

Intervalos de confiança para previsões populacionais municipais baseadas em matrizes de crescimento inter-regional

RESUMO: Entre os desafios para a área de projeções demográficas estão a volatilidade do componente migratório – fundamental para a projeção de pequenas áreas; a compatibilização entre projeções de pequenas e grandes áreas de forma consistente; e a mensuração e inclusão da incerteza em cenários futuros de crescimento demográfico. O objetivo desta pesquisa é lidar com estes três desafios ao propor uma metodologia de previsão populacional baseada na variabilidade dos componentes atrelados ao crescimento vegetativo e migratório de pequenas áreas. O método de projeção proposto utiliza matrizes de crescimento inter-regionais propostas por Andrei Rogers (1968, 1971, 1975, 1995) combinando-as com técnicas de análise Bayesiana e métodos de simulação Monte Carlo para lidar com a incerteza futura das projeções. O método requer apenas três informações disponíveis nos Censos Demográficos: o município de residência anterior, no ano t-n; e as distribuições populacionais por município no ano t-n e t. Como resultado são obtidos intervalos de confiança estatística de 95% para as populações projetadas. Comparações ex-post, entre populações projetadas e observadas, também serão apresentadas para demonstrar a eficácia do método proposto. Utiliza-se como exemplo ilustrativo os dados do estado de São Paulo oriundos dos Censos Demográficos de 1980, 1991 e 2000.

Palavras-chave: projeções; migração; município; crescimento natural; análise espacial

Uma das principais demandas por demógrafos advêm da necessidade de se conhecer o número de pessoas que existirão em um determinado lugar e tempo futuros. Saber onde, quando e quantos indivíduos existirão é essencial para se conhecer o tamanho do mercado consumidor e se planejar a oferta de serviços de saúde, educação, transporte, fornecimento de água e luz, coleta de lixo e infraestrutura, pública e privada. Avanços científicos no instrumental técnico e teórico para se projetar (*project*) e prever (*forecast*) a população nos níveis global, nacional, regional e local tem contribuído para tornar as estimativas mais precisas e menos subjetivas (STOTO, 1983; PFLAUMER, 1988; LEE, 1992; LEE; TULJAPURKAR, 1994; LUTZ et al., 1999; ALHO; SPENCER, 2005; KEYFITZ; CASWELL, 2005). Três desafios, entretanto, ainda permeiam o debate na área de projeções. O primeiro está na dificuldade em se projetar a população de pequenas e pequeníssimas áreas. Isso se dá por causa da volatilidade do componente migratório e da dificuldade em modelá-lo e prevê-lo. O segundo desafio está atrelado à consistência e à integração de projeções feitas de maneira separada, mas que representam parte de um todo. Esse é o caso, por exemplo, de projeções populacionais de pequenas áreas que, quando agregadas, fornecem estimativas distintas daquelas realizadas para a área maior da qual fazem parte. Em teoria, esse problema é resolvido por tabelas multiestado, mas na

prática a aplicação do método é proibitiva quando há muitos “estados” envolvidos na projeção. Por fim, o terceiro desafio para se realizar projeções relaciona-se à incerteza futura e à falta de consenso metodológico para lidar com a mesma.

No Brasil, em particular, estes três desafios têm sido pouco investigados¹. A principal meta desta pesquisa é contribuir para o debate na área de previsões populacionais através do desenvolvimento e teste de uma metodologia alternativa para se prever o crescimento demográfico dos municípios paulistas. O método sugerido tem a virtude de integrar consistentemente as projeções de áreas menores e maiores e, ao mesmo tempo, lidar com a incerteza inerente à dinâmica demográfica futura.

Sobre Projeções Demográficas no Brasil

No Brasil, o método de projeção dominante é o das componentes, que trata separadamente a influência da fecundidade, mortalidade e migração nas taxas de crescimento demográfico. O método foi desenvolvido por Whelpton (1928) e formalizado por Leslie (1945; 1948). Ele baseia-se no pressuposto de que a desagregação do crescimento demográfico em componentes conduz à maior precisão das estimativas populacionais. Essa metodologia, utilizada pelo IBGE, pelo Cedeplar e pela Fundação Seade desde 1970, consiste em observar o comportamento das taxas de fecundidade, mortalidade e migração e, a partir delas, gerar cenários futuros de crescimento para a área-alvo a ser projetada. O método possui dois méritos: a) ser consistente e flexível para lidar com hipóteses alternativas de crescimento e com desagregações por idade, sexo, cor, ou quaisquer outros cortes populacionais de interesse; e b) ser adaptável para levar em conta extrapolações e equações estruturais para predizer o comportamento das componentes. Essa última prática, entretanto, é pouco usual no Brasil. Por outro lado, o método das componentes demográficas é inapropriado para a estimação populacional de pequenas áreas (ex. municípios), já que neste nível de análise as estimativas de fecundidade e mortalidade podem se tornar temporalmente instáveis e de difícil mensuração

1 Waldvogel (1996), Figoli (2000), Figoli et al. (2003), Januzzi (2007), Brito, Cavenaghi e Jannuzzi (2010) e Santos e Barbieri (2015) são algumas exceções.

(FREIRE; ASSUNÇÃO, 1998; CAVENAGHI et al., 2004). Por isso, a solução oficialmente adotada por centros de pesquisa consiste em projetar a população para áreas maiores (ex. unidades da federação) e posteriormente efetuar uma distribuição proporcional ou interpolativa para áreas menores (UNITED NATIONS, 1956; DUCHESNE, 1987; SMITH et al., 2001). No caso da Fundação Seade, “o modelo de projeção considerado adota uma hierarquia que parte da projeção para o total do Estado e se desagrega em regiões administrativas e municípios”², sem considerar as especificidades das trocas migratórias intermunicipais que, em alguns casos, correspondem a uma parcela significativa do crescimento demográfico. Além disso, o crescimento da subpopulação fica determinado pelo da população total (SZWARCOWALD; CASTILHO, 1989).

Os modelos tradicionalmente usados para as pequenas áreas, como o $A_i B_i$ (UNITED NATIONS, 1956; MADEIRA; SIMÕES, 1972), o método “quase-componente” (HAKKERT, 1985), e o ProjPeq (JANUZZI, 2007), não captam a interrelação existente entre pequenas áreas. Para tal, o mais apropriado seria uma abordagem multirregional, capaz de “espelhar as especificidades das trocas populacionais interestaduais, inter-regionais e, particularmente, intra-regional” (BAENINGER; CUNHA, 1996). Como existem diferenças socioeconômicas significativas ao longo do espaço, e como estas diferenças tendem a influenciar a dinâmica demográfica de cada área, a previsão populacional a partir do município de residência é justificável, sobretudo quando o componente migratório possui um papel fundamental na dinâmica de crescimento de áreas menores. Através de previsões populacionais multirregionais consideram-se as heterogeneidades demográficas espaciais e, supostamente, melhora-se a qualidade e precisão das previsões, já que o método assegura consistência entre as projeções dos vários níveis de agregação. Além de considerar as especificidades demográficas de cada região ao permitir que elas tenham suas próprias taxas de crescimento natural e migração, o método multirregional também garante que a projeção das regiões menores será compatível e consistente com a projeção esperada para a região maior. Nesse sentido, a modelagem

2 A metodologia utilizada pela Fundação Seade para projetar a população do estado e dos municípios paulistas está disponível em <http://produtos.seade.gov.br/produtos/projpop/pdfs/projpop_metodologia.pdf>. Acesso em 07 de dez. 2015.

multirregional pode ser entendida como a extensão dinâmica e multidimensional do método das componentes³.

Outro desafio na área de projeções se refere à trajetória futura das componentes que afetam o crescimento demográfico. Para decidir o caminho a ser seguido pela fecundidade, mortalidade e migração, demógrafos têm adotado quatro alternativas (BOOTH, 2006). A primeira consiste em extrapolar tendências históricas e assumir que o comportamento futuro será uma continuação da tendência passada usando metodologias de séries temporais (LEE; CARTER, 1992; LEE; TULJAPURKAR, 1994). A segunda alternativa consiste em definir hipóteses e trajetórias futuras para o período da projeção utilizando o conhecimento e intuição de especialistas. Tipicamente definem-se cenários de crescimento baixo, médio e alto e a partir daí assume-se que o valor esperado estará dentro do intervalo subjetivamente definido (UNITED NATIONS, 1973; WALDVOGEL et al., 2003; IBGE et al., 2004). A terceira consiste em modelar o comportamento dos componentes através do uso de covaráveis exógenas relacionadas. Esse método, entretanto, só se torna superior aos demais quando a capacidade de prever o comportamento futuro das covariáveis relacionadas é maior do que a capacidade de prever o comportamento das variáveis demográficas propriamente ditas (KEYFITZ, 1981; 1982). Por fim, a quarta alternativa consiste em considerar a incerteza através da simulação de trajetórias futuras (PFLAUMER, 1988; LEE, 1992; LEE; TULJAPURKAR, 1994; LEE, 1998; FIGOLI, 2000). O método consiste em simular trajetórias da população através da seleção aleatória de parâmetros e variáveis relacionadas ao crescimento populacional dentro de uma dada distribuição estatística. Para tanto, as estatísticas vitais são geradas pelo método de simulação de Monte Carlo, conforme uma distribuição de probabilidade assumida. De acordo com Figoli (2000), suas vantagens são: “a de ser conceitualmente simples, a de fornecer os resultados a nível da distribuição por idade

3 A abordagem multiestado foi primeiramente introduzida por Andrei Rogers (1968) para estudar o crescimento inter-regional e, posteriormente, desenvolvida por Rogers (1971, 1975, 1995) e Schoen (1988, 2006). Na projeção multiestado, a transição entre diferentes “estados” (ex. regiões, estado conjugal, participação na força de trabalho, condição de saúde) pode ser modelada de acordo com a especificidade de cada estado (BOOTH, 2006). Nos Estados Unidos e na Europa há várias aplicações deste método, mas no caso brasileiro a abordagem multiestado é menos comum. Poucas exceções são os estudos de Machado (1993), em nível nacional, e Fígoli et al (2003), em nível estadual, para projetar a população por idade e sexo utilizando dados do Censo Demográfico de 1991. Muniz (2012) adotou a abordagem multiestado para projetar e simular o crescimento das populações das classes baixa, média e alta, e para mensurar como o crescimento destes três grupos poderiam vir a afetar os níveis de desigualdade de renda no Brasil.

da população e a de permitir o estabelecimento do intervalo de confiança para todos os resultados da projeção.” Esta abordagem aproxima-se da metodologia Bayesiana, tida como a vanguarda dos métodos de previsão populacional (GUIMARÃES, 2014; WISNIOWSKI et al., 2015).

Este artigo introduz um método de previsão populacional capaz de integrar as especificidades locais da dinâmica demográfica municipal, combinando equações lineares bayesianas, variabilidade espacial e métodos de simulação Monte Carlo. O principal resultado são intervalos de confiança para a população do estado e dos municípios paulistas que levam em conta a incerteza futura dos parâmetros associados à dinâmica demográfica. O método desenvolvido utiliza os municípios paulistas, mas poderia também ser aplicado a outros níveis de agregação geográfica e cortes populacionais por idade, sexo, raça, escolaridade etc.

OBJETIVOS

Esta pesquisa tem três objetivos interligados e interdependentes. O **primeiro** é construir matrizes multirregionais de crescimento populacional para os municípios paulistas. Estas matrizes levam em conta as trocas migratórias intermunicipais e o crescimento natural específico de cada município. O **segundo** objetivo caracteriza-se pela identificação, construção e estimação de modelos econométricos preditivos das probabilidades de migração intermunicipal e de crescimento natural (fecundidade e mortalidade). Estes modelos mensuram a associação entre a dinâmica migratória e as taxas de crescimento natural passadas e futuras utilizando métodos bayesianos. Por fim, o **terceiro** objetivo é demonstrar uma metodologia simples e integrada para projetar (e prever) a população dos municípios de São Paulo levando em conta a incerteza inerente ao processo de projeção. Os objetivos descritos são desenvolvidos em quatro módulos marcados por metodologias integradas e contribuições específicas.

DADOS E MÉTODOS

- **Módulo 1. Construção da matriz de projeção multirregional**

A matriz de projeção multirregional integra a fecundidade, mortalidade e mobilidade de diferentes municípios em um instrumento analítico único. Seguindo as ideias de Rogers (1968), a matriz do sistema de projeção multirregional possui dois componentes: i) os elementos na diagonal principal da matriz representam o efeito combinado da fecundidade, mortalidade e emigração de cada município; e ii) os elementos fora da diagonal principal representam as probabilidades de imigração intermunicipal dos municípios j para os municípios i . Simbolicamente, para um intervalo de projeção de cinco anos, a matriz de crescimento inter-regional é representada pela relação:

$$w_{proj}^{t+5} = (B - D + P') * w^t = G * w^t \quad (1)$$

sendo w^t = vetor-coluna cujos elementos denotam a população da região i no tempo t ;

B, D = matrizes-diagonal cujos elementos são não-negativos e representam as taxas brutas de natalidade e mortalidade dos municípios;

P' = transposta da matriz de probabilidade de migrar da região de origem para a região de destino. A matriz quadrada P é composta por elementos p_{ij} , que descrevem a proporção de pessoas que durante um momento específico migram da região j para a região i ;

$G = (B - D + P') =$ matrix de crescimento inter-regional.

Os elementos de P' são calculados utilizando-se a informação sobre o migrante de “data fixa” contida na amostra dos Censos Demográficos de 1991 e 2000. A pergunta sobre o município de residência anterior, cinco anos antes do censo, permite o cálculo das probabilidades de migração intermunicipal (CUNHA, 2005). Para isso, basta estimar uma matriz de fluxos origem-destino e dividir cada um dos elementos da mesma pela soma dos elementos da linha correspondente⁴. Já os elementos das matrizes B e D serão combinados em uma única matriz-diagonal para representar o

⁴ Como os dados de migração resultam de estimativas amostrais, municípios menores tenderão a apresentar probabilidades temporalmente instáveis, sobretudo ao desagrega-las por subgrupos ainda menores, por sexo e idade. Tal variabilidade seria resolvida utilizando-se áreas maiores (ex. microrregiões) ou suavizando-se as probabilidades a partir das taxas de migração das áreas do entorno. Estes procedimentos, no entanto, não serão adotados neste artigo já que o componente migratório é nulo em 93% dos municípios ou representa uma parcela muito pequena do crescimento demográfico observado.

crescimento natural e a emigração dos municípios selecionados na análise. Cada elemento da diagonal principal da matriz G representa o efeito combinado da fecundidade, mortalidade e emigração da região i . O cálculo destes elementos é feito de maneira residual, considerando a diferença entre a população projetada pela matriz P' , $w_{proj}^{t+5, i}$, e a população observada em $w^{t+5, i}$. O crescimento referente à migração dos menores de 5 anos e dos filhos de migrantes também está incluído nos valores da diagonal principal da matriz de crescimento multirregional.

Para ilustrar o funcionamento do sistema de crescimento multirregional, considere a seguinte matriz fictícia de probabilidades de migração entre São Paulo e o resto do Brasil:

Matriz de Transição de Fluxos Migratórios: 1985-90 (em milhões)

De \ Para	São Paulo	Resto do Brasil	População Total em 1985	População Total em 1990
São Paulo (p_{ii}, p_{ij})	26.528 (0,9373)	1.775 (0,0627)	28.303 (1,0000)	30.783
Resto do Brasil (p_{ii}, p_{ij})	1.345 (0,0127)	104.580 (0,9873)	105.925 (1,0000)	113.308

Nota: Adaptou-se as probabilidades para São Paulo a partir de Rogers (1968).

Utilizando-se os dados acima para compor a matriz P' , sem considerar o efeito combinado da fecundidade, mortalidade e emigração, a população em 1990 seria dada por:

$$w_{proj}^{1990} = P' * w^{1985} = \begin{bmatrix} 0,9373 & 0,0127 \\ 0,0627 & 0,9873 \end{bmatrix} \begin{bmatrix} 28.303 \\ 105.925 \end{bmatrix} = \begin{bmatrix} 27.874 \\ 106.354 \end{bmatrix} \quad (2)$$

Os valores da matriz P' mostram que a probabilidade de estar vivendo em São Paulo em 1985 e em 1990 era de 93%, e a probabilidade de ter imigrado de alguma parte do Brasil entre 1985 e 1990

para São Paulo era de 1,27%. Note que como a matriz P' não leva em conta o crescimento vegetativo do período e nem a migração dos menores de cinco anos, a população projetada é menor do que a população de fato observada. Após ajustar os valores da diagonal principal para que a população observada em 1990 seja atingida, tem-se:

$$w^{1990} = G * w^{1985} = \begin{bmatrix} 1,0401 & 0,0127 \\ 0,0627 & 1,0529 \end{bmatrix} \begin{bmatrix} 12.988 \\ 152.082 \end{bmatrix} = \begin{bmatrix} 30.783 \\ 113.308 \end{bmatrix} \quad (3)$$

A matriz G pode então ser utilizada para se realizar previsões populacionais futuras, para se medir a contribuição dos imigrantes para o crescimento específico de cada região, para a análise de populações estáveis, e para se calcular a sensibilidade do crescimento populacional a cada um dos elementos da matriz. Este exemplo simplificado ilustra, portanto, como a mortalidade, fecundidade e mobilidade inter-regionais podem ser combinados para se descrever o crescimento inter-regional, tornando explícito o papel das trocas migratórias.

A matriz de crescimento para se projetar o tamanho futuro dos municípios paulistas tem estrutura similar à descrita acima, mas o número de linhas e colunas da matriz G é igual ao resto do Brasil mais os 572 municípios existentes no estado em 1991. Os 73 novos municípios criados durante a década de 90 foram reagrupados às suas fronteiras políticas iniciais para permitir a comparabilidade geográfica entre os dados dos Censos de 1991 e 2000. A matriz de crescimento gerada em 1991 (G^{1991}) utiliza os dados dos Censos Demográficos de 1991 e 1980, enquanto a matriz de crescimento inter-regional de 2000 (G^{2000}) utilizaria os dados de 1991 e 2000.

- **Módulo 2. Modelagem da migração intermunicipal e do crescimento natural**

O comportamento migratório de cada município paulista é modelado a partir da correlação entre migração passada e futura. Espera-se que a probabilidade passada de migração seja a melhor preditora da probabilidade de se migrar no futuro (GIROSI; KING, 2008). Estima-se regressões lineares bayesianas para garantir que os coeficientes resultem da própria distribuição observada e não de

supostas variações amostrais (como na inferência estatística frequentista). Os coeficientes obtidos poderiam então ser utilizados para prever as probabilidades mínimas e máximas de migração intermunicipal dentro de um intervalo “crível” de confiança estatística de 95%.

Os elementos da diagonal principal da matriz de crescimento inter-regional também serão modelados de acordo com a tendência do crescimento natural entre 1991 e 2000. Estes modelos tendenciais, ainda que simplistas, são factíveis para descrever o crescimento vegetativo municipal, sendo também os mais plausíveis na avaliação de 23 dos 30 especialistas brasileiros consultados em uma pesquisa de opinião realizada por Januzzi (2007, p. 118-119). Os coeficientes bayesianos subsidiam o módulo de projeção e previsão populacional descrito a seguir.

- **Módulo 3. Previsão multirregional da população dos municípios paulistas**

As previsões populacionais possuem dois tipos de incerteza: a primeira associada a amostra censitária utilizada para se calcular as células da matriz de crescimento. Em áreas pequenas pode haver grande instabilidade das probabilidades calculadas. Já a segunda incerteza está atrelada ao valor futuro das células da matriz de crescimento. Para lidar com estes dois tipos de incerteza utilizamos métodos bayesianos para incorporar a própria distribuição de probabilidade das células da matriz de crescimento, permitindo que as mesmas variem segundo o desvio padrão dos elementos específicos dentro e fora da diagonal principal da matriz de projeção.

Em um primeiro momento utiliza-se a matriz gerada em 1991 (G_{1991}) para se projetar a população até 2000. Esse cenário assume que as condições demográficas permanecem as mesmas ao longo da década mantendo-se os elementos da matriz G_{1991} constantes. Esse exercício, apesar de irrealista, é instrutivo para se avaliar como a população de cada município e do estado crescerá se as condições demográficas do passado permanecessem as mesmas.

Em um segundo momento, considera-se a incerteza futura na geração das projeções populacionais entre 1991 e 2000 utilizando-se amostras aleatórias de parâmetros de uma dada distribuição para explorar o comportamento de processos complexos. O método cria milhares de

cenários de projeção demográfica através de mudanças semialeatórias nos elementos da matriz de crescimento, permitindo a obtenção de resultados intervalares de projeção, ou previsões populacionais. Os coeficientes e variáveis preditas para cada período de projeção são gerados conforme a distribuição prévia de probabilidade das variáveis consideradas pelos modelos de regressão propostos.

Considere que para se projetar a população de 1990 para 1995, e de 1995 para 2000, duas matrizes de crescimento multirregional são necessárias. A primeira matriz, G^{1991} , ao ser multiplicada por um vetor-coluna contendo a população dos municípios paulistas em 1990, produz a população projetada em 1995. Realizada desta forma, a projeção assume que o crescimento vegetativo relativo e as probabilidades de mobilidade entre 1990 e 1995 são iguais às observadas entre 1986 e 1991. Para flexibilizar este pressuposto e adicionar alguma incerteza à projeção futura, um elemento de aleatoriedade é adicionado aos valores dos elementos fora (p_{ij}^{1991}) ou dentro (c_i^{1991}) da diagonal da matriz de crescimento calculando-se:

$$p_{ij}^{1991} = p_{ij}^{1991} \varpi, \text{ sendo } \varpi \sim N(1, \sigma) \text{ e aleatório} \quad (4)$$

$$c_i^{1991} = c_i^{1991} \varphi, \text{ sendo } \varphi \sim N(1, \delta) \text{ e aleatório} \quad (5)$$

onde σ e δ são os desvios-padrão dos p_{ij}^{1991} ou c_i^{1991} elementos observados na matriz G^{1991} .⁵ Realizando-se a operação em (4) e (5) por cerca de 1000 vezes, gera-se distribuições de p_{ij}^{1991} e c_i^{1991} aproximadamente normais com médias p_{ij}^{1991} e c_i^{1991} para cada elemento fora ou dentro da diagonal principal da matriz G^{1991} . Para se gerar matrizes de projeção para o período 1990-1995 combinam-se então as matrizes aleatórias, independentes e simultâneas das distribuições de p_{ij}^{1991} e c_i^{1991} . Como resultado deste exercício, 1000 matrizes de crescimento, e conseqüentemente 1000 projeções demográficas probabilísticas são geradas para o período seguinte. Estas matrizes são então novamente

⁵ Esse pressuposto garante que a variabilidade de cada elemento da matriz de crescimento G_{1991} será proporcional ao desvio-padrão das probabilidades de migração municipal observadas no estado de São Paulo.

multiplicadas pelas várias populações projetadas em 1995 para obter-se a previsão populacional para 2000⁶. Mede-se os erros das projeções e da previsão populacional comparando-as com as populações paulistas municipais observadas no Censo de 2000, que servem como referencial para se avaliar a qualidade da previsão média realizada de 1991 para 2000.

RESULTADOS

Descrição das matrizes inter-regionais de crescimento demográfico

As matrizes de projeção multirregional consideram as trocas migratórias intermunicipais e o crescimento natural específico de cada município. As primeiras 572 linhas e colunas correspondem aos municípios existentes em 1991 e 2000 (compatibilizados para 1991), e a 573ª linha/coluna representa as trocas migratórias entre os municípios paulistas e o resto do Brasil. Para a matriz origem-destino de 1991, o maior valor da diagonal (crescimento natural + emigração) foi igual a 1,397. Isso indica que houve um aumento populacional de 39,75% no município de Santana do Parnaíba entre 1986 e 1991. O menor valor da diagonal foi observado no município de Santa Mercedes (0,658). Esse valor indica que 65,8% da população observada em 1986 neste município se encontrava neste mesmo município em 1991. Já no que se refere aos valores fora da diagonal, do total de células presentes na matriz (328.329), cerca de sete por cento (23.181) possuía valores diferentes de zero. Outras medidas descritivas das células contidas dentro e fora da diagonal principal das matrizes de origem-destino de 1991 e 2000 são apresentadas na tabela abaixo:

⁶ Os elementos das matrizes de crescimento incluindo incerteza, G^{1991}_u e G^{1996}_u são gerados a partir da matriz do ano base, G^{1991} . Em 1980 a informação sobre migrante de data fixa não estava disponível, e por isso a construção de uma matriz de crescimento G^{1980} não é possível.

TABELA 1.

Descrição dos elementos das matrizes inter-regionais de origem-destino dos municípios paulistas, 1991 e 2000

Localização das células da matriz de crescimento	Matriz origem-destino	
	1991	2000
Diagonal principal (crescimento natural + emigração)		
Número de células	573	573
Média	0,9345	0,9243
Desvio-padrão	0,0909	0,0996
Mediana	0,9462	0,9324
Máximo	1,3976	1,1852
Mínimo	0,6588	0,4220
Fora da Diagonal principal (probabilidades de imigração)		
% de células diferentes de zero	0,0706	0,0825
Média	0,0027	0,0024
Desvio-padrão	0,0063	0,0057
Mediana	0,0006	0,0005
Máximo	0,1665	0,1300
Mínimo	0,0000	0,0000

Fonte: Censos Demográficos de 1991 e 2000

Nota: As medidas descritivas para os elementos fora da diagonal principal excluem valores iguais a zero.

Modelagem da migração intermunicipal e do crescimento natural

Para se prever as probabilidades de migração intraestadual – células fora da diagonal principal da matriz de crescimento interregional – especificou-se o seguinte modelo linear generalizado:

$$p_{ij}^{2000} = \beta_0 + \beta_1 * p_{ij}^{1991} + \varepsilon_{ij} \quad \text{sendo } p_{ij}^{2000} \sim \Gamma(\kappa; \theta) \quad (6)$$

onde p_{ij} é a proporção de pessoas migrando da região j para a região i . O modelo tem distribuição gamma altamente assimétrica a direita, com p_{ij}^{2000} e parâmetros de forma κ e escala θ maiores que zero; ε_{ij} se refere ao componente residual do modelo. Os coeficientes estimados pelo método bayesiano estão na tabela 2:

TABELA 2.

Coeficientes estimados pelo modelo bayesiano linear generalizado para a probabilidades de migração intermunicipal, São Paulo 1991 e 2000

Variável	Descrição	Média	Desvio-padrão	Erro-padrão da média <i>posterior</i> (MCSE)	Média <i>posterior</i> da regressão Bayesiana (intervalo crível de 95%)
p_{ij}^{1991}	Probabilidades de migração intermunicipais em 1991	0,8379	0,0174	0,000211	(0,8045; 0,8722)
β_0	Intercepto da regressão	0,00099	9,14E-06	1,30E-07	(0,00097; 0,00101)

Nota: As estimativas utilizam 27.089 observações diferentes de zero fora da diagonal principal da matriz de crescimento inter-regional. O modelo de regressão bayesiano utilizou distribuições prévias (*priors*) não informativas. A taxa de aceitação da cadeia de Markov Monte Carlo foi de 37,59% e a taxa média de eficiência foi de 15,42%. Estas taxas indicam boa mistura de cadeia de Markov. Utilizou-se amostragem Metropolis-Hasting com passeio aleatório para a média. Outras medidas de ajuste do modelo bayesiano encontram-se no anexo 1.

Os resultados mostram que a probabilidade média esperada de migração, independentemente daquela observada em 1991, é praticamente zero (0,099%). Levando-se em conta a probabilidade de migração passada, a probabilidade esperada para 2000 altera-se consideravelmente. Quando, por exemplo, a probabilidade de migração em 1991 é igual a 0,01 (ou 1%), a probabilidade de migração em 2000 passa a ser de 0,94% [ou $0,00099 + 0,8379 * 0,01$]. Caso utilizássemos os coeficientes médios estimados para prever as probabilidades de migração em 2010 teríamos um modelo do tipo:

$$\widehat{p_{ij}^{2010}} = 0,00099 + 0,8379 * \widehat{p_{ij}^{2000}} \quad (7)$$

sendo $\widehat{p_{ij}^{2000}}$ o vetor das probabilidades previstas de imigração intermunicipal entre 1995 e 2000 geradas a partir da equação (6), e $\widehat{p_{ij}^{2010}}$ as probabilidades previstas de imigração intermunicipal entre 2005 e 2010. A incerteza futura sobre o valor destas probabilidades previstas é refletida pela variação dos coeficientes bayesianos estimados, contidos no intervalo crível de 95% de confiança estatística.

Já para a diagonal principal, que representa o crescimento natural e a emigração ocorrida entre

1995 e 2000, o modelo preditivo possui a seguinte especificação:

$$c_i^{2000} = a_0 + a_1 * c_i^{1991} + \varepsilon_i \quad \text{sendo } c_i^{2000} \sim N(\mu; \sigma^2) \quad (8)$$

onde c_i são os valores da diagonal principal da matriz de crescimento G para o município i em 1991 ou 2000; e ε_i é termo de erro da regressão. Os valores de c_i^{2000} e c_i^{1991} são calculados seguindo a metodologia descrita no módulo 1. Já os coeficientes a_0 e a_1 (Tabela 3) são estimados utilizando-se métodos bayesianos e mínimos quadrados ordinários⁷:

TABELA 3.

Coefficientes estimados pela regressão bayesiana de mínimos quadrados ordinários para o crescimento natural e emigração intermunicipal, São Paulo 1991 e 2000

Variável	Descrição	Média	Desvio-padrão	Erro-padrão da média posterior (MCSE)	Média posterior da regressão Bayesiana (Intervalo crível de 95%)
c_i^{1991}	Crescimento demográfico natural e emigração entre 1986 e 1991	0,8200	0,0261	0,000847	(0,7692; 0,8707)
a_0	Intercepto da regressão	0,1630	0,0245	0,000796	(0,1156; 0,2106)

Nota: As estimativas baseiam-se em 560 observações da diagonal principal da matriz de crescimento inter-regional (13 observações foram eliminadas da estimativa por terem distância de Cook > 1 ou serem pontos influentes). O modelo de regressão bayesiano utilizou distribuições prévias (*priors*) não informativas. A taxa de aceitação da cadeia de Markov Monte Carlo foi 100% e a taxa média de eficiência foi de 34,6%. Estas taxas indicam que boa mistura de cadeia de Markov. Utilizou-se amostragem Gibbs tanto para a variância quanto para a média. Outras medidas de ajuste do modelo bayesiano encontram-se no anexo 1.

Os coeficientes médios *a posteriori* são próximos daqueles obtidos a partir da abordagem frequentista – o modelo final “explica” quase 70 por cento da variabilidade do crescimento natural municipal no estado de São Paulo em 2000 ($R^2_{ajustado} = 0,66$). A principal vantagem da abordagem

⁷ Restringiu-se a amostra aos casos cujo o peso calculado pela regressão robusta (BERK, 1990; ROUSSEEUW e LEROY, 1987; VERARDI e CROUX, 2009) fosse maior do que zero para diminuir a influência de valores extremos na estimativa dos coeficientes.

bayesiana, no entanto, é fornecer “intervalos críveis” e com interpretação probabilística para os coeficientes. A probabilidade do coeficiente estimado para c_i^{1991} estar contido dentro do intervalo $[0,77; 0,87]$ é de 95%. Utilizando a estimativa da média *posterior*, o modelo mostra que os municípios que cresceram 10% (ou 1,1) entre 1986 e 1991 por causas naturais (e emigração) cresceriam 6,5% (ou $0,163 + 0,82*1,1$) entre 1995 e 2000 de acordo com a predição do modelo. Já os municípios cujo crescimento natural entre 1986 e 1991 foi igual a 1% (ou 1,01), não teriam crescimento demográfico significativo entre 1995 e 2000. Em suma, o modelo mostra que o crescimento natural (mais emigração) no quinquênio seguinte seria, em média, 0,82% daquele observado nos cinco anos anteriores.

Previsão multirregional da população dos municípios paulistas

O quarto objetivo desta pesquisa é ilustrar uma metodologia integrada para projetar (e prever) a população dos municípios de São Paulo levando em conta a incerteza inerente ao processo de projeção. Este módulo da pesquisa introduz o novo método proposto e valida as projeções a partir da comparação com as populações observadas e com os resultados de outro método já consolidado e utilizado pelo IBGE, o A_iB_i .

Duas projeções e uma previsão populacional foram feitas para todos os municípios do estado de São Paulo. A primeira projeção utiliza a matriz de crescimento gerada em 1991 para projetar a população para o ano 2000 assumindo estabilidade nos parâmetros da matriz. Essa projeção determinística resulta da multiplicação das populações observadas em 1990 pelo quadrado da matriz de crescimento previamente calculada:

$$w_{proj}^{2000} = G^{1991} * G^{1991} * w^{1990} \quad (9)$$

w_{proj}^{2000} representa um vetor-coluna das i populações municipais projetadas para o ano 2000; G^{1991} representa a matriz de crescimento observado de 1986 para 1991; e w^{1990} é o vetor-coluna contendo

as i populações observadas de 1990.

A segunda projeção é executada utilizando o método A_iB_i , desenvolvido pelas Nações Unidas (UNITED NATIONS, 1956) e pioneiramente utilizado no Brasil por Madeira e Simões (1972). A aplicação do método requer a projeção da área maior (no caso o estado de São Paulo), e as distribuições dos totais populacionais dos municípios. Para projetar a população até o ano 2000 foram tomadas as populações municipais observadas nos censos de 1980 e 1991, e a estimativa populacional para o estado publicada em Waldvogel and Capassi (1999)⁸. Pelo A_iB_i projeta-se as populações municipais para o ano 2000 da seguinte forma:

$$w_{proj}^{2000,i} = w^{1991,i} + \frac{w^{1991,i} - w^{1980,i}}{W^{1991} - W^{1980}} * (W_{proj}^{2000} - W^{1991}) \quad (10)$$

Como antes, $w^{t,i}$ representa as populações do domínio menor, ou seja, dos municípios i ; $w_{proj}^{t,i}$ a população projetada do município i no tempo t ; e W representa a população do domínio maior, ou seja, do estado de São Paulo.

Para a previsão probabilística populacional considerou-se as equações (4) e (5) para a obtenção de 1000 matrizes de crescimento (e consequentemente 1000 projeções) geradas a partir do método de Monte Carlo e da variabilidade intraestadual (desvio-padrão) dos valores dentro e fora da diagonal principal da matriz de crescimento⁹.

Após combinar os valores gerados dentro e fora da diagonal principal para criar 1000 matrizes de crescimento, utilizou-se o valor médio das projeções para realizar as comparações de precisão do novo método probabilístico. Entre todos os intervalos de projeção populacional gerados para cada

⁸ Na época desta publicação projetava-se que a população do estado de São Paulo seria igual a 36.102.875 pessoas. A população de fato observada foi igual a 36.423.321 indivíduos.

⁹ O anexo 2 apresenta o programa de Stata utilizado para fazer a extração dos parâmetros de variação da matriz de crescimento, apresenta as estatísticas descritivas da distribuição e exemplifica o histograma de ômega (ω) e phi (ϕ) para 1.000 extrações aleatórias das distribuições normais assumidas. Por se tratar de um processo aleatório múltiplas extrações de mesmo tamanho apresentarão distribuições ligeiramente distintas das anteriores e, consequentemente, diferentes intervalos de projeção populacional. A incerteza inerente a este processo, entretanto, não causa variações significativas nas matrizes de crescimento utilizadas. Pelo Teorema Central do Limite, um número maior de extrações apenas torna a distribuição mais próxima da distribuição Gaussiana.

município, constatou-se que 96 por cento deles incluíram o verdadeiro valor observado no Censo Demográfico de 2000. Dos 572 municípios paulistas analisados, 545 tiveram populações previstas dentro dos intervalos produzidos e apenas 27 tiveram populações médias fora do intervalo. Os anexos 3 e 4 ilustram algumas previsões fora e dentro do intervalo de confiança para populações municipais paulistas.

Para mensurar a precisão das projeções e da previsão populacional média utilizou-se o erro médio percentual da projeção (EMP), o erro médio percentual em módulo (EMPM) e a razão entre as populações observada e projetada (ROP). Estes três indicadores são definidos da seguinte forma:

$$EMP_t = \frac{\sum_{i=1,n} \left(\frac{w_{proj}^{t,i} - w^{t,i}}{w^{t,i}} \right) \times 100}{n} \quad (11)$$

$$EMPM_t = \frac{\sum_{i=1,n} \left| \left(\frac{w_{proj}^{t,i} - w^{t,i}}{w^{t,i}} \right) \times 100 \right|}{n} \quad (12)$$

$$ROP_t = \frac{w^{t,i}}{w_{proj}^{t,i}} / n \quad (13)$$

Segundo Devine e Coleman (apud BRITTO; CAVENAGHI; JANNUZZI, 2010, p. 38), o EMP é uma medida de tendência média e pode ser usada como base para testar a presença de tendência média significativa, enquanto o EMPM é uma medida de precisão que indica o quão “perto” as estimativas chegaram do valor observado, em média (DEVINE; COLEMAN, 2003). Já a ROP indica simplesmente o quão maior ou menor a população observada é em relação à população projetada. O quadro comparativo entre as três projeções realizadas e as populações observadas encontra-se abaixo:

TABELA 4.

Erro médio da projeção demográfica, por método utilizado, municípios de São Paulo, 2000

Método	EMP₂₀₀₀	EMPM₂₀₀₀	População estadual projetada para 2000	ROP₂₀₀₀
Projeção determinística usando matriz de crescimento demográfico	4,57 (22,54)	11,16 (20,11)	38.020.794	0,98 (0,14)
A_iB_i	-0,30 (20,86)	10,70 (17,90)	36.075.827	1,03 (0,14)
Valor médio da previsão probabilística de crescimento populacional	5,25 (22,98)	11,57 (20,53)	38.241.176	0,98 (0,14)

Nota: Desvios-padrão encontram-se entre parênteses.

Smith et al. (2001) colocam que para um horizonte de projeção de dez anos, o EMPM aceitável para pequenos domínios, como municípios, tem ficado entre 8% e 14%. No Brasil, os métodos de projeção populacional para pequenos domínios considerados de boa precisão têm sido aqueles em que mais da metade dos municípios apresentam erro percentual entre -10% e 10% (JARDIM, 1995; WALDVOGEL; CAPASSI, 1999).

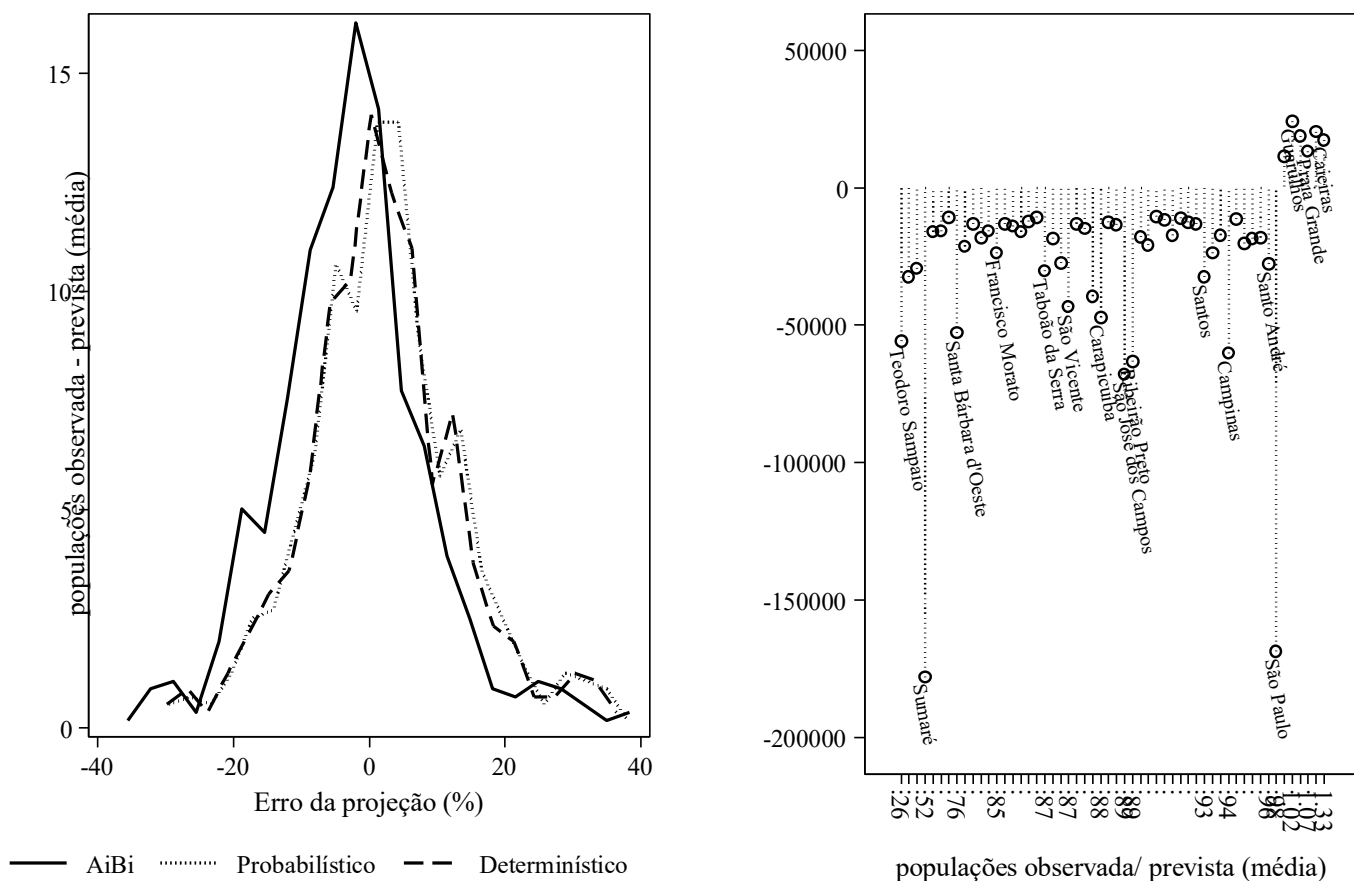
Os resultados obtidos são compatíveis com estes padrões. Apesar da população estadual projetada pelo A_iB_i estar mais próxima da população de fato observada no Censo de 2000 (36.423.321), o erro médio apresentado pelos outros métodos encontra-se dentro dos limites aceitáveis. O método A_iB_i é o que apresenta os menores erros médios (EMP e EMPM), mas todos os métodos possuem precisão aceitável e compatível com aquela sugerida pela literatura. A coluna ROP mostra que, em média, a população dos municípios projetada pelo A_iB_i é três por cento maior do que a população de fato observada. Já os métodos determinístico e probabilístico tendem a projetar populações municipais que são, em média, 98 por cento do tamanho da população de fato observada¹⁰.

¹⁰ No caso da projeção probabilística, especula-se que a população total média do estado seja maior que a observada por causa dos valores extremos projetados, gerados pelo fato das projeções incluírem todos os valores preditos, e não apenas aqueles com 95% de confiança estatística. Pelo método probabilístico a população prevista para o estado estaria entre 24.531.361 e 54.007.595. Este intervalo de confiança seria mais estreito adotando-se níveis de confiança menores (ex. contidos entre aproximadamente dois desvios-padrão da média prevista populacional prevista).

A Figura 1 (à esquerda) apresenta a distribuição de frequência dos erros percentuais da projeção e demonstra que todos os métodos apresentam erro médio em torno de zero. A imagem à direita da figura mostra a distribuição dos erros absolutos dos municípios mais problemáticos, cuja a diferença populacional *ex-post* foi superior a dez mil pessoas.

FIGURA 1.

Erros percentuais e absolutos de projeção populacional, municípios de São Paulo, 2000



Nota: Municípios com erros de projeção maiores que 40% (menos de 3% dos casos) foram omitidos do gráfico à esquerda para facilitar a comparabilidade entre os métodos. Pela mesma razão o gráfico da direita mostra apenas municípios que tiveram erro absoluto de projeção maior que 10.000 pessoas.

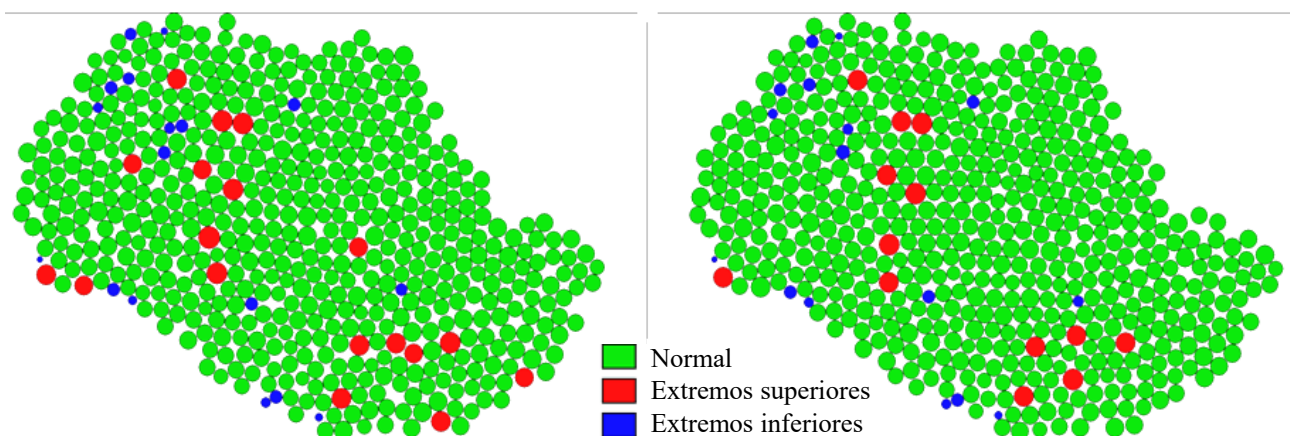
A Figura 2 complementa a análise espacial do erro das projeções identificando municípios cujo crescimento foi discrepante do crescimento dos demais¹¹. O cartograma mostra a localização aproximada dos municípios cuja razão entre a população observada e projetada foi sub ou sobrestimada. O padrão de pontos extremos é similar entre os dois métodos, indicando que os municípios nos quais as populações são sub ou sobrestimadas são os mesmos, independentemente do método de projeção utilizado (o coeficiente de determinação entre os erros das duas projeções foi igual a 0,91).

FIGURA 2.

Cartograma da razão entre as populações observada e projetada em 2000, São Paulo

Método A_iB_i

Método probabilístico



Nota: Os pontos extremos são definidos como aqueles cujos valores são 1,5 vezes aquele do intervalo interquartilício.

DISCUSSÃO E CONTRIBUIÇÕES

No Brasil, a estratégia metodológica dominante para realizar projeções em pequenas áreas consiste em utilizar o método das componentes para projetar a população para áreas maiores e então estimar o número de indivíduos em áreas menores através de métodos de ajuste proporcional ou interpolativo – tais como A_iB_i . A outra opção, conhecida como modelo ProjPeq, “prescinde de investimentos ou esforços técnicos muito vultosos na produção de projeções de totais populacionais para pequenas

¹¹ Os mapas para a projeção determinística apresentaram distribuição similar à da projeção probabilística.

áreas” e “não produz necessariamente estimativas futuras mais precisas do que os outros métodos” (JANUZZI, 2007, p. 111, p. 130). A metodologia probabilística adotada nesta pesquisa, entretanto, difere dos métodos tradicionalmente utilizados. A primeira diferença é fornecer estimativas populacionais que partem de uma área menor para uma área maior de análise, levando em conta as especificidades regionais de maneira integrada. A segunda diferença é utilizar um método de modelagem estrutural que leva em conta a incerteza atribuída ao comportamento futuro da dinâmica demográfica através de métodos bayesianos que utilizam cadeias de Markov e simulações Monte Carlo. Em particular, esta técnica permite a construção de intervalos de confiança estatística para a previsão populacional municipal e para o estado de São Paulo. Os resultados dos modelos estimados neste artigo, entretanto, ainda precisam ser explorados em previsões populacionais de médio e longo prazo (até 2030, por exemplo).

A terceira vantagem do método está na sua simplicidade e facilidade de execução. A matriz de crescimento multirregional pode ser construída a partir de apenas três informações disponíveis nos Censos Demográficos de 1991 e 2000: o município de residência anterior, no ano $t-n$; e as distribuições populacionais por município no ano $t-n$ e t . Como o método não requer a estimação de taxas de fecundidade e mortalidade para pequenas áreas, a sua execução e operacionalização é muito mais simples e direta do que o coorte-componente ou o ProjPeq.

Em quarto lugar, o método sugerido explicita o papel dos fluxos migratórios na dinâmica de crescimento intermunicipal ao separá-los do efeito combinado da mortalidade e da fecundidade. No caso de pequenas áreas, a mensuração e modelagem dos fluxos migratórios são cruciais já que parcela do crescimento demográfico municipal se atrela à dinâmica deste componente. Além disso, a matriz de crescimento multirregional permite a confecção de cenários contrafactuais, simulações e análises de sensibilidade com maior facilidade. Entre as perguntas que a matriz de crescimento populacional ajudaria a responder destacam-se: Qual o efeito marginal da migração e do crescimento vegetativo municipal sobre o crescimento demográfico de São Paulo? Qual seria o crescimento inter-regional e estadual caso os níveis de migração intermunicipal fossem diferentes? Qual seria o nível de

intervenção requerido para redirecionar o sistema de crescimento inter-regional para uma distribuição alvo? Estas perguntas podem ser mais facilmente respondidas através da matriz de crescimento inter-regional.

Em quinto lugar, análises de estabilidade e convergência podem ser feitas sem grandes complicações a partir da matriz de crescimento. Pode-se perguntar, por exemplo, quais seriam as consequências de longo prazo de se manter um calendário de crescimento constante, qual a taxa de crescimento intrínseco, estrutura e distribuição da população estável, e qual é o número de anos de crescimento até a se atingir a estabilidade. Em particular, análises de estabilidade são cruciais já que elas determinam, em grande parte, o quanto os parâmetros da projeção irão variar no futuro. Willekens (1990, p. 17), por exemplo, argumenta que mais importante do que discutir a precisão de projeções de curto e longo prazo, seria discutir quando um sistema demográfico está próximo da estabilidade ou não, já que isso é o que irá determinar e auxiliar a definição de hipóteses futuras para a projeção populacional.

Em sexto lugar, as projeções produzidas poderão ser combinadas com estimativas municipais alternativas, já que previsões populacionais médias podem gerar melhores resultados que previsões monometodológicas. Sanderson (1998) e Ahlburg (1998), por exemplo, exploraram as vantagens de se combinar previsões utilizando diferentes métodos e concluíram que, quando combinados, os modelos econômico-demográficos de países em desenvolvimento produzem estimativas que são tão boas quanto aquelas produzidas pelo método das coorte-componentes. Eles acreditam que a combinação de diferentes previsões populacionais produz resultados mais precisos.

Por fim, as previsões populacionais realizadas através do método proposto tiverem um erro *ex-post* aceitável e similar às projeções feitas pelo $A_i B_i$, indicando que há potencial para que o método seja adotado em outras regiões e para diferentes níveis de agregação, por idade e sexo por exemplo. Em particular, sugere-se que a restrição da variabilidade dos elementos da matriz de crescimento (ϖ e φ) a certas regiões do estado – que apresentaram padrões similares de crescimento demográfico ao longo do espaço – possa vir a aprimorar ainda mais as projeções apresentadas.

REFERÊNCIAS BIBLIOGRÁFICAS

AHLBURG, D. A. Using economic information and combining to improve forecast accuracy in demography, **Working paper, Industrial Relations Center**, University of Minnesota, Minneapolis, 1998.

ALHO, J.; SPENCER, B. **Statistical Demography and Forecasting**. United States, Springer, 2005.

ANSELIN, L. Local Indicators of Spatial Association - Lisa. **Geographical Analysis**, v. 27, n. 2, p. 93-115, 1995.

ANSELIN, L. An Introduction to Spatial Autocorrelation Analysis with GeoDa. **Spatial Analysis Laboratory (SAL)**. Urbana-Champaign, IL, Department of Agricultural and Consumer Economics, University of Illinois, 2003. Disponível em: < <http://www.uni-kassel.de/~rkosfeld/lehre/spatial/spauto.pdf>>. Acesso em 04 de fev., 2015.

BAENINGER, R.; CUNHA, J. M. Migração, Dinâmica Regional e Projeções Populacionais. **São Paulo em Perspectiva**, v. 10, n. 2, p. 102-109, 1996.

BERK, R. A. 1990. A primer on robust regression. In **Modern Methods of Data Analysis**, ed. J. Fox and J. S. Long, 292-324. Newbury Park, CA: Sage.

BOOTH, H. Demographic forecasting 1980 to 2005 in review. **International Journal of Forecasting**, v. 22, n. 3, p. 547-581, 2006.

BRITO, L.; CAVENAGHI, S.; JANNUZZI, P. M. Estimativas e projeções populacionais para pequenos domínios: uma avaliação da precisão para municípios do Rio de Janeiro em 2000 e 2007. **Revista Brasileira de Estudos de População** v. 27, n. 1, p. 35-57, 2010.

CAVENAGHI, S.; POTTER, J.; SCHMERTMANN, C.; ASSUNÇÃO, R. Estimating Total Fertility Rates for Small Areas in Brazil. **Meeting of the Population Association of America**. Boston, MA., 2004.

CUNHA, J. M. Migração e Urbanização no Brasil: alguns desafios metodológicos para análise. **São Paulo em Perspectiva**, v. 19, n. 4, p. 3-20, 2005.

CUNHA, J. M.; JAKOB, A. Segregação socioespacial e inserção no mercado de trabalho na Região Metropolitana de Campinas. **Revista Brasileira de Estudos de População** v. 27, n. 1, 115-139, 2010.

DEVINE, J.; COLEMAN, C. **People might move but housing units don't**: an evaluation of the state and county housing unit estimates. Washington: Population Division Working Paper Series nº 71 - U.S. Census Bureau, 2003.

DUCHESNE, L. **Proyecciones de poblacion por sexo y edad para areas intermedias y menores**. Santiago: Centro Latinoamericano de Demografia, 1987.

FIGOLI, M. Intervalo de confiança para projeção de população baseado no método de Monte Carlo: projeção dos beneficiários urbanos da previdência social. **Revista Brasileira de Estatística**, v. 60, n. 213, p. 25-51, 2000.

FÍGOLI, M. G. B.; WONG, L. L. R.; SAWYER, D. R. T. O.; CARVALHO, J. A. M. Proyección Multirregional: Aplicación en Brasil y sus Unidades Federativas (200-2020). **Notas de Población**, Santiago do Chile, v. 76, p. 7-35, 2003.

FOTHERINGHAM, A.; BRUNSDON, C. Local forms of spatial analysis. **Geographical Analysis**, v. 31, p. 340-358, 1999.

FREIRE, F. H.; ASSUNÇÃO, R. M. Intervalo de Confiança para a Taxa de Fecundidade Total de Pequenas Áreas. **XIV Encontro Nacional de Estudos Populacionais**. Caxambú, MG, Brazil, 1998. Disponível em:

<<http://www.abep.nepo.unicamp.br/docs/anais/PDF/1998/a240.pdf>>. Acesso em 04 de fev., 2015.

GIROSI, F.; KING, G. **Demographic forecasting**. Princeton, NJ: Princeton University Press, 2008.

GUIMARÃES, R. R. Uncertainty in population projections: the state of the art. **Revista Brasileira de Estudos de População**, São Paulo, v. 31, n. 2, 2014.

HAKKERT, R. Quasi-component models for small area population projections: examples from the state of São Paulo, Brazil. **International Union for the Scientific Study of Population**, Florence, 1985.

IBGE; OLIVEIRA, J.; ALBUQUERQUE, F. R.; LINS, I. B. Projeção da População do Brasil por Sexo e Idade para o Período 1980-2050: Revisão 2004 Metodologia e Resultados; Estimativas Anuais e Mensais da População do Brasil e das Unidades da Federação: 1980 - 2020 Metodologia; Estimativas das Populações Municipais: Metodologia, 2004. Disponível em:

<<http://www.ibge.gov.br/home/estatistica/populacao/estimativa2004/metodologia.pdf>>. Acesso em 04 de fev., 2015.

JAKOB, A. Zonas de segregação e vulnerabilidade na Baixada Santista. In: Daniel Arias Vazquez. (Org.). **A Questão Urbana na Baixada Santista: políticas, vulnerabilidades e desafios para o desenvolvimento**. 1ª. ed. São Paulo: Editora Universitária Leopoldianum, p. 65-80, 2011.

JANUZZI, P. M. Cenários futuros e projeções populacionais para pequenas áreas: método e aplicação para distritos paulistanos 2000-2010. **Revista Brasileira de Estudos de População**, vol. 24, n. 01, p. 109-136, 2007.

JARDIM, M. L. **Utilização de variáveis sintomáticas para estimar a distribuição espacial de populações**: aplicação aos municípios do Rio Grande do Sul. Belo Horizonte: UFMG/Cedeplar, 1995.

KEYFITZ, N. The Limits of Population Forecasting. **Population and Development Review**, v. 7, n. 4, p. 579-593, 1981.

_____. Can knowledge improve forecasts? **Population and Development Review**, v. 8, n. 4, p. 729-751, 1982.

KEYFITZ, N.; CASWELL, H. **Applied Mathematical Demography**. New York, NY, Springer, 2005.

LEE, R. D. Stochastic Demographic Forecasting. **International Journal of Forecasting**, v. 8, n. 3, p. 315-327, 1992.

_____. Probabilistic Approaches to Population Forecasting. **Population and Development Review**, v. 24 (Supplement: Frontiers of Population Forecasting) p. 156-190, 1998.

LEE, R. D.; CARTER, L. Modelling and Forecasting the Times series of US Mortality. **Journal of the American Statistical Association**, v. 87, 1992.

LEE, R. D.; TULJAPURKAR, S. Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low. **Journal of the American Statistical Association**, v. 89, n. 428, p. 1175-1189, 1994.

LESLIE, P. H. On the Use of Matrices in Certain Population Mathematics. **Biometrika**, v. 33, n. 3, p. 183-212, 1945.

_____. Some Further Notes on the Use of Matrices in Population Mathematics. **Biometrika**, v. 35, n. 3/4, p. 213-245, 1948.

LUTZ, W.; VAUPEL, J.; AHLBURG, D. A. **Frontiers of population forecasting**. New York, Population Council, 1999.

MACHADO, C. **Projeções Multirregionais de População: o caso brasileiro (1980-2020)**. Tese de

doutorado apresentada no Centro de Desenvolvimento e Planejamento Regional /CEDEPLAR/ FACE/UFMG. Belo Horizonte, MG, 1993.

MADEIRA, J. L.; SIMÕES, C. C. Estimativas preliminares da população urbana e rural segundo as Unidades da Federação, de 1960/1980 por uma nova metodologia. **Revista Brasileira de Estatística**, v. 33, n.129, p. 3-11, 1972.

MIRON, J. Spatial Autocorrelation in Regression Analysis: a Beginner's Guide. In: GAILE, G. L.; WILLMOTT, C. J. (eds.). **Spatial Statistics and Models**. Dordrecht, Holland, D. Reidel Publishing Co., 1984, p. 201-222.

MUNIZ, J. O. Spatial dependence and heterogeneity in ten years of fertility decline in Brazil. **Population Review**, v. 48, n. 2, 2009.

_____. Demographic dynamics of poverty and income inequality: the case of Brazil. **Revista Brasileira de Estudos de População**, v. 29, n. 2, 2012.

PFLAUMER, P. Confidence Intervals for population projections based on Monte Carlo methods. **International Journal of Forecasting**, v. 4, n. 1, p. 135-142, 1988.

ROGERS, A. **Matrix analysis of interregional population growth and distribution**. Berkeley, University of California Press, 1968.

_____. **Matrix Methods in Urban and Regional Analysis**. Holden Day, San Francisco, 1971

_____. **Introduction to Multiregional Mathematical Demography**. John Wiley, New York, 1975.

_____. **Multiregional Demography: Principles, Methods and Extensions**. John Wiley, London, 1995.

ROUSSEEUW, P. J., LEROY, A. M. **Robust Regression and Outlier Detection**. New York: Wiley, 1987.

SANDERSON, W. Knowledge Can Improve Forecasts: A Review of Selected Socioeconomic Population Projection Models. **Population and Development Review**, Supplement: Frontiers of population Forecasting, v. 24, pp. 88-117, 1998.

SANTOS, R. O.; BARBIERI, A. F. Projeções populacionais em pequenas áreas: uma avaliação comparativa de técnicas de extrapolação matemática. **Revista Brasileira de Estudos de População**, v. 32 n. 1, 2015.

SCHOEN, R. **Modeling multigroup populations**. New York: Plenum Press, 1988.

_____. **Dynamic population models**. Dordrecht: Springer, 2006.

SMITH, S.; TAYMAN, J.; SWANSON, D. **State and local population projections: methodology and analysis**. New York: Kluwer, 2001.

STOTO, M. A. The Accuracy of Population Projections. **Journal of the American Statistical Association**, v. 78, n. 381, p. 13-20, 1983.

SZWARCWALD, C. L.; CASTILHO, E. A. Proposta de um modelo para desagregar projeções demográficas de grandes áreas em seus componentes geográficos. **Revista de Saúde Pública**, v. 23, n. 4, p. 269-76, 1989.

UNITED NATIONS. **Manual III**. Methods for population projections by sex and age. New York: Department of International Economic and Social Affairs, 1956.

_____. Dept. of Economic and Social Affairs. **The determinants and consequences of population trends : new summary of findings on interaction of demographic, economic and social factors**. New York, United Nations, 1973.

VERARDI, V.; CROUX, C. Robust regression in Stata. **Stata Journal** 9: 439-453, 2009.

WALDVOGEL, B. C. Projeção Populacional em São Paulo: um método analítico como alternativa. **São Paulo em Perspectiva**, São Paulo, Fundação SEADE, v. 10, n.2, 1996.

WALDVOGEL, B. C.; CAPASSI, R. Cenário da população paulista dos anos 90 ao futuro. **São Paulo em Perspectiva**, São Paulo, Fundação SEADE, v. 13, n.1-2, 1999.

WALDVOGEL, B. C.; FERREIRA, C. E.; YAZAKI, L. M.; GODINHO, R. E.; PERILLO, S. R. Projeção da População Paulista como Instrumento de Planejamento. **São Paulo em Perspectiva**, v. 17, n. 3-4, p. 67-79, 2003.

WILLEKENS, F. **Demographic forecasting: State-of-the-art and research needs**. In Emerging issues in demographic research/ ed by C.A. Hazeu and G.A.B. Frinking. Amsterdam: Elsevier. p. 9-66, 1990.

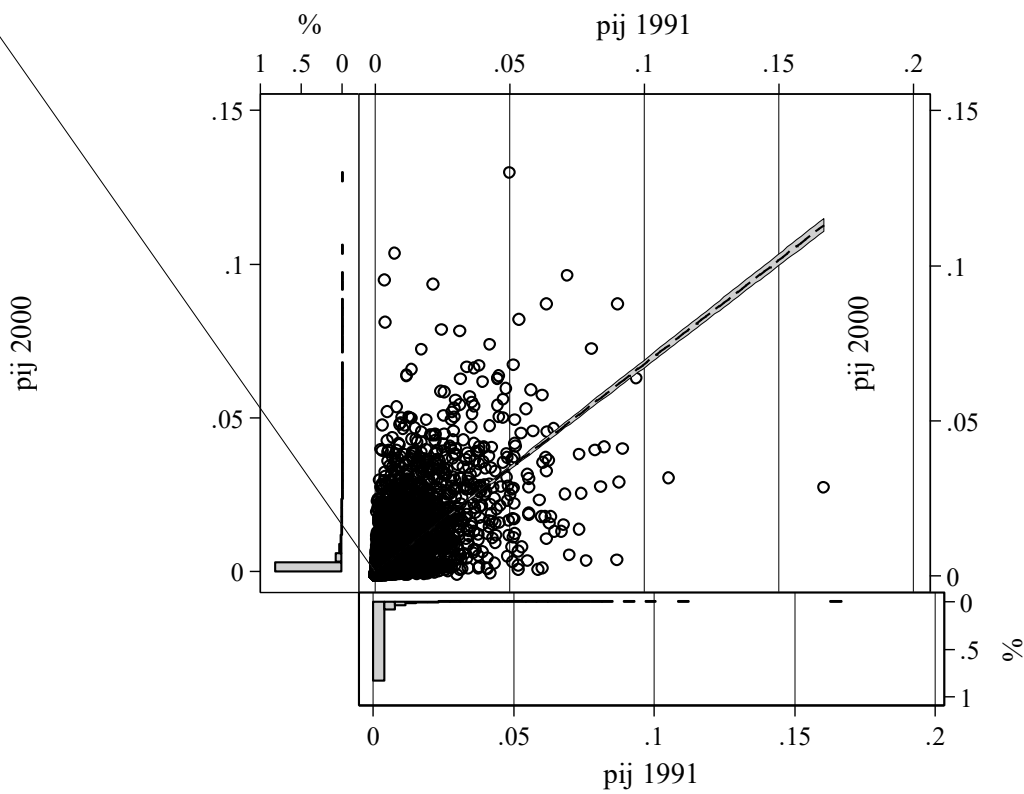
WISNIOWSKI, A.; SMITH, P. W. F.; BIJAK, J.; RAYMER, J.; FORSTER, J. Bayesian Population Forecasting: Extending the Lee-Carter Method. **Demography**, v. 52, n. 3, p. 1035-1059, 2015.

WHELPTON, P. K. Population of the United States, 1925 to 1975. **The American Journal of Sociology**, v. 34, n. 2, p. 253-270, 1928.

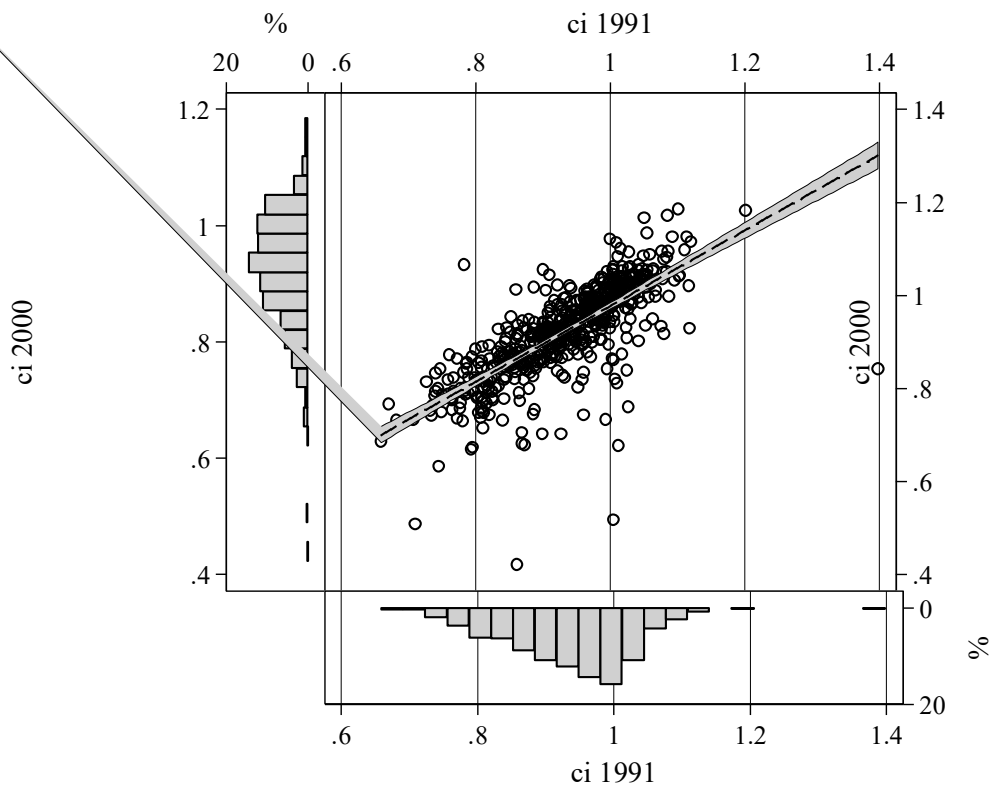
ANEXO 1.

Distribuição e correlação entre os elementos da matriz de crescimento

Elementos fora da diagonal principal (p_{ij}): probabilidades de migração de j para i



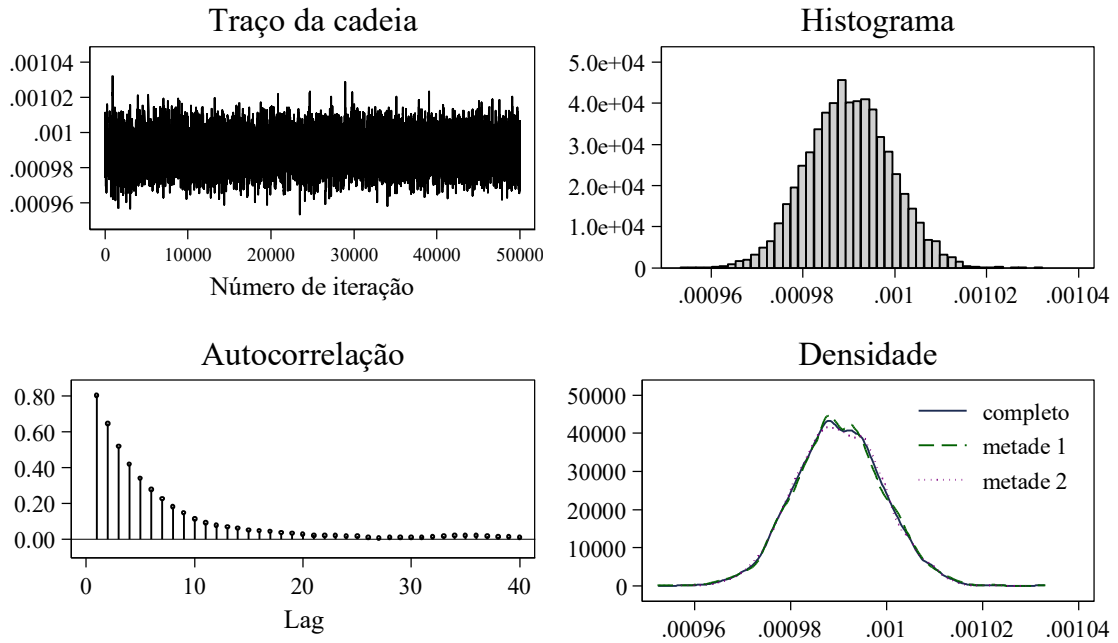
Elementos da diagonal principal (c_i): crescimento vegetativo mais emigração



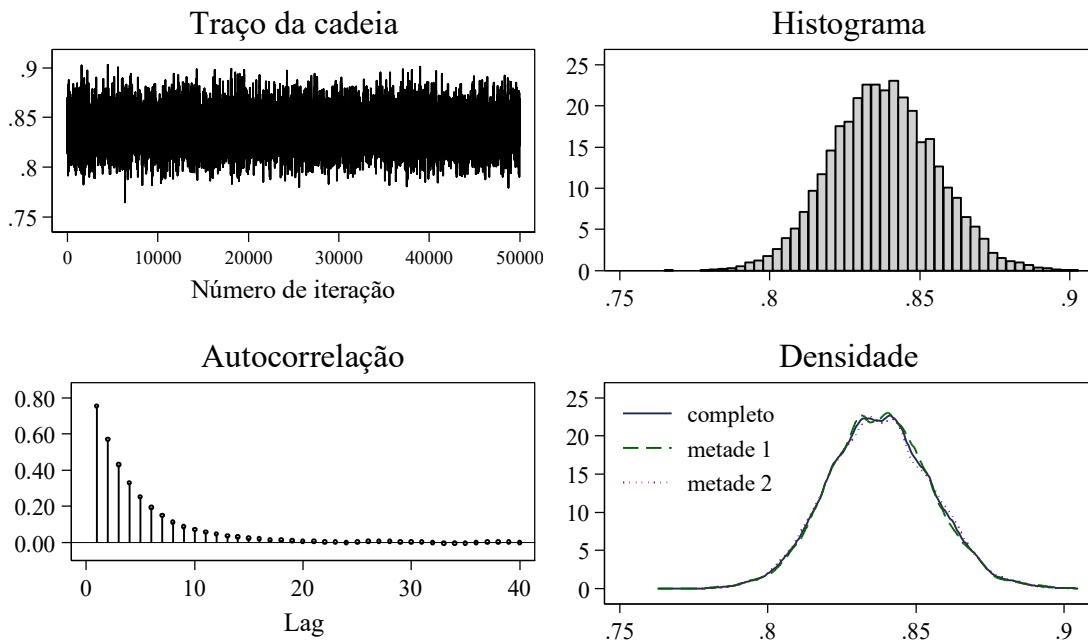
ANEXO 1.1.
Medidas diagnósticas de convergência dos coeficientes bayesianos

Modelo linear generalizado da família gamma com função de ligação identidade

B0



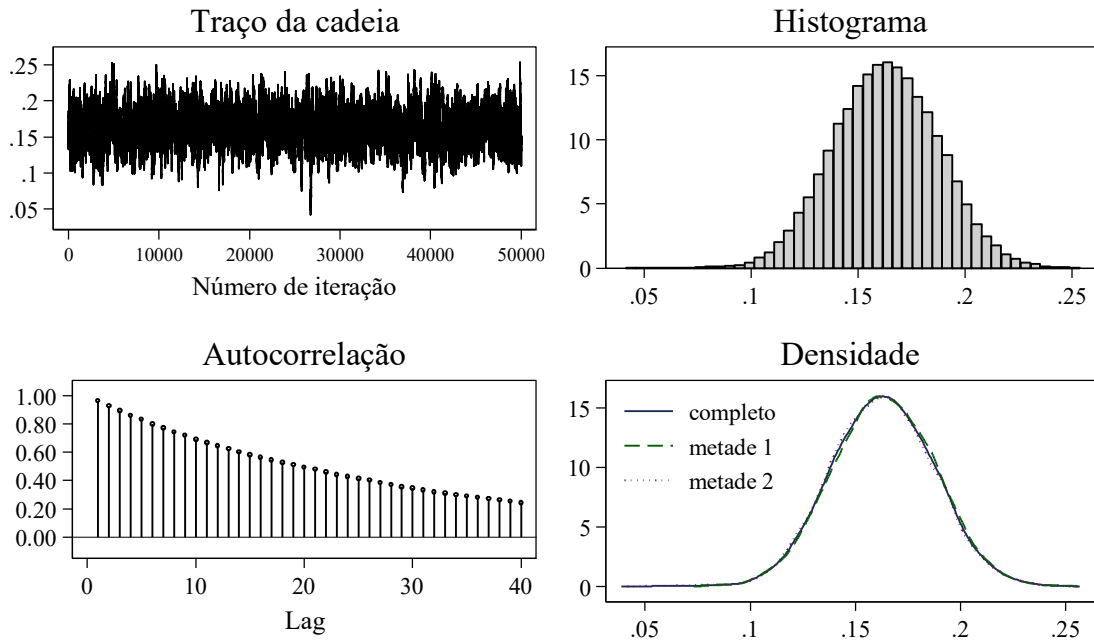
B1 (pij 1991)



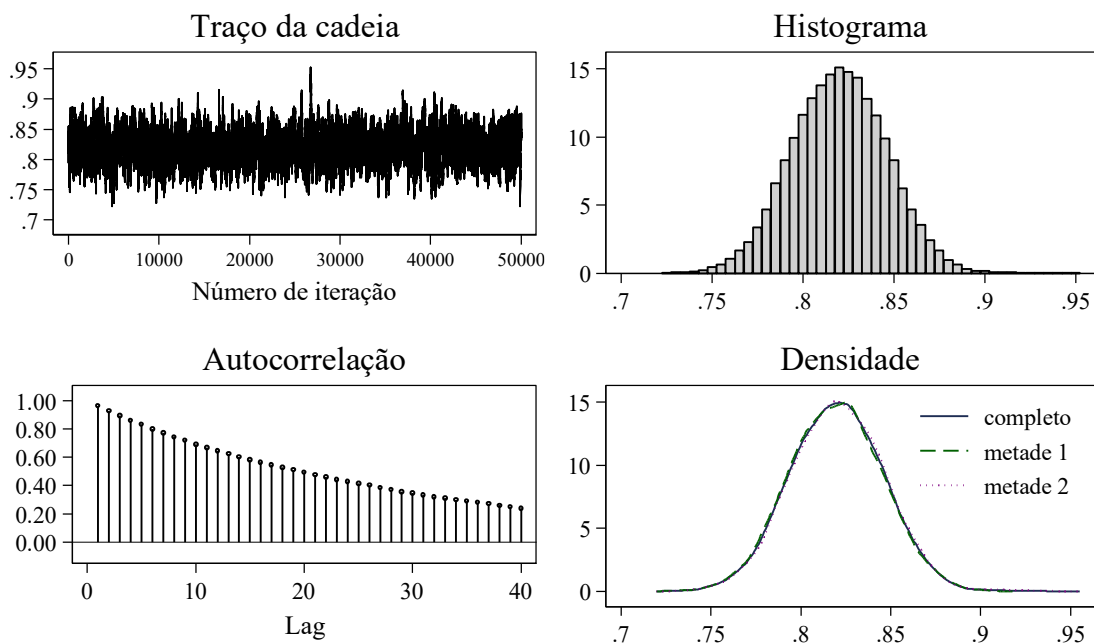
ANEXO 1.1. (cont.)

Modelo de mínimos quadrados ordinários

a_0



a_1 (ci 1991)

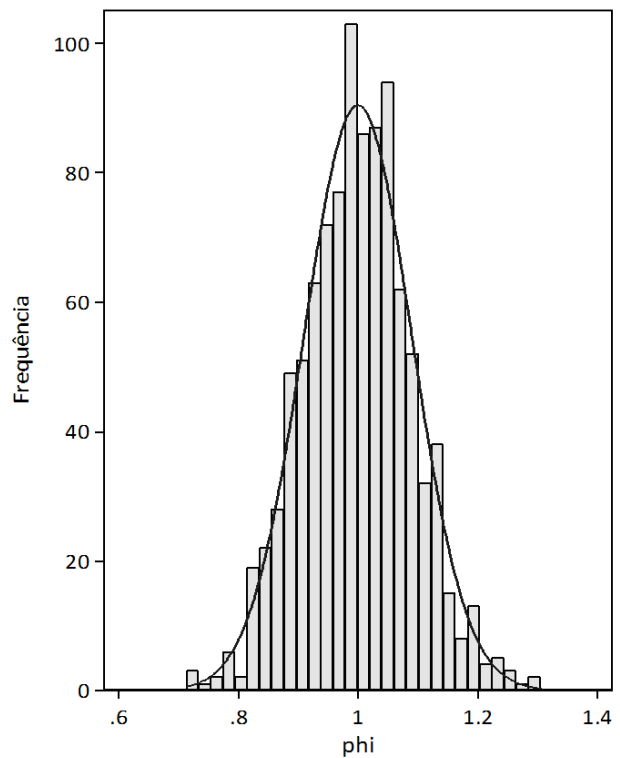
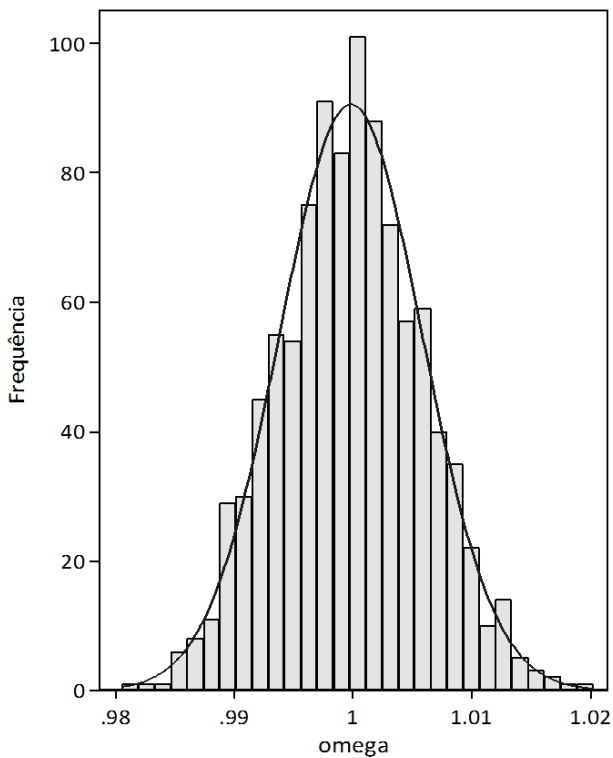


ANEXO 2. Parâmetros de variação da matriz de crescimento

```
*****
*Programa de Stata para gerar 1000 parâmetros de flutuação para as *
matrizes de crescimento e seus respectivos histogramas de frequência *
*****
matrix m = (1,1)
matrix sd = (.0063,.0909) /*define os desvios-padrão empíricos*/
drawnorm omega phi, n(1000) means(m) sds(sd)
summarize
hist omega, freq norm saving(omega_his.gph, replace)
hist phi, freq norm saving(phi_his.gph, replace)
gr combine omega_his.gph phi_his.gph, ycomm
```

$$p_{ij}^{1991} = p_{ij}^{1991} \varpi, \text{ sendo } \varpi \sim N(1; 0,0063)$$

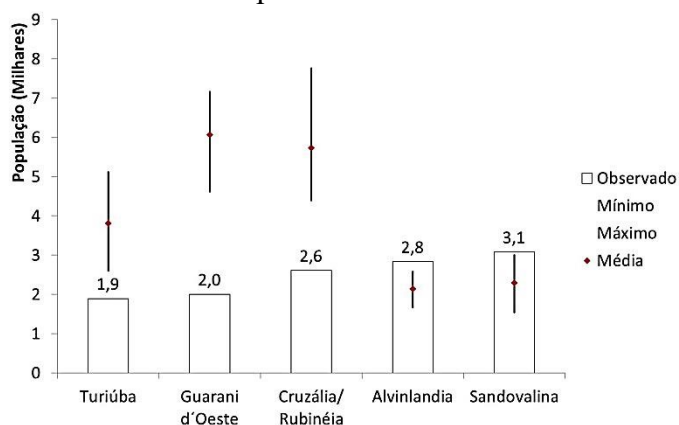
$$c_i^{1991} = c_i^{1991} \varphi, \text{ sendo } \varphi \sim N(1; 0,0909)$$



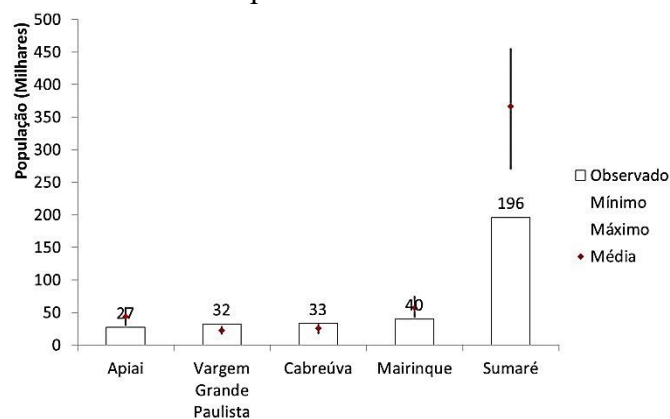
Parâmetro	Obs	Média	Des. Padrão	Min	Max
omega	1000	.9998105	.0060189	.9805955	1.02022
phi	1000	.9988998	.0901226	.7123487	1.304946

ANEXO 3. Previsões populacionais municipais selecionadas

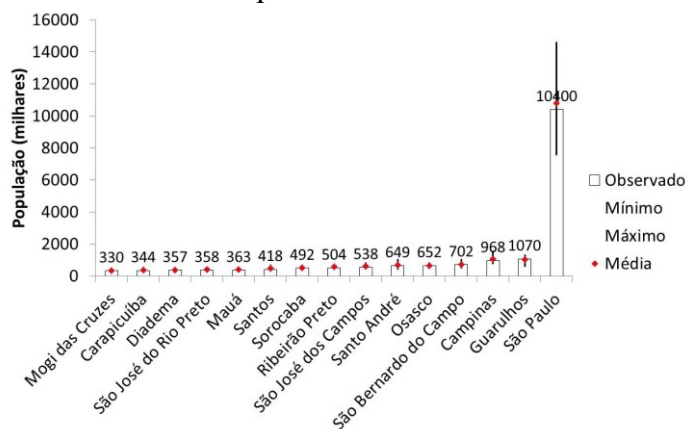
Cinco **menores** municípios com populações observadas **fora** do intervalo populacional probabilístico



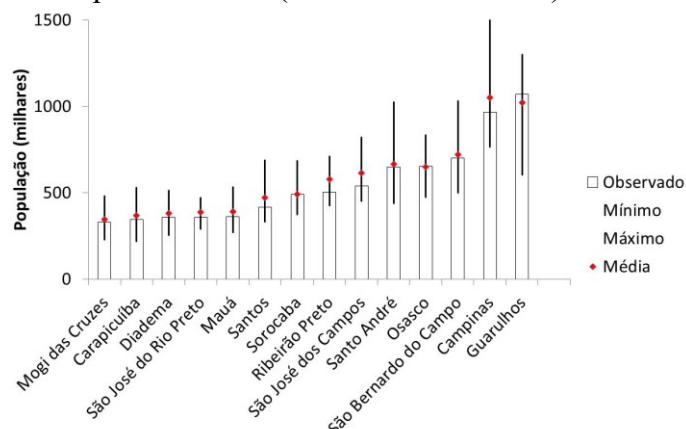
Cinco **maiores** municípios com populações observadas **fora** do intervalo populacional probabilístico



Quinze **maiores** municípios com populações observadas **dentro** do intervalo populacional probabilístico

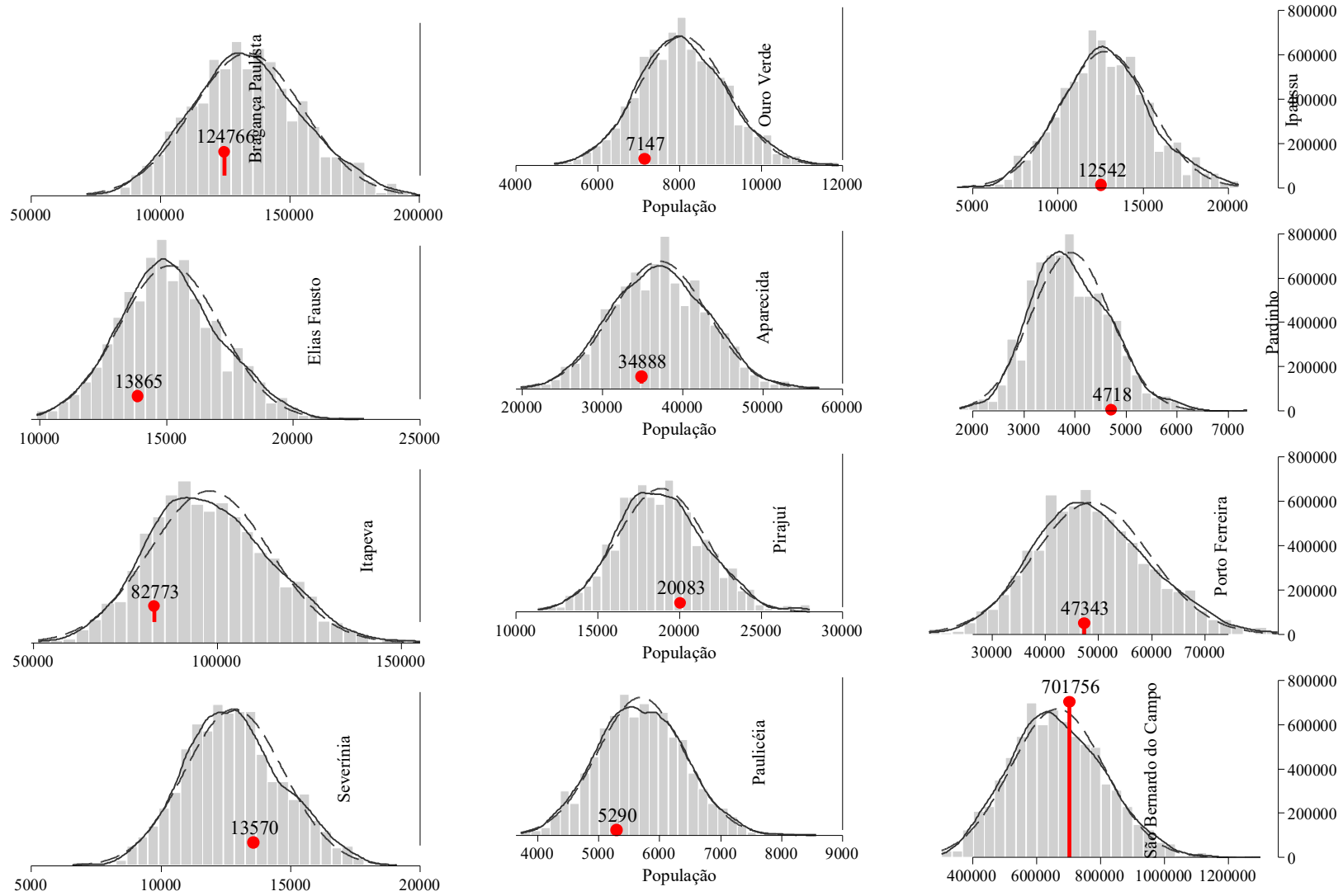


Catorze **maiores** municípios com populações observadas **dentro** do intervalo populacional probabilístico (excluindo São Paulo)



ANEXO 4.

Previsões populacionais *ex-post* para dez municípios paulistas aleatoriamente selecionados



Nota: Os pontos indicam o tamanho das populações municipais observadas em 2000, e os histogramas apresentam a distribuição de frequência das populações projetadas.