Dynamics and Regulation of Bulk Milk Somatic Cell Counts

Ynte H. Schukken, Alfons Weersink, Kenneth E. Leslie and S. Wayne Martin

ABSTRACT

Somatic cell count (SCC) in milk is inversely related to dairy cow productivity and milk quality. In an effort to improve product quality, and indirectly farm productivity, regulatory limits on somatic cell counts have been established by many of the major dairy producing countries. The purpose of this paper was to assess the impact of regulations on bulk milk somatic cell counts in Ontario and to assist producers in meeting regulatory limits through development of prediction models. Through the use of a transfer function model, provincial SCC was found to have dropped by approximately 60,000 as a result of the reduction program. Limits of the regulatory program, seasonality and herd characteristics were found through time series cross-sectional models to have an impact on prediction of SCC at the farm level, but the major influence was historical SCC levels.

RÉSUMÉ

Le comptage des cellules somatiques (CCS) dans le lait est inversement relié à la production et la qualité du lait. Dans le but d'améliorer la qualité du lait, et indirectement la productivité des fermes laitières, plusieurs pays producteurs de lait ont réglementé les limites supérieures du comptage des cellules somatiques. Le but de cet article est d'évaluer l'impact des réglementations du comptage cellulaire du réservoir de lait en Ontario et d'aider les producteurs à rester en deça des limites imposées grâce au développement de modèles de prédiction. En utilisant un modèle combinant l'analyse de régression et un modèle de série chronologique (« transfer function model »), on note une réduction d'environ 60,000 cellules/mL du CCS provincial suite au programme de réduction du CCS. À l'aide de modèles de séries chronologiques transversales on a démontré une influence des limites du programme de réglementation, de l'effet saisonnier et de certaines caractéristiques du troupeau sur la prédiction du CCS au niveau de la ferme, mais la principale influence était le comptage cellulaire observé dans le passé sur la ferme. (*Traduit par Dr Emile Bouchard*)

INTRODUCTION

Dairy cow productivity improves with a lowering of the somatic cell count (SCC) in milk due to the negative correlation between SCC and milk, fat, lactose and casein production (1,2). A low level of bulk milk SCC is also associated with a low prevalence of infection with major mastitis pathogens. Thus, SCC is used as an indicator of the udder health status of cows in the herd (3). In addition to the productivity effects on the farm, a low SCC also increases milk quality and dairy product yields (4). The SCC levels have recently been found to be closely related to milk quality aspects such as plate loop count, freezing point and the presence of inhibitors (5).

In an effort to improve product quality, and indirectly farm productivity, regulatory limits on milk quality, as measured by bulk milk somatic cell count, have become more stringent world wide. In the European Community, the regulatory limit was set at 400,000 cells per mL starting January 1992. In the United States, the federal regulatory limit is currently 1,000,000 cells per mL but decreases to 750,000 in July 1993. In Ontario, a six year step-wise program to lower the regulatory limits from 800,000 to 500,000 was implemented in August 1989. Despite the extent of these programs, no work has been done to evaluate their effectiveness in lowering SCC levels or to develop prediction models to assist producers in meeting SCC limits.

The first objective of this paper is to assess the impact of regulations on bulk milk somatic cell counts in Ontario. Since bulk milk SCC is not a random measurement, understanding the dynamics behind the patterns is necessary to evaluate the effectiveness of quality restrictions and to assist producers in meeting the regulatory limits. The second objective is to determine the effect of individual farm characteristics on the movement of monthly SCC levels. When these movements are understood, it may be possible to predict the next month's SCC reading of a farm based on the previous observations. Accurate prediction would allow preventive measures to be installed before the actual penalty situation has occurred. In order to understand the dynamics of SCC and assess the impacts of the quality program, a unique data set consisting of monthly observations for a six year period on approximately 9,500 farms is used. After a description of the data, the transfer function and time series-cross-sectional econo-

Department of Population Medicine (Schukken, Leslie, Martin) and Department of Agricultural Economics and Business (Weersink), University of Guelph, Guelph, Ontario N1G 2W1.

The financial support of the Bureau of Veterinary Drugs, Health and Welfare Canada, the Ontario Ministry of Agriculture and Food and the Ontario Milk Marketing Board is acknowledged.

Submitted July 9, 1992.

metric methods used to respectively analyze the provincial and pooled data are discussed. Results are then presented followed by conclusions and implications for agencies imposing milk quality regulations and farmers responding to them.

MATERIALS AND METHODS

All milk produced in Ontario is evaluated for its milk quality at the Central Milk Testing Laboratory. From this laboratory, monthly data was obtained on bulk milk somatic cell count, from approximately 9,500 farms over the period January 1985 through September 1991. Data on milk volume and herd characteristics for these farms over the same time span were obtained from the Ontario Milk Marketing Board (OMMB) which sells all milk that is produced in Ontario. These data included kilos of milk sold, herd location, milking system, make of milking equipment, main cattle breed and herd size.

Factors affecting somatic cell count were analyzed through two approaches. The first method examined the monthly provincial SCC average and used a combination of regression and time series analysis. The second approach examined the monthly herd SCC average and subsequently used models that combine time series and cross-sectional data. Both methods are described below.

TRANSFER FUNCTION ANALYSIS

Provincial mean somatic cell counts (SCC) are assumed to be a function of its regulatory limits. These limits, denoted by the variable REGSCC, were 800,000 between August 1989 and August 1990 and 750,000 between August 1990 and August 1991, and 700,000 afterwards. No formal regulations were imposed in Ontario on SCC levels before August 1989. However, it was assumed for analytical purposes that the limit was 1,000,000 cells/mL since the current regulatory limit in the United States is set at this level. In addition to these restrictions, monthly SCC levels have been shown to follow a seasonal pattern that may be captured by fitting a sine and a cosine function to the data (5). This allows a sinusoidal seasonal pattern where the amplitude and the starting point of the sine wave is estimated from the data. The resulting regression equation is then:

$$SCC_t = \beta_0 + \beta_1 REGSCC + \beta_2 sine\left(2\pi\left(\frac{Month}{12}\right)\right) + \beta_3 \cos\left(2\pi\left(\frac{Month}{12}\right)\right) + \epsilon_t \quad [1]$$

where month is 1...12 for January... December, ϵ_t is the additive error term that accounts for the variance in SCC not accounted for by the regulatory limits (REGSCC) and the seasonality variables, and β_0, β_1, β_2, β_3 are the parameters to be estimated. The estimated parameters can be used to forecast the provincial mean somatic cell count with one source of forecast error being the additive noise term ϵ_t. This error term is not likely to be strictly white noise given the biological nature of somatic cells and the presence of unexplained systematic behavior in the residual will increase the forecast error of the regression equation.

Information regarding the future values of ϵ_t can be provided by time series analysis. As opposed to the structural model in equation [1] which relates SCC to a set of other explanatory variables in a causal framework, time series models base predictions of the dependent variable solely on the past behavior of that variable. The behavior may be described by a moving average model involving a weighted sum of current and lagged random disturbances, an autoregressive model involving a weighted sum of past values of the dependent variable and a random disturbance term or a combination of the two models.

A mixed auto-regressive-moving average model can be used to model the residual series of the regression equation [1]. The stochastic process associated with the residual ϵ_t is assumed to be a function of both lagged random disturbance terms and its past values,

$$(1 - \Theta_1\beta - \Theta_2\beta^2 - \ldots - \Theta_p\beta^p)\eta_t = (1 - \Phi_1\beta - \Phi_2\beta^2 - \ldots - \Phi_q\beta^q)\epsilon_t \quad [2]$$

$$or, \quad \Phi^{-1}(\beta)\,\Theta(\beta)\eta_t = \epsilon_t$$

when η_t is a normally distributed error term, and $\Phi(\beta)$ and $\Theta(\beta)$ are the lagged polynomials of the autoregressive and moving average parameters respectively. Such a time series model provides some explanation regarding the future values of ϵ_t and thus can serve to reduce forecast error.

Substituting the above time series model of the variance of SCC not explained by the structural regression into equation [1] results in the following:

$$SCC_t = \beta_0 + \beta_1 REGSCC + \beta_2 sine\left(2\pi\left(\frac{Month}{12}\right)\right) + \beta_3 \cos\left(2\pi\left(\frac{Month}{12}\right)\right) + \Phi^{-1}(\beta)\Theta(\beta)\eta_t \quad [3]$$

The combination of regression analysis with a time series model is referred to as a transfer function model (6). The parameters of the structural regression equation in the transfer equation, β, are estimated simultaneously with the parameters of the time series model, Φ *and* Θ. The lag length on these time series parameters can be determined through an analysis of the residuals of the structural model (7).

TIME SERIES-CROSS-SECTIONAL ANALYSIS

In order to determine the effect of individual farm characteristics on the movement of SCC levels, data on approximately 9,500 farmers over a six year period were analyzed. The use of longitudinal data requires specifying a model which adequately allows for differences in behavior over cross-sectional units or farmers as well as differences over time for each producer. In general such a model can be written as:

$$SCC_{it} = \beta_{1it} + \sum_{k=2}^{K} \beta_{kit} X_{kit} + e_{it} \quad [4]$$

where i=1,2..., N refers to a cross-sectional unit or producer, and t=1,2,..., T refers to a given time period or month. Thus, SCC_{it} is the somatic cell count for producer$_i$ in

month t and X_{kit} is the value of the k_{th} independent variable for the same producer in month t. The error term e_{it} for an individual producer is assumed to have a zero mean and constant variance (8).

Various restrictions are often imposed on the coefficients of the above general model. The most common is to assume a variable intercept term and constant slope coefficients across individuals. Differences in producer behavior are then assumed to be captured by the varying intercept. This model can be written as (8):

$$SCC_{it} = \bar{\beta}_1 + \mu_i + \sum_{k=2}^{K} \beta_k X_{kit} + e_{it} \quad i = 1,2,...N \text{ and } t = 1,2,...,T \qquad [5]$$

where the intercept for the ith producer, β_{1i}, is now equal to the mean intercept $\bar{\beta}_1$, and the difference between this mean and the intercept for the ith producer, μ_i.

The appropriate estimation procedure for the model given by equation [5] depends upon whether the μ_i are assumed to be fixed or random. If the μ_i are fixed, then dummy variables should be specified for each of the cross-sectional units used. If the μ_i are random, then it is an error components model rather than a covariance model. Given the large size of the cross-section and the assumption that μ_i and e_1 are uncorrelated, then the error components model is most applicable (8). The covariance matrix for the vector of random errors in such a model can be expressed as:

$$V = E[(\mu \otimes j_t + e)(\mu \otimes j_t + e)'] \qquad [6]$$

where $\mu = (\mu_1, \mu_2, ... \mu_N)'$, $\otimes$ is the Kronecker product, $\mathbf{j}_t = (1,1,...,1)'$ is a (Tx1) vector and $\mathbf{e} = (e_1, e_2, ...e_N)$. The best linear unbiased estimates of the coefficients, $\hat{\beta}$, with such an error structure are obtained from generalized least squares via:

$$\hat{\beta} = (XV^{-1}X)^{-1}X^{1}V^{-1}SCC \qquad [7]$$

The variance components in the V of the generalized least square estimator of equation [7] were estimated by the fitting of constants method of Fuller and Battese (9).

TABLE I. Provincial bulk somatic cell count (SCC) transfer function regression results

Explanatory variables	Estimate	Standard error	T-ratios
Intercept	171.595	26.869	6.386
REGSCC[a]	0.180	0.029	6.582
Sine (2π * Month/12)[b]	−25.694	4.427	−5.805
Cos (2π * Month/12)[b]	15.764	4.459	3.535
Θ_1(lag 1)[c]	−0.370	0.107	−3.470
Φ_1(lag 3)[d]	0.258	0.111	2.320
$\bar{R}^2$	0.494		
F value	27.055		
Sample size	81		

[a]Ontario regulatory limits on SCC
[b]Month (1=January,...,12=December)
[c]Moving average of past random disturbances
[d]Autoregressive process; weighted average of past observations

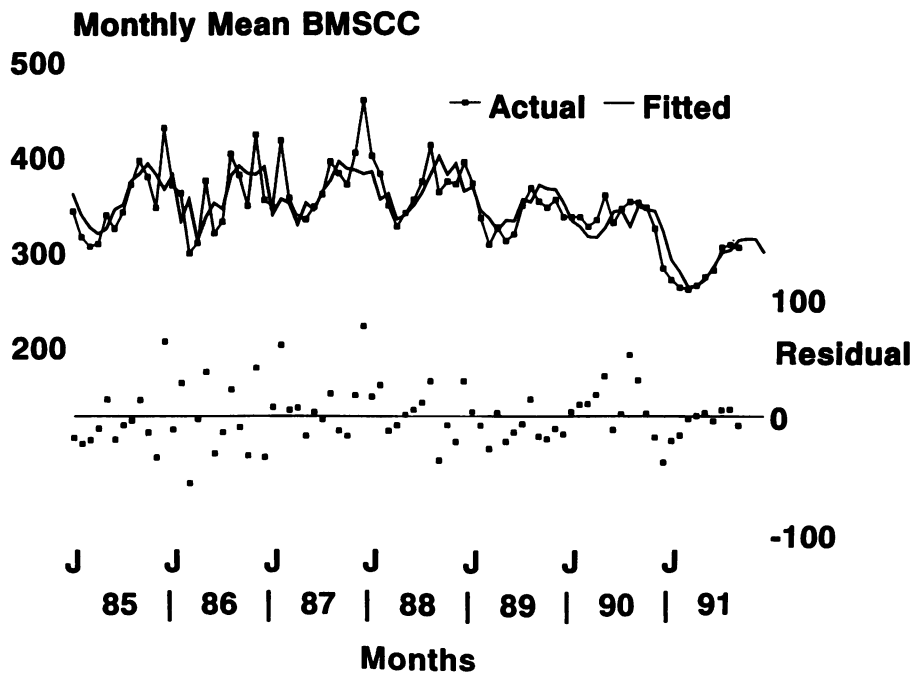

Fig. 1. Actual and predicted Ontario monthly SCC values.

Three variations of the error components model [equation 5] were estimated to determine the effect of individual farm characteristics on the movement of monthly SCC levels and to develop an individual herd SCC prediction model which can be used by regulatory agencies to assist producers in meeting SCC limits. Model 1 utilizes only historical cell count data and would thus be easily employed in herd SCC prediction. Model 2 is the most parsimonious model, in which all variables are significant. Model 3 is the full model containing the regulatory limits (REGSCC) and seasonality variables (sine and cosine) variables as in the provincial transfer function model, plus herd characteristics such as herd size, milking system, breed and autoregressive terms. The models were estimated using the TSCREG procedure in SAS (10) with 700 randomly selected farms of the 9,500 given computational difficulties.

The major purpose of the time series cross-sectional models would be to predict the next month's SCC reading for an individual farm so that the producer may install preventive measures to avoid penalties. Thus, the first two of the three models were also

TABLE II. Pooled herd bulk somatic cell count regression results

Explanatory variables	Model 1	Model 2	Model 3
Intercept	7.24 (0.25)[a]	13.58 (0.52)	7.91 (0.29)
REGSCC[b]	0.075 (2.38)	0.082 (2.91)	0.082 (2.09)
Sine(2π*Month/12)[c]		−18.65 (−4.09)	−18.66 (−4.07)
Cos(2π*Month/12)			−0.90 (−0.20)
Herd size (# of cows)			0.049 (0.76)
Pipe line (yes=1, no=0)		−16.01 (−4.71)	−16.95 (−4.71)
Breed (Holstein=1, Others=0)			4.91 (1.02)
SCC_{-1}[d]	0.429 (91.41)	0.428 (91.28)	0.428 (91.26)
SCC_{-2}	0.115 (22.78)	0.115 (22.72)	0.115 (22.71)
SCC_{-3}	0.070 (14.94)	0.070 (14.93)	0.070 (14.91)
SCC_{-6}	0.066 (16.53)	0.066 (16.49)	0.066 (16.45)
SCC_{-12}	0.115 (28.96)	0.115 (28.78)	0.114 (28.74)
Degrees of freedom	45,070		
Variance component for			
cross-sections	912.47	901.79	905.28
Time series	823.42	645.97	652.84
Error	26106.36	26105.45	26106.03

[a]t-ratios in parentheses
[b]Ontario regulatory limits on SCC
[c]Month (1=January...12=December) of observation
[d]SCC_{-i} is the herd SCC i months previously

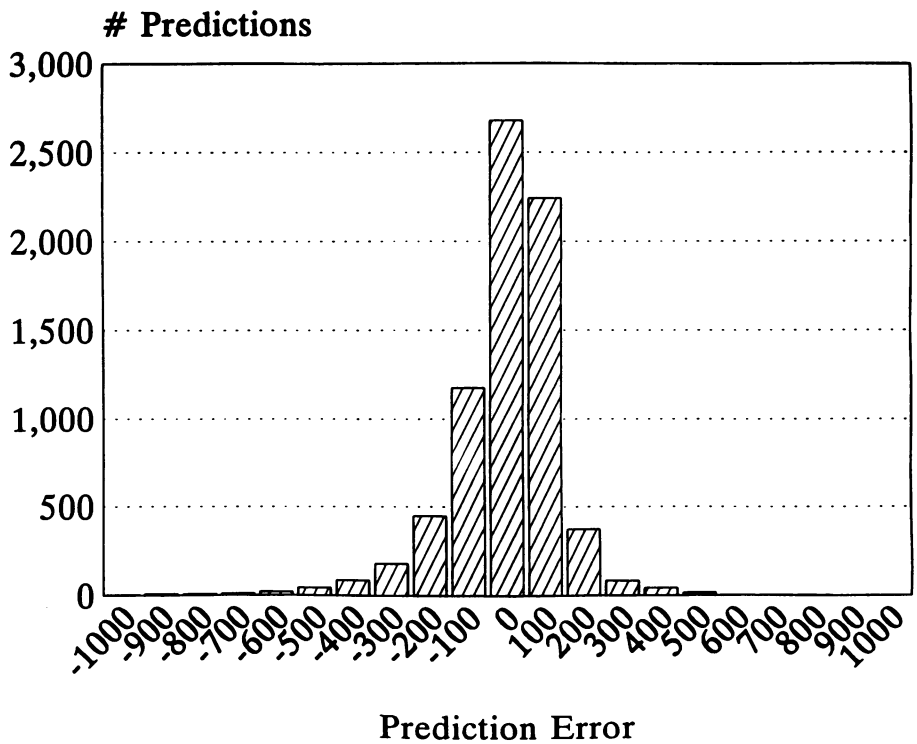

Fig. 2. The prediction error in predicting monthly SCC using Model 2.

evaluated on the basis of their prediction performance relative to a naive prediction where the SCC for a herd next month is assumed to be equal to this month's level. Prediction was done for a random sample of approximately 3500 herds that were not used for the estimation process.

RESULTS

TRANSFER FUNCTION ANALYSIS

The provincial somatic cell count (given in Appendix 1) was first estimated using the structural regression equation [1] with the regulatory limit and seasonality as the explanatory variables. However, autocorrelation was a significant problem in this estimation as indicated by the Chi-square statistic for white noise ($p < 0.001$). Only after fitting the transfer function model [equation 3] with an autoregressive term of lag 3 and a moving average term of lag 1 was the assumption of white noise met as determined by a nonsignificant Chi-square statistic ($p = 0.9$). The resulting model given in Table I explained approximately 50% of the variability in Ontario's somatic cell count. The fit of the model is illustrated in Fig. 1.

The plot of the actual and predicted monthly SCC values in Fig. 1 shows a strong seasonality pattern. This was captured in the model by the sine and cosine variables which were found to be significant and possess realistic signs (Table I). The lowest SCC was expected in April and the highest in October. Figure 1 also illustrates how the Ontario regulatory program initiated in August 1989 has lowered provincial SCC levels. The effect of implementing the regulatory limit for SCC was an approximate decrease of 40,000 cells in the first year of the program and 10,000 cells in every year thereafter. These estimates were obtained by multiplying the regression coefficient in Table I (0.180) by the reduction in the SCC limit in the program. In the first year the reduction was 200 (*1000), and it was 50 (*1000) in the years thereafter.

TIME SERIES-CROSS-SECTIONAL ANALYSIS

The results of the three cross-sectional models are reported in Table II. In both models 1 and 2, the impact of the regulatory program on SCC was an approximate decrease of 15,000 cells, while every next year the SCC decreased by approximately 4,000 cells. These estimates are similar to the provincial SCC model obtained from the transfer function model. However, the impact of the program appears to be smaller with the individual herd data (400,000 versus 15,000, and 10,000 versus 4,000). The estimates in the cross-sectional analysis were obtained after correcting for autoregressive terms. This correction may have decreased the apparent impact of the SCC reduction

TABLE III. Prediction performance of naive model, model 1 and model 2[a]

Parameter	Naive model[b]	Model 1[b]	Model 2[c]
	— Regressing predicted SCC in actual SCC —		
Model R^2	0.816	0.852	0.854
Parameter	0.920	1.000	1.027
		— Prediction error —	
Mean	−8.55	6.35	−3.07
Standard deviation	169.42	149.75	149.19
Relative efficiency	1.0	1.24	1.29
Range	2932	2573	2565

[a]Predictions are on 7411 farms
[b]Prediction of individual herd SCC in month i+1 is equal to level in month i
[c]Parameter estimates given in Table II

program. The presence of a pipeline decreased the farm level SCC by 16,000 cells. This finding is consistent with previous studies. Most dairy farms currently do have pipelines, thus the practical implication of this finding is limited. The term for seasonality (sine) was very similar to the provincial model, with the lowest SCC expected in April and the highest in October. The autoregressive terms were very important in the final models. The first three autoregression terms (lag 1,2,3) are in agreement with the findings in the provincial model. The second two autoregression terms (lag 6 and 12) may reflect repeated annual events on individual farms. For example, some cows with a high SCC may calve every year with a calving interval of approximately 12 months, forcing an annual pattern into the data.

The prediction performance of models 1 and 2 and a naive prediction where the predicted SCC was set equal to the herd's SCC level in the previous month are shown in Table III. All three predictions performed reasonably well since the R^2 of all regressions was greater than 80%. However, model 2 predicted slightly better than model 1, and both these explanatory models predicted substantially better than the naive prediction in terms of prediction error. The relative efficiency (ratio of the variances) was approximately 30% higher in model 1 and 2 versus the naive prediction. Models 1 and 2 were also approximately unbiased since their regression parameters equalled 1.0, while the simple prediction overestimated the herd's SCC level (regression parameter = 0.9). In Fig. 2, a histogram of the prediction error using the predictions of model 2 is shown indicating that the prediction accuracy will usually be within the range of +200 to −200 from the true SCC value.

DISCUSSION

Regulatory limits on somatic cell counts (SCC) in milk have been recently established by many major dairy producing regions in an effort to improve milk quality and herd productivity. Using provincial level data and a transfer function model, it was found that the regulatory program has had a substantial impact on the SCC levels in Ontario. After correction for seasonality and autocorrelation, the program has resulted in a drop of the provincial SCC of approximately 60,000 cells. This represents a substantial effort on behalf of the entire dairy industry. The resulting model may be used to predict future Ontario SCC levels which are useful to guide the industry in the current competitive dairy market.

Prediction of SCC on the farm level based on time series-cross-sectional analysis was found to be mostly dependent on the previous performance of the same farm. Only the regulatory program, seasonality, and the presence of a pipeline, were external sources of information. The final model performed reasonably well in predicting the farm SCC. However, the standard deviation in the prediction error was still approximately 150,000 cells. The fitted models may be utilized to estimate a predicted SCC, when the most current reading is communicated to the farmer. When the predicted SCC is over the regulatory limit, preventive measures can be applied before the regulatory limits are exceeded.

REFERENCES

1. **BARTLETT PC, MILLER GY, ANDERSON CR, KIRK JH.** Milk production and somatic cell count in Michigan dairy herds. Dairy 1990; 73: 2794–2800.
2. **DeLUYKER H.** Milk yield fluctuations associated with mastitis. Flemish Vet J 1991; 62: 207–222.
3. **EMANUELSON U, FUNKE H.** Effect of milk yield on the relationship between bulk milk somatic cell count and prevalence of mastitis. Dairy 1991; 74: 2479–2483.
4. **MITCHELL GE, RODGERS SA, HOULIHAN DB, TUCKER VC.** The relationship between somatic cell count, composition and manufacturing properties of bulk milk. Aust J Dairy Technol 1986; 41: 9–14.
5. **SCHUKKEN YH, LESLIE KE, WEERSINK AJ, MARTIN SW.** Ontario bulk milk somatic cell count reduction program. J Dairy Sci 1992; 75: 3352–3358.
6. **BOX GEP, JENKINS GM.** Time Series Analysis. San Francisco: Holden-Day, 1970.
7. **PINDYCK RS, RUBINFELD DL.** Econometric Models and Economic Forecasts. 3rd ed. New York: McGraw-Hill, 1991.
8. **JUDGE GC, GRIFFITHS WE, HILL RC, LSUTKEPOHL H, LEE T.** The Theory and Practice of Econometrics. 2nd ed. New York: John Wiley and Sons, 1985.
9. **FULLER WA, BATTESE GE.** Estimation of linear models with crossed-error structure. Econometrica 1974; 2: 67–78.
10. **SAS INSTITUTE INC.** SUGI Supplemental Library User's Guide, 5th edition. Cary, North Carolina, 1986.

Appendix 1. Monthly SCC values in Ontario, 1985–1991

	85	86	87	88	89	90	91
Jan	342	369	347	400	371	336	270
Feb	315	361	416	381	335	336	262
Mar	305	298	356	348	307	326	260
Apr	308	309	336	326	325	333	264
May	338	374	334	340	311	359	273
June	324	319	347	354	318	330	280
July	341	331	360	373	348	345	304
Aug	370	402	394	411	366	352	307
Sept	395	380	382	362	352	351	304
Oct	378	348	370	373	346	346	
Nov	346	422	403	370	354	324	
Dec	429	354	458	393	336	282	