Statistics Canada Statistique Canada

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-entreprises

Canada

89-02448/

[redacted]

WORKING PAPER NO. BSMD-86-008E

METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. BSMD-86-008E

MÉTHODOLOGIE

ESTIMATION AND VARIANCE ESTIMATION FOR THE BILATERAL SYB-SYSTEM TO PRODUCE AIR SCHEDULED INTERNATIONAL PASSENGERS ORIGIN AND DESTINATION STATISTICS

by

R. Carpenter, M.A. Hidiroglou and M. Latouche

March 1986

* This is a preliminary version.
Do not quote without author's permission. Comments are welcome.

Table of Contents

		Page
I -	Introduction	1
II -	Objective	2
III -	Methodology	3
IV -	Variables Affecting the Coefficients of Variation	4
V -	Estimation	10
VI -	Reliability and Usefulness of the Estimates	14
VII -	Conclusion and Recommendations	15
VIII -	Bibliography	16

Appendix A - List of Carriers and Cities

Appendix B - Exact Calculation of the Coefficients of Variation

Appendix C - Description of the Data

Appendix D - Statistical Analyses

Appendix E - Parameters

Appendix F - Charts and Tables

ESTIMATION AND VARIANCE ESTIMATION FOR THE BILATERAL SUB-SYSTEM TO PRODUCE AIR SCHEDULED INTERNATIONAL PASSENGER ORIGIN AND DESTINATION STATISTICS

R. Carpenter, M.A. Hidiroglou and M. Latouche

ABSTRACT

The purpose of the Air Scheduled International Passenger Origin and Destination (ASIPOD) system is to provide information on the flow of passengers on international flights.

This information is collected by sampling tickets on a 10% basis and prorating the information to known totals for foreign airlines landing and taking off from Canadian airports. The estimation may be shortly described as post-stratified ratio domain estimation. Given this design and estimation strategy, variance estimates were obtained.

ESTIMATION ET ESTIMATION DE LA VARIANCE POUR LE SOUS-SYSTÈME BILATÉRAL QUI PRODUIT DES STATISTIQUES D'ORIGINE ET DESTINATION POUR LES PASSAGERS SUR LES VOLS INTERNATIONAUX

R. Carpenter, M.A. Hidiroglou et M. Latouche

SOMMAIRE

Le but de l'enquête de l'origine et destination des passagers sur les vols internationaux est de produire des estimations pour le nombre de passagers volant sur des vols internationaux.

Cette information est recueilli en échantillonnant des billets d'avion en utilisant un taux de sondage de 10% et ensuite en ajustant les comptes pondérés à des totaux connus pour les lignes aériennes etrangères aterissant et décolant d'aéroports canadiens. La méthode peut-être décrite comme une méthode d'estimation par domaine par rapport post-stratifié. Etant donné ce plan de sondage et cette méthode d'estimation, les estimateurs de variances sont dérivés.

0 - Summary

To get some idea of the ASIPOD[1] Survey's precision, the Aviation Statistics Centre asked that a simple, rapid method of approximating the coefficients of variation be developed. Such a method, based on passenger volume, carrier, proration level and Canadian gateway, was found. Its efficiency is good, but depends to a large extent on the stability of the population.

1 Air Scheduled International Passenger Origin and Destination.

I - Introduction

A survey should deliver both precise estimates and a measure of that precision. One means of quantifying an estimate's precision is the coefficient of variation.

The coefficient of variation is defined as the ratio of the estimate's standard deviation to the estimate itself. In the Air Scheduled International Passenger Origin and Destination Survey, calculation of the standard deviation, and hence of the coefficient of variation, is a long, complicated process.[1] This is very inconvenient for the users of the estimates (Transport Canada, CTC, External Trade and carriers) who need quick estimates and the coefficient of variation. These users would therefore like to have a rapid method method of computing the coefficient of variation.

---

1 Calculation of the standard deviation is described in Appendix B.

II - Objective

The aim of this study was to find an efficient method of approximating the coefficients of variation without using actual sample data excepting market size. This method was then to be used to construct, for publication, no more than six tables and charts containing the coefficients of variation cross-classified with a limited number of variables. Consequently, the effects that these variables might have on the coefficients of variation were assessed in this study. The following variables were examined: passenger volume, type of flight, airports of origin and destination, carrier and proration percentage.

III - Methodology

The study was based on scheduled international flight data for 1982. The data were separated into two categories: those for direct flights and those for flights via the United States. Direct flights were broken down by Canadian city, foreign city, gateway and carrier. The percentage of assigned passengers and the coefficient of variation were computed for each combination (domain) of these variables. Flights via the United States were broken down in the same manner, with two additional variables, U.S. gateway and gateway carrier. In this way, it was possible to determine the effects of all variables on the coefficients of variation. The formula used to calculate the coefficients of variation is given in Appendix B.

To improve the estimates, the domains were divided into three groups: the domains of carriers that by definition could not have assigned passengers (Group I); all other domains with no assigned passengers (Group II); and domains with assigned passengers (Group III).

IV - Variables Affecting the Coefficients of Variation

The coefficients of variation (CV) in Group I were generally very large. On average, they were 0.71, with a standard deviation of 0.27; 91% of them were greater than 0.28, and 50% were greater than 0.92, as shown in Table I.

TABLE I
Distribution of CVs
Group I

CV (%)		FREQ.	CUM. FREQ.	PERC.	CUM. PERC.
0-4		48	48	0.41	0.41
4-12	*	246	294	2.13	2.54
12-20	**	412	706	3.56	6.10
20-28	***	545	1251	4.71	10.81
28-36	***	664	1915	5.74	16.55
36-44	***	554	2469	4.79	21.34
44-52	***	540	3009	4.67	26.01
52-60	****	735	3744	6.35	32.37
60-68	**********	2025	5769	17.51	49.87
68-76		0	5769	0.00	49.87
76-84		0	5769	0.00	49.87
84-92		0	5769	0.00	49.87
92-100	*****************************	5799	11568	50.13	100.00
	1000 2000 3000 4000 5000				

Legend: FREQ. = frequency
CUM. FREQ. = cumulative frequency
PERC. = percentage
CUM. PERC. = cumulative percentage

In general, a CV greater than 0.30 indicates that the estimate (number of passengers) is of poor quality. Since most of the estimates had unacceptably large CVs, it appeared, initially, that this project was futile. It was noticed, however, that the large CVs were associated exclusively with domains involving less than 100 passengers, as shown in Table II.

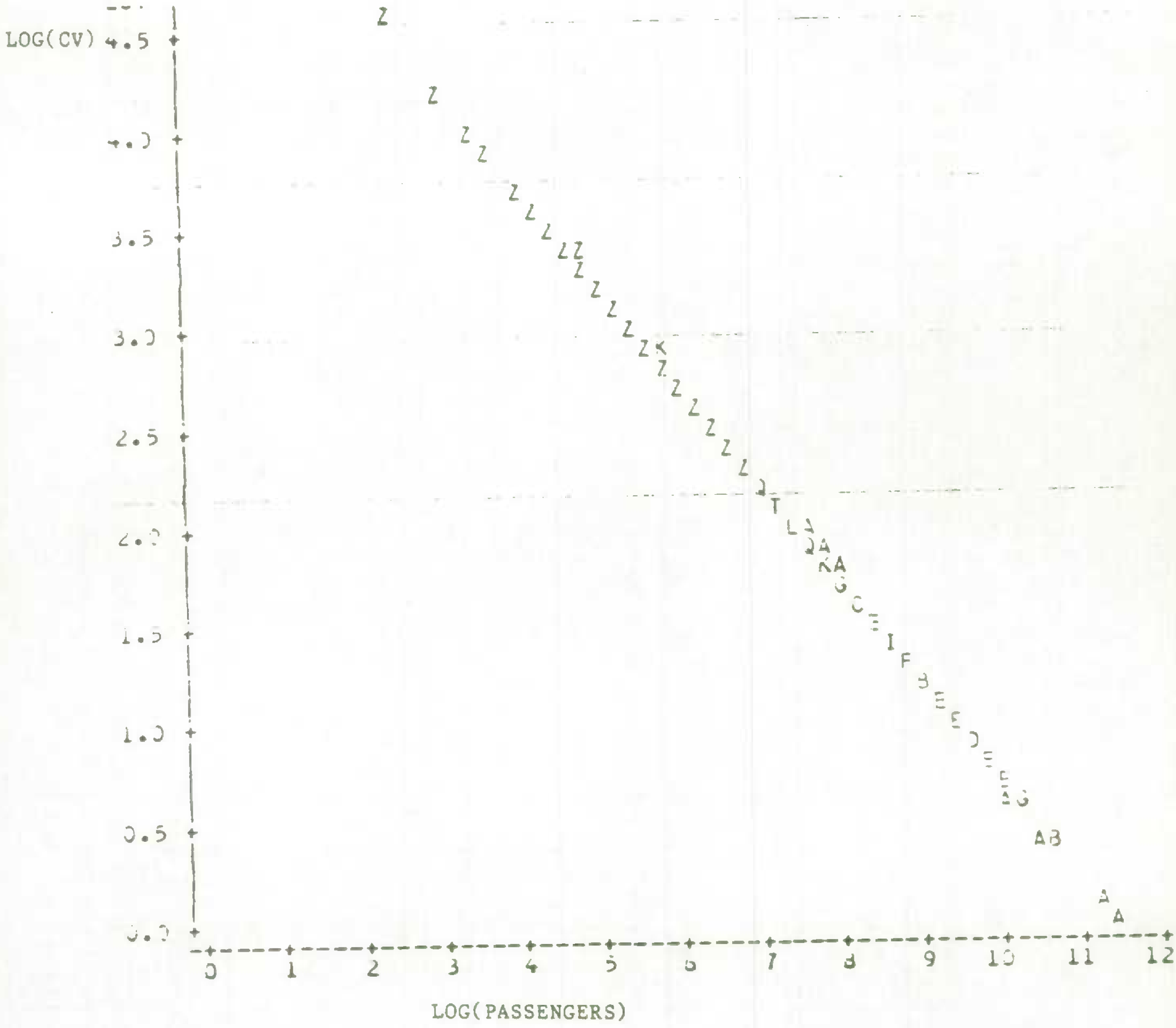
TABLE II
Distribution of Large CVs
by Passenger Volume
Group I

CV	PASSENGERS	0-100	100-300	300-500	500-700	700-PLUS	TOTAL
Good (CV<30%)	FREQ. PERC.	97 0.84	791 6.84	176 1.52	101 0.87	269 2.33	1434 12.40
Poor (CV≥30%)	FREQ. PERC.	10134 87.60	0 0.00	0 0.00	0 0.00	0 0.00	10134 87.60
TOTAL	FREQ. PERC.	10231 88.44	791 6.84	176 1.52	101 0.87	269 2.33	11568 100.00

Legend: FREQ. = frequency (number of domains)
PERC. = percentage

Hence it was worth-while pursuing the study using only those domains with more than 100 passengers. After this screening process, the mean CV was 0.19 with a standard deviation of 0.08, a marked improvement. Chart I provides a clearer picture of the correlation between CV and passenger volume. It reveals a very strong linear correlation between the logarithm of the CV and the logarithm of the number of passengers.

CHART I
Logarithm of the CV
as a Function of the Logarithm
of Passenger Volume
Group I



The CVs of Groups II and III, like those of Group I, were high. The mean was 0.58 with a standard deviation of 0.23.

TABLE III
Distribution of CVs
Groups II and III

CV (%)		FREQ.	CUM. FREQ.	PERC.	CUM. PERC.
0-12	**	117	117	3.42	3.42
12-20	***	151	268	4.41	7.83
20-28	****	200	468	5.84	13.68
28-36	*****	259	727	7.57	21.24
36-44	*****	266	993	7.77	29.02
44-52	******	319	1312	9.32	38.34
52-60	*******	366	1678	10.70	49.04
60-68	***	125	1803	3.65	52.69
68-76	**********	485	2288	14.17	66.86
76-84	******************	895	3183	26.15	93.02
84-92	***	184	3367	5.38	98.39
92-100	*	55	3422	1.61	100.00
	200 400 600 800				

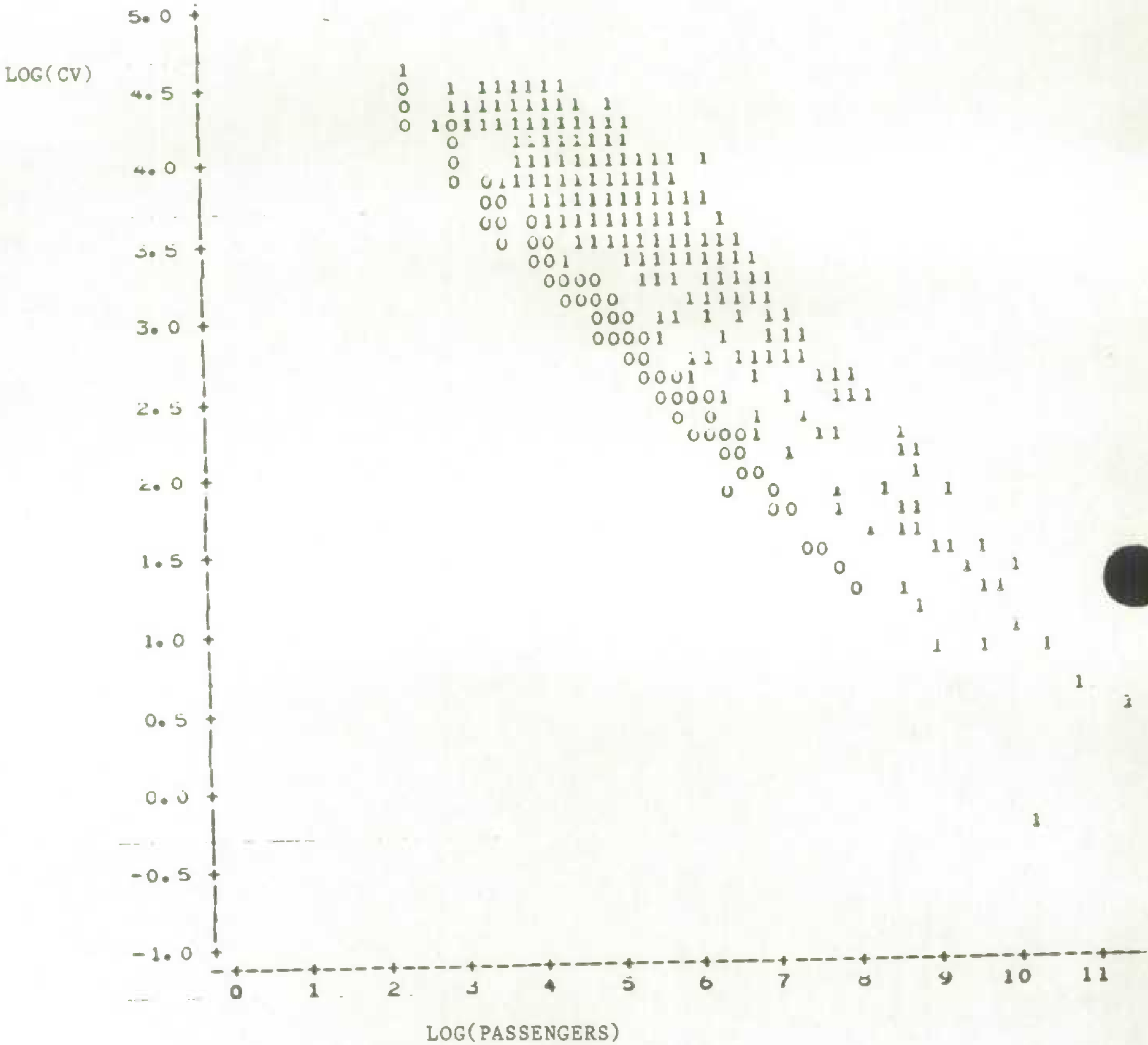
Like the Group I domains, the non-prorated domains with more than 100 passengers had CVs of less than 30%; for prorated domains, on the other hand, it is necessary to go as high as 600 passengers before the CVs become generally acceptable.

TABLE IV
Distribution of Large CVs
by Passenger Volume
Group III
(Prorated Domains)

			PASSENGERS						
			0-100	100-200	200-400	400-600	600-800	800-PLUS	TOTAL
CV	Poor (CV≥30%)	FREQ.	614	329	229	84	17	11	1284
		PERC.	39.18	21.00	14.61	5.36	1.08	0.70	81.94
	Good (CV<30%)	FREQ.	3	4	22	41	31	182	283
		PERC.	0.19	0.26	1.40	2.62	1.98	11.61	18.06
	TOTAL	FREQ.	617	333	251	125	48	193	1567
		POURC.	39.37	21.25	16.02	7.98	3.06	12.32	100.00

Chart II shows the correlation between the logarithm of the CV and the logarithm of the number of passengers for Group III only. The 0's denote non-prorated routes, and the 1's prorated routes. Proration seems to have increased the CVs. Prorated routes are also farther apart than non-prorated routes. For these reasons, the CVs for non-prorated domains (Group II) and prorated domains (Group III) were approximated separately.

CHART II
Logarithm of the CV
as a Function of the Logarithm
of Passenger Volume
Groups II and III



V - Estimation

The main variable used to estimate the CVs was passenger volume. Charts I and II suggested that a log-log model might be appropriate. This model is defined as follows:

$$\log(CV_i) = A + B \times \log(PA_i) + E_i$$

where log = logarithmic function (base e)

CV_i = coefficient of variation for the i^{th} domain

PA_i = passenger volume in the i^{th} domain

E_i = error term for the i^{th} domain. E_i is assumed to be an independent random variable having a normal distribution with mean 0. It is not used for estimation purposes, but only to test certain statistical assumptions.

A and B are the parameters to be estimated; B is the slope of the line and A the X-intercept.

Model (1) was fitted to each of the three groups. Whenever it proved inefficient, it was modified to take into account the effect of one or more qualitative variables (carrier, gateway, origin, destination and so on). For example, the addition of one qualitative variable yielded the following model:

$$\log(CV_{ij}) = A + C_j + B \times \log(PA_{ij}) + E_{ij}$$

where CV_{ij} = coefficient of variation for the i^{th} domain at level j of the qualitative variable in question

C_j = effect due to level j of the qualitative variable. The sum of the effects of all the levels of the qualitative variable is zero.

The practical difference between the two models is that model (1) defines only one line while model (2) defines a set of parallel lines, each of which represents a particular level of the qualitative variable. If, for example, the qualitative variable considered is the carrier, there will be a line (that is, a model) for each carrier.

The qualitative variables used were selected by covariance analysis. The effects of variables such as carrier, proration percentage, Canadian and foreign cities and so on were assessed with this technique. The results of these analyses are presented in Appendix D.

A number of statistical tools are available to measure a model's efficiency. The statistic used here was R^2. The latter ranges between 0 and 1; the larger it is, the more efficient the model is - in other words, the smaller the differences between the observed values and the estimates. When logarithmic scales are used, an R^2 in excess of 0.95 generally indicates that the model is a good one.

1. Estimation for never-prorated carriers

   For this group, model (1) produced a perfect fit. The R^2 for this model was equal to 1.0, which implies that the error terms were zero. After parameters A and B were estimated, the model was:

   $$\log(CV_i) = \log(3) - 0.5 \log(PA_i)$$

   or, on the natural scale,

   $$CV_i = \frac{3}{\sqrt{PA_i}}$$

   PA_i must be greater than or equal to 100.

   Appendix F contains a table and chart showing the CVs as a function of passenger volume for carriers in this group.

2. Estimation for non-prorated domains (carriers other than those in Group I).

   Although the fit was not as good as for the preceding group, model (1) was selected to approximate the CVs of this group as well. Its R^2 was 0.97 on the logarithmic scale and 0.95 on the natural scale. The model for this group was approximately:

   $$\log(CV_i) = 1.04 - 0.54 \log(PA_i)$$

   or

   $$CV_i = \frac{2.82}{PA_i^{0.54}}$$

   PA_i must be greater than or equal to 100. The exact values of the parameters are given in Appendix E.

3. Estimation for prorated domains

   When model (1) was fitted to this group, the R^2 was only 0.80. Model (2) provided a better fit, as the R^2 improved to 0.85 with gateway as the extra variable and 0.92 when the carrier was used. When the gateway and carrier effects were combined, it reached an acceptable 0.96 (0.86 on the natural scale). Thus the following model was selected:

   $$\log(CV_{ijk}) = A + T_j + G_k + B \times \log(PA_{ijk}) + E_{ijk} \qquad (3)$$

where CV_{ijk} is the coefficient of variation for the i^{th} domain of carrier j and gateway k;

T_j is the effect due to carrier j;

G_k is the effect due to gateway k.

The difficulty with this model was that it required the generation of many tables. Because the users had requested a small number of tables, it was necessary to simplify the model. This was accomplished by combining adjacent lines into one line. The line selected to represent the others was the one with the largest CVs.

Four lines were used to represent the Group III model. The equations were approximately as follows:

$$\log(CV_{ij}) = 4.18 - 0.58 \times \log(PA_{ij}) \qquad j = 1$$

$$\log(CV_{ij}) = 2.75 - 0.58 \times \log(PA_{ij}) \qquad j = 2$$

$$\log(CV_{ij}) = 2.43 - 0.58 \times \log(PA_{ij}) \qquad j = 3$$

$$\log(CV_{ij}) = 2.03 - 0.58 \times \log(PA_{ij}) \qquad j = 4$$

or

$$CV_{ij} = \frac{66}{PA_{ij}^{0.58}} \qquad j = 1$$

$$CV_{ij} = \frac{16}{PA_{ij}^{0.58}} \qquad j = 2$$

$$CV_{ij} = \frac{11}{PA_{ij}^{0.58}} \qquad j = 3$$

$$CV_{ij} = \frac{8}{PA_{ij}^{0.58}} \qquad j = 4$$

Note that PA_{ij} must be greater than or equal to 600. These lines were associated with the following carriers:

j = 1: CU gateway YMX
LO gateway YMX
AZ gateway YMX
KL gateway YMX

j = 2: JM gateway YYZ
SN gateway YMX
AF gateway YMX
LH gateway YMX
AT gateway YMX
BW gateway YYZ
BA gateway YVR
BA gateway YYZ
BA gateway YMX

j = 3: KL gateway YYZ
SR gateway YMX
LH gateway YYZ
AF gateway YYZ
AZ gateway YYZ
SU gateway YMX
BA gateway YEG
AR gateway YMX
IB gateway YMX
QF gateway YVR

j = 4: JL gateway YVR
BA gateway YYC
SR gateway YYZ
OK gateway YMX
AY gateway YMX
TP gateway YMX
AI gateway YMX
LY gateway YMX

All other carrier-gateway combinations are associated with the first line (j = 1). This also applies to all new carriers.

VI - Reliability and Usefulness of the Estimates

Though not of optimum quality, the estimates can be considered reliable. Simplification of the Group III model certainly reduced its efficiency, but deliberately maximizing the CV estimators ensured that the actual CVs would not be larger than the estimated CVs. This in turn guaranteed that the quality of the Air Scheduled International Passenger Origin and Destination Survey would not be lower than predicted.

The CV charts and tables in Appendix F can be used to estimate the CVs for subsequent years on condition that the target population remains unchanged and that time has no effect on the correlations among the variables.

It is crucial that the population described in Section IV remain stable from year to year. Events such as the establishment of new carriers or Canadian gateways or the disappearance of existing ones may significantly alter the CVs. A decline in the proration percentage in each domain could reduce the CVs. Furthermore, even though the remaining variables (origin and destination, for example) had no effect on the CVs, it is by no means certain that the results will be identical if additional domains are created.

An important point that could not be analysed in this study is the correlation between CVs and time. If the CVs vary widely from year to year, the tables and charts in Appendix F will have to be recalculated periodically. While it is unlikely that time has a major impact on the CVs, it would be preferable to test this assumption.

VII - Conclusions and Recommendations

It was shown that the coefficients of variation were easy to estimate provided the passenger volume, carrier, proration level and Canadian gateway were known. These variables were used to produce two charts and two tables containing estimated coefficients of variation. The charts and tables for prorated domains were constructed so that the actual coefficients of variation would be less than or equal to the estimates.

The charts and tables can be used to estimate the coefficients of variation for subsequent years on condition that the population remains the same from year to year and that time does not affect the coefficients.

To ensure that the estimates are of the best possible quality, it is recommended that:

1. the charts and tables be recomputed if:
   - five carriers serving at least ten routes during a year are added to or deleted from the survey;

   - at least 30 routes are found to have assigned-passenger proportions of between 0.30 and 0.70;

2. a study be carried out to assess the effect that time has on the correlations between the coefficients of variation and the explanatory variables. This study could be based on a comparison of the 1982 and 1985 coefficients.

VIII - Bibliography

Carpenter, Ron, et Hidiroglou, Michel A., Estimation and Variance Estimation for the Bilateral Sub-system to Produce Air Scheduled International Passenger Origin and Destinaton Statistics. Rapport de la DMEE, Décembre 1985. (Appendice B du présent rapport).

Cochran, William G., Sampling Techniques, John Wiley & Sons Inc., 1977

Gunst, Richard F., et Mason, Robert L., Regression Analysis and its Application, Marcel Dekker Inc., 1980

Montgomery, Douglas C., Design and Analysis of Experiments, John Wiley & Sons Inc., 1986

Neter, John, et Wasserman, William, Applied Linear Statistical Models, Richard D. Irwin Inc., 1974.

APPENDIX A

LIST OF CARRIERS

AND CITIES

GROUP I
Non-prorated Carriers

Carriers:	AC										
	CP										
	PW										
	US: all U.S. carriers.										
	UK: unidentified carriers										
Canadian cities:	YEG			YQR							
	YHZ			YVR							
	YMX			YWG							
	YOW			YXE							
	YQB			YYC							
	YQG			YYZ							
Canadian gateways:	YEG			YUL							
	YHZ			YVR							
	YMX			YWG							
	YOW			YXE							
	YQR			YYC							
	YQX			YYT							
				YYZ							
Foreign cities:	ABJ	ABZ	ACA	ADL	AGP	AKL	ALG	AMM	AMS	ANU	ARN
	ASU	ATH	AUA	AUH	BAG	BAH	BAQ	BCN	BDA	BEG	BER
	BEY	BFS	BJI	BGO	BFW	BHX	BKI	BKK	BKO	BNE	BOD
	BOG	BOM	BON	BRE	BRI	BRU	BSL	BUD	BUE	BUH	BWN
	BZE	CAI	CAS	CCS	CCU	CDG	CGH	CGN	CHC	CKY	CLO
	CMB	CPH	CPT	CTA	CTG	CUN	CUR	CZN	DAC	DAM	DEL
	DHA	DKR	DLA	DOM	DUB	DUR	DUS	DXB	EDI	EIS	EMA
	FAO	FCO	FDF	FNC	FPO	FRA	FUK	GCM	GDL	GEO	GIG
	GLA	GND	GOT	GUA	GVA	GYE	HAJ	HAM	HAV	HEL	HKG
	HKP	HLP	HLL	IST	ISY	ITO	JED	JNB	KAL	KHI	KIN
	KOA	KUL	KWI	LAP	LBA	LBV	LCA	LGW	LHR	LIH	LIM
	LIN	LIS	LGU	LON	LOS	LRT	LUN	LUX	LYS	MAA	MAD
	MAN	MAR	MBJ	MDE	MDZ	MEB	MEX	MHH	MID	MIL	MLA
	MME	MNI	MNL	MOW	MPL	MRS	MTY	MUC	MVD	MZD	NAN
	NAP	NAS	NBO	NCE	NCL	NEV	NIM	NRT	NUE	NWI	OGG
	OPO	ORY	OSA	OSL	OUA	PAP	PAR	PBM	PEK	PEN	PER
	PIK	PMO	POS	PPT	PRG	PTP	PTY	PUS	PVR	RAR	REG
	RIO	ROM	RUH	SAL	SAD	SAY	SCL	SDQ	SEL	SIN	SGD
	SJO	SJU	SKB	SKG	SLU	SMN	SPK	SRZ	STO	STR	STT
	STX	SUF	SUV	SVD	SVG	SXB	SXM	SYD	SZG	TAB	TCB
	THR	TIP	TLS	TLV	TPE	TRS	TUN	TXL	TYO	UIO	UVF
	VCE	VIE	VIG	WAW	YAO	ZAG	ZIH	ZLO	ZRH		

GROUP II
Other Non-prorated Domains

Carriers:

AF	BW	LY
AI	IB	OK
AR	JL	QF
AT	JM	SN
AY	KL	SR
AZ	LH	SU
BA	LO	TP

Canadian cities:

YEG	YQB	YVR	YYC
YHZ	YQG	YWG	YYZ
YOW	YQR	YXE	

Canadian gateways:

YMX
YVR
YYZ

Foreign cities:

ACA	AGP	ALC	AMS	ATH	BCN	BEY	BGI	BIO	BKK	BOM
BRU	BTS	BUE	CAI	CAS	CDG	DEL	FAO	FCO	FNC	FRA
GEO	HEL	HKG	HNL	IEV	KIN	LED	LHR	LIS	LON	MAD
MAN	MBJ	MEB	MEX	MIL	MNL	MOW	MUC	MVD	NRT	OSA
PDL	PMI	POS	PRG	ROM	SEL	SIN	SMA	STO	SUF	SVO
SYD	TER	TLV	TPE	TOY	WAW	ZRH				

GROUP III
Prorated Domains

Carriers:

AF	BA	KL	SN
AI	BW	LH	SR
AR	CU	LO	SU
AT	IB	LY	TP
AY	JL	OK	
AZ	JM	QF	

Canadian cities:

YEG	
YMX	YYC
YVR	YYZ

Canadian gateways:

YEG	YYC
YMX	YYZ
YVR	

Foreign cities:

ABJ	ABZ	ACA	ACC	AGA	AGP	ALC	ALG	AMS	ANU	ATH
BCN	BER	BEY	BGI	BIO	BKK	BOM	BRI	BRU	BSL	CAI
CAS	CCS	CDG	CGN	CPH	CPT	CTA	DAR	DEL	DHA	DUB
DUR	DUS	EDI	FCO	FNC	FRA	FUK	GEO	GLA	GOT	GVA
HAJ	HAM	HAV	HEL	HKG	HLP	HNL	IEV	IST	JNB	KAN
KHI	LBV	LED	LHR	LIM	LIN	LIS	LON	LOS	LYS	MAA
MAD	MAN	MBJ	MEB	MEX	MIL	MNI	MOW	MRS	MUC	MVD
MXP	NAP	NBO	NCE	NRT	OGG	ORY	OSA	OSL	PAR	PDL
PEK	PIK	PMI	PMO	POS	PRG	PUS	REG	ROM	SAY	SEL
SIN	SKB	SKG	SLU	STO	STR	SUF	SVG	SVO	SYD	TAB
TER	TIP	TLS	TLV	TNG	TPE	TRS	TYO	VCE	VIE	WAW
ZRH										

APPENDIX B

CALCULATION OF

COEFFICIENTS OF VARIATION

Estimation and Variance Estimation for the Bilateral Sub-System to Produce Air Scheduled International Passenger Origin and Destination Statistics

by

R. Carpenter and M.A. Hidiroglou

Business Survey Methods Division

1. Introduction

The Bilateral Sub-System and the proposed Air Scheduled International Passenger Origin and Destination Survey (ASIPOD) produce estimates of international scheduled commercial air service passengers by origin and destination for bilateral air negotiations. An international scheduled commercial air service is defined to be an operation which is between points in Canada and points in any other country, and which provides public transportation of persons, goods or mail by aircraft in accordance with a schedule and at a toll or charge per unit of traffic. Such a service is referred to as a "unit toll" service.

The estimation of flow of passengers on routes for scheduled air passengers is an item of importance to officials of the Canadian government from External Affairs, the Canadian Transport Commission, Transport Canada and the Ministry of Industry, Trade and Commerce. These officials use passenger flow estimates in order to ensure that Canada can negotiate a fair market share of commercial air routes with foreign countries.

The Bilateral Sub-System estimation process uses the data from two air traffic surveys to produce estimates of the number of passengers on scheduled international flights between Canadian and foreign markets for various origin-destination combinations. The first of these two surveys, the Revenue Passenger Origin and Destination (O/D) survey, provides a sample (on a 10% basis) of origin-destination data on international journeys with major Canadian or American carriers on at least one leg of the itinerary. These data are collected on statement 35 for the old O/D system and statement 3 (I) on the new O/D system. Additionally, there is O/D data obtained from the United States as a result of an exchange of data agreement between Canada and the United States. The second survey, the Airport Activity Survey (AAS), counts all passengers entering or leaving Canada on major Canadian or foreign scheduled carriers without consideration of the passenger's origin and destination. These data are collected on statement 32 for the old AMS system and statement 6 (I,F) on the new AMS system.

This paper will describe the sampling method used for the Bilateral Sub-System and the proposed ASIPOD system which is in current use, as well as its associated estimation system for the number of passengers travelling between Canadian and foreign airports. The associated variance formulae will then be provided.

2. Description of the Sampling Method and the Estimation System

The Revenue Passenger Origin and Destination data are submitted monthly by major Canadian unit toll air carriers conducting scheduled passenger services. The seven major Canadian carriers contributing information to this survey are Air Canada (AC), Canadian Pacific (CP), Eastern Provincial Airways (PV), Nordair (ND), Pacific Western Airlines (PW), Air Ontario (GX) and Quebecair (QB). The target population of the international estimation system is the set of all tickets with an international (i.e. between a foreign country / United States and Canada) journey. An exchange program on passenger O/D data is maintained between Canada and the United States, whereby the United States gives Canada those records detailing the complete itineraries of the tickets collected in their survey on which:

a) a U.S. and Canadian point is shown in the routing, or

b) a U.S. carrier is recorded as having flown to or from a Canadian point, or

c) a Canadian carrier is recorded as having flown to or from a U.S. point.

As a result of this exchange agreement, the expression "foreign" and "foreign (non-U.S.)" are both used to denote "neither Canadian nor American".

For each participating Canadian and American carriers, a flight coupon is selected from a ticket with a serial number ending in '0', if that carrier is the first participating carrier to fly on a flight segment of that ticket. Hence, the survey reports a 10% sample (assuming that ticket numbers are uniformly distributed between 0 and 9) of unique flight coupons on which there is at least one participating Canadian or American carrier. Information obtained using tickets which have a Canadian gateway as the first/last gateway to foreign countries (excluding the U.S.) will be labelled as "direct". Information obtained using tickets which have an American gateway as the first/last gateway to foreign countries will be labelled "via U.S.". The data for the three months are combined, and duplicates are eliminated. Hence, a file of complete itineraries, Ticket Origin and Destination (TOD) records, is obtained for the quarter.

Passenger origin and destination statistics cannot be compiled directly from TOD data. TOD data is broken into components which are essentially one-way trips using the Directional Origin and Destination (DOD) concept. The DOD concept can be defined as "points of departure and ultimate destination named in the sequence which indicates the direction of travel".

Some information concerning the markets of foreign (non-U.S.) carriers is obtained from the revenue origin and destination survey if connections are made with one of the participating Canadian or U.S. carrier. For example, if a passenger (whose ticket number ends in a '0') travels from Toronto to Montréal with C P Air, then connects with Lufthansa for Munich, the revenue passenger origin and destination survey will capture the trip because a Canadian carrier participated somewhere in the journey. The Canadian carrier would report the complete carrier and routing detail, including the Lufthansa segment. The foreign carrier has therefore interlined with a Canadian carrier on such a trip. However, there are cases when a foreign carrier does not interline with a Canadian carrier. An example of such an itinerary would be that of a passenger flying on Air France from Paris to Montréal and then back to Paris on Air France. This journey would not be reported to the Revenue Passenger Origin and Destination survey. However, the itineraries of such passengers are in the target population of international journeys. This incomplete coverage of the target population is the non-coverage problem for the ASIPOD estimation system.

The existing system takes the Revenue Passenger Origin and Destination survey data and the Airport Activity Survey data and applies a method called the "assignment technique" in order to produce total market estimates. The Airport Activity Survey counts passengers, on a census basis by flight, entering and leaving each Canadian airport. The survey covers all major Canadian, American and foreign scheduled carriers, but it does not consider the passenger's initial origin or final destination. Hence, the Airport Activity Survey provides a count of the total volune of passengers for each carrier by airport in the target population. The assignment technique is a method of estimating the non-coverage volume of passengers and assigning it to origin and destination pairs. We next proceed to describe it in fuller detail.

As mentioned earlier, from the airport activity data, the census traffic flow data are obtained. Traffic flow can be defined as a count, over a certain period of time (quarters in our instance), of the number of persons who are flying on a specific carrier between a Canadian reporting airport and an adjacent point. The adjacent point is called the next stop or the last stop. For the purposes of the assignment technique, only the traffic data for foreign (non-U.S.) carriers are input into the system. The data elements extracted from this survey, and used to determine the O/D international markets, are the number of revenue passengers enplaned and deplaned in Canada, the Canadian gateway carrier, and the Canadian gateway. In this survey the concept of "Canadian gateway" is defined to be that reporting airport in Canada at which a foreign (non-U.S.) carrier enters or leaves Canada. For the Revenue Passenger O/D survey, the Canadian gateway for Canadian and U.S. carriers is the first/last Canadian point in the itinerary for a passenger entering/leaving Canada. For foreign carriers, the Canadian gateway refers to the point inside Canada where the passenger enters or leaves the foreign carrier.

From the Airport Activity Survey, benchmark counts of passengers entering and leaving Canada at Canadian gateway airports are tabulated by carrier and quarter. These benchmarks are adjusted to exclude foreign to foreign passengers.

From the Revenue Passenger Origin and Destination survey, a corresponding number of inbound and outbound passengers on international DOD's can be tabulated by crossing carrier and Canadian gateway airport. An estimate of the number of passengers, carried on foreign carriers for a given Canadian gateway airport is obtained by subtracting from the corresponding benchmark count the corresponding weighted number of international passengers. The next stage allocates the non-coverage volume to origin-destination pairs which are synthetized on the basis of existing pairs. These pairs are called non-interlining DOD's, since they are DOD's, flown by foreign carriers, which do not interline with a participating Canadian carrier. The assignment technique imputes non-interlining DOD's as follows:

(a) All of the DOD's contributing passengers to Canadian carriers interlining with a foreign (non-U.S.) carrier are identified,

(b) The domestic portion of such DOD's (i.e. the portion from the Canadian point to the Canadian gateway city) is eliminated. (The domestic portion of such DOD's would be on a Canadian carrier, and, would therefore be picked up in the revenue passenger O/D survey.) The resultant "truncated DOD's" are, then, non-interlining,

(c) The non-coverage volume is then assigned to the resultant "sample DOD's" in proportion to their original contribution of weighted number of passengers.

An example of how this assignment technique is carried out, would be useful at this point. Suppose that from the Airport Activity Survey it is known that 4,139 passengers have enplaned/deplaned on Lufthansa (LH) at Mirabel (YMX). Suppose that thirty-one passenger DOD's of the form:

Ottawa - AC - Mirabel - LH - Frankfurt - LH - Munich
(YOW - AC - YMX - LH - FRA - LH - MUC)

and ten one-passenger DOD's of the form

Ottawa - CP - Mirabel - LH - Frankfurt
(YOW - CP - YMX - LH - FRA)

have been selected. In this example, the origin-destination pairs are Ottawa - Munich and Ottawa - Frankfurt, with Mirabel being the Canadian gateway. The assignment technique must generate passenger flow counts between Mirabel - Munich and Mirabel - Frankfurt. These are obtained by prorating the relative proportions of Ottawa - Munich and Ottawa - Frankfurt passengers to the number of passengers remaining after the Combined Ottawa - Munich and Ottawa - Frankfurt count has been taken away from the given passenger flow at Mirabel for Lufthansa. The number of passengers to be assigned is:

$$4{,}139 - (30 \times 10 + 10 \times 10) = 3{,}739.$$

Therefore, 3,739 passengers must have assigned records constructed with Mirabel as the origin and Munich and Frankfurt as the destinations. The relative proportions are 300/400 for Ottawa - Munich and 100/400 for Ottawa - Frankfurt. It is assumed that these proportions can be applied to the 3,739 passengers yielding 0.75 X 3,739 = 2,804 passengers for Mirabel - Munich and 0.25 X 3,739 = 935 passengers for Mirabel - Frankfurt. The resulting origin/destination flow table would be:

	Ottawa	Montréal
Frankfurt LH, YMX	100	935
Munich LH, YMX	300	2,804

It must be noted that the assignment technique "tries" to account for non-interlining traffic, by assuming the same pattern as interlining traffic. The estimates are also produced on a non-directional basis.

This is an overall verbal description of the sampling and estimation methods used in the Bilateral Sub-System. In order to associate estimated variances with estimates of the various origin-destination pairs, the above sampling strategy and estimation method must be cast in terms of algebraic symbols.

2. Estimation of Totals and Associated Variances

A number of simplifying assumptions have been made in order to produce estimates of variance. These are:

a) Although TOD's were sampled, they were broken up into DOD's which were used to produce the estimates of passenger flow. In the development of the variance estimation, TOD's are clusters whose elements, the DOD's, are selected with certainty. The present development assumes that DOD's were sampled at the rate of one in ten. If the ASIPOD system had retained the TOD identification, the variance estimation would have been done at the TOD level, using them as clusters.

b) The number of passengers with a given DOD is zero or one (i.e. no group tickets).

c) The selection of tickets with serial numbers ending in the digit '0' produces a systematic sample which is representative.

The entire Bilateral Sub-System estimation system may be described as using a post-stratified domain estimation. Mixtures of two types of this estimation may occur for given domains of interest (origin-destination pairs for given combinations of carriers). For the post-stratum where there has been no adjustment to the estimates using auxiliary information, the estimation will be referred to as domain estimation. For the post-stratum where there has been adjustment to the estimates using auxiliary information, the estimation will be referred to as adjusted ratio estimation. Some notation is required to formalize the estimation process. To this end, the following symbols will be used.

h: a post-stratum (benchmark) which identifies a gateway - gateway carrier -quarter - fifth/non-fifth freedom when the gateway carrier is a foreign national, otherwise it has no distinction;

d: a domain of interest (for non-assigned records) which consists of origin-destination pairs that can be cross-classified with carriers;

d^*: a domain associated with assigned records relative to domain d;

X_h: known number of passengers for all routes / h-th post-stratum; (excludes estimated foreign to foreign passenger volume).

N_h: estimated population number of DOD's / h-th post-stratum;

n_h: sampled number of DOD's / h-th post-stratum;

$_dn_h$: sampled number of DOD's belonging to the d-th domain / h-th post-stratum;

y_{hi}: number of passengers on the i-th sampled DOD (i = 1, ..., n_h) / h-th post-stratum: this is a zero-one variable;

$_dy_{hi}$: number of passengers on the i-th sampled DOD (i = 1, ..., n_h / h-th post-stratum belonging to the d-th domain (not-assigned);

$_{d^*}y_{hi}$: corresponding number of passengers for assigned DOD's;

Note

$$_dy_{hi} = \begin{cases} y_{hi} & \text{if } i \in d \\ 0 & \text{otherwise} \end{cases}$$

f: the total sampling fraction, i.e. n/N = 0.10;

H: the total number of post-strata.

A. Domain Estimation (non-assigned)

This type of estimation is done for both Canadian carriers and for all carriers that go via U.S. For a given domain "d" of interest, the estimate is

$$ {}_d\hat{Y}_1 = \sum_{h=1}^{H} {}_d\hat{Y}_h $$

where

$$ {}_d\hat{Y}_h = \sum_{i=1}^{n_h} {}_d y_{hi} \times 1/f $$

Note that the summation is done over all post-strata. If for a particular post-stratum "h", no passengers have been sampled, then ${}_d\hat{Y}_h$ is simply zero. If the above estimation is regarded as post-stratified on h, the corresponding estimator of variance for ${}_d\hat{Y}_1$ is given by:

$$ {}_d v_1 = v({}_d\hat{Y}_1) $$

$$ = (f^{-1}-1) \sum_{h=1}^{H} N_h \; {}_d S^2_{1h} + f^{-2} \sum_{h=1}^{H} (1 - N_h/N) \; {}_d S^2_{1h} $$

where N is the total estimated number of population DOD's across all domains in scope to the Bilateral Sub-System, carriers and gateways;

$$ {}_d S^2_{1h} = \frac{1}{n_h - 1} \left[\sum_{i=1}^{{}_d n_h} y^2_{hi} - n_h \left(\sum_{i=1}^{{}_d n_h} y_{hi} / n_h \right)^2 \right] $$

$$ = \frac{1}{n_h - 1} \left({}_d n_h - {}_d n^2_h / n_h \right) $$

$$ = {}_d\hat{R}_{1h} \left(1 - {}_d\hat{R}_{1h}\right) $$

with

$$ {}_d\hat{R}_{1h} = {}_d n_h / n_h \,. $$

B. Adjusted Ratio Estimation (Assigned)

This type of estimation is done in order to obtain passenger flow estimates for foreign (non-U.S.). Using known post-strata passenger counts X_h, tickets which have been sampled and which have Canadian or U.S. carriers interlining with foreign carriers are prorated at the first/last Canadian gateway if the ticket is inbound/outbound. Two cases of this estimation will be considered: i) One in which the level of the origin-destination pair of the assigned record is not equal to the original non-assigned record ($d^* \neq d$) and ii) the other where the level of the origin-destination pair of the assigned record corresponds to the original non-assigned record ($d^* = d$). An example to illustrate the above cases would be useful at this point. Suppose that a ticket sampled at the Quebec airport provides the following itenary

Quebec - AC - Mirabel - AF - Paris

This ticket will give rise to a non-assigned origin-destination pair (Quebec, Paris) and an assigned origin-destination pair (Mirabel, Paris): in this case $d^* \neq d$. However, if the origin-destination level is country, the origin-destination pair is (CANADA, FRANCE), and $d^* = d$.

a. Case where $d^* \neq d$

The estimate for a given domain of interest "d*" is:

$$_{d^*}\hat{Y}_2 = \sum_{h=1}^{H} (X_h - \hat{Y}_h) \; ({}_{d^*}\hat{Y}_h \,/\, \hat{Y}_h)$$

where $\hat{Y}_h$ is the sum of weighted up number of passengers flying in/out of the gateway amounted with post-stratum h on a foreign carrier interlining with a Canadian or a U.S. carrier, and $_{d^*}\hat{Y}_h$ is similarly defined for the domain "d". Letting D_h^* being the union of all records to be assigned within the h-th post-stratum, and m_h as the number of tickets to be assigned within the h-th post-stratum,

$$_{d^*}\hat{Y}_h = \sum_{i=1}^{m_h} {}_{d^*}y_{hi} \times 1/f$$

and

$$\hat{Y}_h = \sum_{d^* \varepsilon D_h^*} {}_{d^*}\hat{Y}_h \ .$$

The corresponding variance estimator for ${}_{d^*}\hat{Y}_2$ is

$${}_{d^*}v_2 = v({}_{d^*}\hat{Y}_2)$$

$$\doteq (f^{-1}-1) \sum_{h=1}^{H} N_h \; {}_{d^*}S_{2h}^2 + f^{-2} \sum_{h=1}^{H} (1 - N_h/N) \; {}_{d^*}S_{2h}^2$$

where

$${}_{d^*}S_{2h}^2 = \frac{1}{m_h - 1} \left[\sum_{i=1}^{m_h} {}_{d^*}w_{hi}^2 - m_h \left(\sum_{i=1}^{{}_{d^*}m_h} y_{hi} \,/\, m_h \right)^2 \right]$$

with

$${}_{d^*}w_{hi} = \hat{R}_{2h} \left({}_{d^*}y_{hi} - {}_{d^*}\hat{R}_{3h} \, y_{hi} \right) - {}_{d^*}y_{hi}$$

$$\hat{R}_{2h} = X_h \,/\, \hat{Y}_h \quad \text{and} \quad {}_{d^*}\hat{R}_{3h} = \left({}_{d^*}\hat{Y}_h \,/\, \hat{Y}_h \right) .$$

It can be shown that ${}_{d^*}S_{2h}^2$ simplifies to:

$${}_{d^*}\hat{R}_{3h} \left(1 - {}_{d^*}\hat{R}_{3h}\right) X$$

$$\left[{}_{d^*}\hat{R}_{3h} \; \hat{R}_{2h} \left(\hat{R}_{2h} - 2\right) + 1 \right]$$

b. Case where d* = d

This case could occur, for instance, when the origin or destination of a ticket that is to be prorated is a Canadian gateway. The estimate for a given domain of interest d* is:

$$ {}_{d*}\hat{Y}_3 = \sum_{h=1}^{H} X_h \left({}_{d*}\hat{Y}_h \,/\, \hat{Y}_h \right) $$

If there are k_h tickets in the post-stratum and that ${}_{d*}k_h$ of these tickets belong to the "d*" domain, the corresponding variance estimator for ${}_{d*}\hat{Y}_3$ is:

$$ {}_{d*}v_3 = v({}_{d*}\hat{Y}_3) $$

$$ \doteq (f^{-1}-1) \sum_{h=1}^{H} N_h \; {}_{d*}S^2_{3h} + f^{-2} \sum_{h=1}^{H} (1 - N_h/N) \; {}_{d*}S^2_{3h} $$

where

$$ {}_{d*}S^2_{3h} = \frac{1}{k_h - 1} \left[\sum_{i=1}^{{}_{d*}k_h} y^2_{hi} - 2\,{}_{d*}\hat{R}_{4h} \sum_{i=1}^{{}_{d*}k_h} y^2_{hi} + {}_{d*}\hat{R}^2_{4h} \sum_{i=1}^{k_h} y^2_{hi} \right] $$

It can be shown that ${}_{d*}S^2_{3h}$ simplifies to:

$$ {}_{d*}\hat{R}_{4h} \left(1 - {}_{d*}\hat{R}_{4h}\right) \quad \text{where} \quad \hat{R}_{4h} = \left({}_{d*}k_h \,/\, k_h \right) . $$

Denoting as ${}_{d}\hat{Y}_{2h}$, ${}_{d}\hat{Y}_{2h}$ and ${}_{d}\hat{Y}_{3h}$ as the estimates for post-stratum h and domain d* for the previously mentioned cases, it is possible that an estimate of the number of passengers for the given post-stratum h and domain d* be the sum of the three components. That is,

$$ {}_{d}\hat{Y}_h = {}_{d}\hat{Y}_{1h} + {}_{d}\hat{Y}_{2h} + {}_{d}\hat{Y}_{3h} . $$

It can be shown that the estimate of variance for ${}_{d}\hat{Y}_h$ is

$${}_d S_h^2 = \hat{P}_h \{ {}_d \hat{R}_{2h} [1-({}_d \hat{R}_{3h} + {}_d \hat{R}_{4h})] [{}_d \hat{R}_{2h} ({}_d \hat{R}_{3h} + {}_d \hat{R}_{4h}) - 2\,{}_d \hat{R}_{3h}]$$

$$+ {}_d \hat{R}_{3h} (1 - {}_d \hat{R}_{3h}) \}$$

$$+ \hat{Q}_h \, {}_d \hat{R}_{1h} (1 - {}_d \hat{R}_{1h})$$

$$+ \hat{P}_h \hat{Q}_h ({}_d \hat{R}_{3h} + {}_d \hat{R}_{1h})^2$$

where $\hat{P}_h = n_h / (k_h + m_h + n_h)$ is the ratio of the number of tickets that have been assigned to the total number of tickets including assigned tickets in a given post-stratum h. $\hat{Q}_h$ is simply $1 - \hat{P}_h$.

An estimate for a given domain d, is then

$${}_d \hat{Y} = \sum_{h=1}^{H} {}_d \hat{Y}_h$$

and the corresponding variance is

$${}_d v \doteq (f^{-1}-1) \sum_{h=1}^{H} N_h \; {}_d S_h^2 + f^{-2} \sum_{h=1}^{H} (1-N_h/N) \; {}_d S_h^2 .$$

The associated estimated coefficient of variation is

$$cv\,({}_d \hat{Y}_h) = ({}_d v)^{1/2} \; / \; {}_d \hat{Y}.$$

4. Acknowledgement

The authors would like to thank Gord Baldwin of the Transportation Division for many helpful comments related to the ASIPOD estimation system.

References

[1] Carpenter, R.: "Specifications for Estimates of the Reliability for the Air Scheduled International Passenger Origin and Destination Estimation", prepared in the Business Survey Methods Division of Statistics Canada, revised January 1982.

[2] Hunter, Greg and DiPiétro, Lisa (1983): "The Methodology of the Canadian Air Scheduled International Passenger Origin and Destination Estimation System", pp. 27-49, Survey Methodology, Vol. 9, No. 1.

APPENDIX C

DESCRIPTION OF THE DATA

APPENDIX C - Description of the Data

A. Never-prorated Carriers (Group I)

This group consisted of the following five carriers: AC, CP, PW, UK[1] and US[2]. All PW flights were via the United States, while the other four also had direct flights.

The domains for this group were various combinations of the five carriers, 240 foreign cities, 12 Canadian cities and 13 gateways. Via-U.S. domains were also combined with an unknown number of carriers and U.S. gateways. The cities are listed in Appendix A. In all, this group had 11,568 domains. The coefficient of variation and the proration percentage were estimated for each domain.

Passenger volume varied widely from domain to domain; the smallest domain had 10 passengers and the largest 91,000. The mean number of passengers was 155, but the standard deviation was 1,641. The direct domains, however, when taken separately, had an average of 262 passengers, compared with only 50 for the via-U.S. domains. Nevertheless, the distribution of passengers was similar for both types of flights. Table V shows the distribution of passengers by domain. Note that there were more small domains than large ones.

---

1 UK was used to denote unidentified carriers.
2 U.S. carriers were all denoted by US.

TABLE V
Distribution of Passengers
(Group I)

Number of passengers		FREQ.	CUM. FREQ.	PERC.	CUM. PERC.
0 - 80	********************	10008	10008	86.51	86.51
80 - 240	**	892	10900	7.71	94.23
240 - 400		219	11119	1.89	96.12
400 - 560		116	11235	1.00	97.12
560 - 880		72	11307	0.62	97.74
880 - 1040		51	11358	0.44	98.18
1040 - 1200		27	11385	0.23	98.42
1200 - 1360		22	11407	0.19	98.61
1360 - 1520		19	11426	0.16	98.77
1520 - 1680		13	11439	0.11	98.88
1680 - 1840		7	11446	0.06	98.95
1840 - 2000		11	11457	0.10	99.04
2000 - 2160		12	11469	0.10	99.14
2160 - 2320		9	11478	0.08	99.22
2320 - 2480		5	11483	0.04	99.27
2480 - 2640		8	11491	0.07	99.33
2640 - 2800		1	11492	0.01	99.34
2800 - 2960		4	11496	0.03	99.38
2960 - 3120		3	11499	0.03	99.40
3120 - 3280		0	11499	0.00	99.40
3280 - 3400		4	11503	0.03	99.44
3440 - 3600		0	11503	0.00	99.44
3600 - 3760		1	11504	0.01	99.45
3760 - 3920		2	11506	0.02	99.46
3920 and more		0	11506	0.00	99.46
		62	11568	0.54	100.00
	4000 8000				

Legend: FREQ. = frequency
CUM. FREQ. = cumulative frequency
PERC. = percentage
CUM. PERC. = cumulative percentage

B. Non-prorated and Prorated Carriers (Groups II and III)

The 1,855 domains in Group II were the result of combinations of 22 carriers, 62 foreign cities, 11 Canadian cities and 3 gateways. Group III was composed of 1,567 domains involving 22 carriers, 125 foreign cities, 5 Canadian cities and 5 gateways.

Where passengers were assigned, the proration proportions were very large.

TABLE VII
Distribution of Proration Proportions

		FREQ.	CUM. FREQ.	PERC.	CUM. PERC.
0.0	*************************	1855	1855	54.21	54.21
0.1		3	1858	0.09	54.30
0.2		0	1858	0.00	54.30
0.3		0	1858	0.00	54.30
0.4		0	1858	0.00	54.30
0.5		1	1859	0.03	54.32
0.6		8	1867	0.23	54.56
0.7	*	97	1964	2.83	57.39
0.8	***	193	2157	5.64	63.03
0.9	*******	523	2680	15.28	78.32
1.0	**********	742	3422	21.68	100.00
	600 1200 1800				

This table shows that the domains can be separated into two classes: one with zero proration and the other with very high proration proportions. These two classes were associated with passenger volume; zero prorations were associated with small domains.

TABLE VIII
Distribution of Prorated and Non-prorated Routes
by Passenger Volume

			PASSENGERS					
			0-100	100-300	300-500	500-600	700+	TOTAL
PRORA-TED	WITH	FREQ.	617	496	156	83	215	1567
		PERC.	18.03	14.49	4.56	2.43	6.28	45.79
	NON	FREQ.	1702	97	26	14	16	1855
		PERC.	49.74	2.83	0.76	0.41	0.47	54.21
	TOTAL	FREQ.	2319	593	182	97	231	3422
		PERC.	67.77	17.33	5.32	2.83	6.75	100.00

Legend: FREQ. = frequency
PERC. = percentage

Groups II and III had one feature in common with Group I: small domains far outnumbered large ones. However, mean passenger volume was higher in Groups II and III than in Group I (414 compared with 155).

TABLE IX
Distribution of Passengers
Group II and III

		FREQ.	CUM. FREQ.	PERC.	CUM. PERC.
0 - 80	**********************	2187	2187	63.91	63.91
80 - 240	******	633	2820	18.50	82.41
240 - 400	**	192	3012	5.61	88.02
400 - 560	*	118	3130	3.45	91.47
560 - 720	*	71	3201	2.07	93.54
720 - 880		32	3233	0.94	94.48
880 - 1040		25	3258	0.73	95.21
1040 - 1200		22	3280	0.64	95.85
1200 - 1360		17	3297	0.50	96.35
1360 - 1520		12	3309	0.35	96.70
1520 - 1680		6	3315	0.18	96.87
1680 - 1840		7	3322	0.20	97.08
1840 - 2000		12	3334	0.35	97.43
2000 - 2160		6	3340	0.18	97.60
2160 - 2320		7	3347	0.20	97.81
2320 - 2480		3	3350	0.09	97.90
2480 - 2640		2	3352	0.06	97.95
2640 - 2800		5	3357	0.15	98.10
2800 - 2960		1	3358	0.03	98.13
2960 - 3120		2	3360	0.06	98.19
3120 - 3280		3	3363	0.09	98.28
3280 - 3440		3	3366	0.09	98.36
3440 - 3600		1	3367	0.03	98.39
3600 - 3760		1	3368	0.03	98.42
3760 - 3920		1	3369	0.03	98.45
3920 and more		53	3422	1.55	100.00
	500 1000 1500 2000				

APPENDIX D

STATISTICAL ANALYSES

Statistical Analyses

For all groups, the first model fitted was the following:

$$CV_i{}^2 = A + \frac{B}{PA_i} + E_i \qquad (4)$$

where E is the error term. There was a good reason for trying this model first: the formulas developed in Appendix B implicitly assumed that the passenger volume in a given domain i had a multinomial distribution with mean NP_i, where N is the number of passengers on scheduled international flights and P_i is the proportion of passengers who are in domain i $(0 \quad P_i \quad 1)$. Under these conditions, the CV is given by

$$CV_i{}^2 = \frac{1}{nP_i} - \frac{1}{n} \qquad (5)$$

where n is the sample size. Equation (5) is similar to equation (4) except that in the former, P_i replaces PA_i (passenger volume).

For never-prorated carriers (AC, CP, PW, UK and US), model (4) yielded a perfect fit no matter whether the flight was direct or via the United States. For other carriers, however, it was preferable to use a more general model of the form

$$\log(CV_i) = A + B \log(PA_i) + E_i \qquad (6)$$

Chart II suggested that for these carriers, proration had a definite effect on the CVs. Consequently, separate analyses were performed for non-prorated domains.

A. Non-prorated domains (Group II)

Even though model (6) was selected to estimate the CVs of these domains, a covariance analysis suggested that the carrier had a significant effect on the CV (1% level).

The remaining variables studied (city of origin, province of origin, city of destination, country of destination and destination area code) had no significant impact. In addition, there appeared to be no interaction between these variables and passenger volume; that is, the slopes did not vary significantly from variable to variable.

B. Prorated domains (Group III)

As in the case of Group II, it was preferable to use a model similar to (6) rather than model (4). To improve the estimates, it was necessary to incorporate the carrier effect and the gateway effect, both of which were significant at the 1% level. Although the model's R^2 was lower for this group (0.96) than for the other groups, no other variable had a significant impact. Interaction between passenger volume and the other variables was also non-existent.

APPENDIX E

PARAMETERS

GROUP I

equation: $\log(CV_i) = A + B \times \log(PA_i)$

$A = \log(3)$

$B = -0.5$

or $CV_i = \frac{3}{\sqrt{PA_i}}$ $PA_i \geqslant 100$

GROUP II

equation: $\log(CV_i) = A + B \times \log(PA_i)$

$A = 1.036851524$

$B = 0.54100134$

or $CV_i = \frac{M}{PA_i^N}$ $PA_i \geqslant 100$

$M = 2.820323301$

$N = 0.54100134$

GROUP III

equation: $\log(CV_{ijk}) = A + T_j + G_k + B \times \log(PA_{ijk})$ $PA_{ijk} \geqslant 600$

$A = 1.571377254$

$B = -0.57837335$

$T_j + G_k =$	Carrier	Gateway	Line (L)
2.61210291	CU	YMX	1
1.50198300	LO	YMX	1
1.45320426	AZ	YMX	1
1.32524031	KL	YMX	1
1.18536220	JM	YYZ	2
1.18366674	SN	YMX	2
1.14969342	AF	YMX	2
1.08081594	LH	YMX	2
1.05349382	AT	YMX	2
1.01996794	BW	YYZ	2
0.97313382	BA	YVR	2
0.95485891	BA	YYZ	2
0.91587089	BA	YMX	2
0.86792814	KL	YYZ	3
0.83238768	SR	YMX	3
0.79997656	LH	YYZ	3

0.76638625	AF	YYZ	3
0.76216066	AZ	YYZ	3
0.74530637	SU	YMX	3
0.65271190	BA	YEG	3
0.63040008	AR	YMX	3
0.62294799	IB	YMX	3
0.60126050	QF	YVR	3
0.46588120	JL	YVR	4
0.39397847	BA	YYC	4
0.35975005	SR	YYZ	4
0.16005781	OK	YMX	4
0.09502567	AY	YMX	4
0.00000000	TP	YMX	4
-0.31049330	AI	YMX	4
-0.58781219	LY	YMX	4

Approximation:

equation: $\log(CV_{ij}) = A + L_j + B \times \log(PA_{ij})$

$A = 1.571377254$

$B = -0.57837335$

$L_1 = 2.61210291$

$L_2 = 1.18536220$

$L_3 = 0.86792814$

$L_4 = 0.46588120$

or $$CV_{ij} = \frac{A \times L_j}{PA_{ij}^{\ B}}$$

with

$A = 4.81327273$

$L_1 = 13.62767852$

$L_2 = 3.271871679$

$L_3 = 2.381970628$

$L_4 = 0.331046625$

$B = 0.578373350$

APPENDIX F

CHARTS AND TABLES

CHARTS AND TABLES

The coefficients of variation (CV), presented in the following two tables and two charts, are based on passenger volume. The CVs for non-prorated domains are shown in Table I and Chart I; those for prorated domains are shown in Table II and Chart II.

Where a CV for a particular passenger volume is not in the table, refer to the corresponding chart.

A. Non-prorated domains

CVs are given separately for carriers that never have proration (grouping 1) and carriers that may have proration (grouping 2). Only the CVs for non-prorated domains are given in the chart and table below. The CVs for prorated domains are presented in part B of this appendix.

The groupings of non-prorated domains are composed of the following carriers:

Grouping 1: AC
CP
PW
UK (unidentified carriers)
US (all U.S. carriers)

Grouping 2:	AF	BW	LY
	AI	IB	OK
	AR	JL	QF
	AT	JM	SN
	AY	KL	SR
	AZ	LH	SV
	BA	LO	TP

If estimates are required for a non-prorated domain and the carrier is not in either of the above groupings (i.e. a new carrier), it is assumed to belong to grouping 1.

TABLE I
COEFFICIENT OF VARIATION
NON-PRORATED DOMAINS

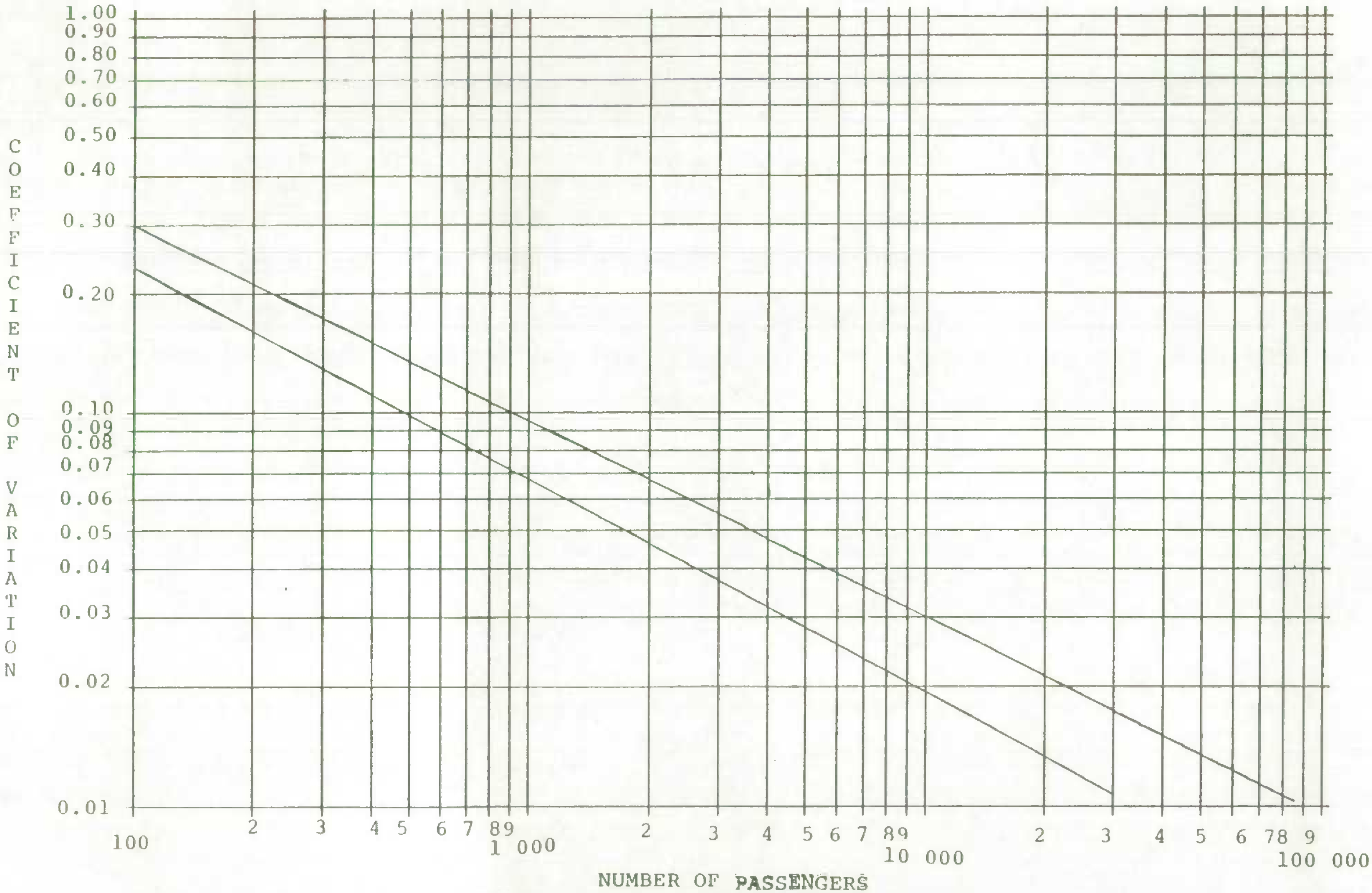
Number of passengers	Grouping 1	Grouping 2
100	0.300	0.234
200	0.212	0.160
300	0.173	0.129
400	0.150	0.110
500	0.134	0.098
600	0.122	0.089
700	0.113	0.081
800	0.106	0.076
900	0.100	0.071
1000	0.095	0.067
1500	0.077	0.054
2000	0.067	0.046
2500	0.060	0.041
3000	0.055	0.037
3500	0.051	0.034
4000	0.047	0.032
4500	0.045	0.030
5000	0.042	0.028
5500	0.040	0.027
6000	0.039	0.025
6500	0.037	0.024
7000	0.036	0.023
7500	0.035	0.023
8000	0.034	0.022
8500	0.033	0.021
9000	0.032	0.020
9500	0.031	0.020
10000	0.030	0.019
15000	0.024	0.016
20000	0.021	0.013
25000	0.018	0.012
30000	0.017	0.011
35000	0.016	0.010
40000	0.015	0.009
45000	0.014	0.009
50000	0.013	0.008
55000	0.013	0.008
60000	0.012	0.007
65000	0.012	0.007
70000	0.011	0.007
75000	0.011	0.007
80000	0.011	0.006
85000	0.010	0.006
90000	0.010	0.006
95000	0.010	0.006
100000 or more	0.009	0.006

Grouping 1: Carriers that never have proration (AC, CP, PW, UK, US)
Grouping 2: Carriers that may have proration.

GRAPH I

GRAPH OF CV VERSUS PASSENGERS

FOR DOMAINS WITHOUT PRORATION



LINE 1 CARRIER AC CP PW US AND UK
LINE 2 ALL OTHER CARRRIERS

B. Prorated domains

For prorated domains, the CVs vary not only by passenger volume and by carrier, but also by Canadian gateway. As a result, the following four groupings have been defined, each containing a limited number of gateway-carrier combinations:

Grouping 1:
- CU gateway YMX
- LO gateway YMX
- AZ gateway YMX
- HL gateway YMX

Grouping 2:
- JM gateway YYZ
- SN gateway YMX
- AF gateway YMX
- LH gateway YMX
- AT gateway YMX
- BW gateway YYZ
- BA gateway YVR
- BA gateway YYZ
- BA gateway YMX

Grouping 3:
- KL gateway YYZ
- SR gateway YMX
- LH gateway YYZ
- AF gateway YYZ
- AZ gateway YYZ
- SU gateway YMX
- BA gateway YEG
- AR gateway YMX
- IB gateway YMX
- QF gateway YVR

Grouping 4:
- JL gateway YVR
- BA gateway YYC
- SR gateway YYZ
- OK gateway YMX
- AY gateway YMX
- TP gateway YMX
- AI gateway YMX
- LY gateway YMX

New carriers and any gateway-carrier combination not listed above is to be associated with grouping 1.

For example, if the CVs for carrier BA at gateway YYZ are required, the CVs for grouping 2 are used. If the CVs for the same carrier at gateway YYC are required, the CVs for grouping 4 are used.

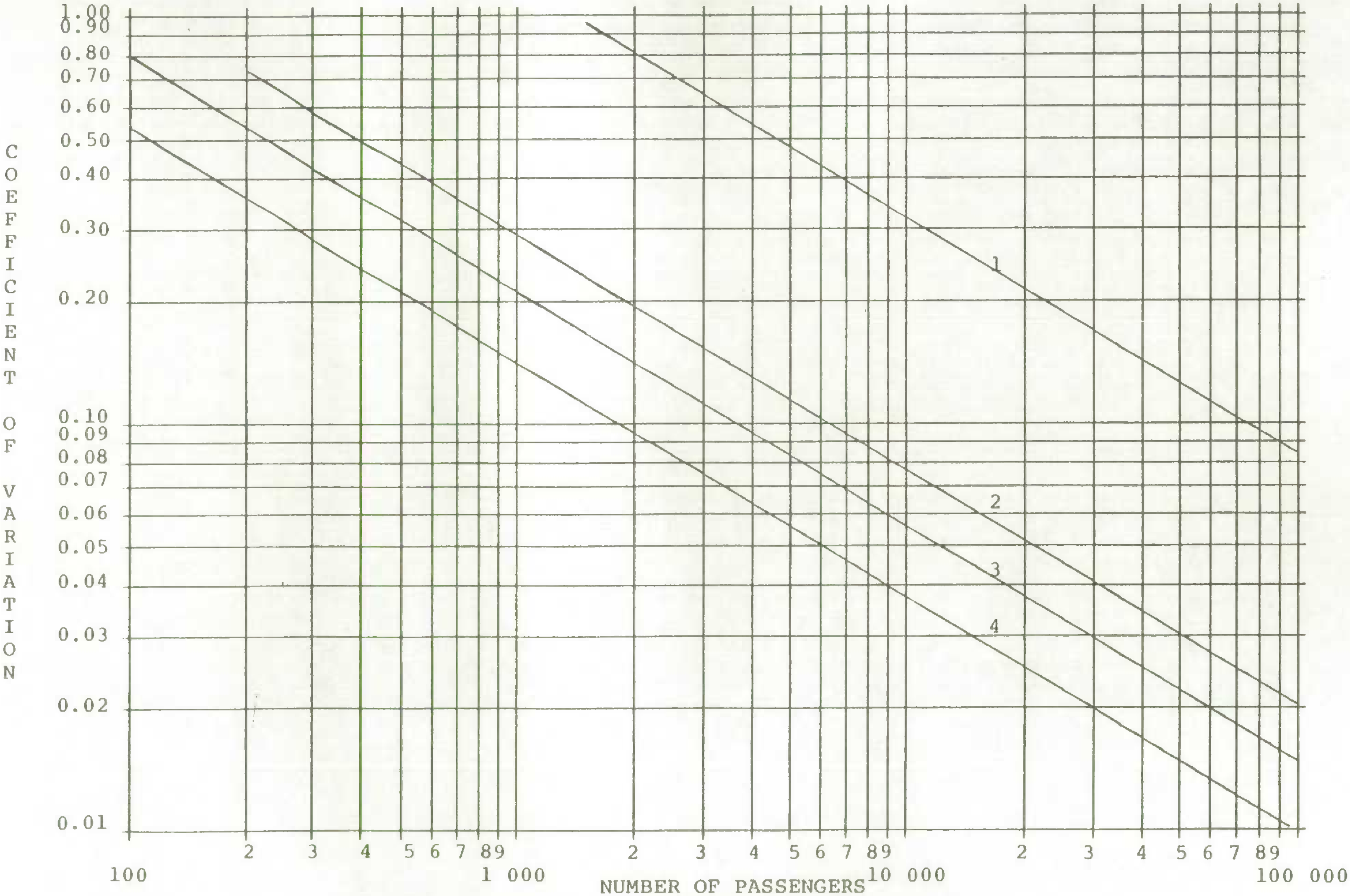
TABLE II
COEFFICIENT OF VARIATION
PRORATED DOMAINS

Number of passengers	Grouping			
	1	2	3	4
600			0.284	0.190
700			0.259	0.174
800			0.240	0.161
900			0.224	0.150
1000		0.290	0.210	0.141
1500		0.229	0.167	0.112
2000		0.194	0.141	0.095
2500		0.171	0.124	0.083
3000		0.154	0.112	0.075
3500		0.140	0.102	0.068
4000		0.130	0.095	0.063
4500		0.121	0.088	0.059
5000		0.114	0.083	0.056
5500		0.108	0.079	0.053
6000		0.103	0.075	0.050
6500		0.098	0.072	0.048
7000		0.094	0.069	0.046
7500		0.090	0.066	0.044
8000		0.087	0.063	0.042
8500		0.084	0.061	0.041
9000		0.081	0.059	0.040
9500		0.079	0.057	0.038
10000		0.077	0.056	0.037
15000	0.252	0.061	0.044	0.030
20000	0.213	0.051	0.037	0.025
25000	0.188	0.045	0.033	0.022
30000	0.169	0.041	0.029	0.020
35000	0.154	0.037	0.027	0.018
40000	0.143	0.034	0.025	0.017
45000	0.134	0.032	0.023	0.016
50000	0.126	0.030	0.022	0.015
55000	0.119	0.029	0.021	0.014
60000	0.113	0.027	0.020	0.013
65000	0.108	0.026	0.019	0.013
70000	0.103	0.025	0.018	0.012
75000	0.099	0.024	0.017	0.012
80000	0.096	0.023	0.017	0.011
85000	0.092	0.022	0.016	0.011
90000	0.089	0.021	0.016	0.011
95000	0.087	0.021	0.015	0.010
100000 or more	0.084	0.020	0.015	0.010

Excessively large CVs are useless and have therefore been omitted.

GRAPH OF CV VERSUS PASSENGERS

FOR DOMAINS WITH PRORATION



SEE NEXT PAGE FOR CARRIER-GATE LIST

C. Composite CVs

It is also possible to obtain a CV for a city pair linked by more than one carrier and/or via more than one gateway. This CV is a weighted sum of the CVs of all possible carriers and/or gateways for this city pair. The square of the CV is given by

$$CV^2 = \frac{CV_1{}^2 \times PA_1{}^2 + CV_2{}^2 \times PA_2{}^2 + CV_3{}^2 + \ldots + CV_t{}^2 \times PA_t{}^2}{(PA_1 = PA_2 + PA_3 + \ldots + PA_t)^2}$$

where $CV_i{}^2$ is the square of the CV for the i^{th} carrier or the i^{th} gateway;

PA_i is the number of passengers using the i^{th} carrier or the i^{th} gateway.

D. Examples

1. Calculate the CV for the city pair Montreal-Paris (YMX-PAR). It is served only by AC and AF, the former carrying 1,000 passengers and the latter 2,500, 1,300 of which are assigned.

   First find the CVs for AC and AF.

   For AC, you must use Table I (non-prorated), grouping 1 (never-prorated carriers). For AF, you must use Table II (prorated), grouping 2, which includes the combination AF-YMX. This gives us

   CV_{AC} for 1,000 passengers: 0.095

   CV_{AF} for 2,500 passengers: 0.171

   Then we use equation (1):

$$CV^2 = \frac{CV_{AC}^2 \times PA_{AC}^2 + CV_{AF}^2 \times PA_{AF}^2}{(PA_{AC} + PA_{AF})^2}$$

   Finally, we take the square root and we get a CV of 0.126.

2. Calculate the CV for the city pair Ottawa-Paris, linked by AF; 2,500 passengers went to Paris through Mirabel and 1,200 through Toronto.

   Since there are no assigned passengers, we can use Table I and Chart I.

   As AF is in grouping 2, we have

   CV_{AF-YMX} for 2,500 passengers = 0.041

   For Toronto, we obtain the CV from Chart I; for 1,200 passengers, it is 0.066.

Using equation (1), we obtain

$$CV^2 = \frac{CV^2_{AF-YMX} \times PA^2_{AF-YMX} + CV^2_{AF-YYZ} \times PA^2_{AF-YYZ}}{(PA_{AF-YMX} + PA_{AF-YYZ})^2}$$

$$= \frac{0.041^2 \times 2500^2 + 0.066^2 \times 1200^2}{(2500 + 1200)^2}$$

$$= 0.001$$

Taking the square root, we get a CV of 0.03.