Survey Methodology, December 1990
Vol. 16, No. 2, pp. 217–237
Statistics Canada

Robust Small Area Estimation Combining Time Series and Cross-Sectional Data

D. PFEFFERMANN and L. BURCK[1]

ABSTRACT

The common approach to small area estimation is to exploit the cross-sectional relationships of the data in an attempt to borrow information from one small area to assist in the estimation in others. However, in the case of repeated surveys, further gains in efficiency can be secured by modelling the time series properties of the data as well. We illustrate the idea by considering regression models with time varying, cross-sectionally correlated coefficients. The use of past relationships to estimate current means raises the question of how to protect against model breakdowns. We propose a modification which guarantees that the model dependent predictors of aggregates of the small area means coincide with the corresponding survey estimators and we explore the statistical properties of the modification. The proposed procedure is applied to data on home sale prices used for the computation of housing price indexes.

KEY WORDS: Kalman filter; Linear constraints; State-space models.

1. INTRODUCTION

Statistical Bureaus are often confronted with the demand to provide reliable estimators for small area means. The problem with the production of such estimators is that the sample sizes within those areas are usually too small to allow the use of direct survey estimators. As a result, new estimators have been proposed in recent years which combine auxiliary information (obtained from a census or administrative records) with the survey data obtained from all the small areas. The common feature of these estimators is that they can be structured in general as a linear combination of two components: a "synthetic estimator" of the form $\bar{\underset{\sim}{X}}_i'\hat{\underset{\sim}{\beta}}$ where $\bar{\underset{\sim}{X}}_i$ represents the average auxiliary information at the small area level and $\hat{\underset{\sim}{\beta}}$ is a vector of estimated regression coefficients; and a "correction factor" of the form $(\bar{y}_i - \bar{\underset{\sim}{x}}_i\hat{\beta})$ where $\bar{y}_i$ and $\bar{\underset{\sim}{x}}_i$ are the sample means of the target and the auxiliary variables. The correction factors are used to account for the variability of the small area means not explained by the auxiliary variables. The major difference between the various estimators is in the approach followed to determine the weights assigned to the two components in the linear combination, ranging from a "design based approach" (Särndal and Hidiroglou 1989) to "empirical Bayes" (Fay and Herriot 1979) and "mixed linear models" (Battese, Harter and Fuller 1989, Pfeffermann and Barnard 1991).

Very few studies are reported in the literature on the possible use of the time series relationships of the data to further increase the efficiency of the small area estimators. This is despite the fact that many of the small area estimators are derived from repeated surveys such as labour force surveys. The econometric literature contains a vast number of studies on the combined modelling of time series and cross-sectional data, see *e.g.* Rosenberg (1973b), Johnson (1977, 1980), Maddala (1977, Chapter 7), Dielman (1983) and Pfeffermann and Smith (1985) for reviews. However, none of these studies is directed to the problem of estimating (predicting) small area means from survey data. Fitting time series models to survey data has been considered

[1] D. Pfeffermann, Department of Statistics, Hebrew University, Jerusalem 91905. L. Burck, Unit for Statistical Analysis, Central Bureau of Statistics, Jerusalem 91130.

in the context of estimating aggregate population means, see the review papers of Smith (1979) and Binder and Hidiroglou (1988) and the more recent articles by Binder and Dick (1989), Tiller (1989) and Pfeffermann (1991). But again, these methods are not in routine use mainly because the classical survey estimators of the aggregate means are often almost as efficient when the models hold and more robust when the models fail to hold.

The situation is clearly different when dealing with a small area estimation problem; it seems to us that for this kind of problem, the use of time series models can be of great advantage. Although the exact nature of the model to be used in a particular application is obviously 'data dependent', the class of models we consider in the next section is broad enough to apply to many, if not most of the small area estimation problems arising in practice. These models have the further advantage that their estimation is relatively simple. Estimation issues are discussed in Section 3.

The use of a model always raises the question of how to protect against possible model failures and this question becomes even more sensitive when considering the use of a model for the production of official statistics. In Section 4 we consider this issue and propose a modification to the model dependent predictors which guarantees that for aggregates of the small area means for which the direct survey estimators can be trusted, the modified model predictors coincide with the survey estimators. The statistical properties of the modified predictors are explored. We conclude the article in Section 5 with empirical results which illustrate the performance of the model with and without the proposed modification. The data used for the illustrations are the sale prices of homes in the city of Jerusalem during the months of September 1985 through November 1989. These data are used routinely by the Central Bureau of Statistics in Israel for the computation of housing price indexes.

2. REGRESSION WITH CROSS-SECTIONALLY AND TIME VARYING COEFFICIENTS

2.1 A General Class of Models

In what follows we denote by $\underset{\sim}{Y}_{tk}$ the $n_{tk} \times 1$ vector of observations on a target variable Y, pertaining to an area k at time t, $k = 1, \ldots, K$, $t = 1, 2, \ldots$. We assume for convenience that $n_{tk} \geq 1$ but as becomes evident later on, the model permits that some of the areas not be observed at certain times. Let X_{tk} define the corresponding $n_{tk} \times (p + 1)$ design matrix of the auxiliary variables with a vector of ones as its first column. In many applications, the same row vector $\underset{\sim}{x}'_{tk}$ of auxiliary values applies to all the Y values of a given time so that $X_{tk} = \underset{\sim}{1}_{n_{tk}}\underset{\sim}{x}'_{tk}$ where $\underset{\sim}{1}_{n_{tk}}$ is a column vector of ones of length n_{tk}. This is the case when the only available data are the small area survey estimators. Confidentiality as well as processing costs often preclude the use of micro data on individual survey respondents. The theory described in this article is not restricted to the availability of the micro data (see the example in Section 2.2) but data availability has an obvious effect on model specifications and precision of estimation.

The regression model holding in area k at time t is defined as

$$\underset{\sim}{Y}_{tk} = X_{tk}\underset{\sim}{\beta}_{tk} + \underset{\sim}{\epsilon}_{tk};\ E(\underset{\sim}{\epsilon}_{tk}) = 0,\ E(\underset{\sim}{\epsilon}_{tk}\underset{\sim}{\epsilon}'_{tk}) = \sigma_k^2 I_{n_{tk}} \tag{2.1}$$

where $\underset{\sim}{\beta}'_{tk} = (\beta_{tk0}, \beta_{tk1}, \ldots, \beta_{tkp})$.

We define the (superpopulation) mean of the target variable values in area k at time t to be

$$\Theta_{tk} = E(M_{tk} \mid \underset{\sim}{\beta}_{tk}) = \bar{\underset{\sim}{X}}_{tk}\underset{\sim}{\beta}_{tk} \tag{2.2}$$

where

$$M_{tk} = \frac{1}{N_{tk}} \sum_{i=1}^{N_{tk}} Y_{tki} \quad \text{and} \quad \bar{\underset{\sim}{X}}_{tk} = \frac{1}{N_{tk}} \sum_{i=1}^{N_{tk}} \underset{\sim}{x}'_{tki}$$

with $i = 1, \ldots, N_{tk}$ indexing the population units. Obviously, when $\underset{\sim}{x}'_{tki} \equiv \underset{\sim}{x}'_{tk}$, then $\bar{\underset{\sim}{X}}_{tk} = \underset{\sim}{x}'_{tk}$.

Let $\hat{\underset{\sim}{\beta}}_{tk}$ define an estimator for $\underset{\sim}{\beta}_{tk}$. Then $\hat{\Theta}_{tk} = \bar{\underset{\sim}{X}}_{tk} \hat{\underset{\sim}{\beta}}_{tk}$ and

$$\hat{M}_{tk} = \frac{1}{N_{tk}} \left[\sum_{i=1}^{n_{tk}} Y_{tki} + \sum_{i=n_{tk}+1}^{N_{tk}} \underset{\sim}{x}'_{tki} \hat{\underset{\sim}{\beta}}_{tk} \right] = \hat{\Theta}_{tk} + \frac{1}{N_{tk}} \left(\sum_{i=1}^{n_{tk}} (Y_{tki} - \underset{\sim}{x}'_{tki} \hat{\underset{\sim}{\beta}}_{tk}) \right)$$

implying that in the usual case of small sampling rates within the areas, $\hat{\Theta}_{tk}$ can also be considered as an estimator of the finite population mean M_{tk}. For this reason we no longer distinguish between the finite and superpopulation means.

The notable feature of (2.1) is that the coefficients $\underset{\sim}{\beta}_{tk}$ are allowed to vary both cross-sectionally and over time. The following equations specify the variation of the coefficients over time:

$$\begin{bmatrix} \beta_{tkj} \\ \beta_{kj} \end{bmatrix} = T_j \begin{bmatrix} \beta_{t-1,kj} \\ \beta_{kj} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \eta_{tkj}, \; j = 0, \ldots, p \tag{2.3}$$

where we use the notation β_{kj}, $j = 0, 1, \ldots, p$, to define fixed coefficients which we interpret below, and T_j to define fixed (2×2) matrices and where the residuals $\{\eta_{tkj}\}$ satisfy

$$E(\eta_{tkj}) = 0, \; E(\eta_{tkj}\eta_{tk\ell}) = \delta_{j\ell}, \; E(\eta_{tkj}\eta_{t-d,k\ell}) = 0 \quad \text{for} \quad d > 0. \tag{2.4}$$

The implication of (2.4) is that residuals of different coefficients pertaining to the same time t are allowed to be correlated but the serial and cross serial correlations are assumed to be zero.

Next, we illustrate the use of (2.3) by considering some simple cases:

(a) $T_j = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ implies that $\beta_{tkj} = \beta_{kj} + \eta_{tkj}$ so that β_{kj} represents, in this case, a common mean. This is the well known Random Coefficient Regression Model (Swamy 1971) which is often used in econometric applications. Obviously, by postulating, $\mathrm{var}(\eta_{tkj}) = 0$, the model reduces to the case of a fixed regression coefficient over time.

(b) $T_j = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ implies that $\beta_{tkj} = \beta_{t-1,kj} + \eta_{tkj}$ which is the familiar random walk model, see *e.g.* Cooley and Prescott (1976) and LaMotte and McWhorter (1977) for application of this model in econometric studies. In this case the coefficient β_{kj} is redundant and should be omitted so that $T_j \equiv 1$.

(c) $T_j = \left[\begin{smallmatrix} \rho, & 1-\rho \\ 0, & 1 \end{smallmatrix}\right]$ implies the first order autoregressive relationship $(\beta_{tkj} - \beta_{kj}) = \rho(\beta_{t-1,kj} - \beta_{kj}) + \eta_{tkj}$ considered by Rosenberg (1973a).

(d) $T_j = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ implies that $\beta_{tkj} = \beta_{t-1,kj} + \beta_{kj} + \eta_{tkj}$ which defines a local approximation to a linear trend (Kitagawa and Gersch 1984). The coefficient β_{kj} represents, in this case, a fixed slope.

It should be emphasized that different matrices T_j can be used for different coefficients β_{tkj}. In fact, by defining $\alpha'_{tk} = (\beta_{tk0}, \beta_{k0}, \beta_{tk1}, \beta_{k1}, \ldots, \beta_{tkp}, \beta_{kp})$; $\tilde{T} = \text{diag}[T_0, T_1, \ldots, T_p]$, a block diagonal matrix with T_j as the j-th block; $\tilde{G} = I_{p+1} \otimes \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ where I_{p+1} is the identity matrix of order $p + 1$ and $\otimes$ defines the Kronecker product and $\eta'_{tk} = (\eta_{tk0}, \eta_{tk1}, \ldots, \eta_{tkp})$, the combined model holding for the coefficients β_{tk} can be written as

$$\alpha_{tk} = \tilde{T}\alpha_{t-1,k} + \tilde{G}\eta_{tk}; \quad E(\eta_{tk}) = 0, E(\eta_{tk}\eta'_{t-d,k}) = A_d\Delta \tag{2.5}$$

where $A_d = 1$ for $d = 0$ and $A_d = 0$ otherwise, and $\Delta = [\delta_{ij}]$ is defined by the variances and covariances δ_{ij} (equation 2.4).

The model defined by (2.5) specifies the variation of the regression coefficients of a specific area over time. The common approach to account for cross-sectional relationships between small area means is to allow for random small area effects which are time invariant $\{u_k\}$. The general model defined by (2.1) and (2.3) includes this case by writing $Y_{tk} = 1_{n_{tk}}u_{tk} + X_{tk}\beta_{tk} + \epsilon_{tk} = X^*_{tk}\beta^*_{tk} + \epsilon_{tk}$, say, and specifying $u_{tk} = u_{t-1,k} + \eta_{tk}$ with $u_{ok} = 0$, $\text{var}(\eta_{1k}) = \sigma^2_\eta$ and $\text{var}(\eta_{tk}) = 0$ for $t > 1$ (compare with case (b) above). By assuming in addition the autoregressive relationship defined by case (c) for the intercept variable and fixing the other regression coefficients (case (a) with zero residual variances), the resulting model is similar to the model considered by Choudhry and Rao (1989) except that in their general formulation of the model the observation residuals of equation (2.1) are allowed to be serially correlated. Notice that equation (2.1) now contains two random "intercept terms" but the model is nonetheless identifiable. Choudhry and Rao assume that the only available data are the survey estimators so that the estimation of the serial correlations needs to be carried out externally, using the micro observations. Alternatively, a model accounting for the serial correlations can be postulated. Choudhry and Rao assume an AR(1) model in their study.

A more general way to account for the cross-sectional relationships between the small area means is to allow for non zero correlations between the residual terms η_{tkj} and η_{tmj} of the models specifying the time series variation of the regression coefficients β_{tkj} and β_{tmj} operating in areas k and m (equation 2.4). Often it is reasonable to assume that the correlations decay as the distance between the areas increases. This can be formulated as, $E(\eta_{tkj}, \eta_{tmj}) = \delta_{jj}\rho_j f_j(k,m)$, $k \neq m$, where $f_j(k,m)$ is a monotonic decreasing function of the distances $D(k,m)$. The case of geometrically decaying correlations is obtained by defining $f_j(k,m) = \rho_j^{|k-m|-1}$. The case of fixed correlations is obtained by specifying $f_j(k,m) \equiv 1$ and in what follows we consider this case only. Allowing for fixed cross-sectional correlations for all the regression coefficients can be formulated as

$$E(\eta_{tk}\eta'_{tm}) = D(\Delta)\emptyset, \quad k \neq m \tag{2.6}$$

where $D(\Delta)$ is the diagonal matrix with the variances δ_{jj} on the main diagonal and $\emptyset$ is another diagonal matrix composed of the correlations ρ_j.

Before concluding this section we present the model defined by (2.1), (2.5) and (2.6) in a state-space form. Presenting the model in this form has important computational advantages.

Let $Y'_t = (Y'_{t1}, \ldots, Y'_{tK})$ represent the vector of observations of length $n_t = \sum_k n_{tk}$ for all the areas at time t and let $\epsilon'_t = (\epsilon'_{t1}, \ldots, \epsilon'_{tK})$ represent the corresponding regression residuals. Define $Z_{tk} = [1_{ntk}, 0_{ntk}, x_{tk1}, 0_{ntk}, \ldots, x_{tkp}, 0_{ntk}]$ where 0_{ntk} is a vector of zeroes of length n_{tk} and x_{tkj} is the vector of values for the j-th auxiliary variable, $j = 1, \ldots, p$. Let Z_t be the block diagonal matrix composed of the matrices Z_{tk}. The matrix Z_t is of order $n_t \times [K \times 2 \times (p + 1)]$. Define also $\alpha'_t = (\alpha'_{t1}, \ldots, \alpha'_{tK})$, $\eta'_t = (\eta'_{t1}, \ldots, \eta'_{tK})$, $\Sigma_t = \text{Diag}[\sigma^2_1 1'_{nt1}, \ldots, \sigma^2_K 1'_{ntK}]$, $T = I_K \otimes \tilde{T}$, and $G = I_K \otimes \tilde{G}$.

Using this notation, the model defined by (2.1), (2.5) and (2.6) can be written compactly as

$$\underset{\sim}{Y}_t = Z_t \underset{\sim}{\alpha}_t + \underset{\sim}{\epsilon}_t; \; E(\underset{\sim}{\epsilon}_t) = \underset{\sim}{0}, E(\underset{\sim}{\epsilon}_t \underset{\sim}{\epsilon}_t') = \Sigma_t \tag{2.7}$$

$$\underset{\sim}{\alpha}_t = T\underset{\sim}{\alpha}_{t-1} + G\underset{\sim}{\eta}_t; \; E(\underset{\sim}{\eta}_t) = \underset{\sim}{0}, E(\underset{\sim}{\eta}_t \underset{\sim}{\eta}_t') = \Lambda, \tag{2.8}$$

where $\Lambda = [\Lambda_{k\ell}]$, $k, \ell = 1, \ldots, K$ with $\Lambda_{k\ell} = \Delta$ when $k = \ell$ and $\Lambda_{k\ell} = D(\Delta)\emptyset$ when $k \neq \ell$. The matrices $\Lambda_{k\ell}$ are $(p + 1) \times (p + 1)$.

The model defined by (2.7) and (2.8) conforms to the classical state-space formulation, see, *e.g.* Anderson and Moore (1979) and Harvey (1984). By this formulation, (2.7) is the observation equation and (2.8) is the state equation with $\underset{\sim}{\alpha}_t$ defining the state vector. The apparent advantage of restructuring the model in a state space form is that the vectors $\underset{\sim}{\alpha}_t$, and hence the population means Θ_{tk}, as well as the estimation error variances can be estimated conveniently by means of the Kalman filter. We discuss the use of the filter in sections 3 and 4.

2.2 Explicit Estimators of the Small Area Means

In order to illustrate how past and neighbouring data are used under the model to "strengthen" the small area estimators we consider the case where the same vector $\underset{\sim}{x}_{tk}$ of auxiliary values applies to all the units of a given area at a given time. In this case the observation equation can be formulated in terms of the sample means, *i.e.*

$$\bar{Y}_{tk} = \underset{\sim}{x}_{tk}' \underset{\sim}{\beta}_{tk} + \bar{\epsilon}_{tk}; \; E(\bar{\epsilon}_{tk}) = 0, \; E(\bar{\epsilon}_{tk}^2) = \sigma_k^2 / n_{tk}, \; k = 1, \ldots, K. \tag{2.9}$$

Suppose that the regression coefficients follow a random walk (case (b) of equation 2.3) so that for area k

$$\beta_{tkj} = \beta_{t-1,kj} + \eta_{tkj}; \; E(\eta_{tkj}) = 0, \; E(\eta_{tkj}\eta_{tk\ell}) = \delta_{j\ell}, j, \ell = 1, \ldots, p \tag{2.10}$$

and for areas $k \neq m$,

$$E(\eta_{tkj}\eta_{tmj}) = \delta_{jj}\rho_j; \quad E(\eta_{tkj}\eta_{tm\ell}) = 0, j \neq \ell. \tag{2.11}$$

The random walk model implies that the coefficients drift slowly away from their initial value with no inherent tendency to return to a mean value. Obviously, for residuals η_{tkj} such that $E(\eta_{tkj}^2) = 0$ the corresponding regression coefficients are fixed over time. Notice also that since $\underset{\sim}{\beta}_{tk} = \underset{\sim}{\beta}_{t-1,k} + \underset{\sim}{\eta}_{tk}$, the predictor of $\underset{\sim}{\beta}_{tk}$ at time $(t - 1)$ is the same as the predictor $\hat{\underset{\sim}{\beta}}_{t-1,k}$ of $\underset{\sim}{\beta}_{t-1,k}$.

Using the Kalman filter equations presented in section 3, it is shown in the Appendix that the estimator $\hat{\Theta}_{tk}$ of the small area mean Θ_{tk} (equation 2.2) can be structured in this case in the following form

$$\hat{\Theta}_{tk} = \underset{\sim}{x}_{tk}'\hat{\underset{\sim}{\beta}}_{t-1,k} + \left(1 - \frac{\sigma_k^2}{n_{tk}v_k^2}\right)\left(\bar{Y}_{tk} - \underset{\sim}{x}_{tk}'\hat{\underset{\sim}{\beta}}_{t-1,k}\right) + \frac{\sigma_k^2}{n_{tk}v_k^2}\sum_{\substack{m=1\\m\neq k}}^{K}\gamma_{km}\left(\bar{Y}_{tm} - \underset{\sim}{x}_{tm}'\hat{\underset{\sim}{\beta}}_{t-1,m}\right) \tag{2.12}$$

where the coefficients $\{\gamma_{km}\}$ are the partial regression coefficients in the regression of $e_{tk} = (\bar{Y}_{tk} - \underset{\sim}{x}'_{tk}\hat{\underset{\sim}{\beta}}_{t-1,k})$ against the prediction errors $\{e_{tm} = (\bar{Y}_{tm} - \underset{\sim}{x}'_{tm}\hat{\underset{\sim}{\beta}}_{t-1,m})\}$ obtained in the other areas and v_k^2 is the residual (unexplained) variance in the regression.

The estimator $\hat{\Theta}_{tk}$ is composed of three components: the "synthetic" estimator, $\underset{\sim}{x}'_{tk}\hat{\underset{\sim}{\beta}}_{t-1,k}$, where $\hat{\underset{\sim}{\beta}}_{t-1,k}$ is the optimal predictor of $\underset{\sim}{\beta}_{tk}$ based on all the observations up to and including time $t-1$, the "correction factor" $(\bar{Y}_{tk} - \underset{\sim}{x}'_{tk}\hat{\underset{\sim}{\beta}}_{t-1,k})$ based on the prediction error in area k, and an "adjustment factor" based on the prediction errors observed for the other areas. The first two components correspond to the components of the classical small area estimators discussed in the introduction. Notice that the smaller the sample size n_{tk}, the smaller is the weight assigned to the current sample mean $\bar{Y}_{tk}$ in the estimation of Θ_{tk} and the larger is the weight assigned to the time series predictor $\underset{\sim}{x}'_{tk}\hat{\underset{\sim}{\beta}}_{t-1,k}$. The third component in the right hand side of (2.12) represents the information borrowed from neighbouring areas. The weight assigned to this component depends on the magnitude of the correlations ρ_j between the corresponding error terms $\{\eta_{tkj}\}$ in the models holding for the regression coefficients (equation 2.11). Obviously, when the regressions in the various areas are independent so that $\rho_j = 0$ for all j and hence $\gamma_{km} = 0$ for all m, the third component vanishes and the predictor $\hat{\Theta}_{tk}$ reduces to a weighted average of the current mean $\bar{Y}_{tk}$ and the time series predictor $\underset{\sim}{x}'_{tk}\hat{\underset{\sim}{\beta}}_{t-1,k}$.

3. MODEL ESTIMATION AND INITIALIZATION USING THE KALMAN FILTER

3.1 Estimation of the Regression Coefficients by Means of the Kalman Filter

In this section we present the Kalman filter equations for the updating and smoothing of the state vectors $\underset{\sim}{\alpha}_t$ defined by the equations (2.7) and (2.8) (the area regression coefficients in our case). We assume that the V-C matrices Σ_t and Λ are known. Estimation of these matrices is considered in section 3.2. The theory of the Kalman filter is developed in numerous publications (see *e.g.* Anderson and Moore 1979 and Meinhold and Singpurwalla 1983) and so we restrict the discussion to aspects most germane to the small area estimation problem.

Let $\hat{\underset{\sim}{\alpha}}_{t-1}$ be the best linear unbiased predictor (blup) of $\underset{\sim}{\alpha}_{t-1}$ based on all the data observed up to time $(t-1)$. Since $\hat{\underset{\sim}{\alpha}}_{t-1}$ is blup for $\underset{\sim}{\alpha}_{t-1}$, $\hat{\underset{\sim}{\alpha}}_{t|t-1} = T\hat{\underset{\sim}{\alpha}}_{t-1}$ is the blup of $\underset{\sim}{\alpha}_t$ at time $(t-1)$. Furthermore, if $P_{t-1} = E(\hat{\underset{\sim}{\alpha}}_{t-1} - \underset{\sim}{\alpha}_{t-1})(\hat{\underset{\sim}{\alpha}}_{t-1} - \underset{\sim}{\alpha}_{t-1})'$ is the V-C matrix of the prediction errors at time $(t-1)$, $P_{t|t-1} = TP_{t-1}T' + G\Lambda G'$ is the V-C matrix of the prediction errors $(\hat{\underset{\sim}{\alpha}}_{t|t-1} - \underset{\sim}{\alpha}_t)$. (Follows straightforwardly from 2.8).

When a new vector of observations $[\underset{\sim}{Y}_t, Z_t]$ becomes available, the predictor of $\underset{\sim}{\alpha}_t$ and the V-C matrix P_{t-1} are updated according to the formulae

$$\begin{aligned} \hat{\underset{\sim}{\alpha}}_t &= \hat{\underset{\sim}{\alpha}}_{t|t-1} + P_{t|t-1}Z_t'F_t^{-1}(\underset{\sim}{Y}_t - \hat{\underset{\sim}{Y}}_{t|t-1}) \\ P_t &= (I - P_{t|t-1}Z_t'F_t^{-1}Z_t)P_{t|t-1} \end{aligned} \tag{3.1}$$

where $\hat{\underset{\sim}{Y}}_{t|t-1} = Z_t\hat{\underset{\sim}{\alpha}}_{t|t-1}$ is the blup of $\underset{\sim}{Y}_t$ at time $(t-1)$ so that $\underset{\sim}{e}_t = (\underset{\sim}{Y}_t - \hat{\underset{\sim}{Y}}_{t|t-1})$ is the vector of innovations with V-C matrix $F_t = (Z_tP_{t|t-1}Z_t' + \Sigma_t)$.

The new data observed at time t can be used also for the updating (smoothing) of past estimators of the state vectors and hence for the updating of past estimators of the small area means. Denoting by t^* the most recent month with observations, the smoothing is carried out using the equations

$$\hat{\alpha}_{t|t^*} = \hat{\alpha}_t + P_t T' P_{t+1|t}^{-1}(\hat{\alpha}_{t+1|t^*} - T\hat{\alpha}_t)$$

$$P_{t|t^*} = P_t + P_t T' P_{t+1|t}^{-1}(P_{t+1|t^*} - P_{t+1|t})P_{t+1|t}^{-1} T P_t; \; t = 2, 3, \ldots, t^* \tag{3.2}$$

where $P_{t|t^*}$ is the V-C matrix of the prediction errors $(\hat{\alpha}_{t|t^*} - \alpha_t)$. Notice that $\hat{\alpha}_{t^*|t^*} = \hat{\alpha}_{t^*}$ and $P_{t^*|t^*} = P_{t^*}$ define the starting values for the smoothing equations.

Estimators of the small area means or aggregates of the means are obtained from the filtered (or smoothed) estimators of α_t in a straightforward manner using the relationship $\hat{\theta}_{tk} = \bar{X}_{tk}\hat{\beta}_{tk} = \bar{Z}'_{tk}\hat{\alpha}_{tk} = \bar{Z}'_{tk}A_{tk}\hat{\alpha}_t$ where $\bar{Z}'_{tk} = (1, 0, \bar{X}_{tk1}, 0, \ldots, \bar{X}_{tkp}, 0)$ and A_{tk} is the appropriate indicator matrix. Hence, if $\theta_t^w = \sum_{k=1}^K w_k \theta_{tk}$, then $\hat{\theta}_t^w = \sum_{k=1}^K w_k \bar{Z}'_{tk} A_{tk} \hat{\alpha}_t = a'_{tw}\hat{\alpha}_t$, say. For given V-C matrices Σ_t and Λ, the MSE's of the estimation errors are obtained as

$$E(\hat{\theta}_{tk} - \theta_{tk})^2 = \bar{Z}'_{tk}A_{tk}P_t A'_{tk}\bar{Z}_{tk} \quad \text{and} \quad E(\hat{\theta}_{tk}^w - \theta_{tk}^w) = a'_{tw}P_t a_{tw}. \tag{3.3}$$

Notice that the MSE's in (3.3) are with respect to the joint distribution of the observations $\{Y_{tk}\}$ and the vectors of coefficients $\{\beta_{tk}\}$ so that they represent average MSE's over the possible realizations of the area means.

3.2 Estimation of the V-C Matrices and Initialization of the Filter

The actual application of the Kalman filter requires the estimation of the unknown elements of the matrices Σ_t and Λ and the initialization of the filter, that is, the estimation of the vector α_o and the corresponding V-C matrix P_o of the estimation errors. In this section we describe simple estimation procedures which can be used for these purposes.

Assuming a normal distribution for the residual terms ϵ_t and η_t of equations (2.7) and (2.8), the log likelihood function of the vectors $Y_{m+1}, \ldots, Y_{t^*}$, conditional on the first m vectors $Y_1, \ldots, Y_m$, can be formulated as

$$L(\lambda) = \text{constant} - \frac{1}{2}\sum_{t=m+1}^{t^*} (\log |F_t| + e_t' F_t^{-1} e_t) \tag{3.4}$$

where λ contains the unknown model variances and covariances written in a vector form. The scalar m defines the number of time periods needed to construct initial values for the Kalman filter. (For the random walk model considered in section 2.2, $m = 1$, provided that sufficient data are available in every area to allow the computation of the OLS estimators of the vectors of coefficients). The expression in (3.4) follows from the prediction error decomposition, see Schweppe (1965) and Harvey (1981) for details. For given matrices Σ_t and Λ, the innovations e_t and the V-C matrices F_t can be obtained by application of the Kalman filter equations (3.1).

The computation of the likelihood function requires the initialization of the Kalman filter which can be carried out most conveniently by application of the approach proposed by Harvey and Phillips (1979). By this approach, the nonstationary components of the state vector are initialized with very large error variances which corresponds to postulating a noninformative prior distribution so that the corresponding state estimates can conveniently be taken as zeroes. (For the random walk model, initializing with a noninformative prior yields the OLS estimators after one time period, see Meinhold and Singpurwalla 1983, for a Bayesian formulation of the Kalman filter). The stationary components of the state vector are initialized by the corresponding unconditional means and variances which may be part of the unknown parameters defining the arguments of the likelihood function.

Maximization of the likelihood function (3.4) can be implemented using the method of scoring with a variable step length. In particular, let $\underset{\sim}{\lambda}_{(o)}$ define initial estimates of the unknown elements in $\underset{\sim}{\lambda}$. Then the method of scoring consists of solving iteratively the set of equations

$$\underset{\sim}{\lambda}_{(i)} = \underset{\sim}{\lambda}_{(i-1)} + r_i\{I[\underset{\sim}{\lambda}_{(i-1)}]\}^{-1}g[\underset{\sim}{\lambda}_{(i-1)}] \tag{3.5}$$

where $\underset{\sim}{\lambda}_{(i-1)}$ is the estimator of $\underset{\sim}{\lambda}$ as obtained in the $(i-1)$-th iteration, $I[\underset{\sim}{\lambda}_{(i-1)}]$ is the information matrix evaluated at $\underset{\sim}{\lambda}_{i-1}$ and $g[\underset{\sim}{\lambda}_{(i-1)}]$ is the gradient of the log likelihood evaluated at $\underset{\sim}{\lambda}_{i-1}$. The coefficient r_i is a variable step length introduced to guarantee that $L[\underset{\sim}{\lambda}_{(i)}] \geq L[\underset{\sim}{\lambda}_{(i-1)}]$ in every iteration. The value of r_i can be determined by a grid search procedure in the region $[0,1]$. The formulae for the k-th element of the gradient vector and the $k\ell$-th element of the information matrix are given in Watson and Engle (1983).

Having estimated the model variances and covariances, these estimates can be substituted for the true parameters in the Kalman filter equations (3.1) - (3.2) to yield the estimators of the regression coefficients and the V-C matrices and hence the small area estimators and their variances (see equation 3.3). Notice however that the estimated V-C matrices ignore the variability induced by the need to estimate the unknown elements contained in $\underset{\sim}{\lambda}$. Ansley and Kohn (1986) propose correction factors of order $1/t^*$ to account for this extra variation in state space modelling using first order Taylor approximations. Hamilton (1986) proposes a Monte Carlo procedure which consists of sampling from a multivariate normal distribution with mean given by the maximum likelihood estimator of the vector $\underset{\sim}{\lambda}$ and V-C matrix defined by the inverse of the information matrix, and estimating the state vectors for each random realization of the parameter values. This procedure is more flexible in terms of the assumptions involved and provides further insight into the sensitivity of the Kalman filter estimators to errors in the variance and covariance estimators. However, it is computationally more intensive.

4. MODIFICATIONS TO PROTECT AGAINST MODEL BREAKDOWNS

4.1 Description of the Problem and Proposed Modifications

The use of a model for small area estimation seems inevitable in view of the small sample sizes within the areas. However it raises the question of how to protect against model breakdowns. Testing the model every time that new data becomes available is often not practical, requiring instead the development of a "built-in mechanism" to ensure the robustness of the estimators when the model fails to hold.

One possibility is to modify the regression estimators derived in the various time periods so that they satisfy certain linear constraints obtained by equating aggregate means of the raw data with their expected fitted values under the model. More precisely, we propose to augment the model equation (2.1) by linear constraints of the form

$$\sum_k W_{tk}^{(\ell)} \sum_i Y_{tki} = \sum_k W_{tk}^{(i)} \sum_i \underset{\sim}{x}'_{tki} \underset{\sim}{\beta}_{tk} \quad \ell = 1, 2, \ldots, L(t),\ t = 1, \ldots, t^* \tag{4.1}$$

where the coefficients $W_{tk}^{(\ell)}$ are fixed, standardized weights such that $\Sigma_k n_{tk} W_{tk}^{(\ell)} = 1$. An example for such a constraint would be the equation

$$\sum_{k=1}^{K} N_{tk}\hat{M}_{tk} \Big/ \sum_{k=1}^{K} N_{tk} = \sum_{k=1}^{K} N_{tk}(\underset{\sim}{x}'_{tk}\underset{\sim}{\beta}_{tk}) \Big/ \sum_{k=1}^{K} N_{tk} \tag{4.2}$$

where $\hat{M}_{tk}$ is the direct, survey estimator in area k. For $\underset{\sim}{x}_{tk} \simeq \underset{\sim}{\bar{X}}_{tk}$, the equation (4.2) guarantees that the model dependent predictor of the aggregate population mean coincides with the corresponding survey estimator. Such a constraint can be justified by arguing that the survey estimators, although not reliable enough for estimating the small area means due to the small sample sizes, can be trusted when being combined for estimating the aggregate mean. Notice that "adding up" constraints are ordinarily imposed on statistical agencies anyway. Battese, Harter and Fuller (1988) and Pfeffermann and Barnard (1991) use a similar constraint for analysing cross-sectional surveys. Often, the small areas can be grouped into broader groups, with sufficient data in each of the groups to justify the use of the survey estimators for estimating the corresponding group means. In this case, one can impose several constraints of the form (4.2) where the summation is now over the areas belonging to the same group. Notice in this respect that in view of the correlations between the regression coefficients operating in the various areas, a constraint applied to a sub-set of the areas will modify the regression estimates in all the areas. We illustrate this property in the empirical study.

It is important to emphasize that the set of constraints in (4.1) does not represent external information about possible values of the regression coefficients. Rather, it serves as a "control system" to guarantee that the model estimators adjust themselves more rapidly to possible changes in the behavior of the regression coefficients. As a result, the variances of the modified regression estimators are slightly larger than the variances of the optimal estimators under the model. Obviously, when no such changes occur and the variances of the aggregate means are sufficiently small, one would expect the constraints to be satisfied approximately even without imposing them explicitly. As mentioned above, it is possible to incorporate several separate constraints in each time period but it is imperative that the variances of the corresponding aggregate means will be small enough to ensure that the modifications are indeed needed and do not interfere with the random fluctuation of the raw data.

4.2 Inference Incorporating the Linear Constraints

In Section 4.1 we proposed to amend the model equations (2.1) by imposing the set of constraints (4.1) thereby ensuring the robustness of the regression estimators against sudden drifts in the values of the coefficients.

Computationally, this can be implemented most conveniently by augmenting the vectors $\underset{\sim}{Y}_t$ of equation (2.7) by the scalars $\sum_k W_{tk}^{(\ell)} \sum_i Y_{tki}$, augmenting the matrices Z_t by the corresponding row vectors $(W_{t1}^{(\ell)} 1'_{nt1} Z_{t1}, \ldots, W_{tK}^{(\ell)} 1'_{ntK} Z_{tK})$ and setting the respective variances of the residual terms to zero. The augmented set of equations, together with (2.8), form a pseudo state-space model which could be estimated using the Kalman filter equations (3.1). Notice that the pseudo V-C matrix $\Sigma_t^{(P)}$ of the augmented residual vector is no longer positive definite (the last $L(t)$ rows and columns of $\Sigma_t^{(P)}$ consist of zeroes) but this does not cause computational difficulties.

The drawback of applying the Kalman filter to the pseudo model is that the V-C matrices of the regression estimators fail to account for the actual variability of the aggregate means appearing in the left hand side of (4.1). In order to deal with this problem, we propose to amend the formula for the updating of the V-C matrix P_t (equation 3.1) so that the variances and covariances of the aggregate means will be taken into account.

Let $\underset{\sim}{Y}_t^{(A)}$ and $Z_t^{(A)}$ represent the augmented Y vector and Z matrix at time t and denote by $\Sigma_t^{(A)}$ the actual V-C matrix of the residual terms $[\underset{\sim}{Y}_t^{(A)} - Z_t^{(A)}\underset{\sim}{\alpha}_t]$. The matrix $\Sigma_t^{(A)}$ is of order $[n_t + L(t)]$ with Σ_t in the first n_t rows and columns and the variances and covariances of the means $\Sigma_k W_{tk}^{(\ell)} \Sigma_i Y_{tki}$ among themselves and with the vector $\underset{\sim}{Y}_t$ in the remaining rows and columns. Denoting by $\hat{\underset{\sim}{\alpha}}_{t-1}^{(A)}$ the robust predictor of $\underset{\sim}{\alpha}_{t-1}$ as obtained at time $(t - 1)$ using the pseudo model and by $P_{t-1}^{(A)}$ the actual V-C matrix of the errors $(\hat{\underset{\sim}{\alpha}}_{t-1}^{(A)} - \underset{\sim}{\alpha}_{t-1})$, the modified state estimator at time t is obtained as

$$\hat{\underset{\sim}{\alpha}}_t^{(A)} = T\hat{\underset{\sim}{\alpha}}_{t-1}^{(A)} + P_{t|t-1}^{(A)} Z_t^{(A)\prime} \left(F_t^{(P)}\right)^{-1} \left[\underset{\sim}{Y}_t^{(A)} - Z_t^{(A)} T \hat{\underset{\sim}{\alpha}}_{t-1}^{(A)}\right] \tag{4.3}$$

where $P_{t|t-1}^{(A)} = (TP_{t-1}^{(A)}T' + G\Lambda G')$ and $F_t^{(P)} = Z_t^{(A)} P_{t|t-1}^{(A)} Z_t^{(A)\prime} + \Sigma_t^{(P)}$ (Compare with 3.1). It is shown in the Appendix that the actual V-C matrix $P_t^{(A)}$ of the errors $(\hat{\underset{\sim}{\alpha}}_t^{(A)} - \underset{\sim}{\alpha}_t)$ satisfies the recursive equation

$$P_t^{(A)} = \left[I - K_t^{(P)} Z_t^{(A)}\right] P_{t|t-1}^{(A)} + K_t^{(P)} \left[\Sigma_t^{(A)} - \Sigma_t^{(P)}\right] K_t^{(P)}, \tag{4.4}$$

where $K_t^{(P)} = P_{t|t-1}^{(A)} Z_t^{(A)\prime} (F_t^{(P)})^{-1}$ is the pseudo Kalman gain. The first expression on the right hand side of (4.4) corresponds to the usual updating formula of the Kalman filter (compare with 3.1)). The second expression is a correction factor which accounts for the actual variances and covariances of the means $\Sigma_k W_{tk}^{(\ell)} \Sigma_i Y_{tki}$, not taken into account in the first expression.

The amended Kalman filter defined by the equations (4.3) and (4.4) produces robust predictors $\hat{\underset{\sim}{\alpha}}_t^{(A)}$ instead of the optimal, model dependent predictors, $\hat{\underset{\sim}{\alpha}}_t$ but otherwise uses the correct V-C matrices under the model. Thus, this filter can be used for the routine estimation of the vectors of coefficients and hence for the estimation of the small area means, and when the model holds it will give similar results to those obtained under the optimal filter. In periods where the model fails to hold, the updating formula (4.4) could be incorrect (depending on the particular model failures) but the predictors $\hat{\underset{\sim}{\alpha}}_t^{(A)}$ will nonetheless satisfy the linear constraints (4.1). The smoothing equations (3.2) can likewise be modified to satisfy the linear constraints.

5. EMPIRICAL RESULTS

5.1 Description of the Data and Model Fitted

In order to illustrate the important features of the class of models defined in Section 2, we fitted such a model to home sale prices in Jerusalem. The sale prices are recorded on a monthly basis and are routinely used by the Central Bureau of Statistics in Israel for the computation of monthly housing price indexes (HPI) adjusted for changes in quality. The HPI is computed separately for each city or group of cities and for each house size defined by the number of rooms, ranging from 1 to 5. The number of transactions carried out each month is very small in many of these cells and for 1 room apartments it occasionally happens that there are no transactions. The mean and standard deviation (S.D.) of the monthly number of transactions carried out during the period July 1987 - November 1989 are listed below.

Size	1	2	3	4	5
Mean	2.7	29.0	101.9	39.7	5.6
S.D.	2.6	12.9	50.4	18.8	3.5

The need to adjust for changes in quality results from the fact that the transactions performed are not under control, giving rise to large differences in quality from one month to the other particularly in the small cells. The following quality measure variables (QMV) are recorded for every transaction: $\tilde{X}^{(1)}$ - the apartment floor area, $\tilde{X}^{(2)}$ - the age of the apartment, $X^{(3)}$, $X^{(4)}$ - dummy variables defining districts within the city.

The problems involved in the computation of the HPI and the method used in Israel are discussed at length in a recent article by Pfeffermann, Burck and Ben-Tuvia (1989). The following model was proposed by the authors as an alternative to the model in current use. The triple index "tki" defines the i-th transaction of size k in month t with Y_{tki} standing for the log of the sale price and $X_{tki}^{(j)} = \log(\tilde{X}_{tki}^{(j)}), j = 1, 2$.

$$Y_{tki} = \beta_{tk0} + \beta_{tk1}X_{tki}^{(1)} + \beta_{tk2}X_{tki}^{(2)} + \beta_{tk3}X_{tki}^{(3)} + \beta_{tk4}X_{tki}^{(4)} + \epsilon_{tki} \tag{5.1}$$

$$\begin{aligned} \beta_{tk0} &= \beta_{t-1,k0} + \beta_{k0} + \eta_{tk0} \\ \beta_{tkj} &= \beta_{t-1,kj} + \eta_{tkj}, \; j = 1, \ldots, 4, \end{aligned} \tag{5.2}$$

with the error terms ϵ_{tki} and η_{tkj} satisfying the assumptions (2.1), (2.4) and (2.5). Notice that the model assumed for the intercept term is the local approximation to a linear trend defined under case (d) of Section (2.1). The model assumed for the other coefficients is the random walk model defined under case (b).

The regression defined by (5.1) forms the basis for the construction of an HPI adjusted for changes in quality. By fixing the values of the QMV's at their average population values which are constant over time, (the values of these variables are adjusted approximately every five years), average sale prices can be computed using (5.1) and these averages are comparable between months since they refer to homes of similar qualities.

Pfeffermann, Burck and Ben-Tuvia discuss the considerations in selecting the model defined by (5.2) for the regression coefficients. They show empirical results which validate the fitness of the model. However, the results of that study were obtained by fitting the model to each cell separately, that is, without accounting for the cross-sectional relationships of the regression coefficients. This aspect of the model is explored in the present study. Another major purpose of the empirical study is to illustrate the performance of the modifications proposed in Section 4 to protect against model breakdowns.

5.2 Estimation of the Model

The model defined by (5.1) and (5.2) can be put in a state-space form similar to (2.7) and (2.8). In fact, the vectors $\underset{\sim}{\alpha}_t$ and the matrices Z_t, T and G assume, in this case, simple structures, since for $j = 1, \ldots, 4$, $\beta_{kj} \equiv 0$ (see case (b) of Section 2.1). Thus, $\underset{\sim}{\alpha}_{tk}' = (\beta_{tk0}, \beta_{k0}, \beta_{tk1}, \ldots, \beta_{tk4})$, $Z_{tk} = [\underset{\sim}{1}_{ntk}, \underset{\sim}{0}_{ntk}, \underset{\sim}{X}_{tk}^{(1)}, \ldots, \underset{\sim}{X}_{tk}^{(4)}]$, $\tilde{T} = [\underset{\sim}{e}_1, \underset{\sim}{e}_1 + \underset{\sim}{e}_2, \underset{\sim}{e}_3, \ldots, \underset{\sim}{e}_6]$, a 6×6 matrix with $\underset{\sim}{e}_j$ having a one in position j and zeroes elsewhere and $\tilde{G} = [\underset{\sim}{e}_1, \underset{\sim}{e}_3, \ldots, \underset{\sim}{e}_6]$ which is 6×5. The matrix Δ is defined as in (2.5). The vector $\underset{\sim}{\alpha}_t$ and the matrices Z_t, T, G and Λ are obtained from the vectors $\{\underset{\sim}{\alpha}_{tk}\}$ and the matrices $\{Z_{tk}\}$, $\tilde{T}$, $\tilde{G}$ and Δ in the same way as in (2.7) and (2.8).

Having set the model in a state-space form we next attempted to estimate the unknown variances and covariances using the method of scoring algorithm described in Section 3.2. As it turned out, however, the computer time needed for convergence was way beyond the capacity of the IBM 1481 mainframe used for this study. Notice that the number of unknown parameters of the combined state-space model is $\dim(\underset{\sim}{\lambda}) = 25$ whereas the dimension of the state vectors and hence the dimension of the corresponding V-C matrices is $\dim(\underset{\sim}{\alpha}_t) = 30$. The total number of observations per month ranges from 55 to 353. The computer program written for this study uses numerical derivatives so that each iteration of the method of scoring requires a separate sweep through all the data with each sweep involving $[\dim(\underset{\sim}{\lambda}) + 1]$ computations of the state vector $\hat{\underset{\sim}{\alpha}}_t$ and the V-C matrix P_t (equation 3.1) at each point in time. These computations are needed in order to evaluate the log likelihood functions and hence the corresponding derivatives. It is clear therefore that the computational costs increase with the length of the series, the number of observations, the size of the state vector and the number of unknown parameters.

In order to deal with this problem we estimated the variance σ_k^2 (equation 2.1) and the matrix Δ (equation 2.5) separately for each of the five apartment sizes using the time series of observations corresponding to each size and then estimated the correlations ρ_j (equation 2.6) by a crude, grid search procedure. We found that setting $\rho_j = ½$ for every j gives satisfactory results both in terms of the behaviour of the innovations (the one step ahead prediction errors) and in terms of the smoothness of the regression coefficients corresponding to apartments of size one and five where the monthly sample sizes are very small. Notice that by estimating the variances and covariances defining the time series relationships of the regression coefficients separately for each size, one is more flexible in terms of the model assumptions although there is some loss of efficiency if the variances and covariances are indeed the same across the different sizes.

5.3 Results

Pfeffermann, Burck and Ben-Tuvia (1989) illustrate the adequacy of the time series models fitted to the various apartment sizes. As mentioned earlier, our purpose in this study is to compare the results obtained with and without the accounting for the cross-sectional correlations and to illustrate the performance of the modifications (4.1) in protecting against model breakdowns.

In order to sharpen the comparisons as much as possible, we deliberately inflated the Y-values by 5 percent in each of the following four months: October 1987, November 1988, January 1989 and May 1989. Thus all the Y-values of all the apartment sizes corresponding to the months October 1987 - October 1988 were inflated by 5 percent, the Y-values corresponding to November 1988 - December 1988 were inflated by 10.25 percent (5 percent on top of the previous 5 percent) and so forth. These kinds of model breakdowns (although obviously not in such magnitudes) may result from intentional devaluations of the currency and are of main concern when modeling sale prices. See Pfeffermann, Burck and Ben-Tuvia for further discussion. Similar model breakdowns may occur, for example, with series of unemployment rates in periods of abrupt economic recessions.

Table 1 shows the average mean squared errors (AMSE) of the model residuals $\hat{\epsilon}_{tki} = (Y_{tki} - \hat{\beta}_{tko} - \sum_{j=1}^{4} X_{tki}^{(j)} \hat{\beta}_{tkj})$ and the model innovations $e_{tki} = [Y_{tki} - (\hat{\beta}_{t-1,k0} + \beta_{k0}) - \sum_{j=1}^{4} X_{tki}^{(j)} \hat{\beta}_{t-1,kj}]$ (see equations 5.1 and 5.2), separately for each of the five apartment sizes. The AMSE's were computed as $\text{AMSE}_k(\epsilon) = 1/N \sum_{t=1}^{N} (1/n_t \sum_{i=1}^{n_t} \hat{\epsilon}_{tki}^2)$; $\text{AMSE}_k(e) = 1/N \sum_{t=1}^{N} (1/n_t \sum_{i=1}^{n_t} e_{tki}^2)$ where $t = 1, \ldots, N$ indexes the months of July 1987 - November 1989. We distinguish between four different estimators of the regression coefficients as defined by whether the model accounts for the cross-sectional correlations $(\rho_j \equiv ½)$, $(\rho_j \equiv 0)$ and by whether or not the estimators are modified to protect against the model breakdowns (abbreviated as "Rob. Inc." and "No Rob." in the table). The modifications were carried out by augmenting the observation equation of each month by three linear constraints of the form 4.2. These constraints forced the aggregate means of the fitted values in each of the three

Table 1

Average Mean Squared Errors of Residuals and Innovations With and Without the Accounting for Cross-sectional Correlations and the Inclusion of the Robustness Modifications, by Size

Apt. Size	Mean Squared Errors of Innovations				Mean Squared Errors of Residuals			
	$\rho \equiv 1/2$		$\rho \equiv 0$		$\rho \equiv 1/2$		$\rho \equiv 0$	
	Rob. Inc.	No Rob.	Rob. Inc.	No Rob.	Rob. Inc.	No Rob.	Rob. Inc.	No Rob.
1	.141	.134	.176	.218	.021	.027	.056	.092
2	.070	.090	.084	.123	.021	.039	.023	.070
3	.065	.090	.070	.197	.017	.042	.019	.143
4	.067	.123	.072	.198	.019	.066	.021	.141
5	.067	.114	.077	.193	.023	.033	.065	.106

districts to coincide with the corresponding means of the observed values. When incorporating the constraints, the model was fitted using the amended Kalman filter as defined by the equations (4.3) and (4.4).

In order to illustrate the performance of the four sets of regression estimators in the various months and in particular, in and around the months where we inflated the data, we plotted the monthly MSE's of the innovations and residuals as obtained for 3 and 5 room apartments. The plots are shown in Figures 1 to 4. Notice that the values of Table 1 for 3 and 5 room apartments are correspondingly the averages of the values shown in the four figures.

The main conclusions from the table and the graphs are as follows:

Accounting for the cross-sectional correlations and including the linear constraints to protect against the model breakdowns yields better results than in the other cases considered. This outcome is most prominent in the cells of 1 and 5 room apartments where the sample sizes in each month are very small. In the other three cells, there are only small differences between the case ($\rho \equiv 1/2$, Rob. Inc.) and the case ($\rho \equiv 0$, Rob. Inc.) which could be expected since as the number of observations in each month increases, there is less borrowing of information from neighbouring cells (small areas in the more general context). The situation is different, however, when the linear constraints are removed. Accounting for the cross-sectional correlations yields in this case much better results than when not accounting for them and this is true for all the apartment sizes. Thus, by borrowing information from one cell to the other, the estimators of the regression coefficients adapt themselves much more rapidly to the sudden drifts in the data as seen also more directly in the figures [The four peaks in each graph are in the months where the data were inflated and as can be seen, the graphs corresponding to the case ($\rho \equiv 1/2$, No Rob.) return to their normal level of the months before the inflation much faster than the graphs representing the case ($\rho \equiv 0$, No Rob.)

Another interesting comparison is between the case where the linear constraints are included and the case where they are not. Clearly, the inclusion of the constraints improves the results substantially when accounting for the serial correlations and the improvements are even more prominent when the serial correlations are set to zero. It is interesting to compare in this context the figures exhibiting the monthly MSE's of the innovations with the figures exhibiting the monthly MSE's of the residuals. In the four months where we inflated the data the MSE's of the innovations are high which is obvious since the innovations are the differences between the observations and their predictors from previous months. Still, when the linear constraints are included, the MSE's return to their normal level right after the months of inflation. As

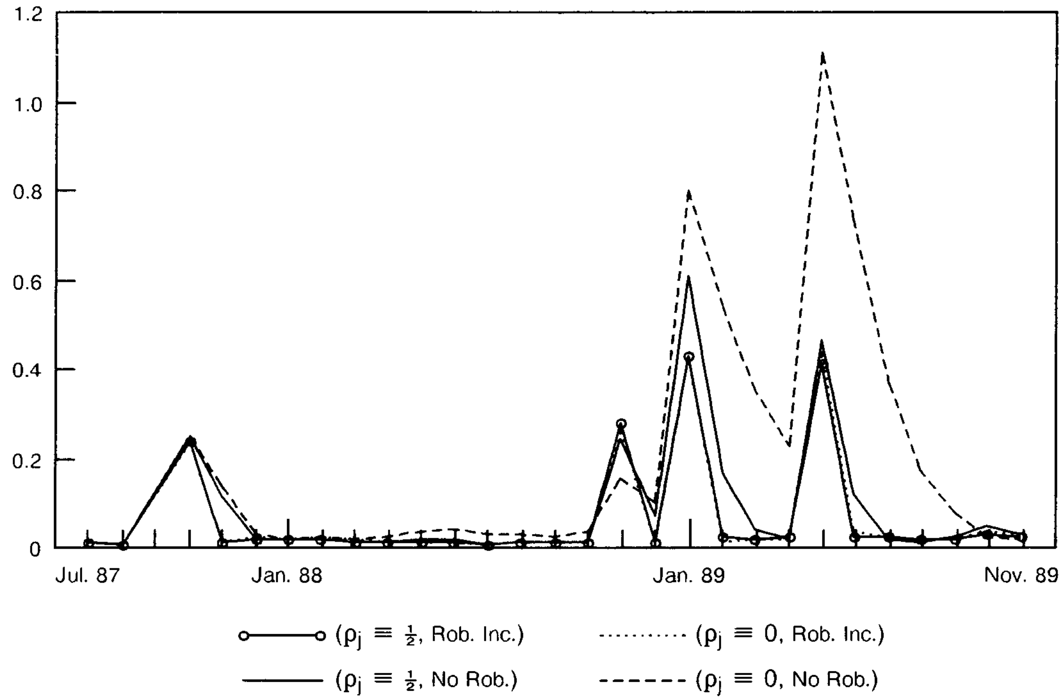

Figure 1 Monthly Mean Squared Errors of Innovations, 3 Room Apartments

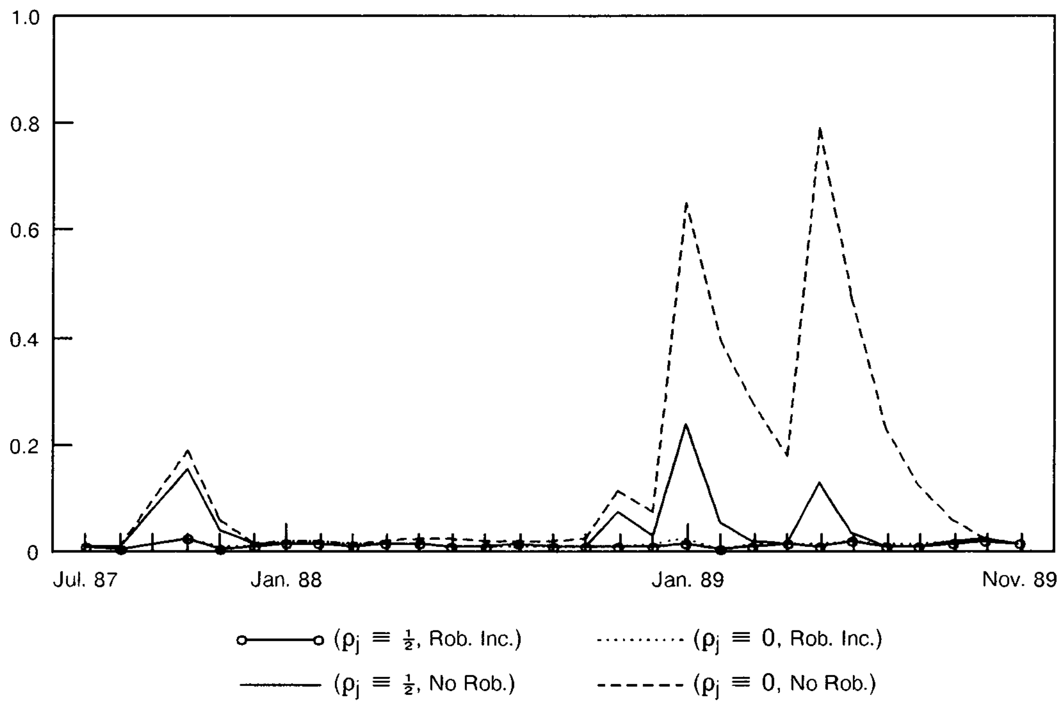

Figure 2 Monthly Mean Squared Errors of Residuals, 3 Room Apartments

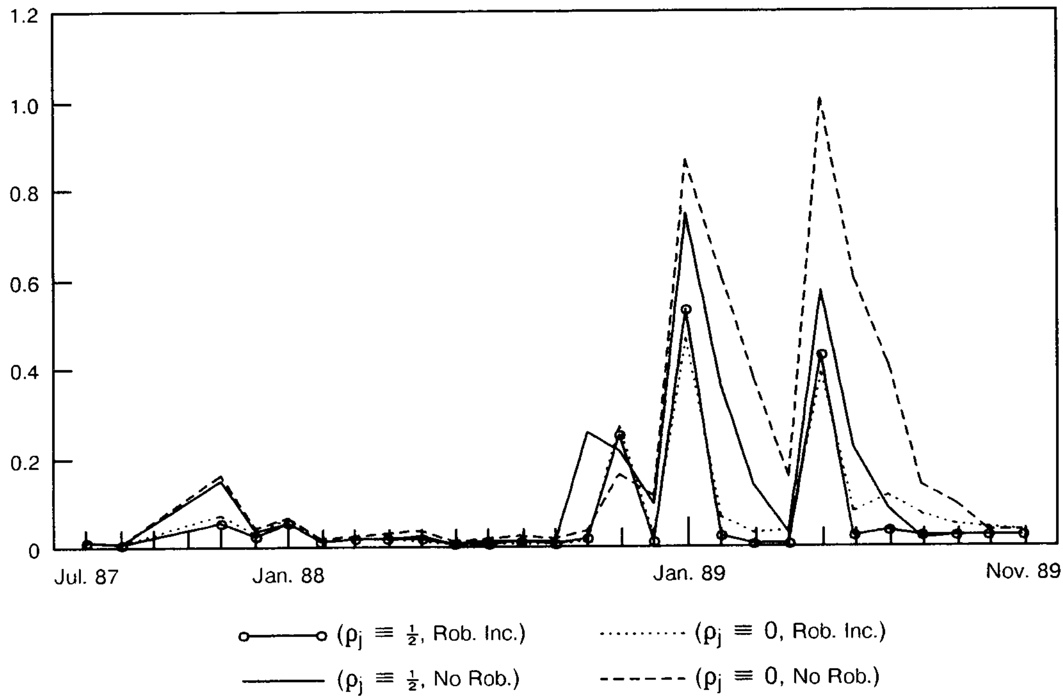

Figure 3 Monthly Mean Squared Errors of Innovations, 5 Room Apartments

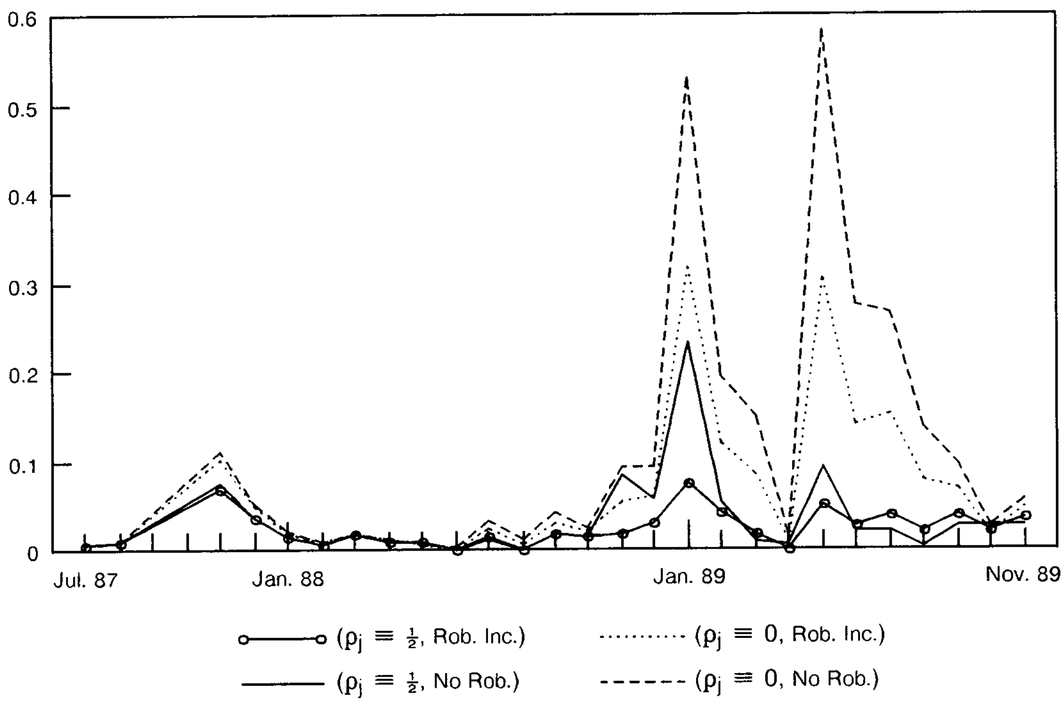

Figure 4 Monthly Mean Squared Errors of Residuals, 5 Room Apartments

for the residuals, once the linear constraints are included, there is practically no increase in the MSE values in the months of inflation in the case of 3 room apartments and, when accounting for the serial correlations, only a slight increase in the case of 5 room apartments. However, when ignoring the serial correlations, the residual MSE's for 5 room apartments are much larger in the months of inflation than in the other months even when imposing the constraints. This outcome has a simple explanation. The linear constraints are imposed on the aggregate means of the fitted values in each district but since the number of observations in 5 room apartments is a small fraction of the total number of observations, the constraints alone have a relatively small effect on the estimated regression coefficients in this cell. On the other hand, the constraints have a large effect on the estimated coefficients in the other cells so that when accounting for the cross-sectional correlations, the estimators corresponding to 5 room apartments are also modified since they are correlated with the other coefficients.

The way by which the linear constraints protect against sudden drifts in the data is illuminated in Figure 5 where we plotted the monthly intercept estimates for 3 room apartments.

As can be seen, with the linear constraints included, the intercept adapts itself to the new level of the data in the same month that the inflation occurs. Without the inclusion of the constraints, the adaption to the new level of the data takes several months. The plot of the monthly intercept estimates of 5 room apartments does not have this nice pattern since with the small sample sizes observed each month, the effect of the inflation is to alter also the other regression coefficients.

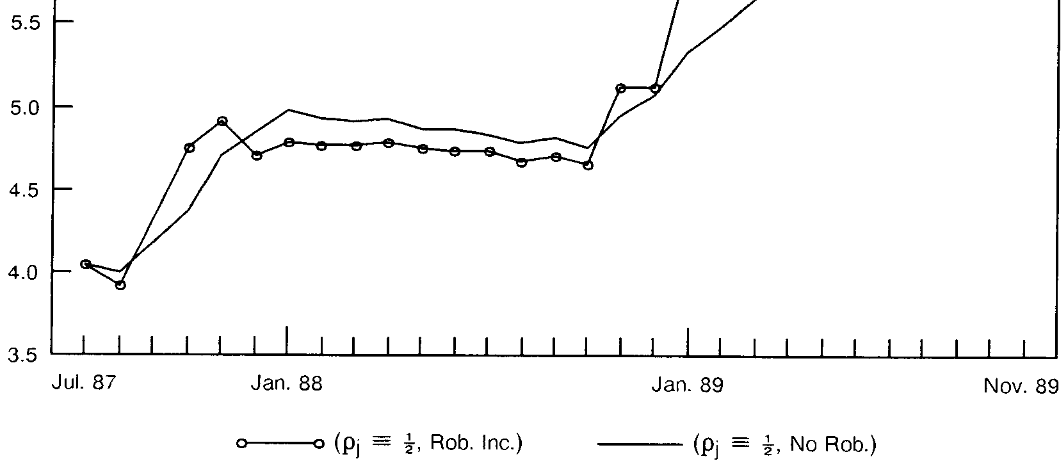

Figure 5 Monthly Estimates of Intercept, 3 Room Apartments

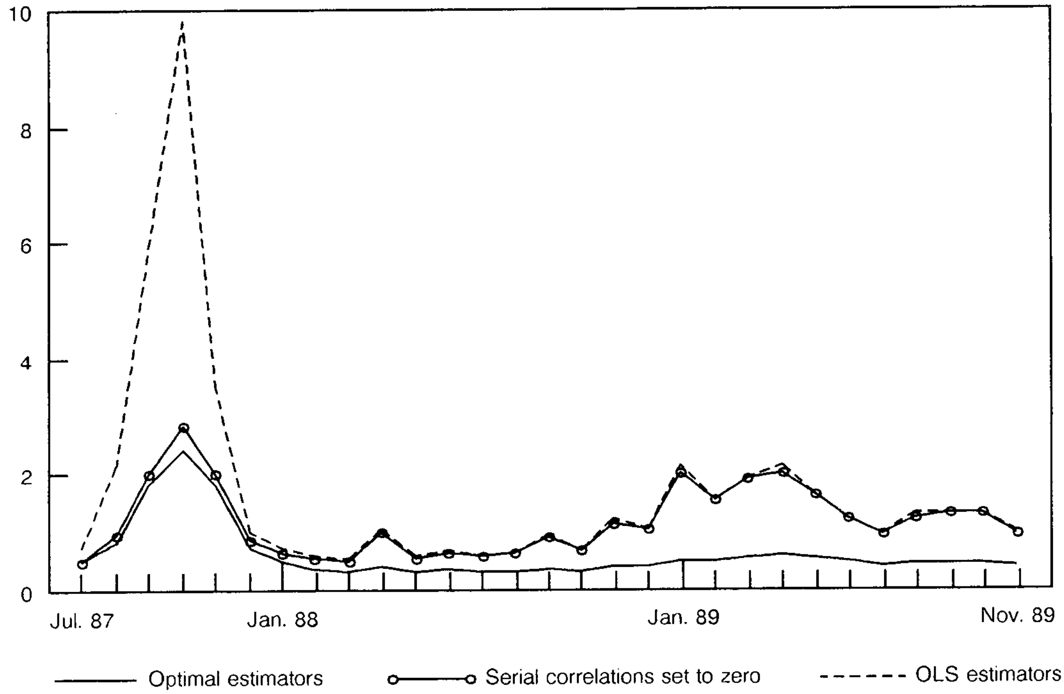

Figure 6 Variances of Estimators of Cell Means ($\times 10^4$), 3 Room Apartments

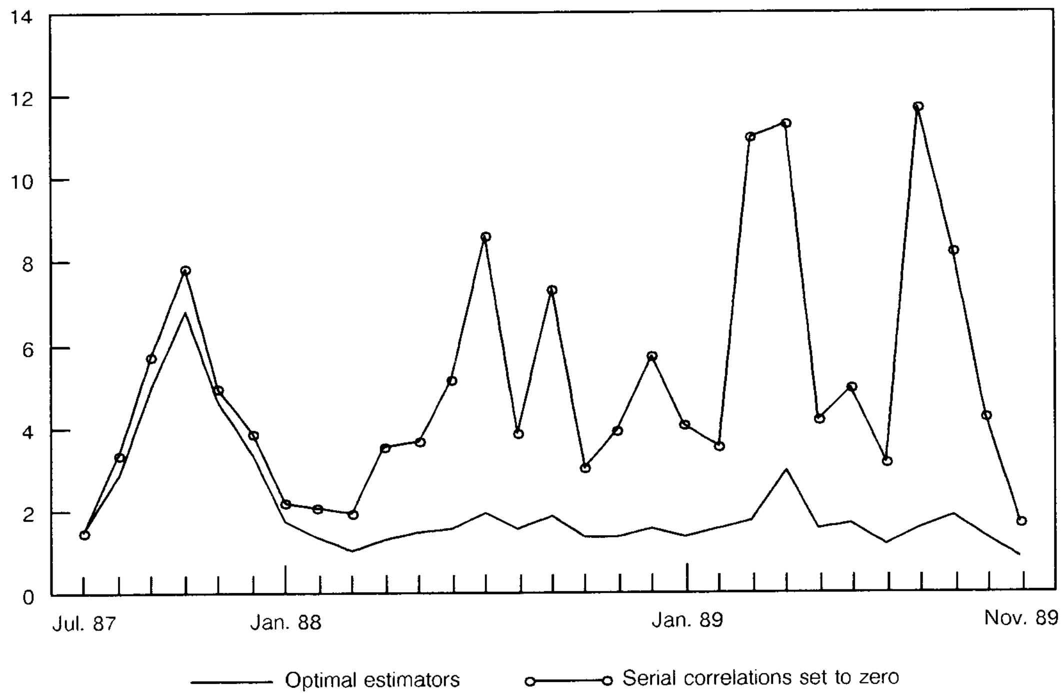

Figure 7 Variances of Estimators of Cell Means ($\times 10^4$), 5 Room Apartments

Our discussion so far centered on the empirical distribution of the model residuals and innovations. A major application of small area estimation is the prediction of the small area means (equation 2.2). Clearly, when a model yields residuals with well behaved properties it can also be expected to yield good estimators for the population means. Nevertheless, it is interesting to compare the theoretical variances of the small area means estimators as obtained with and without the accounting for the cross-sectional correlations, under the model which accounts for these correlations with $\rho_j \equiv \frac{1}{2}$. This comparison permits the assessment of the loss in efficiency when the serial correlations are ignored.

Figures 6 and 7 show the monthly variances of the cell mean estimators as obtained for 3 and 5 room apartments. (The variances have been multiplied by 10^4.) The figure for 3 room apartments also contains the variances of the ordinary least squares (OLS) estimators of the population means, that is, the variances of the estimators when estimating the regression coefficients in each month by OLS. These estimators are not operational in the case of 5 room apartments because of the very small monthly sample sizes.

The important conclusion drawn from the two figures is that by accounting for the cross-sectional correlations the variances of the resulting estimators can be reduced quite substantially, depending on the sample sizes. This is obviously the case in the case of 5 room apartments but is also true for 3 room apartments despite the fact that the sample sizes in these cells are relatively very large. The large sample sizes ordinarily obtained for 3 room apartments make the OLS estimators quite comparable to the estimators obtained when ignoring the cross-correlations in the estimation of the population means. Notice however the big gap between the variance of the OLS estimator and the variance of the other two estimators in October 1987. In this month there were only 10 observations of 3 room apartments and it is here where the use of the past data has its main impact even when ignoring the cross-sectional correlations. (The number of observations for 3 room apartments in November 1987 is 28; in all the other months there are at least 46 observations.)

Another important outcome arising from the two figures is the much greater stability of the variances of the optimal estimators under the model as compared to the variances of the estimators which ignore the cross-sectional correlations. Notice in this respect that the differences in the variances from one month to the other depend not only on the sample sizes in each month but also on the values of the explanatory variables (the design matrix) and the amount of past data observed. Still, it is the sample sizes which mostly explains the differences in the variances of the estimators particularly towards the end of the series.

ACKNOWLEDGEMENT

This article was written while the first author was on sabbatical leave at Statistics Canada under its Research Fellowship program. The authors would like to thank a referee for helpful comments.

APPENDIX

a) Derivation of Equation (2.12)

When $\underset{\sim}{x}_{tki} = \underset{\sim}{x}_{tk}$, $\hat{\theta}_{tk} = \underset{\sim}{x}'_{tk}\hat{\underset{\sim}{\beta}}_{tk} = \underset{\sim}{z}'_{tk}\hat{\underset{\sim}{\alpha}}_{tk}$ so that $\hat{\underset{\sim}{\theta}}_t = (\hat{\theta}_{t1}, \ldots, \hat{\theta}_{tK})' = Z_t\hat{\underset{\sim}{\alpha}}_t$.
Also, for the random walk model the matrix T is the identity matrix and by equation (3.1)

$$Z_t\hat{\underset{\sim}{\alpha}}_t = Z_t\hat{\underset{\sim}{\alpha}}_{t-1} + \left(Z_tP_{t|t-1}Z_t'\right)F_t^{-1}\left(\underset{\sim}{Y}_t - Z_t\hat{\underset{\sim}{\alpha}}_{t-1}\right) =$$
$$\left(I - \Sigma_tF_t^{-1}\right)\underset{\sim}{Y}_t + \Sigma_tF_t^{-1}Z_t\hat{\underset{\sim}{\alpha}}_{t-1} \tag{A1}$$

since $F_t = (Z_tP_{t|t-1}Z_t' + \Sigma_t)$. Suppose for convenience that $k = 1$ and define

$$F_t = \begin{bmatrix} f_{11}, \underset{\sim}{f}_1' \\ \underset{\sim}{f}_1, F_{22} \end{bmatrix} \quad \text{and} \quad H_t = F_t^{-1} = \begin{bmatrix} h_{11}, \underset{\sim}{h}_1' \\ \underset{\sim}{h}_1, H_{22} \end{bmatrix} \quad \text{were } f_{11} \text{ and } h_{11}$$

are scalars, $\underset{\sim}{f}_1'$ and $\underset{\sim}{h}_1'$ are $[1 \times (K-1)]$ and F_{22} and H_{22} are $[(K-1) \times (K-1)]$.
Using this notation, it follows from (A1) that

$$\hat{\theta}_{t1} = \left(1 - \frac{\sigma_1^2}{n_{t1}}h_{11}\right)\bar{Y}_{t1} + \frac{\sigma_1^2}{n_{t1}}h_{11}\left(\underset{\sim}{x}'_{t1}\hat{\underset{\sim}{\beta}}_{t-1,1}\right) - \frac{\sigma_1^2}{n_{t1}}\sum_{k=2}^{K} h_{11}\frac{h_{1k}}{h_{11}}\bar{e}_{tk}. \tag{A2}$$

Let $\gamma_1' = (\gamma_{12}, \ldots, \gamma_{1K}) = \underset{\sim}{f}_1'F_{22}^{-1}$ defines the partial regression coefficients in the regression of $\bar{e}_{t1}$ on $(\bar{e}_{t2}, \ldots, \bar{e}_{tK})$ and $v_1^2 = (f_{11} - \underset{\sim}{f}_1'F_{22}^{-1}\underset{\sim}{f}_1)$ define the residual variance in the regression.

Equation (2.12) follows directly from (A2) since

$$\underset{\sim}{f}_1'F_{22}^{-1} = -\frac{1}{h_{11}}\underset{\sim}{h}_1'; \quad \left(f_{11} - \underset{\sim}{f}_1'F_{22}^{-1}\underset{\sim}{f}_1\right)^{-1} = h_{11} \tag{A3}$$

by well known properties of the inverse of a partitioned matrix.

b) Derivation of Equation (4.4)

By (4.3),

$$\hat{\underset{\sim}{\alpha}}_t^{(A)} = \left(I - K_t^{(P)}Z_t^{(A)}\right)T\hat{\underset{\sim}{\alpha}}_{t-1}^{(A)} + K_t^{(P)}\underset{\sim}{Y}_t^{(A)}. \tag{A4}$$

Hence,

$$\hat{\underset{\sim}{\alpha}}_t^{(A)} - \underset{\sim}{\alpha}_t = \left(I - K_t^{(P)}Z_t^{(A)}\right)\left(T\hat{\underset{\sim}{\alpha}}_{t-1}^{(A)} - \underset{\sim}{\alpha}_t\right) + K_t^{(P)}\left(\underset{\sim}{Y}_t^{(A)} - Z_t^{(A)}\underset{\sim}{\alpha}_t\right). \tag{A5}$$

The prediction errors $(T\hat{\underset{\sim}{\alpha}}_{t-1}^{(A)} - \underset{\sim}{\alpha}_t)$ are independent of the residuals $(\underset{\sim}{Y}_t^{(A)} - Z_t^{(A)}\underset{\sim}{\alpha}_t)$ and so,

$$P_t^{(A)} = E\left[\left(\hat{\underset{\sim}{\alpha}}_t^{(A)} - \underset{\sim}{\alpha}_t\right)\left(\hat{\underset{\sim}{\alpha}}_t^{(A)} - \underset{\sim}{\alpha}_t\right)'\right] = Q_tP_{t|t-1}^{(A)}Q_t' + K_t^{(P)}\Sigma_t^{(A)}K_t^{(P)\prime} \tag{A6}$$

where we denote for convenience $Q_t = (I - K_t^{(P)}Z_t^{(A)})$.

By definition of the matrix $F_t^{(P)}$ (see below 4.3), equation (A6) can be written in the form

$$P_t^{(A)} = Q_t P_{t|t-1}^{(A)} - P_{t|t-1}^{(A)} Z_t^{(A)\prime} K_t^{(P)\prime} + K_t^{(P)} F_t^{(P)} K_t^{(P)\prime}$$

$$+ K_t^{(P)} \left(\Sigma_t^{(A)} - \Sigma_t^{(P)} \right) K_t^{(P)\prime} \qquad \text{(A7)}$$

which implies the relationship (4.4) by straightforward algebra.

REFERENCES

ANDERSON, B.O.D., and MOORE, J.B. (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.

ANSLEY, C.F., and KOHN, R. (1986). Prediction mean squared error for State Space models with estimated parameters. *Biometrika*, 73, 467-473.

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

BINDER, D.A., and DICK, J.P. (1989). Modelling and estimation for repeated surveys. *Survey Methodology*, 15, 29-45.

BINDER, D.A., and HIDIROGLOU, M.A. (1988). Sampling in time. *Handbook of Statistics*, Vol. 6, (Eds.), P.R. Krishnaiah and C.R. Rao, Amsterdam: Elsevier Science, 187-211.

CHOUDHRY, G.H., and RAO, J.N.K. (1989). Small area estimation using models that combine time series and cross-sectional data. *Analysis of Data in Time*. Ottawa: Statistics Canada (to appear).

COOLEY, T.F., and PRESCOTT, E.C. (1976). Estimation in the presence of stochastic parameter variation. *Economitrica*, 44, 167-184.

DIELMAN, T.E. (1983). Pooled cross-sectional and time series data: A survey of current statistical methodology. *The American Statistician*, 37, 111-122.

HAMILTON, J.D. (1986). A standard error for the estimated state vector of a State-Space model. *Journal of Econometrics*, 33, 388-397.

HARVEY, A.C. (1981). *Time Series Models*. Deddington, Oxford: Philip Allan.

HARVEY, A.C., and PHILLIPS, G.D.A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, 66, 49-58.

KITAGAWA, G., and GERSCH, W. (1984). A smoothness priors State-Space modeling of time series with trend and seasonality. *Journal of the American Statistical Association*, 79, 378-389.

JOHNSON, L.W. (1977). Stochastic parameter regressions: An annotated bibliography. *International Statistical Review*, 45, 257-272.

JOHNSON, L.W. (1980). Stochastic parameter regression: An additional annotated bibliography. *International Statistical Review*, 48, 95-102.

LaMOTTE, L.R., and McWHORTER, A. (1977). Estimation, testing and forecasting with random coefficient regression models. *Proceedings of the Section on Business and Economic Statistics, American Statistical Association*, 814-817.

MADDALA, G.S. (1977). *Econometrics*. Kogakusta: McGraw-Hill.

MEINHOLD, R.J., and SINGPURWALLA, N.D. (1983). Understanding the Kalman filter. *The American Statistician*, 37, 123-127.

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 BARNARD, C. (1991). Some new estimators for small area means with application to the assessment of farmland values. *Journal of Business and Economic Statistics* (to appear).

PFEFFERMANN, D., BURCK, L., and BEN-TUVIA, S. (1989). A time series model for estimating housing price indexes adjusted for changes in quality. *Analyisis of Data in Time*. Ottawa: Statistics Canada.

PFEFFERMANN, D., and SMITH, T.M.F. (1985). Regression models for grouped populations in cross-section surveys. *International Statistical Review*, 53, 37-59.

ROSENBERG, B. (1973a). The analysis of cross section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement*, 2, 399-428.

ROSENBERG, B. (1973b). A survey of stochastic parameter regression. *Annals of Economic and Social Measurement*, 2, 381-397.

SÄRNDAL, C.E., and HIDIROGLOU, M.A. (1989). Small domain estimation: A conditional analysis. *Journal of the American Statistical Association*, 84, 266-275.

SCHWEPPE, F. (1965). Evaluation of likelihood functions for gaussian signals. *IEE Transactions on Information Theory*, 11, 61-70.

SMITH, T.M.F. (1978). Principles and problems in the analysis of repeated surveys. *Survey Sampling and Measurement*, (Ed.) N.K. Nawboodivi, New York: Academic Press, 201-216.

SWAMY, P.A.V.B. (1971). *Statistical Inference in Random Coefficient Regression Models*. Berlin: Springer-Verlag.

TILLER, R. (1989). A Kalman filter approach to labour force estimation using survey data. *Proceedings of the Section on Business and Economic Statistics, American Statistical Association* (to appear).

WATSON, M.W., and ENGLE, R.F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23, 385-400.