Staff Papers Series

Model Selection with Temporal and Spatial Aggregation:

Alternative Marketing Margin Models

by

Charles C. Lyon

and

Gary D. Thompson

Lᴍ

Department of Agricultural and Applied Economics

University of Minnesota
Institute of Agriculture, Forestry and Home Economics
St. Paul, Minnesota 55108

Model Selection with Temporal and Spatial Aggregation:

Alternative Marketing Margin Models

by

Charles C. Lyon

and

Gary D. Thompson

The authors are a Graduate Research Assistant, Department of Agricultural and Applied Economics, University of Minnesota and an Assistant Professor of Agricultural Economics, The University of Arizona, respectively. The authors would like to acknowledge the helpful comments of Steven C. Blank.

Model Selection with Temporal and Spatial Aggregation: Alternative Marketing Margin Models

Introduction

The effects of data aggregation on the adequacy and performance of a given econometric model have been extensively examined (Theil; Ironmonger; Nerlove; Grunfeld and Griliches; Bryan; Zellner and Montmarquette). The general approach taken in these studies involves assessing the potential information loss or bias resulting from aggregation by repeatedly estimating the same econometric specification under different aggregation schemes (spatial or temporal).[1] Broadly speaking, the studies demonstrate that the statistical performance (and inferences based thereupon) of a particular specification is sensitive to the chosen unit of observation. This result suggests that for a given econometric model, there may exist an optimal level of aggregation for which the model is "correctly" specified, with model misspecification occurring at other aggregation levels. This possibility indicates a dimension of research which is seldom, if ever, addressed in the aggregation literature; that is, how the level of aggregation may influence the choice of model specification.[2]

This paper investigates this issue by specifying several competing econometric models (each of which may plausibly contain some elements of the "truth" embodied in the data-generating process) and determining how spatial and temporal aggregation affects specification selection. Classical non-nested hypothesis tests (Pesaran; Pesaran and Deaton; Davidson and MacKinnon; Bernanke, et. al.) are performed to determine model preference. Because nearly all previous empirical studies employing non-nested tests consider only one level of aggregation, this study extends earlier research by exploring the relationship between "truth" (i.e., model performance) and aggregation. Thus two hitherto distinct branches of literature are conflated and research questions addressed in each branch lead to a new research question. Simply, the research question posed in the aggregation literature has been: Maintaining a single econometric specification, what is the "correct" unit of observation? while the question posed in the non-nested literature is: Given a unit of observation, what is the "correct"

model? This paper demonstrates that the choice of specification and dataset are closely linked and that a relevant question for research thus becomes: As aggregation levels change, what is the change in the appropriate model?

Recent contributions to the empirical research on agricultural marketing margins provide an opportune vehicle for the analysis. Several alternative models of marketing margins have been specified and, in some cases, compared using non-nested hypothesis tests (Wohlgenant and Mullen (WM); Thompson and Lyon (TL); Faminow and Laubscher; Powers). The margin models have been employed as single equation components in structural models or estimated as "reduced form" equations. In structural models, a simple markup model has usually been assumed to reflect accurately the relationships between farm-gate and retail prices (Freebairn and Rausser; Arzac and Wilkenson) or has been justified theoretically but not statistically (Kinnucan and Forker). In circumstances where adequate data for a structural specification were not available, margin models have been estimated as single equation "reduced form" models (TL; Faminow and Laubscher). In these latter studies, non-nested tests were used to select the margin specification most consistent with the data employed. Although the studies estimated similar models, the non-nested tests did not indicate preference for the same model in both papers. Whether this outcome reflects unique market structure or product characteristics for different commodities or results from the use of different types of data (i.e., weekly versus monthly and city versus national) remains unclear.

In the present study, fluid whole milk was chosen as the commodity for analysis because it is a homogeneous product across spatially distinct markets. Hence, any detected differences in model choice at various levels of data aggregation should not be attributable to product heterogeneity but rather to marketing agent behavior and the nature of the marketing technology involved. In choosing fluid whole milk, two desirable product characteristics across markets were considered: (1), minimal variation in inherent product quality; and (2), minimal regulatory or policy differences. Few, if any, agricultural products produced and consumed on a national basis embody both these characteristics. But while some dairy policy differences exist across markets, federal milk quality standards ensure that milk is of a strictly homogeneous quality across markets. Thus three alternative models of marketing margins for

2

whole milk are specified using data for the markets of Minneapolis, MN, Kansas City, MO and Philadel-phia, PA. Monthly data for these cities form the "base" dataset and are aggregated over time to "simulate" quarterly and semi-annual datasets and over space to obtain a "regional" dataset. Use of the same base dataset to obtain the aggregate datasets ensures that empirical results will be strictly comparable. Two non-nested hypothesis test statistics are derived and calculated to determine whether model preference is related to temporal or spatial aggregation. Findings indicate that model performance and preference are sensitive to temporal and spatial aggregation. Thus in applied research, the choice of margin specification appears to be inextricably linked to the choice of the economic unit of observation.

The remainder of this paper is organized as follows. The next section reviews the theory underlying the alternative models. The third section describes the data and the model specifications. The fourth section presents the econometric results and is followed by a description of the non-nested testing procedures and results. The final section summarizes the study.

Theoretical Considerations

Marketing margins may be defined as either: (1), the difference between the price paid by consumers and that obtained by producers; or (2), the price of a collection of marketing services which is an outcome of the supply and demand for those services (Tomek and Robinson). Given the first definition, the margin is simply the difference between two market-clearing prices corresponding to the intersection of the primary demand and derived supply curves at the retail level and the intersection of the derived demand and primary supply curves at the farmgate level. The difference between these two prices represents the cost of marketing the product. Thus a food item at the retail level consists of two components: the farm product and the marketing services.

Changes in the margin for a retail product depend on the relationship between the quantity supplied of the product at the farm level and the supply of marketing services. If the marketing services supply function has a positive slope, an increase in farm quantities marketed will increase the margin. In some cases, the supply function may be perfectly elastic or negatively sloped. In the latter case, economies of scale may exist in the production of marketing services so that increased quantities of farm products lead to decreased margins.

3

The second definition states that margins are determined by the prices of the collection of marketing services necessary to convey the product from farm gate to retail market. These services may include packing, processing, transportation, and retailing. The prices of the individual services are functions of the supply and demand conditions in their respective markets. Changes in the prices of these inputs produce a change in the margin for the food product.

Given these two definitions, price spreads can be related to retail prices, farm prices, and the cost of marketing inputs. One model specification with a long pedigree in empirical work but with little formal theoretical justification is the markup model (Waugh; George and King). Waugh argued that consumer demand (primary demand) is the determining factor in the relationship between farm and retail prices; thus food prices are determined at the primary level and farm prices are simply the retail prices minus the costs of marketing inputs (Waugh, p. 20). An empirical specification consistent with this notion is

(1) $M = f(PR, Z)$ $M = PR - PF$

where M is the farm to retail margin, PR is the retail price, PF is the farm price and Z represents other variables such as marketing input costs, time trends or binary variables. This formulation permits spreads to consist of either absolute or percentage markups or a combination thereof.

The markup model also provides the basis for Heien's dynamic approach to margin modelling, whereby "store managers apply a markup over costs...in order to arrive at price" (Heien, p. 11). In the short-run, the plausible assumption of a fixed-proportion (Leontief) food retailing production function with farm and marketing services inputs leads to a markup model. In the longer run, when substitution between farm and marketing services inputs likely occurs and a Leontief production function is no longer tenable, markup pricing still prevails in Heien's dynamic model so long as the given technology displays constant returns to scale (CRTS).

In addition to the markup model, WM and TL specified and estimated the "Relative" (RL) model which is consistent with the structural analysis presented by Gardner. The model is derived from the inverse derived demand curve for farm output facing the agricultural processor. This specification

4

defines margin as a function of retail price, quantity, and input costs. It is expressed as

$$(2) \qquad M = f(PR, PR^*Q, IC)$$

where PR and M are defined as before, PR^*Q is an interaction term between retail price and the total quantity marketed and IC are marketing costs. The model is linearly homogeneous in input and output prices (WM).

Two important assumptions underlie the RL model and Gardner's analysis. Although Gardner does not explicitly treat short-run versus long-run adjustments in his static equilibrium model, sufficiently long-run changes must presumably occur so that retail demand and farm supply equations can shift. The marketing technology embodied in the model is assumed to display CRTS while the elasticity of substitution between farm and marketing inputs is not constrained to be zero.

The last specification considered, the "marketing cost" (MC) model, is derived from the second definition of a margin, that is, as being composed of a "bundle of marketing services." In this case, marketing service firms are hypothesized to provide services until the marginal cost of the service is equal to marginal revenue. Marketing service costs are determined solely by the quantity of farm output and the firm's cost function. Therefore, the specification includes only a quantity and an input cost variable and is expressed as

$$(3) \qquad M = f(Q, IC)$$

where all variables are defined as in (1) and (2).

These three margin models are quite parsimonious and therefore somewhat limited in their ambitiousness. They seek to describe the relationship existing between price levels rather than to provide a complete analysis of marketing margin behavior, for which a full structural model would probably be required. Nevertheless, they possess considerable explanatory power, achieved with minimal modelling and data efforts. These attributes are likely responsible for their historically extensive use and the continued interest in these models. Relationships between model preference and level of aggregation are suggested when one considers the relative merits and underlying assumptions of the competing margin models. For example, Gardner (p. 406) criticized the MU model by demonstrating that no simple markup pricing rule can accurately capture the relationship between farm and retail prices when these prices are

determined by simultaneous farm supply and retail demand shifts. The markup model, however, is able to capture changes in the margin which occur as a result of adjustments in either supply or demand schedules, but not both. The RL model allows simultaneous changes in demand and supply conditions through the inclusion of the interaction term, PR*Q (WM). One implication of these arguments is that with short-run agricultural data (e.g., weekly or monthly observations), the MU specification may adequately capture margin behavior since it is highly likely that retail demand schedules will not shift substantially. Alternatively, temporally aggregated data (e.g., quarterly or annual observations) may embody shifts in both demand and supply schedules, thereby favoring the RL specification and possibly invalidating the MU model. Finally, the MC model will be valid if marketing firms adhere to a marginal cost pricing rule for marketing services. While this appears to be a relatively innocuous assumption at any level of aggregation, firms may deviate from this rule in the short-run if, for example, adjustment costs associated with expanding or contracting output are prohibitive. In the presence of such adjustment costs, however, one would expect that in the long-run, firms will be able to fully adjust output in order to equate marginal revenue and marginal cost, thereby suggesting that the MC specification will gain plausibility as the time interval between observations lengthens.

The nature of the marketing technology assumed in each of the three models also suggests that model performance may vary depending upon the level of data aggregation. Two aspects of the marketing technology are particularly relevant: returns to scale and the elasticity of substitution. Both Heien and Gardner assumed CRTS so that the MU and RL models may be appropriate when CRTS are present. In the MC model, no assumptions regarding the nature of long-run average costs are made; the MC model could therefore be consistent with any type of returns to scale. Diewert rigorously demonstrated that the asymptotic technology of an industry will be approximately consistent with CRTS even when the technologies of individual firms display non-constant returns to scale.[3] Thus, spatially aggregated data might tend to embody CRTS as the returns to scale characteristics of various firms are implicitly aggregated across markets. The elasticity of substitution between farm and marketing service inputs (σ) is assumed to be zero in the short-run markup model but could be non-zero in the long-run MU version as well as in the MC or RL models. Diewert also demonstrated that the asymptotic technology of the industry displays at least as much input substitutability as the technology of the firm and, in the limiting

6

case, that the industry technology possesses input substitutability even when all firms have distinct non-homothetic, Leontief production functions. Hence, the more spatially aggregated the data, the more untenable becomes the assumption of zero input substitutability. Consequently, with spatially and temporally aggregated data, the short-run MU model based on no substitutability between farm inputs and marketing services inputs would likely be rejected.

Although the foregoing discussion does not allow unambiguous conclusions regarding the relationship between choice of margin models and temporal and spatial aggregation, a few tendencies emerge. First, the MU model may be incompatible with temporally and spatially aggregated data if simultaneous supply and demand shifts or input substitutability are important. Second, the MC model could be more consistent with temporally and spatially disaggregate data if non-CRTS technology plays an important role. Finally, the RL model would likely perform well for more aggregate data in which non-zero elasticity of substitution, CRTS, and simultaneous shifts in retail demand and farm supply are important features.[4] The validity of these theoretical tendencies linking model performance to data aggregation level can be explored by statistically testing the alternative models against one another at different levels of aggregation.

None of the marketing margin models in equations (1) - (3) is necessarily a nested case of any of the others so that nested tests of linear restrictions are ruled out. Instead, classical non-nested hypothesis techniques must be employed. Non-nested testing is motivated by the notion that competing models may each have some probability of being the "true" model. Thus competing specifications are evaluated on the basis of how well they explain misspecifications in alternative models (Pesaran and Deaton; Bernanke, et. al.). The models are compared in a pair-wise manner against one another to determine if any single model emerges as preferred. Conditional upon the power and size of the test statistics, non-nested tests permit the data to resolve the issue of which model is most appropriate for a given dataset. Of course, non-nested tests may indicate acceptance or rejection of all candidate models because null and alternative models are juxtaposed in the pair-wise comparisons.

A variety of test statistics for non-nested testing of single equation and multivariate analysis have been proposed. For brevity, a closer discussion of the tests utilized in this paper is postponed until the data and alternative model specifications are presented.

Data and Model Specifications

The data employed in this study consist of monthly observations on retail fluid whole milk prices, processor level prices for milk, total quantities of whole milk sold and dairy processor wages for the 10 years 1979 to 1988. The retail and processor price data are for the markets of Kansas City, Minneapolis, and Philadelphia. The wage data is for the states of Missouri, Minnesota and Pennsylvania. The quantity data are for the relevant federal milk marketing order within which each city resides.[5] Dairy products other than fluid whole milk are not considered due to a lack of monthly market-level retail prices. The three markets were selected based upon availability of a continuous monthly time series for all variables and to capture any elements of margin behavior related to spatial factors. Dairy processing wages were chosen to proxy marketing input costs because they represent the largest single component of operating costs for milk processors (Jones) and were available on a monthly market-by-market basis.[6] The wage data displayed a large degree of variability so, following Heien (p. 15), wages were converted to unit labor costs. Other important marketing cost variables, such as transport costs and packaging costs, could not be included due to the lack of appropriate data.

Processor prices for milk are reported in terms of dollars per hundred weight (cwt.). To determine the margin on a retail unit basis (i.e., gallons), the farm level price per retail unit was calculated using a formula which accounts for both the value of non-fat solids and the value of fat solids (see Table 1). The non-fat solids value was calculated using the Announced Class I Cooperative Price. This price is the amount paid by milk processors (handlers) to milk producers for fluid grade milk and in certain cases may reflect charges to processors for services performed by suppliers. In general, however, the Announced Class I Price closely approximates the cost of fluid grade raw milk delivered to the processor.

To test the relationship between model specification and aggregation level, each of the three models in equations (1) - (3) was specified empirically as linear seemingly unrelated regressions (SUR):

$$(4) \qquad M_{it} = a_1 + a_2 PR_{it} + a_3 W_{it} + a_4 D + e_{1it}$$

$$(5) \qquad M_{it} = b_1 PR_{it} + b_2 (PR^*Q)_{it} + b_3 W_{it} + b_4 D + e_{2it}$$

$$(6) \qquad M_{it} = c_1 + c_2 Q_{it} + c_3 W_{it} + c_4 D + e_{3it}$$

8

where i = Kansas City, Minneapolis, and Philadelphia; t is the time index, t = 1,...,T; and equations (4), (5) and (6) are the MU, RL, and MC models, respectively. All variables are defined in Table 1. The SUR estimation technique was used because contemporaneous correlation between error terms for each market was hypothesized. Thus the vector of error terms for each equation is assumed to be distributed $N(0, \Sigma \otimes I_T)$, where Σ is the (m x m) covariance matrix, and m denotes the number of equations in each system (m = 3, in this case).

The relationship between length of observation and model preference was explored by estimating each of these models in turn using monthly, quarterly, and semi-annual data. The quarterly (semi-annual) data were obtained from the monthly data by calculating the mean of the prices and wages over the previous 3 (6) months and by summing the quantities for the relevant period. This method of "simulating" quarterly and semi-annual observations from the monthly data has two justifications. First, it eliminates any potential biases or information losses which could arise from choosing a single monthly observation upon which to base the quarterly or semi-annual observation. Second, it resembles the manner in which datasets available to the applied researcher are usually constructed. Government agencies such as the Bureau of Labor Statistics often calculate average retail food prices by computing a weighted average of samples conducted over time and space (Rabil). Thus the simulated datasets employed here mimic the quality of price data often employed by applied economists. The sum rather than the average of the quantity data was used in order to reflect aggregate supply conditions over the relevant period.

To test the sensitivity of model preference to spatial aggregation, the market-level data were aggregated into a single weighted average series for the monthly, quarterly, and semi-annual data. Because the Middle Atlantic marketing order far surpasses the Greater Kansas City and Upper Midwest marketing orders in quantities sold, aggregate prices and wages were calculated by weighting market prices and wages by per capita quantities sold. The specifications for these models are identical to those in equations (4) through (6), except that the i subscript is deleted. These spatially-aggregated models will be referred to as equations (7) - (9). The "experimental design" of the non-nested tests at different levels of aggregation is depicted by the six cells in Figure 1.[7]

9

Estimation Results

The models in equations (4) through (6) were estimated using the iterative Zellner efficient technique, which is computationally equivalent to maximum likelihood estimation (Kmenta and Gilbert). The single equation versions in equations (7) to (9) were estimated by OLS. In total, eighteen regressions were performed—three SUR models and three single equation models were estimated for each of the three data categories. The SUR results are displayed in Table 2. Also presented are the Wald statistic values testing for contemporaneous correlation among equations. In all cases except one, the null hypothesis of no contemporaneous correlation was rejected. Thus for these cases, the SUR estimation procedure is consistent with the data. In the case where the null hypothesis could not be rejected (the MC model with semi-annual data), the SUR technique was used to maintain uniformity in the estimation procedure; therefore standard errors for the coefficients in this model may be overstated relative to those of OLS estimates (Kmenta and Gilbert). For reasons to be stated below, however, the SUR estimator is still desirable so as to account for the existence of a cross-equation error structure.

In all cases, the SUR models were tested for the existence of a within- and across-equation first-order autoregressive structure (Guilkey). Given m equations, the error structure is

$$(10) \qquad e_t = Re_{t-1} + v_t$$

where e_t is an (m x 1) vector containing the tth error terms from all equations; R is an (m x m) matrix whose elements, ρ_{ij}, are the autocorrelation coefficients between disturbances in equation i and those of equation j with i,j = 1, 2, 3; and v_t is an (m x 1) vector of white-noise residuals. If all $\rho_{ij} = 0$, there exists contemporaneous correlation, but no autocorrelation; if $\rho_{ij} = 0$ only for $i \neq j$, there exists an AR(1) process within each equation but not across equations; and if $\rho_{ij} \neq 0$ for $i \neq j$, there exists a first order vector autoregressive scheme both within and across equations. Wald test values for the two separate null hypotheses of $R = 0$ (no autocorrelation) and R diagonal [AR(1) process only] provide indirect evidence of the nature of the error structure and are reported in Table 2. The Wald test values indicate the existence of a first order vector autoregressive [VAR(1)] scheme in all but two of the models. Thus for the VAR(1) cases, the procedure outlined by Guilkey and Schmidt to obtain efficient estimates was employed. In the two cases where diagonality of R was not rejected (the MU model for quarterly and semi-

10

annual data), indicating an AR(1) process but not a VAR(1) process, the AR(1) estimation procedure outlined in Judge, et al. was used.

The primary focus of this paper is on the non-nested hypothesis test results and attendant implications; thus the estimation results are only briefly reviewed. First, in the SUR models, the retail price coefficients in each regression have the expected positive sign and are very significant. The magnitudes of the retail price coefficients in the MU models indicate that margins are more sensitive to retail price changes in Minneapolis than in the other two markets. In addition, a spatial pattern is suggested since the magnitude of the retail price coefficients varies inversely with the distance of the markets from the Eau Claire WI milkshed, the center of the federal milk marketing order pricing system and a region of surplus milk production.[8] Because larger retail price coefficients imply less sensitivity of processor prices to retail price changes, this pattern may also reflect the increased responsiveness of processor prices to retail price changes as the Class I differential increases and excess supplies diminish. Indeed, in markets awash in excess supplies (such as Minneapolis), processor prices may be expected to be less responsive to changes in the determinants of primary demand. Alternatively, the pattern may simply be a manifestation of margin rigidity in the Philadelphia market owing to the minimum price schema imposed on milk marketers by Pennsylvania milk marketing regulations. These minimum prices (or "resale price controls") are essentially designed to fix marketing margins (Manchester). The real significance of the pattern may also be debatable since it is not consistently demonstrated in the RL model results.

The estimated coefficients for the Wage variable are positive and significant in most cases. In six cases, the estimated coefficient is negative, but significantly so in only three cases. An explanation for this unexpected result is multicollinearity between the wage and quantity variables present in the same specifications. This possibility is suggested by the nearly uniformly positive and significant coefficients for the Wage variable in the MU specifications. Regarding quantities, the MC estimation results exhibit a mixed effect with quantities affecting margins positively in Philadelphia and negatively (when significant) in Minneapolis and Kansas City. This is suggestive of an upward-sloping marginal cost curve for processors in the Philadelphia market and economies of scale existing in the processing industry in the other two markets.

11

The behavior of margins during 1988, the first year of an across-the-board 50 cents per cwt. decrease in the federal support level for milk effective Jan. 1, 1988, is captured by the DLOW dummy variable. Coefficients on the DLOW variable in nearly all models indicate that during 1988, margins increased in Kansas City and Philadelphia and decreased in Minneapolis. The increased level of significance of these coefficients at the quarterly and semi-annual levels and their insignificance at the monthly level (except in the MU model) demonstrates the remedial influence which data aggregation apparently has on the regression results.

The single equation estimation results largely echo those of the SUR models (Table 3). Of particular interest is the pattern in the retail price coefficients indicating the increasing responsiveness of margins to retail price changes as observation period lengthens. Relatively smaller retail price coefficients with the monthly data indicate greater margin rigidity and suggest the existence of some type of fixed markup behavior in the short-run. As observation period lengthens, retailers probably adjust prices with less reliance on short-run aberrations in processor prices and with more reliance on expected long-run price trends, resulting in less margin rigidity and relatively larger retail price coefficients. Also notable is the relatively poor performance of the Wage variable, which enters negative (but insignificant) in three cases and is significant in only the MU model. Finally, in contrast to the SUR results, the binary variable, DLOW, is insignificant in all but two cases.

Hypothesis Test Results

Originating with Theil's (1954) work on linear aggregation, the effects of temporal and spatial aggregation have usually been analyzed by estimating a single specification across different levels of aggregation. Griliches observed that different "true" models appear plausible for different levels of aggregation, suggesting that alternative specifications ought to be tested against one another at each level of aggregation to determine acceptable specifications for each level. In this vein, two related non-nested test statistics are presented below.

The non-nested tests employed in this analysis are the "P2" and "P3" tests derived in Bernanke, Bohn and Reiss (BBR, p. 302) for the spatially aggregated, single equation models (7) through (9) and a

12

multivariate analog to the P2 and P3 tests for the spatially disaggregated, seemingly unrelated regression (4) through (6). Although BBR present six separate tests, the P2 and P3 tests are the most computationally straightforward, with the P3 test having size and power properties comparing favorably to the more complicated tests. The six BBR tests differ from conventional non-nested tests in that they are designed for models exhibiting serially correlated error structures. Prior to development of the BBR tests, researchers confronted with the task of testing model specification with non-nested hypothesis tests sometimes used generalized least-squares (GLS) techniques to obtain efficient estimators and then applied conventional non-nested tests conditional upon the GLS estimates. BBR note that this procedure is inherently inconsistent because autoregressive errors and model misspecification may be closely associated.

Autoregressive error structures may result from either a true serial correlation process or from model misspecification. If either of these conditions are present but not acknowledged, OLS estimates will be biased and inconsistent (in the latter case) or inefficient (in the former case). Misspecification implies that estimated error terms may contain relevant omitted variables. Transformation of data to account for the AR(1) or the VAR(1) process in the case of SUR models, process prior to instituting the non-nested tests amounts to an a priori presumption, however, that the model exhibits a natural serial correlation process and that error terms do not contain relevant omitted variables. Conventional non-nested tests applied to transformed data could therefore fail to recognize this potential source of specification error.[9]

The BBR tests (termed "GLS" tests by BBR) provide a framework within which estimator inefficiency and inconsistency resulting from both misspecification and serial correlation can be tested simultaneously. No a priori assumptions regarding the source of the serial correlation process need be made; therefore the BBR tests appear to be more generally applicable to models which contain AR(1) or VAR(1) processes than conventional tests (termed "OLS" tests by BBR). In Monte Carlo experiments with the OLS and GLS tests when significant autocorrelation was present, BBR found the GLS tests to have generally smaller size bias than the OLS tests and the GLS statistics were more closely distributed to the asymptotic normal distribution. The GLS versions of the P2 and P3 tests were found to be more powerful than their OLS counterparts. Whether these Monte Carlo results carry over to multivariate models

13

remains uncertain because VAR(1) processes have not been analyzed previously in the context of multi-variate P tests (Chalfant and Finkelshtain).

We turn now to the non-nested hypothesis testing of the alternative margin models for each of the SUR data categories—monthly, quarterly and semi-annual—and for the single equation models. The SUR P2 and P3 test results and the single equation hypothesis test results are reported in Table 4. For the SUR models, the test statistics were calculated by extending the procedure for single equation models with an AR(1) structure outlined in Bernanke, et al., to the multivariate models considered here. The extension involves replacing the scalar estimate of ρ in the BBR equations with its multivariate analog, denoted R. Given two competing models, define the null and alternative hypotheses

$$(11) \quad \begin{aligned} H_0: \ & y = X\beta + u_0 & u_{0,t} = R_\bullet \, u_{0,t-1} + v_{0,t} & \quad v_0 \sim N(0, \Sigma_0) \\ H_1: \ & y = Z\gamma + u_1 & u_{1,t} = R_1 \, u_{1,t-1} + v_{1,t} & \quad v_1 \sim N(0, \Sigma_1) \end{aligned}$$

where the regressors, X and Z, are (mT x K) matrices of non-stochastic variables with at least one column of X not contained in Z; y is a (mT x 1) vector of dependent variables; $R_\bullet$ and R_1 are the (m x m) matrices of autocorrelation coefficients defined above in (10); β and γ are conformable parameter vectors; and u_0, u_1, v_0, and v_1 are conformable vectors of disturbances, all normally distributed. The P2 test value is defined as the t-statistic for λ obtained in the following multivariate regression:

$$(12) \quad \hat{u}_{0,t} = \hat{u}_{0,t-1}\pi_1 + \left[X_t - \left(I \otimes \hat{R}_0 \right) X_{t-1} \right]\pi_2$$
$$+ \left(I \otimes \Sigma_0 \Sigma_1^{-1} \right) * \left(\hat{v}_{0,t} - \hat{v}_{1,t} \right)\lambda + \eta_t$$

where π_1, π_2 and λ are parameters to be estimated (see Appendix for a derivation). The P3 test value is obtained from the t-statistic for λ after substituting $\hat{v}_{10t}$ for $\hat{v}_{1t}$ in eq. (12), where

$$(13) \quad \hat{v}_{10,t} = \left[y_t - \left(I \otimes \hat{R}_0 \right) y_{t-1} \right] - \left[Z_t - \left(I \otimes \hat{R}_1 \right) Z_{t-1} \right]\hat{\gamma}_{10}$$

and $\hat{\gamma}_{10}$ is the maximum likelihood parameter vector obtained from the auxiliary regression of $\left[X_t - \left(I \otimes \hat{R}_0 \right) X_{t-1} \right]\hat{\beta}$ on $\left[Z_t - \left(I \otimes \hat{R}_1 \right) Z_{t-1} \right]$. The observed P2 and P3 test values are compared to the tabled t-statistics for the appropriate degrees of freedom. A test statistic value greater than the tabled critical t-value indicates that the model represented as the null hypothesis (the "maintained" model) is rejected by the model represented by the alternative hypothesis (the "alternative" model) given the data;

14

that is, the alternative model has sufficient explanatory power to reject the maintained model. Pair-wise comparisons of the two models are performed by interchanging H_0 and H_1 in eq. (11) and recomputing the regression in eq. (12). Based on Monte Carlo evaluations, BBR found that the P3 statistic may fail to reject a false null model more frequently than the P2 statistic. They caution, however, that in finite samples, the P3 statistics may have an advantage over the P2 statistics, as they "explicitly recognize the truth of the null" (p. 302).

The preceding tests are "pair-wise" in the sense that each model is tested against a single alternative. It is also possible to test the maintained model against more than one alternative by calculating a regression based on eq. (12) with as many regressors like $(I \otimes \hat{\Sigma}_0 \hat{\Sigma}_i^{-1}) * (\hat{v}_{0,t} - \hat{v}_{i,t}) \lambda_i$ as there are alternative models. The statistic based on this test has an asymptotic F distribution with the null hypothesis that all the λ_i's are equal to zero (Davidson and MacKinnon, 1983). Despite their intuitive appeal, the joint tests have apparently not been used much, if at all, in the non-nested testing literature.

We focus now on the non-nested tests, the results of which are displayed in Tables 4 and 5. Table 4 reports the t-statistics for the pair-wise hypothesis tests while Table 5 reports the results of the joint hypothesis tests, whereby each model is tested against both alternatives simultaneously. For clarity of presentation, the pair-wise tests in Table 4 will be examined first; discussion of the joint tests will follow. Considering first the t-statistics for the monthly SUR specifications, the pair-wise P2 test indicates rejection of all maintained models by all alternatives. Thus a model preference under the P2 statistic at the monthly level does not emerge. Under the P3 test, however, the MU and MC models cannot be rejected by one another while the RL model is rejected by both alternatives. Thus the choice at the monthly level reduces to a judgement between the MC model and the MU model. Comparing the relative magnitudes of the t-statistics, the MU model emerges as the marginally preferred SUR specification at the monthly level.

For the quarterly SUR specifications, the MU and MC models are rejected by both alternatives under the P2 test while the RL model is rejected by only the MC model. Under the P3 test all models are rejected by only one of the alternatives. Considering both the P2 and P3 tests together, the RL model is rejected twice while the MU and MC models are each rejected three times. Thus the RL model displays

15

the strongest performance at this level of aggregation. At the semi-annual level, the RL model clearly emerges as the preferred model as both the MU and MC models are rejected by each alternative under both the P2 and P3 statistics, while the RL model is rejected by only one alternative under the P3 statistics.

For the single equation models, the pair-wise P2 and P3 tests favor the RL model at the quarterly and semi-annual levels, while at the monthly level neither the RL and MU models can be rejected by any alternative under either test. The rejection of the MU models at the quarterly and semi-annual levels, however, suggests that temporal aggregation has an important effect on the adequacy of the MU model, similar to that observed with the SUR models. While the RL model emerges at this level as the unambiguously preferred model, the evolution of the MC model into a more tenable specification at this level (relative to its performance at the monthly level) is noteworthy. The improved performance of the MC model is consistent with the hypothesis that margins based on the marginal cost pricing rule becomes a more viable proposition as observation length increases.

The joint hypothesis tests are reported in Table 5. For computational ease, the SUR models were jointly tested by calculating likelihood ratio statistics based on unrestricted regressions of the form given by an augmented version of eq. (12) and regressions performed after imposing the restriction that all $\lambda_i = 0$. The single equation models were tested by calculating the appropriate F-statistic. Asymptotically, both statistics have the χ^2 distribution. For the SUR models, all monthly models are once again rejected by their two alternatives, although the relatively low value of the likelihood ratio statistic for MU as the maintained model tends to corroborate the marginal preference for this model indicated by the pair-wise tests. At the quarterly and semi-annual levels, the MU model is overwhelmingly rejected, while the RL model clearly emerges as the uniquely preferred SUR specification at both levels.

For the joint hypothesis single equation tests, the results largely echo those of the pair-wise t-tests with the RL model emerging as the preferred specification at the quarterly and semi-annual levels and with both the MU and RL models acceptable at the monthly level. Once again the MC model grows progressively more tenable as the length of observation grows; however, while the model fails to be rejected at the 5% significance level, it is rejected at the 10% level with the semi-annual data.

16

To summarize these results, the effect of spatial aggregation at the monthly level has been to shift from tentative acceptance of the MU model in the SUR framework to acceptance of the MU or RL models in the single equation setup. An intuitive explanation for the acceptance of the MU model is that aggregation across markets results in "average" margin behavior which is identical to margin behavior in the individual markets. An indirect method for testing this phenomenon is to assess the equality of coefficients across markets in the SUR models; this test is equivalent to Zellner's test for aggregation bias with micro-data.[10] If all coefficients are statistically equivalent, then average margin behavior would be expected to approximate margin behavior in each market. Equality of coefficients across equations was tested by imposing linear equality restrictions on the coefficients in each model and then calculating the χ^2 statistic. For all restrictions, the null hypothesis of equality was rejected at the 5% level. Thus, the monthly MU model does not pass one test for spatial aggregation even though the MU model cannot be completely rejected by the alternative models for regional or spatially aggregated monthly data.

Generally then, the results of the P2 and P3 tests provide mutually supportive evidence in favor of the hypothesized relationship between model preference and aggregation level. Although non-nested tests performed on the monthly SUR data do not provide as unambiguous a preference for any single model as do tests at the quarterly or semi-annual levels, one clear implication emerges: temporal aggregation results in acceptance of the RL specification as the preferred margin model. This is consistent with the hypothesis that for longer periods of observation, changes in both demand and supply schedules become more likely, thus favoring the RL model. Failure of the tests to indicate a stronger preference at the monthly SUR level may simply reflect observed tendencies of the statistics to overly reject (accept) a true (false) null model with larger sample sizes. The findings, however, are consistent with the notion that the RL model is preferred when margin behavior is affected by changes in both sides of the market, whereas the MU model is a valid specification when adjustments occur on only one side of the market.

For the spatially aggregated data, the RL model rejects the MU and MC models at the quarterly and semi-annual levels; at the monthly level, the MU and RL models are acceptable. In an indirect manner, this suggests that the asymptotic (i.e., regional) marketing technology in the fluid whole milk industry is consistent with both CRTS and a positive elasticity of substitution between farm inputs and

17

marketing inputs. These results help explain why TL chose the MU model when employing weekly data for three U.S. cities while with monthly, national data (Faminow and Laubscher; Powers) and annual, national data (WM) the RL model has at least been cautiously accepted over the other two specifications.

Finally, the marginal preference for the MU model over the RL model at the monthly level is noteworthy in light of the WM results. WM cite simultaneous beef industry trends of declining real beef prices and decreased production as evidence that both supply and demand changes are important in explaining price movements. For this reason, the RL model was postulated and emerged as the preferred specification in their study. The whole milk data employed in this study, however, exhibit the same trends of declining real prices and decreasing production. Thus the failure of the RL model to emerge in this study as the preferred specification at the monthly level lends additional weight to the contention that model performance is data-specific.

Conclusions

This study demonstrates that temporal and spatial aggregation of data affect the choice of specification for marketing margin models. Because econometric specification and the level of data aggregation are both chosen (either explicitly or implicitly) by the researcher, the relationship between choice of specification and aggregation level should not be ignored. The non-nested hypothesis test results for most levels of data aggregation seem to confirm Gardner's criticism of the MU model, but also suggest that for data of a shorter time-span—weekly or monthly observations—and for spatially disaggregated markets (cities or urban areas) the MU model remains a defensible specification. This is consistent with the findings of studies in which alternative specifications of marketing margin models have been compared using non-nested hypothesis tests but with data constructed at different levels of temporal and spatial aggregation. Finally, the marginal preference for the MU model at the monthly level by the P2 statistic tends to support the historical use of this specification (with monthly data) by other applied researchers (e.g., Kinnucan and Forker).

Several caveats now deserve mention. First, stronger results at the most disaggregate level might have been obtained had weekly data been available. Following the trends evident in the findings reported here, it is anticipated that non-nested tests performed on weekly data would show a strong

18

preference for the MU model. Second, the tests were conducted on data for only one commodity; future research should consider the aggregation/specification relationship for other agricultural commodities. Third, the semi-annual regressions were performed with only 20 observations per equation. The paucity of observations at this level may have adversely influenced regression results; a lengthier data series would have been desirable. Finally, Bernanke, et. al. present six potential non-nested statistics for testing models with AR(1) specifications. Only the two statistics most easily implemented were calculated here. Thus future research should consider implementing the remaining test statistics to determine if they corroborate results obtained in this paper.

Notes

[1] One strain of the literature examines how the performance of distributed lag and time series models is affected by temporal aggregation (Amemiya and Wu; Blank and Schmiesing; Brewer; Geweke; Mundlak; Tiao and Wei; and Wei). The effects of spatial aggregation are seldom addressed in these studies.

[2] In a penetrating discussion of misspecification and aggregation, Griliches (p. 737) remarked that there likely exist different "truths" at different levels of aggregation, an idea apparently originating with Houtthaker (1955). Accordingly, he suggested abandoning the approach, due to Theil (1954), of comparing a single specification across different levels of aggregation.

[3] In all of the alternative models considered here, competitive factor markets are assumed to prevail. Diewert's results also require that firms minimize costs.

[4] In the context of estimating the demand for farm output, Wohlgenant found that the elasticity of substitution for dairy output and marketing inputs was different from zero and that the marketing technology displayed constant returns to scale.

[5] The U.S. Dairy Industry is composed of 41 federal milk marketing orders which regulate the handling and pricing of approximately 70% of the Nation's milk supply. Underlying this federal structure is a system of state regulations governing trade practices, milk promotion and minimum milk prices. For the markets and time period considered here, Missouri had trade practice regulations, Minnesota had milk promotion and trade practice regulations and Pennsylvania had milk promotion regulations and had established minimum prices at the producer, wholesale, and retail levels.

[6] Operating costs do not include the cost of materials for processing and resale, i.e., raw milk and cream.

[7] The power of the test statistics used throughout the following may be affected by the different samples sizes in each cell of the experimental design. None of the results, however, appears to be biased as sample size changes.

[8] The Federal Milk Marketing Order pricing system sets minimum prices for milk contingent upon end-use: fluid or manufactured. The minimum price for fluid use milk is based upon the Minnesota-Wisconsin price (M-W price) for manufactured grade milk and involves adding a price differential to the M-W price. The price differential increases with increased distance from Eau Claire, WI, so that markets farther from Eau Claire will generally have higher Class I prices, subject to local market conditions.

20

[9] Pesaran (1982) suggests that models with autocorrelated errors could first be transformed to obtain non-autocorrelated error terms after which standard multivariate non-nested would then be applied. As Bernanke, et. al., points out, such an approach would ignore potential sources for misspecification. The results from applying Pesaran's approach to the models and datasets considered here, which for brevity are not reported, would have resulted in different model choices at some levels of aggregation. Hence, accounting for autocorrelation structures and misspecification explicitly in the hypothesis tests appears to matter in this empirical application.

[10] Pesaran, et. al., refer to Zellner's test as a test of "micro-homogeneity". Their more general test for perfect aggregation—spatial aggregation in the context of the present study—could not be implemented here because the necessary rank condition for their test statistic was not met (Pesaran, et. al., p.868).

References

Amemiya, Takeshi and Roland Y. Wu. "The Effect of Aggregation on Prediction in the Autoregressive Model." J. Amer. Stat. Assoc. 67(1972):628-32.

Arzac, Enrique R. and Maurice Wilkinson. "A Quarterly Econometric Model of United States Livestock and Feed Grain Markets and Some of Its Policy Implications." Amer. J. Agr. Econ. 61(1979):297-308.

Bernanke, Ben, Henning Bohn and Peter C. Reiss. "Alternative Non-Nested Specification Tests of Time-Series Investment Models." J. Econometrics 37(1988):293-326.

Blank, S.C. and B.H. Schmiesing. "Whimsical Aggregation of Temporal Data, Market Identification, and Fragile Results." Agribusiness 6(1990):33-40.

Brewer, K.R.W. "Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX Models." J. Econometrics 1(1973):133-54.

Bryan, William R. "Bank Adjustments to Monetary Policy: Alternative Estimates of the Lag." American Economic Review 57(1967):855-864

Chalfant, J.A., and I. Finkelshtain. "Small Sample Properties of Multivariate Non-nested Hypothesis tests." In: Proceedings of the Business and Economic Statistics Section of the American Statistical Association, 1987:379-384.

Davidson, R., and J.G. MacKinnon. "Several Tests for Model Specification in the Presence of Alternative Hypotheses." Econometrica 49(1981):781-93.

Davidson, R., and J.G. MacKinnon. "Testing the Specification of Multivariate Models in the Presence of Alternative Hypotheses." J. Econometrics 23(1983):301-313.

Diewert, W. E. "The Comparative Statics of Industry Long-run Equilibrium." Canadian J. Econ. 14(1981): 78-92.

Faminow, M. and J.M. Laubscher. "Empirical Testing of Alternative Price Spread Models in the South African Maize Market." J. Agr. Econ. (forthcoming).

Freebairn, J.W. and Gordon C. Rausser. "Effects of Changes in the Level of U.S. Beef Imports." Amer. J. Agr. Econ. 57(1975):676-88.

Gardner, Bruce L. "The Farm-Retail Price Spread in a Competitive Food Industry." Amer. J. Agr. Econ. 64(1975):399-409.

George, P.S., and G.A. King. Consumer Demand for Food Commodities in the United States with Projections for 1980. Giannini Foundation Monograph No. 26, Univ. of Calif., Berkeley, 1971.

Geweke, J. "Temporal Aggregation in the Multiple Regression Model." Econometrica 46(1978):643-61.

Godfrey, L.G., and M.H. Pesaran. "Tests of Non-Nested Regression Models: Small Sample Adjustments and Monte Carlo Evidence." J. Econometrics 21(1982):133-54.

Griliches, Zvi. "Discussion" of "Effects of Aggregation Over Time on Dynamic Characteristics of an Econometric Model," R.F. Engle and T. Liu in Econometric Models of Cyclical Behavior, Volume II, B.G. Hickman, ed., New York: Columbia Univ. Press, 1972.

Grunfeld, Y. and Z. Griliches. "Is Aggregation Necessarily Bad?" **Rev. Econ. Stat.** 62(1960):1-13.

Guilkey, David K. "Alternative Tests for a First-Order Vector Autoregressive Error Specification." **J. Econometrics** 2(1974):95-104.

Guilkey, David K., and Peter Schmidt. "Estimation of Seemingly Unrelated Regressions with Vector Autoregressive Errors." **J. Amer. Stat. Assoc.** 68(1973):642-47.

Heien, Dale M. "Markup Pricing in a Dynamic Model of the Food Industry." **Amer. J. Agr. Econ.** 62(1980):10-18.

Ironmonger, D.S. "A Note on the Estimation of Long-Run Elasticities." **J. Farm Econ.** 41(1959):626-632.

Houthakker, H.S. "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis." **Rev. Econ. Studies** 23(1955):27-31.

Jones, W. Webster. Milk Processor-Distributors' Sales, Costs, and Margins, 1979. Econ. Res. Serv. ERS-666. Washington, D.C.: October 1981.

Judge, G.G., W.E Griffiths, R. Carter Hill, H. Lutkepohl, and Tsoung-Chao Lee. **The Theory and Practice of Econometrics.** 2nd. ed. New York: Wiley, 1985.

Kinnucan, Henry W., and Olan D. Forker. "Asymmetry in Farm-Retail Price Transmission for Major Dairy Products." **Amer. J. Agr. Econ.** 69(1987):285-92.

Kmenta, J. **Elements of Econometrics.** Second Edition. MacMillan Publishing Company. New York, 1986.

Kmenta, J., and Roy F. Gilbert. "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions." **J. Amer. Stat. Assoc.** 63(1968):1180-1200.

Manchester, Alden C. **The Public Role in the Dairy Economy: Why and How Governments Intervene in the Milk Business.** Westview Press, 1983.

Mundlak, Yair. "Aggregation Over Time in Distributed Lag Models." **Int. Econ. Rev.** 2(1961):154-63.

Nerlove, Marc. "On the Estimation of Long-Run Elasticities: A Reply." **J. Farm Econ.** 41(1959):632-640.

Pesaran, M.H. "On the General Problem of Model Selection." **Rev. Econ. Studies** 41(1974):153-171.

Pesaran, M.H. "On the Comprehensive Method of Testing Non-Nested Regression Models." **J. Econometrics** 18(1982):263-274.

Pesaran, M.H., and A.S. Deaton. "Testing Non-Nested Nonlinear Regression Models." **Econometrica** 46(1978):677-94.

Pesaran, M.H., R.G. Pierse, and M.S. Kumar. "Econometric Analysis of Aggregation in the Context of Linear Prediction Models." **Econometrica** 57(1989):861-88.

Powers, Nicholas J. "Marketing Order Impacts on Farm-Retail Price Spreads: Comment." **Amer. J. Agr. Econ.** forthcoming, May 1991.

Rabil, Floyd A. "Average Retail Food Prices: A Brief History of Methods." Technical Note. **Bureau of Labor Statistics.**

Theil, H. **Linear Aggregation of Economic Relations,** Amsterdam: North-Holland, 1954.

Thompson, Gary D., and Charles C. Lyon. "Marketing Order Impacts on Farm-Retail Price Spreads: The Suspension of Prorates on California-Arizona Navel Oranges." **Amer. J. Agr. Econ.** 71(1989):647-660.

Tiao, G.C. and W.S. Wei. "Effect of Temporal Aggregation on the Dynamic Relationship of Two Time Series Variables." **Biometrika** 63(1976):513-23.

Theil, H. **Principles of Econometrics,** New York: John Wiley and Sons, 1971.

Tomek, William G., and Kenneth L. Robinson. **Agricultural Product Prices.** Ithaca, NY: Cornell University Press, 1981.

United States Department of Agriculture, Agricultural Marketing Service. Federal Milk Order Market Statistics, Annual Summary, various issues.

United States Department of Agriculture, Agricultural Marketing Service, Dairy Division. Unpublished Retail Price data, various years.

United States Department of Labor, Bureau of Labor Statistics. Unpublished Wage data, various years.

United States Department of Labor, Bureau of Labor Statistics. BLS Handbook of Methods, Volume II: The Consumer Price Index. Bulletin 2134-2. Washington, D.C., April 1984.

Waugh, F.V. Demand and Price Analysis: Some Examples from Agriculture. Washington, D.C.: U.S.D.A. Tech. Bull. No. 1316, 1964.

Wohlgenant, Michael K. "Demand for Farm Output in a Complete System of Demand Functions." **Amer. J. Agr. Econ.** 71(1989):241-52.

Wohlgenant, Michael K., and John D. Mullen. "Modeling the Farm-Retail Price Spread for Beef." **W. J. Agr. Econ.** 12(1897):119-25.

Wei, W.W.S. "The Effect of Temporal Aggregation on Parameter Estimation in Distributed Lag Model." J. **Econometrics** 8(1978):237-46.

Zellner, A. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." **J. Amer. Assoc. Stat.** 57(1962):348-68.

Zellner, A. and C. Montmarquette. "A Study of Some Aspects of Temporal Aggregation Problems in Econometric Analysis." **Rev. Econ. Stat.** 53(1971):335-42.

Appendix: Multivariate "P2" Test

The P2 tests developed by Bernanke et al., is given by

$$(A.1) \quad \hat{u}_{0t} = \hat{u}_{0t,t-1}\pi + \left(X_t - \hat{\rho}_0 X_{t-1}\right)\pi_2 + \lambda\left[\hat{v}_{0t} - \hat{v}_{1t}\right] + m$$

where $\hat{u}_{0t}$ are correlated residuals from the null single equation model, $\hat{v}_{0t}$ and $\hat{v}_{1t}$ are uncorrelated residuals from the null and alternative models, $\hat{\rho}_0$ is the estimated autocorrelation coefficient for the null model, and X_t are explanatory variables from the null models. Heuristically, if the difference between the predicted dependent variables under the null and alternative hypotheses is significant, then a large t-statistic value associated with $\hat{\lambda}$ indicates that the alternative model possesses enough explanatory power to reject the null model.

The derivation of a multivariate analog to the P2 test of Bernanke et al., begins by differencing the null and alternative seemingly unrelated equation models.

$$(A.2) \quad H_0^*: \quad y_t = \left(I \otimes R_0\right)y_{t-1} + \left[X_t - \left(I \otimes R_0\right)X_{t-1}\right]\beta + v_{0t}$$
$$= y_{0,t-1}^* + X_t^*\beta + v_{0t}$$

$$H_1^*: \quad y_t = \left(I \otimes R_1\right)y_{t-1} + \left[Z_t - \left(I \otimes R_1\right)Z_{t-1}\right]\gamma + v_{1t}$$
$$= y_{1,t-1}^* + X_t^*\gamma + v_{1t}$$

where all observations in each of the variables have been reordered such that the first observation from each equation represent the first m observations, the second observation from each equation the second m observations, and so on. After differencing the models, the weighted exponential likelihood approach used by Pesaran (1982) may be employed to generate a composite model. In particular, the exponentiated likelihood functions of H_0^* and H_1^* are mixed using the weights $\left(1-\lambda\right)$ and λ, respectively. A convex combination of exponential likelihood functions is generated in this manner. After suitable manipulation of the weighted exponential likelihood function, the resulting composite multivariate model is obtained

$$(A.3) \quad H_\lambda: \quad y_t = (1-\lambda)G^{-1}\left(I \otimes \Sigma_0\right)\left(y_{0,t-1}^* + X_t^*\beta\right) + \lambda G^{-1}\left(I \otimes \Sigma_1\right)\left(y_{1,t-1}^* + Z_t^*\gamma\right) + v$$

where $G \equiv (1-\lambda)I \otimes \Sigma_0 + \lambda I \otimes \Sigma_1$. Note that G^{-1} is a nonlinear function of λ because of the inverse operator.

The multivariate P2 test is derived by taking an approximation to (A.3) evaluated at $\lambda = 0$. Differentiating the right-hand side of (A.2) with respect to λ and evaluating at $\lambda = 0$ yields.

(A.4) $\left(I \otimes \Sigma_0 \Sigma_1^{-1}\right)\left[\left(y^*_{1,t-1} + Z^*_t \gamma\right) - \left(y^*_{0,t-1} + X^*_t \beta\right)\right]$

Note that $\partial G^{-1}/\partial\lambda = -G^{-1} \, \partial G/\partial\lambda \, G^{-1} = -G^{-1}\left[\left(I \otimes \Sigma_1^{-1}\right) - \left(I \otimes \Sigma_0^{-1}\right)\right]G^{-1}$ (Theil, 1971, p. 33). Based upon

(A.4), the P2 test approach of MacKinnon and Davidson (1981, 1983) gives the multivariate regression

(A.5) $\hat{u}_{0t} = \hat{u}_{0,t-1}\pi_1 + \hat{X}^*_t \pi_2 + \lambda\left(I \otimes \Sigma_0 \Sigma_1^{-1}\right)\left(\hat{v}_{0t} - \hat{v}_{1t}\right) + \eta$

where the P2 test value is given by the t-statistic on $\hat{\lambda}$. The estimated parameters $\hat{\pi}_1$ are an (m x m) matrix which corresponds to R_0 in (A.2). $\hat{\pi}_2$ is an (K x 1) vector of parameters to be estimated which correspond to β in (A.2).

The multivariate P2 test in (A.5) differs from the univariate P2 test (A.1) in two important ways. First, all "natural" autocorrelation within <u>and</u> across equations must be purged by including $\hat{u}_{0,t-1}$ (an m x m matrix) as an explanatory variable in (A.5). In general, m^2 possible sources of within- and across-equation autocorrelation must be accounted for. In the univariate case, only a single source of autocorrelation must be purged. Secondly, and more importantly, both the deterministic portion—the two models—and the stochastic portion—the particular combination of covariances—must be weighted (Pesaran, 1982). To verify that both deterministic and stochastic portions of the models must be weighted, observe that (A.5) may be rewritten as

(A.6) $\hat{u}_{0t} = \hat{u}_{0,t-1}\pi_1 + \hat{X}^*_t \pi_2 + \lambda\left(I \otimes \Sigma_0 \Sigma_1^{-1}\right)\left[\left(y^*_{1,t-1} + Z^*_t \gamma\right) - \left(y^*_{0,t-1} + X^*_t \beta\right)\right] + \eta$

In the univariate case with scalar variances in each model, weighting the variances results in a test regression equivalent to weighting only the deterministic portion of the models (Pesaran 1982, p. 267).

Significance of λ in the P2 Test

An intuitive rationale for relying on the significance of λ in the P2 test to assess the null specification can be obtained by considering limiting cases. First, assume H_0 to be the "true" model and H_1 "false". Then $y \sim (X\beta, \cdot)$ and of interest is the expected value of λ, $E(\lambda)$, since if $E(\lambda) > 0$, then λ will have a significant t-statistic in the P2 test regression. For notational simplicity consider only the univariate regression given in (A.1); the interpretation extends directly to the multivariate case. Now, subtract $U_{0,t-1}\pi_1$ from the LHS of (A.1). Then it can be shown that

(A.7) $E\left(\hat{\lambda}\right) = \begin{array}{l} a_{21}\left[X^{*'} E(y - X\beta) - X^{*'} E(y-X\beta)_{-1}\pi_1\right] + \\ a_{22}\left[(\hat{v}_A - \hat{v}_0)'\left[E(y-X\beta) - E(y-X\beta)_{-1}\right]\pi_1\right]\end{array}$

26

where $a_{2,1}$ and $a_{2,2}$ are the (2,1) and (2,2) elements of $[X'X]^{-1}$. Since $E\left(y - X\hat{\beta}\right) = E(y) - E\left(X\hat{\beta}\right) = X\beta - XE\left(\hat{\beta}\right)$ and under H_0 true, $E\left(\hat{\beta}\right) = \beta$, then $E\left(y - X\hat{\beta}\right) = 0$ and thus $E\left(\hat{\lambda}\right) = 0$. Analogously, let H_1 be the "true" model and H_0 be "false". Then $y \sim \left(Z\gamma, \cdot\right)$ and still $E\left(\hat{\lambda}\right)$ is given by (A.7) but now $E(y) = Z\gamma$ and clearly $Z\gamma \neq XE\left(\hat{\beta}\right)$ so that $E(\lambda) \neq 0$. These two limiting cases provide heuristic justification for the P2 and P3 tests.

It may also be shown that the value of $E(\lambda)$ when H_0 is maintained but H_1 is "true", depends on the difference $\left(X\beta - Z\gamma\right)$. To see this, consider the quantity $E_{X\beta}[\lambda_{H0}] - E_{X\beta}[\lambda_{H1}]$ where $E_{X\beta}$ is defined as the expectation operator when H_0 is maintained and $\hat{\lambda}_{H0}$ is the value of $\hat{\lambda}$ under H_0 and H_0 is true. Analogously, $\hat{\lambda}_{H1}$ is the value of $\hat{\lambda}$ under H_1 as the true model. Thus $E_{X\beta}\left[\hat{\lambda}_{H0}\right]$ is the expected value of $\hat{\lambda}$ when H_0 is both the maintained model and the true model and $E_{X\beta}\left[\hat{\lambda}_{H1}\right]$ is the expected value of $\hat{\lambda}$ when H_0 is maintained but H_1 is true. It can be shown that the following holds

$$(A.8) \quad E_{X\beta}\left[\hat{\lambda}_{H0}\right] - E_{X\beta}\left[\hat{\lambda}_{H1}\right] \;=\; a_{21}\left[X^{*'}\left(X\beta - Z\gamma\right) - X^{*'}\left(X_{-1}\beta - Z_{-1}\gamma\right)\pi_1\right]$$
$$+ \; a_{22}\left[\hat{v}^1(X\beta - Z\gamma) - \hat{v}^1(X_{-1}\beta - Z_{-1}\gamma)\pi_1\right].$$

Because $E_{X\beta}\left[\hat{\lambda}_{H0}\right]$ is the expectation of $\hat{\lambda}$ when H_0 is both maintained and true, then $E_{X\beta}\left[\hat{\lambda}_{H0}\right] = 0$ by the analysis above. Thus (A.8) demonstrates that the expected value of $\hat{\lambda}$ when H_0 is the maintained model but H_1 is the "true" model is a function of $\left(X\beta - Z\gamma\right)$ or the difference between the "information" contained in the vectors $X\beta$ and $Z\gamma$.

Table 1. Variable Definitions

PR	Retail price per gallon of fluid whole milk; i = Kansas City, MO, Minneapolis, MN, and Philadelphia, PA; in cents (Source: Dairy Division, Agricultural Marketing Service, USDA);
W	Per unit labor costs for dairy processors; with $$WAGE = \frac{DMWAGE}{Q};$$ where DMWAGE is the total monthly wage paid by dairy processors and Q is the total quantity of whole milk sold within the relevant market order; in dollars (Wage Source: Bureau of Labor Statistics; Quantity Source: Federal Milk Marketing Order (FMMO) Statistics, USDA);
PF	Farm value of raw milk, cents per gallon; with $$PF = FVFS + FVNFS$$ where FVFS is the Farm Value of Fat Solids per retail unit; and FVNFS is the Farm Value of Nonfat Solids per retail unit. The two components of PF are calculated as follows: 1. VFS/cwt. = BD x 3.5. 2. VNFS/cwt. = PI - VFS/cwt. 3. VNFS/lb. = VNFS/cwt. : 8.99 4. #FS/gal. = %FS x 8.6 5. FVFS/gal. = #FS/gal. x BD 6. #NFS/gal. = %NFS * 8.6 7. FVNFS/gal. = #NFS/gal. x VNFS/lb. where VFS/cwt. is Value of Fat solids per cwt.; BD is Butterfat Differential (Source: FMMO Statistics); VNFS/cwt. is Value of Nonfat Solids per cwt.; PI is the Announced Class 1 Cooperative Price (Source: FMMO Statistics); 8.99 is the conversion factor Nonfat Solids from cwt. to lb.; #FS/gal. is the lbs. of Fat Solids per gallon of whole milk; %FS is the percentage of Fat Solids in whole milk; 8.6 is the number of lbs. per gallon; FVFS/gal. is the Farm Value of Fat Solids per gallon; #NFS/gal. is the lbs. of Nonfat Solids per gallon; and %NFS is the percentage of Nonfat Solids in whole milk;
Q	Total quantity of fluid whole milk sales within the relevant federal milk marketing order; in millions of pounds (Source: FMMO Statistics); data for Kansas City is from the Greater Kansas City Markng Order (MO), Minneapolis data is from the Upper Midwest MO, and Philadelphia data is from the Mid-Atlantic MO.
DLOW	A (0, 1) binary variable; set equal to 1 for 1988 and set equal to zero, otherwise.

All prices are deflated by the CPI-U.

Table 2. Estimates of Spatially Disaggregate Margin Models with Temporal Aggregation.

Markup Model

Monthly Data

Dep Var[a]	Constant	Price	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	-0.160 (-6.083)	0.534 (17.319)	0.00195 (5.329)	0.0296 (4.540)	85.82	155.74 / 13.79
MMN	-0.264 (-6.294)	0.754 (16.259)	0.00183 (4.828)	-0.00377 (-0.063)		
MPH	0.0203 (1.008)	0.289 (9.411)	0.000455 (0.200)	0.0251 (3.171)		

Quarterly Data

Dep Var	Constant	Price	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	-0.44 (-20.058)	0.990 (39.003)	0.000616 (2.238)	0.107 (7.261)	28.89	40.96 / 3.63
MMN	-0.415 (-20.335)	1.006 (41.865)	0.000474 (1.726)	0.0828 (5.119)		
MPH	-0.306 (-11.135)	0.814 (27.430)	-0.00519 (-1.175)	0.105 (6.696)		

Semi-Annual Data

Dep Var	Constant	Price	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	-0.157 (-1.808)	0.459 (5.447)	0.00334 (2.788)	0.0249 (2.449)	15.12	13.51 / 8.43
MMN	-0.282 (-4.010)	0.641 (8.653)	0.00517 (5.090)	-0.00379 (-0.381)		
MPH	-0.0705 (-1.257)	0.263 (5.652)	0.0323 (2.900)	0.0104 (1.182)		

Relative Model

Monthly Data

Dep Var[a]	Price	P*Q	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	0.383 (17.362)	0.00166 (1.468)	0.000292 (0.632)	0.0107 (1.358)	84.04	323.80 / 25.57
MMN	0.867 (15.165)	-7.2E-05 (-0.519)	-0.00020 (-0.613)	-0.00178 (-0.239)		
MPH	0.345 (9.176)	-0.00041 (-2.377)	-0.00555 (-3.986)	0.00186 (0.179)		

Quarterly Data

Dep Var	Price	P*Q	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	0.790 (9.100)	-0.00477 (-5.554)	-0.00160 (-2.331)	0.0127 (1.534)	31.19	96.09 / 11.37
MMN	0.422 (4.687)	0.000077 (0.356)	0.00152 (2.574)	-0.0319 (-3.405)		
MPH	0.553 (8.050)	-0.00049 (-3.523)	0.000773 (0.220)	0.00677 (0.860)		

Semi-Annual Data

Dep Var	Price	P*Q	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	0.477 (12.289)	0.00127 (4.660)	0.000351 (0.934)	0.0323 (3.907)	10.62	36.02 / 16.08
MMN	0.742 (6.287)	0.00002 (0.792)	0.000255 (0.568)	-0.0163 (-2.154)		
MPH	0.495 (5.486)	-0.00021 (-2.360)	0.00557 (1.468)	0.0118 (1.512)		

Marketing Cost Model

Monthly Data

Dep Var[a]	Constant	Quantity	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	0.196 (5.568)	-0.00149 (-1.622)	-1.4E-05 (-0.022)	0.00858 (0.768)	38.78	261.21 / 11.88
MMN	0.408 (4.910)	0.000291 (0.415)	-0.0029 (-0.391)	-0.00253 (-0.216)		
MPH	0.145 (4.565)	0.000387 (2.209)	0.00233 (0.490)	0.00744 (1.045)		

Quarterly Data

Dep Var	Constant	Quantity	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	0.079 (0.953)	-0.00125 (-2.768)	-0.00032 (0.359)	0.0159 (1.248)	11.28	91.13 / 13.10
MMN	-1.790 (-2.660)	-0.00049 (-3.050)	0.00327 (4.092)	-0.0422 (-4.066)		
MPH	0.129 (1.308)	0.000206 (1.138)	0.00146 (0.265)	0.00833 (0.914)		

Semi-Annual Data

Dep Var	Constant	Quantity	Wage	DLOW	SUR[b]	VAR/AR[c]
MKC	0.335 (6.612)	0.000259 (0.340)	-0.00204 (-2.252)	0.0279 (1.701)	3.44	33.36 / 14.01
MMN	-0.467 (-1.245)	-0.00027 (-1.931)	0.00298 (3.002)	-0.0333 (-2.885)		
MPH	0.0438 (0.406)	0.000188 (1.819)	0.00454 (0.828)	0.0168 (1.489)		

[a] Dependent variables in each set of equations are the marketing margins (PR-PF) in the three urban areas of Kansas City (MKC), Minneapolis (MMN) and Philadelphia (MPH).

[b] Wald statistic values of testing contemporaneous correlation across equations ($\chi^2_{3,\,.10} = 6.25$).

[c] Wald statistic values for testing a VAR (1) structure ($\chi^2_{9,\,.10} = 14.68$) or an AR(1) structure ($\chi^2_{6,\,.10} = 10.64$). The null hypothesis for a VAR (1) structure is that R = 0 while the null hypothesis of AR (1) is R is a diagonal matrix.

[d] T-statistics below coefficient estimates.

Table 3. Estimation Results for Spatially Aggregated, Single Equation Models.

Markup Model

Dep Var	Constant	Price	Wage	DLOW	R_2	RHO	DW[a]	EST.[b]
M (MON)	-0.165 (2.74)c	0.614 (7.97)	-0.00005 (0.08)	0.00303 (0.48)	0.77	0.982 (57.9)	1.98	AR(1)
M (QTR)	-0.152 (2.13)	0.453 (6.36)	0.0057 (2.72)	-0.00049 (0.05)	0.84	0.799 (8.3)	1.63	AR(1)
M (SEMI)	-0.335 (4.69)	0.577 (9.43)	0.012 (5.57)	0.01 (1.25)	0.88	n.a.	1.88	OLS

Relative Model

Dep Var	Price	P*Q	Wage	DLOW	R_2	RHO	DW[a]	EST.[b]
M (MON)	0.61 (6.9)	0.00005 (0.14)	-0.00004 (0.04)	0.0019 (0.32)		0.999 (304)		AR(1)
M (QTR)	0.666 (5.55)	-0.0008 (2.82)	-0.0015 (0.64)	-0.007 (0.82)		0.990 (45.5)		AR(1)
M (SEMI)	0.947 (3.73)	-0.00082 (1.91)	-0.0026 (0.76)	0.0035 (0.35)		0.996 (50.85)		AR(1)

Marketing Cost

Dep Var	Constant	Quantity	Wage	DLOW	R_2	RHO	DW[a]	EST.[b]
M (MON)	0.195 (4.33)	0.00046 (1.61)	0.0014 (1.12)	-0.00014 (0.02)	0.19	0.971 (45)	2.18	AR(1)
M (QTR)	0.334 (2.699)	-0.0003 (0.923)	0.00048 (0.107)	-0.0111 (0.996)	0.28	0.971 (24.9)	1.08	AR(1)
M (SEMI)	-0.536 (1.6)	0.0013 (2.55)	0.0107 (1.6)	0.017 (0.99)	0.62	0.610 (3.36)	2.01	AR(1)

[a] Durbin-Watson statistic value after AR(1) correction has been performed.
[b] Estimation method for each equation.
[c] T-statistics for estimated coefficients.

Table 4. Non-nested Test Results for All Levels of Spatial and Temporal Aggregation.

Regional, Seemingly Unrelated Models

Alternative Model	*Monthly Data* —Maintained Model—			*Quarterly Data* —Maintained Model—			*Semi-Annual Data* —Maintained Model—		
	MU	RL	MC	MU	RL	MC	MU	RL	MC
MU		9.47 / 9.97	26.77 / 1.19a		-0.66 / 3.90	-2.15 / 0.33		-0.30 / -0.62	2.18 / 3.86
RL	11.74 / -4.47		22.02 / 15.04	-6.12 / -1.??		9.95 / 3.13	9.77 / -19.65		5.11 / -2.68
MC	8.58 / -1.23	-3.64 / -4.55		-6.25 / 4.94	2.60 / 0.70		-2.40 / -2.82	0.52 / 2.21	

Spatially Aggregated, Single Equation Models

Alternative Model	*Monthly Data* —Maintained Model—			*Quarterly Data* —Maintained Model—			*Semi-Annual Data* —Maintained Model—		
	MU	RL	MC	MU	RL	MC	MU	RL	MC
MU		-0.42 / -0.26	6.86 / -6.88		-0.35 / -1.79	2.60 / 1.11		-0.16 / 0.03	1.87 / 1.42
RL	1.72 / 1.32		-6.84 / 4.82	3.00 / -1.79		3.19 / 3.75	3.06 / -2.38		2.35 / 0.33
MC	0.50 / -0.43	0.94 / 1.35		0.21 / 2.25	-0.44 / -0.60		0.23 / -2.12	0.41 / -0.72	

a Boldface numbers indicate cases for which the alternative model failed to reject the maintained model. The upper number is the P2 test value and the lower is the P3 test value.

Table 5. Non-nested Test Results for Multiple Alternative Hypotheses.

		-------_Seemingly Unrelated Models, Likelihood Ratio Tests_-------		
		Monthly Data	_Quarterly Data_	_Semi-Annual Data_
	MU^a	107.85[b]	57.84	76.52
		24.71	67.34	102.74
Maintained Model	RL	68.47	**5.07**[c]	**0.53**
		78.35	**2.29**	**3.87**
	MC	207.27	44.87	14.48
		119.12	**9.21**	14.60

		----------------_Single Equation Models, F-tests_----------------		
	MU	4.24[d]	6.98	4.06
		0.62	7.25	3.98
Maintained Model	RL	**0.42**	**0.23**	**1.03**
		0.99	**0.79**	**0.69**
	MC	22.44	5.10	3.68
		23.16	6.47	2.88

[a] The maintained model is tested jointly against the other two models.

[b] The upper value is the P2 test result and the lower is the P3 test value. $\chi^2_{6\,(.05)} = 12.59$

[c] Bold faced values indicate that the maintained model could not be rejected by the alternative models jointly.

[d] F-statistic critical values are $F_{(.05, 2, 116)} \cong 3.07$ (monthly), $F_{(.05, 2, 36)} = 3.26$ (quarterly), and $F_{(.05, 2, 16)} = 3.63$, $F_{(.10, 2, 16)} = 2.67$ (semi-annual).