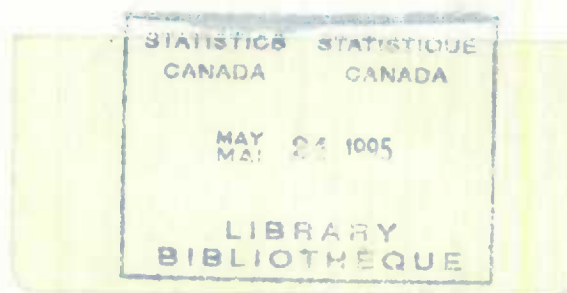Statistics   Statistique
Canada   Canada

Methodology Branch

Social Survey
Methods Division

Direction de la méthodologie

Division des méthodes
d'enquêtes sociales

Canada

WORKING PAPER NO. SSMD 95-04 E

METHODOLOGY BRANCH

Time Series Methods for Postcensal Estimation of Household Counts

Susana Rubin-Bleuer

SSMD 95-04 E

Abstract

We are concerned with the estimation of a population characteristic Y_t which changes over time, using data from a repeated sample survey carried out at regular time intervals. We propose to achieve accuracy by careful modelling of the survey error, acknowleging a bias component in the direct survey estimator, and the use of auxiliary information represented by a 'known' survey error in estimating another characteristic X_t. We will use state space models and the Kalman Filter for the estimation. We also address the problem of internal consistency between the sum of the published estimates of individual domains and the estimate of the aggregate. We impose calibration conditions which can be expressed in state space form. We present an application to postcensal estimation of the current number of households: the calibration equations induce a 'joint model' which effectively reduce the variability of the estimators. A measure of the model bias is obtained by the relative error in estimation at June 1991 when the corresponding census counts are available.

Résumé

Dans cet article, nous utilisons les données d'une enquête répétée à intervalles réguliers pour estimer une caractéristique d'une population qui évolue avec le temps. Pour obtenir des estimations plus précises, nous modélisons l'erreur d'échantillonnage, en reconnaissant explicitement un biais d'échantillonnage dans les estimés de l'enquête, et nous utilisons l'information auxiliaire sur la vraie valeur de l'erreur d'échantillonnage dans l'estimation d'une autre caractéristique. Nous utilisons les vecteurs d'état et le filtrage de Kalman pour l'estimation. De plus, nous proposons des solutions au problème de l'accord entre la somme des estimés des domaines et l'estimé de l'agrégation des domaines. Nous imposons des conditions d'étalonnage que nous exprimons sous forme de vecteur d'états. Nous présentons une application à l'estimation post-censitaire du nombre ponctuel de ménages: les équations d'étalonnage introduisent un modèle conjoint qui réduit la variabilité des estimés. Une mesure du biais de modélisation s'obtient en calculant l'erreur relative de l'estimation du mois de Juin 1991, quand les comptes du recensement deviennent disponibles.

•1. Introduction

The Canadian Census of Population produces, among other things, counts on the distribution of the population and households. During the five-year period between censuses, the Population Estimates Program of Statistics Canada provides population counts by province, and these estimates are constructed from the previous census by adding the components of change (births, deaths and migration) obtained from administrative records. Unlike postcensal estimates of population, estimates of the number of households cannot be obtained from a combination of census counts and administrative data, since there are no administrative records complete enough for this task.

However, reliable postcensal estimates of the number of households (or occupied private dwellings) are very much in demand for the production of statistics on housing, marketing research, etc. Also, good estimates of the number of households by number of persons could be used for developing sampling weights for household surveys.

We have to turn to other sources to construct such estimates. One possible source is the Canadian Labour Force Survey (CLFS). The CLFS yields monthly estimates of both the current number of households and the LFS population, but the estimators are subject to sampling variability and bias and therefore model-based smoothing and benchmarking with census data may produce more efficient and less biased estimators.

The provincial estimators of the number of households have small coefficients of variation. For example, from July to September 1993 they ranged between 0.5% and 1.2%. Even so, the estimates are not reliable enough for our purpose: we seek for estimators with a degree of accuracy comparable to the Postcensal Estimates of Population. Hence we resort to borrow information from other time periods. Most available techniques assume that direct survey estimators are approximately unbiased; but we will acknowledge a bias component in the direct survey estimator because of the accuracy demanded.

In general, we are concerned with the estimation of some population characteristic Y_t, such as the population total, which changes over time. Suppose that the data available are from partially overlapping repeated sample surveys carried out on the population of interest at regular time intervals. Let y_t represent a direct sample survey estimate of Y_t based on the sample at time t alone, so that

$$y_t = Y_t + \epsilon_t \tag{1.1}$$

where ϵ_t denotes the survey error at time t. Suppose further, that as a by-product, the survey

yields the estimate x_t of an auxiliary characteristic for which we know the true value X_t. For example, we let y_t be the monthly CLFS estimate of the number of households, x_t the monthly estimate of the population count and X_t the corresponding population count produced by the Postcensal Population Estimates Program of Statistics Canada.

Here y_t and x_t are subweighted estimators (design based estimators before any postratification is done) of households and population counts respectively. Postratification (or calibration) is done in an effort to deal with the coverage error. Calibration of y_t to the distribution of the population obtained from the Postcensal Estimates yields both the combined ratio estimator $y_{Rt} = y_t \cdot X_t / x_t$ and the separate ratio (or "final weight") estimator y_{ft}.

The postcensal estimates of the population are considered very accurate and could be treated as the true counts of the population. Hence, the data can be viewed as a bivariate series of estimates

$$\begin{pmatrix} y_t \\ x_t - X_t \end{pmatrix} \qquad (1.2)$$

where the first component is the survey estimate of the characteristic of interest and the other component could represent the survey error in estimating the auxiliary quantity X_t.

In order to improve the estimation, modelling different responses could be considered. The bivariate series in (1.2) can be modelled taking advantage of its correlation structure if we assume that there is an association between the survey error in estimating the number of households and the survey error in estimating the population count, $x_t - X_t$. Among the univariate responses we can look at the CLFS subweighted estimator y_t, at the combined ratio estimator y_{Rt} or at the final weight (poststratified) estimator y_{ft}.

Previous proposals to solve this problem considered the headship rate $h_t = y_t / x_t$ (Lemaitre, 1989) or the ratio estimate y_{Rt} (Ghangurde, 1991) as a response and used global models (e.g. a deterministic polynominal); they did not obtain satisfactory results probably due to the lack of flexibility of the models.

The CLFS is a rotating monthly panel survey in which five-sixths of the sample (five panels out of six) is retained from one month to the next. A detailed description of the design of the CLFS and the calculation of the survey estimators can be found in Singh, Drew, Gambino and Mayda (1990). The partial overlapping of the CLFS creates serial correlations between the errors of the survey estimators.

We propose to obtain accurate estimators by means of a time series method that

a)   uses structural modelling, that is, the main components of the series are modelled explicitly and have a direct interpretation,

b)   accounts for the serial correlations of the survey estimators over time,

c)   corrects for bias by benchmarking with previous census counts (the benchmarking equation makes the model identifiable when we introduce a bias term in the survey error),

d)   uses the available auxiliary information, represented by the CLFS and Postcensal Estimates of Population at month t ,

       and

e)   calibrates the sum of domain estimates to coincide with the model-based estimate of the aggregate in order to achieve of internal consistency.

Point c) is a modelling feature not usually found in the analysis of survey data and point e) represents a new technique of calibration with a model-based estimator.

We consider only models with a state space configuration and obtain the model-based estimators via the Kalman Filter algorithms. This approach yields estimators of the unobserved components of interest (the population parameter Y_t, the survey errors ϵ_t, etc.), their mean square errors and easily computed measures of goodness of fit. If, in addition, the model includes benchmarking conditions, the state space framework and the Kalman Filter provide a most natural way to incorporate them.

In order to measure the error in estimation, variance alone may be misleading and we need to have as well an estimate of the bias (when the model does not hold). The bias here refers to the model bias and should not be confused with the bias of the survey estimator, which will be accounted for in the model. In general, the magnitude of the model bias is difficult to estimate. When more accurate estimates exist at specified time periods, these can be used to estimate the bias in the model-dependent estimators. In our case, not only can we estimate the variability of the estimator under different models, but also we can obtain an overall measure of the error in estimation at the time period of the last census, June 1991, because we know the true value of the number of households in June 1991 (the estimators incorporate data until May 1991, including previous census counts).

So the model-based methods will be evaluated in terms of goodness of fit statistics and of the relative error in estimation (error of closure) at June 1991.

The population targets of the CLFS and the Census of Population and Households do not coincide, and hence the census figures were modified to account for this. The Postcensal Estimates of Population used in the calculation of the CLFS estimates and the modified census counts were adjusted using the appropriate census net undercoverage rates (cf. Bleuer and Declos, (1995) Appendix II).

Section 2 describes the design considerations that lead us to the class of models that we propose for the postcensal estimation of the current number of households.

In section 3 we present the empirical results obtained when fitting the models to the ten provinces of Canada, we show that a pragmatic method of estimation is a simple univariate approach that models the 'poststratified' CLFS estimate y_p: it is "good" in terms of model bias and (model-based) variance and from the perspective of production, it is rather straightforward in its implementation. We also summarize our recommendations.

Section 4 describes a method for the estimation of the current number of households by size and the issues associated with internal consistency of domain estimators: calibration of the sum of the separate domain estimates to the model-based estimate of the aggregate (number of households of all sizes) implies that the error resulting from the (model-based) estimation has to be appropriately expressed; optimal and sub-optimal (binding) methods of calibration are presented. The separate domain estimators are not as reliable as the aggregate estimators but the optimal method of calibration proposed yields domain estimators that have lower variability. Empirical results verify that the optimal method yield internally consistent and more efficient estimators than the estimators before calibration.

•2.   Considerations for Modelling

We follow the approach of Pfeffermann and Bleuer (1993) who look at time series models for survey data as a combination of two distinct models: the model describing the evolution of the characteristics of interest, and the survey errors model representing the time series relationships between the errors of the survey estimators.

The responses considered here are the bivariate series $(y_t, x_t - X_t)'$ and the univariate series of the final weight estimators y_{ft}. At previous census times τ, the census values c_τ coincide with the number of households Y_τ; hence not only Y_τ but also the survey errors ϵ_τ in estimating households are known:

$$Y_\tau = c_\tau, \; \epsilon_\tau = y_\tau - Y_\tau = y_\tau - c_\tau \tag{2.1}$$

Figure 2.1 below shows the series of estimates of the current number of households in the provinces of Prince Edward Island (P.E.I.), Quebec, Saskatchewan and British Columbia (B.C.) from March 1985 to May 1991.

[Insert Figure 2.1]

We observe an increasing trend and a variability that can be due to sampling errors and perhaps to seasonal movements. This is typical of most provinces. Thus, the model postulated for the population parameters Y_t is the following structural model: Y_t is described as the sum of a trend component, a seasonal effect and an irregular factor:

$$Y_t = L_t + S_t + I_t \tag{2.2}$$

where L_t is the trend level, assumed to be locally linear, S_t is the seasonal effect and I_t is the irregular term, assumed to be white noise with zero mean and variance σ_I^2; I_t represents the sum of errors left over in the population value Y_t after accounting for trend and seasonal effects. The trend level and seasonal effects are also allowed to vary stochastically with time:

$$L_t = L_{t-1} + R_t$$

$$R_t = R_{t-1} + \eta_{Rt}$$

$$S_t = -S_{t-1} - S_{t-2} - \cdots S_{t-11} + \eta_{St}$$

where R_t is the stochastic slope of the trend L_t and $\{\eta_{Rt}\}$ and $\{\eta_{St}\}$ are assumed to be white noise with means zero and respective variances σ_R^2 and σ_S^2.

- 5 -

Structural models are simple and flexible and have performed well when fitted to numerous empirical series. Moreover, this model was used successfully when analyzing data from labour force surveys by Pfeffermann (1991), and Pfeffermann and Bleuer (1993), among others.

Let us assume for now that $x_t - X_t$ represents the survey error in the CLFS estimation of the population. This error contains the sampling error, which exists by design, and the non-sampling error, which is difficult to control.

In order to understand the nature of the non-sampling error, the CLFS looks at the slippage (that is, the relative error in the estimation of the population with respect to the postcensal estimates, $100 \times (x_t - X_t)/X_t$). The observed monthly sample slippages in the last few years show that the estimates are subject to large negative bias or undercoverage: Figure 2.2 depicts the slippages from March 1985 to May 1991 for the provinces of P.E.I., Quebec, Saskatchewan and B.C.

[Insert Figure 2.2]

The CLFS has historically undercovered its target population by four to six percent relative to the Canadian Census (Clark, Kennedy and Wysocki, 1993). The slippages depicted in Figure 2.2 are larger because they are relative to the Postcensal Estimates of Population X_t, which are adjusted for net undercoverage rate. This coverage error is induced by many factors (missing households, missing persons within a household, misclassification of vacant dwellings, etc.) and is not constant; for example, seasonal and census effects may be important in its make-up; also in a recent study Clark et al. (1993) concluded that most of the large fluctuations in the slippage over time are due to the rotation of clusters that have experienced large growth, thus suggesting the presence of random variation in the bias over time.

Hence it is reasonable to suspect that the subweighted estimators of any characteristic of the CLFS are subject to significant bias due to undercoverage.

Indeed, when we compare the subweighted and final weight estimators of the household counts with previous census values, we observe that the subweighted estimators y_t are negatively biased and that the final weight estimators y_{ft} are positively biased though this bias is considerably smaller.

Therefore this bias must be accounted for when modelling any direct survey estimator. However, the components of the survey error process and their interactions are difficult to understand and model. Since five sixths of the households remain in the sample from one month

- 6 -

to the next, the series of survey errors from the CLFS over time are usually assumed to follow a stationary process or a process of the form $\epsilon_t = k_t * e_t$ where k_t is a constant proportional to the level of the series and e_t is stationary (see for example Binder and Dick (1989) or Pfeffermann and Bleuer (1993)). However, the design-based standard errors of the survey estimators of the number of households, which may be used in the calculation of k_t, are not available. Hence we will assume that the survey error in the CLFS estimation of the population can be expressed as the sum of 2 uncorrelated processes:

$$x_t - X_t = c_t + d_t \tag{2.3}$$

where d_t is stationary and c_t is non-stationary.

Similarly, let $\{\epsilon_t\}$ denote the series of survey errors in the CLFS estimation of the number of households at month t, whether the corresponding direct estimator is y_t or y_{ft}; we assume that

$$\epsilon_t = b_t + e_t, \tag{2.4}$$

where, as above, e_t is a stationary process and b_t is non-stationary.

Thus it is reasonable to assume that the processes d_t and e_t have a stationary and invertible autoregressive moving-average (ARMA) representation:

$$d_t, e_t \sim ARMA(p, q). \tag{2.5}$$

A random walk model for the processes c_t and b_t could account for large variations only a few months after the events but it is the simplest model we can impose with our present knowledge. Therefore we assume

$$c_t = c_{t-1} + \eta_{ct}, \text{ and } b_t = b_{t-1} + \eta_{bt} \tag{2.6}$$

where $\{\eta_{ct}\}, \{\eta_{bt}\}$ are white noise with mean zero and variance σ_c^2 and σ_b^2 respectively. Equations (2.3) and (2.4) do not imply that d_t and e_t contain only the sampling errors; $\{c_t\}$ and $\{b_t\}$ will be referred to as bias processes and represent only part of the non-sampling errors.

We should remark that usually, we do not know the initial value for the trend level component of the model L_0, and hence it is confounded with the initial bias b_o. However the model becomes identifiable at the first census time τ, when the constraint (2.1) is incorporated in the model.

•3. Models for the Current Number of Households: Empirical Results

Our aim is to reduce the bias and the variability present in the direct survey estimators by means of a model that is simple and consistent throughout the provinces and perhaps also throughout subprovincial regions. The class of models proposed can be expressed in a compact state space configuration and the estimators are obtained via the Kalman Filter Algorithms. The model parameters are assumed unknown and are estimated from the data by the method of maximum likelihood (see Appendix I).

If the model bias is reasonably small, we could estimate the variability of the resulting estimators by the model-based estimates of the mean square errors obtained from the Kalman Filter algorithms. For the purpose of comparison we will use the series of coefficients of variation, (the c.v. is defined by the square root of the estimated mean square error divided by the corresponding estimate of the number of households). The error of closure is defined by $100 * (\hat{Y}_\tau - c_\tau)/c_\tau$ and will give an indication of the bias incurred in the estimation by assessing the overall error at the time point of June 1991.

We used the series of estimates from Saskatchewan to illustrate our search for an appropriate model for this problem.

Figure 3.1 shows the series of subweighted and poststratified CLFS estimates of the number of households in the province of Saskatchewan from June 1986 to October 1992. The horizontal lines represent the respective census values in 1986 and 1991.

[Insert Figure 3.1]

Postratification incorporates the auxiliary information contained in $x_t - X_t$ and not only reduces the bias but the variability as well. We want to improve upon the poststratified estimates y_{ft} by borrowing strength across time and by accounting for the left over bias. The question remains if one could obtain better results by using the auxiliary information at time t and the extra information across time simultaneously, for example by directly modelling the bivariate series $(y_t, x_t - X_t)'$.

In Bleuer and Declos (1995) we showed that simple models describing the association between the survey errors ϵ_t and $x_t - X_t$ yield poor estimators. Indeed, Table 3.1 lists measures of model bias and variability (maximum coefficient of variation over the period of analysis) resulting from fitting four models to the series of estimates from Saskatchewan. In all four

models the process Y_t of the number of households at month t follows the structural model and the census constraint defined in the previous section. The first 3 models are nested; the bivariate models 1 and 2 assume a non-zero constant correlation between the stationary components of the survey errors ϵ_t and $x_t - X_t$, but model 2 relaxes this condition to a shorter period of time:

$$d_t = \theta d_{t-1} + u_t,$$
$$e_t = \varphi e_{t-1} + v_t,$$
$$\rho = corr(u_t, v_t) \text{ and } \omega = corr(\eta_{ct}, \eta_{bt}) \quad \forall \, t,$$

where θ and φ are fixed constants and $\{u_t\}$ and $\{v_t\}$ are assumed white noise with zero mean and variances σ_u^2 and σ_v^2 respectively and $\{\eta_{ct}\}$ and $\{\eta_{bt}\}$ are the residual errors of the random walks $\{c_t\}$ and $\{b_t\}$ respectively.

Model 3 fits the univariate subweighted estimates y_t with no utilization of the auxiliary information ($\rho = \omega = 0$) and model 4 fits the postratified estimate y_{pt}.

Table 3.1: Coefficients of Variation and Errors of Closure by Method of Estimation

Model	Maximum c.v. %	E.O.C. (June 1991) %
1. Bivariate ($\rho \neq 0$, $\omega \neq 0$, *March* $85 \leq t \leq$ *May* 91)	0.50	2.23
2. Bivariate ($\rho \neq 0$, $\omega \neq 0$, *March* $85 \leq t \leq$ *Aug* 1990)	0.75	1.98
3. Univariate Subweighted ($\rho = 0$, $\omega = 0$)	1.10	0.25
4. Univariate, Poststratified (AR(3) model for the survey error)	0.43	1.71

The numbers above show that as the amount of auxiliary information used increases, the model-based variability of the estimators decreases, as we expected; the model bias, however, increases.

Ideally, the correlation between the sampling error processes in the estimation of the number of households and population should be positive and fairly constant, but in practice it is not so. Models 1 and 2 impose a constant correlation between the stationary elements of the survey errors for a fixed period of time and this is a very strong assumption.

The absence of model bias depends on how well the model explains the processes underlying the production of the estimates, that is, the consequences of the design and the

operations; the results above give an indication of how sensitive the estimators are to a deviation from the model.

Any attempt to relax the constant correlation assumption and still use the auxiliary information contained in $x_t - X_t$ would complicate the model and would not necessarily yield a better proposition.

Figure 3.1 shows that the subweighted estimates of the number of households are closer to the census values than the final weight estimates are. This is not true in general but it may explain the low error of closure obtained by the univariate model with the subweighted response.

The univariate models defined by the subweighted and final weight (poststratified) series respectively are approximately equivalent (in terms of bounds for bias^2 + variance) for the province of Saskatchewan. But we can argue that a response that is less variable and biased in general, and that yields a smoother time series, is the better response for an estimation method. And indeed this is shown when we fit the series of subweighted and poststratified estimates of household counts for every province. Not only the coefficients of variation yielded by the final weight model were considerably smaller but the error of closure as well. The errors of closure are comparable to those of the Postcensal Estimates of Population. Table 3.2 below lists the results of fitting model 4 to the ten provincial series.

Table 3.2: Maximum Coefficients of Variation and Errors of Closure

Province	Household Counts		Postcensal Population Estimates
	MCV %	EOC %	EOC %
Newfoundland	0.70	0.89	1.32
PEI	3.00	1.55	2.15
Nova Scotia	0.60	1.26	0.52
New Brunswick	0.40	−0.01	−0.17
Quebec	0.50	0.73	0.10
Ontario	0.50	−0.95	−0.06
Manitoba	0.80	−0.14	0.90
Saskatchewan	0.40	1.71	1.30
Alberta	1.20	0.93	0.59
British Columbia	0.70	0.24	0.22

All of this suggests the adoption of the univariate model that fits the poststratified estimator

- 10 -

for the Postcensal Estimation of the Current Number of Households and assumes a structural model for the number of households and a survey error that contains an autoregressive process and a bias component.

When fitting this model to each of the ten provincial series of survey estimates, we observed some common features that could suggest a simpler method for this type of data.

The trend level in the number of households was approximately (globally) linear in most provinces but the assumption of local linear trend remains important to capture change.

There was little evidence of seasonal effects and when they were present they could be considered as part of the survey errors. This is reasonable: people may move from province to province in a seasonal pattern but households are usually stable; and the undercoverage errors represented in ϵ_t are sometimes suspect of having a seasonal component in their make up. To investigate this theory we fitted a model where the seasonal term was included as a component of the survey error rather than the current number of households (see Bleuer and Declos (1995), Appendix II). The provincial errors of closure (in %) at June 1991 remained unchanged up to 2 decimals. Hence whether the seasonal effects were components of the parameter of interest or of the survey errors, they seemed to be negligible.

The model assumes that the direct survey estimator is affected by bias. We then look at the bias ratios (bias of the survey estimator over the number of households), from 1986 to June 1991. The estimated bias ratios yielded by the model were relatively small and varied from a maximum (over the months) of 1% in Nova Scotia to 4% in Manitoba. However we aimed at a much higher accuracy than 4% (the past errors of closure corresponding to the Postcensal Estimates of Population were 2% and under). By accounting for this bias we could obtain errors of closure ranging from 0.1% to 1.71% in June 1991 for the proposed estimates of the provincial number of households.

The variance of the bias process was relatively small compared with the level of the number of households in all provinces, while the variance of the stationary process was quite large. A significant variance component in the non-stationary part (assumed to be a random walk) of the survey error would imply that the variation in the non-sampling error increased with time. The resulting estimates of variability meant that if there was variation, it has been under control as part of the stationary process (this is compatible with the current wisdom on the Canadian Labour Force Survey).

The incorporation of the random walk term in the survey error has played a role in the

validation of the model, as a kind of verification that the model does indeed describe a survey process under control. But we may set the variance of the bias process to zero in a further simplification of the model as long as we are confident that the non-sampling errors remain stable.

Finally, the estimate to be published at month t should be based on all of the available data until time t, including the previous census counts.

•4. Estimation of the Number of Households by Size and Calibration

4.1 Estimation

Once we decide upon a model and obtain accurate estimators of the current number of households Y_t our next concern is that of obtaining reliable estimators of the number of households by size and by province Y_{it} and such that the sum of the estimates over the sizes coincide with the provincial aggregate $\hat{Y}_t$.

Figure 4.1 is a plot of the CLFS estimates of the number of households of size 1 from March 1985 to June 1991. The horizontal lines represent the respective census values in June 1986 and June 1991. We can observe that the CLFS figure overestimates the number of households of size 1 and that the series of estimates $\{y_{it}\}$ becomes widely variable from sometime in 1989 on.

[Insert Figure 4.1]

The true number of households of size 1 changes more than the number of households of all sizes and the survey error becomes more erratic. Since five sixths of the households of size 1 remains in the sample from one month to the next, the sampling error might depend on the previous error in the same manner than for the aggregate. Now contact with households of size 1 is more difficult and hence the survey error model might be inadequate to explain the monthly variations of this process.

We fitted the model proposed in Section 2 to the series $\{y_{it}\}$ of CLFS final weight estimates of the number of households of size i at month t in Canada[1], for sizes 1, 2 and 3+ (3 or more persons in the household).

The series of the domain estimates $\{y_{it}\}$ are in general more variable and this is reflected in the numbers shown in Table 4.1.

---

[1]   by Canada we mean the union of the 10 provinces since there is no Labour Force data from the territories.

- 13 -

Table 4.1: Diagnostics for sizes 1, 2 and 3+

Size	Maximum Coefficient of Variation %	E.O.C. (1991) %	MB	MARE %
1	3.92	-2.68	324	0.75
2	2.65	0.67	-539	0.52
3+	1.65	0.73	453	0.34

The variability of the resulting estimators and the mean bias (MB) and mean absolute relative bias (MARE, see Appendix I) suggest a fit not as good as we obtained for the provincial estimates of the aggregate number of households. The large error of closure for size 1 reflects the inability of the model to explain the survey error in the months before closure. However for the sake of consistency and simplicity we will adopt this model for the estimation of the household counts by size.

Next we compare the sum of the separate estimators $\sum_i \hat{Y}_{it}$ with the aggregate estimator $\hat{Y}_t$ in terms of the series of estimated coefficients of variation and errors of closure. Figures 4.2 and 4.3 show that the series of estimates corresponding to the sum of the separate domains coincides with the aggregate estimates (relative difference oscillating between -0.2% and 0.1%) but that the variability of the first series substantially larger. This is due to the relative poor fit of the separate models. The aggregate model $\{\hat{Y}_t\}$ series is more reliable and we should choose it for publication. Our next section deals with the problems associated with this situation.

[Insert Figures 4.2 and 4.3]

4.2   Calibration of Domain Estimators

Whenever we produce model-based estimates for different areas or domains we have to deal with the problem of internal consistency between the published estimates for the larger domain and the sum of the estimates of the individual domains within it.

This is a common problem in the application of small area and seasonal adjustment techniques. But the main body of solutions proposed so far lie in the field of small area methods. Among these we have methods proposed by Battese, Harter and Fuller (1988), Pfeffermann and Burck (1990) and Rao and Choudhry (1993).

Since usually the direct survey estimator for a large area is reliable and this is what will eventually be published, the individual small area estimates are constrained to add up to the direct survey estimator of the larger area. In the field of seasonal adjustment the problem is

- 14 -

slightly different: the seasonally adjusted estimators of the large area and the separate small area within the larger one are all model based estimators. So often before we deal with internal consistency of the estimates to be published we have to decide which is the "better" estimator: the one derived from fitting the aggregate survey estimator or the sum of the separate small area estimators. In some cases both estimators are statistically alike and we do not have a problem. However, if the seasonally adjusted estimator of the aggregate is more accurate then we resort to calibration of the separate estimators.

Now, the method we use for the Postcensal Estimation of the Number of Households yields the estimates of the number of households and of the mean square errors of the corresponding estimators. So we can evaluate which estimator is better, the aggregate estimator $\hat{Y}_t$ or the sum of the separate estimators $\sum \hat{Y}_{ti}$, by looking at the coefficients of variations of the estimators at month t and at the error of closure at t = june 1991 (census time).

If we decide for $\hat{Y}_t$ and we have to calibrate there are two issues to be concerned with:

1) to obtain calibrated estimators such that

$$\tilde{Y}_t = \sum_i \tilde{Y}_{ti} \quad \text{(internal consistency)}$$

and $\tilde{Y}_t$ is good in terms of bias and variance.

2) the individual benchmarked estimators $\tilde{Y}_{ti}$ should compare well with the separate model estimators $\hat{Y}_{ti}$.

Within the framework of the proposed model-based estimators, state-space modelling offer a natural extension to calibration. Pfeffermann and Burck (1990) used this technique for small area estimators to add up to the reliable design-based estimator of the large area, thus providing a robust mechanism to guard against model failure.

The method that we propose is similar to that of Pfeffermann and Burck (1990) in that we consider the joint model of the domain estimators but we constrain them to add up to the model-based estimator of the union of domains rather than the design estimator. The expression of this model in state-space form is not straightforward because the model-based estimator for the aggregate is a contrast of the design-based estimators for all the previous time periods up to the current time t inclusive.

With this set up, we derive two different calibration models, a "non-binding" model which yields optimal estimators under the model and a "binding" model, which yields sub-optimal estimators.

- 15 -

The following are the same considerations for both methods. We assume that there are d individual domains which add up to the aggregate domain. The model for each individual domain i, $i = 1, 2, ..., d$, is

$$y_{ti} = Y_{ti} + \epsilon_{ti}, \tag{4.1}$$

where Y_{ti} is the true number of households in domain i at month t, and for every domain i, $\{Y_{ti}\}$ follows a structural model; ϵ_{ti} represents the survey error in the design-based estimator of the number of households in domain i at month t. For every domain i, the survey error is assumed to be the sum of a random walk process and an autoregressive model of order 3. Thus,

$$Y_{ti} = L_{ti} + S_{ti} + I_{ti}, \tag{4.2}$$

where L_{ti}, S_{ti} and I_{ti} follow the same laws established for the trend levels, seasonal components and irregular terms in (2.2).

The survey error model is given by

$$\epsilon_{ti} = b_{ti} + e_{ti} \tag{4.3}$$

where b_{ti} is as in (2.6) and e_{ti} is an autoregressive process of order 3. At census times, the model has the added constraints, for every domain i,

$$c_{\tau i} = Y_{\tau i}. \tag{4.4}$$

The model for the aggregate is the following: let

$$Y_t = Y_{t1} + ... + Y_{td}$$

be the true number of households in the union of all domains at time t; we assume that the sum of the separate trends $L_{t1} + ... + L_{td}$, separate seasonal effects $S_{t1} + ... + S_{td}$ and separate irregular factors $I_{t1} + ... + I_{td}$ represent the corresponding structural components for the aggregate process Y_t; and at census times we assume

$$c_\tau = Y_\tau, \tag{4.5}$$

where

$$c_\tau = c_{\tau 1} + ... + c_{\tau d}. \tag{4.6}$$

The aggregate survey estimate will be denoted by $y_{t\,agg}$ and we do not necessarily assume that $y_{t\,agg}$ coincides with the sum of the individual domain survey estimates. As with the separate domains we assume

$$y_{t \, agg} = Y_t + b_t + e_t \tag{4.7}$$

Let the state-space model for the individual domains be expressed as

$$y_{ti} = Z_{ti} \, \alpha_{ti}$$
$$\alpha_{ti} = T^i \, \alpha_{t-1i} + \eta_{ti} \tag{4.8}$$

with

$$Q_i = E(\eta_{ti} \cdot \eta'_{ti}), \quad i = 1, 2, \dots, d .$$

The state-space model for the aggregate estimator is

$$y_{t \, agg} = Z_t \, \gamma_t$$
$$\gamma_t = T^o \, \gamma_{t-1} + \eta_{to} \tag{4.9}$$

with

$$Q_o = E(\eta_{to} \cdot \eta'_{to}).$$

Note: for the estimation of the number of households, Z_{ti} and Z_{to} are $1 \times p$ vectors at regular months t, and $2 \times p$ matrices at census times τ, when the measurement vector is $y_{\tau i} = (y_{\tau i}, c_{\tau i})'$, where $y_{\tau i}$ denotes the direct survey estimator of the number of households in domain i at month τ and $c_{\tau i}$ denotes the census number of households in domain i at month τ.

We propose a joint model consisting in the separate domain models plus a calibration equation:

$$y_{ti} = Y_{ti} + \epsilon_{ti} \tag{4.10}$$

$i = 1, 2, \dots, d$, and either

$$\hat{Y}_t = Y_{t1} + Y_{t2} + \dots + Y_{td} + \delta_t \tag{4.11}$$

or

$$\hat{Y}_t = Y_{t1} + Y_{t2} + \dots + Y_{td}, \tag{4.12}$$

where $\hat{Y}_t$ is the model-based estimator obtained from fitting (4.5) and (4.6) given the data $D_t = (y_1, y_2, \dots, y_t)$, and δ_t represents the error in the model-based estimation of the true

number of households at time t. At census times τ, the usual benchmarking equations (4.4) and (4.5) are satisfied.

The model given by (4.10) and (4.11) yields optimal estimators and is non-binding; the model determined by (4.10) and (4.12) is binding and yields suboptimal estimators.

We should note that (4.10) assumes zero correlation between the different domain components. This assumption does not necessarily hold. For example, given an approximate stable population, as the number of households of size larger than one decreases, the number of households of size 1 may increase. But when we have more than two domains (e.g. sizes 1, 2 and 3 or more), the correlation structure could be intractable and thus is simpler to assume independence of domain estimators.

In general, a linear constraint like equations (4.11) or (4.12) derived from adding up the information contained in the separate models, does not provide more information; but in our case equations (4.11) and (4.12) induce a correlation between the components of the separate domains and thus establishes effectively a "joint model" with implicit extra information about the relationships between domains. Hence the variability of the resulting estimators should decrease.

In order to express the joint models in state-space form we let

$$\hat{\gamma}_t = E(\gamma_t | D_t)$$

under (4.5) and (4.6), that is, $\hat{\gamma}_t$ is the predictor of the state-space vector γ_t given the data D_t under the aggregate model. Hence,

$$\hat{\gamma}_t - \gamma_t = U_{to} T^o(\hat{\gamma}_{t-1} - \gamma_{t-1}) - U_{to}\eta_{to} \qquad (4.13)$$

where

$$U_{to} = I(p) - K_{to} Z_{to}$$

and

$$K_{to} = P_{t/t-1}^o Z_{to}' F_{to}^{-1}$$

is the Kalman Gain matrix for the aggregate model. If h is the $1\times p$ vector of zeros and ones defined such that $Y_{st} = h\alpha_{st}$, where α_{st} is the $p\times p$ state-vector corresponding to the model for

domain i, then

$$\hat{Y}_{it} = h \, \hat{\alpha}_{it} \tag{4.14}$$

is the model-based estimator of Y_{it}, with $\hat{\alpha}_{it} = E(\alpha_{it}|(y_{i1}, ..., y_{it}))$. Similarly, the model-based estimator of Y_t under the aggregate model is given by

$$\hat{Y}_t = h \, \hat{\gamma}_t. \tag{4.15}$$

The state-vector for the joint models (4.10) and (4.11) or (4.10) and (4.12) is defined as the $(d+1)\, p \times 1$ vector composed by the d state vectors α_{it} corresponding to the separate models plus the vector $\hat{\gamma}_t - \gamma_t$:

$$\alpha_t = (\alpha'_{1t}, ... , \alpha'_{dt}, \hat{\gamma}'_t - \gamma'_t)'. \tag{4.16}$$

The transition matrix T_t is the $(d+1)p \times (d+1)p$ time-varying matrix defined by the separate transition matrices T^i in the block diagonals and the matrix $U_{to} \, T^o$ in the last $p \times p$ elements. The transition error η_t is defined by the $(dp+2) \times 1$ vector including the d transition errors η_{it} plus the transition errors of the aggregate model that correspond only to the survey error:

$$\eta_t = (\eta'_{t1}, ..., \eta'_{td}, \eta_{bto}, \eta_{eto})';$$

and R_t is an $(d+1)p \times (dp+2)$ matrix such that $R_t \, \eta_t \, [dp+1: dp+d] = -U_{to} \, \eta_{to}$ and $Q = E[\eta_t \cdot \eta'_t]$ is assumed diagonal.

The transition equation is the same for both models:

$$\alpha_t = T_t \alpha_{t-1} + R_t \, \eta_t \tag{4.17}$$

for every month t.

4.3   Optimal Method

The non-binding model (4.10) and (4.11) differ from the binding model in its measurement equation:

$$y_t = Z_t \cdot \alpha_t \tag{4.18}$$

where

$$y_t = (y_{t1}, ..., y_{td}, \hat{Y}_t)' \tag{4.19}$$

and Z_t is a $(d+1) \times (d+1)p$ matrix with the matrixes Z_{it} in the block diagonals, $i=1, 2, ..., d$ and the last row, the $1 \times (d+1)p$ vector defined by

- 19 -

$$h_{ab} = (h, h, ..., h)$$

where h was defined in (4.14). Let $\mathcal{D}_t$ denote the joint data set until time t, i.e., $\mathcal{D}_t = D_t \bigcup \{y_{1t}, ..., y_{dt}, i=1, ..., i=1, ..., d\}$. Equations (4.17) and (4.18) define a state-space model and the Kalman-Filter applied to this model will yield

$$\hat{\alpha}_t = E(\alpha_t / \mathcal{D}_t)$$

that are Empirical Best Linear Unbiased Predictors (EBLUP's) under the model.

4.4   Sub-optimal method

The binding model defined by (4.10) and (4.12) can be expressed in state-space form and the measurement equation is written as

$$y_t = Z_p \alpha_t, \tag{4.20}$$

where y_t is as in (4.19) and Z_p coincides with Z_t except in the last p elements of the last row, which are set to zero. The difference between the two measurement matrices will be denoted by Z_D and it is a $(d+1) \times (d+1)p$ matrix whose elements are all zeros except in the last row, the $1 \times (d+1)p$ vector defined by

$$h_D = (0, 0, ..., 0, h),$$

where h is the $1 \times p$ vector defined in (4.14) and $\mathbf{0}$ represents the $1 \times p$ zero vector. We should remark that even though the subscript t is not explicit in this measurement matrix, Z_p also depends on time, being constant for regular months t and at census times τ it coincides with Z_τ.

In order to fit the binding model defined by (4.20) and obtain unbiased predictors and correct estimators of their mean square errors we have to modify the Kalman Filter. The modified equations are described in Appendix II.

4.5   Empirical Results

We fitted the joint series of separate domain estimates using the optimal method. Table 4.2 lists the range of coefficients of variation and errors of closure for each size.

Table 4.2 Comparison between the separate and joint models

Size	MCV % (separate)	MCV % (joint)	eoc % (separate)	eoc % (joint)
1	3.92	3.00	-2.68	-2.73
2	2.65	2.20	0.67	0.51
3⁺	1.65	1.40	0.73	0.70

The estimates of the household counts by size obtained from the joint model are almost the same as those produced by the separate models but the coefficients of variation are much lower for the joint model. This indicates that in our problem calibration is advantageous not only for the sake of internal consistency and as a robust constraint that protects against model failure but for improved efficiency of the resulting estimators.

The sum of the separate domain estimates $\sum_i \tilde{Y}_{i} = \tilde{Y}_{t}$, produced by the joint model is not exactly equal to the aggregate estimate $\hat{Y}_{t}$, since the optimal method is not binding. However the difference between the series $\{\tilde{Y}_{t}\}$ and $\{\hat{Y}_{t}\}$ is negligible and its respective series of coefficients of variation are also similar. Figure 4.4 is a plot of the estimated coefficients of variation for the three models yielding estimates $\sum_i \hat{Y}_{i}$, $\hat{Y}_{t}$ and $\tilde{Y}_{t}$.

A binding model would produce domain estimates that add up to the aggregate: $\sum_i \tilde{Y}_{i} = \hat{Y}_{t}$, and this would be convenient. However we seek accurate estimators and the binding condition introduces more variability in the resulting estimators. The formulas in Appendix II show the extra variance terms associated with the corresponding predictors. Hence the use of the suboptimal method is not advised.

[Insert Figure 4.4]

- 21 -

Appendix I: State Space Representation and Estimation

I.1   The State Space Model and the regular Kalman Filter

The class of models given by (1.1) and (2.1) to (2.6), can be expressed in a compact state space model, similarly to the model representation in Pfeffermann (1991).

The series of observations will consist in vectors y_t of time varying dimension r_t; r_t will usually double its size at census times when we add the constraining equations. The measurement matrix will therefore be of time varying dimensions, $r_t \times m$ where m is the dimension of the state vector and in all applications we will assume, without loss of generality, that the measurement errors are zero: any errors expressed in the model will be included in the state vector even if they are uncorrelated over time, as done with the irregular terms I_t. The bias b_t will also be included as part of the state vector.

The state space representation of the models enables us to update, smooth or predict the state vectors at any given time t by means of the Kalman Filter. By "updating" we mean the prediction of the state vector at time t, based on all the data until time t. "Smoothing" refers to the prediction of the state vector based on all the available data before and after time t. "Prediction" applies to postsample periods.

We will refer to Harvey (1989) for the recursive Kalman Filter equations. The components of the state vector, the measurement matrix and the transition matrix are described with detail in Bleuer and Declos (1995), Appendix I.

I.2   Estimation of the model parameters and adjustments to the covariance matrix of the estimation error

The coefficients of the transition matrix, the measurement matrices and the variance-covariance matrix are usually unknown and have to be estimated from the available data. Assuming that the error terms in the model for $\{Y_t\}$ and the survey errors $\{\epsilon_t\}$ have a normal distribution, the unknown model parameters are estimated by the method of maximum likelihood.

Once the unknown model parameters have been estimated, the Kalman Filter is applied with the true parameter values replaced by the parameter estimates. Then the covariance matrices of the estimation error produced by the Kalman Filter underestimate the true variances because they ignore the extra variation implied by the parameter estimation.

In order to correct for this under estimation, a modification of the Monte Carlo method

developed by Hamilton (1986) could be used. Recently Singh, Stukel and Pfeffermann (1993) pointed out that there is an important term missing from the Hamilton (1986) approximation. The corrections based on Singh *et al.* (1993) should be considered.

I.3   Measures to Evaluate the Model Performance

Once the hyperparameters are estimated and the model is defined we fit the model to the different data series. The goodness of fit of the models is assessed with plots and diagnostic statistics based on the innovations or one-step ahead prediction errors in predicting the survey estimator y_t at time t. Among them, the mean bias (MB) refers to the average of the innovations and the mean absolute relative (MARE) refers to the mean absolute relative error in predicting the survey estimators. For more details see Bleuer and Declos (1995), Appendix I.

Appendix II: Modified Kalman Filter Algorithm to fit the pseudo model (4.20)

In an argument similar to that of Pfefferman and Burck (1990), let α_{t-1}^P denote the constrained predictor of α_{t-1} as obtained at time $t-1$ using the pseudo model and let P_{t-1} denote the variance-covariance matrix of the errors $\alpha_{t-1}^P - \alpha_{t-1}$.

We define $\alpha_{t|t-1}^P = T_t \, \alpha_{t-1}^P$. Hence the unconditional expectation of $\alpha_{t|t-1}^P - \alpha_t$ is zero:

$$E(\alpha_{t|t-1}^P - \alpha_t) = E(T_t \, \alpha_{t-1}^P - T_t \, \alpha_{t-1} - R_t \, \eta_t)$$

$$= T_t E(\alpha_{t-1}^P - \alpha_{t-1}) - R_t E(\eta_t)$$

$$= 0,$$

by the definition of α_{t-1}^P and the model assumptions on η_t. The unconditional M.S.E. of $\alpha_{t|t-1}^P - \alpha_t$ is given by

$$P_{t|t-1} = E\{(\alpha_{t|t-1}^P - \alpha_t)(\alpha_{t|t-1}^P - \alpha_t)'\}$$

$$= T_t P_{t-1} T_t' + R_t Q R_t'.$$

Now consider the innovation

$$e_t^* = y_t - Z_p \, \alpha_{t|t-1}^P.$$

Given that the "true" model (4.18) holds rather than (4.20), we have

$$e_t = Z_t \, \alpha_t - Z_p \, \alpha_{t|t-1}^P$$

$$= Z_p(\alpha_t - \alpha_{t|t-1}^P) + Z_D \, \alpha_t \tag{II.1}$$

Where $Z_D = Z_t - Z_p$ as defined above. Hence the unconditional expectation of the innovation is zero:

$$E(e_t) = E(Z_D \, \alpha_t) = (0, E[h(\hat{\gamma}_t - \gamma_t)]) = 0,$$

since by definition $\hat{\gamma}_t$ is the EBLUP of γ_t under the aggregate model.

We write

$$\alpha_t = \alpha_{t|t-1}^P + (\alpha_t - \alpha_{t|t-1}^P)$$

$$y_t = Z_P \, \alpha_{t|t-1}^P + e_t$$

Hence the vector

$$\begin{pmatrix} \alpha_t \\ y_t \end{pmatrix}$$

has mean

$$\begin{pmatrix} \alpha_{t|t-1}^P \\ Z_P \, \alpha_{t|t-1}^P \end{pmatrix}$$

with covariance matrix

$$\begin{pmatrix} P_{t|t-1} & P_{t|t-1} Z_P' + W_t \, Z_D' \\ \\ Z_P \, P_{t|t-1} + Z_D \, W_t' & F_P \end{pmatrix}$$

where F_P is the covariance matrix of the innovation

$$F_P = E(e_t . e_t')$$

$$= Z_P \, P_{t|t-1} Z_P' + Z_D \, W_t' \, Z_P' + Z_P \, W_t \, Z_D' + X_P,$$

with X_P defined as a $(d+1) \times (d+1)$ matrix whose elements are all zeros except the element $(d+1, d+1)$ which is the variance of $\hat{Y}_t$ under (4.18):

$$varagg: \ = E(\hat{Y}_t - Y_t)^2,$$

and $W_t = E[(\alpha_t - \alpha_{t|t-1}^P) \, \alpha_t']$.

Since

$$W_t = P_{t|t-1}$$

we have

$$F_p = Z_t P_{t|t-1} Z_p' + Z_p P_{t|t-1} Z_D' + X_p, \qquad (\text{II.2})$$

and hence the mean of α_t conditional on y_t or updated predictor of α_t is

$$
\begin{aligned}
\alpha_t^P &= \alpha_{t|t-1}^P + P_{t|t-1} Z_t' F_p^{-1} . e_t \\
&= \alpha_{t|t-1}^P + K_p e_t.
\end{aligned}
\qquad (\text{II.3})
$$

The covariance of α_t conditional on y_t or mean square error matrix of is

$$
\begin{aligned}
P_t &= E(\alpha_t^P - \alpha_t)(\alpha_t^P - \alpha_t)' \\
&= P_{t|t-1} - K_p Z_t P_{t|t-1} \\
&= (I - K_p Z_t) P_{t|t-1}.
\end{aligned}
\qquad (\text{II.4})
$$

Thus the method yields unconditionally unbiased estimators but not optimal since the pseudo-model (4.20) is not the correct model; we impose it in order to obtain domain estimators $\hat{Y}_{dt}$ that add up to the original estimator $\hat{Y}_t$ obtained from the aggregate model.

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error-components model for prediction of country crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Binder, David, Dick P. (1989), "Modelling and Estimation for Repeated Surveys", *Survey Methodology*, 15, 29-45.

Bleuer, S.R. and Declos, M. (1995), "Postcensal Estimation of the Number of Households", *Technical Report*, SSMD, Statistics Canada.

Bleuer, S.R. (1993), "Time Series Modelling Using Auxiliary Information With Application To The Postcensal Estimation of Households", *Proceedings of the Social Statistics Section, American Statistical Association*.

Clark, C., Kennedy, B., and Wysocki, M. (1993). "Coverage Error in the Canadian Labour Force Survey", Internal Report, Statistics Canada.

Ghangurde (1991). "Postcensal Estimation of Households", *Internal Report*, Social Survey Methods Division, Statistics Canada.

Hamilton, J.D. (1986), "A Standard Error for the Estimated State Vector of a State-Space Model", *Journal of Econometrics*, 33, 387-397.

Harvey, A.C. (1989). "Forecasting Structural time series models and the Kalman Filter", Cambridge University Press, Cambridge.

Lemaître, G. (1989). "Postcensal Estimation of Households", *Internal Report*, Demography Division, Statistics Canada.

Pfeffermann, D. (1991). "Estimation and seasonal adjustment of population means using data from repeated surveys, *Journal of Business and Economic Statistics*, 163-175.

Pfeffermann, D. and Bleuer, S.R. (1993). "Robust joint modelling of Canadian Labour Force series of small areas, *Survey Methodology*, 19, 149-163.

Pfefferman, D. and Burck, L. (1990). "Robust small area estimation combining time series and cross-sectional data, *Survey Methodology*, 16,217-237.

Rao, J.N.K., and Choudhry, G.H. (1993). Small area estimation: overview and empirical study. *Monograph proceedings of the International Conference on Establishment Surveys, Buffalo, June 1993*.

Singh, A.C., Stukel, D.M. and Pfeffermann, D. (1993). "Bayesian versus frequentist measures of uncertainty for small area estimators", *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Singh, M.P., Drew, J.D., Gambino, J.G., and Mayda, F. (1990). *Methodology of the Canadian Labour Force Survey*. Catalogue No. 71-526, Statistics Canada, Ottawa, Ontario, K1A 0T6.
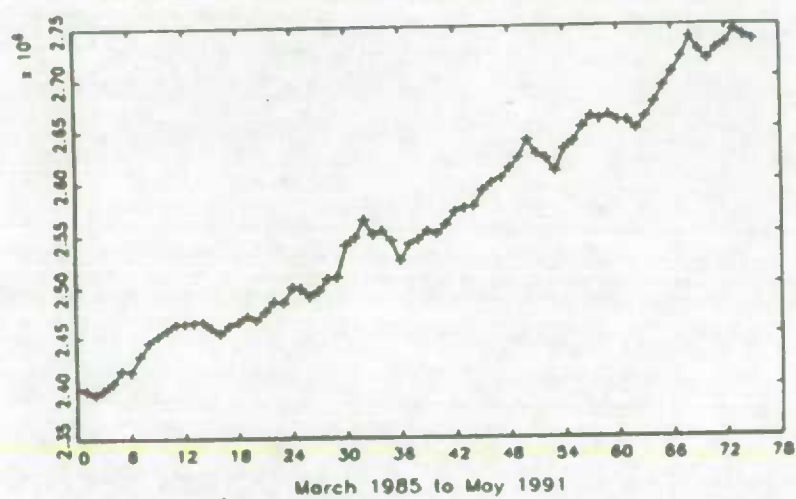
Figure 2.1: CLFS Number of Households, poststratified estimator, P.E.I.

March 1985 to May 1991
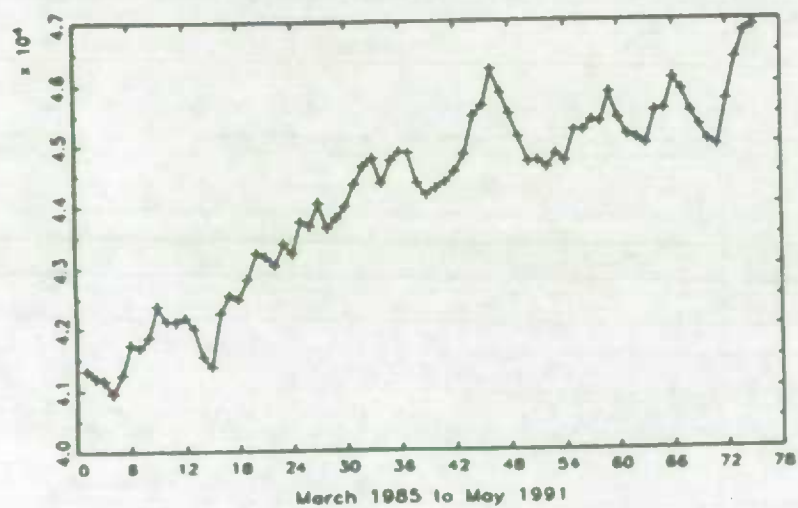
CLFS Number of Households, poststratified estimator, B.C.

March 1985 to May 1991
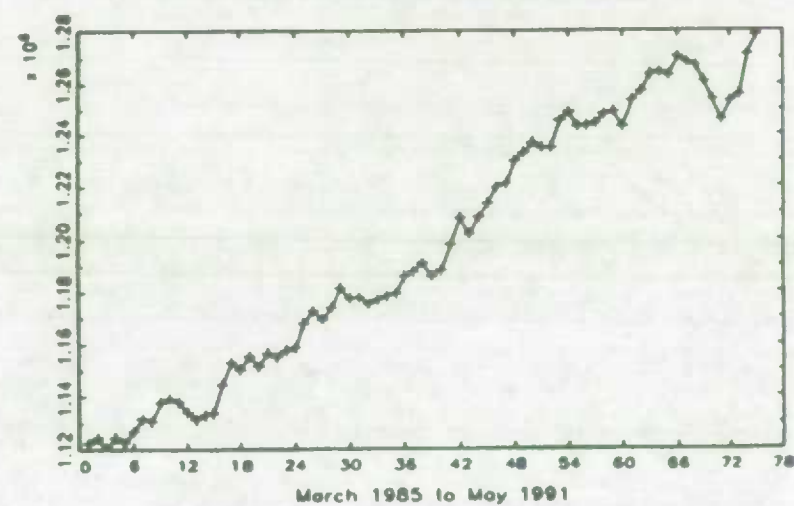
CLFS Number of Households, poststratified estimator, Quebec

March 1985 to May 1991

CLFS Number of Households, poststratified estimator, Saskatchewan
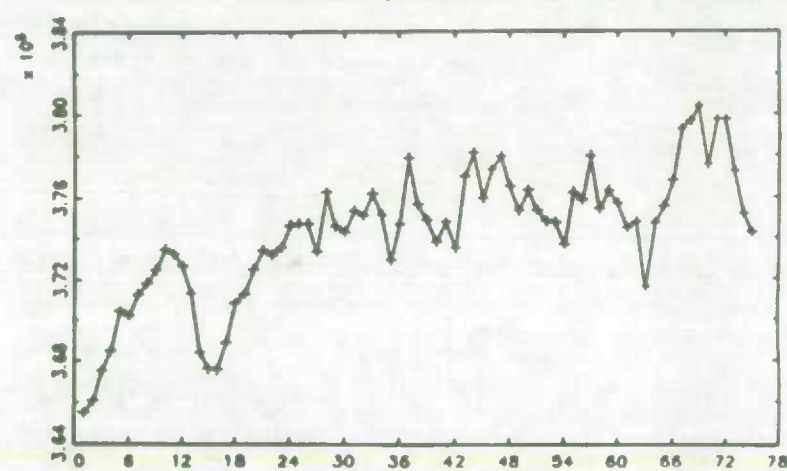
Figure 2.2:Relative Error in the Estimation of the Population (%), P.E.I.
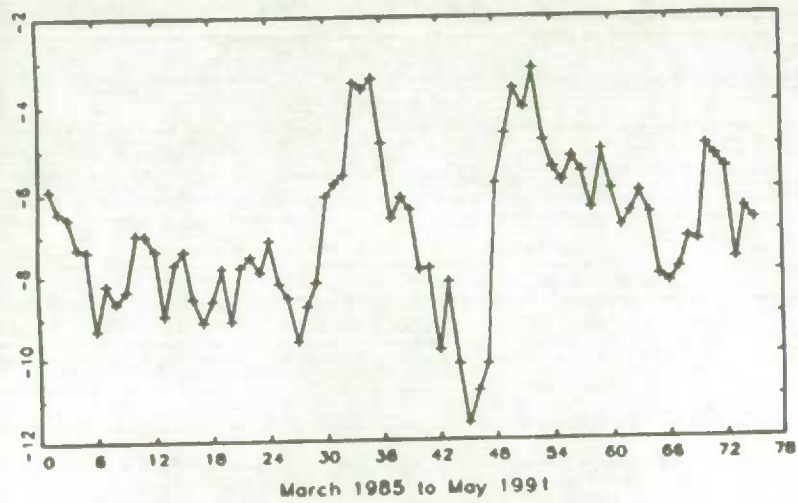
March 1985 to May 1991

Relative Error in the Estimation of the Population (%), B.C.

March 1985 to May 1991

Relative Error in the Estimation of the Population (%), Quebec

March 1985 to May 1991

Relative Error in the Estimation of the Population (%), Saskatchewan

March 1985 to May 1991
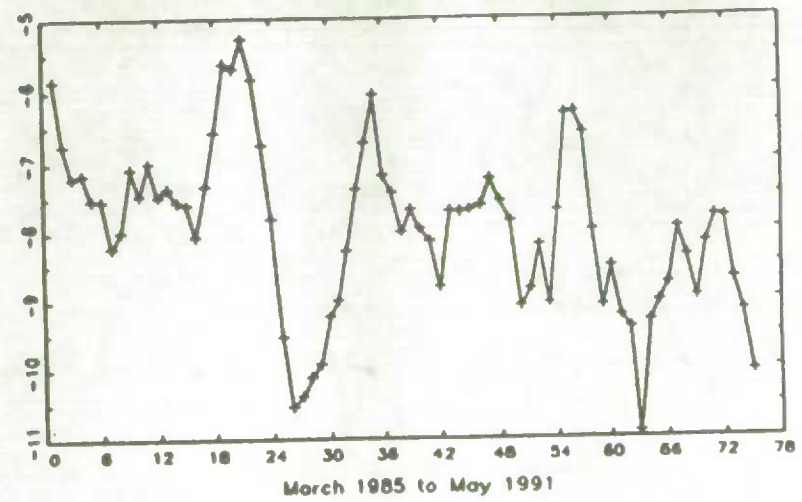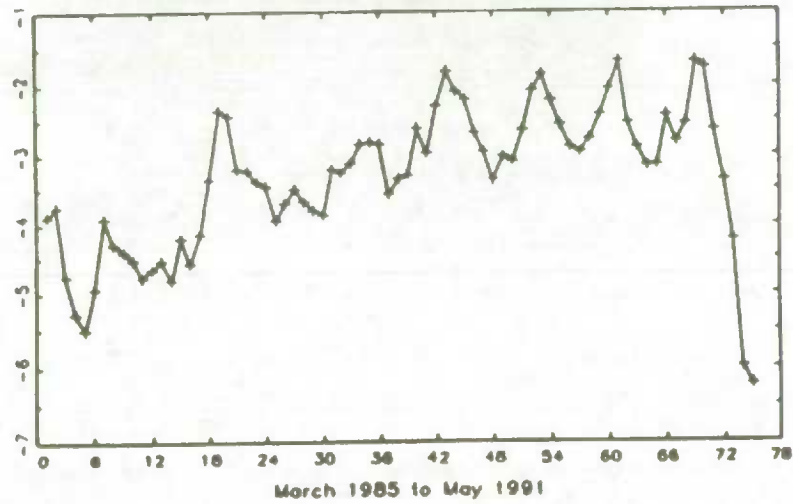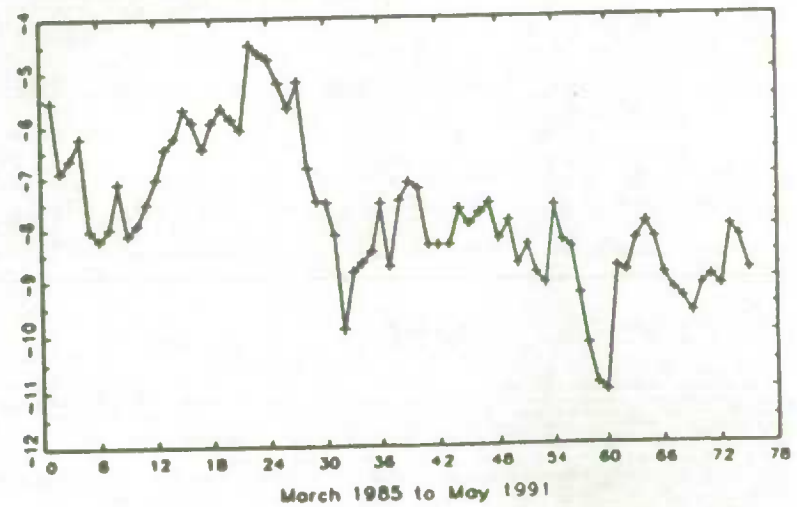
Figure 3.1: LFS Number of Households, subweighted & final, Saskatchewc
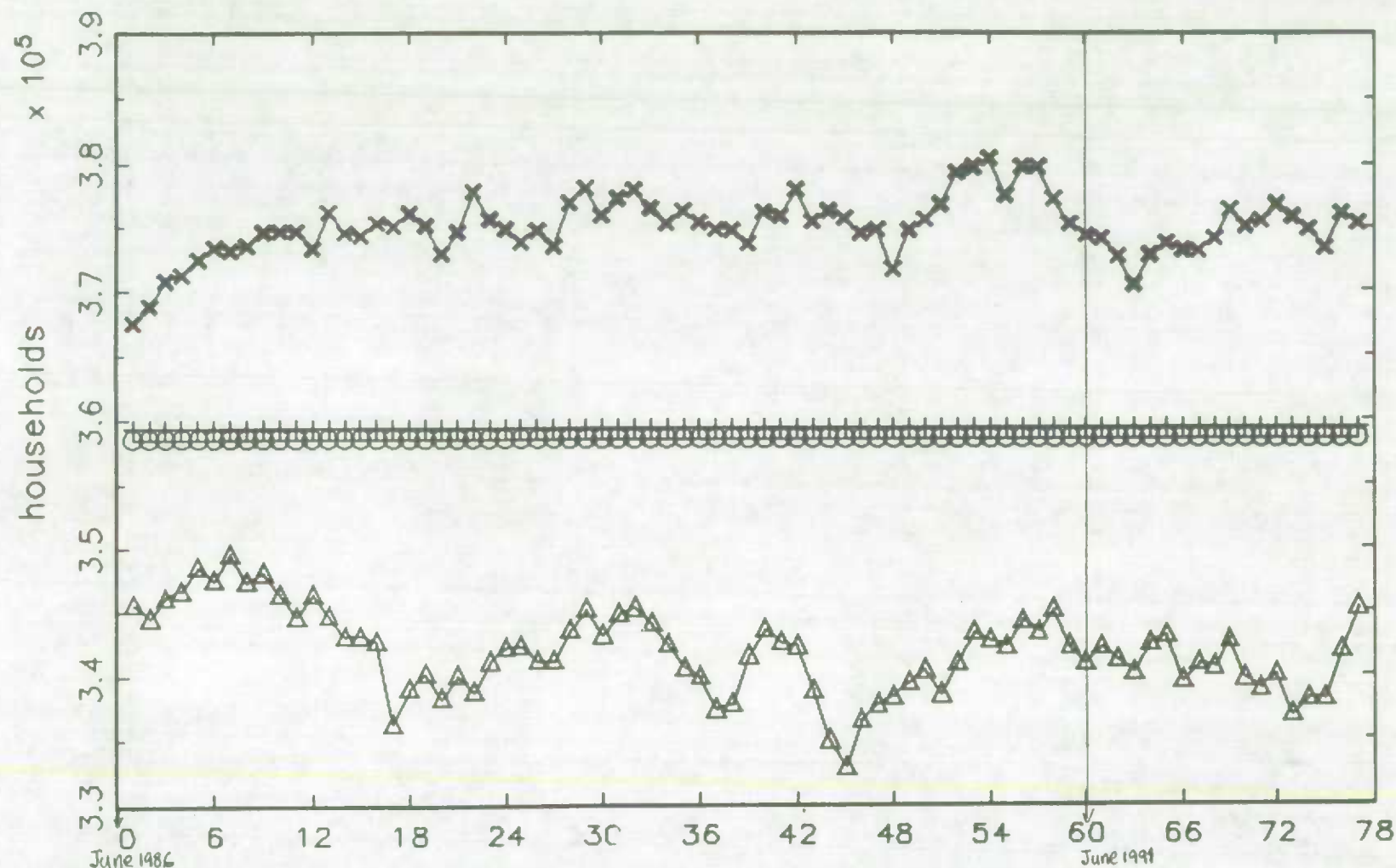
6/1986−10/1992; △subweighted, ✖ final weight, + census 1986, o census 1991

Figure 4.1: LFS Number of Households of Size 1 and Census Counts

June 1991
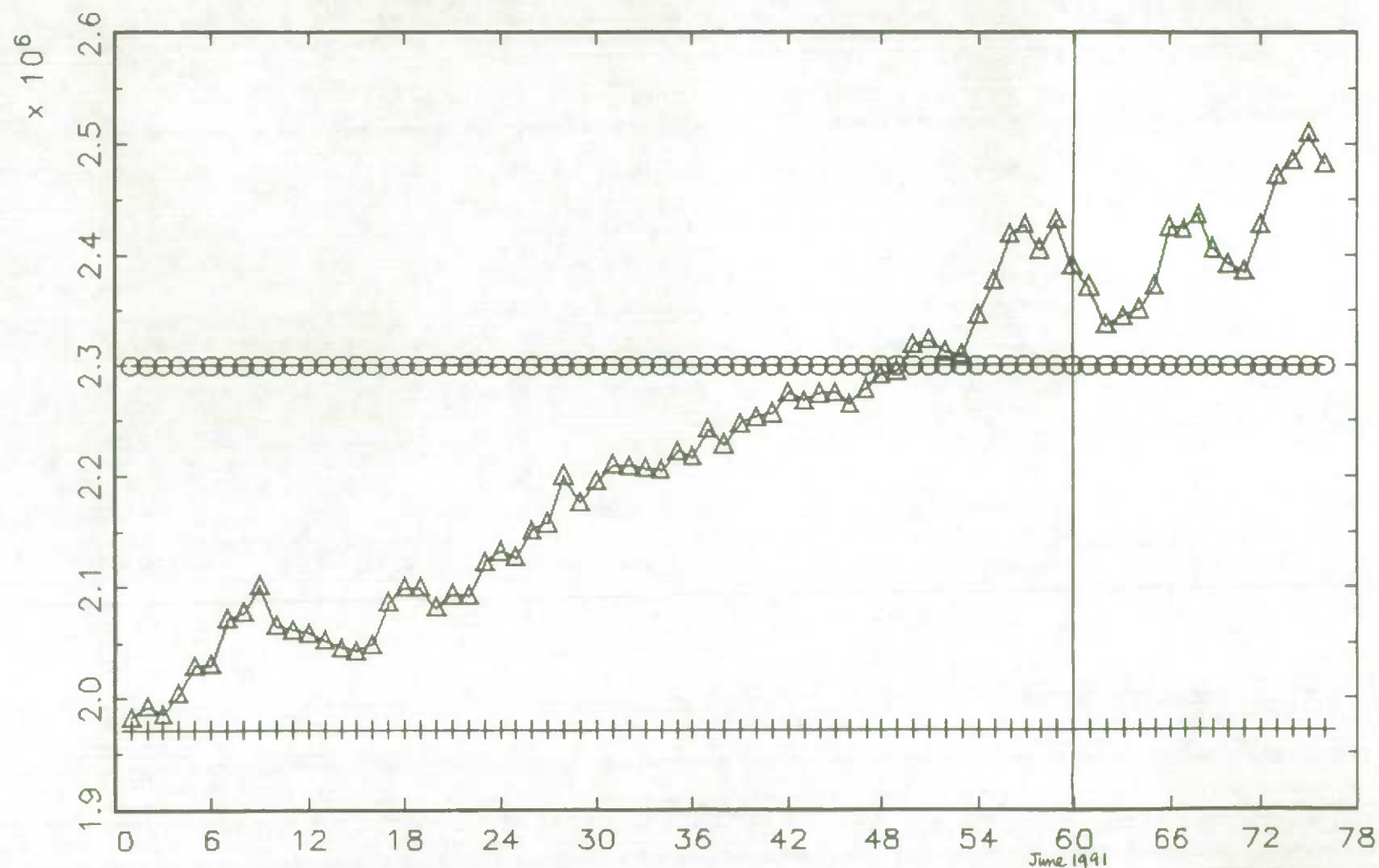
3/1985–6/1991; triangle LFS estimate, + census 1986, o census 1991

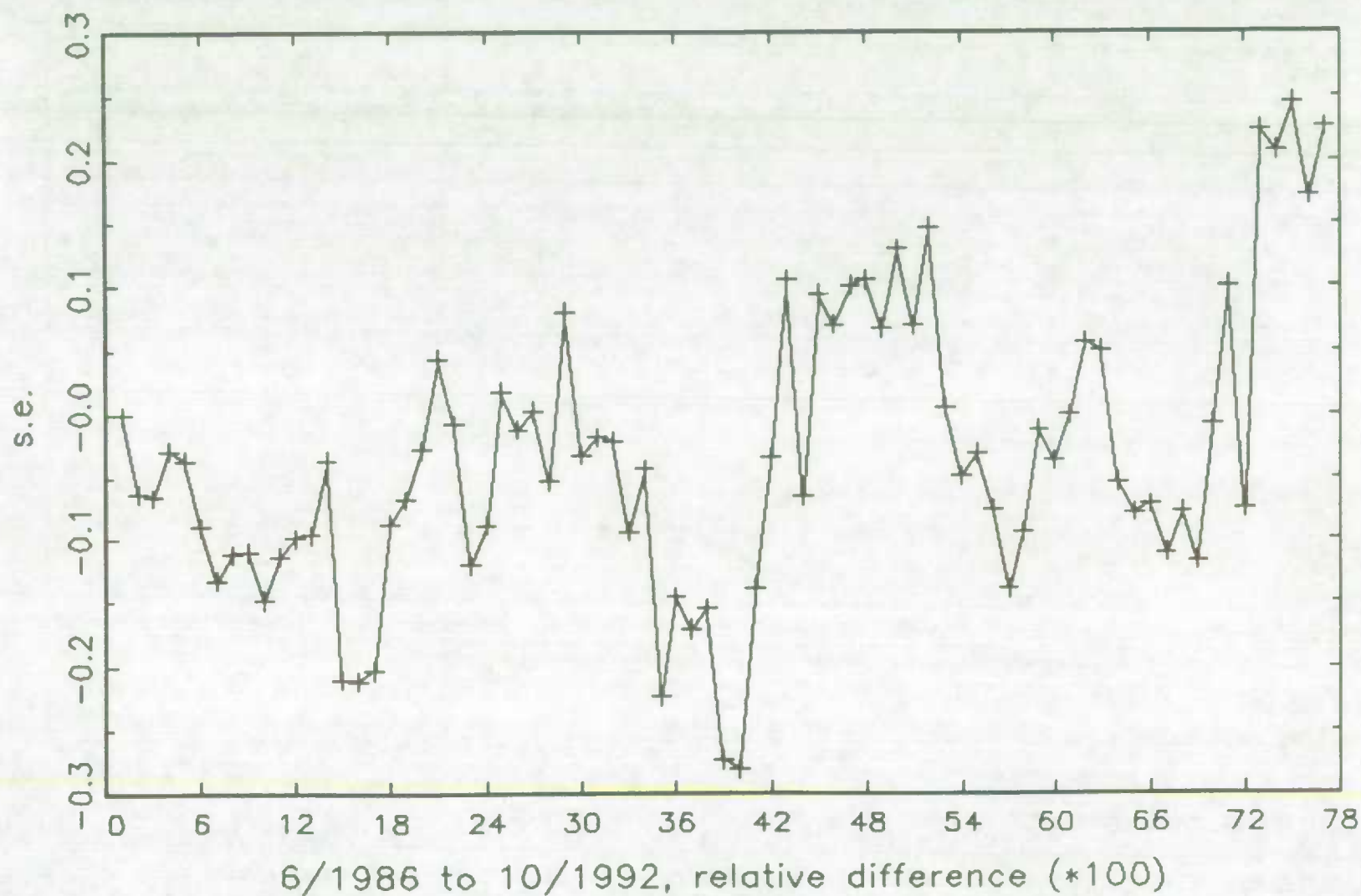Figure 4.2: Relative Difference Between Aggregate & Sum of Separate Estim

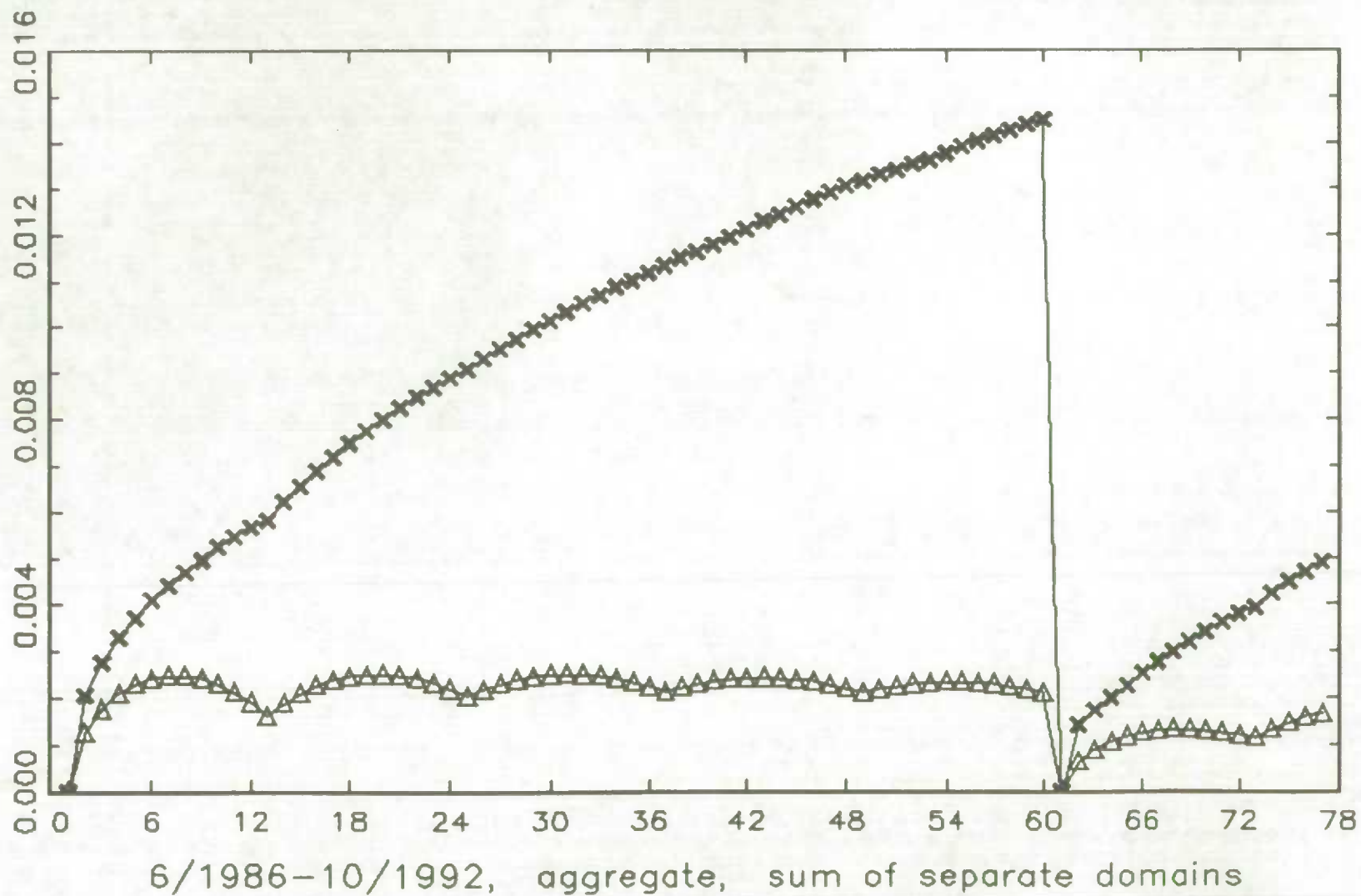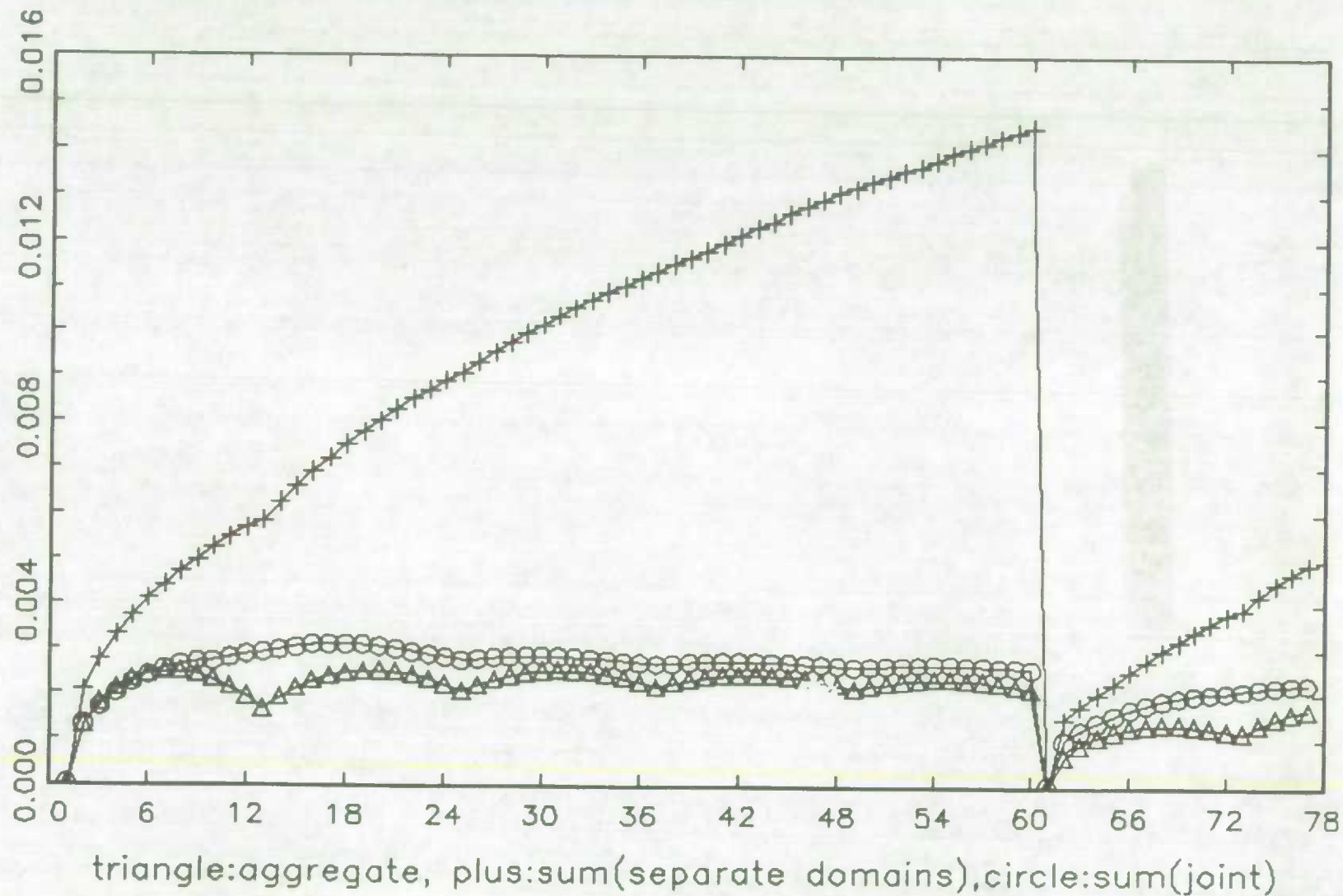6/1986 to 10/1992, relative difference (*100)

Figure 4.3: Coefficients of Variation, Aggregate & Sum of Separate Domains

6/1986-10/1992,   aggregate,   sum of separate domains

C. V.'s, Aggregate of domains, Sum of Separate Domains & Sum of Joint

triangle:aggregate, plus:sum(separate domains),circle:sum(joint)