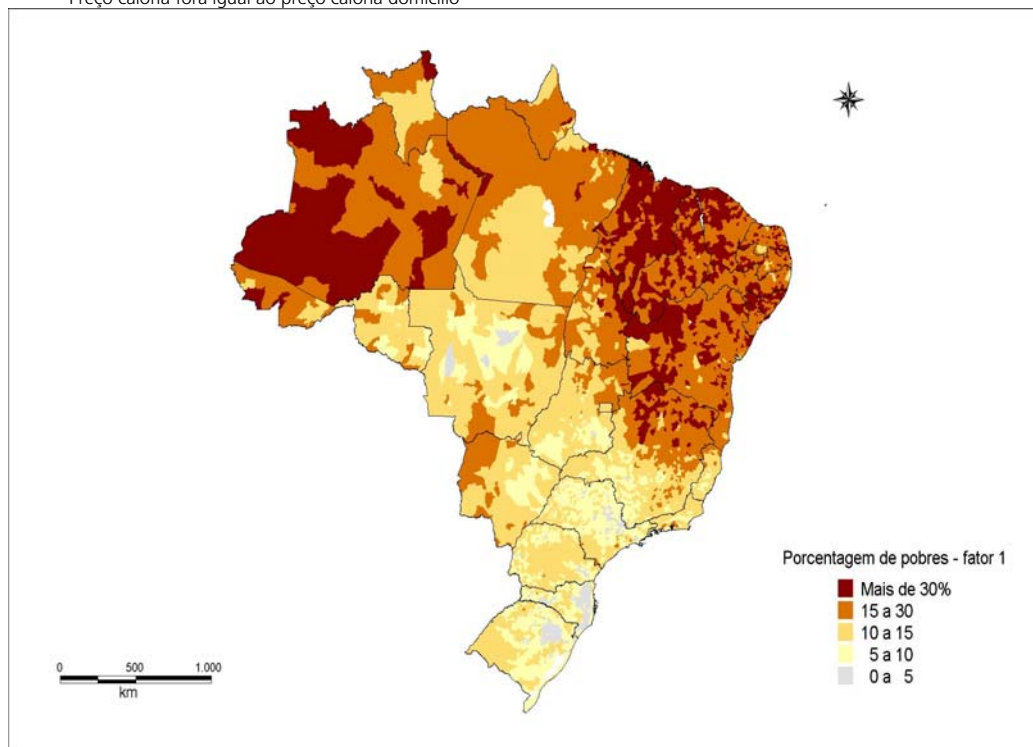
Nessa direção, a análise da incidência da pobreza presente no mapa 3 – em que se considerou o fator 1 de preço da caloria – indica que nos Estados do Rio Grande do Sul, Santa Catarina e São Paulo a quase totalidade dos municípios apresenta baixa participação de pobres na população, chamando particularmente a atenção as áreas de concentração da pequena propriedade familiar – Vale do Itajaí e Serra Gaúcha. De outra parte, desperta também a atenção a concentração 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. De fato, a pobreza tem elevados níveis de concentração nos Estados do Maranhão, Piauí, Ceará, Bahia, Alagoas e Amazonas. Neste último caso, deve-se ter o cuidado em razão do tamanho dos municípios, que por serem bastante extensos fornecem falsa impressão de concentração de populações pobres. Nesse sentido, deve-se sublinhar, mais uma vez, que o número absoluto de pobres no Estado de São Paulo e, de modo geral, na região Sudeste, ultrapassa 3 e 8 milhões de pessoas, respectivamente. Isso significa participação, da ordem de 1/3, da região Sudeste no total de pobres do país (tabela 5 e mapa 8).

Analisando o mapa 4, que diz respeito à pobreza quando se considera que a alimentação fora de casa tem custo duas vezes maior do que dentro, algumas alterações na distribuição e na intensidade da pobreza são observadas. O número de municípios cuja participação dos pobres na população supera 30% torna-se bem maior, cobrindo, contudo, as mesmas áreas já apontadas anteriormente. Ou seja, chamam a atenção os elevados níveis de concentração da pobreza nas regiões Nordeste e Norte, onde se destacam, em termos negativos, os Estados do Maranhão, do Piauí e do Amazonas.

MAPA 3

Proporção de pobres por município

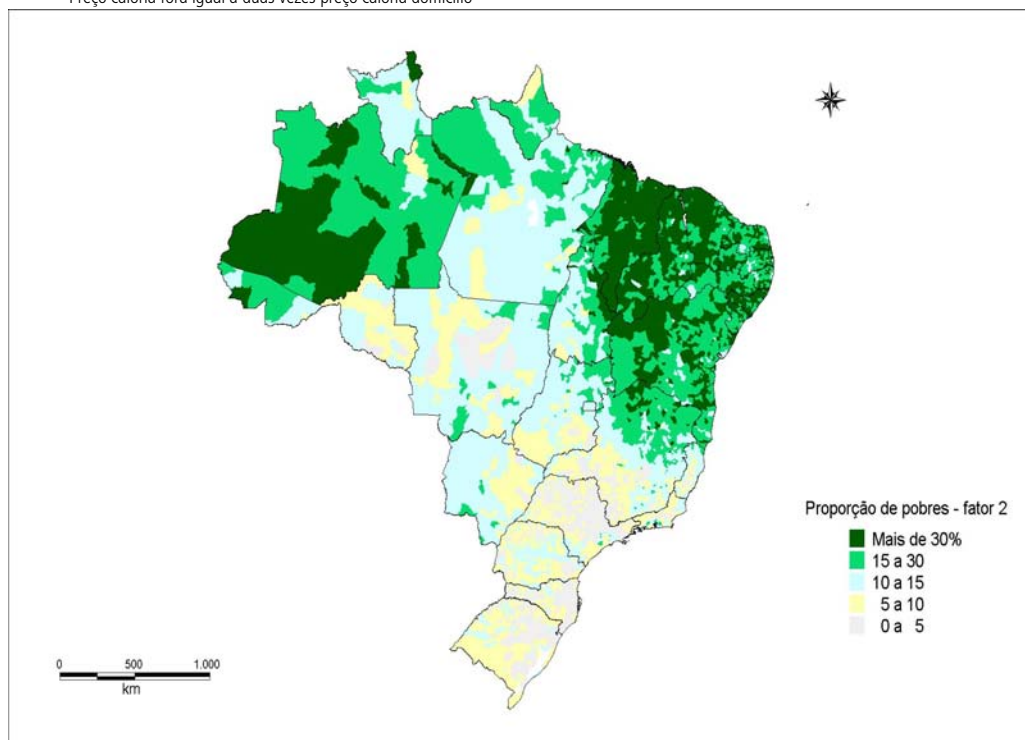
Preço caloria fora igual ao preço caloria domicílio



MAPA 4

Proporção de pobres por município

Preço caloria fora igual a duas vezes preço caloria domicílio



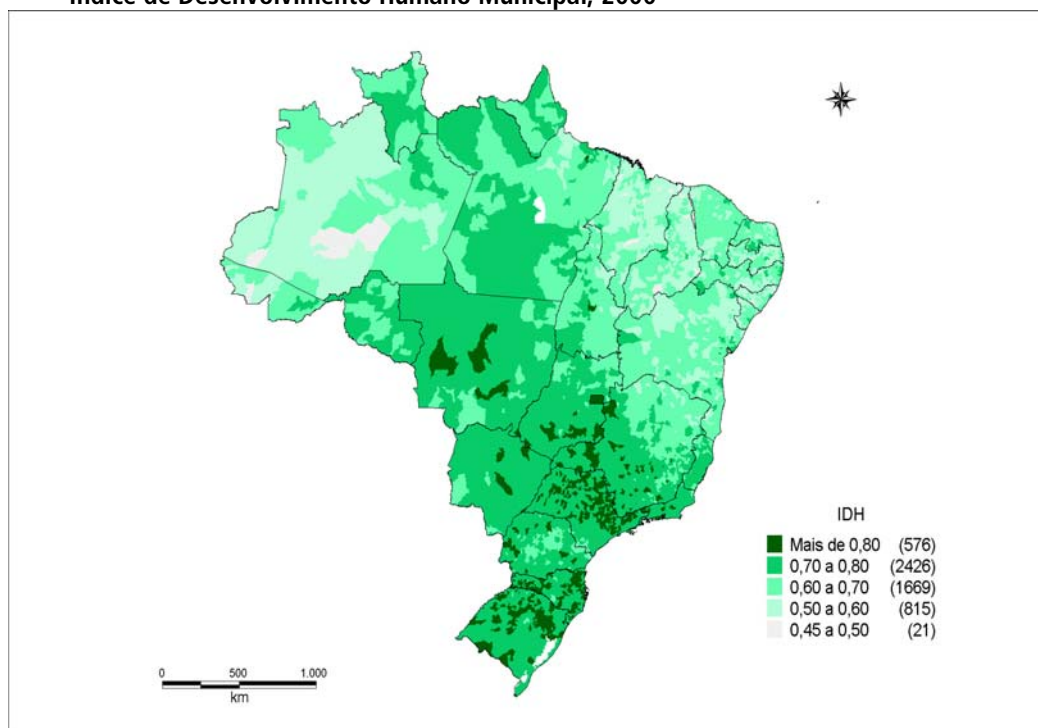
É interessante notar que o mapa 4 coloca o foco nas regiões em que a proporção de pobres supera entre 10% e 15%, como no Centro-Oeste, particularmente nas áreas de fronteira com Paraguai e Bolívia, no nordeste do Mato Grosso e Goiás, onde a questão fundiária e a existência de populações de quilombolas e indígenas são fatos reconhecidos. É também evidente a clivagem no Estado de Minas, com sua região norte apresentando indicadores de incidência de pobres que atingem 30%, enquanto no Triângulo e na porção sul a participação dos pobres na população não ultrapassa 10%.

Com o intuito de avaliar o quão próximos de outros indicadores de bem-estar populacional encontram-se os resultados aqui alcançados acerca da espacialização da indigência e da pobreza, são apresentados os mapas do IDH-M, da esperança de vida ao nascer e da renda domiciliar *per capita*. Dessa forma, tornar-se-á possível verificar o grau de aderência entre indigência e pobreza e indicadores de desenvolvimento de riqueza. Esses mapas foram elaborados com base nos dados do Atlas do Desenvolvimento Humano do Brasil, elaborado pelo *pool* Pnud/Ipea/FJP a partir dos dados do Censo Demográfico de 2000. De forma a aperfeiçoar a análise, foram calculadas as correlações dos dados municipais entre incidência de indigentes e pobres e esses indicadores, bem como com índices de concentração da renda, mais especificamente o Gini e o L de Theil (tabela 9).

Os mapas 5, 6 e 7, em que são apresentados, respectivamente, os perfis espaciais do IDH-M, da esperança de vida ao nascer e da renda domiciliar *per capita*, retratam, grosso modo, situações semelhantes à da indigência e pobreza. Concretamente, parece haver uma linha divisória entre o Centro-Sul e o Norte-Nordeste do país. Isso é corroborado pelos elevados coeficientes de correlação, superiores, em módulo, a 0,7; de fato, os coeficientes de correlação são negativos e bastante expressivos. É de se sublinhar, todavia, algumas diferenças importantes, quais sejam: a associação é mais expressiva com o IDH-M do que com a renda domiciliar *per capita* e a esperança de vida, e observa-se maior aderência – negativa – com a pobreza do que com a indigência. Assim, há correlação negativa bastante grande entre a presença de pobres e indigentes e o desenvolvimento humano e, em menor grau, com a riqueza. Todavia, existe algum descompasso entre tais associações negativas, dado que os coeficientes situam-se entre -73% e -90%, ou seja, a indigência, principalmente, e a pobreza, em menor grau, não caminham tão *pari passu* com o desenvolvimento e a riqueza.

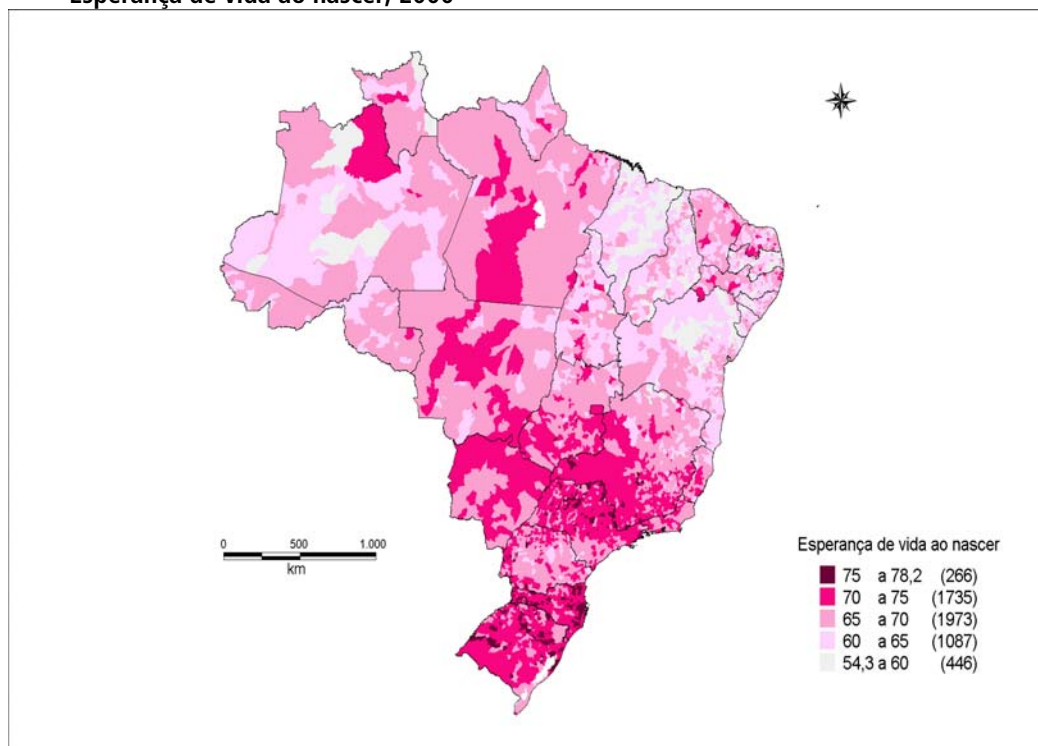
MAPA 5

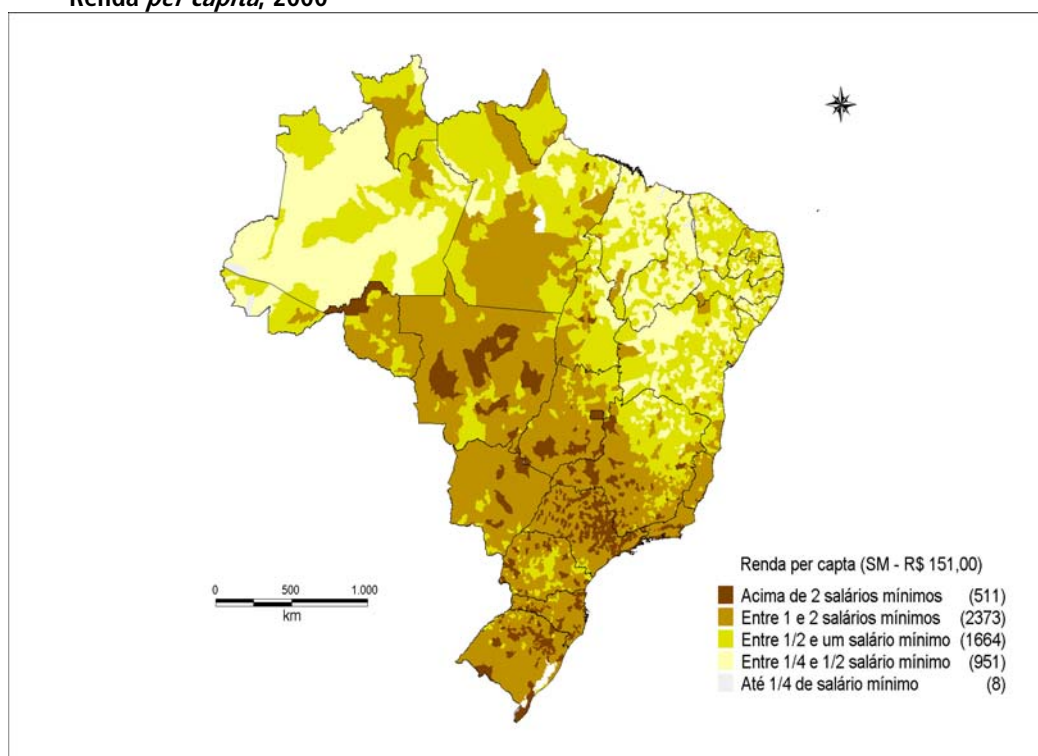
Índice de Desenvolvimento Humano Municipal, 2000



Mapa 6

Esperança de vida ao nascer, 2000



Renda *per capita*, 2000

Já no que concerne à desigualdade, observa-se que a correlação positiva com a pobreza é pouco significativa, haja vista que os coeficientes de correlação entre as distribuições de indigentes e pobres e os índices de Gini e o L de Theil situam-se ao redor de 20%. Depreende-se desse fato que as políticas de combate à pobreza e as voltadas para a diminuição da desigualdade não se dirigem, em termos mais gerais, às mesmas regiões. Efetivamente, a incidência da pobreza é claramente concentrada espacialmente, enquanto a desigualdade é uma realidade que está disseminada por todo o território nacional.

TABELA 9

Coeficientes de correlação entre a incidência da indigência e pobreza e indicadores selecionados

População	Renda <i>per capita</i>	IDH-M	Esperança de vida	Índice de Gini	L de Theil
Indigentes					
Fator 1	-0.726	-0.838	-0.746	0.496	0.205
Fator 2	-0.730	-0.849	-0.746	0.454	0.190
Pobres					
Fator 1	-0.812	-0.898	-0.793	0.465	0.216
Fator 2	-0.799	-0.889	-0.791	0.439	0.206

Elaboração dos autores.

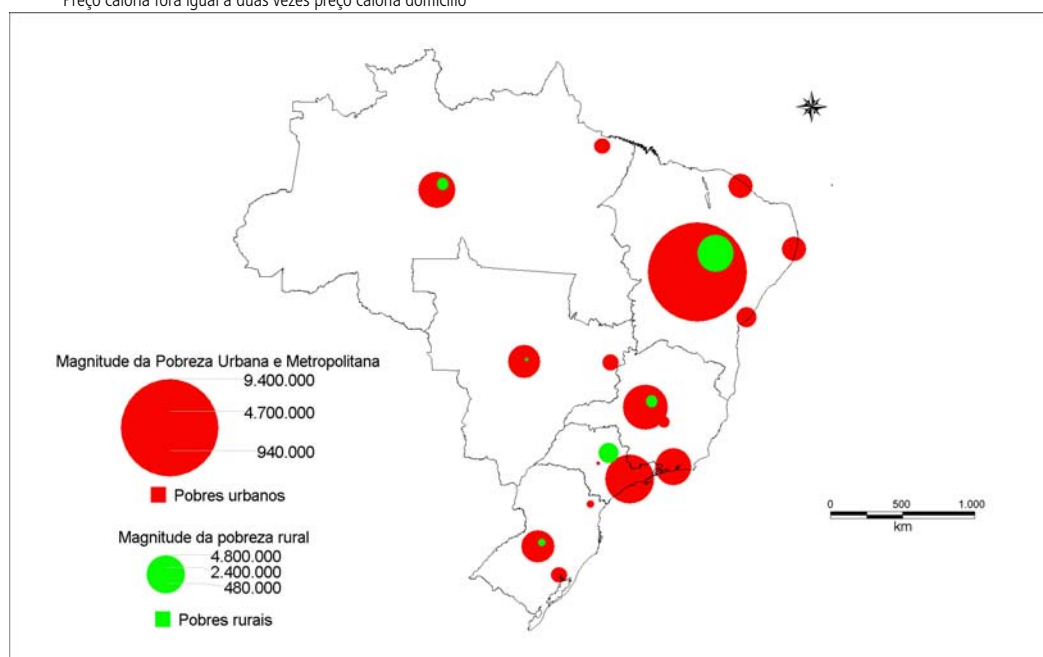
O mapa 8 ilustra a magnitude da pobreza, levando-se em conta o fator 2, nos domínios geográficos considerados no estudo, explicitando, o que já foi sublinhado seguidas vezes, que expressivos contingentes de pobres encontram-se nas Regiões Metropolitanas de São Paulo e Rio de Janeiro e nas áreas urbanas do Nordeste e Sudeste, ainda que nas áreas rurais as proporções de pobres sejam mais significativas. Assim, as políticas públicas de combate à pobreza devem ter presente esse quadro, ao

que se soma o fato de serem de envergadura os diferenciais de custo de vida e as pressões orçamentárias. Efetivamente, as linhas de pobreza metropolitanas e, em menor grau, as urbanas superam em muito as rurais, o que significa custos bastante diversos para os programas de transferência de renda. Nesta direção, valores universais para os programas de transferência de renda repercutem de modo bastante distinto, advindo de tal fato, como se sabe, a defesa de uma segmentação – diferenciação. Bolsas diferentes, ainda que defensáveis, podem validar o retorno pouco eficaz da política de salários mínimos regionais.

MAPA 8

População de pobres por contexto geográfico

Preço caloria fora igual a duas vezes preço caloria domicílio



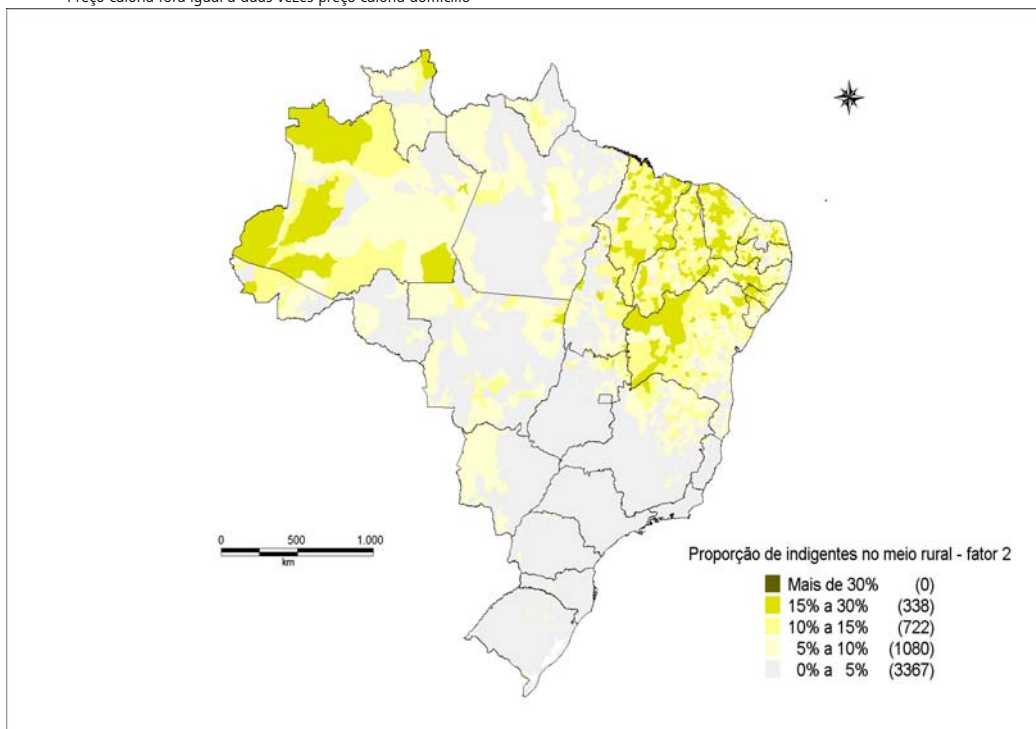
Por último, resta analisar a distribuição espacial da indigência e da pobreza rural no que concerne à sua incidência na população rural dos municípios. Como se pode notar pela comparação entre os mapas 2 e 4 e entre os de número 9 e 10, a seguir, o padrão distributivo é muito semelhante. Há concentração espacial em termos de maiores proporções de indigentes e pobres nas populações no Nordeste e Norte do país, bem como no norte mineiro, na região do Pantanal, no nordeste e noroeste mato-grossense e no nordeste goiano.

Por outro lado, os municípios do Sul e Sudeste apresentam baixos níveis de incidência da indigência e da pobreza rural, valendo lembrar, ademais, que no caso da região Sul a distribuição das populações indigentes e pobres entre os meios urbano, rural e metropolitano é diversa da observada para a população com um todo. Nessa região, as áreas urbanas contribuem com proporção maior de indigentes e pobres *vis-à-vis* à parcela de sua população que aí se domicilia. Rio Grande do Sul e Santa Catarina, onde predomina a pequena propriedade familiar e são baixos os níveis de concentração fundiária, são os estados em que a participação de indigentes e pobres na população é menor. Pode-se dizer que São Paulo também apresenta baixos níveis de incidência de pobreza rural, mas diferentemente desses outros estados, sua população rural é muito pouco expressiva.

MAPA 9

Proporção de indigentes por município – área rural

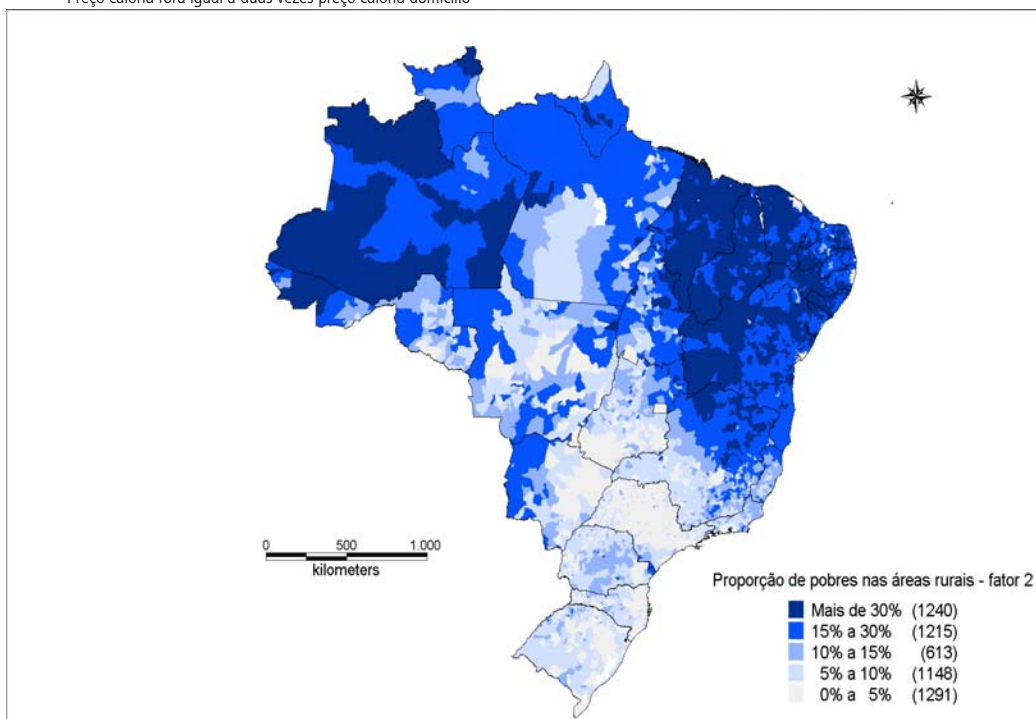
Preço caloria fora igual a duas vezes preço caloria domicílio



MAPA 10

Proporção de pobres por município – área rural

Preço caloria fora igual a duas vezes preço caloria domicílio



REFERÊNCIAS

- ARIAS, A. *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, dez. 2003. (publicação restrita). Mimeo.
- BARRETTO, 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.; ANDERSON, D. *Model selection and inference*. A practical information-theoretic approach. Springer, 1998.
- COMISSÃO ECONÔMICA PARA A AMÉRICA LATINA E O CARIBE (CEPAL). *Medición de la pobreza en Brasil: una estimación de las necesidades de energía y proteínas de la población*. Santiago: Cepal, 1996.
- CHEN, X.; CONLEY, T. A new semiparametric spatial model for panel time series. *Journal of Econometrics*, n. 105, p. 59-83, 2001.
- COCHRAN, W. *Sampling techniques*. Wiley, 1977.
- DAVISON, A.; HINKLEY, D. *Bootstrap methods and their applications*. Cambridge University Press, 1997.
- DRAPER, N.; SMITH, H. *Applied regression analysis*. Wiley-Interscience, 1998.
- FAO/WHO/UNU. *Necessidades de energia e proteínas*. Genebra: FAO/WHO, 1985. Revisão do mesmo estudo elaborado em 1973 e disponibilizado em 1974 (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.; FRIEDMAN, J. *The elements of statistical learning*. Data mining inference and prediction. Springer, 2001.
- HILLIER, F. S.; LIEBERMAN, G. J. *Introduction to operations research*. McGraw-Hill Science, 2004.
- HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. 2. ed. Wiley-Interscience, 2000.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Pesquisa de Orçamentos Familiares 2002-2003: primeiros resultados – Brasil e Grandes Regiões*. 2. ed. Rio de Janeiro, 2004a.
- _____. *Pesquisa de Orçamentos Familiares 2002-2003: análise da disponibilidade domiciliar de alimentos e do estado nutricional no Brasil*. Rio de Janeiro, 2004b.
- _____. *Segurança alimentar: 2004*. Rio de Janeiro, 2006a.
- _____. *Pesquisa de Orçamentos Familiares 2002-2003: antropometria e análise do estado nutricional de crianças e adolescentes no Brasil*. Rio de Janeiro, 2006b.

- INSTITUTO DE PESQUISA ECONÔMICA APLICADA (Ipea). *Contribuições para a construção de linhas de indigência para o Brasil*. Brasília, 06 de outubro de 1997. Mimeografado da Comissão Mista Ipea/Cepal/IBGE.
- KLEINBAUM, D. G.; KLEIN, M.; PRYOR, E. R. *Logistic regression*. 2. ed. Springer, 2005.
- LEHMANN, E. *Elements of large-sample theory*. Springer, 1999.
- LEITENSTORFER, F.; TUTZ, G. *Generalized monotonic regression based on B-Splines with an application to air pollution data*. Ludwig-Maximilians Universitat, maio 2005 (Discussion Paper, n. 444).
- LUSTOSA, T. *Cálculo de las necesidades energéticas de la población brasileña para la construcción de una línea de pobreza*. Santiago: Cepal, 1999 (Apresentado no IV 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, S. Medindo a pobreza no Brasil: evolução metodológica e requisitos de informação básica. In: LISBOA, M.; MENEZES FILHO, N. (Orgs.). *Microeconomia e sociedade no Brasil*. Rio de Janeiro: Contracapa, 2001.
- _____. *Pobreza no Brasil: afinal do que se trata?* Rio de Janeiro: Editora FGV, 2003.
- TUTZ, G.; LEITENSTORFER, F. *Generalized smooth monotonic regression*. Ludwig-Maximilians Universitat, mar. 2005 (Discussion Paper, n. 417).
- WINSTON, W. L. *Operations research: applications and algorithms*. Duxbury Press, 2003.
- WOOLDRIDGE, J. *Econometric analysis of cross section and panel data*. The MIT Press, 2002.

ANEXO

ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS UTILIZADAS NO MATCHING (CENSO DEMOGRÁFICO E POF)

Variável/Estatística	Censo demográfico				Pesquisa de Orçamento Familiar			
	Mínimo	Máximo	Média	Desvio-padrão	Mínimo	Máximo	Média	Desvio-padrão
Número de moradores	1	43	3,8	2,0	1	20	3,8	1,9
Número de moradores ajustados	1	43	3,8	2,0	1	20	3,7	1,9
Número de moradores homens	0	30	1,9	1,3	1	12	2,0	1,1
Número de mulheres	0	22	1,9	1,2	1	13	2,0	1,1
Renda monetária mensal familiar	0	1.011.958,00	1.042,34	2.970,37	0	384.675,67	1.307,14	2.980,34
Renda monetária mensal familiar <i>per capita</i>	0	701.000,00	345,01	1.229,03	0	54.953,67	431,74	957,24
Renda monetária do trabalho mensal familiar	0	1.010.500,00	841,70	2.715,80	0	367.555,42	961,73	2.478,32
Renda monetária do trabalho mensal familiar <i>per capita</i>	0	700.000,00	265,36	1.073,85	0	52.507,92	304,99	715,62
Renda monetária de aposentadorias e pensões	0	50.000,00	165,16	548,85	0	27.250,00	173,44	666,27
Renda monetária de aposentadorias e pensões <i>per capita</i>	0	50.000,00	64,47	262,21	0	21.800,00	68,63	332,34
Tipo de residência (*)	1	3	-	-	1	3	-	-
Número de cômodos	1	30	5,7	2,3	1	30	5,8	2,3
Número de dormitórios	1	9	2,0	0,9	1	9	2,0	0,9
Número de banheiros	0	9	1,1	0,8	0	13	1,2	0,7
Forma de abastecimento de água (*)								
Escoadouro sanitário (*)								
Condição de ocupação (*)								
Existência de energia elétrica (**)	0	1	93%	0,3	0	1	94%	0,2
Presença de rádios (**)	0	1	87%	0,3	0	1	28%	0,5
Presença de geladeiras e freezer (**)	0	1	81%	0,4	0	1	82%	0,4
Presença de videocassete (**)	0	1	32%	0,5	0	1	22%	0,4
Presença de máquinas de lavar roupas (**)	0	1	30%	0,5	0	1	32%	0,5
Presença de microondas (**)	0	1	17%	0,4	0	1	11%	0,3
Presença de microcomputadores (**)	0	1	9%	0,3	0	1	9%	0,3
Presença de TVs (**)	0	1	85%	0,4	0	1	85%	0,4
Presença de automóveis (**)	0	1	31%	0,5	0	1	1%	0,1
Presença de ar-condicionado (**)	0	1	7%	0,2	0	1	8%	0,3
Idade do chefe	10	130	45,6	15,4	13	102	46	15
Gênero do chefe (***)			77%				74%	
Cor do chefe	1	5			1	5		
Anos de estudo do chefe	0	17	5,4	4,5	0	17	5,3	4,6
Crianças de 7 a 14 anos na escola (***)								
Total de crianças de 7 a 14 anos (***)								
Pessoas de 15 anos e mais	0	23	2,7	1,3	1	8	1,6	0,8
Alfabetizados com mais de 15 anos (***)								
Anos de estudos das pessoas de 15 anos e mais	0	177	15,7	12,2	0	63	11,4	7,8

Fonte: Censo Demográfico 2000 e POF 2002-03.

Notas: (*) Variáveis categóricas para as quais não se aplicam as estatísticas descritivas.

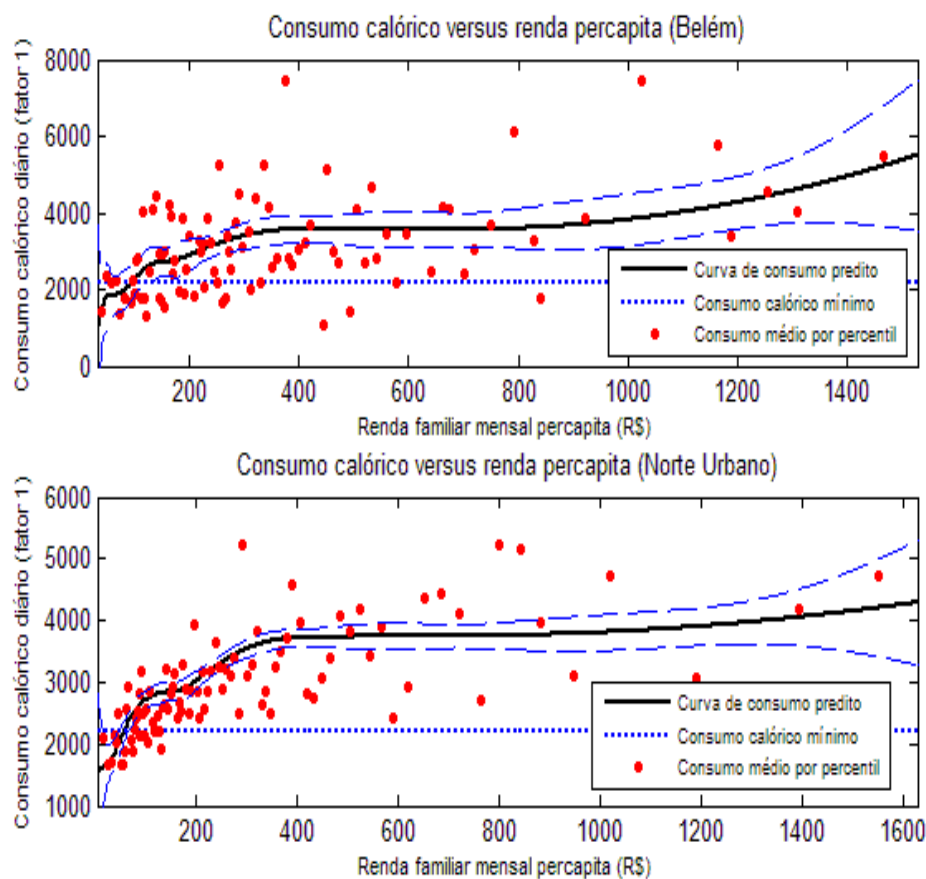
(**) Variáveis binárias – ter ou não ter – para as quais o valor médio corresponde à proporção de domicílios com a presença do bem ou do serviço.

(***) Variável binária e categórica que relaciona o valor 1 a homem e 0 a mulher, com a média referindo-se à proporção de homens.

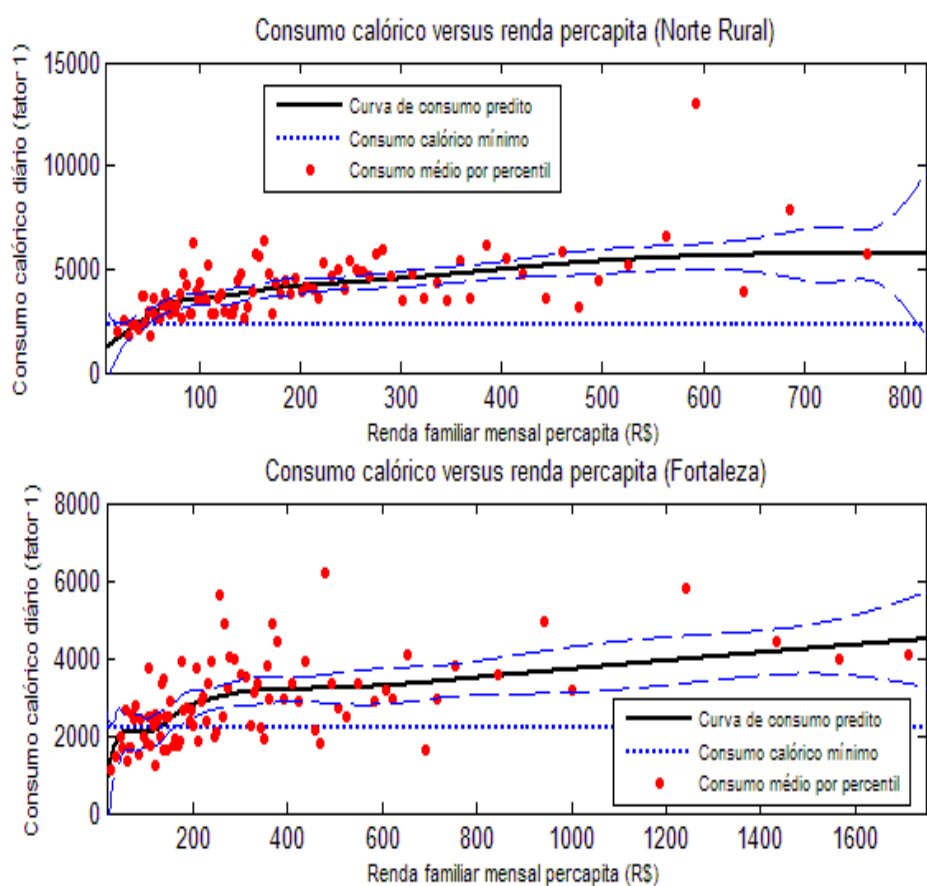
APÊNDICE A

VALORES IGUAIS PARA O PREÇO DA CALORIA CONSUMIDA FORA E DENTRO DO DOMICÍLIO

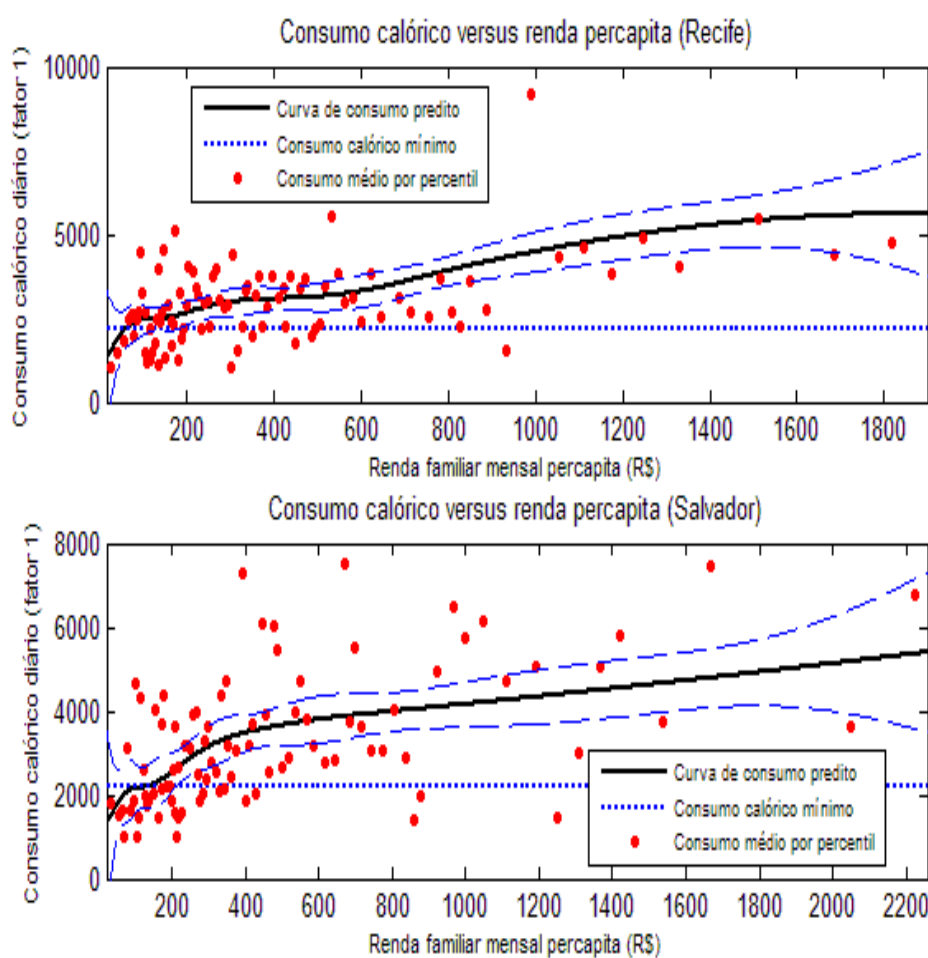
A1. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Belém e do Norte Urbano



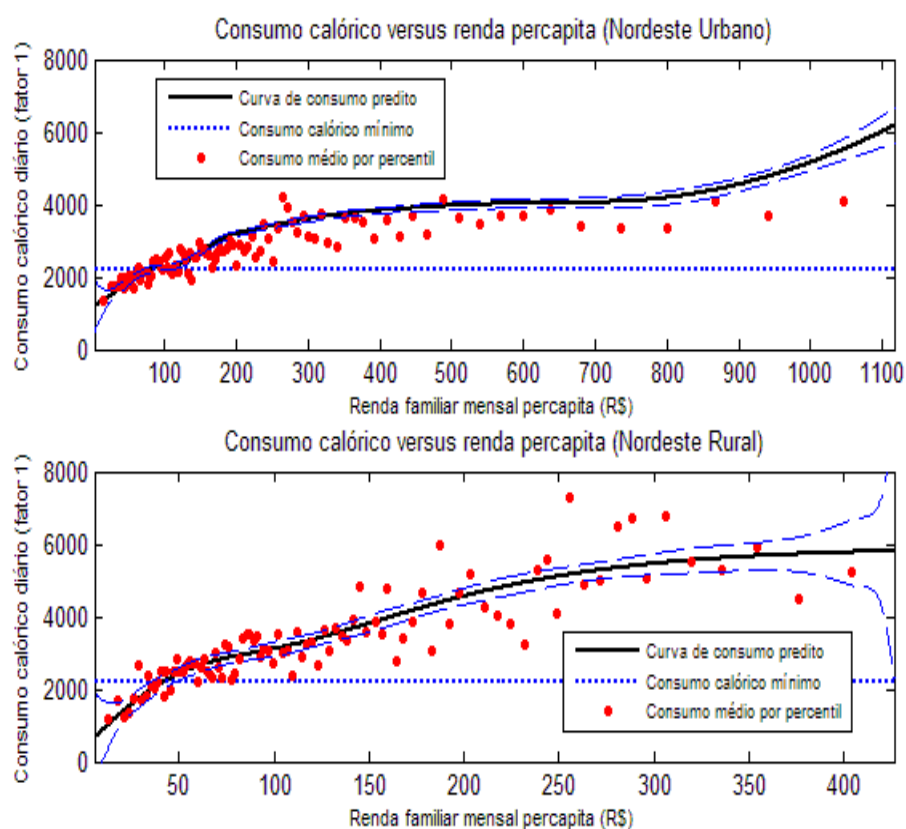
A2. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Norte Rural e de Fortaleza



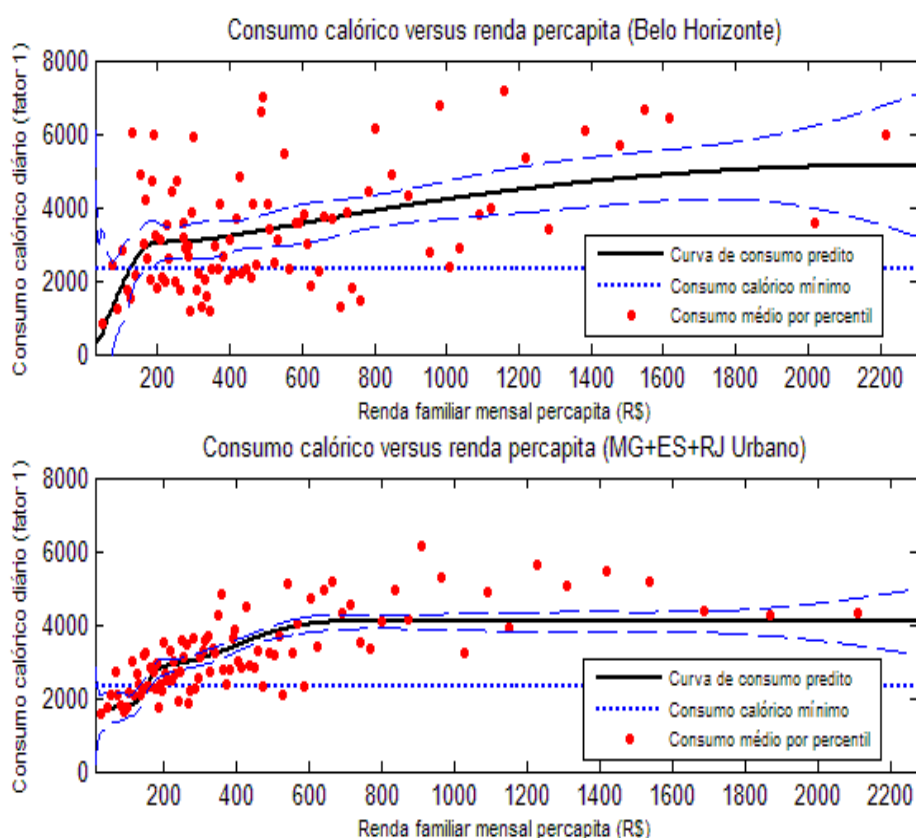
A3. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Recife e de Salvador



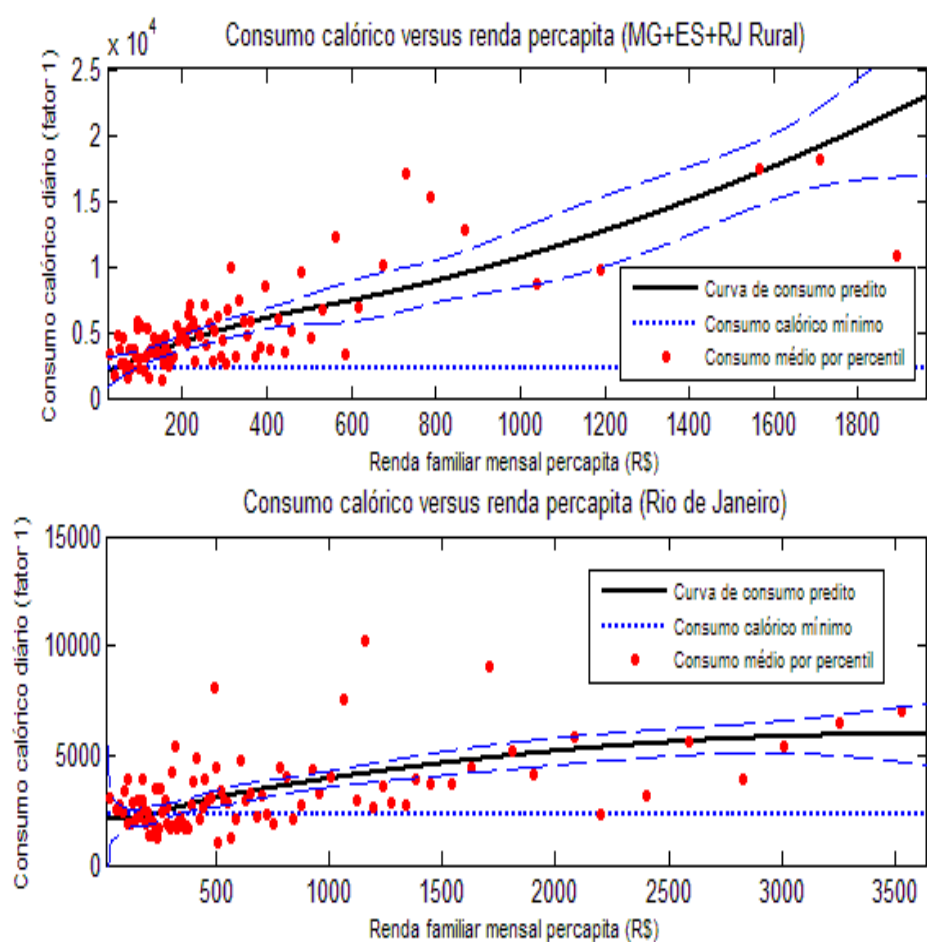
A4. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Nordeste Urbano e do Nordeste Rural



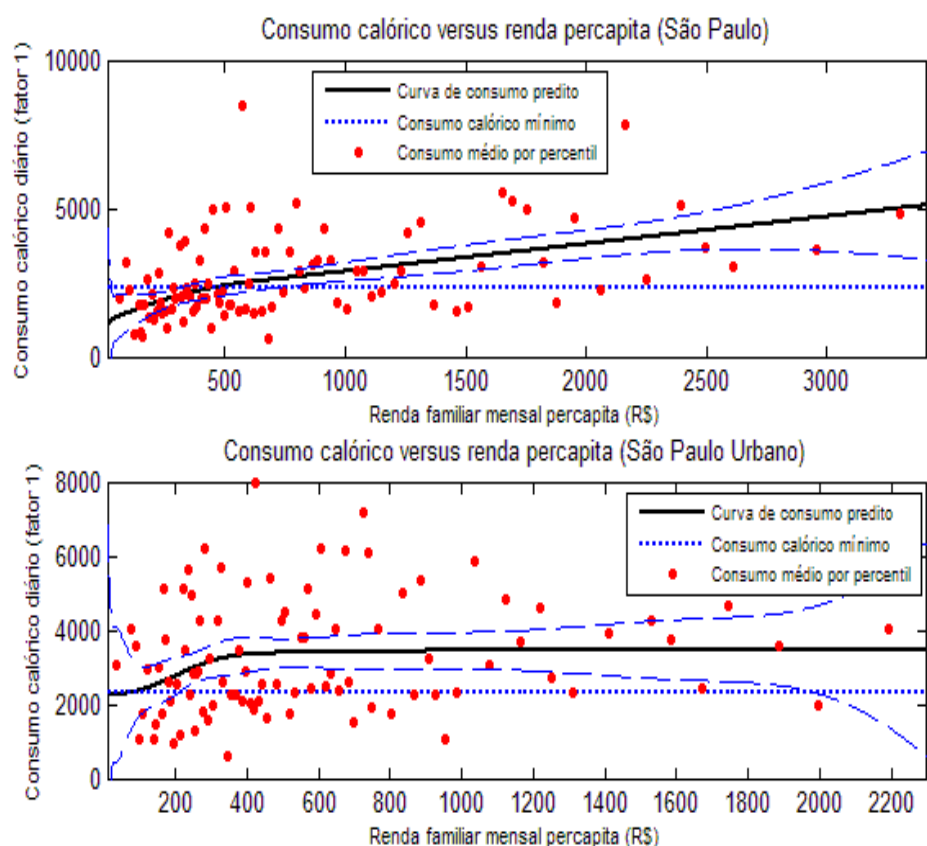
A5. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Belo Horizonte e do Sudeste (excluindo São Paulo) Urbano



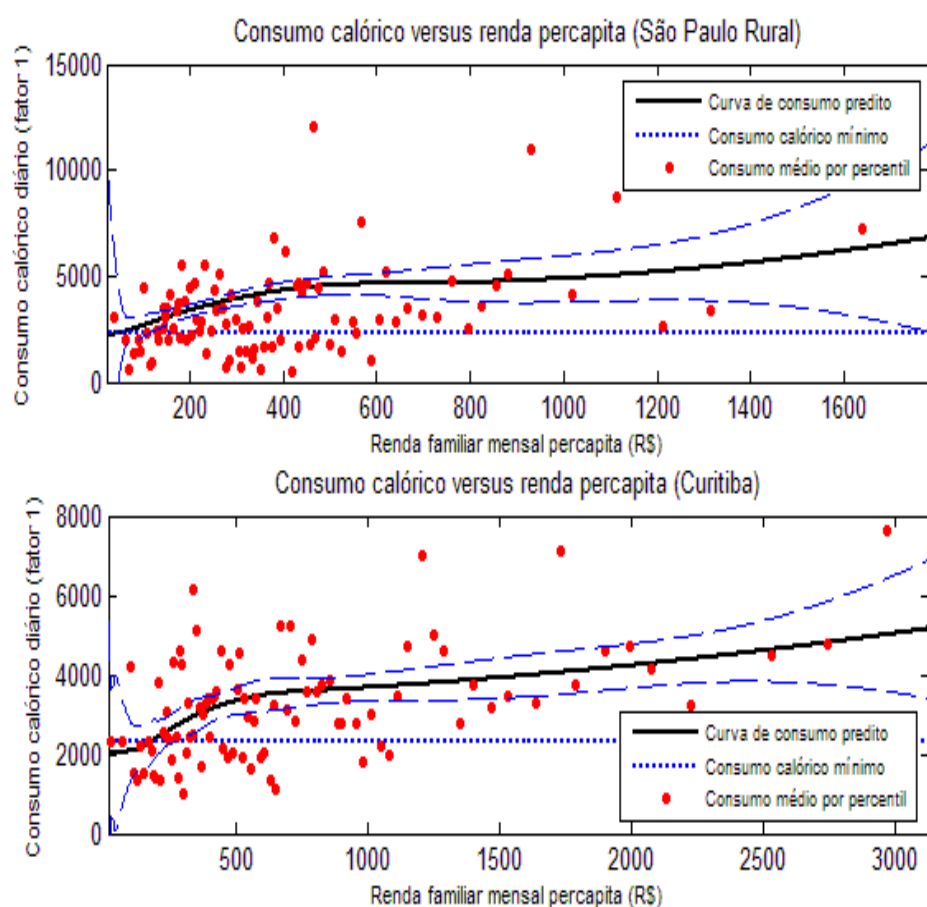
A6. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Rio de Janeiro e do Sudeste (excluindo São Paulo) Rural



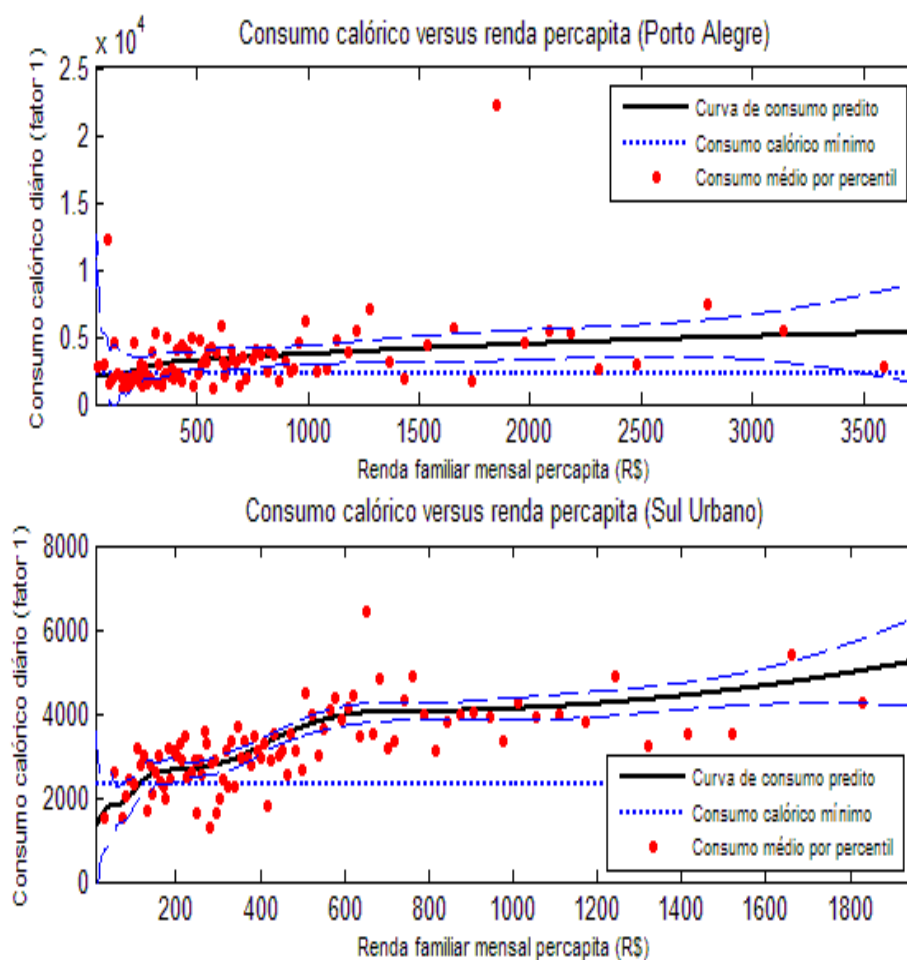
A7. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de São Paulo (Região Metropolitana) e São Paulo Urbano (excluindo a Região Metropolitana)



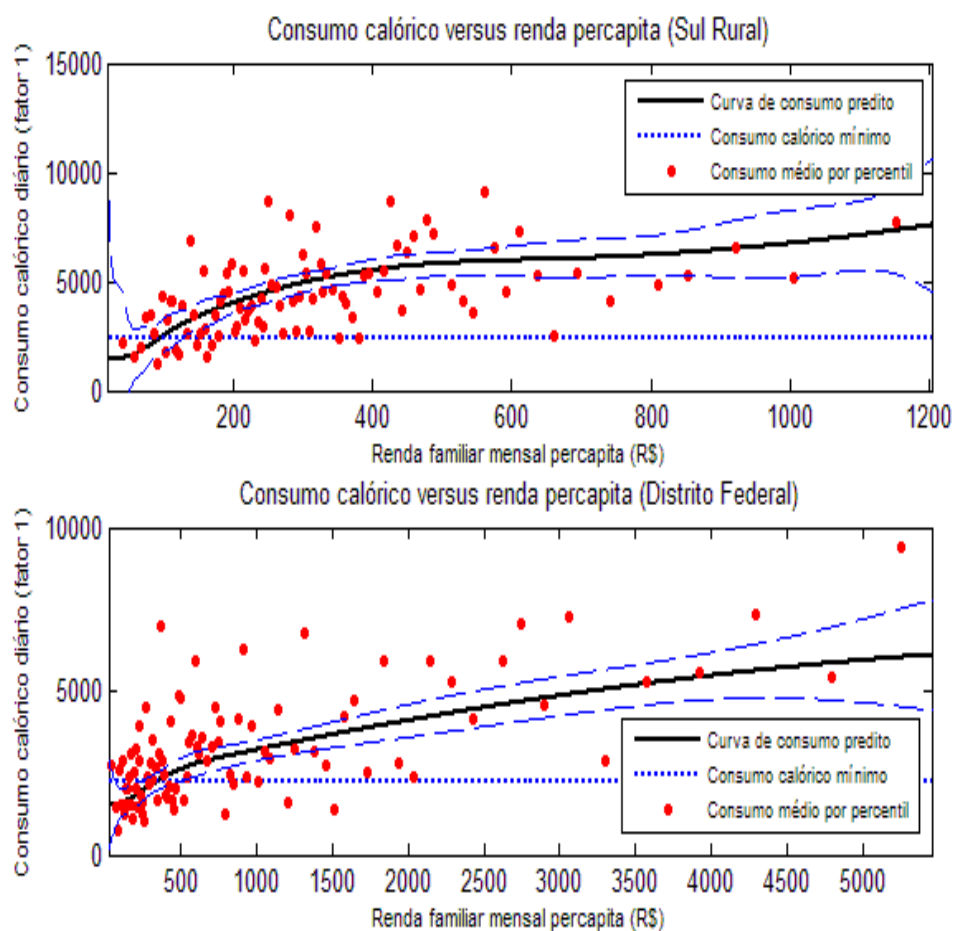
A8. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Curitiba e São Paulo Rural



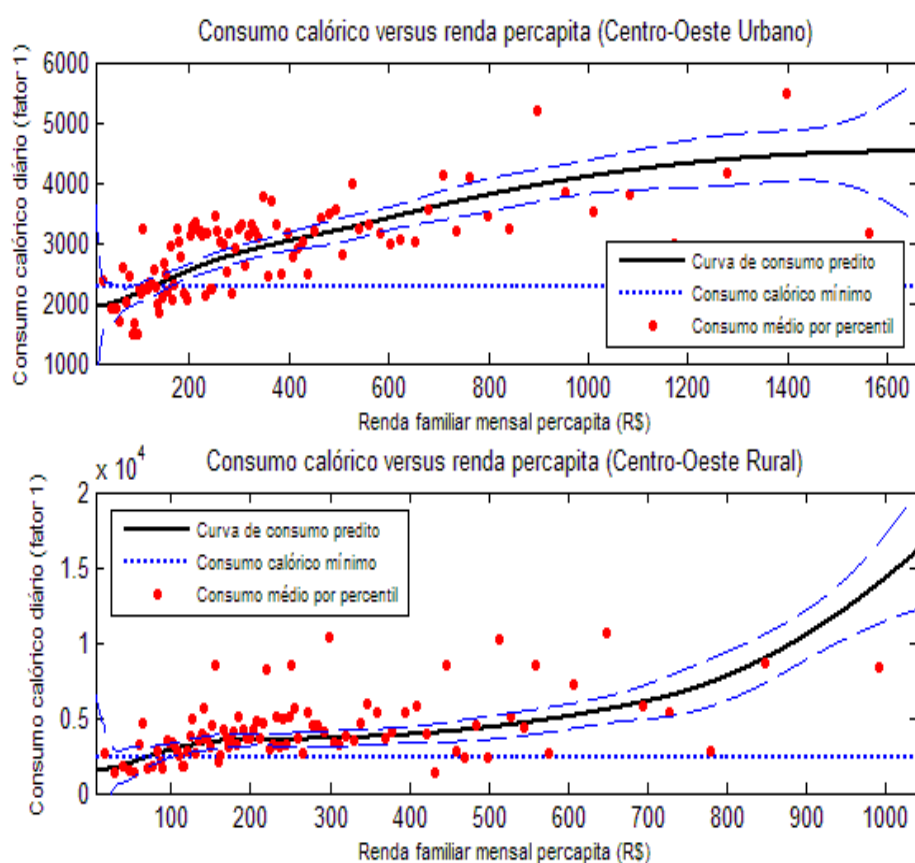
A9. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Porto Alegre e Sul Urbano



A10. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Sul Rural e do Distrito Federal



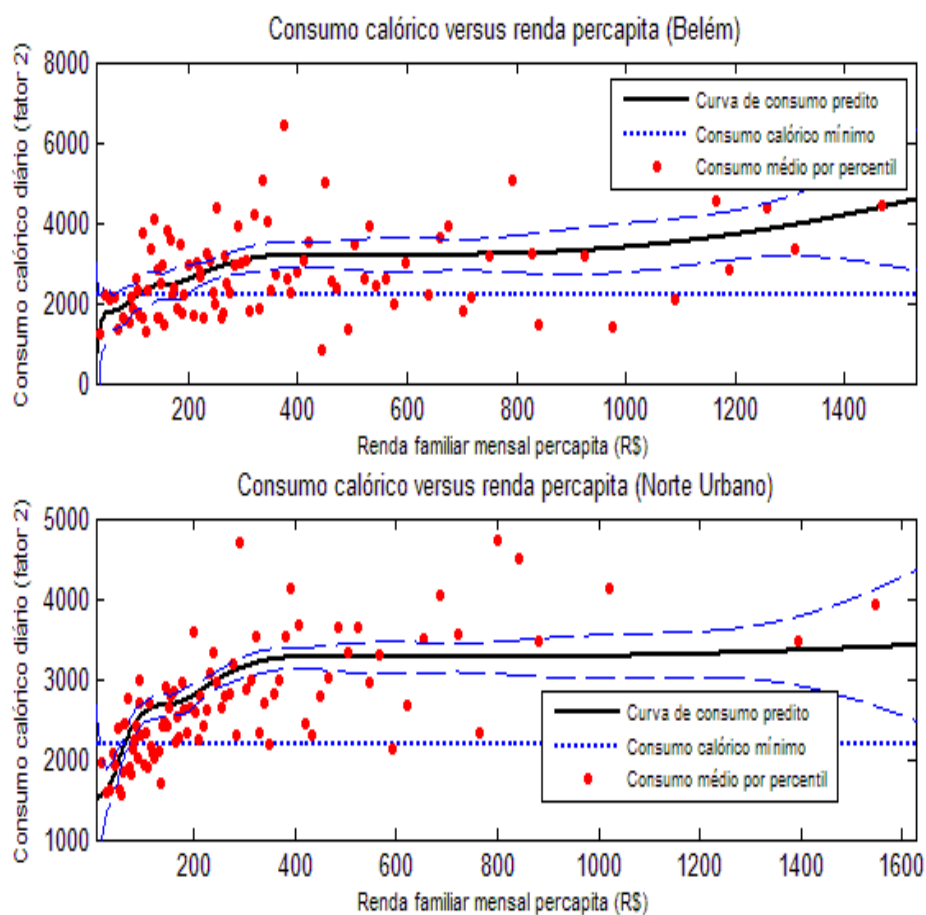
A11. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Centro-Oeste Urbano e do Centro-Oeste Rural



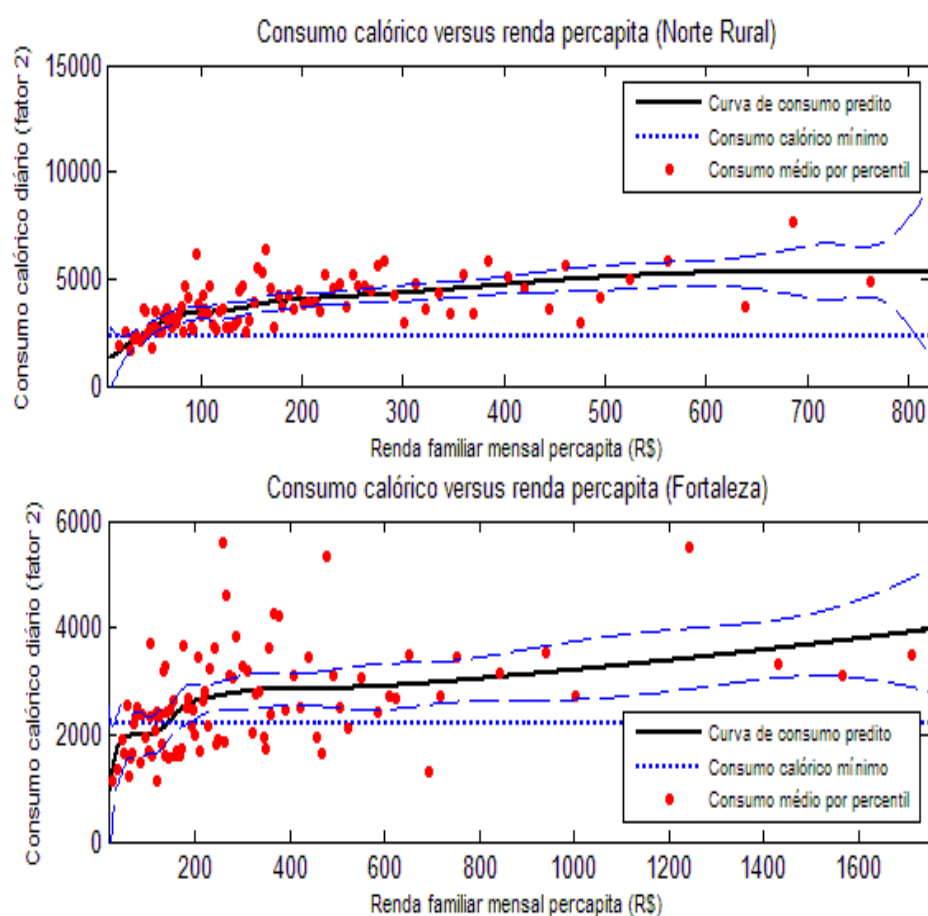
APÊNDICE B

VALOR DA CALORIA FORA DO DOMICÍLIO DUAS VEZES O VALOR DA CALORIA CONSUMIDA EM CASA.

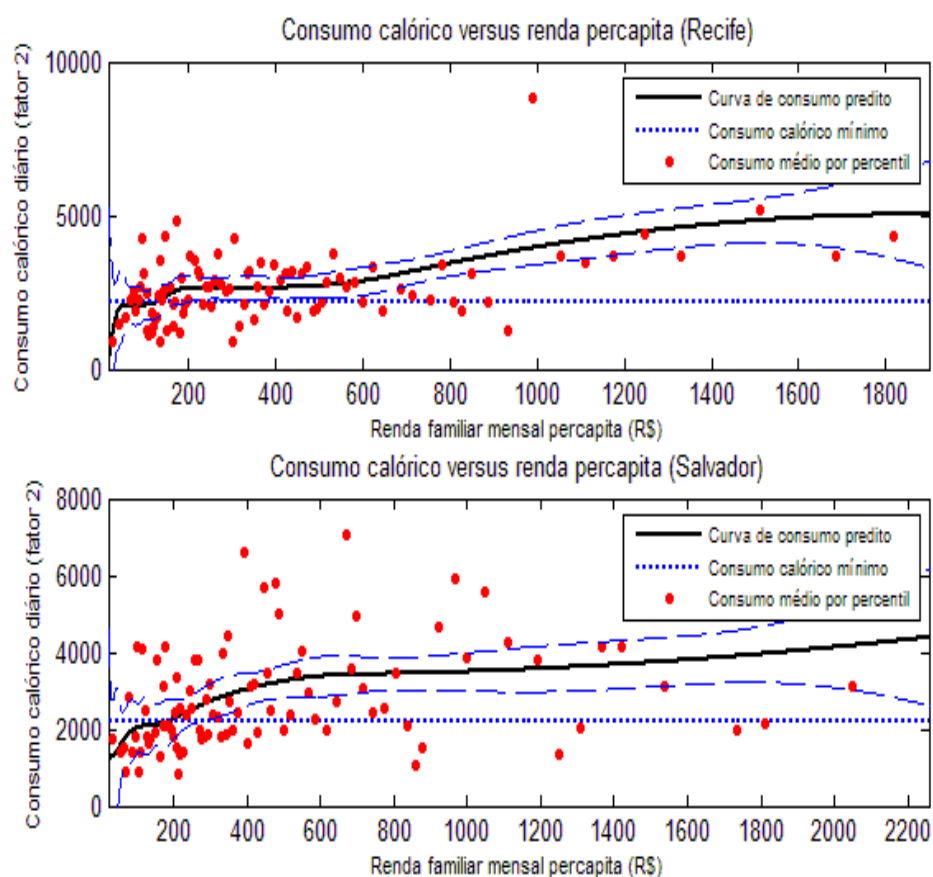
B1. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Belém e do Norte Urbano



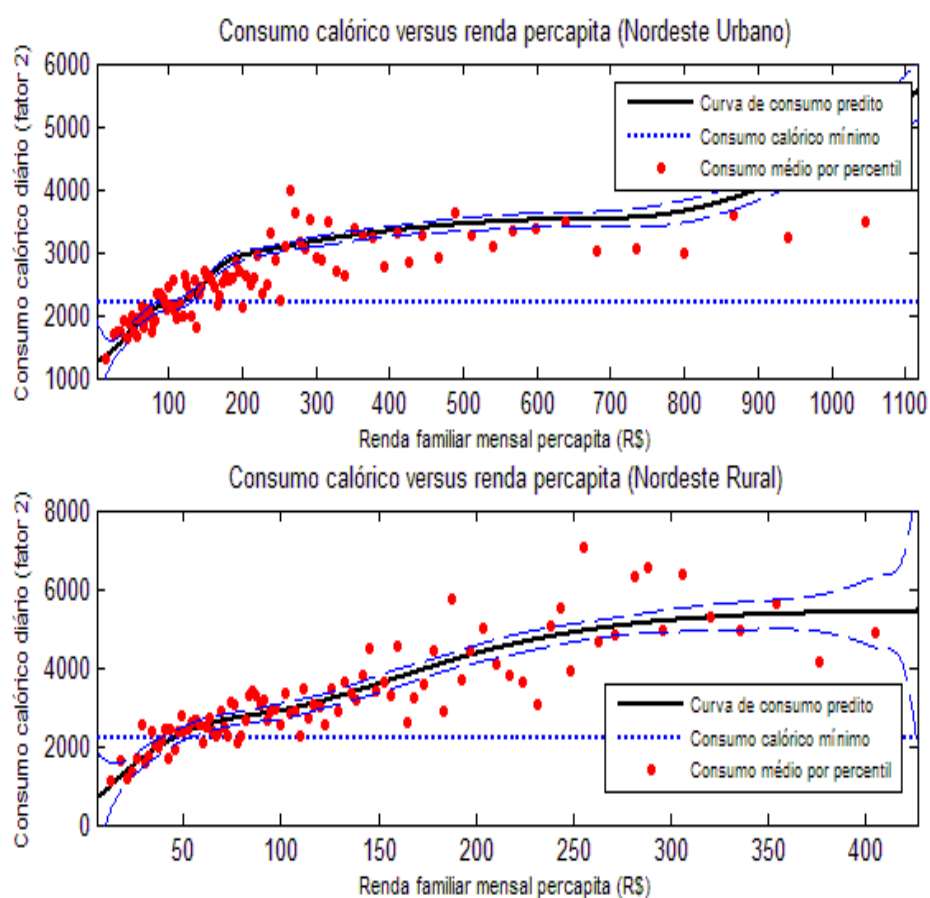
B2. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Norte Rural e de Fortaleza



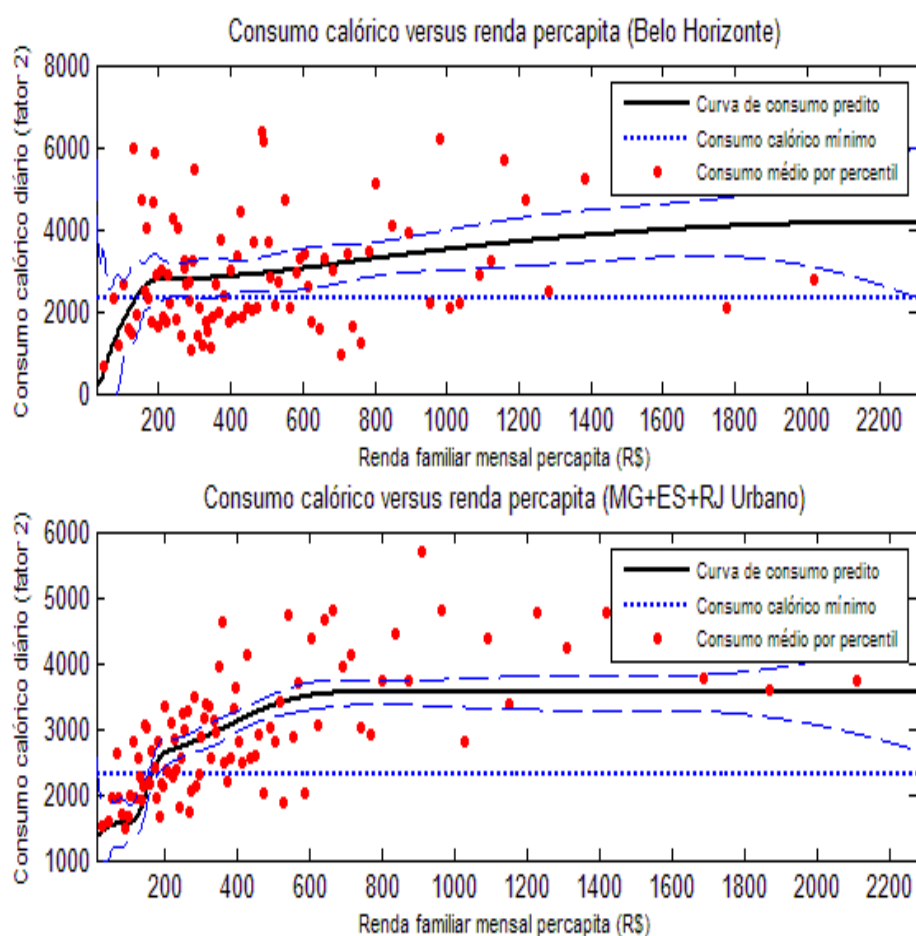
B3. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Recife e de Salvador



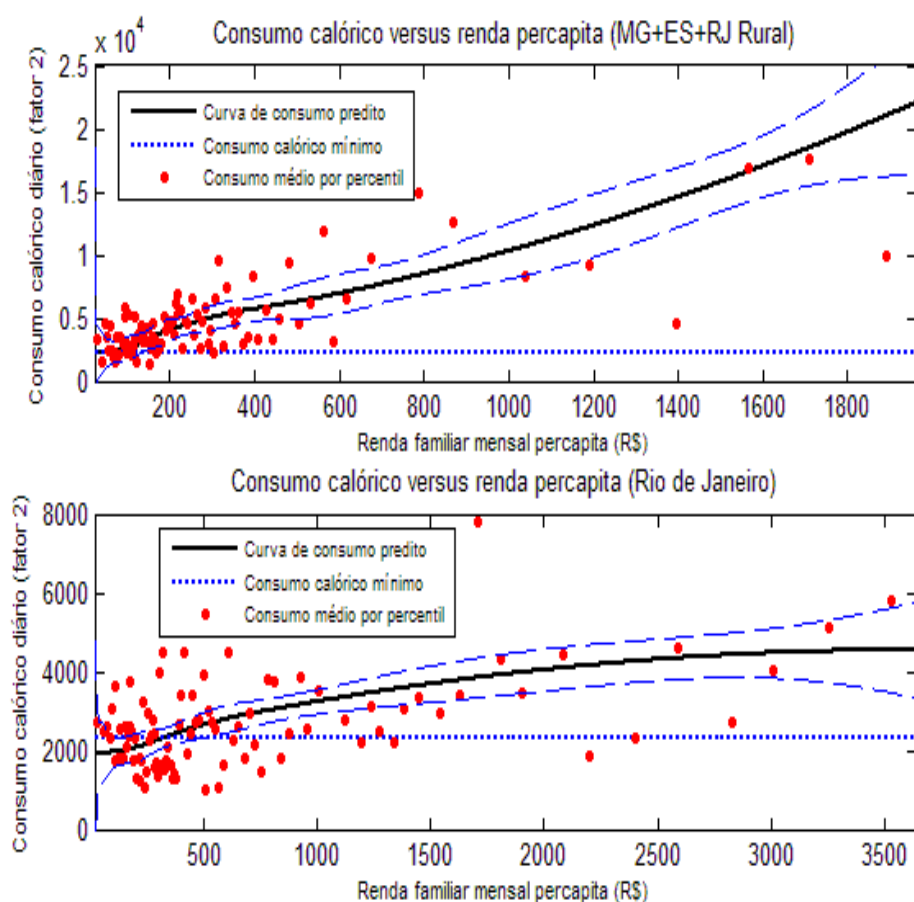
B4. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Nordeste Urbano e do Nordeste Rural



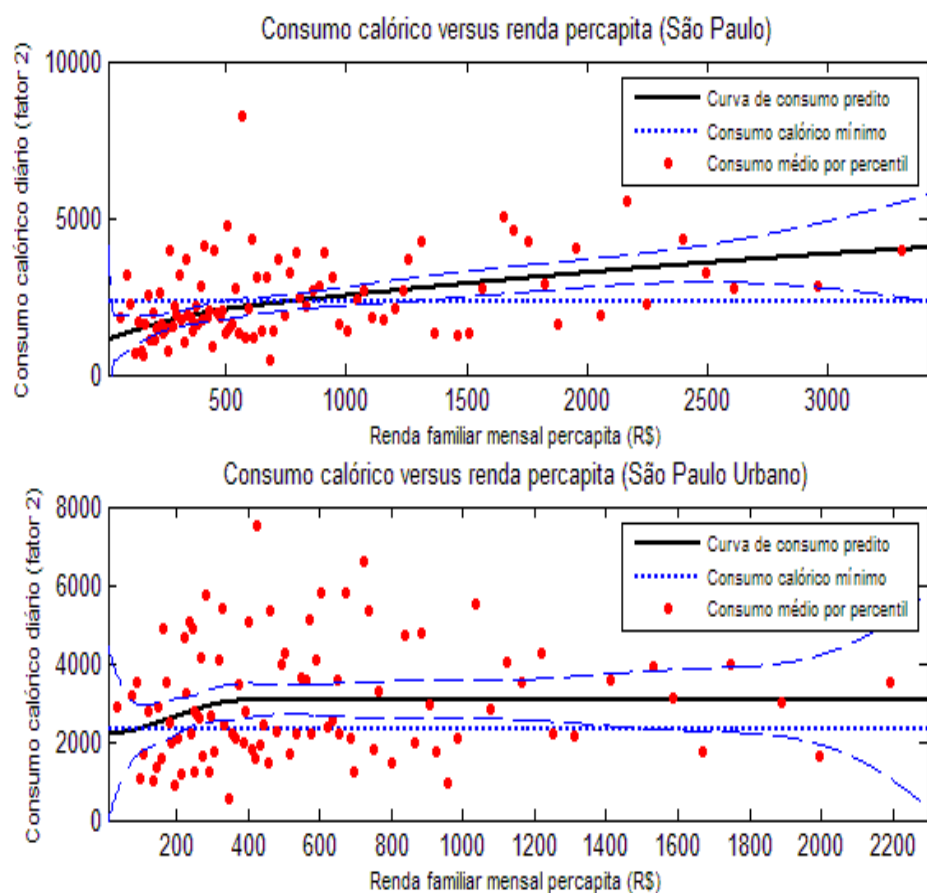
B5. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Belo Horizonte e do Sudeste (excluindo São Paulo) Urbano



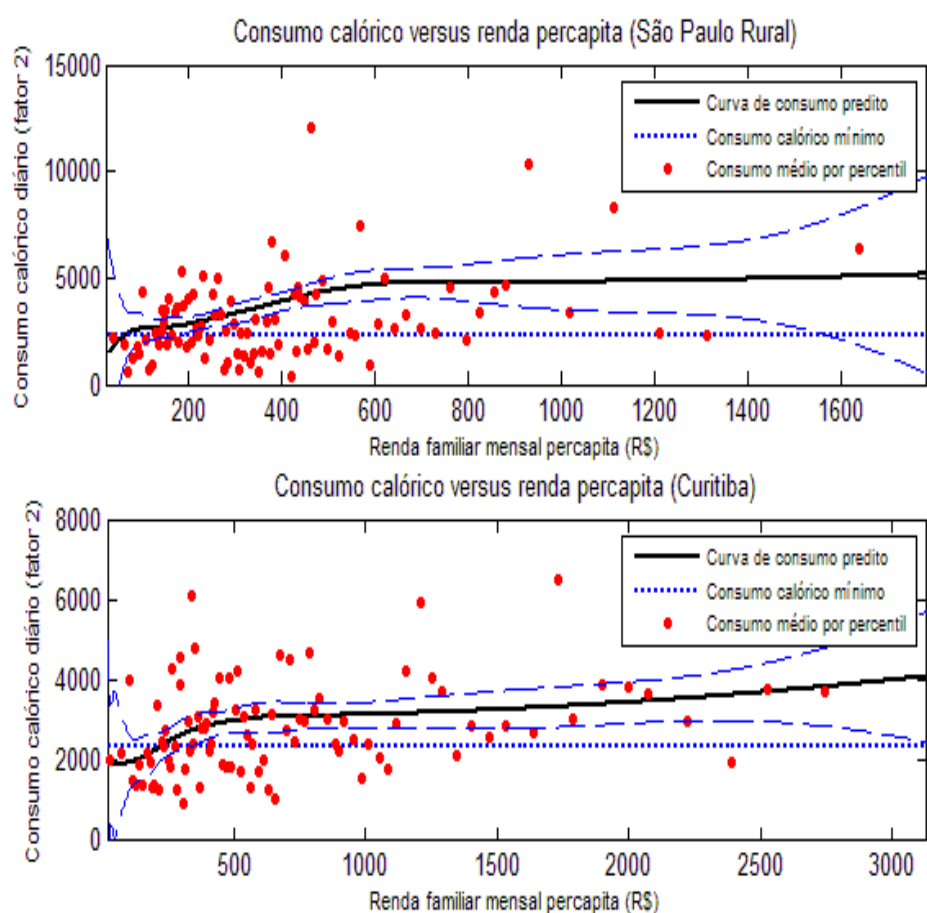
B6. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Rio de Janeiro e do Sudeste (excluindo São Paulo) Rural



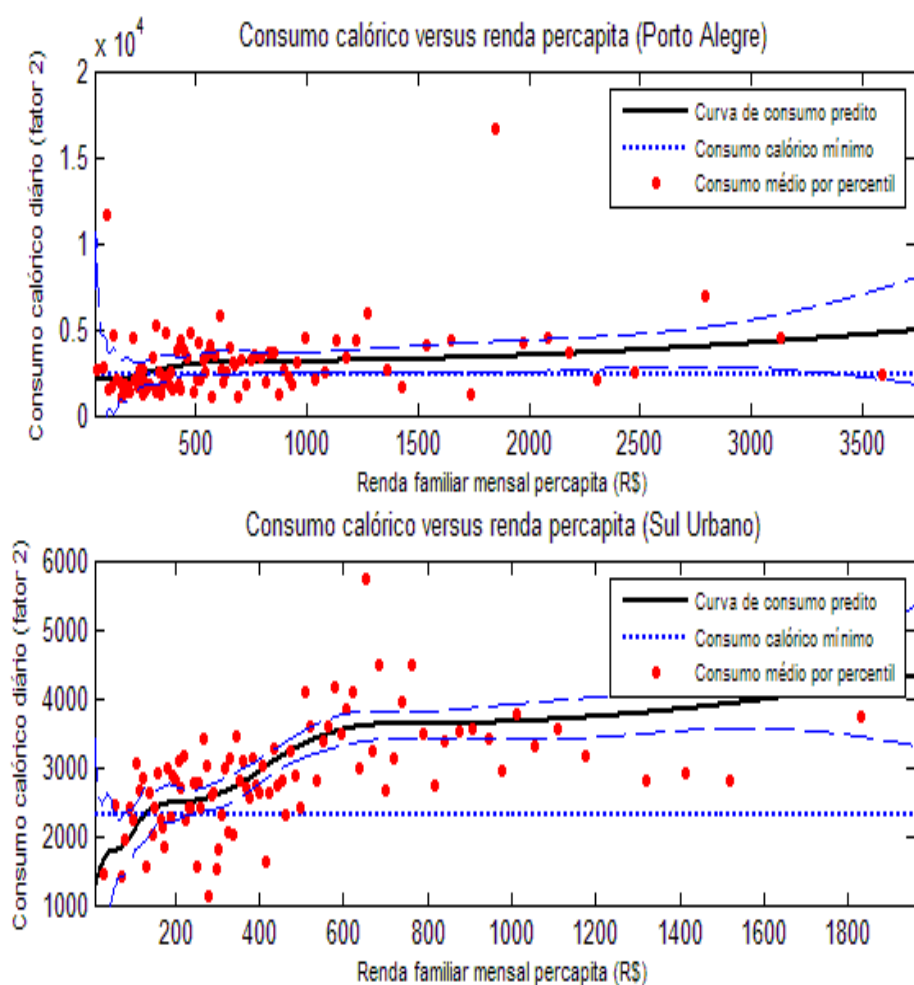
B7. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de São Paulo (Região Metropolitana) e São Paulo Urbano (excluindo a Região Metropolitana)



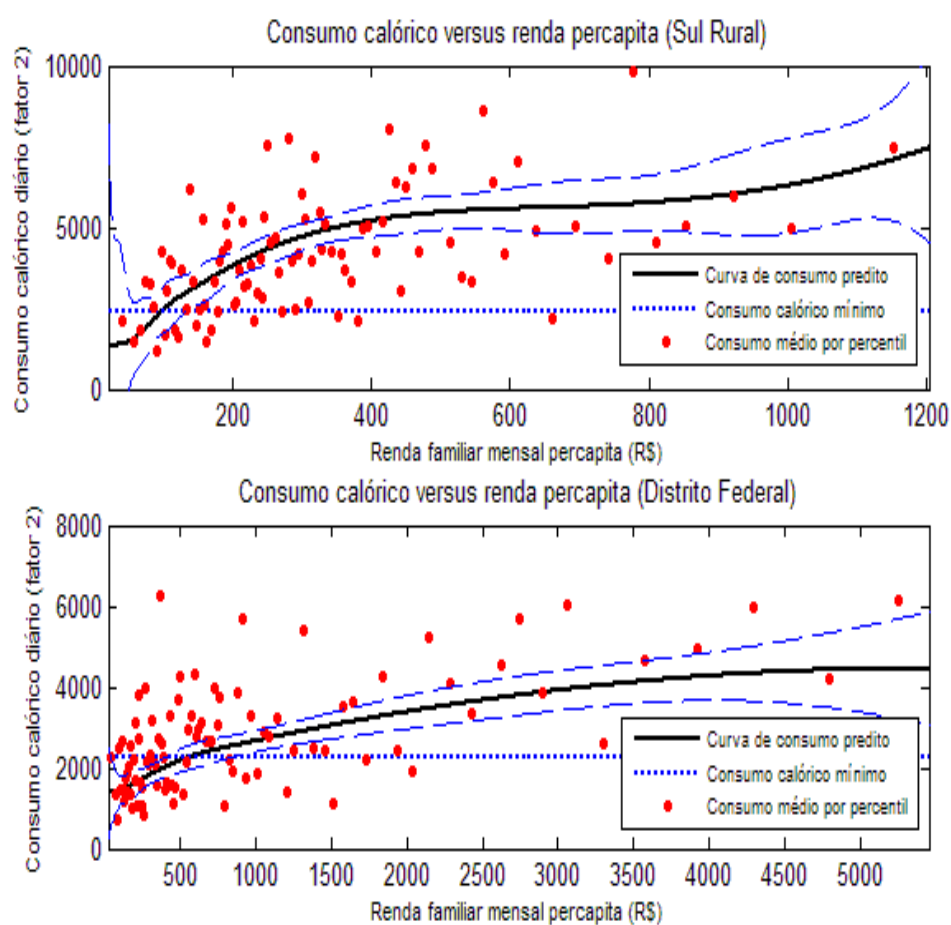
B8. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Curitiba e São Paulo Rural



B9. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos de Porto Alegre e Sul Urbano



B10. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Sul Rural e do Distrito Federal



B11. Gráficos de consumo calórico *versus* renda *per capita* para os contextos geográficos do Centro-Oeste Urbano e do Centro-Oeste Rural

