#829.0

A New Estimation Approach Using Alternately Actual and Synthesized Observations: Application to a Quarterly Cow Inventory Model

1990

By

Bruce L. Dixon, John W. Goodwin and Danny R. Pippin*

Cattle - statistics

*Bruce L. Dixon and John W. Goodwin are Professor and University Professor of Agricultural Economics at the University of Arkansas, Fayetteville; Danny R. Pippin is a farm management specialist with the Tennessee Cooperative Extension Service. The expert programming assistance of Ms. Siew Goh and Ms. Diana Danforth is gratefully acknowledged.

AAEA 1990

\#8290

A New Estimation Approach Using Alternately Actual
and Synthesized Observations: Application to a Quarterly Cow Inventory Model

An efficient method for estimating regression parameters when observations on the dependent variable are alternately actual observations and synthesized observations is developed. A quarterly cow inventory model is estimated using this technique. Increases in real, short term interest rates are shown to have had a major impact on cow inventories.

Key words: synthesized observations, efficient estimation, cow inventories.

A New Estimation Approach Using Alternately Actual and Synthesized Observations: Application to a Quarterly Cow Inventory Model

The ability to estimate with a combination of observed and synthetic data and to be able to obtain complete statistical efficiency is particularly important for agriculture with its large variety of livestock populations. However, such a method is also important for estimating models explaining the behavior of any stock variable that is measured and reported on a periodic basis where the frequency of measurement is not fully appropriate for estimation or decision making. This assumes that reasonable methods can be used to synthesize observations for unreported periods.

Observed and synthesized observations could be regressed on appropriate independent variables treating both types of dependent variables as having been generated in the same way. From an econometric point of view such an approach is unsatisfactory in principle because it may lead to inefficient estimates due to heteroscedasticity. The dependent variables for time series data could easily be characterized by two different and independent patterns of autocorrelation. The objective of this study is to develop an efficient estimator for models with both synthesized and actual observations.

Cattle inventory data are not available at sufficiently short intervals to allow accurate modelling of inventories and their response to price and weather variations in the short term. Cattle inventory models by Bobst and Davis and Rucker et al. use annual data. Semi-annual data have been available only since 1973.

Cattle inventory estimates are reported every January 1 and July 1. Pippin and Goodwin detail a methodology for synthesizing the inventories for unreported quarters (April 1 and October 1). Their basic approach is to take the inventory levels from a semiannually reported date and then adjust those numbers for the next quarter on the basis of imports, exports, slaughter, death loss, calf crop and natural biological progression from one class to another. The estimation technique developed here is applied to estimating a model to explain quarterly cow inventory variations.

Statistical Model

Suppose the model of interest is given as:

$$y_t = X_t\beta + \epsilon_t \qquad \epsilon_t = \rho\epsilon_{t-1} + u_t \qquad u_t \sim (0,\sigma_u) \tag{1}$$

but only alternate values of y_t are observed. However, an observation for the unobserved periods can be synthesized as a function of the previously observed period, say y_1 to obtain a value for y_2, which is not observed. To get the synthesized value, assume that y_{t-1}, an actual observation, is adjusted to give $\hat{y}_t$, an estimate of the unobserved y_t based on some known relationship:

$$\hat{y}_t = y_{t-1} + Z_t\gamma \tag{2}$$

where Z_t is a vector of explanatory variables, some of which might also be in X, the nxk matrix of regressors, and γ is a vector, assumed known.

The term $Z_t\gamma$ is called the adjustments process because it adjusts y_{t-1} to current levels. In estimating the β in (1) $\hat{y}_t$ is regressed on

$X_t\beta$ for y_t that are not observable. To obtain an efficient estimator it is necessary to derive the distributions of the regression error terms. To develop the error distribution note that $\hat{y}_t$ is being regressed on $X_t\beta$ so,

$$\hat{y}_t - X_t\beta = \omega_t$$

where ω_t is used in lieu of ϵ_t since ϵ_t is not a part of the process generating the value of $\hat{y}_t$. To write ω_t in terms of stochastic components, substitute for $\hat{y}_t$ in terms of (2) to get

$$y_{t-1} + Z_t\gamma - X_t\beta = \omega_t \tag{3}$$

and then substitute again for y_{t-1} to get

$$X_{t-1}\beta + \epsilon_{t-1} + Z_t\gamma - X_t\beta = \omega_t. \tag{4}$$

In designing a synthesizing process for the unobserved variables a reasonable objective would be to require that on the average the $\hat{y}_t$ equaled the expected value of y_t, which is $X_t\beta$. That is, the objective of the synthesis is to compute $\hat{y}_t$ such that:

$$E(\hat{y}_t) = X_t\beta. \tag{5}$$

If (5) is a characteristic of the synthesizing process then it is straightforward to obtain the distribution of ω_t. From (2)

$$\begin{aligned} E(\hat{y}_t) &= E(y_{t-1}) + Z_t\gamma \\ &= E(X_{t-1}\beta + \epsilon_{t-1}) + Z_t\gamma \\ &= X_{t-1}\beta + Z_t\gamma \end{aligned} \tag{6}$$

Thus, equating (5) and (6), implies

$$X_t\beta = X_{t-1}\beta + Z_t\gamma. \tag{7}$$

It would seem a highly unlikely situation for the synthesis to be perfect in every period since some of the components of Z_t could be random in repeated samples as they are in our cow inventory application. Thus it is more reasonable to assume that the randomness in the adjustment ($Z_t\gamma$) should be represented by an additive error term, θ_t, so that

$$\hat{y}_t = y_{t-1} + Z_t\gamma + \theta_t. \qquad (8)$$

Now replacing (2) with (8) and assuming (7) holds, ω_t is written as:

$$\omega_t = \epsilon_{t-1} + \theta_t$$

where it is assumed that θ_t has mean zero. For the same general reasons that there could be autocorrelation in the ϵ_t there could also be autocorrelation in the θ_t. For our application it is assumed

$$\theta_t = \phi\theta_{t-1} + v_t \qquad v_t \sim (0,\sigma_v).$$

Further it is assumed that all v_t are distributed independently of all u_t which implies θ_t and ϵ_t are independent of each other. Then the variance of ω_t can be written as:

$$\frac{\sigma_u}{(1-\rho^2)} + \frac{\sigma_v}{(1-\phi^2)} \qquad (9)$$

or $\sigma_\epsilon + \sigma_\theta$ where $\sigma_\epsilon = \dfrac{\sigma_u}{(1-\rho^2)}$ and $\sigma_\theta = \dfrac{\sigma_v}{(1-\phi^2)}$.

Efficient estimation requires developing the structure of the covariance matrix of the error terms. Essentially there are two different error terms, ϵ_t and ω_t, with variances σ_ϵ and $\sigma_\epsilon + \sigma_\theta$. The relevant covariances are, where s is an even integer and w is an odd integer:

$$E(\epsilon_t\epsilon_{t-s}) = \rho^s\sigma$$

$$E(\omega_t\ \omega_{t-s}) = E[(\epsilon_{t-1} + \theta_t)\ (\epsilon_{t-s-1} + \theta_{t-s})],$$

$$= \rho^s\sigma_\epsilon + \phi^s\sigma_\theta$$

$$E(\omega_t \epsilon_{t-w}) = E[(\epsilon_{t-1} + \theta_t)\ (\epsilon_{t-w})]$$
$$= \rho^{w-1}\sigma_\epsilon .$$

Finally, if there is an odd number of observations it is necessary to compute

$$E(\omega_t \epsilon_{t+1}) = E[(\epsilon_{t-1} + \theta_t)(\epsilon_{t+1})]$$
$$= \rho^2\sigma .$$

Empirical Estimation Procedure

Clearly, maximum likelihood (ML) estimation could be applied if the distributions of u_t and v_t were known. In this application an estimated generalized least squares (EGLS) estimator is used because it is computationally simpler than ML and does not require assuming a specific distribution. If consistent estimates of ϕ^2, ρ^2, σ_u and σ_v can be obtained, then β can be estimated consistently using EGLS by substituting the consistent estimate of the error covariance matrix into the generalized least squares formula (Goldberger).

Transformation to White Noise

As a means of estimating ϕ^2 and ρ^2, first consider transforming the model to obtain non-autocorrelated errors. Assume a sample of n observations where odd numbered observations are actual observations and even numbered observations denote synthesized values. Consider the subsample of the n/2 (or (n+1)/2 if n is odd) observations with actual observations on y_t. Let two such observations be:

$$y_t = X_t\beta + \epsilon_t \tag{10}$$

$$y_{t+2} = X_{t+2}\beta + \epsilon_{t+2}. \tag{11}$$

Multiply (10) by ρ^2 and subtract from (11) to get

$$y_{t+2} - \rho^2 y_t = X_{t+2}\beta - \rho^2 X_t\beta + \epsilon_{t+2} - \rho^2\epsilon_t. \tag{12}$$

By the assumed pattern of autocorrelation,

$$\epsilon_{t+2} - \rho^2\epsilon_t = \rho u_{t+1} + u_{t+2}. \tag{13}$$

The two terms on the right of (13) are, by assumption, a linear combination of independent random variables so that their sum is independent. Thus ordinary least squares (OLS) on the transformed data in (12) would give consistent estimates of β. However this would ignore the observations from the synthesizations.

To utilize the synthetic observations consider the difference of $\hat{y}_t$ and y_{t-1}, i.e.

$$\begin{aligned} \hat{y}_t - y_{t-1} &= X_t\beta + \omega_t - X_{t-1}\beta - \epsilon_{t-1} \\ &= (X_t - X_{t-1})\beta + \theta_t. \end{aligned} \tag{14}$$

It is apparent from (14) that the intercept (if there is one) is subtracted out of the model and is therefore unestimable from this subset of the data. However the intercept can be estimated from the data subset with the observed dependent variable. Since θ_t is autocorrelated (14) can be differenced in the same way as (10) and (11), i.e.

$$\begin{aligned} \hat{y}_t - y_{t-1} - \phi^2(\hat{y}_{t-2} - y_{t-3}) = {} & ((X_t - X_{t-1}) - \phi^2(X_{t-2} - X_{t-3}))\beta \\ & + v_t + \phi^2 v_{t-2}. \end{aligned} \tag{15}$$

Estimating ρ^2 and ϕ^2

The parameters ρ^2 and ϕ^2 are seldom known in empirical work so they must be estimated. They can be obtained in a straightforward manner. For example ρ^2 can be obtained by regressing the OLS residuals of (10) on their lagged values. Similarly the OLS residuals of (14) are regressed on their lagged values to get an estimate of ϕ^2.

Correcting for Heteroscedasticity

Given an estimate of ρ^2, $\sigma_u(1 + \rho^2)$ can be estimated consistently from OLS applied to (12) with $\hat{\rho}^2$ replacing $\hat{\rho}^2$. Similarly OLS applied to the observations defined in (15) with $\hat{\phi}^2$ replacing $\hat{\phi}^2$ will generate a consistent estimate of $\sigma_v(1 + \phi^2)$. Given these two estimates, both (12) with $\hat{\rho}^2$ replacing ρ^2 and (15) ϕ^2 replacing ϕ^2 can be divided through by the roots of $\hat{\sigma}_u(1 + \rho^2)$ and $\hat{\sigma}_v (1 + \phi^2)$, respectively, to correct for implied heteroscedasticity between the two different sets of data. Combining the two sets of transformed data into one sample and then applying OLS will yield consistent estimates of β and its standard errors.

Specification and Estimation of the Cow Inventory Model

Cattle prices, milk prices, input prices, and pasture conditions are important factors for explaining cattle inventories. Because of the quarterly nature of the data, a search procedure was undertaken by Goodwin and Pippin to identify the appropriate lags for the explanatory variables. Futures prices were not considered since futures contracts are not available for weaned calves. The search procedure specified the cow inventory (thousands) as:

$CWINV=f(FST45_{12-14}, HAYPR_{12-14}, MILKPR_{4-6}, RLINT_{2-4}, PST_{4-7}, Q1, Q2, Q3)$

(16)

where

CWINV = Quarterly cow inventory, thousand head.

FST45 = Deflated feeder steer price, $/cwt 400-500 pound Choice Steers at Kansas City.

HAYPR = Deflated hay price, average received by U.S. farmers, $/ton.

MLKPR = Deflated milk price, received by U.S. farmers, all milk sold to plants, $/cwt.

RLINT = Real interest rate: Prime 4-6 months less annualized rate of inflation.

PST = Pasture and Range Conditions Index.

Q1 = Binary variable for January 1 observation.

Q2 = Binary variable for April 1 observation.

Q3 = Binary variable for July 1 observation.

The subscripts indicate the length of lag of the particular independent variables in quarters. For example, 12-14 would indicate the average of the variable over the period 12-14 quarters previously. The sample data range from the first quarter of 1973 to the first quarter of 1987 for 57 observations.

The major assumption is that cow inventories are related to the price of Kansas City choice steers between 400 and 500 pounds. The value of weaned calves is the primary price series to which cow-calf producers respond in their decisions to expand or reduce brood cow inventories. Data sources for all the variables are described in detail in Goodwin and Pippin but essentially all of the data come from various USDA sources. All prices in the model are deflated by the CPI and adjusted to first quarter 1987 dollars, the last quarter of the sample.

The cow inventory equation was estimated as described earlier in the paper. In addition, the results below are from a seemingly unrelated system that also had calf, steer and heifer inventory equations. Thus the estimates below are also influenced by the errors of the other inventory equations. The results are:

$$\begin{aligned} CWINV = \underset{(3729)}{30588} &+ \underset{(7.917)}{26.39}FST45_{12\text{-}14} - \underset{(15.91)}{81.81}HAYPR_{12\text{-}14} - \underset{(82.67)}{276.2}RLINT_{2\text{-}4} \\ &+ \underset{(148.3)}{894.2}MLKPR_{4\text{-}6} + \underset{(37.07)}{114.8}PST_{4\text{-}7} + \underset{(308.4)}{125.9}Q1 + \underset{(330.5)}{1222}Q2 + \underset{(142.6)}{1427}Q3 \end{aligned}$$

All of the coefficients have the expected sign and are strongly statistically significant. The coefficient of determination is .87. As a further test of the model's validity eight observations from April 1 1987 through January 1, 1989 were predicted. The percent root mean square error was 5.8 percent.

Feed prices are conspicuously absent. Corn prices were used in preliminary estimation and found to be statistically insignificant at all lag lengths and therefore were not included in the model. In the beef cow equation by Bobst and Davis feed price was also found to be insignificant at the .05 level for a two sided test. Likewise corn price does not appear in the model by Rucker et al. Feed prices are likely to be more important in a beef supply model than an inventory model.

<u>Impact of Government Policies</u>

The estimated regression model can be used to explain the apparent change in direction of the trend in cow numbers and the apparent change in the phasing of the cycle in those numbers during the late 1970's and early 1980's. The two most volatile explanatory variables in this model are real interest rates and milk prices -- both of which are largely determined by government policy. One of the most striking economic policy changes of the very late 1970's and the early 1980's was in the philosophy that drove monetary policy. By 1978, inflation had become of such overriding concern that the general objective of monetary policy became limiting the money supply in order to reduce inflation.

When real interest rates are fixed at their pre-1978 averages for the sample and projected into the period from 1978 on, the model calls the up-turn in cow numbers in the same quarter and at the same level as was the case with post-1978 real interest rates at their historical level as shown in Figure 1. This is to be expected in view of the average lag of three quarters on real interest rates. But an additional five quarters of <u>increasing</u> numbers is observed under this scenario. That is, the 1978 change in monetary policy was a <u>major</u> factor in the apparent restructuring of the cycle in cattle numbers. Also, the <u>level</u> of the inventory is such that the 1986 simulated cow inventory is still above the level of the previous trough in numbers, whereas the actual 1986 inventory had declined below the levels of the <u>two</u> previous troughs.

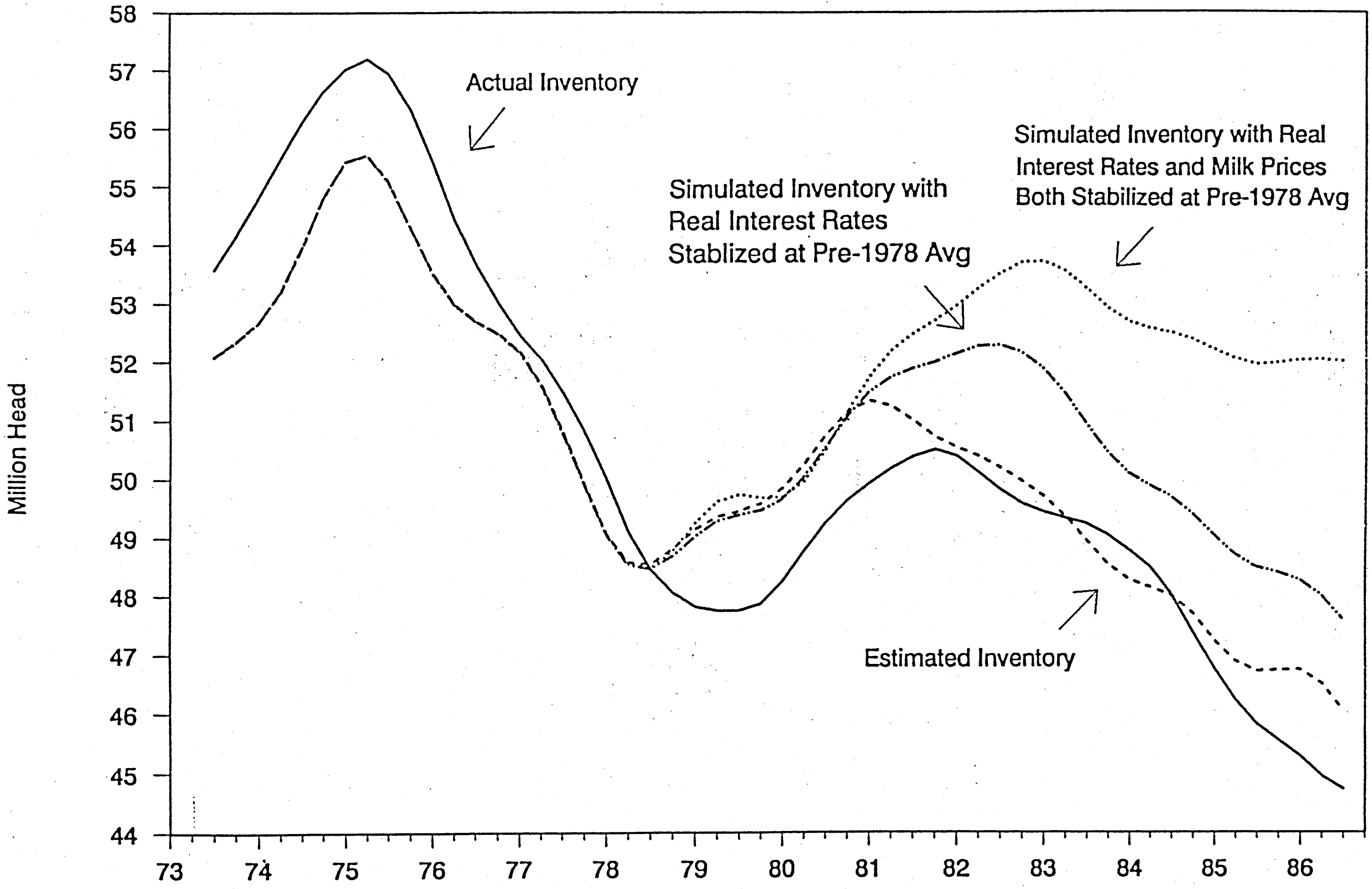
When both real interest rates and milk prices are stabilized and projected to the post-1978 period, the phasing of the cycle in cow numbers is the same as was observed when real interest rates alone were stabilized. But the level of the inventory is held to substantially higher levels. Thus, at least some of the apparent reversal in the direction of the overall trend in cow inventories may be due to reduced milk price supports.

Conclusions

So long as acceptable assumptions can be made about the dynamics of a stock under observation and sufficient data are available to allow computation of reasonable estimates of the unobserved variable, a regression model can be constructed that is a correct specification of the underlying process. In the model specified here the error in the synthesization of the unobserved values is directly incorporated into the model. This leads to heteroscedasticity and autocorrelation which can be accommodated with a GLS estimator.

While there is obviously a strong biological component to the cattle cycle, it is clear that government policy has had a substantial impact on inventories. The increase in real interest rates explains an important portion of the change in the cattle cycle from its pattern over the last eight decades. Past studies have ignored this cost component. The large variations in the real interest rates that occurred in the late seventies and early eighties were likely necessary to make it possible to detect this impact.

Figure 1: Seasonally Adjusted Quarterly Inventory of Cows, U.S., 1973-1986, Actual and Estimated Inventories, and Simulated Inventories with Stablilized Real Interest Rates and Milk Prices



REFERENCES

Bobst, Berry W. and Joe T. Davis, "Beef Cattle Numbers, Crop Acreage, and Crop Policy," American Journal of Agricultural Economics, 69(Nov 1987),771-776.

Goldberger, A. S., Econometric Theory (New York: Wiley, 1964).

Goodwin, John W. and Danny R. Pippin, The Cycle in U.S. Cattle Numbers Related to Selected Supply Variables, Arkansas Agricultural Experiment Station, Report Series No. 314., February 1990.

Pippin, Danny R. and John W. Goodwin, "Synthesizing Quarterly Estimates of the U.S. Cattle Inventory," Arkansas Agricultural Experiment Station Special Report No. 141, March 1990.

Rucker, R. R., O. R. Burt and J. T. LaFrance, "An Econometric Model of Cattle Inventories," American Journal of Agricultural Economics, 66(Feb. 1984),131-144.