

การเปรียบเทียบการใช้ตัวแบบอัตราณณะ เพื่อการพยากรณ์อัตราณณะโดยข้อมูลจากประเทศไทย Comparison of Mortality Rate Modelling and Forecasting by Data from Thailand

จันทิศา บุญมหาสิทธิ์* และสำรวม จงเจริญ

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

ถนนเสรีไทย แขวงคลองจั่น เขตบางกะปิ กรุงเทพมหานคร 10240

Chanthida Bunmahasit* and Samruam Chongcharoen

Graduate School of Applied Statistics, National Institute of Development Administration,

Serithai Road, Khlong Chan, Bang Kapi, Bangkok 10240

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบการใช้ตัวแบบอัตราณณะเพื่อการพยากรณ์อัตราณณะของประชากรไทยในอีก 10 ปีข้างหน้า โดยศึกษาด้วยตัวแบบ 2 ตัวแบบ คือ ตัวแบบลี-คาร์เตอร์ที่ประมาณค่าพารามิเตอร์ด้วยวิธีการแยกด้วยค่าเจาะจง ตัวแบบลี-คาร์เตอร์ที่ประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดและตัวแบบลี-มิลเลอร์ แล้วเปรียบเทียบความแม่นยำของค่าประมาณของอัตราณณะที่ได้จากแต่ละตัวแบบด้วยวิธีการหาความคลาดเคลื่อนสัมบูรณ์ ซึ่งข้อมูลที่ใช้ในการศึกษา คือ จำนวนประชากรและข้อมูลการตายจำแนกตามอายุและเพศของประชากรไทยปี พ.ศ. 2546-2555 จากสำนักงานสถิติแห่งชาติ สำนักงานปลัดกระทรวงกระทรวงสาธารณสุข และกรมการปกครอง กระทรวงมหาดไทย ผลการศึกษาพบว่าตัวแบบลี-คาร์เตอร์ที่ประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดเป็นตัวแบบที่เหมาะสมที่สุดในการพยากรณ์อัตราณณะของประชากรไทยในอีก 10 ปีข้างหน้า นอกจากนี้ยังพบว่าค่าพยากรณ์อัตราณณะในอีก 10 ปีข้างหน้าของทั้ง 2 ตัวแบบ ให้ค่าที่เป็นในทิศทางเดียวกัน คือ มีลักษณะลดลงอย่างช้า ๆ

คำสำคัญ : อัตราณณะ; ตัวแบบลี-คาร์เตอร์; ตัวแบบลี-มิลเลอร์; การแยกด้วยค่าเจาะจง; วิธีภาวะน่าจะเป็นสูงสุด

Abstract

This research aims to comparison of mortality modelling and forecasting the mortality rate of the Thai population in the next 10 years by using 2 projection methods: Lee-Carter model estimation of the parameter by the singular value decomposition (SVD), Lee-Carter model

estimation of the parameter by maximum likelihood estimation (ML) and Lee-Miller models. In addition to comparing the estimates from 2 methods using absolute error. Data used in the study are the number of population and the number of death by age and sex of the year 2003-2012 from the National Statistical Office, Ministry of Public Health and the Ministry of Interior respectively. The results showed that the Lee-Carter model estimation of the parameter by maximum likelihood estimation is the best model to forecast the mortality for Thai population in the next 10 years. Moreover, it also found that the projections of the mortality rates over the next 10 years from 2 methods are in the same feature and slowly decreasing.

Keywords: mortality rate; Lee-Carter model, Lee-Miller model; singular value decomposition; maximum likelihood estimation

1. บทนำ

ในอดีตประชากรไทยมีอัตราการเกิดและการตายแต่ละปีอยู่ในระดับสูงใกล้เคียงกัน เนื่องจากผู้นิยามมีลูกกันมาก การแพทย์และสาธารณสุขยังไม่เจริญ แต่ในภายหลังจากการแพทย์และสาธารณสุขมีความก้าวหน้าเพิ่มขึ้น เป็นผลให้สุขภาพอนามัยของประชากรไทยดีขึ้น ดังนั้นการตายของประชากรจึงลดลงและชีวิตของประชากรไทยยืนยาวขึ้น อีกทั้งการเปลี่ยนแปลงนี้ยังแสดงให้เห็นถึงการพัฒนาด้านเศรษฐกิจ สังคม สิ่งแวดล้อมของประเทศได้ ด้วยเหตุนี้การตายจึงเป็นองค์ประกอบพื้นฐานของการเปลี่ยนแปลงขนาดและโครงสร้างของประชากร เพราะฉะนั้นจึงมีการวิเคราะห์รูปแบบแนวโน้มสาเหตุการตาย ซึ่งสามารถนำไปประยุกต์ใช้ศึกษาในสาขา เช่น ประชากรศาสตร์ สาธารณสุขศาสตร์ เศรษฐศาสตร์ อีกทั้งในธุรกิจ ประกันชีวิต จำเป็นต้องศึกษาแนวโน้มการเปลี่ยนแปลงการตายของประชากร อัตราการตายของประชากร เพื่อกำหนดเป้าหมาย นโยบายทางด้านต่าง ๆ [1]

ในอดีตมีนักวิชาการได้สร้างและนำเสนอตัวแบบทางคณิตศาสตร์หลายท่าน เช่น ในปี ค.ศ. 1992 Lee และ Carter [2] ได้เสนอวิธีการในการพยากรณ์ระดับและรูปแบบของการตายของอายุต่าง ๆ ในระยะ

ยาวโดยนำเอาพื้นฐานด้านการวิเคราะห์อนุกรมเวลา (time series analysis) และวิธีการถดถอยอย่างง่าย (simple method) มาประยุกต์รวมกันเพื่อหาข้อสรุปเกี่ยวกับการแจกแจงของอัตราการตายตามอายุ ต่อมาในปี ค.ศ. 2001 Lee และ Miller [3] ได้ดัดแปลงวิธีของลี-คาร์เตอร์เพื่อปรับปรุงแก้ไขความผิดพลาดและปัญหาจากของตัวแบบลี-คาร์เตอร์ โดยปรับให้ดัชนีการเปลี่ยนแปลงของเวลาให้มีการแจกแจงแบบปัวซอง นอกจากนี้ในปี ค.ศ. 2002 Brouhns และคณะ [4,5] ได้นำเสนอการประมาณค่าพารามิเตอร์ของตัวแบบลี-คาร์เตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (maximum likelihood method) ซึ่งให้จำนวนการตายมีการแจกแจงแบบปัวซอง ส่วนในประเทศไทยเมื่อปี พ.ศ. 2549 ณิช และ สุวณี [6] ได้ใช้ข้อมูลจำนวนประชากรและจำนวนตายระหว่างปี พ.ศ. 2503-2547 จำแนกตามเพศและอายุ เพื่อพยากรณ์อัตราการตายของประชากรไทยโดยใช้ตัวแบบพยากรณ์อัตราการตายของลี-คาร์เตอร์ รวมถึงมีนักวิชาการอีกหลาย ๆ คนศึกษาตัวแบบอัตราการตายที่เหมาะสมกับแต่ละประเทศนั้น ๆ ดังนั้นในงานวิจัยนี้จะนำเสนอการตัวแบบอัตราการตายของประชากรไทยจากตัวแบบ 2 ตัวแบบ ดังนี้

1.1 ตัวแบบลี-คาร์เตอร์ (Lee-Carter model,

LC) ที่ประมาณพารามิเตอร์ด้วย

1.1.1 วิธีการแยกด้วยค่าเจาะจง (singular value decomposition, SVD)

1.1.2 วิธีภาวะน่าจะเป็นสูงสุด (maximum likelihood method, ML)

1.2 ตัวแบบลี-มิลเลอร์ (Lee-Miller model, LM)

โดยเลือกศึกษาและเปรียบเทียบระหว่างการประมาณพารามิเตอร์ด้วยวิธีการแยกด้วยค่าเจาะจง (LC-SVD) กับการประมาณพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (LC-ML) ว่าตัวแบบใดให้ค่าประมาณค่าพารามิเตอร์ที่ใกล้เคียงกันมากกว่า และวิธีใดให้ค่าคลาดเคลื่อนน้อยกว่ากัน และตัวแบบลี-มิลเลอร์ที่ถูกพัฒนาจากตัวแบบลี-คาร์เตอร์นั้นให้ค่าการประมาณค่าอัตราการมอดิฟายหรือไม่ และวิธีใดให้ค่าคลาดเคลื่อนน้อยกว่ากัน หลังจากนั้นเปรียบเทียบความถูกต้องและความแม่นยำของการประมาณค่าอัตราการมอดิฟายของตัวแบบ ว่าการประมาณค่าอัตราการมอดิฟายของตัวแบบใดสามารถพยากรณ์ค่าอัตราการมอดิฟายได้ใกล้เคียงกับค่าจริงมากที่สุด แล้วนำตัวแบบนั้นมาพยากรณ์ค่าอัตราการมอดิฟายของประชากรไทยในอนาคต

2. อุปกรณ์และวิธีการ

2.1 ข้อมูลและแหล่งที่มาของข้อมูล

การศึกษาประสิทธิภาพของตัวแบบจะใช้ข้อมูลจำนวนประชากรและจำนวนการตายที่ใช้ในการศึกษาทั้งเพศชายและเพศหญิงเป็นข้อมูลทุติยภูมิจำนวน 10 ปี คือ พ.ศ. 2546-2555 ซึ่งได้มาจากสำนักงานสถิติแห่งชาติ สำนักงานปลัดกระทรวงกระทรวงสาธารณสุข และกรมการปกครอง กระทรวงมหาดไทย โดยแยกตามเพศและรายอายุ ตั้งแต่อายุต่ำกว่า 1 ปี จนถึงอายุมากกว่า 100 ปี รวมมีทั้งหมด 102 รายอายุ

2.2 ตัวแบบลี-คาร์เตอร์

ลี-คาร์เตอร์ [2] เสนอตัวแบบที่แสดงแนวโน้มการตายในอนาคตจากอัตราการมอดิฟายปีของอนุกรม โดยกำหนดให้ $m_{x,t}$ แทนอัตราการมอดิฟายปีของประชากรอายุ x ปี ในปี t [6] นั่นคือ

$$m_{x,t} = e^{a_x + b_x k_t + \varepsilon_{x,t}} \quad (1)$$

โดยที่ a_x แทนค่าคงที่ที่กลางปีประชากรอายุ x ปี; b_x แทนอัตราเสื่อมของดัชนีเวลาที่อายุ x ปี; k_t แทนดัชนีการเปลี่ยนแปลงของเวลาในปี t ; $\varepsilon_{x,t}$ แทนความคลาดเคลื่อนของลอการิทึมของอัตราการมอดิฟายปีของประชากรอายุ x ปี ในปี t ที่กำหนดให้มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น σ_ε^2 ของประชากรในแต่ละอายุ x ปี; t แทนปีที่ใช้ในการศึกษา ซึ่ง $t = t_1, t_2, \dots, T$; x แทนอายุที่ใช้ในการศึกษา ซึ่ง $x = x_1, x_2, \dots, X$

และสามารถเขียนอยู่ในรูปของลอการิทึมของอัตราการมอดิฟายปี $[\ln(m_{x,t})]$ ของประชากรอายุ x ปี ในปี t ได้ดังนี้

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} \quad (2)$$

การประมาณค่าพารามิเตอร์

ตัวแบบลี-คาร์เตอร์ข้างต้นพบว่า a_x , b_x และ k_t เป็นค่าคงที่ที่ไม่ทราบค่า ดังนั้นในการหาค่าอัตราการมอดิฟายปี ($m_{x,t}$) ก่อน ซึ่งหาได้จากข้อมูลของสำนักงานสถิติแห่งชาติ สำนักงานปลัดกระทรวงกระทรวงสาธารณสุข และกรมการปกครอง กระทรวงมหาดไทย นั่นคือ

$$m_{x,t} = \frac{D_{x,t}}{L_{x,t}} \quad (3)$$

โดยที่ $D_{x,t}$ แทนจำนวนการตายกลางปีของประชากรอายุ x ปี ในปี t ; $L_{x,t}$ แทนจำนวนประชากรกลางปีของประชากรอายุ x ปี ในปี t ; $m_{x,t}$ แทนอัตราการมอดิฟายปีของประชากรอายุ x ปี ในปี t

การประมาณค่าพารามิเตอร์ a_x , b_x และ k_t จากตัวแบบลิ-คาร์เตอร์ เพื่อให้ได้คำตอบที่แน่นอนและมีเพียงคำตอบเดียวจึงกำหนดเงื่อนไขเพิ่มเติม 2 เงื่อนไข คือ $\sum_{t=t_1}^T k_t = 0$ และ $\sum_{x=x_1}^X b_x = 1$

บทความวิจัยนี้จะเสนอวิธีการประมาณค่าพารามิเตอร์ 2 วิธี ดังนี้

2.2.1 วิธีการแยกด้วยค่าเจาะจง (singular value decomposition, SVD)

วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีการแยกด้วยค่าเจาะจง มีวิธีการดำเนินงานดังนี้

กำหนดให้ $\hat{m}_{x,t}$ แทนค่าประมาณของอัตราการณกลางปีของประชากรอายุ x ปี ในปีที่ t ประมาณค่าพารามิเตอร์ $\hat{a}_x$ จาก

$$\hat{a}_x = \frac{1}{T} \sum_{t=1}^T \ln(m_{x,t}) \quad (4)$$

โดยที่ $\hat{a}_x$ แทนค่าประมาณค่าเฉลี่ยลอการิทึมของอัตราการณกลางปีของประชากรในแต่ละอายุ x ปี; T แทนจำนวนปีของข้อมูลการศึกษา

จากสมการประมาณค่าของตัวแบบ

$$\ln(m_{x,t}) = \hat{a}_x + b_x k_t \quad (5)$$

และสามารถเขียนตัวแบบใหม่ได้ดังนี้

$$Z_{x,t} = \ln(m_{x,t}) - \hat{a}_x = b_x k_t \quad (6)$$

งานวิจัยของลิ-คาร์เตอร์ [2] นั้นนำเสนอวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีการแยกด้วยค่าเจาะจง ซึ่งวิธีนี้จะเป็นการแก้ปัญหาโดยการแยกเมตริกซ์ Z ซึ่งเป็นเมตริกซ์ของสมการเชิงเส้นเป็น 3 เมตริกซ์ คือ U, W และ V โดยให้ x เป็นแถว แทนอายุ $x = x_1, x_2, \dots, X$ และ t เป็นคอลัมน์ แทนจำนวนปี $t = t_1, t_2, \dots, T$ ดังนั้นวิธี SVD เมตริกซ์ Z เป็นเมตริกซ์ที่มีขนาด $X \times T$ สามารถเขียนในรูปผลการคูณของ 3 เมตริกซ์ นั่นคือ เมตริกซ์ทแยงตั้งฉากแนวตั้ง (column-orthogonal matrix)

U ขนาด $X \times T$ คูณกับเมตริกซ์ทแยง (diagonal matrix) W ขนาด $T \times T$ ที่มีสมาชิกเป็นจำนวนบวกหรือศูนย์เท่านั้น และคูณกับทรานสโพสของเมตริกซ์ตั้งฉากแนวนอน (transposed orthogonal matrix) V' ขนาด $T \times T$ [7] ได้ดังนี้

$$Z_{X \times T} = U_{X \times T} W_{T \times T} V'_{T \times T} \quad (7)$$

โดยที่ U แทนเมตริกซ์ตั้งฉากปกติ (orthonormal matrix) ที่มาจากเวกเตอร์เจาะจงของ ZZ' ; V แทนเมตริกซ์ตั้งฉากปกติ (orthonormal matrix) ที่มาจากเวกเตอร์เจาะจงของ $Z'Z$; W แทนบล็อกซ์เมตริกซ์ทแยง (diagonal matrix) ที่ w_j ; $j = 1, 2, \dots, r$ คือรากที่สองของค่าเจาะจง (eigenvalues) ของ Z

$$Z = \begin{pmatrix} \ln(m_{x_1,t_1}) - \hat{a}_{x_1} & \ln(m_{x_1,t_2}) - \hat{a}_{x_1} & \dots & \ln(m_{x_1,T}) - \hat{a}_{x_1} \\ \ln(m_{x_2,t_1}) - \hat{a}_{x_2} & \ln(m_{x_2,t_2}) - \hat{a}_{x_2} & \dots & \ln(m_{x_2,T}) - \hat{a}_{x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \ln(m_{x,T_1}) - \hat{a}_X & \ln(m_{x,T_2}) - \hat{a}_X & \dots & \ln(m_{x,T}) - \hat{a}_X \end{pmatrix}$$

$$= \begin{pmatrix} u_{1,1} & u_{1,2} & \dots & u_{1,T} \\ u_{2,1} & u_{2,2} & \dots & u_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ u_{X,1} & u_{X,2} & \dots & u_{X,T} \end{pmatrix} \begin{pmatrix} w_{1,1} & 0 & \dots & 0 \\ 0 & w_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{T,T} \end{pmatrix} \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,T} \\ v_{2,1} & v_{2,2} & \dots & v_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ v_{T,1} & v_{T,2} & \dots & v_{T,T} \end{pmatrix}'$$

จากข้างต้นจะได้ว่า $\hat{b}_x$ คือ คอลัมน์ที่ 1 ของเมตริกซ์ U สำหรับทุกค่า x (อายุ) นั่นคือ $u_{1,1}, u_{2,1}, u_{3,1}, \dots, u_{X,1}$; $\hat{k}_t$ คือ แถวที่ 1 คอลัมน์ที่ 1 ของเมตริกซ์ W คูณกับคอลัมน์ที่ 1 ของเมตริกซ์ V สำหรับทุก ๆ ค่า t (ปี) นั่นคือ $w_{1,1}v_{1,1}, w_{1,1}v_{2,1}, w_{1,1}v_{3,1}, \dots, w_{1,1}v_{T,1}$

2.2.2 วิธีที่น่าจะเป็นสูงสุด (maximum likelihood method, ML)

Natacha และคณะ [5] นำเสนอวิธีการประมาณค่าพารามิเตอร์ของตัวแบบลิ-คาร์เตอร์แบบใหม่ นั่นคือ วิธีที่น่าจะเป็นสูงสุด โดยวิธีนี้กำหนดให้จำนวนการตายกลางปีของประชากร ($D_{x,t}$) มีการแจก

แจกแจงแบบปัวซอง นั่นคือ $D_{x,t} \sim \text{Poisson}(L_{x,t}, m_{x,t})$ จะได้ฟังก์ชันภาวะน่าจะเป็น เท่ากับ

$$L(a_x, b_x, k_t) = \prod_{x,t} P(D_{x,t}) = \prod_{x,t} \frac{(L_{x,t} m_{x,t})^{D_{x,t}} e^{-L_{x,t} m_{x,t}}}{D_{x,t}!} = \prod_{x,t} \frac{(L_{x,t} e^{a_x + b_x k_t})^{D_{x,t}} e^{-L_{x,t} e^{a_x + b_x k_t}}}{D_{x,t}!} \quad (8)$$

สามารถประมาณค่าพารามิเตอร์ a_x , b_x และ k_t โดยการหาค่าสูงสุดจากลอการิทึมของ

$$\log L(a_x, b_x, k_t) = \sum_{x,t} [D_{x,t} (a_x + b_x k_t) - L_{x,t} e^{a_x + b_x k_t}] \quad (9)$$

ซึ่งหาได้โดยหาอนุพันธ์ (Differentiate) $\log L(a_x, b_x, k_t)$ เทียบกับพารามิเตอร์ a_x , b_x และ k_t แล้วให้ผลลัพธ์เท่ากับศูนย์ จะได้สมการเชิงเส้น $v+1$ สมการที่สามารถหาค่าของพารามิเตอร์ได้จากการทำวนซ้ำ (Iteration) ด้วยวิธีของนิวตัน-ราฟสัน (Newton-Raphson method) โดยมีวิธีการดำเนินงานดังนี้

$$\hat{\theta}^{(v+1)} = \hat{\theta}^{(v)} - \frac{\frac{\partial L^{(v)}}{\partial \theta}}{\frac{\partial^2 L^{(v)}}{\partial \theta^2}} \quad \text{โดยที่ } L^{(v)} = L^{(v)}(\hat{\theta}^{(v)}) \quad (10)$$

กำหนดให้ $\hat{a}_x^{(0)} = 0$, $\hat{b}_x^{(0)} = 1$ และ $\hat{k}_t^{(0)} = 0$

$$\hat{a}_x^{(v+1)} = \hat{a}_x^{(v)} - \frac{\sum_{t=1}^T (D_{x,t} - \hat{D}_{x,t}^{(v)})}{-\sum_{t=1}^T \hat{D}_{x,t}^{(v)}}, \quad \hat{b}_x^{(v+1)} = \hat{b}_x^{(v)}, \quad \hat{k}_t^{(v+1)} = \hat{k}_t^{(v)} \quad (11)$$

$$\hat{k}_t^{(v+2)} = \hat{k}_t^{(v+1)} - \frac{\sum_{x=1}^X (D_{x,t} - \hat{D}_{x,t}^{(v+1)}) \hat{b}_x^{(v+1)}}{-\sum_{x=1}^X \hat{D}_{x,t}^{(v+1)} (\hat{b}_x^{(v+1)})^2}, \quad \hat{a}_x^{(v+2)} = \hat{a}_x^{(v+1)}, \quad \hat{b}_x^{(v+2)} = \hat{b}_x^{(v+1)} \quad (12)$$

$$\hat{b}_x^{(v+3)} = \hat{b}_x^{(v+2)} - \frac{\sum_{t=1}^T (D_{x,t} - \hat{D}_{x,t}^{(v+2)}) \hat{k}_t^{(v+2)}}{-\sum_{t=1}^T \hat{D}_{x,t}^{(v+2)} (\hat{k}_t^{(v+2)})^2}, \quad \hat{a}_x^{(v+3)} = \hat{a}_x^{(v+2)}, \quad \hat{k}_t^{(v+3)} = \hat{k}_t^{(v+2)} \quad (13)$$

โดยที่ $\hat{D}_{x,t}^{(v)} = L_{x,t} e^{\hat{a}_x^{(v)} + \hat{b}_x^{(v)} \hat{k}_t^{(v)}}$ เราจะทำวนซ้ำไปเรื่อย ๆ โดยการสมมติค่าคำตอบในตอนเริ่มต้นใหม่เพื่อนำไปปรับค่าคำตอบในรอบถัดไป จะทำงานกว่าคำตอบมีค่าเปลี่ยนแปลงน้อยมากหรือมีค่าลู่ออกเข้าหาค่าเดียวกัน

การพยากรณ์ค่าประมาณ $\hat{k}_t$

หลังจากประมาณค่าพารามิเตอร์ $\hat{a}_x$, $\hat{b}_x$ และ $\hat{k}_t$ ได้แล้ว เนื่องจากเราต้องการศึกษาอัตราการมรณะในอีก 10 ปีข้างหน้า ทำให้เราต้องพยากรณ์ค่าประมาณของพารามิเตอร์ $\hat{k}_t$ ซึ่งมีอยู่ทั้งหมด T ค่า ในการพยากรณ์ค่าประมาณ $\hat{k}_t$ จะใช้ตัวแบบ $ARIMA(p, d, q)$ ที่เขียนอยู่ในรูปของ

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q} \quad (14)$$

การพยากรณ์ค่าอัตราการระยะกลางปีในอีก

10 ปีข้างหน้า

นำค่าพยากรณ์ $\hat{k}_{t+s}; s=1,2,...,10$

มาคำนวณหาค่าพยากรณ์อัตราการระยะกลางปีในอีก 10 ปีข้างหน้า ด้วยตัวแบบลี-คาร์เตอร์ ซึ่งกำหนดให้

$\tilde{m}_{x,t+s}$ แทนค่าพยากรณ์อัตราการระยะกลางปีของ

ประชากรอายุ x ปี ในปี t ดังนี้

$$\tilde{m}_{x,t+s} = e^{\hat{a}_x + \hat{b}_x \hat{k}_{t+s}}; s=1,2,...,10 \quad (15)$$

2.3 ตัวแบบลี-มิลเลอร์

ลี-มิลเลอร์ [3] ได้ปรับปรุงแก้ไขตัวแบบอัตราการระยะของตัวแบบลี-คาร์เตอร์ เพื่อลดความผิดพลาดของการพยากรณ์ค่าอัตราการระยะ โดยการกำหนดวิธีการประมาณค่าพารามิเตอร์ k_t ใหม่ นั่นคือ k'_t ซึ่งให้มีการแจกแจงแบบปัวซอง สามารถเขียนตัวแบบลี-มิลเลอร์อยู่ในรูปดังต่อไปนี้

$$\ln(m_{x,t}) = a_x + b_x k'_t \quad (16)$$

โดยที่ $m_{x,t}$ แทนอัตราการระยะกลางปีของประชากรอายุ x ปี ในปี t ; a_x แทนค่าคงที่กลางปีประชากรอายุ x ปี; b_x แทนอัตราเสื่อมของดัชนีเวลาที่อายุ x ปี; k'_t แทนดัชนีการเปลี่ยนแปลงของเวลาในปี t ซึ่งเป็นระดับของการมรณะ โดยกำหนดให้มีการแจกแจงแบบปัวซอง

การประมาณค่าพารามิเตอร์

จากตัวแบบลี-มิลเลอร์ข้างต้น พบว่าค่าพารามิเตอร์ a_x และ b_x ของตัวแบบลี-มิลเลอร์นั้นจะเป็นค่าเดียวกับการประมาณค่าพารามิเตอร์ a_x และ b_x จากตัวแบบลี-คาร์เตอร์ที่ประมาณค่าพารามิเตอร์ด้วยวิธีการแยกเงาเงา แต่ค่าพารามิเตอร์ k'_t มีการกำหนดให้เป็นตัวแบบถดถอยปัวซอง ดังนั้นจะประมาณค่าพารามิเตอร์ k'_t ด้วยวิธีภูวณะน่าจะเป็นสูงสุด โดยการแก้สมการด้วยวิธีของนิวตัน-ราฟสัน (Newton-Raphson method) โดยมีวิธีการเดียวกับตัวแบบลี-คาร์เตอร์ในสมการที่ (12)

การพยากรณ์ค่าประมาณ k'_t

การพยากรณ์ค่าอัตราการระยะกลางปีใน

อีก 10 ปีข้างหน้า นั้น เราต้องทำพยากรณ์ค่าดัชนีเวลา k'_t ในอนาคต ซึ่งทำในทำนองเดียวกับตัวแบบลี-คาร์เตอร์ในสมการที่ (14)

การพยากรณ์ค่าอัตราการระยะกลางปีในอีก

10 ปีข้างหน้า

นำค่าพยากรณ์ $\hat{k}'_{t+s}; s=1,2,...,10$

มาคำนวณหาค่าพยากรณ์อัตราการระยะกลางปีในอีก 10 ปีข้างหน้า ด้วยตัวแบบลี-มิลเลอร์ ซึ่งกำหนดให้ $\tilde{m}_{x,t+s}$ แทนค่าพยากรณ์อัตราการระยะกลางปีของประชากรอายุ x ปี ในปี t ดังนี้

$$\tilde{m}_{x,t+s} = e^{\hat{a}_x + \hat{b}_x \hat{k}'_{t+s}}; s=1,2,...,10 \quad (17)$$

2.4 เกณฑ์การประเมินความแม่นยำของการประมาณค่า

พิจารณาการเปรียบเทียบค่าอัตราการระยะระหว่างค่าอัตราการระยะที่ได้ในแต่ละตัวแบบกับค่าอัตราการระยะจริงที่ได้ข้อมูลจากสำนักงานสถิติแห่งชาติ สำนักงานปลัดกระทรวง กระทรวงสาธารณสุข และกรมการปกครอง กระทรวงมหาดไทย โดยวิธีการหาค่าความคลาดเคลื่อนสัมบูรณ์ นั่นคือ

$$\mathcal{E} = |A_t - F_t| \quad (18)$$

โดยที่ A_t แทนค่าอัตราการระยะจากข้อมูลจริง; F_t แทนค่าประมาณของอัตราการระยะที่ได้ในแต่ละตัวแบบ; $\mathcal{E}$ แทนผลต่างค่าอัตราการระยะที่ได้ในแต่ละตัวแบบกับค่าอัตราการระยะจริง

จากนั้นเปรียบเทียบว่าในแต่ละรายอายุตัวแบบใดให้ค่าคลาดเคลื่อนที่น้อยที่สุด แล้วนับว่าแต่ละตัวแบบมีค่าคลาดเคลื่อนน้อยที่สุดที่รายอายุ

3. ผลการวิจัย

3.1 ผลการประมาณค่าอัตราการระยะจากตัวแบบ 2 ตัวแบบ

หลังจากคำนวณค่าอัตราณกระจกกลางปีที่ได้ข้อมูลจากสำนักงานสถิติแห่งชาติ สำนักงานปลัดกระทรวง กระทรวงสาธารณสุข และกรมการปกครอง กระทรวงมหาดไทย กับค่าประมาณอัตราณกระจกกลางจากตัวแบบ LC-SVD, LC-ML และ LM พบว่าในทุก ๆ ปีของการศึกษาค่าอัตราณกระจกกลางปีในช่วงอายุต่ำกว่า 1 ปี จนถึงอายุต่ำกว่า 90 ปี จากตัวแบบทั้ง 2 ตัวแบบ มีค่าต่างจากค่าจริงน้อยมาก แต่เมื่ออายุ 90 ปีขึ้นไป ค่าอัตราณกระจกกลางปีของ 2 ตัวแบบ มีค่าต่างจากค่าจริงอย่างเห็นได้ชัด งานวิจัยนี้ขอเสนอตัวอย่างค่า

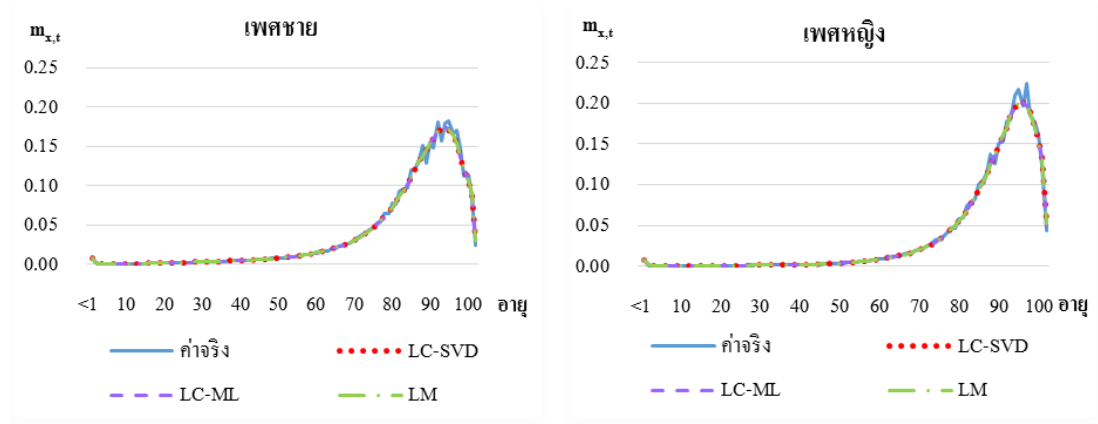
อัตราณกระจกกลางปีเปรียบเทียบกับค่าประมาณอัตราณกระจกจากตัวแบบ LC-SVD, LC-ML และ LM ของเพศชายและเพศหญิงในปี พ.ศ. 2550 ตามตารางที่ 1 และรูปที่ 1

3.2 ผลการประเมินประสิทธิภาพของแต่ละตัวแบบ

เมื่อนำค่าอัตราณกระจกของแต่ละตัวแบบหาความคลาดเคลื่อนแต่ละรายอายุของทุก ๆ ปี ที่ศึกษาพบว่าจากตารางที่ 2 ในปี พ.ศ. 2546 ค่าคลาดเคลื่อนที่ใกล้เคียงกับค่าอัตราณกระจกจริงมากที่สุดเพศ

ตารางที่ 1 เปรียบเทียบของอัตราณกระจกกลางปีกับ LC-SVD, LC-ML และ LM ทั้งเพศชายและเพศหญิงแต่ละอายุ ในปี พ.ศ. 2550

อายุ	เพศชาย				เพศหญิง			
	ค่าจริง	LC-SVD	LC-ML	LM	ค่าจริง	LC-SVD	LC-ML	LM
น้อยกว่า 1 ปี	0.008463	0.008207	0.008205	0.008198	0.006847	0.006926	0.006948	0.006945
1 ปี	0.001091	0.001164	0.001192	0.001156	0.000952	0.001037	0.001102	0.001056
2 ปี	0.000738	0.000837	0.000844	0.000833	0.000524	0.000588	0.000606	0.000595
3 ปี	0.000660	0.000664	0.000670	0.000661	0.000500	0.000479	0.000492	0.000485
4 ปี	0.000637	0.000622	0.000633	0.000619	0.000443	0.000429	0.000440	0.000434
5 ปี	0.000565	0.000623	0.000636	0.000620	0.000396	0.000421	0.000433	0.000426
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90 ปี	0.147314	0.159715	0.159887	0.159758	0.153453	0.161335	0.161217	0.161611
91 ปี	0.180708	0.168357	0.168811	0.168571	0.178437	0.169119	0.169217	0.169078
92 ปี	0.156576	0.171479	0.172222	0.171639	0.180290	0.185351	0.185672	0.185290
93 ปี	0.179691	0.175112	0.175402	0.175269	0.210073	0.194262	0.194792	0.194097
94 ปี	0.182954	0.170956	0.171792	0.171327	0.216588	0.202271	0.202967	0.201703
95 ปี	0.165416	0.164200	0.164471	0.164665	0.196649	0.200841	0.201275	0.199933
96 ปี	0.169991	0.153402	0.153505	0.154015	0.223881	0.196491	0.196500	0.195138
97 ปี	0.148493	0.136281	0.136291	0.136999	0.184927	0.190053	0.189503	0.188059
98 ปี	0.111927	0.120973	0.121646	0.122031	0.176086	0.171239	0.170927	0.168835
99 ปี	0.114114	0.105045	0.105747	0.105965	0.162700	0.154478	0.153714	0.152254
100 ปี	0.083265	0.088351	0.088526	0.089132	0.123921	0.134347	0.133796	0.131849
มากกว่า 100 ปี	0.023835	0.029859	0.028343	0.030219	0.043244	0.053355	0.049785	0.051922



รูปที่ 1 เปรียบเทียบอัตราภาระจริงกลางปีกับ LC-SVD, LC-ML และ LM ของเพศชายและเพศหญิง ในปี พ.ศ. 2550

ชาย คือ ตัวแบบ LC-ML ซึ่งมีจำนวน 49 รายอายุ จากทั้งหมด 102 รายอายุ ดังนั้นสามารถสรุปได้ว่าตัวแบบที่มีค่าคลาดเคลื่อนรายอายุน้อยที่สุดของเพศชาย คือ ตัวแบบ LC-ML ซึ่งมีจำนวน 7 ปี คือ พ.ศ. 2546-2548

และปี พ.ศ. 2552-2555 ส่วนในเพศหญิง คือ ตัวแบบ LC-ML ซึ่งมีจำนวน 6 ปี คือ พ.ศ. 2547-2548 และปี พ.ศ. 2552-2555

ตารางที่ 2 จำนวนรายอายุที่ให้ค่าอัตราภาระของตัวแบบ LC-SVD, LC-ML และ LM ใกล้เคียงกับค่าอัตราภาระจริงมากที่สุดของเพศชายและเพศหญิง

ปี	ชาย			หญิง		
	LC-SVD	LC-ML	LM	LC-SVD	LC-ML	LM
พ.ศ. 2546	33	49	20	43	42	17
พ.ศ. 2547	31	50	21	33	52	17
พ.ศ. 2548	31	51	20	34	44	24
พ.ศ. 2549	48	33	21	33	28	41
พ.ศ. 2550	39	28	35	38	27	37
พ.ศ. 2551	42	29	31	53	16	33
พ.ศ. 2552	31	52	19	31	50	21
พ.ศ. 2553	39	43	20	23	47	32
พ.ศ. 2554	30	42	30	33	41	28
พ.ศ. 2555	35	57	10	23	46	33

3.3 ผลการพยากรณ์อัตราภาระ

ตารางที่ 2 พบว่าตัวแบบที่เหมาะสมในการพยากรณ์อัตราภาระกลางปีของประชากรไทยใน พ.ศ. 2556-2565 คือ ตัวแบบลี-คาร์เตอร์ที่ประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (LC-ML) ในงานวิจัยนี้ขอนำเสนอตัวอย่างค่าพยากรณ์อัตราภาระกลางปีตัวแบบ LC-ML เพศชายและเพศหญิงแต่ละอายุในปี พ.ศ. 2557, 2560 และ 2563 ตามตารางที่ 3 จะได้ว่าในทุก ๆ ปี ของการพยากรณ์อัตราภาระพบว่า

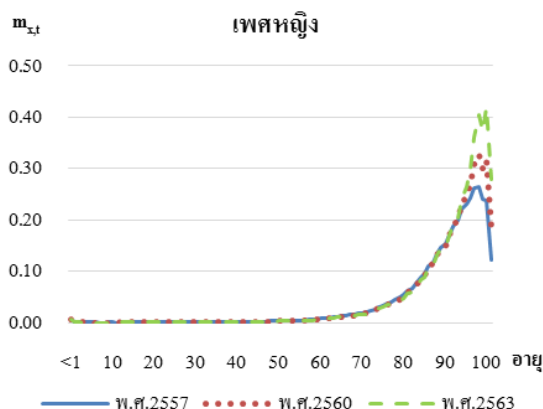
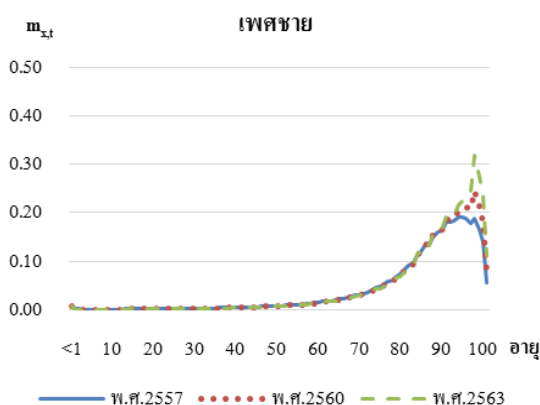
ที่ช่วงอายุต่ำกว่า 1 ปี จนถึงอายุ 90 ปี จะมีค่าพยากรณ์อัตราภาระที่ใกล้เคียงกัน แต่หลังจากอายุ 90 ปี เมื่อปีที่พยากรณ์เพิ่มขึ้นค่าอัตราภาระมีแนวโน้มเพิ่มขึ้น และเมื่อเปรียบเทียบระหว่างเพศชายและเพศหญิงพบว่าในช่วงอายุต่ำกว่า 1 ปี จนถึงอายุ 90 ปี เพศชายจะมีค่าพยากรณ์อัตราภาระสูงกว่าเพศหญิง แต่หลังจากอายุ 90 ปี เพศหญิงจะมีค่าพยากรณ์อัตราภาระสูงกว่าเพศชาย

ตารางที่ 3 ค่าพยากรณ์อัตราภาระกลางปี ($\tilde{m}_{x,t}$) ตัวแบบ LC-ML เพศชายและเพศหญิงแต่ละอายุใน พ.ศ. 2557, 2560 และ 2563

อายุ	เพศชาย			เพศหญิง		
	พ.ศ. 2557	พ.ศ. 2560	พ.ศ. 2563	พ.ศ. 2557	พ.ศ. 2560	พ.ศ. 2563
น้อยกว่า 1 ปี	0.007662	0.007394	0.007134	0.006393	0.006157	0.005929
1 ปี	0.000791	0.000639	0.000516	0.000677	0.000543	0.000435
2 ปี	0.000656	0.000576	0.000505	0.000431	0.000370	0.000317
3 ปี	0.000520	0.000456	0.000400	0.000353	0.000304	0.000262
4 ปี	0.000455	0.000383	0.000322	0.000326	0.000284	0.000248
5 ปี	0.000462	0.000391	0.000331	0.000321	0.000281	0.000245
⋮	⋮	⋮	⋮	⋮	⋮	⋮
90 ปี	0.160075	0.160173	0.160272	0.154060	0.150929	0.147862
91 ปี	0.177773	0.182628	0.187616	0.167235	0.166347	0.165464
92 ปี	0.179840	0.183939	0.188132	0.183847	0.183027	0.182211
93 ปี	0.181610	0.184930	0.188311	0.196023	0.196582	0.197142
94 ปี	0.188517	0.197862	0.207671	0.214128	0.219373	0.224747
95 ปี	0.186026	0.198348	0.211487	0.222755	0.233205	0.244144
96 ปี	0.184186	0.202522	0.222683	0.227338	0.242825	0.259367
97 ปี	0.175601	0.200377	0.228649	0.241940	0.270191	0.301740
98 ปี	0.181351	0.223274	0.274890	0.240132	0.280025	0.326545
99 ปี	0.161995	0.202290	0.252608	0.221960	0.262066	0.309418
100 ปี	0.137008	0.172000	0.215930	0.215795	0.267851	0.332465
มากกว่า 100 ปี	0.052372	0.072106	0.099276	0.097306	0.131738	0.178355

นอกจากนี้สามารถสรุปได้ว่าในทุก ๆ ปี ค่าอัตราการณกลางปีสำหรับทั้งเพศชายและเพศหญิงมีแนวโน้มในลักษณะเดียวกัน คือ ตั้งแต่ช่วงอายุต่ำกว่า 1 ปี จนถึงอายุ 1 ปี ค่าพยากรณ์อัตราการณมีแนวโน้ม

ลดลงและจะมีแนวโน้มคงที่จนถึงอายุ 50 ปี ต่อมาจะเพิ่มขึ้นอย่างรวดเร็วจนถึงอายุ 95 ปี และหลังจากนี้ค่าพยากรณ์อัตราการณมีแนวโน้มลดลง ดังรูปที่ 2



รูปที่ 2 ค่าพยากรณ์อัตราการณกลางปี ($\tilde{m}_{x,t}$) ตัวแบบ LC-ML ของเพศชายและเพศหญิง ใน พ.ศ. 2557, 2560 และ 2563

4. สรุป

การศึกษาตัวแบบลิ-คาร์เตอร์ที่ประมาณพารามิเตอร์ด้วยวิธีการแยกด้วยค่าเจาะจง ตัวแบบลิ-คาร์เตอร์ที่มีการประมาณพารามิเตอร์ด้วยวิธีภูวณะน่าจะเป็นสูงสุด และตัวแบบลิ-มิลเลอร์ พบว่าพารามิเตอร์ a_x และ b_x มีค่าใกล้เคียงกันมาก แต่พารามิเตอร์ k_t มีค่าแตกต่างกันเล็กน้อย และเปรียบเทียบอัตราการณจริงกับค่าประมาณอัตราการณของแต่ละตัวแบบ พบว่าตัวแบบที่ให้อัตราการณที่เหมาะสมสำหรับประชากรไทยทั้งเพศชายและเพศหญิงในปี พ.ศ. 2546-2555 คือ ตัวแบบลิ-คาร์เตอร์ที่ประมาณพารามิเตอร์ด้วยวิธีภูวณะน่าจะเป็นสูงสุด เนื่องจากเป็นตัวแบบที่ให้ค่าความคลาดเคลื่อนจากค่าจริงน้อยที่สุด ดังนั้นจึงพยากรณ์ค่าอัตราการณกลางปีในปี พ.ศ. 2556-2565 ด้วยตัวแบบลิ-คาร์เตอร์ที่ประมาณพารามิเตอร์ด้วยวิธีภูวณะน่าจะเป็นสูงสุด จะได้ว่าค่า

พยากรณ์อัตราการณทั้งเพศชายและเพศหญิงเป็นไปในทิศทางเดียวกันคือมีค่าลดลงอย่างช้า ๆ ต่อเนื่องไป นอกจากนี้ยังพบว่าแนวโน้มอัตราการณที่ได้จากตัวแบบมีความสอดคล้องกับลักษณะการเปลี่ยนแปลงทางด้านการแพทย์ของประเทศไทยที่มีการพัฒนาก้าวหน้าขึ้นทุกปี ทำให้อัตราการเสียชีวิตของประชากรมีแนวโน้มที่จะลดลงตามไปด้วย

5. ข้อเสนอแนะ

5.1 จากข้อมูลจำนวนการตายพบว่ามีส่วนของข้อมูลที่ไม่ทราบอายุ ซึ่งผู้ศึกษาจึงต้องปรับข้อมูลที่ไม่ทราบอายุ โดยใช้วิธีการกระจายข้อมูลในกลุ่มที่ไม่ทราบอายุนี้อย่างอายุต่าง ๆ ตามน้ำหนักในแต่ละอายุนั้น ๆ ซึ่งอาจเกิดความคลาดเคลื่อนได้ เพราะ ฉะนั้นหากเป็นไปได้ผู้ที่ต้องการศึกษาต่อควรศึกษาหาวิธีการปรับข้อมูลที่เหมาะสม

5.2 จากการศึกษาในส่วนของการพยากรณ์ผู้ศึกษาพบปัญหาในเรื่องการขาดแคลนข้อมูลในส่วนของดัชนีเวลาเนื่องจากการศึกษาครั้งใช้ข้อมูลประชากรและข้อมูลการตายของประชากรย้อนหลังเพียง 10 ปี ซึ่งหากต้องการค่าพยากรณ์ที่มีความแม่นยำมากขึ้น จำเป็นเพิ่มจำนวนปีในการศึกษาย้อนหลังมากกว่านี้

5.3 การเลือกตัวแบบที่เหมาะสมที่สุดในการพยากรณ์อัตราการมรณะในอนาคต ในการศึกษานี้ผู้ศึกษาเลือกพยากรณ์อัตราการมรณะโดยเลือกจากค่าความคาดเคลื่อนระหว่างค่าอัตราการมรณะของแต่ละตัวแบบกับค่าจริงที่ต่ำที่สุดในแต่ละรายอายุ แต่เพื่อให้มีความแม่นยำมากขึ้นอาจแบ่งพยากรณ์ค่าอัตราการมรณะเป็นช่วงอายุตามตัวแบบที่เหมาะสมที่สุด เช่น วัยเด็ก วัยกลางคน วัยสูงอายุ

6. เอกสารอ้างอิง

- [1] ปราโมทย์ ประสาทกุล และปัทมา ว่าพัฒน์วงศ์, 2549, ภาวะการตายและความยืนยาวของชีวิตประชากรไทย, น. 22-33, IPSR Annual Report 2006 : Mortality, the Reflection of Population Security, สถาบันวิจัยประชากรศาสตร์ และสังคม มหาวิทยาลัยมหิดล, กรุงเทพฯ.
- [2] Lee, R.D. and Carter, L.R., 1992, Modeling and forecasting U.S. mortality, J. Amer. Stat. Assoc. 87: 659-671.
- [3] Lee, R.D. and Miller, T., 2001, Evaluating the performance of the Lee-Carter method for forecasting mortality, Demography 38: 537-549.
- [4] Wilmoth, J.R., 1993, Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change, Technical Report, Department of Demography, University of California, Berkeley, U.S.A.
- [5] Brouhns, N., Denuit, M. and Vermunt, J.K., 2002, A Poisson log-bilinear regression approach to the construction of projected lifetables, Insur. Math. Econ. 31: 373-393.
- [6] นิชา ราชฤทธิ์ และสุวณี สุรเสียงสังข์, 2549, ตัวแบบพยากรณ์อัตราการมรณะของประชากรไทย โดยวิธีของลีและคาร์เตอร์, ว.ประชากรศาสตร์ 22(2): 25-42.
- [7] Baker, K., 2013, Singular Value Decomposition Tutorial, Available Source: http://www.ling.ohiostate.edu/~kbaker/pubs/Singular_Value_Decomposition_Tutorial.pdf, November 17, 2014.