「한국 시스템다이내믹스 연구」 제13권 제2호(2012. 7) : 47-71

Comparison of Three Optimization Methods Using Korean Population Data

오덕교*
Oh, Deok-Kyo

Abstract

The purpose of this research is the examination of validity of data as well as simulation model, i.e. to simulate the real data in the SD model with the least error using the adjustments for the faithful reflection of real data to the simulation. In general, SD programs (e.g. VENSIM) utilize the Euler or Runge-Kutta method as an algorithm. It is possible to reflect the trend of real data via these two estimation methods however can cause the validity problem in case of the simulation requiring the accuracy as they have endogenous errors. In this article, the future population estimated by the Korea National Statistical Office (KNSO) to 2050 is simulated by the aging chain model, dividing the population into three cohorts, 0-14, 15-64, 65 and over cohorts by age and offering the adjustments to them. Adjustments are calculated by optimization with three different methods, optimization in EXCEL, manual optimization with iterative calculation, and optimization in VENSIM DSS, the results are compared, and at last the optimal adjustment set with the least error are found among them. The simulation results with the pre-determined optimal adjustment set are validated by methods proposed by Barlas (1996) and other alternative methods. It is concluded that the result of simulation model in this research has no significant difference from the real data and reflects the real trend faithfully.

Keywords: adjustment, optimization, validity, statistical validity test, significance

---

* Research fellow at Korea Corporate Governance Service (KCGS), Tel: +82-2-3775-3701, e-mail: deokkyo@cgs.or.kr.

Ⅰ. INTRODUCTION

> "The simulation should mimic the reality"
>
> - Anonymous

When we design and implement the simulation model, the most important thing is how well to reflect the reality. Insufficient reflection of reality reduces the confidence for the simulation model itself, and furthermore decreases the application and usage of simulation.

Therefore, proper reflection of reality is essential in implementing the simulation and model developers should try to improve the validity of simulation results, modifying and complementing the model continuously (Forrester 1994).

However, it is very difficult to reflect the real data with high accuracy in system dynamics. As one of popular system dynamics softwares, VENSIM basically utilizes the Euler and Runge-Kutta methods for approximated estimation (Barton and Tobias 1998), the error in the simulation result is inevitable. Therefore, this research is purposed to find the way to obtain the simulation result with high accuracy, adapting the adjustments to the simulation model, as well as to properly validate the simulation result by the right methods.

Ⅱ. LITERATURE REVIEW

Optimization is regarded as one of the key future challengeable field in system dynamics (Yügel and Barlas 2006; Richardson 1999; Coyle 2000) and only several studies have been conducted on simulation-based optimization in the system dynamics domain (Yügel and Barlas 2006). The main reason for the rare optimization application is quoted to be directly related to one of the main characteristics of SD approach: importance of the dynamics pattern observed rather than a value that a system variable takes a point during simulation (Yügel and Barlas 2006). Within system dynamics, the optimization has been utilized primarily in identifying the best range of parameter values for policies in any given model, based on a specified objective function (Duggan 2008).

Model validation is an important, yet controversial aspect of any model-based methodology in general, and system dynamics in particular (Barlas 1996). Validation is the process of establishing confidence in the soundness and usefulness of a model, begins as the model builder accumulates the confidence that a model behaves plausibly and generates problem symptoms or modes of behavior seen in the real system (Forrester and Senge 1980). The ultimate objective of system dynamics model validation is to establish the validity of model structure and the general logical order of validation is, first to test the validity of the structure, and then start testing the behavior accuracy, only after the structure of model is perceived adequate, increasing the confidence in system dynamics models by a wide variety of tests (Barlas 1996; Forrester and Senge 1980).

Statistical significance testing in the model validation is on controversy: In system dynamics validation, there is very little use of statistical significance testing and system dynamics has often been criticized for it and system dynamicists have responded by arguing that statistical significance can contribute very little to model validation in significance (Barlas 1996). As system dynamics models are usually pattern-based rather than accuracy-based, the statistical significance testing is less meaningful than in other simulation methodologies. Sometimes, in “big picture” models, it is very difficult to validate the models due to lots of variables to be tested.

Ⅲ. SIMULATION

1. Data to be used

In this research, the future population estimation results by the Korea National Statistical Office (KNSO) shall be used. According to the future population estimation result report (KNSO 2006), the total population in Korea is estimated to reach 42.3 million in 2050 and 38% of total population will be aged 65 and over (refer to 〈Table 1〉). The total population is estimated to increase until 2018, and then decrease, as shown.

〈Table 1〉 shows us the estimated population structure. The population is classified into three cohorts according to age: age 0-14, 15-64, and 65 and over. The cohort of “0-14” is estimated to decrease while “65 and over” cohort is estimated to increase year by year. The working age population cohort, “15-64”, is estimated to have a different trend. It is expected to increase slightly until 2018 and then drop down to 53% in 2050.

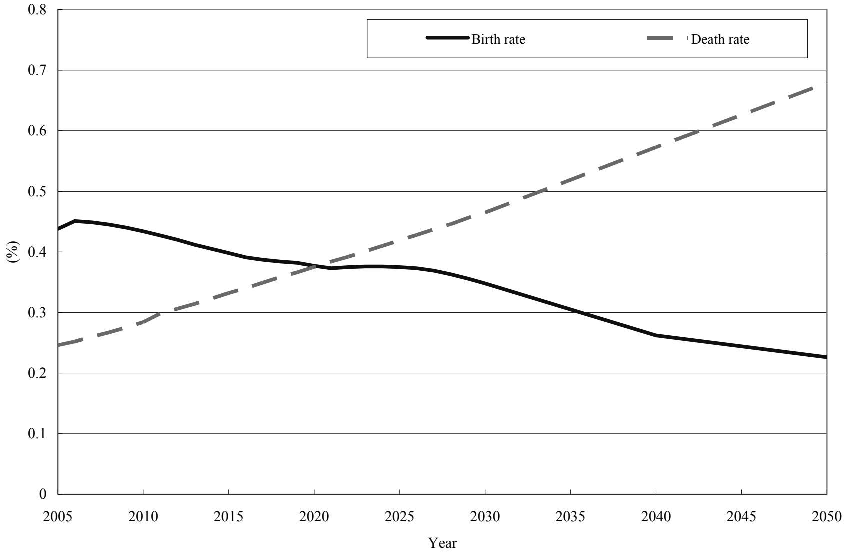
There are three rates to be considered in estimating population: birth rate, death rate

〈Table 1〉 Estimation of total population and growth rate

Year		1980	1990	2000	2005	2010	2018	2020	2030	2050
Total population (In thousands)		38,124	42,869	47,008	48,138	48,875	49,340	49,326	48,635	42,343
Cohort structure (%)	0～14	34	25.6	21.1	19.2	16.2	12.7	12.4	11.4	8.9
	15～64	62.2	69.3	71.7	71.7	72.9	72.9	72	64.4	53
	65 and over	3.8	5.1	7.2	9.1	11	14.3	15.6	24.3	38.2

Source: KNSO (2006).

(mortality rate), and international transfer rate. The birth rate is estimated to decrease while death rate is estimated to increase as shown in [Figure 1] below where the dotted line indicates the death rate and the solid line is the birth rate. The increase in the death rate is caused mainly by an increase in the population of “65 and over” cohort. Although the death rate of “0-14” and “15-64” cohorts decreases, the total population is estimated to decrease.



[Figure 1] Future birth rate and death rate

Data Source: KNSO(2006).

The regression analysis with the KNSO data is performed to obtain equations representing the trend of these two rates for simulation in the time domain. The equations obtained are as follows:

$$BR_t = 9.92 - 4.7 * 10^{-3} * t, \qquad (1)$$

$$DR_t = -18.59 + 9.39 * 10^{-3} * t, \qquad (2)$$

where t is time, and BR_t and DR_t are total birth rate and death rate at time t.

The international transfer rate is estimated to decrease as time progresses (KNSO 2006), which means that Korean people who transfer to foreign countries will decrease while foreigners reside in Korea will increase. The equation for this rate is as follows:

$$ITR_t = 1122531 - 541.289 * t \qquad (3)$$

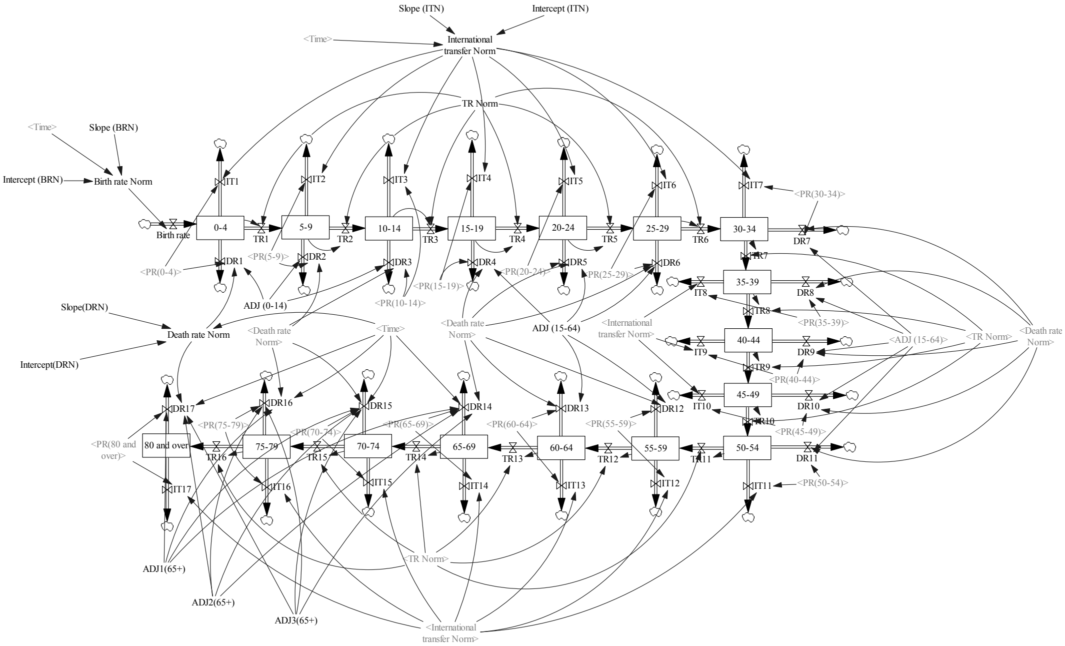
where ITR_t is international transfer rate at time t.

2. Stock and Flow Diagram(SFD)

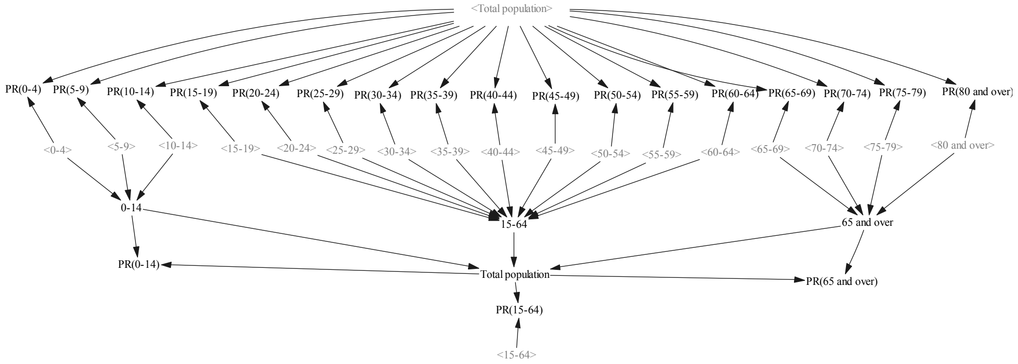
This simulation shall be performed in VENSIM^{TM} program developed by Ventana systems Inc. First, the aging chain model which depicts how the population moves from one stock of age group to the next as they get older (Kim and Goggi 2005) shall be applied to formulate the simulation model. Subordinate cohorts of 5-age interval shall be planted in this model. That is to say, even though, as mentioned above, there are three large cohorts of age “0-14”, “15-64”, and “65 and over.” To analyze, the “0-14” cohort shall be divided into “0-4”, “5-9”, and “10-14.” The “15-64” shall be divided into 10 subordinate cohorts, and the “65 and over” shall be divided into “65-69”, “70-74”, “75-79”, and “80 and over” subordinate cohorts. The population sector in this model is depicted in [Figure 2] and [Figure 3] below. This cohort model starts from the birth rate which is calculated by the birth rate equation (Equation (1)) listed above. The transfer rates between subordinate cohorts can be simply expressed by

$$\frac{P_i^t}{LT}$$

where P_i^t = the population of ith cohort at time t and LT = the lead time = 5 years. The international transfer rates shall be calculated using the same method as the death rate but without adjustment. The total international transfer rate calculated by Equation (3) shall be distributed to each cohort proportional to the population ratio which is calculated by the formula of (cohort population/total population).



[Figure 2] Population sector



[Figure 3] Population ratio

The total death rate shall be calculated by the death rate equation (Equation (2)). The death rate for each cohort cannot be found so the alternative is to calculate each death rate by ***TotalDeathRate**PopulationRatio*adjustment*** and apply it to this model. This alternative equation can be used to distribute the total death rate to the death rate of each cohort proportional to the population ratio.

The populations of subordinate cohorts except the first “0-4” cohort are calculated by

$$P_t^i = \int_{t0}^{t} (TR_t^{i-1} - TR_t^i - DR_t^i - ITR_t^i)dt + P_0^i \quad (i \geq 2 \text{ and } t \geq t_0)$$

where P_t^i is the population of the i$^{\text{th}}$ cohort at time t, P_0^i is the initial value of the i$^{\text{th}}$ cohort, TR_t^{i-1} is the inflow transfer rate from the (i-1)$^{\text{th}}$ cohort to the i$^{\text{th}}$ cohort at t, TR_t^i is the outflow transfer rate of the i$^{\text{th}}$ cohort to the (i+1)$^{\text{th}}$ cohort at t, DR_t^i is the death rate of the i$^{\text{th}}$ cohort at t, and ITR_t^i is the international transfer rate of the i$^{\text{th}}$ cohort at t. We understand that $TR_t^i = 0$ for the “80 and over” cohort because it is the last cohort and t_0=2007 in this model.

The population of the “0-4” cohort is expressed as

$$P^1 = \int_{t_0}^{t} (BR_t - TR_t^1 - DR_t^1 - ITR_t^1)dt + P_0^1$$

where BR_t is the birth rate at t. This model utilizes BR_t instead of TR_t^{i-1} for this cohort since it is the first cohort.

The cohort population of age 65 and over is calculated by

$P_t^{65+} = \sum_{i=14}^{17} P_t^i$ where i means the ordinal number of the subordinate cohort.

Since $P_t^i = \int_{t0}^{t} (TR_t^{i-1} - TR_t^i - DR_t^i - ITR_t^i)dt + P_0^i$, the equation above can be expressed by

$$P_t^{65+} = \sum_{i=14}^{17} \left(\int_{t_0}^{t} (TR_t^{i-1} - TR_t^i - DR_t^i - ITR_t^i)dt + P_0^i \right).$$

This equation can be re-written as

$$P_t^{65+} = P_0^{65+} + \int_{t_0}^{t} \sum_{i=14}^{17} (TR_t^{i-1} - TR_t^i - DR_t^i - ITR_t^i)dt$$

where $P_0^{65} = P_0^{14} + P_0^{15} + P_0^{16} + P_0^{17}$.

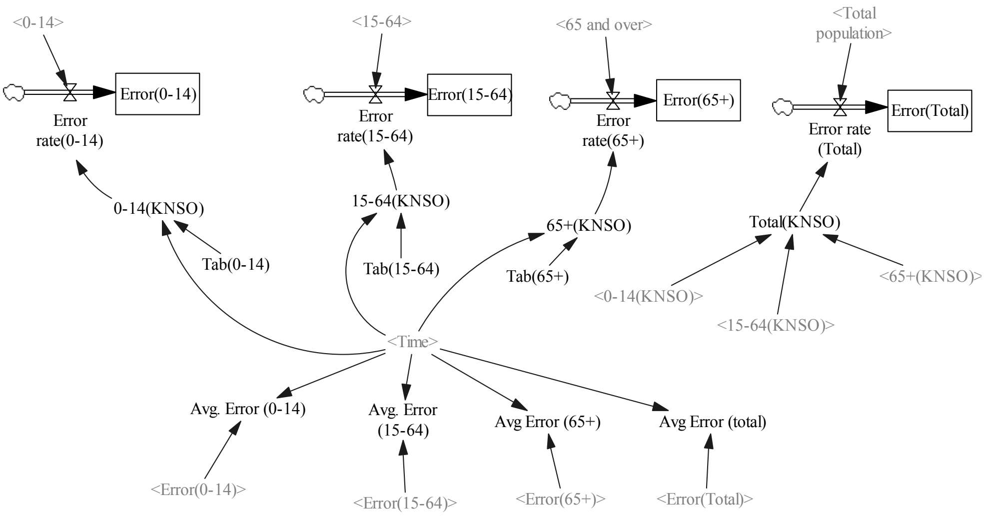
If $OF_t^{65+} = \sum_{i=14}^{17} (DR_t^i + ITR_t^i)$, the equation above yields

$$P_t^{65+} = P_0^{65+} + \int_{t0}^{t} (TR_t^{13} - OF_t^{65+})dt.$$

The flow rate of “65 and over” cohort is

$$\dot{P}_t^{65+} = TR_t^{13} - OF_t^{65+}.$$

To calculate the error in VENSIM, the sub-model for error calculation is inserted into this model, as illustrated in [Figure 4] below. The cohort population data estimated by KNSO are input into the table functions for four categories, “0-14” cohort, “15-64” cohort, “65 and over” cohort, and total population (the sum of cohort population) for the period of 2007 to 2050 years and the simulation results are compared with the KNSO data. The average error is finally calculated by dividing the cumulative error by the elapsed time.

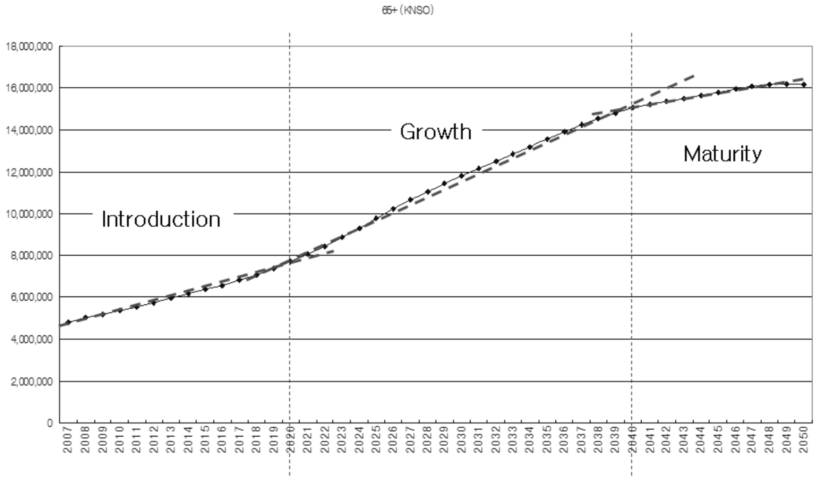


[Figure 4] Error calculation sub model

At last, this simulation is set to start at the year of 2007 and finish at 2050 with the time interval of 0.25 year.

Ⅳ. ADJUSTMENT SETTING

The “adjustment” is necessary to simulate matched with the KNSO data with accuracy, at least 95% accuracy. In other words, adjustment is calculated to estimate with the least error within 5% compared with the KNSO data. To improve the accuracy of this model, the adjustment shall be differentiated by the three large cohorts and by the timeframe for “65 and over” cohort because this cohort is estimated to have different slopes in the back and forth of 2020 and 2040 (refer to [Figure 5]) and two cohorts of “0-14” and “15-64” are revealed to have relatively low errors with the single adjustment in the preliminary test.



[Figure 5] Trend of "65 and over" cohort population

Levin & Roberts(1976) propose the lifecycle of an agency by four stages: start-up, growth, maturity, and decay (pp. 25-26). As shown above, this population is fit to the lifecycle of an agency of Levin and Roberts(1976).

The start-up (introduction) stage shall be the period of 2007 to 2020 years, as the population increases with the slope less than that in the growth stage is defined as the period of 2021 to 2040 years and the period of 2041 to 2050 years is the maturity stage. The approximate slope for each stage is shown in [Figure 5] above. The red dotted lines represent the simplified approximate slopes in three stages.[1] Next, three adjustments for "65 and over" cohort shall be allocated according to the divided timeframes. Optimized adjustments shall be obtained by minimizing the objective functions specified by the errors of "0-14", "15-64", and "65 and over" cohorts, as shown in 〈Table 2〉 below. The Mean Absolute Percent Error (MAPE) is adapted to calculate the model error, which is defined as

$$MAPE = \frac{1}{n}\sum \frac{|X_m - X_d|}{X_d}$$

where X_d is the estimated value by KNSO and X_m is the estimated value in this model (Russell & Taylor 2006 p. 499; Sterman 2000 p. 875).[2]

---

[1] The period selection is implemented by the regression analysis to find the period maximizing R^2.

[2] The popular Mena Squared Error (MSE) can be considered as an alternative (i.e. least square method) however does not express the model error to be compared itself but the error term for calculation. As the model error is the objective to be compared in the long haul, the MAPE is appropriate in this research.

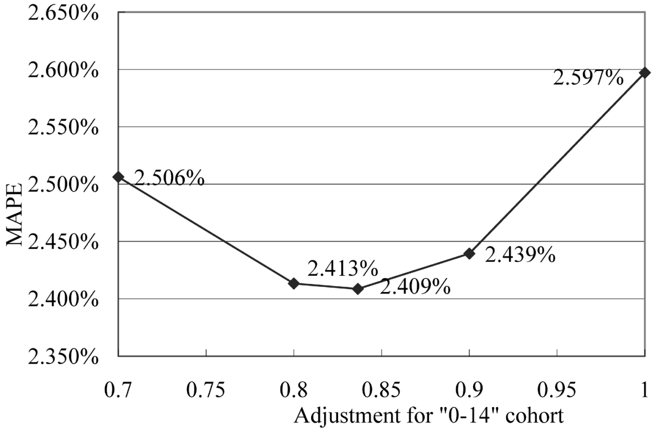
The calculation is implemented from the first inflow in this model, that is, “0-14” cohort. Sequentially, after obtaining the adjustment for the “0-14” cohort, the adjustment for next cohort is calculated. This means as the calculation continues, the adjustments obtained from the previous calculations shall be excluded from the constraints (see the constraints in 〈Table 2〉). The optimization is performed by three methods, the linear programming in MS-EXCEL software, the manual optimization by the iterative calculation in VENSIM, and the automatic optimization in VENSIM DSS. To run the linear programming in MS-EXCEL, the cohort population is also modeled on the spreadsheet. The optimized adjustments are shown in 〈Table 2〉.

〈Table 2〉 Adjustment optimization

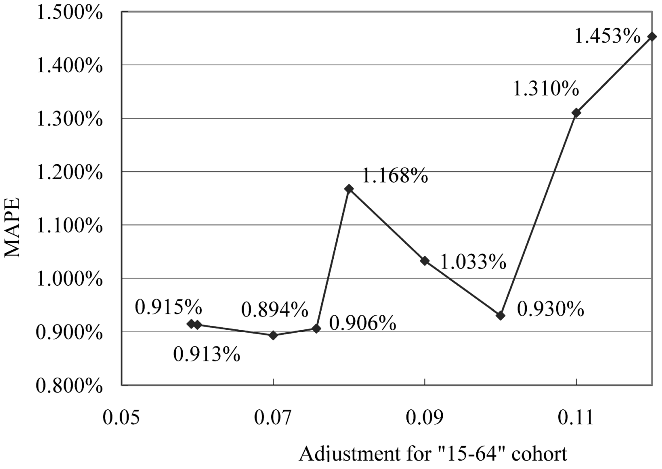
Cohort	0-14	15-64	65 and over
Objective Function	$Min\frac{1}{n}\sum\frac{\lvert P_t^{K1}-P_t^{E1}\rvert}{P_t^{K1}}$	$Min\frac{1}{n}\sum\frac{\lvert P_t^{K2}-P_t^{E2}\rvert}{P_t^{K2}}$	$Min\frac{1}{n}\sum\frac{\lvert P_t^{K3}-P_t^{E3}\rvert}{P_t^{K3}}$
Constraints	$0 \le a \le 2$, $0.05 \le b \le 1$, and c1, c2, c3 ≥ 0	$0.05 \le b \le 1$, and c1, c2, c3 ≥ 0	c1, c2, c3 ≥ 0
Optimal Adjustments (MS-EXCEL)	0.8365 (a)	0.0592 (b)	8.30 (c1) 3.12 (c2) 2.54 (c3)
Optimal Adjustments (Iterative calculation)	0.8365 (a)	0.07 (b)	8.40 (c1) 3.27 (c2) 2.55 (c3)
Optimal Adjustments (VENSIM DSS)	0.838426 (a)	0.0687369 (b)	8.43131 (c1) 3.13579 (c2) 2.45554 (c3)

P_t^{K1} =“0-14” cohort population estimated by KNSO, P_t^{K2} =“15-64” cohort population estimated by KNSO, P_t^{K3} =“65 and over” cohort population estimated by KNSO, P_t^{E1} =“0-14” cohort population estimated in this research, P_t^{E2} =“15-64” cohort population estimated in this research, P_t^{E3} =“65 and over” cohort population estimated in this research, a=adjustment of “0-14” cohort, b=adjustment of “15-64” cohort, c1=adjustment of “65 and over” cohort during 2007-20, c2=adjustment of “65 and over” cohort during 2021-40, and c3=adjustment of “65 and over” cohort during 2041-50.

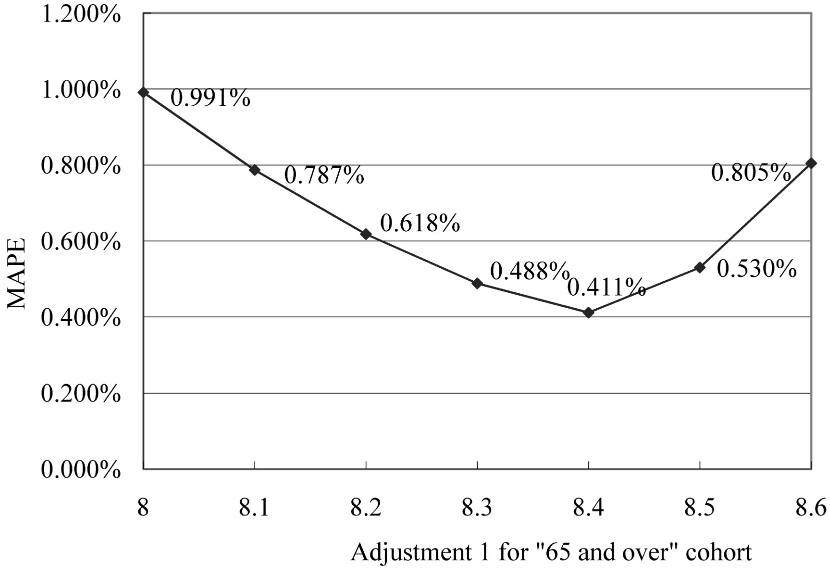
First, we obtained the five adjustments by the optimization on MS-EXCEL: a=0.8365 b=0.0592, c1=8.3, c2=3.12 and c3=2.54. However, as the VENSIM software utilizes the Euler method (in this research) in estimating the values these adjustment values might be tuned for the VENSIM program. We input these adjustments as bases and tuned them with the interval of 0.1 or 0.01. With the iterative calculations, the optimal adjustment set is obtained as a=0.8365, b=0.07, c1=8.4, c2=3.27, and c3=2.25. The corresponding local errors of these adjustments on VENSIM are shown in [Figure 6] through [Figure 10].



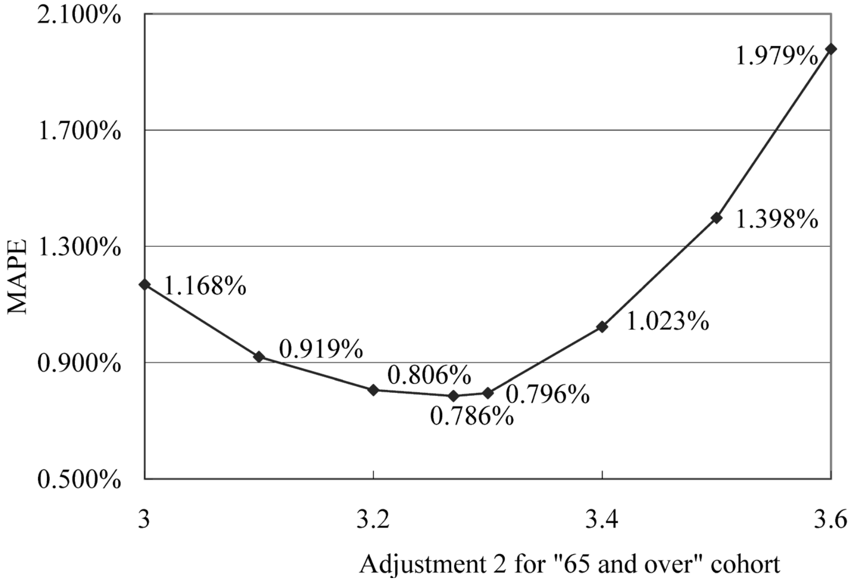
[Figure 6] Error and adjustment for “0-14” cohort



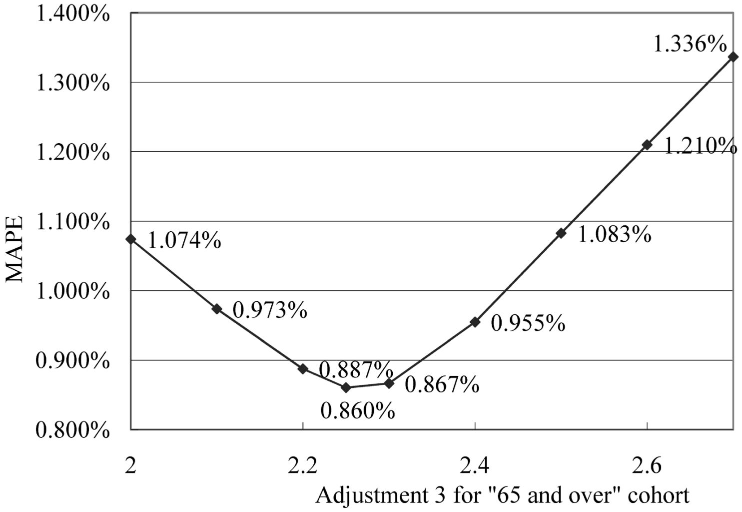
[Figure 7] Error and adjustment for “15-64” cohort



[Figure 8] Error and adjustment (c1) for “65 and over” cohort



[Figure 9] Error and adjustment (c2) for ”65 and over” cohort



[Figure 10] Error and adjustment (c3) for "65 and over" cohort

Finally, the optimization was performed on the DSS version of VENSIM. The objective functions and constraints in ⟨Table 2⟩ were input for optimization. The optimization result is a = 0.838426, b = 0.0687369, c1 = 8.43131, c2 = 3.13579, and c3 = 2.45554, as illustrated in ⟨Table 2⟩. All parameters except c3 are similar with the results of iterative calculation and as the estimation of c3 is the final optimization task so the optimal value of c3 was mitigated a little as other parameters were slightly changed to minimize the error.

We compared the errors of three methods to choose one with the least errors, based on the formulated error calculation algorithm in [Figure 5]. ⟨Table 3⟩ below depicts the errors of three methods. First, as shown, all calculated errors are within 5% and the method with the least error shall be chosen next. The best case with the least errors for cohorts is shown to be the third case that optimized in VENSIM as what to investigate in this research is mainly the cohort population, however the first method shows the least error for the total population because there are compensations between cohort populations as total population is the sum of three cohort populations.

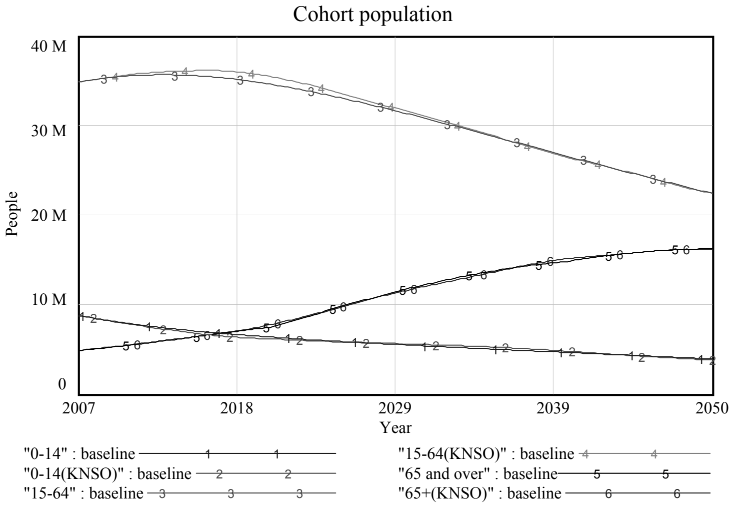
〈Table 3〉 Comparison of optimization methods

Optimization method	0-14 cohort	15-64 cohort	65 and over cohort
MS-EXCEL	2.4%	0.92%	0.85%
Iterative Calculation	2.4%	0.89%	0.86%
VENSIM DSS	2.4%	0.89%	0.76%

Ⅴ. RESULTS AND ANALYSIS

1. Cohort population comparison

Populations of the three large cohorts are compared in [Figure 11]. It tells us that the population of the “0-14” cohort decreases continuously, that of the “15-64” cohort increases for the next 10 years but then decreases until 2050, and that of the “65 and over” increases continuously. Specifically, for the “0-14” cohort, the population eventually decreases from 8.73 million in 2007 to 3.89 million in 2050. For the “15-64” cohort, the population in 2007 is 34.9 million, the maximum is estimated to be 35.7 million in 2012.75, and it is estimated to be 22.4 million in 2050. The population of the “65 and over” cohort in 2050 is estimated to be 16.24 million, over 3 times the cohort population in 2007.



[Figure 11] Comparison of cohort population

2. Validity test

Validity tests are largely categorized into the structure and behavior validity (Barlas 1996) and sometimes three: structure, behavior, and policy (Forrester and Senge 1980). As this simulation model is the purely correlational model built primarily for forecasting purpose (such as time-series or regression models), it can be assessed to be valid if its output matches the real output within some special range of accuracy, without any questioning of the validity of the individual relationships that exist in the model, since there is no claim of causality in structure (Barlas 1996). Therefore, the validity test in this research focuses on the behavior test.

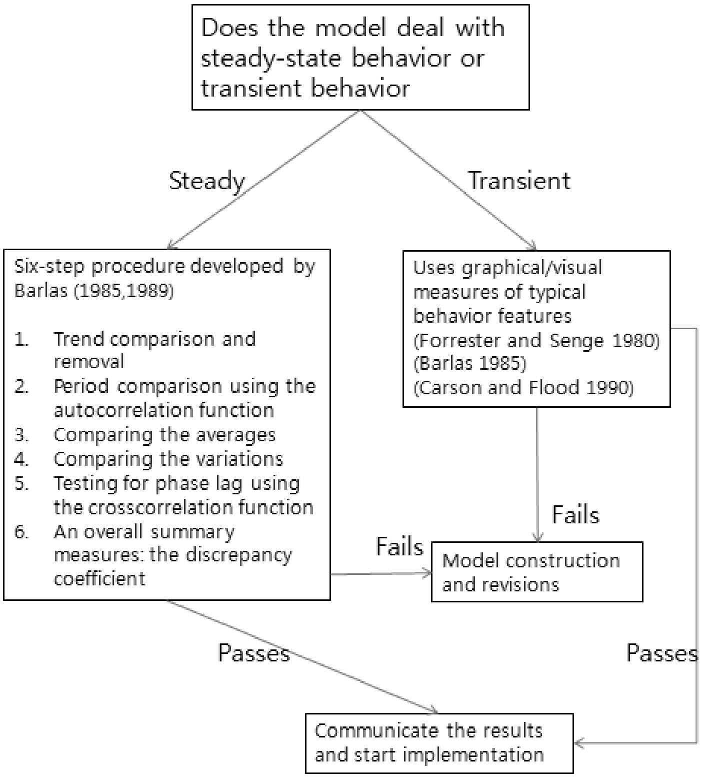
To know the conformity of the simulated value to the actual value (Coyle and Exelby 2000), the behavior pattern tests are performed for the validation in this research, specifically composed of six-step procedure: trend comparison, period comparison (autocorrelation function test), average comparison, variation comparison, phase lag test and overall summary measure (Barlas 1996)(refer to [Figure 12]). For the trend comparison, different from Barlas (1996)'s suggestion, the regression analysis is applied and coefficients of two data are compared to examine how well the trend of simulated data is coincided with that of original KNSO data, and t-and F-tests are utilized for average and variation comparison, respectively, similar with ANOVA. In details of trend comparison, coefficients of simulated data are statistically examined to be within the confidence interval of coefficients of real data with the significance level of $\alpha=0.05$ as regression analysis results. As the population of "15-64" cohort is convex-shaped, the second-order equation is adopted for parameter estimation while for other cohorts the linear equation is adopted. As a result, it is proved that the simulated data are not significantly different from the original data.

In the phase lag test, two cross correlation functions are statistically tested by t-test to verify their conformity3). Test result shows no significant difference between two cross correlation function values.

〈Table 4〉 below illustrates the overall test results, notifying that the simulated data are not significantly different from the original data (for details, see APPENDIX 1 as well as Barlas (1996)).

---

3) Barlas (2006) states that the maximum of the cross correlation function occurs at lag 0 if the output of the model and the actual system are perfectly in phase. In this research, the maximum value occurs at lag 0 and the t-test is implemented additionally for more stringent validation.



[Figure 12] Logical sequence of behavior pattern valuation (Barlas 1996)

〈Table 4〉 Validity test results

Test	Test method	Result
Trend comparison	Linear regression analysis	No significant difference
Period comparison	Sample autocorrelation function	No significant difference
Average comparison	t-test	No significant difference
Variation comparison	F-test	No significant difference
Phase lag	Cross-correlation function	No significant difference
Overall summary measure	Discrepancy coefficient	No significant difference

Ⅵ. CONCLUSIONS

This research has the implication regarding the methodology about adjustment setting and calculation and validity test. The adjustments were adapted for fitting the simulation model to the reality (actual data) with the least errors and three different adjustment sets are compared: the optimization in EXCEL and VENSIM DSS, and manual optimization by iterative calculation. This comparison lets us understand the estimation algorithm in VENSIM which causes the difference from the reality. Furthermore, the stringent validity tests for the simulation model with the optimized adjustment setting were implemented and this simulation model is approved. Throughout these efforts, the accuracy of SD model was maximized enough for the model to reflect the reality statistically. However all SD models do not have to equip the accuracy, this articles has focused on the accuracy improvement in the SD model where the least error or reality reflection is required for validity. The model error should be sufficiently analyzed and discussed in validating the simulation model. Insufficient validation and discussion will cause the model not to be trusted by other people. In this research the example of sufficient validation and discussion has provided, though there are things to be solved statistically.

Though we implemented statistical validity tests, the statistical significance testing has the limitation which means it is very useful only when the null hypothesis is rejected in the case of "Xm=Xt" null hypothesis where Xm represents some measure of the model and Xt corresponds to the same measure of the real systems (Barlas 1996). In this research, the purpose of statistical testing is not to reject but to fail to reject the null hypothesis, as the purpose of testing is to find the simulation data mimicking the actual data sufficiently. Therefore, the estimate which can make this problematic purpose reconcile with the statistical testing is needed for the stringent validity test (Barlas 1996).

In this paper, the model is a kind of agent-based system dynamics model as does not include the feedback loop. It will be pretty challengeable to perform the optimization and find the optimal parameters in the feedback loop model to obtain the simulation result that has no significant difference from the actual data, though it is difficult to design the model.

【REFERENCES】

Akkermans, Henk and Nico Dellaert. (2005). The discovery of industrial dynamics: the contribution of system dynamics to supply chain management in a dynamic and fragment world. *System Dynamics Review* 21(3): 173-86.

Anderson, Virginia, and Lauren Johnson. (1997). *Systems thinking basics: From concepts to causal loops.* MA: Pegasus communications.18-21.

Barlas, Yaman. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review* 12(3): 183-210.

Barlas, Yaman. (2006). BTSII: behavior validity testing software II. User's manual.

Barton, P. M. and A.M. Tobias. (1998). Notes and Insights: Accurate estimation of performance measure for system dynamics models. System Dynamics Review 14(1): 85-94.

Burns, James and Balaji Janamanchi. (2006). Optimal control and optimization of system dynamics models: some experiences and recommendations. *Proceedings of 24th international conference of the system dynamics society*

Cavana, Robert Y. and Edwin D. Mares. (2004). Integrating critical thinking and systems thinking: from premises to causal loops. *System Dynamics Review* 20(3): 223-235.

Cohen, Jacob. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1): 37-46.

Coyle, Geoff. (1998). The practice of system dynamics: milestones, lessons and ideas from 30 years experience. *System Dynamics Review* 14(4): 343-365.

Coyle, Geoff and David Exelby. (2000). The validation of commercial system dynamic models. *System Dynamics Review* 16(1): 27-41.

Dangerfield, Brian and Carole Roberts. (1999). Optimization as a statistical estimation tool: An example in estimating the AIDS treatment-free incubation period distribution. *System Dynamics Review* 15(3): 273-291.

Dogan, Gokhan. (2007). Bootstrapping for confidence interval estimation and hypothesis testing for parameters of system dynamics models. System Dynamics Review 23(4): 415-436.

Duggan, Jim. (2008). Equation-based policy optimization for agent-based system dynamics models. System Dynamics Review 24(1): 97-118.

Forrester, Jay W. (1994). System dynamics, systems thinking, and soft OR. *System Dynamics Review* 10(2): 245-256.

Forrester, Jay W. and Peter M. Senge. (1980). *Tests for building confidence in system dynamics models*. TIMS Studies in the management sciences 14, 209-228.

George, Janet. (2008). "Changing concepts in ageing: a social policy perspective." *Asian Journal of Gerontology & Geriatrics* 3, 40-45.

Homer, Jack and Rogelio Oliva. (2001). Maps and models in system dynamics: a response to Coyle. *System Dynamics Review* 17:4, 347-355.

Johnson, Richard A. and Dean W. Wichern. (2002). *Applied Multivariate Statistical Analysis*. 5th ed. NJ: Prentice-Hall. 354-416, 543-569.

Kim, Hyunjung and James Goggi. (2005). System dynamics modeling for long term care policy. *Proceedings of 23rd international conference of the system dynamics society*.

Korea National Statistical Office. (2005). *Future population estimation result report*.

Lee, Tsuey-Ping. 2005. Examining local social welfare policy implementation using system dynamics perspective. *Proceedings of 23rd international conference of the system dynamics society.*

Levin, Gilbert and Edward B. Roberts. 19976. *The dynamics of human service delivery.* MA: Ballinger publishing company, 26.

Lyneis, James M. (1999). System dynamics for business strategy: a phased approach. *System Dynamics Review* 15(1): 37-70.

Lyneis, James M., and David N. Ford. (2007). System dynamics applied to project management: a survey, assessment, and directions for future research. *System Dynamics Review* 23(2/3): 157-189.

Peterson, David W. and Robert L. Eberlein. (1994). Reality check: a bridge between systems thinking and system dynamics. *System Dynamics Review* 10:2/3, 159-174.

Repenning, Nelson P. (2000). A dynamic model of resource allocation in multi-project research and development systems. *System Dynamics Review* 16(3): 173-212.

Rodrigues, Alexandre. (1994). The role of system dynamics in project management: A comparative analysis with traditional models. *Proceedings of international conference of the system dynamics society*.

Russell, Roberta S. and Bernard W Taylor III. (2006). *Operations management: Quality and competitiveness in a global environment*. 5th edition, NJ: John Wiley and Sons.

Sterman, John D. (2000). *Business dynamics: Systems thinking and modeling for a complex world.*

McGraw-Hill.

Yügel, Gönenç and Yaman Barlas. (2006). Pattern-based system design/optimization. *Proceedings of 24th international conference of the system dynamics society*.

【APPENDIX 1. Validity tests】

1. Trend Comparison

Linear or non-linear trends of populations of three cohorts in both KNSO and simulated data are estimated by the regression analysis and coefficients are compared and examined to know whether coefficients in simulated data are within the confidence interval of coefficients of the raw data. The analysis result shows that as all coefficients of simulated data are within the confidence interval of coefficients of raw data two data are not different with the statistical significance.

〈Table A-1〉 Trend Comparison

	Coefficient of original data	Standard error	t value	P-value	Confidence Interval (95%)		Coefficient of simulated data
					Lower limit	Upper limit	
"0-14" cohort							
Intercept	7775409.46	92229.99	84.31	1.73E-48	7589281.80	7961537.11	7909246.58
time	-95360.94	3693.64	-25.82	2.14E-27	-102814.99	-87906.88	-100622.41
"15-64" cohort							
Intercept	36261273.5	284287.25	127.55	6.31E-55	35687143.77	36835403.23	36006538.22
time (t)	-32094.95	30582.12	-1.05	0.30	-93856.80	29666.90	-61254.85
time (t2)	-7360.88	687.66	-10.70	1.92E-13	-8749.64	-5972.11	-6347.60
"65 and over" cohort							
Intercept	4307360.52	156036.62	27.61	1.50E-28	3992465.87	4622255.16	4299900.88
time	307309.66	6248.97	49.18	9.30E-39	294698.73	319920.59	307447.19

2. Period comparison

Sample autocorrelation function which represents the ratio of the covariance to the variance is applied in this test. To examine whether the values of sample autocorrelation functions are equal of both the original and simulated data, the confidence interval of the difference between function values of two data is utilized (for details, see Barlas (1996)). If the distance (difference) is within the confidence interval, we can't say two data are different, in other words, two are not rejected significantly. The maximum lag must be 1/4 of the number of data, as a rule of thumb (Dogan 2007, Barlas 2006). Therefore autocorrelation function values and test statistic values are calculated for lags k=1 to 10. From 〈Table A-2〉 through 〈Table A-4〉, all distances are within the confidence interval so two data are not significantly different in all cohorts.

〈Table A-2〉 Period comparison for "0-14" cohort

k	D	0-14 (KNSO)	0-14 (SIM)	Se(D)	Lower limit	Upper limit	Within the confidence interval
1	0.01254	0.00058	0.00058	0.03411	- 0.06823	0.03411	Yes
2	0.02604	0.00238	0.00238	0.06899	- 0.13798	0.06899	Yes
3	0.03930	0.00668	0.00543	0.11003	- 0.22007	0.11003	Yes
4	0.05128	0.01163	0.00941	0.14506	- 0.29012	0.14506	Yes
5	0.06117	0.01747	0.01386	0.17701	- 0.35403	0.17701	Yes
6	0.06846	0.02377	0.01828	0.20505	- 0.41010	0.20505	Yes
7	0.07288	0.02977	0.02223	0.22803	- 0.45606	0.22803	Yes
8	0.07401	0.03390	0.02440	0.24146	- 0.48292	0.24146	Yes
9	0.07053	0.03592	0.02509	0.24701	- 0.49402	0.24701	Yes
10	0.06266	0.03650	0.02527	0.24852	- 0.49704	0.24852	Yes

〈Table A-3〉 Period Comparison for “15-64” cohort

k	D	15-64 (KNSO)	15-64 (SIM)	Se(D)	Lower limit	Upper limit	Within the confidence interval
1	- 0.0038	0.0006	0.0006	0.0341	- 0.0682	0.0341	Yes
2	- 0.0068	0.0024	0.0024	0.0690	- 0.1380	0.0690	Yes
3	- 0.0090	0.0020	0.0023	0.0653	- 0.1306	0.0653	Yes
4	- 0.0106	0.0037	0.0042	0.0886	- 0.1771	0.0886	Yes
5	- 0.0118	0.0057	0.0064	0.1101	- 0.2203	0.1101	Yes
6	- 0.0127	0.0077	0.0089	0.1290	- 0.2581	0.1290	Yes
7	- 0.0133	0.0096	0.0114	0.1450	- 0.2900	0.1450	Yes
8	- 0.0129	0.0105	0.0128	0.1526	- 0.3052	0.1526	Yes
9	- 0.0113	0.0109	0.0133	0.1555	- 0.3110	0.1555	Yes
10	- 0.0082	0.0117	0.0140	0.1603	- 0.3205	0.1603	Yes

〈Table A-4〉 Period comparison for “65 and over” cohort

k	D	65 and over (KNSO)	65 and over (SIM)	Se(D)	Lower limit	Upper limit	Within the confidence interval
1	- 0.0006	0.0006	0.0006	0.0341	- 0.0682	0.0341	Yes
2	- 0.0011	0.0024	0.0024	0.0690	- 0.1380	0.0690	Yes
3	- 0.0013	0.0019	0.0020	0.0626	- 0.1253	0.0626	Yes
4	- 0.0012	0.0036	0.0036	0.0851	- 0.1702	0.0851	Yes
5	- 0.0013	0.0056	0.0057	0.1059	- 0.2119	0.1059	Yes
6	- 0.0015	0.0075	0.0078	0.1239	- 0.2478	0.1239	Yes
7	- 0.0017	0.0093	0.0098	0.1382	- 0.2763	0.1382	Yes
8	- 0.0016	0.0101	0.0104	0.1432	- 0.2864	0.1432	Yes
9	- 0.0007	0.0102	0.0103	0.1434	- 0.2867	0.1434	Yes
10	- 0.0012	0.0108	0.0105	0.1463	- 0.2925	0.1463	Yes

3. Mean Comparison

Statistical t-test is applied to compare the means between actual and simulated data. As shown, no cohorts show the significant difference in the mean. The comparison result tells us that there is no significant difference between averages of two data.

〈Table A-5〉 Mean Comparison for three cohorts

	0-14		15-64		65 and over	
	Simulated	Actual	Simulated	Actual	Simulated	Actual
Mean	5745864.773	5725149.318	30731827.27	30981725	10910015.45	10914518.18
Variance	1.739E+12	1.595E+12	1.947E+13	2.162E+13	1.590E+13	1.585E+13
Observation	44	44	44	44	44	44
d.f.	86		86		86	
t-value	0.0753		-0.2586		-0.0053	
P(T<=t) (two-tailed)	0.9402		0.7966		0.9958	

4. Variation Comparison

F-test is applied for the variation comparison. There is no variation difference between actual and simulated data.

〈Table A-6〉 Variation comparison for three cohorts

	0-14		15-64		65 and over	
	Simulated	Actual	Simulated	Actual	Simulated	Actual
Mean	5745864.77	5725149.32	30731827.27	30981725.00	10910015.45	10914518.18
Variance	1.739E+12	1.595E+12	1.947E+13	2.162E+13	1.590E+13	1.585E+13
Observation	44	44	44	44	44	44
d.f.	43	43	43	43	43	43
F-ratio	1.0904		0.9006		1.0030	
P(F<=f)	0.3889		0.3665		0.4961	

5. Phase lag test

The cross correlation function is measured and analyzed for this test. After calculating the cross correlation functions (Simulated-to-Actual data (SA) and Actual-to-Simulated data (AS)), they are statistically compared by t-test (for details, see Barlas (1996)). No pairs are revealed to be significantly different.

〈Table A-7〉 Cross correlation function comparison

	0-14		15-64		65 and over	
	Csa	Cas	Csa	Cas	Csa	Cas
Mean	-0.11985	-0.13676	0.29284	0.33091	-0.69320	-0.69513
Variance	1.62355	1.66798	4.98463	4.65788	1.48792	1.49028
Observation	44	44	44	44	44	44
d.f.	86		86		86	
t-value	0.061835		-0.081338		0.00742	
P(T<=t) (two-tailed)	0.9508		0.9354		0.9941	

6. Overall summary measure

The discrepancy coefficient is utilized for this measure (for details, see Barlas (1996)) and represents two data become consistent as it comes close to zero. The calculated measure is illustrated in 〈Table A-8〉 below and all values are revealed to be close to zero.

〈Table A-8〉 Discrepancy measure

Cohort	0-14	15-64	65 and over
Discrepancy Coefficient	0.0615	0.0364	0.0150