Pork

1980



Series E, Agricultural Economics

# A RANDOM PARAMETER REGRESSION APPROACH TO ESTIMATING A NORTH AMERICAN PORK SUPPLY MODEL

by

Bruce L. Dixon, Larry J. Martin

and Ellen W. Goddard

July 1980 No. 80 E-128

Bruce L. Dixon is an Assistant Professor of Agricultural Economics at the University of Illinois; Larry J. Martin is Associate Professor of Agricultural Economics at the University of Guelph; and Ellen W. Goddard is a graduate student in Agricultural Economics at Latrobe University, Victoria Australia.

Paper presented at the AAEA annual meetings, selected papers session, Urbana, Illinois, July 28-30, 1980.

ABSTRACT

A Random Parameter Regression Approach to
Estimating a North American Pork Supply Model

Bruce L. Dixon, Larry J. Martin
and Ellen W. Goddard

Equations modelling pork supply for North America are estimated assuming coefficients change both randomly and systematically. Results show predictive superiority of time varying models over least squares for sample period. Ex post forecasting does not exhibit similar superiority. Measuring structural change and effect of additional observations on estimates are considered.

# A RANDOM PARAMETER REGRESSION APPROACH TO ESTIMATING A NORTH AMERICAN PORK SUPPLY MODEL

In their review of price analysis and outlook, Tomek and Robinson observe that demand structures can change over time. They also note that some estimates of demand model parameters often exhibit substantial sensitivity to the deletion or addition of one or two observations. They suggest that because of this latter phenomena, a random coefficients model might be appropriate for various price analysis problems. While Tomek and Robinson's comments refer to demand models, it can certainly be argued that the same estimation problems arise in estimating supply equations.

In this paper the estimation results from employing a Kalman filter type model on the estimation of four supply equations for the North American pork market are presented and analyzed. The Kalman filter model specifies that the vector of slope coefficients in a regression model changes over time due to both systematic forces and as a result of random disturbances. Thus it provides a means of coping directly with the two problems noted by Robinson and Tomek: change in the structure of the process being modelled and sensitivity of the estimates to the addition or deletion of a few observations. This latter problem is handled in the way suggested by Tomek and Robinson, i.e. random parameter regression. Tomek and Robinson further state that as of 1973, a random parameter regression technique had not been applied to agricultural commodity analysis. To our knowledge, this remains true as of early 1980.

In the first section the Kalman filter regression model is presented and aspects of parameter estimation are considered. In the second section the pork supply model is briefly discussed and the third section presents some of the more prominent and interesting results of estimation and model validation.

## KALMAN FILTER REGRESSION MODEL

There is an extensive literature on models with randomly varying parameters as discussed in Rosenberg. In random parameter models the coefficient vector for each observation is assumed to be a realization from an underlying stochastic process. Depending on the model specification, the moments of the underlying population may be assumed constant or time varying. The random parameter model in this study assumes that the observed dependent variable, $Y_t$, is related to a 1xK vector of independent variables, $X_t$, as

$$(1) \quad Y_t = X_t B_t + e_t \quad t=1, 2,---, T$$

Where $B_t$ is a kx1 vector of regression parameters and $e_t$ is a normally distributed error term with mean zero and variance $\sigma^2$. It is assumed that $E(e_i e_j) = 0, i \neq j$. The $B_t$ are specified to be unobservable but related over time as

$$(2) \quad B_{t+1} = AB_t + u_t \qquad t=1, 2,---, T-1.$$

where A is an unknown and constant matrix and $u_t$ is a normally distributed error term with mean zero and positive semidefinite covariance matrix $\sigma^2 Q$. If some of the parameters in $B_t$ are constants then the rows and columns of $\sigma^2 Q$ that correspond to these parameters are specified to be null.

From (2) it can be seen that change in $B_t$ can be attributed to two sources: random components as represented by the additive error term $u_t$ and systematic components as given by the vector $AB_t$. Thus this model differs very distinctly from other time varying models such as that of Hildreth and Houck. As noted by Sarris (1973), a first-order process as given by (2) may often be implausible but dynamic relationships of higher order may provide satisfactory time paths for the coefficients. Modelling higher order processes in the framework of (2) is straightforward as demonstrated by Sarris (1973).

If A in (2) is set equal to the identity matrix then any change in $B_t$ is due to purely random forces. Alternatively, if Q is null, then $B_t$ varies systematically over time. In the case of a null Q, if $B_1$ is considered to be constant, then all the $B_t$ are constants. When both A=I and Q is null, (1) and (2) then constitute the classical regression model to which ordinary least squares (OLS) estimation becomes appropriate. Hence when A is not the identity or Q is not null, the OLS model is inappropriate.

If $\sigma^2$, Q and A are known then the mean and covariance of $B_t$ can be estimated in a straightforward fashion. It should be emphasized that due to the time varying nature of $B_t$, estimation results in a time trajectory for $B_t$, i.e. a value of the vector $B_t$ for each time period that changes with respect to time. Sarris (1974) observes that economists rarely have good estimates of $\sigma^2$, Q and A. To circumvent this difficulty, Sarris (1974) proposes maximum likelihood estimation of $\sigma^2$, Q and A based on the normality of $u_t$ and $e_t$. The log likelihood function is nonlinear and requires an iterative solution as described in Kirsch and Wall. Given estimates of $\sigma^2$, Q and A and then considering these estimates to be constants, estimates of the mean of $B_t$ and their covariances can be computed as given in Cooley and Wall.

THE SUPPLY MODEL

The supply model estimated is a subset of the larger supply and demand model formulated by Martin and Zwart. The four supply equations estimated have similar specifications. There is one equation for the U.S. and three for Canada; one each for Western Canada, Quebec, and Ontario. Each equation has a set of seasonal dummy variables to account for variation due to the particular quarter. Also, each equation has lagged hog price, feed price and the endogenous variable lagged one period. The price lags are all five quarters except for the Quebec equation in which the lags are seven quarters.

The supply equations are estimated using quarterly data beginning with the first quarter of 1964. To test the forecasting power of the model for short term forecasts and to measure the sensitivity of the estimates to additional observations, two data sets are used. The first is 52 observations, starting with the first quarter of 1964 through the fourth quarter of 1976. The second set augments the first set by adding five additional observations from 1977 through the first quarter of 1978. Thus we generate some empirical evidence to evaluate the conjecture of Tomek and Robinson that random parameter models might be a more appropriate specification.

Since the model in (1) and (2) reduces to some interesting special cases, we specify four different general specifications. The first is using OLS to estimate the equations. This is done for the sake of comparison. The next model is the purely systematic variation (PSV) model where $B_t$ is assumed to be time varying on a deterministic path, i.e., $A \neq I$ but the covariance of the error vector in (2), $\sigma^2 Q$, is specified to be null. The next two models allow A to be different from the identity and Q to be non-null. For reasons discussed shortly, we have two such random and systematic variation models, RSV1 and RSV2 where they differ only with respect to the elements of Q.

ESTIMATION AND SIMULATION RESULTS

In theory the estimation of A and Q is sraightforward. However, between A and Q there are potentially 77 independent parameters to be estimated. The non-linear optimization routine simply cannot handle such a heavy numerical burden. Thus A and Q are specified to be diagonal which we have no *a priori* reason to reject. Even this simplification leads to substantial difficulties in estimation. Since starting values are required in the iterative procedure, we specified A=I and Q to be null, the OLS hypothesis. The diagonal components of Q displayed a prediliction toward becoming increasingly negative. To circumvent this problem the diagonal elements of Q are set at two different a priori levels which exhibited varying degrees of plausibility. In RSV1, the three components of Q relating to the three slope coefficients are specifie to be smaller than those set in RSV2 by a factor of 10. Otherwise the element of Q are identical in RSV1 and RSV2.

Estimation Results for the PSV Models

In Table 1 the large sample chi-square tests of the null hypothesis that A equals the identity matrix are given for the PSV models (the test uses a gen likelihood ratio). In Quebec the null hypothesis is rejected for both data set (52 observations and then 57 observations) and in the U.S. and Ontario the null hypothesis is not rejected for either data set. Western Canada rejects for 52 observations but not for 57 observations. Oddly enough, the hypothesis is not rejected for the models in which the $R^2$ is lowest in the OLS estimation. Thus, low $R^2$ is not necessarily an indicator of systematic change.

The estimated values of A are displayed in Table 2. Note that even when the hypothesis of A=I is not rejected, this does not mean that all the diagonal components of A will be relatively near one. In the Ontario model the componen of A associated with the hog price coefficient is .9453 for 57 observations. T is a very severe degree of change considering that the model is quarterly. A surprising result is that some components of A exhibit a rather drastic change with the incorporation of additional observations into the sample. For example the Quebec model the feed price component goes from .9141 to 1.030 with the addition of five observations. Overall, the time trajectories of the coefficie for the PVM models are theoretically satisfying since the coefficients have the proper signs and the magnitudes of the coefficients are quite plausible. Howeve there is a definite shifting of some of the estimates for the same time period as a result of the additional observations.

To test the fit of the PSV models historical simulations were conducted ov the sample period where the values of the lagged endogenous variables were thos generated by the model after the initial period. The root mean square errors (RMSE) are presented in Table 3, in the second and sixth rows. Observe that th PSV models do better than their OLS counterparts and that the degree of improvement is greatest where the hypothesis of A=I was rejected. This suggests that the PSV model may be appropriate means of accounting for structural change in some cases.

The Models with Random and Systematic Change

For the RSV1 and RSV2 specification, the components of A were reestimated with Q set as stated earlier. In general, the diagonal elements have values si to those in the PSV models but with a few glaring exceptions. Additionally, the level of some of the estimates when the five observations are added in the RSV1 model showed substantial change. These changes are evident in all four equatic so that at least for the models and data considered here, a random parameter regression has not lead to insensitivity of the parameter estimates to the addition of a few observations. A further aspect of the RSV1 and the RSV2 model is that the coefficients are not satisfactory in terms of conventional economic wisdom. Coefficients of the lagged endogenous variables sometimes become negati and the coefficients of prices also sometimes have an improper sign. This problem is much more prevalent in the RSV2 model. The reason for persisting with the RSV2 models is vividly illustrated in the third, fourth and seventh lines of Table 3. The performance of the RSV2 model in terms of RMSE and percentage RMSE is astounding. In all four of the equations the RSV2 specification appears to have substantial predictive and explanatory power. Surely if were interested in better forecasting ability, and for the last three or four yea hog price and supply forecasting track records have not been enviable, a few peculiar signs would be a small price to pay for greatly enhanced accuracy.

Forecasting Performance

Three sets of forecasts were computed. The first is a 10 period forecast f the first quarter of 1977 through the second quarter of 1979. The second foreca is for the five quarters beginning with the first quarter of 1977. The last forecast is from the second quarter of 1978 through the second quarter of 1979. This latter forecast uses updated estimates based on 57 observations. The root mean squares of these forecasts are given in Table 4. The contrast between Table 4 and Table 3 in terms of the superiority of the random variable models a time varying models is startling. None of the alternatives to OLS is consisten superior to OLS nor do any of the alternatives stand out as being superior to a of the other alternatives. However, one of the three alternative methods is usually superior to the OLS model.

The question of why the RSV models do so well in the historical period and have a much more mixed and certainly less distinctive ex post forecasting performance can be answered in the following way: first, recall that the diffe between RSV1 and RSV2 is that three of the variances in the coefficient error vector co-variance are increased. This says that there is more randomness in th period-to-period transition in the coefficient vector, i.e., given the coeffici today, we can say less about what they will be in future periods. Thus, by increasing Q, the $B_t$ are freer to move to levels that will result in the predic dependent historical variable closer to the observed $Y_t$. Hence RSV2 does bett on the historical simulation than RSV1 because the coefficients have a wider ra of values they can assume. This result is somewhat analogous to the fact that adding more variables to an OLS regression will never decrease $R^2$ whether or no they are relevant. That is, increasing Q will likely give a better fit to samp observations whether or not it is a valid specification. However, since this randomness cannot be forecasted, the RSV models do not forecast well. Thus, fo forecasting a random variable model may not do better than OLS unless there is a way to predict accurately the random components..

## CONCLUSIONS

The random and systematically varying models used in this study proved to a substantial improvement over OLS for replicating sample data. However, thei forecasting performance vis-a-vis OLS was mixed. In the case of the systemati varying model (PSV) the coefficient values were all reasonable but they display substantial sensitivity to expansion of the sample by a few additional observat This change due to the addition of observations is also evident in the RSV1 and RSV2 models. Hence the use of random variable models, at least of the type specified in (1) and (2), does not appear to solve the problem of parameter sensitivity to additional observations. Perhaps other random variable models might be more appropriate. Even though the RSV and PSV models indicate strong evidence of parameter change over time, the paterns of change are not well enou established to forecast clearly better than OLS. This may result from the actual parameter change of being of an abrupt nature rather than the smooth pattern indicated by (2). While the models discussed in this paper show promise as an alternative to OLS, much research remains to be done.

Table 1. Chi-Square Tests of the Hypothesis A=I For the Purely Systematic Variation Model*

| | U.S. | Western C. | Quebec | Ontario |
|---|---|---|---|---|
| 52 observations | 9.126 | 18.40 | 41.28 | 11.47 |
| 57 observations | 8.066 | 9.414 | 29.28 | 11.5 |

*Critical value of $\chi^2$ at 95% with seven degrees of freedom is 14.07.

Table 2. Maximum Likelihood Estimates of the Diagonal Elements of A for the Purely Systematic Variation Model*

| U.S. | | Western | |
|---|---|---|---|
| 57 obs. | 52 obs. | 57 obs. | 52 obs. |
| 1.006 | 1.002 | 1.007 | 1.018 |
| .9946 | .9894 | 1.0295 | 1.052 |
| .9851 | .9793 | .9256 | 1.047 |
| .9961 | .9945 | .9917 | .996 |
| .991 | 1.004 | .9737 | .9697 |
| 1.003 | 1.001 | .9893 | .9926 |
| .9972 | .9970 | 1.000 | .9979 |

| Quebec | | Ontario | |
|---|---|---|---|
| 57 obs. | 52 obs. | 57 obs. | 52 obs. |
| 1.023 | 1.013 | 1.008 | 1.007 |
| .9917 | 1.021 | .9751 | .9755 |
| .9709 | .9771 | .9999 | .9963 |
| 1.015 | 1.027 | 1.005 | 1.012 |
| .9863 | .9463 | .9453 | .9542 |
| 1.030 | .9141 | .9563 | .9633 |
| .9950 | 1.005 | .9948 | .9971 |

*For each model the first element in the column corresponds to the element in A that multiplies the intercept term, the next three numbers correspond to elements in A that multiply the seasonal dummies, the next number multiplies hog price, the next number multiplies feed price and the last number multiplies the lagged endogenous variable.

Table 3. Root Mean Square Errors Computed from Dynamic Simulations for the Four Model Specifications*,**

| | U.S. | Western C. | Quebec | Ontario |
|---|---|---|---|---|
| OLS 52 obs | 167.5 (6.62) | 14.80 (13.9) | 5.878 (8.94) | 5.411 (5.95) |
| PSV 52 obs | 143.7 (5.66) | 8.696 ( 8.19) | 2.448 (3.27) | 4.417 (4.80) |
| RSV1 52 obs | 104.3 (4.13) | 2.334 (2.15) | .9798 (1.43) | 1.237 (1.38) |
| RSV2 52 obs | 58.28 (2.33) | .3863 (.359) | .2446 (.370) | .2927 (.328) |
| OLS 57 obs | 167.8 (6.62) | 14.89 (14.13) | 5.963 (8.96) | 5.425 (6.00) |
| PSV 57 obs | 142.2 (5.55) | 10.85 (10.1) | 3.570 (4.97) | 4.229 (4.62) |
| RSV1 57 obs | 103.9 (4.09) | 2.470 (2.27) | 1.150 (1.56) | 1.294 (1.46) |

*In each row the number is the root mean square error defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (P_t - A_t)^2}$$

where $P_t$ is the simulated or predicted value of an endogenous variable and $A_t$ is the actual value of the endogenous variable. In parentheses below each RMSE is the percentage RMSE defines as:

$$PRMSE = 100\sqrt{\frac{1}{n} \sum_{t=1}^{n} [(P_t - A_t)/A_t]^2}$$

**For the U.S. equation pork supply is in hundreds of millions of pounds and for the Canadian equations the units of measurement are millions of pounds.

Table 4. Root Mean Squares of Forecasts*

Root Mean Squares of Ten Period Forecasts

| | U.S. | Western C. | Quebec | Ontario |
|---|---|---|---|---|
| OLS 52 obs | 332.6 (11.73) | 21.12 (22.45) | 17.66 (14.23) | 12.44 (11.95) |
| PSV 52 obs | 338.1 (12.0) | 63.58 (65.1) | 19.96 (15.8) | 13.94 (12.71) |
| RSV1 52 obs | 228.2 (8.11) | 64.37 (65.2) | 15.83 (13.0) | 12.91 (11.8) |
| RSV2 52 obs | 177.0 (6.33) | 69.04 ( 7.03) | 13.74 (11.8) | 13.06 (11.6) |

Root Mean Squares of Five Period Ex Post Forecasts
Using 52 Observations for Sample

| | U.S. | Western C. | Quebec | Ontario |
|---|---|---|---|---|
| OLS 52 obs | 217.8 (8.14) | 19.11 (21.8) | 10.14 (10.8) | 9.836 (11.7) |
| PSV 52 obs | 236.4 (8.94) | 42.18 (46.8) | 6.834 (7.42) | 7.570 (9.17) |
| RSV1 52 obs | 174.0 (6.54) | 41.60 (45.5) | 7.544 (8.41) | 7.206 (8.70) |
| RSV2 52 obs | 224.2 (8.11) | 45.08 (49.2) | 9.42 (10.4) | 5.75 (6.98) |

Root Mean Squares on Five Period Ex Post Forecasts
Using 57 Observations for Sample

| | U.S. | Western C. | Quebec | Ontario |
|---|---|---|---|---|
| OLS 57 obs | 239.2 (8.32) | 4.716 (4.50) | 28.32 (21.3) | 19.08 (16.0) |
| PSV 57 obs | 180.0 (6.29) | 18.40 (18.1) | 28.45 (21.8) | 21.29 (18.0) |
| RSV1 57 obs | 162.7 (5.69) | 2.946 (2.84) | 20.71 (15.8) | 17.90 (15.1) |

*The definition of root mean squares is given at the bottom of Table 3.

# REFERENCES

Cooley, T.F. and K.D. Wall. "A Note on Optimal Smoothing for Time Varying Coefficient Problems." NBER Working Paper 128; 1976.

Hildreth, C. and J.P. Houck. "Some Estimators for a Linear Model with Random Coefficients." JASA 63 (1968): 584-95.

Kirsch, J. and K.D. Wall. "Estimation of Time-Varying Parameters." NBER TROLL Experimental Programs, D0070, Cambridge, MA, May 1976.

Martin, L.J. and A.C. Zwart. "A Spatial and Temporal Model of the North American Park Sector for the Evaluation of Policy Alternatives." American Journal of Agr. Economics. 57(1975): 55-66.

Rosenberg, B. "A Survey of Stochastic Parameter Regression." Annals of Econ. and Soc. Meas. 4(1973): 381-398.

Sarris, A.H. "A Bayesian Approach to Estimation of Time Varying Regression Coefficients." Annals of Econ. and Soc. Meas. 4(1973): 501-24.

Sarris, A.H. "A General Algorithm for Simultaneous Estimation of Constant and Randomly Varying Parameters in Linear Relations." NBER Working Paper 38; 1974.

Tomek, W.G. and K.L. Robinson. "Agricultural Price Analysis and Outlook." A Survey of Agricultural Economics Literature, L.R. Martin, ed., University of Minn. Press, Minneapolis, 1977.