SSC Annual Meeting, June 1997
Proceedings of the Survey Methods Section

TREND ESTIMATION FOR THE CANADIAN LABOUR FORCE SURVEY

M. Yu and H. Mantel[1]

ABSTRACT

In this paper we compare two methods of trend estimation, the X11-ARIMA method and a new method proposed by Pfeffermann, Feder and Signorelli (1997), using data from the Canadian Labour Force Survey (LFS). The new method is a two step procedure designed to be used with data from a rotating panel survey. In the first step the time series of panel estimates are used to get an unbiased estimate of the autocorrelation function of the survey errors, from which an ARMA model for the survey errors is identified. In the second step a state space model, which incorporates the ARMA model for the survey errors, is fit to the series of overall survey estimates. This model decomposes the series of survey estimates into four components: survey error, a locally linear trend component, seasonal effects, and an irregular term. Thus the trend estimate may be considered to be a smoothed, deseasonalized series. We apply the method to unemployment rate series from Canadian LFS data for a number of census metropolitan areas of various sizes, and compare the resulting trend estimates to those obtained by X11-ARIMA.

KEY WORDS: Labour Force Survey; state space modelling; trend estimation.

RÉSUMÉ

Dans cet article, nous comparons deux méthodes d'estimation de la tendance, la méthode X11-ARIMA et une nouvelle méthode proposée par Pfeffermann, Feder and Signorelli (1997), en utilisant des données provenant de l'Enquête canadienne sur la population active (EPA). La nouvelle méthode est une procédure en deux étapes utilisée avec des données de sondage en panel rotatif. Dans la première étape, on utilise des séries chronologiques des estimations de panel afin d'obtenir une estimation sans biais de la fonction d'autocorrélation des erreurs d'échantillonnage, pour laquelle un modèle ARMA pour l'erreur d'échantillonnage est identifié. Dans la second étape, on ajuste un modèle espace-état qui incorpore le modèle ARMA pour les erreurs d'échantillonnage aux séries des estimations de sondage globaux. Ce modèle décompose la série d'estimations d'enquête en quatre composantes: une erreur d'échantillonnage, une composante de tendance linéaire locale, les effects saisonniers et un terme irrégulier. Ainsi, l'estimateur de tendance peut être considéré comme une série chronologique lisse désaisonnalisée. Nous appliquons cette méthode à la série du taux de chômage de l'EPA pour un certain nombre de régions métropolitaines du recensement de tailles variables et l'on compare l'estimateur de tendance résultant à ceux obtenus par X11-ARIMA.

MOTS CLÉS: Enquête canadienne sur la population active; modèlisation espace-état; estimation de tendance.

1. INTRODUCTION

In this paper we study a new method of trend estimation for the Canadian Labour Force Survey (LFS) by considering trend estimates of the unemployment rates of nine census metropolitan areas (CMAs). We compare these trend estimates to those produced by the X11-ARIMA program in terms of smoothness of the estimated trends and convergence of the trend estimates as more months of data are added.

The new method was originally proposed for the Australian Labour Force Survey by Pfeffermann, Feder and Signorelli (1997); we will subsequently refer to this as the PFS method. In studying female participation and unemployment rates for three regions of Sydney, they found that the PFS method yielded trend estimates which were smoother than those produced by X11-ARIMA. The X11-ARIMA trend estimates showed undesirable up and down movement which was due to the fact that X11-ARIMA does not account for autocorrelations in the survey error process over time. Such autocorrelations are present because of the rotating panel structure which is commonly used in labour force surveys.

The PFS method involves two steps. In the first step the time series of panel estimates are used to get unbiased estimates of the autocorrelation function of the survey errors. A key observation here is that although

[1]M. Yu and H.J. Mantel, Household Survey Methods Division, Statistics Canada, Tunney's Pasture, Ottawa, Ontario, Canada K1A 0T6, mingyu@statcan.ca, manthar@statcan.ca

the panel survey errors, $e_t^{(j)} = y_t^{(j)} - Y_t$, are not directly observable, the panel survey pseudo-errors, $e_{tp}^{(j)} = e_t^{(j)} - \bar{e}_t = y_t^{(j)} - \bar{y}_t$, are observable and, under mild assumptions of independence of survey errors for different panels, the autocovariances of the survey errors can be expressed as linear combinations of the autocovariances of the panel survey pseudo-errors. An appropriate ARMA model is then identified for the survey errors based on the estimated survey error autocorrelation function. In the second step a state space model, which incorporates the ARMA model identified at the first step for the survey errors, is fit to the series of survey estimates. This state space model decomposes the series of survey estimates into four components: survey error, a locally linear trend component, seasonal effects, and an irregular term.

The X11-ARIMA program gives a similar decomposition of the series of survey estimates but it does not distinguish between the survey error and the irregular term. Furthermore, as noted above, it does not account for possible autocorrelations in the survey error process which can result in spurious up and down motion in the estimated trend over time.

The remainder of this paper is organized as follows. Section 2 gives a detailed description of the PFS method as it applies to the Canadian LFS. Section 3 describes our empirical study and presents some results and discussion.

2. THE PFS METHOD

In this section we give a detailed description of the PFS method in the context of the Canadian LFS. We first describe the notation and assumptions. Next we describe the estimation of the survey error autocorrelation function using the pseudo-errors. Finally we describe the state space model for the series of survey estimates, which incorporates the model for the series of survey errors.

2.1 Notation and Assumptions

The Canadian LFS is a monthly household survey which uses a rotating panel survey design in which each sample dwelling remains in the sample for six consecutive months before being replaced by a new sample dwelling; thus one sixth of the sample (one panel) is replaced each month. We often refer to the six panels of the sample as rotation groups.

We regard the six LFS rotation groups as being six independent samples all drawn using the same sampling design. In fact, this assumption is not entirely valid since in some cases more than one rotation group may be represented in the same primary sampling unit; however, we assume that the inter-panel correlations are ignorable. Since non-response adjustment and weight calibration is done separately within each rotation group, the overall estimates of totals are then averages of six independent estimates. The estimate of the unemployment rate, being a ratio of estimates of totals, cannot be regarded as an average of six independent panel estimates; nevertheless, we assume that this is approximately true.

Let Y_t be the population value of a characteristic at time t and y_t the direct sample estimate derived from the data collected at time t, and so $e_t = y_t - Y_t$ is the survey error. In the PFS method, the series of survey errors $\{e_t, t=1,...,N\}$ is assumed to be stationary and the survey error autocorrelations (SEA) are defined as

$$\rho_k = Corr(e_t, e_{t-k}); \quad k=1,2,... .$$

The estimate y_t can be decomposed (to a sufficiently close approximation) as $y_t = \sum_{j=1}^{6} y_t^{(j)}/6$, where $y_t^{(j)}$ is the sample estimate based on the panel which is in the sample for the jth time at month t. Denote by $e_t^{(j)} = (y_t^{(j)} - Y_t)$ the survey error corresponding to the jth panel estimate. Then we have $e_t = \sum_{j=1}^{6} e_t^{(j)}/6$. Note that the $e_t^{(j)}$, and hence e_t, are not observable. For estimating ρ_k, Pfeffermann *et al.* (1997) define the following so called panel survey pseudo-errors:

$$e_{tp}^{(j)} = (y_t^{(j)} - y_t) = (y_t^{(j)} - Y_t) - (y_t - Y_t) = (e_t^{(j)} - e_t)$$

We also denote by $e_{t-k}^{(j,t)}$ the survey error term at month t-k for the panel corresponding to the panel enumerated for the jth time at month t (If $k<j$ then this will be the same as the panel that is enumerated for the jth time at month t, otherwise it will be different).

The following assumptions on the panel survey errors are required by the PFS method:

Assumption A: $Cov(e_t^{(j)}, e_{t-k}^{(m,t)})=0$ for $m \neq j$, $k=0,1,...$

Assumption B: $Cov(e_t^{(j)}, e_{t-k}^{(j,t)}) = \gamma_k^j$, a function of k and j but not of t.

The validity of both assumptions seems to be plausible for the Canadian LFS, see Lee (1990).

2.2 Computation and Estimation of the SEA

The PFS method of estimation of the survey error autocorrelation ρ_k is based on the following lemma from Pfeffermann *et al.* (1997).

Lemma: Let $C_k^j = Cov(e_{tp}^{(j)}, e_{t-k,p}^{(j,t)})$ denote the lag k autocovariance of the jth month in sample survey pseudo-errors. Then:

$$\rho_k = Corr(e_t, e_{t-k}) = \sum_{j=1}^{6} C_k^j / \sum_{j=1}^{6} C_0^j$$

Based on this lemma, ρ_k can be estimated by

$$\hat{\rho}_k = \sum_{j=1}^{6} \hat{C}_k^j \Big/ \sum_{j=1}^{6} \hat{C}_0^j \qquad (1)$$

where $\hat{C}_k^j = \sum_{t=k+1}^{N} (e_{tp}^{(j)} - e_p^{(j)})(e_{t-k,p}^{(j,t)} - e_p^{(j,k)})/N$, $e_p^{(j)} = \sum_{t=1}^{N} e_{tp}^{(j)}/N$ and $e_p^{(j,k)} = \sum_{t=k+1}^{N+k} e_{t-k,p}^{(j,t)}/N$. This estimate can be shown to be consistent under some regularity conditions. The estimate of the partial autocorrelation (PAC) function is derived from $\{\hat{\rho}_k\}$ using the Yule-Walker equations.

The estimated autocorrelation and corresponding partial autocorrelation functions are then used to identify an ARMA model for the survey error series. For example, in their study of female participation rates for Sydney using Australian LFS data, Pfeffermann *et al.* (1997) concluded that an autoregressive regression model of order 2 (AR(2)) was good fit for their survey error series, that is:

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + u_t \qquad (2)$$

where u_t is white noise with zero mean and variance σ_u^2. Parameter estimates for the ARMA model are derived from the autocorrelations as estimated in (1); for example, for the model (2) we have

$$\hat{\phi}_1 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}, \quad \hat{\phi}_2 = \frac{1 - \hat{\rho}_2}{1 - \hat{\rho}_1^2}\hat{\rho}_1 \quad \text{and} \quad \hat{\sigma}_u^2 = (1 - \hat{\rho}_1\hat{\phi}_1 - \hat{\rho}_2\hat{\phi}_2)\sum_{j=1}^{6}\hat{C}_0^j/5.$$

2.3 Basic Structural Model for Trend Estimation

The PFS method assumes a Basic Structural Model for the unknown population values Y_t, which consists of the following set of equations:

$$L_t = L_{t-1} + R_{t-1} + \eta_{Lt}; \quad R_t = R_{t-1} + \eta_R$$
$$S_t = -\sum_{j=1}^{11} S_{t-j} + \eta_{St} \qquad (3)$$
$$Y_t = L_t + S_t + I_t$$

where L_t is the trend level, R_t is the increment in the trend, S_t is the seasonal effect and I_t is the irregular component at time *t*. The error terms $\{\eta_{Lt}\}$, $\{\eta_{Rt}\}$, $\{\eta_{St}\}$ and $\{I_t\}$ are assumed to be independent white noise processes with zero means and variances σ_L^2, σ_R^2, σ_S^2 and σ_I^2 respectively.

Model (3) and the ARMA model for e_t are combined into a state space representation of general form

$$y_t = Y_t + e_t = (L_t + S_t + I_t) + e_t = Z\alpha_t; \quad \alpha_t = T\alpha_{t-1} + \eta_t, \qquad (4)$$

where Z and T are fixed design and transition matrices, α_t is the state vector, and η_t is a vector of independent error terms. For the AR(2) model in (2), the state vector takes the form $\alpha_t = (L_t\ R_t\ S_t\ S_{t-1} \ldots S_{t-11}\ I_t\ e_t\ e_{t-1})^T$.

The unknown model parameters σ_L^2, σ_R^2, σ_S^2 and σ_I^2 are estimated by the method of maximum likelihood. Taking the parameters of the ARMA model for the survey error process to be equal to their estimates based on the panel survey pseudo-error series, as described in Section 2.2, the Kalman Filter can be used to evaluate the log-likelihood at different values of σ_L^2, σ_R^2, σ_S^2 and σ_I^2.

Finally, current and smoothed estimates of the state vector α_t in (4), which includes the trend L_t, are produced using, succesively, the Kalman Filter and de Jong's backward smoothing algorithm.

Pfeffermann *et al.* (1997) found the model based trend estimates to be smoother than those produced by X11-ARIMA. However, after adjusting the observations y_t by subtracting the model based estimates of the survey error e_t, the X11-ARIMA based trend estimates matched the model based estimates very closely.

3. APPLICATION OF THE PFS METHOD TO THE CANADIAN LFS

In this section, we apply the PFS method of trend estimation to the unemployment rate series of the Canadian

LFS for selected CMAs. CMA level estimates were chosen for this study since it was expected that the survey errors for these estimates would be large enough to present serious problems for trend estimation. Of the 25 CMAs for which estimates are currently published, we chose the three largest, Toronto, Vancouver and Montreal, as well as a sample of six of the smaller CMAs. The other CMAs were stratified based on population size into small, medium and large strata. Calgary and Winnipeg were randomly chosen from the large stratum, London and Windsor from the medium stratum, and Saskatoon and Trois Rivières from the small stratum. The general conclusions in the next two subsections are based on all nine of the study CMAs; however, due to space restrictions results are shown for only Vancouver, Calgary and Windsor.

3.1 Estimation of survey error autocorrelation

Based on the PFS algorithm, the estimates of the survey error autocorrelations, as well as the corresponding partial autocorrelations (PAC), of the unemployment rates series for the CMAs of Windsor, Calgary and Vancouver are presented in Table 1 up to lag 11. Note that the partial autocorrelations of lag higher than 2 are very small, less than 0.05. This suggests that the survey error series is close to an AR(2) process; in fact, for some of the study CMAs an AR(1) model for the survey errors would be adequate.

Table 1. Autocorrelations and Partial Autocorrelations of Survey Errors for the Unemployment rate series in Windsor, Calgary and Toronto

CMA	Lag→	1	2	3	4	5	6	7	8	9	10	11
Windsor	SEA	.529	.424	.306	.229	.153	.088	.083	.042	.036	.023	.003
	PAC	.529	.201	.030	.008	-.022	-.035	.035	-.018	.008	.001	-.022
Calgary	SEA	.542	.352	.246	.217	.175	.159	.106	.063	.057	.026	.034
	PAC	.542	.082	.038	.075	.019	.040	-.027	-.023	.019	-.033	.026
Vancouver	SEA	.487	.324	.233	.183	.138	.088	.145	.128	.129	.146	.104
	PAC	.487	.114	.050	.040	.013	-.015	.109	.018	.036	.058	-.023

3.2 Comparison of trend estimation methods

We adopt the basic structural model (3) for the true unemployment rate series and the AR(2) process (2) for the survey error, as in Pfeffermann *et al.* (1997). Parameters of the survey error process and the basic structural model were estimated separately for each of the study CMAs' unemployment rate series, and smoothed estimates of the trends were obtained, as described in Section 2.

Figures 1, 5 and 8 present the unemployment rates, the state space model based trend estimates, and the X11-ARIMA trend estimates for the CMAs of Calgary, Vancouver and Windsor, respectively. Note that the X11-ARIMA trend estimates are much more responsive to movements in the raw series. As mentioned in the introduction, some of the movement of the X11-ARIMA trend estimates, relative to the model based estimates, may be due to the former not accounting for autocorrelations in the survey error process. On the other hand, the model based trend estimates are much smoother.

Figure 2 shows the raw series for Calgary adjusted by the state space model based estimate of the survey error series, along with the model based and X11-ARIMA based estimates of trend derived from this adjusted series. The estimated survey error for the adjusted data should be zero, and hence autocorrelations in the survey error process are not an issue, leading to the two trend estimates being practically identical. This demonstrates that if there is no autocorrelation in the survey error series then the two methods yield the same trend estimates. Nevertheless, it should be noted that even when there is no autocorrelation in the survey error series, the PFS approach gives a more informative decomposition of the series since it gives separate estimates of the survey error and the irregular term in the state space model; this is possible because the PFS method uses information in the panel estimates to estimate the survey error variance.

To investigate the change in the trend estimates as more data become available, we derived the state space model based and X11-ARIMA based estimates of trend using successively more data. Data from the following four time periods were used: from January 1987 up to February 1993, up to August 1993, up to February 1994, and the full data from January 1987 up to August 1994. The model based estimates for Calgary, Vancouver and Windsor are presented in, respectively, Figures 3, 6 and 9. The corresponding X11-ARIMA based estimates are given in Figures 4, 7 and 10, respectively. In general, the X11-ARIMA based trend estimates settle down to their final value more quickly than the model based estimates. This is not surprising in light of the much greater smoothness of the model based estimates; the price of this smoothness is that the value of the trend estimate for a particular point in time is affected by observations in the raw series that are quite far removed in time.

REFERENCES

Dagum, E.B. (1988). *The X-11 ARIMA/88 Seasonal Adjustment Method--Foundations and Users' Manual.* Time Series Research and Analysis Division, Statistics Canada. Ottawa

Lee, H. (1990). Estimation of Panel Correlations for the Canadian Labour Force Survey. *Survey Methodology*, 16, 283-292.

Pfeffermann, D., Feder, M. and Signorelli, D. (1997). Estimation of Autocorrelations of Survey Errors with Application to Trend Estimation in Small Areas, *Journal of Business and Economic Statistics* (to appear).

Figure 1. Raw Series, Model trend and X-11 Trend for Calgary

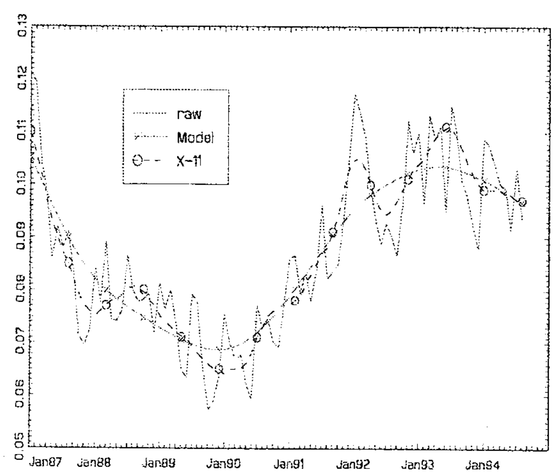

Figure 2. Adjusted Raw Series, Model trend and X-11 Trend for Calgary

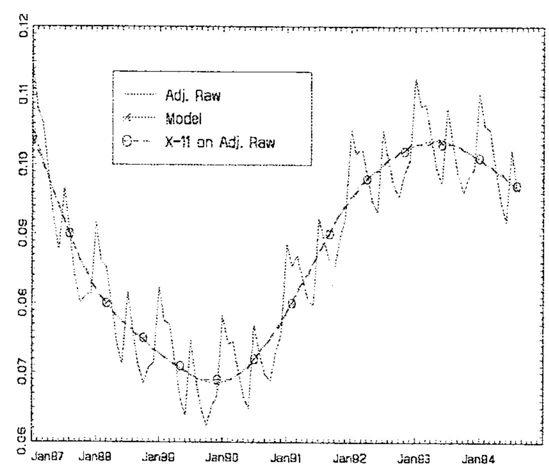

Figure 3. Model trends based on different periods for Calgary

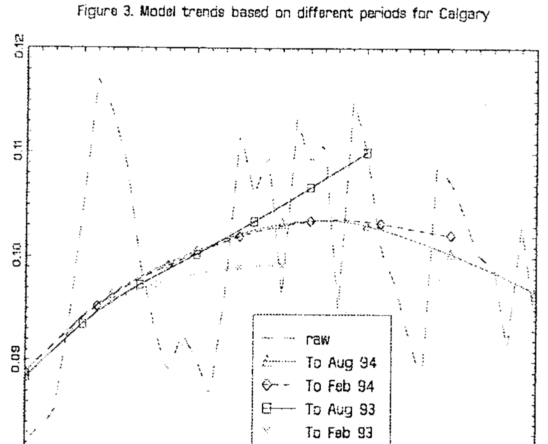

Figure 4. X-11 trends based on different periods for Calgary

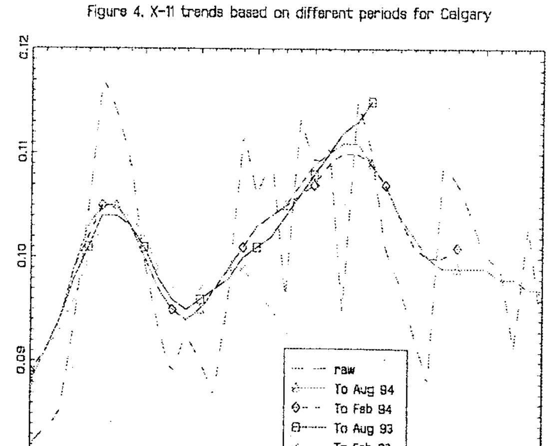

Figure 5. Raw Series, Model trend and X-11 Trend for Vancouver

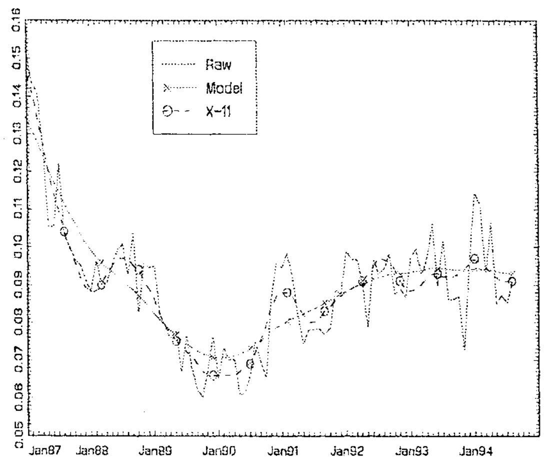

Figure 6. Model trends based on different periods for Vancouver

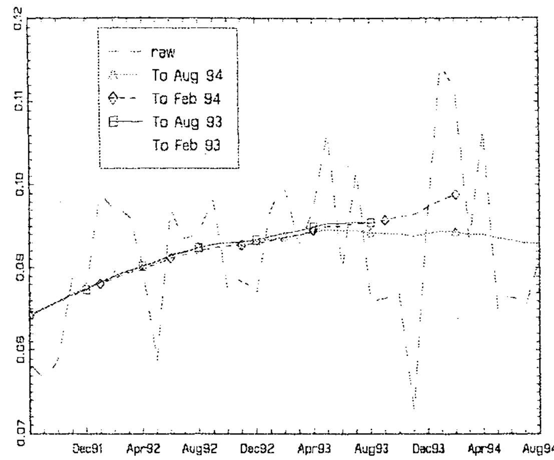

Figure 7. X-11 trends based on different periods for Vancouver

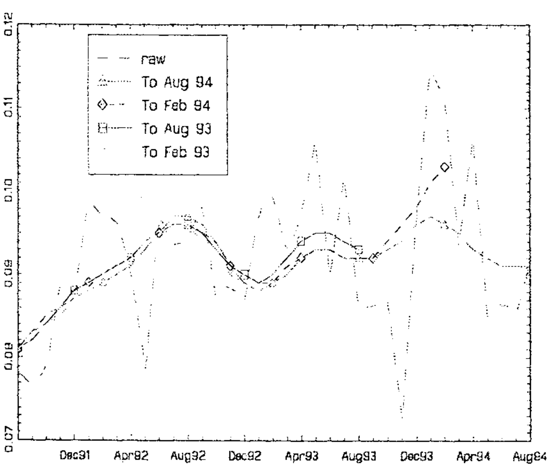

Figure 8. Raw Series, Model trend and X-11 Trend for Windsor

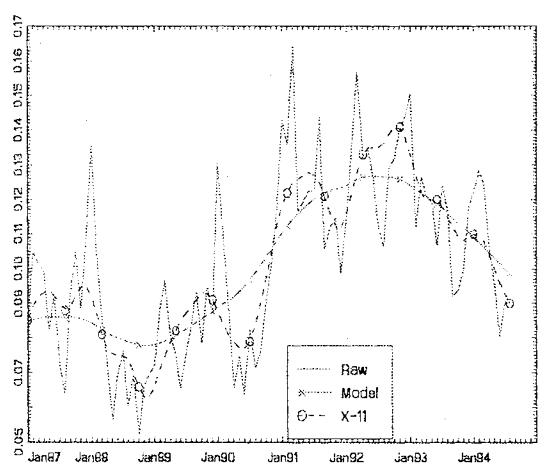

Figure 9. Model trends based on different periods for Windsor

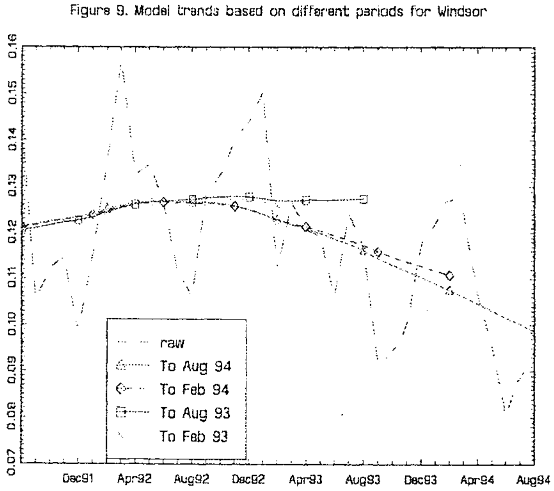

Figure 10. X-11 trends based on different periods for Windsor