Mixing Unbiased Estimators

Leigh J. Halliwell, FCAS, MAAA
leigh@lhalliwell.com

CASE Fall 2011 Meeting
Georgia State University
October 5, 2011

Simple Problem

- Several uncorrelated random variables of the same mean, but of different variances:

$$X_i \sim \left[\mu, \sigma_i^2\right]$$

Question 1: What linear combination of them will create the best estimator of μ?

♬

- Are you BLUE? Sing the BLUEs 'till BLUE in the face!
  - The <u>L</u>inear <u>E</u>stimator (combination) must be <u>U</u>nbiased
  - The less the variance, the better the <u>LUE</u>. Least is <u>B</u>est.

Answer

- Weight the RVs inversely to their variances:

$$\hat{\boldsymbol{\mu}} = \frac{\dfrac{X_1}{\sigma_1^2} + \mathbf{K} + \dfrac{X_n}{\sigma_n^2}}{\dfrac{1}{\sigma_1^2} + \mathbf{K} + \dfrac{1}{\sigma_1^2}}$$

“Hope y’all knew that!”

Too Easy?? Don’t go home yet …

- Question 2: What’s the estimator’s variance?
- Question 3: Answer Q1&2, if RVs correlated.

Don’t Fret: It’s Multivariate!

AY	Ultimate μ_1		Std Dev	Prediction-Error Variance Σ_1									
1998	20,663,109	±	224,359	5.03E+10	1.13E+09	1.33E+09	1.44E+09	1.50E+09	2.09E+09	2.85E+09	3.64E+09	5.49E+09	8.25E+09
1999	20,824,378	±	228,971	1.13E+09	5.24E+10	2.22E+09	2.32E+09	2.35E+09	2.96E+09	3.74E+09	4.57E+09	6.55E+09	9.38E+09
2000	22,264,182	±	241,730	1.33E+09	2.22E+09	5.84E+10	3.38E+09	3.38E+09	4.04E+09	4.89E+09	5.82E+09	8.03E+09	1.11E+10
2001	21,486,528	±	246,708	1.44E+09	2.32E+09	3.38E+09	6.09E+10	4.33E+09	5.00E+09	5.84E+09	6.80E+09	9.09E+09	1.21E+10
2002	20,487,825	±	249,060	1.50E+09	2.35E+09	3.38E+09	4.33E+09	6.20E+10	5.59E+09	6.41E+09	7.37E+09	9.67E+09	1.27E+10
2003	20,572,601	±	291,372	2.09E+09	2.96E+09	4.04E+09	5.00E+09	5.59E+09	8.49E+10	1.06E+10	1.18E+10	1.46E+10	1.79E+10
2004	20,579,410	±	344,815	2.85E+09	3.74E+09	4.89E+09	5.84E+09	6.41E+09	1.06E+10	1.19E+11	1.74E+10	2.10E+10	2.46E+10
2005	21,381,119	±	398,343	3.64E+09	4.57E+09	5.82E+09	6.80E+09	7.37E+09	1.18E+10	1.74E+10	1.59E+11	2.74E+10	3.16E+10
2006	23,895,269	±	513,711	5.49E+09	6.55E+09	8.03E+09	9.09E+09	9.67E+09	1.46E+10	2.10E+10	2.74E+10	2.64E+11	4.79E+10
2007	25,571,140	±	666,133	8.25E+09	9.38E+09	1.11E+10	1.21E+10	1.27E+10	1.79E+10	2.46E+10	3.16E+10	4.79E+10	4.44E+11

AY	Ultimate μ_2		Std Dev	Prediction-Error Variance Σ_2									
1998	20,780,207	±	175,099	3.07E+10	5.18E+08	7.82E+08	9.10E+08	9.84E+08	1.16E+09	1.63E+09	2.11E+09	2.78E+09	4.69E+09
1999	21,062,679	±	207,420	5.18E+08	4.30E+10	7.25E+09	7.22E+09	7.07E+09	7.36E+09	7.93E+09	8.71E+09	1.02E+10	1.25E+10
2000	22,247,195	±	245,423	7.82E+08	7.25E+09	6.02E+10	1.25E+10	1.22E+10	1.26E+10	1.33E+10	1.43E+10	1.64E+10	1.92E+10
2001	21,577,395	±	265,162	9.10E+08	7.22E+09	1.25E+10	7.03E+10	1.49E+10	1.53E+10	1.60E+10	1.72E+10	1.96E+10	2.25E+10
2002	20,667,840	±	279,038	9.84E+08	7.07E+09	1.22E+10	1.49E+10	7.79E+10	1.69E+10	1.77E+10	1.89E+10	2.15E+10	2.44E+10
2003	20,860,557	±	313,738	1.16E+09	7.36E+09	1.26E+10	1.53E+10	1.69E+10	9.84E+10	2.13E+10	2.27E+10	2.58E+10	2.89E+10
2004	20,871,758	±	403,318	1.63E+09	7.93E+09	1.33E+10	1.60E+10	1.77E+10	2.13E+10	1.63E+11	3.29E+10	3.71E+10	4.09E+10
2005	21,630,036	±	485,897	2.11E+09	8.71E+09	1.43E+10	1.72E+10	1.89E+10	2.27E+10	3.29E+10	2.36E+11	4.87E+10	5.31E+10
2006	23,858,441	±	590,139	2.78E+09	1.02E+10	1.64E+10	1.96E+10	2.15E+10	2.58E+10	3.71E+10	4.87E+10	3.48E+11	7.04E+10
2007	25,642,727	±	857,783	4.69E+09	1.25E+10	1.92E+10	2.25E+10	2.44E+10	2.89E+10	4.09E+10	5.31E+10	7.04E+10	7.36E+11

- Estimates of AY ultimate losses from independent paid and incurred models. Mix this!
- Don’t try this at home, kids. I’m a trained professional.

Multivariate Mixture

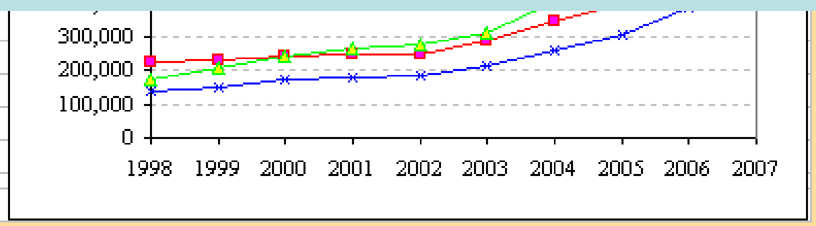
AY	Paid Ult	Incd Ult	Differ	Mixed Ult		Z_{Paid}	Z_{Incd}
1998	20,663,109	20,780,207	117,097	20,736,089		38%	62%
1999	20,824,378	21,062,679	238,301	20,949,478		48%	52%
2000	22,264,182	22,247,195	-16,987	22,238,079		-54%	154%
2001	21,486,528	21,577,395	90,867	21,514,773		69%	31%
2002	20,487,825	20,667,840	180,015	20,557,311		61%	39%
2003	20,572,601	20,860,557	287,957	20,697,355		57%	43%
2004	20,579,410	20,871,758	292,348	20,696,138		60%	40%

$$\Sigma_0^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}$$

$$\Sigma_0^{-1}\hat{\boldsymbol{\mu}}_0 = \Sigma_1^{-1}\hat{\boldsymbol{\mu}}_1 + \Sigma_2^{-1}\hat{\boldsymbol{\mu}}_2$$

You can't mix apples and oranges!

2003	291,372	313,738	212,523
2004	344,815	403,318	261,584
2005	398,343	485,897	307,737
2006	513,711	590,139	387,044
2007	666,133	857,783	525,932
Total	1,475,942	1,845,299	1,146,036



- Inverse-variance matrix weights work
  - Should always test $H_0 : \hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2 = 0_{10\times1} \sim \chi^2_{10}$, because, as we all know …

The Gauss-Markov Theorem

- Minimize $\Phi(\mathrm{X}) = \mathrm{X}'\Sigma\mathrm{X}$ subject to $\mathrm{A}_{m\times n}\mathrm{X}_{n\times p} = \mathrm{B}_{m\times p}$.
  - $\Sigma_{n\times n}$ is positive-definite
  - Rows of A are linearly independent, or $rank(\mathrm{A}) = m$.

- Minimizing a quadratic form subject to a linear constraint
  - If $Var[\mathbf{Z}] = \Sigma$, then $Var[\mathrm{X}'\mathbf{Z}] = \mathrm{X}'\Sigma\mathrm{X}$
  - So, in effect, minimizing variance subject to constraint

- Solution: $$\mathrm{X}_0 = \Sigma^{-1}\mathrm{A}'\left(\mathrm{A}\Sigma^{-1}\mathrm{A}'\right)^{-1}\mathrm{B}$$

$$\Phi(\mathrm{X}_0) = \mathrm{B}'\left(\mathrm{A}\Sigma^{-1}\mathrm{A}'\right)^{-1}\mathrm{B}$$

Corollary: The Linear Statistical Model

$$\begin{bmatrix} \mathbf{y}_1 \\ -- \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ --- \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} \mathbf{e}_1 \\ -- \\ \mathbf{e}_2 \end{bmatrix}, \text{ where } Var \begin{bmatrix} \mathbf{e}_1 \\ -- \\ \mathbf{e}_2 \end{bmatrix} = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]$$

X_1 is of full column rank (columns independent)

Var is symmetric non - negative - definite; Σ_{11} is positive - definite

Solution of the Linear Statistical Model

$$\hat{\mathbf{y}}_2 = X_2\hat{\boldsymbol{\beta}} + \Sigma_{21}\Sigma_{11}^{-1}\left(\mathbf{y}_1 - X_1\hat{\boldsymbol{\beta}}\right)$$

$$Var\left[\mathbf{y}_2 - \hat{\mathbf{y}}_2\right] = \left(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1\right)Var\left[\hat{\boldsymbol{\beta}}\right]\left(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1\right)' + \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

where

$$\hat{\boldsymbol{\beta}} = \left(X_1'\Sigma_{11}^{-1}X_1\right)^{-1}X_1'\Sigma_{11}^{-1}\mathbf{y}_1$$

$$Var\left[\hat{\boldsymbol{\beta}}\right] = \left(X_1'\Sigma_{11}^{-1}X_1\right)^{-1}$$

"Unfortunately, this [slide] is too narrow to contain the proof." Proof in Appendix C, 1997 paper of next slide

Conjoint Prediction

- A composite of paid and incurred submodels whose unpaid and IBNR predictions are so designed that paid+unpaid = incurred+IBNR (by exposure period, e.g., AY)

- Done the hard way in “Conjoint Prediction of Paid and Incurred Losses,” *CAS Forum*, Summer 1997, vol 1

- Shortcut (with flaws) in “Modeling Paid and Incurred Losses Together,” *CAS E-Forum*, Spring 2009

- This presentation condensed from “Mixed Estimation and Conjoint Prediction: Three Equivalent Forms of Paid-Incurred Modeling,” recently submitted to *Variance* (So it might appear in 2013)

Abstractly, How to Design It

- If we didn't know better, random vector $\mathbf{x}$ would have (vector) mean μ and (matrix) variance Σ.
- However, we know that $Prob\left[\mathrm{A}_{m\times n}\mathbf{x}_{n\times 1} = \mathrm{b}_{m\times 1}\right] = 1$
- We must revise its mean and variance:

$$1)\quad \mu \to \mu^{*}, \text{ where } \mathrm{A}\mu^{*} = \mathrm{b}$$

$$2)\quad \Sigma \to \Sigma^{*} = \Sigma - \Sigma\mathrm{A}'\left(\mathrm{A}\Sigma\mathrm{A}'\right)^{-1}\mathrm{A}\Sigma$$

1) constrains the expected value (the model parameters)
2) tightens the variance (takes away m degrees of freedom)

Example: Sched P WorkComp 2007

- See Mixed Estimation.xls, Exhibit 6
  - 1) Expected-value constraint row 133 (x6a): $\begin{bmatrix} 1'_{11} & -1'_{11} \end{bmatrix} \begin{bmatrix} \beta_{\text{Paid}} \\ \beta_{\text{Incd}} \end{bmatrix} = 0$
  - 2) 133 × 133 tightened variance matrix (x6b)
  - Solution (x6c) requires advanced linear algebra
- Solution summarized in Exhibit 9:

AY	Paid	CaseIncd	Unpaid		Std Dev	IBNR		Std Dev	Ultimate
1998	17,738,999	19,765,070	2,997,087	±	138,034	971,016	±	138,034	20,736,086
1999	17,557,257	19,812,936	3,392,217	±	153,110	1,136,538	±	153,110	20,949,474
2000	18,432,885	20,739,446	3,805,189	±	171,119	1,498,629	±	171,119	22,238,074
2001	17,289,118	19,811,447	4,225,650	±	179,522	1,703,321	±	179,522	21,514,768
2002	15,899,281	18,589,679	4,658,025	±	184,857	1,967,628	±	184,857	20,557,306
2003	15,283,538	18,314,350	5,413,813	±	212,522	2,383,000	±	212,522	20,697,351
2004	14,134,508	17,632,172	6,561,625	±	261,583	3,063,960	±	261,583	20,696,133
2005	12,789,801	17,221,257	8,686,404	±	307,737	4,254,949	±	307,736	21,476,205
2006	10,724,002	16,822,179	13,146,034	±	387,042	7,047,859	±	387,043	23,870,036
2007	5,211,936	12,931,177	20,391,191	±	525,931	12,671,951	±	525,931	25,603,127
Total	145,061,325	181,639,713	73,277,235	±	1,146,012	36,698,849	±	1,146,011	218,338,560

An Easier Way to Paid-Incurred Agreement

- Start with semi-conjoint model, Exhibit 7
  - Has expected-value constraint, but original variance (x7a)
  - With reduced parameters, can be solved (x7b) in Excel
  - Complicated prediction-error variance due to e-v constraint; positive covariance throughout. Excerpt of x7b:

AY AGE	y2hat		StdPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr	VarPrdErr
1998 Unpd	2,931,158	±	223,565	5.00E+10	7.26E+08	8.53E+08	9.28E+08	9.65E+08	1.35E+09	1.84E+09	2.34E+09	3.54E+09	5.31E+09	2.22E+08	5.35E+08
1999 Unpd	3,275,075	±	227,979	7.26E+08	5.20E+10	1.69E+09	1.74E+09	1.74E+09	2.12E+09	2.59E+09	3.11E+09	4.34E+09	6.07E+09	2.50E+08	6.04E+08
2000 Unpd	3,840,647	±	240,431	8.53E+08	1.69E+09	5.78E+10	2.70E+09	2.67E+09	3.05E+09	3.54E+09	4.10E+09	5.43E+09	7.20E+09	2.94E+08	7.10E+08
2001 Unpd	4,207,585	±	245,201	9.28E+08	1.74E+09	2.70E+09	6.01E+10	3.56E+09	3.93E+09	4.38E+09	4.93E+09	6.26E+09	7.90E+09	3.20E+08	7.72E+08
2002 Unpd	4,599,117	±	247,447	9.65E+08	1.74E+09	2.67E+09	3.56E+09	6.12E+10	4.47E+09	4.89E+09	5.43E+09	6.74E+09	8.26E+09	3.33E+08	8.02E+08
2003 Unpd	5,303,808	±	288,687	1.35E+09	2.12E+09	3.05E+09	3.93E+09	4.47E+09	8.33E+10	8.46E+09	9.04E+09	1.05E+10	1.17E+10	4.64E+08	1.12E+09
2004 Unpd	[illegible]	±	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8.46E+09	1.16E+11	1.37E+10	1.54E+10	1.62E+10	6.33E+0	[illegible]
2005 Unpd	8,617,002	±	392,367	2.34E+09	3.11E+09	4.10E+09	4.93E+09	5.43E+09	9.04E+09	1.37E+10	1.54E+11	2.03E+10	2.09E+10	8.08E+0	[illegible]
2006 Unpd	13,210,032	±	503,124	3.54E+09	4.34E+09	5.43E+09	6.26E+09	6.74E+09	1.05E+10	1.54E+10	2.03E+10	2.53E+11	3.17E+10	1.22E+0	[illegible]
2007 Unpd	20,417,432	±	647,646	5.31E+09	6.07E+09	7.20E+09	7.90E+09	8.26E+09	1.17E+10	1.62E+10	2.09E+10	3.17E+10	4.19E+11	1.83E+0	[illegible]
1998 IBNR	1,010,744	±	174,704	2.22E+08	2.50E+08	2.94E+08	3.20E+08	3.33E+08	4.64E+08	6.33E+08	8.08E+08	1.22E+09	1.83E+09	3.05E+10	[illegible]
1999 IBNR	1,239,148	±	205,473	5.35E+08	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.42E+09	1.84E+08	4.22E+10
2000 IBNR	1,491,745	±	241,656	8.08E+08	9.12E+08	1.07E+09	1.17E+09	1.21E+09	1.69E+09	2.31E+09	2.94E+09	4.44E+09	6.68E+09	2.79E+08	6.03E+09
2001 IBNR	1,747,321	±	260,433	9.40E+08	1.06E+09	1.25E+09	1.36E+09	1.41E+09	1.97E+09	2.69E+09	3.43E+09	5.17E+09	7.77E+09	3.24E+08	5.80E+09
2002 IBNR	2,058,023	±	273,783	1.02E+09	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8.40E+09	3.51E+08	5.54E+09
2003 IBNR	2,522,465	±	307,235	1.20E+09	1.35E+09	1.59E+09	1.73E+09	1.80E+09	2.51E+09	3.42E+09	4.37E+09	6.59E+09	9.90E+09	4.13E+08	5.55E+09
2004 IBNR	3,206,286	±	393,347	1.68E+09	1.90E+09	2.23E+09	2.43E+09	2.52E+09	3.52E+09	4.80E+09	6.13E+09	9.25E+09	1.39E+10	5.80E+08	5.40E+09
2005 IBNR	4,365,683	±	472,007	2.18E+09	2.46E+09	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.20E+10	1.80E+10	7.50E+08	5.44E+09
2006 IBNR	6,979,319	±	570,119	2.87E+09	3.24E+09	3.81E+09	4.15E+09	4.31E+09	6.01E+09	8.21E+09	[illegible]	1.58E+10	2.38E+10	9.91E+08	5.83E+09
2007 IBNR	12,615,607	±	818,443	4.84E+09	5.47E+09	6.43E+09	6.99E+09	7.27E+09	1.01E+10	1.38E+10	1.77E+10	2.66E+10	4.00E+10	1.67E+09	5.18E+09

To the tune of "Kung Fu Fighting:"

Actuaries were spreadsheet fighting

They were keying as fast as lightning

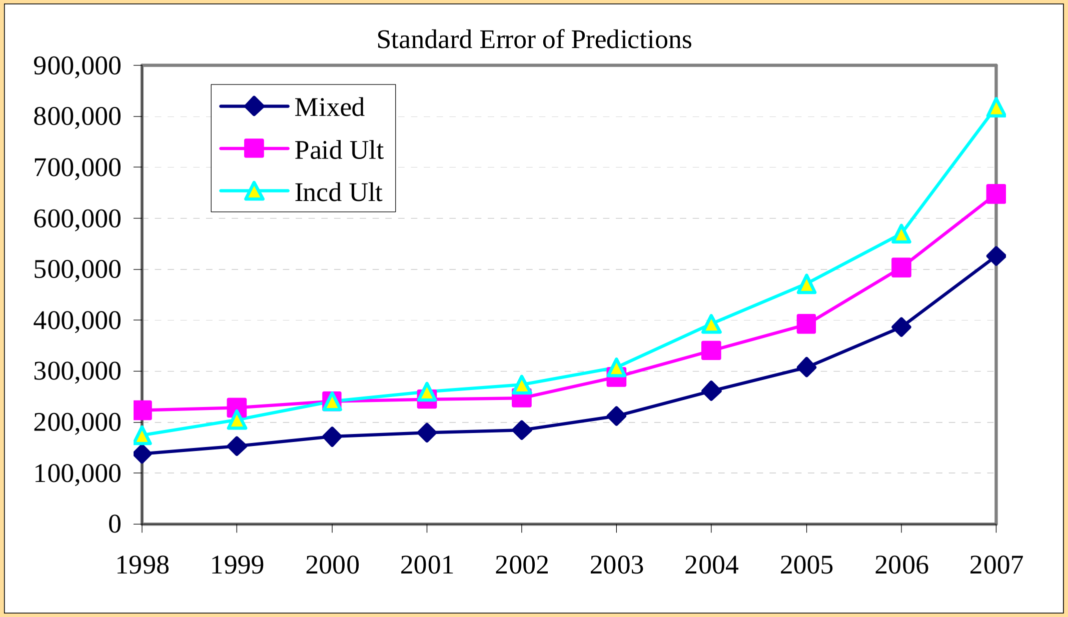
It was a little bit frightening

Those cats were data mining

Mix the Paid-Incurred Estimates

- Gauss-Markov finds optimal 10×10 weights, Exhibit 8
- With W_0, form the mixed estimator and its variance, x8b
- Results agree with conjoint prediction (Exhibit 9):

AY	μ		Std
1998	20,736,086	±	138,034
1999	20,949,474	±	153,110
2000	22,238,075	±	171,119
2001	21,514,768	±	179,522
2002	20,557,306	±	184,856
2003	20,697,350	±	212,522
2004	20,696,133	±	261,583
2005	21,476,205	±	307,736
2006	23,870,037	±	387,043
2007	25,603,128	±	525,931
Total	218,338,561	±	1,146,012
Paid Ult	217,928,209	±	1,372,673
Incd Ult	218,876,055	±	1,630,573



Summary of Linear Models

AY	Paid	CaseIncd	Paid Ultimate	Paid Std Dev	Incurred Ultimate	Incurred Std Dev
1998	17,738,999	19,765,070	20,663,109	± 224,359	20,780,207	± 175,099
1999	17,557,257	19,812,936	20,824,378	± 228,971	21,062,679	± 207,420
2000	18,432,885	20,739,446	22,264,182	± 241,730	22,247,195	± 245,423
2001	17,289,118	19,811,447	21,486,528	± 246,708	21,577,395	± 265,162
2002	15,899,281	18,589,679	20,487,825	± 249,060	20,667,840	± 279,038
2003	15,283,538	18,314,350	20,572,601	± 291,372	20,860,557	± 313,738
2004	14,134,508	17,632,172	20,579,410	± 344,815	20,871,758	± 403,318
2005	12,789,801	17,221,257	21,381,119	± 398,343	21,630,036	± 485,897
2006	10,724,002	16,822,179	23,895,269	± 513,711	23,858,441	± 590,139
2007	5,211,936	12,931,177	25,571,140	± 666,133	25,642,727	± 857,783
Total	145,061,325	181,639,713	217,725,562	± 1,475,942	219,198,836	± 1,845,299

AY	Semi-Conjoint Paid Ult	Semi-Conjoint Std Dev	Semi-Conjoint Incd Ult	Semi-Conjoint Std Dev	Conjoint / Mixed Ultimate		Conjoint / Mixed Std Dev
1998	20,670,157	± 223,565	20,775,814	± 174,704	20,736,086	±	138,034
1999	20,832,332	± 227,979	21,052,084	± 205,473	20,949,474	±	153,110
2000	22,273,532	± 240,431	22,231,191	± 241,656	22,238,075	±	171,119
2001	21,496,703	± 245,201	21,558,768	± 260,433	21,514,768	±	179,522
2002	20,498,398	± 247,447	20,647,702	± 273,783	20,557,306	±	184,856
2003	20,587,346	± 288,687	20,836,815	± 307,235	20,697,350	±	212,522
2004	20,599,535	± 340,582	20,838,458	± 393,347	20,696,133	±	261,583
2005	21,406,803	± 392,367	21,586,940	± 472,007	21,476,205	±	307,736
2006	23,934,034	± 503,124	23,801,498	± 570,119	23,870,037	±	387,043
2007	25,629,368	± 647,646	25,546,784	± 818,443	25,603,128	±	525,931
Total	217,928,209	± 1,372,673	218,876,055	± 1,630,573	218,338,561	±	1,146,012