Estimação jackknife ponderada do erro quadrático médio de predição do EBLUP temporal

Luís Nobre Pereira
Universidade do Algarve - *Escola Superior de Gestão, Hotelaria e Turismo* - `Lmper@ualg.pt`

Pedro Simões Coelho
Universidade Nova de Lisboa - *Instituto Superior de Estatística e Gestão de Informação* - `Psc@isegi.unl.pt`

Resumo: A metodologia baseada na melhor predição linear empírica não enviesada (*Empirical Best Linear Unbiased Prediction*), consagrada com o acrónimo EBLUP, é muito utilizada na estimação de parâmetros para pequenos domínios. Apesar da relativa facilidade de dedução dos EBLUPs, mesmo num contexto de um modelo longitudinal, a medição da sua qualidade é um problema complexo devido à dificuldade de estimação do erro quadrático médio de predição (EQMP) de tais preditores. Neste trabalho utiliza-se um estimador de parâmetros de interesse em pequenos domínios assistido pelo modelo temporal de Rao-Yu (Rao e Yu, 1994). O EBLUP temporal é apresentado e é revisitada a aproximação analítica assimptótica do EQMP do EBLUP temporal proposta por Rao e Yu (1994). Sob o modelo de Rao-Yu, é proposta uma metodologia *jackknife* ponderada para estimar o EQMP do EBLUP, desenvolvida a partir dos trabalhos de Chen e Lahiri (2008). Foi realizado um estudo por simulação com o objectivo de comparar o desempenho do estimador proposto com o obtido por via da aproximação analítica do EQMP.

Palavras–chave: *jackknife*, EQMP do EBLUP temporal, métodos de reamostragem, estimação em pequenos domínios

Abstract: The empirical best linear unbiased prediction approach is a popular method for the estimation of small area parameters. While empirical best linear unbiased predictors (EBLUP) are fairly easy to obtain, even under a longitudinal model, measuring its quality is a challenging problem due to difficulties on estimating the mean squared prediction error (MSPE) of such estimators. This work assumes that the small area parameters of interest follow a cross-sectional and time-series stationary model due to Rao and Yu (1994). This paper presents the temporal EBLUP and reviews the asymptotic analytical approximation of the MSPE of the EBLUP proposed by Rao and Yu (1994). Under the model due to Rao and Yu (1994), a weighted jackknife method is introduced to estimate the MSPE of the temporal EBLUP using similar ideas as in Chen and Lahiri (2008). A Monte Carlo simulation study was carried out in order to compare the performance of the proposed estimator with the estimator based on the asymptotic analytical approximation of the MSPE.

different measures of uncertainty of the temporal EBLUP with this estimator.

Keywords: jackknife, MSPE of the temporal EBLUP, resampling methods, small area estimation

1 Introdução

Nas últimas décadas tem-se assistido a um crescimento acentuado da procura de estimativas fiáveis para parâmetros de variáveis de interesse ao nível de pequenos domínios, tendo em vista produzir informação estatística que permita um melhor planeamento, afectação mais equitativa de recursos nas áreas da saúde, educação, protecção social, entre outras, ou mesmo a realização de análises explicativas. Os pequenos domínios referem-se tipicamente a pequenas áreas geográficas ou pequenas subpopulações, para as quais a informação obtida através de inquéritos por amostragem é escassa, e para as quais se pretende estimar determinados parâmetros. Nestes casos, os estimadores tradicionais (estimadores directos baseados no desenho amostral e que utilizam valores da variável de interesse exclusivamente do domínio de estudo) não produzem estimativas com um nível de precisão adequado, uma vez que são baseadas geralmente num reduzido número de observações amostrais disponíveis em cada domínio. No limite, poderão eventualmente existir domínios sem qualquer observação amostral, para os quais não é possível sequer produzir estimativas directas.

Para resolver este problema, tem vindo a ser desenvolvida uma classe de estimadores indirectos baseados na metodologia EBLUP, a qual se tem revelado bastante adequada num vasto conjunto de aplicações na área da estimação em pequenos domínios e de problemas relacionados. A referida metodologia utiliza um modelo linear geral misto adequado para capturar as características principais do desenho amostral e para combinar no processo de estimação a informação amostral disponível com informação auxiliar relevante proveniente de outras fontes, de forma a compensar a escassez de informação amostral relativa a cada um dos domínios que são alvo de inferência. Dados administrativos, bem como dados amostrais exteriores ao domínio e/ou ao momento de referência, são comummentemente utilizados neste tipo de estimação. Desta forma, os modelos utilizados na estimação em pequenos domínios permitem produzir estimativas para os parâmetros de interesse com adequados níveis de precisão. Entre os modelos de nível área contemporâneos, encontra-se o conhecido modelo de Fay e Herriot (1979). Contudo, a utilização mais frequente de inquéritos longitudinais veio permitir utilizar dados de natureza cronológica, muitos deles apresentando autocorrelação temporal significativa. Foi neste contexto que Rao e Yu (1994) propuseram um modelo temporal de nível área para estimação em pequenos domínios.

O modelo temporal de Rao-Yu é um caso particular do modelo linear geral misto envolvendo efeitos aleatórios autocorrelacionados com uma estrutura de covariância homogénea. Utilizando a metodologia EBLUP, Rao e Yu (1994) derivaram o EBLUP temporal e uma aproximação analítica do EQMP desse estimador, seguindo as ideias de Kackar e Harville (1984). Contudo, essa aproximação corrigida até à segunda ordem exige que se verifique a normalidade dos erros do modelo, bem como que se esteja perante um pequeno número de períodos de tempo e um grande número de pequenos domínios para aproximar correctamente os verdadeiros valores do EQMP do EBLUP temporal.

As técnicas de reamostragem são actualmente aceites como boas alternativas às aproximações assimptóticas. Elas são atractivas porque são conceptualmente simples e de fácil aplicação mesmo para modelos mais complexos. Para além disso, essas técnicas exigem que se verifiquem menos pressupostos e o seu desempenho é satisfatório mesmo para um reduzido número de pequenos domínios. Neste contexto, vários métodos de reamostragem têm vindo a ser propostos no contexto da estimação em pequenos domínios. Veja-se por exemplo o método *jackknife* proposto por Jiang, Lahiri e Wan (2002) e o método *bootstrap* introduzido por Butar e Lahiri (2003) para estimação da incerteza do EBLUP do parâmetro de interesse em pequenos domínios. Veja-se também os recentes desenvolvimentos do método *jackknife* apresentadas por Chen e Lahiri (2008), do *bootstrap* paramétrico apresentadas por González-Manteiga et al (2005, 2007) e por Hall e Maiti (2006a), e do *bootstrap* não paramétrico apresentadas por Hall e Maiti (2006b). Contudo, os referidos métodos ainda não foram utilizados no contexto de modelos temporais de nível área para estimação em pequenos domínios. Este trabalho apresenta uma extensão da metodologia *jackknife* ponderada para estimar o EQMP do EBLUP temporal, utilizando ideias semelhantes às de Chen e Lahiri (2008). Um estudo por simulação compara o desempenho dos estimadores *jackknife* com a aproximação analítica do EQMP do EBLUP temporal, para diferentes valores dos parâmetros de variância.

O artigo está organizado da seguinte forma. Na secção 2 é apresentado o modelo de Rao-Yu. Na secção 3 é discutida a estimação do EQMP do EBLUP temporal e apresentada a sua aproximação analítica. Na secção 4 é introduzida a metodologia *jackknife* ponderada para estimar o EQMP do EBLUP temporal. O estudo por simulação com o objectivo de comparar o desempenho dos estimadores *jackknife* com a aproximação analítica do EQMP do EBLUP temporal é apresentado na secção 5. Finalmente, a secção 6 contém as principais conclusões deste trabalho.

2 Modelo de Rao-Yu

O modelo de nível área proposto por Rao e Yu (1994) é o seguinte modelo em três níveis:

$$y_{it} = \theta_{it} + e_{it} \tag{1}$$

$$\theta_{it} = \mathbf{x}'_{it}\beta + \nu_i + u_{it} \tag{2}$$

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1 \tag{3}$$

onde θ_{it} é o parâmetro de interesse referente ao i-ésimo pequeno domínio no período t (i=1,..., m; t=1,..., T) e representa-se por y_{it} o seu estimador directo não enviesado no desenho, e_{it} são os erros da sondagem normalmente distribuídos, conhecido θ_{it}, com média 0 e variância conhecida σ^2_{it}, $\mathbf{x}_{it}(p \times 1)$ é um vector coluna de variáveis auxiliares e $\beta(p \times 1)$ é um vector coluna de parâmetros de regressão. Tem-se ainda que ν_i são efeitos aleatórios específicos de domínio independentes com $\nu_i \sim N(0, \sigma_\nu^2)$ e u_{it} são efeitos aleatórios específicos de domínio-tempo independentes com $\varepsilon_{it} \sim N(0, \sigma^2)$, seguindo um processo AR(1) para cada i. Assume-se que os erros $\{e_{it}\}$, $\{\nu_i\}$ e $\{u_{it}\}$ são mutuamente independentes. Combinando o modelo de erro de sondagem (1) com o modelo de ligação (2) e com o processo AR(1), Rao e Yu (1994) obtiveram o seguinte modelo:

$$y_{it} = \mathbf{x}_{it}'\beta + \nu_i + u_{it} + e_{it}, \qquad u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1. \tag{4}$$

Rao e Yu (1994) mostraram que o modelo (4) pode ser apresentado na forma matricial como:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{v} + \mathbf{u} + \mathbf{e} \tag{5}$$

onde $\mathbf{y} = col_{1\leq i\leq m}(\mathbf{y}_i)$, $\mathbf{y}_i = col_{1\leq t\leq T}(y_{it})$, $\mathbf{X} = col_{1\leq i\leq m}(\mathbf{X}_i)$, $\mathbf{X}_i = col_{1\leq t\leq T}$ $(\mathbf{x}_{it}')$, $\mathbf{Z} = \mathbf{I}_m \otimes \mathbf{1}_T$, $\mathbf{v} = col_{1\leq i\leq m}(\nu_i)$, $\mathbf{u} = col_{1\leq i\leq m}(\mathbf{u}_i)$, $\mathbf{u}_i = col_{1\leq t\leq T}(u_{it})$, $\mathbf{e} = col_{1\leq i\leq m}(\mathbf{e}_i)$, $\mathbf{e}_i = col_{1\leq t\leq T}(e_{it})$, $\mathbf{I}_m$ é uma matriz identidade de ordem m e $\mathbf{1}_T(T\times 1)$ é um vector coluna de uns. Adicionalmente, $\mathbf{e}$, $\mathbf{v}$ e $\mathbf{u}$ são mutuamente independentes, com $\mathbf{u} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}_m \otimes \mathbf{\Gamma}\right)$, $\mathbf{v} \sim N\left(\mathbf{0}, \sigma_\nu^2\mathbf{I}_m\right)$ e $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, onde $\mathbf{\Gamma}(T\times 1)$ é uma matriz com elementos $\rho^{|i-j|}/\left(1-\rho^2\right)$ e $\mathbf{R} = diag_{1\leq i\leq m;1\leq t\leq T}\left(\sigma_{it}^2\right)$. Assumindo que os erros da sondagem são independentes satisfazendo $e_{it} \sim N(0,1)$, pode-se verificar que o modelo (5) é um caso particular do modelo linear geral misto com estrutura de covariância homogénea diagonal por blocos, $Cov(\mathbf{y}) = \mathbf{V} = blockdiag_{1\leq i\leq m}(\mathbf{V}_i)$, com $\mathbf{V}_i = \sigma^2\mathbf{\Gamma} + \sigma_\nu^2\mathbf{J}_T + \mathbf{I}_T$. Assumindo que o vector das componentes de variância $\psi = \left(\sigma_\nu^2, \sigma^2, \rho\right)'$ é conhecido, o estimador BLUP de θ_{it} é dado por (Rao e Yu, 1994):

$$\widetilde{\theta}_{it}(\psi) = \mathbf{x}_{it}'\widetilde{\beta} + \left(\sigma_\nu^2\mathbf{1}_T + \sigma^2\gamma_T\right)'\mathbf{V}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\widetilde{\beta}\right) \tag{6}$$

onde $\widetilde{\beta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ e γ_t é a t-ésima linha de Γ. Rao e Yu (1994) estimaram σ_ν^2 e σ^2 através de um método dos momentos (extensão do método III de Henderson (Henderson, 1953)) e consideraram ρ conhecido, de forma a obter o estimador EBLUP de θ_{it}, $\widehat{\theta}_{it}(\widehat{\psi})$. Por esta razão, considera-se daqui em diante que $\psi = \left[\sigma_\nu^2(\rho), \sigma^2(\rho)\right]'$.

3 Aproximação analítica do EQMP do EBLUP temporal

A partir dos trabalhos de Kackar e Harville (1984) e de Prasad e Rao (1990), Rao e Yu (1994) deduziram uma aproximação de segunda ordem do EQMP do EBLUP temporal, quando as componentes de variância são estimadas pelo referido método dos momentos. Assumindo que os erros $\{e_{it}\}$, $\{\nu_i\}$ e $\{u_{it}\}$ são normalmente distribuídos e ρ é conhecido, essa aproximação é dada por:

$$EQMP_{it}(\widehat{\theta}_{it}) \doteq g_{1it}(\psi) + g_{2it}(\psi) + g_{3it}(\psi) \tag{7}$$

onde

$$g_{1it}\left(\psi\right)=\sigma_{\nu}^{2}+\sigma^{2}/\left(1-\rho^{2}\right)-\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)'\mathbf{V}_{i}^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right), \quad (8)$$

$$g_{2it}\left(\psi\right)=\left[\mathbf{x}_{it}-\mathbf{X}_{i}'\mathbf{V}_{i}^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)\right]'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left[\mathbf{x}_{it}-\mathbf{X}_{i}'\mathbf{V}_{i}^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)\right], \quad (9)$$

$$g_{3it}\left(\psi\right)=tr\left(\mathbf{A}_{t}\mathbf{\Sigma}^{*}\right), \quad (10)$$

sendo $\mathbf{\Sigma}^{*}(2\times 2)$ uma matriz de variâncias-covariâncias dos estimadores não enviesados das componentes de variância, $\widehat{\sigma}_{\nu}^{2}(\rho)$ e $\widehat{\sigma}^{2}(\rho)$, e $\mathbf{A}_t$ uma matriz simétrica da mesma dimensão com elementos diagonais $a_{11}=\left[\gamma_t-\mathbf{\Gamma}\mathbf{V}_i^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)\right]'$ $\mathbf{V}_i^{-1}\left[\gamma_t-\mathbf{\Gamma}\mathbf{V}_i^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)\right]$ e $a_{22}=\left[\mathbf{1}_t-\mathbf{J}_t\mathbf{V}_i^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)\right]'\mathbf{V}_i^{-1}$ $\left[\mathbf{1}_t-\mathbf{J}_t\mathbf{V}_i^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)\right]$ e com elementos fora da diagonal $a_{12}=a_{21}=$ $\left[\gamma_t-\mathbf{\Gamma}\mathbf{V}_i^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)\right]'\mathbf{V}_i^{-1}\left[\mathbf{1}_t-\mathbf{J}_t\mathbf{V}_i^{-1}\left(\sigma_{\nu}^{2}\mathbf{1}_{T}+\sigma^{2}\gamma_{t}\right)\right]$. Segundo Rao e Yu (1994), a avaliação dos elementos de $\mathbf{\Sigma}^{*}$, $V\left(\widehat{\sigma}_{\nu}^{2}\right)$, $V\left(\widehat{\sigma}^{2}\right)$ e $Cov\left(\widehat{\sigma}_{\nu}^{2};\widehat{\sigma}^{2}\right)$, é baseada na avaliação da covariância de duas formas quadráticas normalmente distribuídas. Assumindo que ρ é conhecido e que a estimação das componentes de variância é efectuada pelo método III dos momentos de Henderson (1953), $\widehat{\psi}=\left[\widehat{\sigma}_{\nu}^{2}(\rho),\widehat{\sigma}^{2}(\rho)\right]'$, então um estimador aproximadamente centrado do EQMP do EBLUP, com valor esperado corrigido até à ordem $o(m^{-1})$, é dado por (Rao e Yu, 1994):

$$eqmp_{it}^{RY}\left(\widehat{\theta}_{it}\right)=g_{1it}(\widehat{\psi})+g_{2it}(\widehat{\psi})+2g_{3it}(\widehat{\psi}). \quad (11)$$

4 Aproximação *jackknife* ponderada do EQMP do EBLUP temporal

Nesta secção é introduzida uma metodologia alternativa que permite obter uma aproximação do estimador do EQMP do EBLUP temporal através de um procedimento jackknife ponderado. Note-se que a expressão do EQMP dada em (7), envolve expressões exactas para as parcelas g_{1it} e g_{2it}. Contudo, a parcela g_{3it}, a qual representa a variabilidade adicional presente no EBLUP temporal devida à estimação das componentes de variância, não pode ser calculada analiticamente de forma exacta, sendo necessário utilizar aproximações. É neste contexto que é apresentada uma metodologia *jackknife*, baseada nos trabalhos gerais de Jiang, Lahiri e Wan (2002) e Chen e Lahiri (2008), a qual está desenhada para determinar estimativas para g_{3it}. É então deduzida uma aproximação *jackknife* ponderada para o estimador do EQMP do EBLUP sob o modelo temporal de nível área estacionário devido a Rao e Yu (1994). O método *jackknife* consiste genericamente em retirar inicialmente as observações referentes ao j-ésimo pequeno domínio, j=1, ..., m, em seguida estimar, para cada pequeno domínio excluído, os parâmetros do modelo com base no conjunto de dados remanescente e, por último, estimar as diferentes componentes do MSPE.

Considerando-se o modelo de Rao-Yu e a estimação das componentes de variância pelo método dos momentos, uma aproximação *jackknife* ponderada para o estimador do EQMP do EBLUP temporal é dada por:

$$\begin{aligned} eqmp^J_{it}(\widehat{\theta}_{it}) &= g_{1it}(\widehat{\psi}) + g_{2it}(\widehat{\psi}) - \widehat{\mathbf{c}}'_{WJ,t}(\widehat{\psi})\nabla g_{1it}(\widehat{\psi}) + tr(\mathbf{A}_t \upsilon_{WJ,t}) + \\ &+ tr\left\{\mathbf{L}_t(\widehat{\psi})\left[\mathbf{y}_i - \mathbf{X}_i\widehat{\beta}(\widehat{\psi})\right]\left[\mathbf{y}_i - \mathbf{X}_i\widehat{\beta}(\widehat{\psi})\right]'\mathbf{L}'_t(\widehat{\psi})\upsilon_{WJ,t}\right\} \end{aligned} \tag{12}$$

onde $g_{1it}(\psi)$ é dado por (8), $g_{2it}(\psi)$ é dado por (9) e $\mathbf{A}_t$ é uma matriz simétrica de dimensão 2×2 tal como descrito acima. Para além disso, $\nabla g_{1it}(\psi) = \left(\frac{\partial g_{1it}}{\partial\sigma^2}, \frac{\partial g_{1it}}{\partial\sigma^2_\nu}\right)'$ é o gradiente de $g_{1it}(\psi)$ em σ^2 e σ^2_ν, o qual é um vector de dimensão 2×1 com elementos:

$$\frac{\partial g_{1it}}{\partial\sigma^2} = \frac{1}{1-\rho^2} + \left[(\sigma^2_\nu \mathbf{1}_T + \sigma^2\gamma_t)'\mathbf{V}_i^{-1}\mathbf{\Gamma} - 2\gamma'_t\right]\mathbf{V}_i^{-1}(\sigma^2_\nu\mathbf{1}_T + \sigma^2\gamma_t) \tag{13}$$

e

$$\frac{\partial g_{1it}}{\partial\sigma^2_\nu} = 1 + \left[(\sigma^2_\nu \mathbf{1}_T + \sigma^2\gamma_t)'\mathbf{V}_i^{-1}\mathbf{J}_T - 2\mathbf{1}'_T\right]\mathbf{V}_i^{-1}(\sigma^2_\nu\mathbf{1}_T + \sigma^2\gamma_t). \tag{14}$$

Tem-se ainda que $\mathbf{L}_t(\psi) = \left(\frac{\partial \mathbf{b}_t}{\partial\sigma^2}, \frac{\partial \mathbf{b}_t}{\partial\sigma^2_\nu}\right)'$ é uma matriz por blocos de ordem $2T \times 1$, onde $\mathbf{b}_t = \mathbf{V}_i^{-1}\left(\sigma^2_\nu\mathbf{1}_T + \sigma^2\gamma_t\right)$, e com blocos de ordem $T \times 1$ dados por:

$$\frac{\partial \mathbf{b}_t}{\partial\sigma^2} = \mathbf{V}_i^{-1}\left[\gamma_t - \mathbf{\Gamma}\mathbf{V}_i^{-1}(\sigma^2_\nu\mathbf{1}_T + \sigma^2\gamma_t)\right] \tag{15}$$

e

$$\frac{\partial \mathbf{b}_t}{\partial\sigma^2} = \mathbf{V}_i^{-1}\left[\mathbf{1}_T - \mathbf{J}_T\mathbf{V}_i^{-1}(\sigma^2_\nu\mathbf{1}_T + \sigma^2\gamma_t)\right]. \tag{16}$$

Por último, $\widehat{\mathbf{c}}_{WJ,t} = \sum_{e=1}^m w_{et}(\widehat{\psi}_{-e} - \widehat{\psi})$ é um estimador *jackknife* ponderado do enviesamento de $\widehat{\psi}$, i.e., $E(\widehat{\psi}) - \psi$; e $\widehat{\upsilon}_{WJ,t} = \sum_{e=1}^m w_{et}(\widehat{\psi}_{-e} - \widehat{\psi})(\widehat{\psi}_{-e} - \widehat{\psi})'$ é um estimador *jackknife* ponderado da matriz de covariâncias de $\widehat{\psi}$, ambos corrigidos até à ordem $o(m^{-1})$, e onde ψ_{-e} é o estimador de ψ depois de eliminar os dados referentes ao e-ésimo pequeno domínio e w_{et} são ponderadores a satisfazerem a seguinte condição $w_{et} = 1 + o(m^{-1})$.

Como existem diferentes possibilidades de escolha dos ponderadores, decidiu-se utilizar duas possibilidades utilizadas por Chen e Lahiri (2008) no contexto do modelo de Fay-Herriot: $w_{1,et} = (m-1)/m$ e $w_{2,et} = 1 - \mathbf{x}'_{et}\left(\sum_{i=1}^m \sum_{t=1}^T \mathbf{x}_{it}\mathbf{x}'_{it}\right)^{-1}$ $\mathbf{x}_{et}$. Naturalmente que o uso de diferentes ponderadores resulta em diferentes estimadores *jackknife* do EQMP do EBLUP temporal.

5 Estudo por simulação de Monte Carlo

Neste estudo, foram considerados 28 domínios (m=28) e 7 períodos de tempo (T=7). Foi considerado um modelo com p=2, incluindo a constante. Os valores de X_1 foram gerados com base em $X_1 \sim Unif(0;1)$. Admitiu-se que os verdadeiros valores dos efeitos fixos são $\beta = (1,2)'$. Os efeitos aleatórios ν_i foram gerados de forma independente satisfazendo $\nu_i \sim N(0;\sigma_\nu^2)$ com $\sigma_\nu^2 \in \{0,5;1,0\}$. Os efeitos aleatórios u_{it} foram obtidos com base nos valores gerados de forma independente para $\varepsilon_{it} \sim N(0;\sigma^2)$ com $\sigma^2 \in \{0,25;0,5;1,0\}$, e considerando $\rho \in \{0,2;0,4;0,8\}$. O estudo por simulação de Monte Carlo foi realizado com base em L=1000 conjuntos de dados gerados tal como descrito acima. Para cada conjunto de dados foram produzidas estimativas "analíticas"e estimativas *jackknife* com pesos $w_{1,et}$ e $w_{2,et}$ para a terceira componente do EQMP, $g_{3it}(\widehat{\psi})$, i=1,..., 28, t=1,..., 7. Em seguida foram calculadas as estimativas $eqmp_{it}^{RY}(\widehat{\theta}_{it})$, $eqmp_{it}^{J1}(\widehat{\theta}_{it})$ e $eqmp_{it}^{J2}(\widehat{\theta}_{it})$, e o seu respectivo enviesamento relativo (ER), EQM relativo (EQMR) e rácio de enviesamento (RE):

$$eqmp_{it}^{a} = L^{-1}\sum_{l=1}^{L} eqmp_{it}^{a(l)} \tag{17}$$

$$ER_{it} = L^{-1}\sum_{l=1}^{L} \frac{eqmp_{it}^{a(l)} - EQMP_{it}}{EQMP_{it}} \times 100 \tag{18}$$

$$EQMR_{it} = L^{-1}\sum_{l=1}^{L} \frac{(eqmp_{it}^{a(l)} - EQMP_{it})^2}{EQMP_{it}} \times 100 \tag{19}$$

$$RE_{it} = \frac{L^{-1}\sum_{l=1}^{L} eqmp_{it}^{a(l)} - EQMP_{it}}{\sqrt{L^{-1}\sum_{l=1}^{L}(eqmp_{it}^{a(l)} - EQMP_{it})^2}} \tag{20}$$

onde $a \in \{RY, J1, J2\}$ representa os diferentes estimadores do EQMP e $EQMP_{it}$ representa as estimativas empíricas do EQMP do EBLUP. As estimativas empíricas do EQMP do EBLUP, as quais são tomadas como valores de referência, foram calculadas previamente utilizando R=5000 conjuntos de dados, de forma a assegurar melhor precisão nos resultados. Com o objectivo de sintetizar os resultados obtidos no estudo por simulação, foram produzidas três medidas globais para o conjunto dos mT domínios de interesse: a média do valor absoluto do enviesamento relativo, $MVAER = (mT)^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}|ER_{it}|$; a média do EQM relativo, $MEQMR = (mT)^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T} EQMR_{it}$; e a média do rácio de enviesamento, $MER = (mT)^{-1}\sum_{i=1}^{m}\sum_{t=1}^{T}|RE_{it}|$.

Na tabela 1 são apresentados os resultados referentes ao MVAER, MEQMR e MRE dos estimadores em análise quando $\rho = 0,4$. Os resultados correspondentes a $\rho \in \{0,2;0,8\}$ são omitidos porque permitem retirar conclusões globalmente semelhantes às da situação exposta na tabela 1. Os resultados indicam que não existe um estimador que seja uniformemente melhor do que os restantes, em termos de enviesamento, de EQM e de rácio de enviesamento, para todas as combinações de componentes de variância. Os resultados evidenciam também que as aproximações *jackknife* do EQMP do EBLUP temporal apresentam níveis de enviesamento relativo e de EQM relativo da mesma ordem de grandeza que a aproximação analítica, a qual apresenta um melhor desempenho nestas medidas. Pelo contrário, é ainda de salientar que as aproximações *jackknife* tendem a apresentar um melhor desempenho ao nível do rácio de enviesamento.

Tabela 1: MVAER, MEQMR e MRE dos estimadores do EQMP do EBLUP temporal, $\rho = 0,4$

		MVAER(%)			MEQMR(%)			MRE		
σ^2	σ_ν^2	RY	J1	J2	RY	J1	J2	RY	J1	J2
0,25	0,5	31,607	38,391	38,503	4,587	7,559	7,632	0,856	0,815	0,814
	1,0	28,119	33,028	33,086	3,872	5,812	6,005	0,842	0,808	0,794
0,50	0,5	22,858	25,819	25,864	3,394	4,854	4,881	0,812	0,766	0,766
	1,0	20,858	23,231	23,248	3,002	4,809	4,849	0,800	0,702	0,700
1,00	0,5	12,236	13,477	13,508	1,176	1,439	1,446	0,807	0,806	0,806
	1,0	11,549	12,758	12,785	1,108	1,351	1,357	0,798	0,799	0,800

6 Conclusão

Neste artigo é apresentado um procedimento jackknife ponderado para estimar o EQMP do EBLUP temporal. Os resultados indicam que não existe um estimador que seja uniformemente melhor do que os restantes em todas as situações. Contudo, é de salientar que os resultados numéricos também demonstram que os estimadores jackknife constituem uma verdadeira alternativa à aproximação analítica, mesmo para um pequeno número de períodos de tempo e um número moderado de pequenos domínios. Em particular, o uso deste tipo de estimadores por reamostragem é promissor num contexto de estimação em pequenos domínios assistida por modelos longitudinais mais complexos, para os quais é impossível deduzir aproximações analíticas do EQMP do EBLUP.

Referências

[1] Butar, F.B. e Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small area estimators. *Journal of Statistical Planning and Inference*, 112, 63-76.

[2] Chen, S. e Lahiri, P. (2008). On mean squared prediction error estimation in small area estimation problems. *Communications in Statistics - Theory and Methods*, 37, 1792-1798.

[3] Fay, R.E. e Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

[4] González-Manteiga,W., Lombardía,M., Molina,I., Morales,D. e Santamaría,L.(2008).Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, 5, 443-462.

[5] Hall, P. e Maiti, T. (2006a). Nonparametric estimation of mean squared prediction error in nested-error regression models. *Annals of Statistics*, 34, 1733-1750.

[6] Hall, P. e Maiti, T. (2006b). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society: Series B*, 68, 221-238.

[7] Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics*, 9, 226-252.

[8] Jiang, J., Lahiri, P. e Wan, S.-M. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *Annals of Statistics*, 30), 1782-1810.

[9] Kackar, R.N. e Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, 79, 853-862.

[10] Prasad, N.G.N. e Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.

[11] Rao, J.N.K. e Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *The Canadian Journal of Statistics*, 22, 511-528.