United States
Department of
Agriculture

Economic
Research
Service

Technical
Bulletin
Number 1932

January 2012

A Quarterly Econometric Model for Short-Term Forecasting of the U.S. Dairy Industry

Roberto Mosheim

Recommended citation format for this publication:

Mosheim, Roberto. *A Quarterly Econometric Model for Short-Term Forecasting of the U.S. Dairy Industry,* TB-1932, U.S. Department of Agriculture, Economic Research Service, January 2012.

Cover photo credit: Shutterstock

USDA

United States Department of Agriculture

Technical Bulletin Number 1932

January 2012

A Report from the Economic Research Service

www.ers.usda.gov

A Quarterly Econometric Model for Short-Term Forecasting of the U.S. Dairy Industry

Roberto Mosheim

Abstract

This research evaluates the econometric approaches employed by USDA's Economic Research Service (ERS) to contribute to the dairy sector forecasts published in the monthly USDA *World Agricultural Supply and Demand Estimates* (WASDE) report. To generate the estimates, a quarterly model for the U.S. dairy industry is specified using data for fourth-quarter 1998 (Q4/1998) to first-quarter 2009 (Q1/2009), and it is estimated and validated employing data for Q2/2009 to Q1/2010. Different forecasts are generated using a variety of single equation and system methods, and then evaluated in terms of forecasting precision or predicting turning points in the data. Different approaches, however, more effectively forecast different variables. Vector autoregression with exogenous variables outperforms structural regression models when forecasting prices, but single and system estimations of structural models are superior to time series models when forecasting some items on farm supply and commodity balance sheets.

Keywords: U.S. dairy industry, forecasts, simultaneous equations, vector autoregression

Acknowledgments

Various versions of this study benefited from the comments from USDA, Economic Research Service economists Don Blayney, Paul Westcott, and Joe Cooper and four anonymous reviewers. Erik Dohlman and Amy Hagerman (both from ERS) provided valuable editorial advice. The author also thanks USDA, Economic Research Service administrators for their feedback and support, and others who helped complete this research. Special thanks also to Angela Anderson for editorial support and Cynthia A. Ray for design assistance.

Contents

Summary iii

Introduction 1

Salient Features of the U.S. Dairy Industry 2

Comparison of a Selection of Relevant U.S. Dairy Sector Forecasting Models 7

Model Specification, Estimation, and Validation 8

- Data Sources and Variable Construction. 8
- Specification 8
- Estimation and Validation. 20

Summary and Conclusion 26

References. 27

Appendix A 29

Appendix B 30

Appendix C. 31

Appendix D 32

Summary

What Is the Issue?

This report documents the ongoing forecasting activities by USDA's Economic Research Service (ERS) that combine judgment-based forecasting with rigorous econometric estimations and data construction that can provide better forecasts than a mostly judgment-based system alone. The formal modeling process gives the forecasting activity at ERS transparency and full documentation on the specification, estimation, and validation procedures employed. Specifically, the econometric models generate the monthly dairy sector forecasts that contribute to the USDA *World Agricultural Supply and Demand Estimates* (WASDE) report. In addition, the model's estimates potentially can be used to examine the structure of the sector and the influence of policy-relevant variables. The merit of various econometric and time series models, however, is more about their ability to forecast effectively than for their potential contribution to policy analysis.

What Did the Study Find?

Various estimation methods successfully forecast different endogenous variables, a situation that might change as the sector evolves or as additional data become available and the applied econometric model improves. The ERS model generated projections that outperformed the consensus forecast by USDA's Interagency Commodity Estimates Committee (ICEC) in roughly half of the instances in terms of accuracy and predicting turning points in the data of interest.

The results demonstrate how different methods are preferable, depending on the variable. To produce forecasts in the dairy sector, we cannot rely on a single estimating method. Moreover, the findings suggest that composite forecasting or forecast blending will play a significant role in this process. Careful econometric specification and data development will ensure a successful transition from the mostly judgment-based forecasting system previously used at ERS.

The ERS forecasting model relies on specific characteristics not seen in previous studies of this kind. Specifically, it:

- Uses a variety of methods to estimate endogenous variables;
- Employs both time series and structural models; and
- Uses quarterly data and ex-post forecasting (seldom seen in this type of research).

The estimations highlight certain characteristics of the U.S. dairy industry:

- Milk production per cow is seasonal and increases over time.
- Herd size movements are cyclical and tied to fluctuations in the all-milk price.
- As the margin (all-milk price minus feed cost) decreases (increases), herd sizes decrease (increase) after a number of periods.

- Price movement in the all-milk price is correlated to the price of cheese and butter more than whey and nonfat dry milk (NDM).
- The dairy sector is highly interlinked (reflected in block recursive structure where variables at one stage serve as determinants for the next).

How Was the Study Conducted?

The model is divided into 4 blocks comprised of 15 behavioral equations and 1 block comprised of 5 identities. Most of the behavioral equations are specified in logarithmic form that permits interpretation of estimated coefficients as elasticities. Each equation within a block forecasts a variable required by ICEC. The blocks are linked in a block-recursive fashion such that the first one generates estimates of variables that are then employed as predetermined variables in the second, third, and fourth. The resulting structure produces consistent forecasts across different sections of the dairy sector.

Economic theory typically defines the structural equations in models, although that practice is not as useful in the case of time series models where all variables within each block influence each other. The ERS model is based on quarterly data, beginning with the first quarter when all necessary variables are available (fourth-quarter 1998 or Q4/1998). Possible limitations, with respect to degrees of freedom for estimation, were the main reasons that the system of equations were divided into blocks and also explains why some modeling choices, such as the econometric specification of dairy product prices by means of inverse (price dependent) product supply equations, were made.

Prior to estimating, the various blocks were identified by the rank condition to ensure that unique values of the structural parameters could be derived from the reduced form of the system. Estimations of the model were conducted by various simultaneous equations and time series methods that generate different values for the endogenous variables. These results were validated by withholding four quarters of known data and estimating the model, generating an ex-post forecast. The ex-post forecasts were compared with the known values of the withheld data to determine how well the models performed based on data available at the end of first-quarter 2009 (Q1/2009). These projections were also compared with those agreed upon by USDA in early April 2009 for the four successive quarters ending in first-quarter 2010 (Q1/2010).

Introduction

The econometric model supports the ongoing forecasting activities performed by USDA's Economic Research Service (ERS) in preparation of the monthly USDA *World Agricultural Supply and Demand Estimates* (WASDE) report. ERS analysts combine judgment-based forecasting with both formal econometric modeling and qualitative elements to produce *quarterly* projections of commodity balance sheets (in this case supply, use, and price variables for the dairy sector). These quarterly estimates are then presented and discussed at the monthly meetings of the Interagency Commodity Estimates Committee (ICEC), the members of which, representing several USDA agencies, vet the quarterly projections. ICEC's consensus forecast is then presented in the monthly WASDE report.

This study documents the specification, estimation, and validation (as well as the data development) of the ERS dairy sector forecasting model. Single- and multiple-equation models were employed to estimate the *entire* system, including ordinary least squares (OLS), two-stage least squares (2SLS), three-stage least squares (3SLS), seemingly unrelated regression (SURE), and two time series models (unrestricted and restricted vector autoregression with exogenous variables or VARX). These methods produced different values for the endogenous variables.

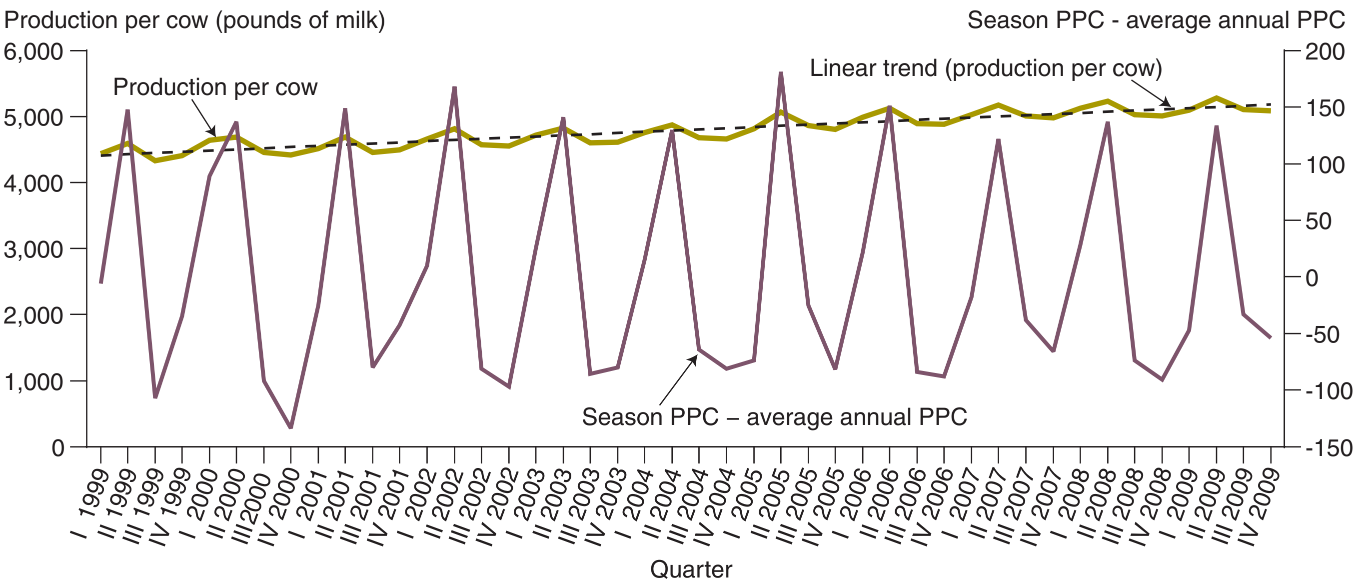
The forecasts produced by the various time series (TS) and structural models (SM) were validated by withholding four quarters of data when specifying and estimating the model. This process generated an ex-post forecast (see Greene, 2008, p. 101) to determine how well the models predicted variables, the values of which were known from information available at the end of first-quarter 2009 (Q1/2009). These projections were also compared with the four quarters from Q2/2009 to Q1/2010, as agreed upon by ICEC members at the beginning of April 2009.

Salient Features of the U.S. Dairy Industry

The relationship between aggregate milk supply and aggregate milk demand (use) drives the dairy industry toward equilibrium based on changes in the all-milk price (AMP). AMP is a weighted average of the price of milk used in fluid milk products and the price of milk used in manufactured dairy products, such as cheese, butter, nonfat dry milk (NDM), and whey. In general, the incentive price to dairy farmers is AMP. Total quantity of farm milk supplied is derived as the product of total number of cows and milk production per cow. Recent trends in cow productivity (milk produced per cow) illustrate the seasonality of production per cow (Q2 is spring) and an increasing trend in production per cow over time (fig. 1). The number of cows, in contrast, has shown a cyclical pattern related to, among other things, movements in AMP and feed costs (FC). A strong inverse correlation (significant at the 1-percent level) can be seen between the number of cows and the real difference (in 2005 dollars) between the AMP and feed costs (calculated from corn, soybean meal, and hay prices) (fig. 2). Changes in this real margin are associated with movements in the number of cows over time. A positive correlation (significant at the 10-percent level) can be seen after five quarters—the most significant positive correlation found in the data.

The all-milk price is linked to manufactured dairy product prices, as noted by Alston et al. (2006), by market clearing conditions. A strong positive association can be seen between the movement of AMP and changes in the prices of cheese and butter (less so for NDM and whey) (fig. 3). These relationships are complemented by the estimates of the calculated pairwise correlations between prices (table 1). All of the correlations between AMP and product prices are positive and significant at the 1-percent level. Cheese has the strongest correlation with the all-milk price, and whey has the weakest. These four

Figure 1
U.S. dairy production per cow and season

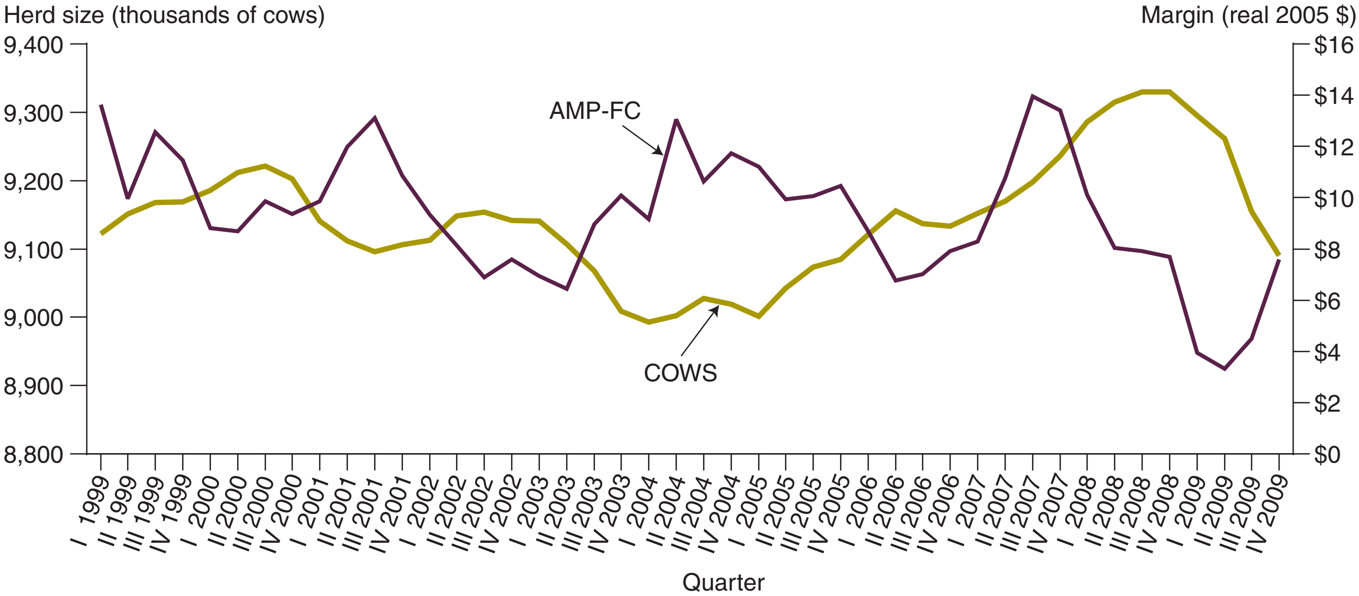


Regression of trend: *PPC* = 439.8 + 17.9* quarter, R^2 = 081.
PPC=Production per cow (pounds of milk).

Source: USDA, Economic Research Service calculations based on USDA, National Agricultural Statistics Service data.

Figure 2

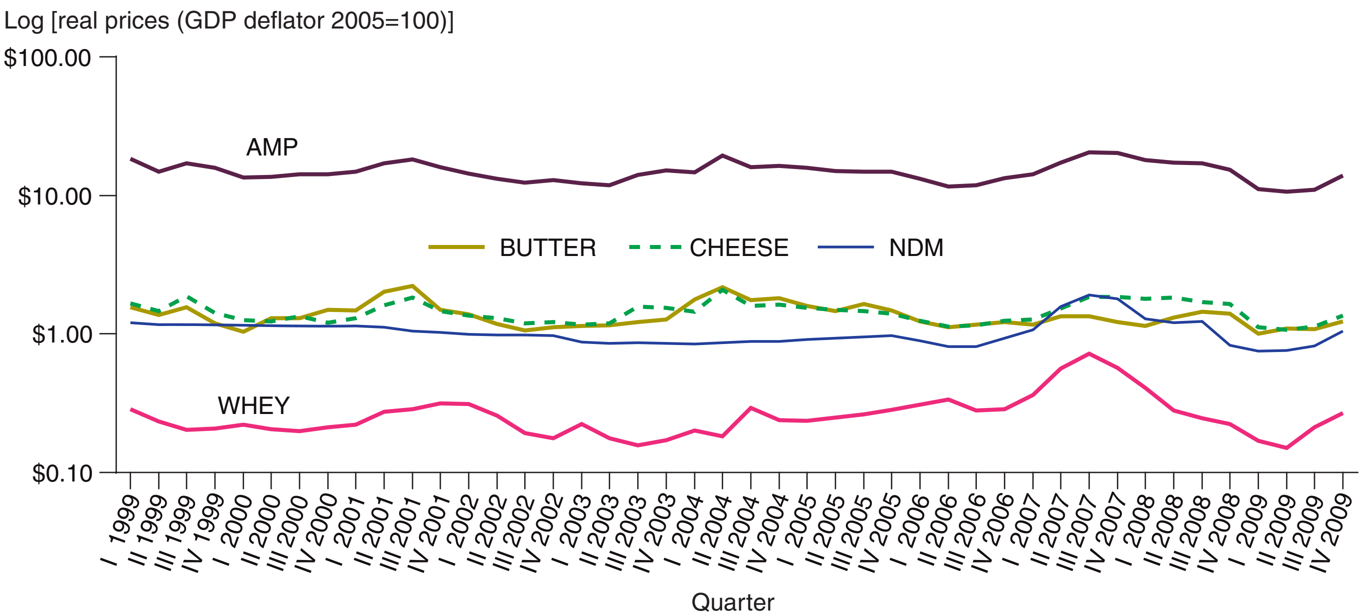
Herd size and margin (all-milk price minus feed cost), Q1/1999-Q4/2009



AMP=All-milk price.
FC=Feed cost.
HS=Herd size.
Correlation: Herd size at t (quarter) and margin (AMP-FC 2005 $) at t = -0.37, significant at 1-percent level.
Correlation: Herd size at t and margin at t-5 = 0.27, significant at 8-percent level.
Source: USDA, Economic Research Service calculations based on USDA, National Agricultural Statistics Service data.

Figure 3

U.S. dairy product prices and all-milk price



AMP=All-milk price.
NDM=Nonfat dry milk.
GDP=Gross domestic product.
Source: USDA, Economic Research Service calculations based on USDA, National Agricultural Statistics Service data.

Table 1
Pairwise correlation of all-milk price and dairy product prices*

Dairy product	All-milk price	Cheese	Butter	NDM
Cheese	0.92*	NA	NA	NA
Butter	0.56*	0.62*	NA	NA
NDM	0.68*	0.48*	-0.003	NA
Whey	0.38*	0.21	0.05	0.60*

NDM=Nonfat dry milk. NA=Not applicable.
*Significant at the 1-percent level.

Source: USDA, Economic Research Service estimates.

products (butter, cheese, NDM, and whey) are especially important because they are used to calculate Class III and Class IV prices,[1] which are part of the ICEC's forecasting activities reported in the WASDE.

Managing milk production and use becomes a crucial industry objective when milk and dairy product prices move in unexpected ways. Public policy focuses on the balance of production and use through Government entities like the Commodity Credit Corporation (CCC), which buys cheese, butter, and NDM, and recently, with concerted action by producers, such as with the herd retirement program of Cooperatives Working Together (CWT).[2]

The dairy sector—from production to final use, including trade and inventory balance—is complex (fig. 4). The analysis highlights five blocks within the dairy industry. The first block deals with the production of farm milk, processing fluid products and manufacturing cheese, NDM, butter, and whey from different economic agents in the sector—farmers and processors. The use of milk for dairy products beyond these four commodities is taken into account in other parts of the model when analyzing aggregate dairy sector components. Both farm milk and dairy products are measured in pounds. The second, third, and fourth blocks model aggregate quantities of dairy products, which help define equilibrium in the dairy sector as a whole (fig. 5). These aggregates are measured on a milk equivalent fat or skim solids basis as pointed out previously. Notably, the milk equivalent measure, which is employed to construct the dairy product aggregates, also allows one to "deconstruct" dairy products back into the original milk they came from.

Two important definitions or identities are usually employed when analyzing the dairy sector as a whole (these variables are measured on a milk-equivalent basis).

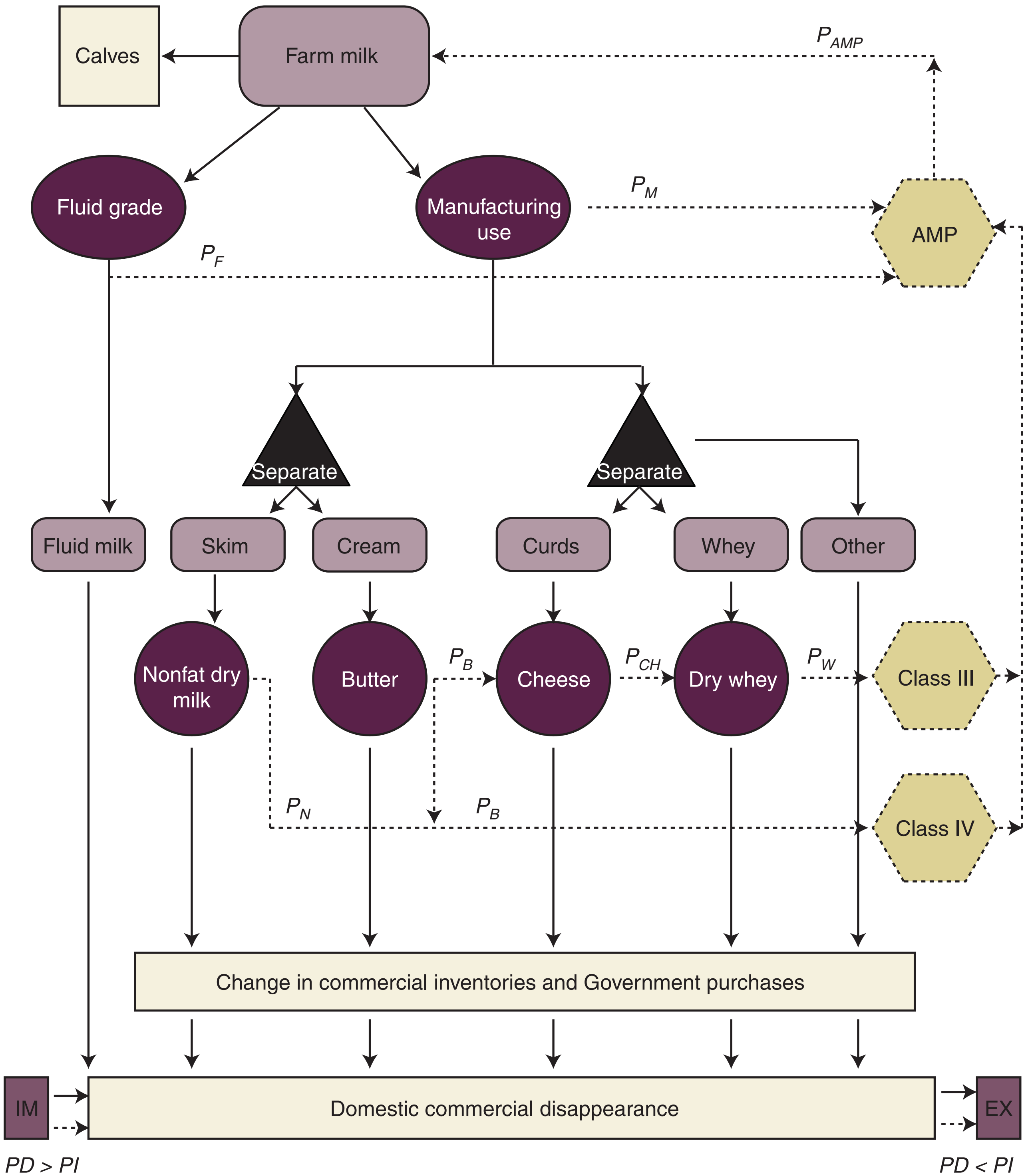
- ***Total supply:*** Milk marketing (cow inventory multiplied by production per cow minus farm use of milk) plus beginning stocks and imports of dairy products.
- ***Domestic commercial disappearance:*** Total supply minus ending commercial stocks, exports, and Government net removals.

Combining these two identities leads to an expression for the difference between overall marketed production (aggregate domestic quantity supplied) and domestic commercial disappearance (aggregate domestic quantity demanded). Excess aggregate supply (or aggregate demand) becomes a

[1]Information on class prices and specific price formulas for the year available at http://www.ams.usda.gov/AMSv1.0/ams.fetchTemplateData.do?template=TemplateG&navID=IndustryMarketingandPromotion&leftNav=IndustryMarketingandPromotion&page=MilkPrices&description=Prices&acct=dmktord.

[2]Balancing milk supply with demand (use) as a public policy objective means converting dairy products into milk equivalent units to aggregate them and to derive aggregate supply and demand for the dairy sector as a whole. Critics of aggregating dairy products into milk equivalent units point to issues like double counting because there is not complete agreement about which products can be considered dairy products and uncertainty about the milk content in some. However, because aggregation is performed by formula and the various components are updated regularly, any systematic errors will be corrected by statistical agencies eventually (Blayney, 2010). Nevertheless, distortions are implied in the aggregation. For example, variable construction procedures assume that imported products have the same composition of fats and skim solids as domestically manufactured ones. A classic exposition for the rationale behind aggregating into milk equivalents is offered in Miller (1989).

Figure 4
Structure of the U.S. dairy sector

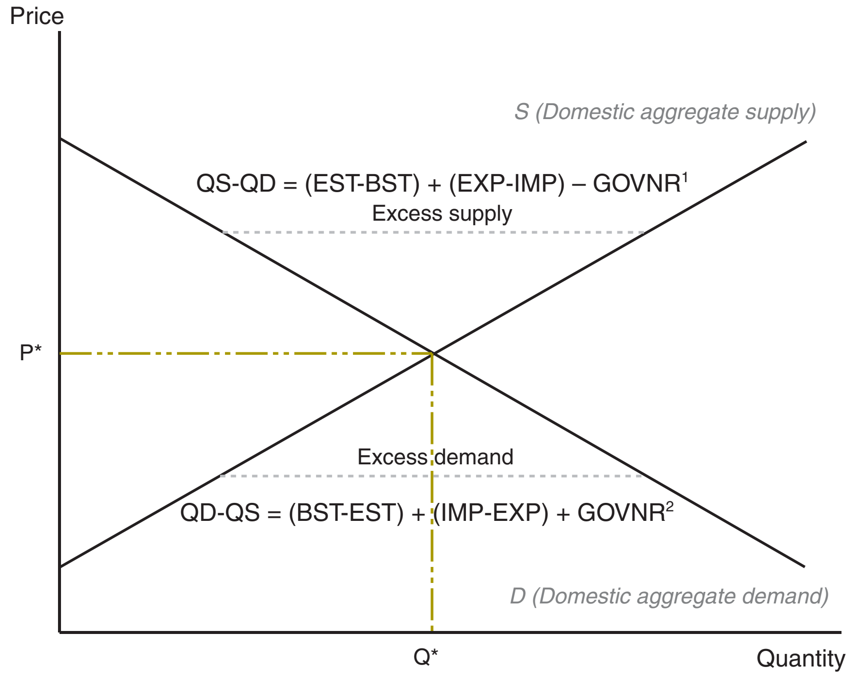


P_{AMP} =All-milk price.
P_F =Price fluid milk.
P_M =Price milk used for dairy products.
P_N =Price nonfat dry milk.
P_B =Price butter.
P_{CH} =Price cheese.
P_W =Price whey.
IM=Dairy product, commercial imports.
EX=Dairy product, commercial exports.
PD=Dairy products, domestic price.
PI=Dairy product, international price.

Source: USDA, Economic Research Service.

combination of increases (decreases) in stocks, exports, and/or Government removals. Figure 5 illustrates aggregate demand and supply for the dairy sector as a whole together with the decomposition of excess aggregate demand and supply. The economic relations and patterns are used to specify a model of the dairy sector, which will be discussed after reviewing a critical selection of dairy forecasting models where the analysis goes beyond correlation and trends to formal modeling in which movements of variables are considered within a *ceteris paribus* (all other things equal) framework.

Figure 5
U.S. dairy industry aggregate supply-demand



P=Price.
Q=Quantity.
S=Supply.
D=Demand.
EST=End of period stocks.
IMP=Imports.
GOVNR=Government net removals.
BST=Beginning of period stocks.
[1]Excess supply ends up as increased stocks, trade surplus, and/or Government Commodity Credit Corporation purchases.
[2]Excess demand is met by decreases in stocks, trade deficit, and/or Government Commodity Credit Corporation sales.

Source: USDA, Economic Research Service.

Comparison of a Selection of Relevant U.S. Dairy Sector Forecasting Models

The ERS model can be distinguished from others, such as that of the USDA's Agricultural Marketing Service (AMS) (2006), in that it is intended to provide short-run forecasts. AMS (2006), in contrast, approximates long-run equilibrium conditions of supply and demand and thus can be used for long-run projections. This type of model is more appropriate for policy analysis where the derived equilibrium multipliers determine the long-term effect associated with a permanent change of an exogenous variable.[3] Bailey et al. (2006) and Bailey (2009) do, however, provide short-term models similar to ERS's in that the dynamic system implied is not solved until the system achieves long-run equilibrium.

The ERS model also differs from Jesse and Schuelke (2002), AMS (2006), Westcott (1986), and Premakumar and Chaudhary (1996) in that it uses a variety of methods to estimate the endogenous variables of interest and selects a "best" forecasting method, or one that most successfully forecasts the variable. In ERS's model, the estimating methods employed are not assumed to be the best *a priori*; they could vary with changes in market conditions or the structure of the sector. Also, ERS uses SM and TS, specifically vector autoregression with exogenous variables (VARX).[4] In VARX models, all variables included in a system of equations (both endogenous and exogenous) affect a particular endogenous variable. In a structural model, by contrast, economic theory plays a key role in the specification of any particular equation in the system.

ERS's model is validated using ex-post forecasting, a procedure that permits the evaluation of the out-of-sample forecasting performance. ICEC requires forecasts for up to eight quarterly periods ahead of the current period. This report forecasts four quarters. Among studies reviewed here, only Westcott (1986) uses out-of-sample or ex-post forecast evaluation. The current ERS model only uses 1 year to validate the estimations and, unlike many other forecasting models of the U.S. dairy sector, the ERS model is quarterly rather than monthly. However, the quarterly timeframe captures the implicit seasonality of the sector and reduces the volatility that is more likely to appear when employing monthly data. Nevertheless, the use of a quarterly model limits the degrees of freedom to specify and estimate the model.[5] As a result, ERS's model specifies only those variables required for the intended purpose—providing estimates required for the WASDE report—as endogenous.

ERS's forecasting model is recursive, as is Jesse and Schuelke's (2002), in that endogenous variables determined at one stage serve as determinants for the next stage. However, the recursiveness is in a block form where sets of endogenous variables determined simultaneously serve as determinants for the next set of equations. Thus, it is an integrated model that encompasses all major sectors of the dairy industry while remaining tractable.

[3]Formally, a simultaneous system of equations (structural model) is one in which the endogenous variables in one equation may appear as an explanatory variable in another equation. A so-called reduced form is obtained when this system is solved so that all endogenous variables are expressed only in terms of the exogenous variables. The system, however, can have lagged endogenous variables, which are considered exogenous in this context. When the system is further solved so that every equation is expressed only in terms of the current values of the endogenous variables and a set of initial conditions, we have a long-run model (see Greene, 2008, pp. 389-94). A broader meaning of a structural model is that it is one derived from economic theory. In contrast, a more data-driven model, such as vector autoregression with exogenous variables (VARX), uses time series where any variable can be introduced just because it works.

[4]In general, in vector autoregression models (both with and without exogenous variables or moving average components), all variables in a system of equations affect a particular endogenous variable. The restriction can be imposed so that variables with insignificant coefficients are set to zero. In an "econometric model," in contrast, specification of an equation plays an important role, and it is easier to test the particular equation's conformity with economic theory. However, as Greene (2008, p. 695) notes, a vector autoregression model can be viewed as a particular kind of seemingly unrelated regression model with all variables included or as a reduced form of a simultaneous equations system. The distinction between the two approaches is blurred since one approach can be made to look exactly like the other. A more general distinction is one between behavioral (derived from a choice problem) and nonbehavioral models.

[5]Models, such as generalized conditional autoregressive heteroskedasticity (GARCH), that Engle (2001) employed to analyze patterns of volatility clustering in financial and similar types of data may be more appropriate to analyze monthly data in the dairy sector.

Model Specification, Estimation, and Validation

Data Sources and Variable Construction

Table 2 provides descriptive statistics for the variables employed in the econometric model, and table 3 presents the associated data sources. The endogenous variables constitute the set of variables required by ICEC to carry out its role. An ICEC-identified data source is provided shortly before each monthly meeting. These variables constitute information that goes directly into the quarterly model to generate forecasts. Another set of exogenous variables not provided by ICEC also must be forecasted forward for the ERS model. For example, the model uses supply equations to forecast prices, which require forecasts of market quantities of cheese, whey, NDM, and butter. The model does not estimate price and quantity simultaneously[6] for dairy products but it does estimate overall supply and demand for the dairy sector (see fig. 5). The model specification identifies supply equations for manufactured dairy products so as to forecast their prices. The model also requires international prices to forecast forward exports, imports, and stocks that are crucial components of aggregate supply and demand (see fig. 5). Except for the energy index (see Appendix A), the variables in table 2 require little manipulation from the sources specified in table 3. The VARX methodology (see Appendix B) is employed to forecast this latter set of variables.

[6]As more data becomes available, the model will be respecified to contain both supply and demand for the dairy products in block 1.

Specification

The SM model is composed of five blocks of behavioral equations and identities (table 4). The first block contains farm milk supply (equations 1.1 and 1.2); dairy product—cheese, NDM, butter, and whey—inverse supply functions (equations 1.3-1.6); and an all-milk price equation (equation 1.7), which is an inverse-derived demand equation for farm milk. The second, third, and fourth blocks provide estimates of stocks, imports, exports, and net removals of aggregate dairy products on a fat basis and a skim solid basis (see fig. 5) (equations 2.1, 2.2, 2.3, 3.1, 3.2, 3.3, 4.1, and 4.2). Finally, the fifth block defines milk production and marketing (equations 5.1 and 5.2), aggregate supply (equation 5.3), domestic commercial disappearance (equation 5.4), and the overall balance or equilibrium (equation 5.5) (see fig. 5).

The VARX model structure is similar in structure to the SM. Endogenous and exogenous variables in each block are the same for both approaches. For VARX models, however, every variable directly influences every other variable. Two time series models are estimated—a restricted and an unrestricted version. In the restricted model, the insignificant coefficients from the unrestricted model have been eliminated.

The blocks are interrelated in a variety of ways in both the SM and the VARX models. At the specification stage, for example, the same butter and NDM domestic prices used in block 1 are used in blocks 2 and 3 to calculate domestic-international price differentials or are used in block 4 to calculate domestic-support price differences. In the forecasting stage, the prices forecasted in block 1 are connected in the same way to blocks 2, 3, and 4 as in the specification stage, (i.e., they are employed to calculate price differentials). Overall, there are 15 behavioral equations and 5 identities.

Table 2
Summary statistics for variables used in specifying quarterly dairy forecasting models, Q4/1998-Q1/2009

Definition and (variable)	Mean	Standard deviation	Definition and (variable)	Mean	Standard deviation
Endogenous variables					
Number of cows (COW)	9143	88	Beginning stocks fat (STBFAT)	9.98	2.19
Production per cow (PPC)	4758	252	Imports fat (IMPFAT)	1.21	0.21
Cheese price (CHP)	1.31	0.23	Exports fat (EXPFAT)	0.63	0.30
Butter price (BTRP)	1.25	0.28	Net removals fat (GOVNRF)	0.07	0.12
Nonfat dry milk (NDM) price (NDMP)	0.94	0.22	Beginning stocks skims (STBSS)	9.13	1.07
Whey price (WHP)	0.24	0.10	Imports skims (IMPSS)	1.15	0.18
All Milk Price (AMP)	13.85	1.89	Exports skims (EXPSS)	3.64	1.56
			Net removals stock skims (GOVNRS)	1.23	1.14
Exogenous and Auxiliary Variables					
Price 16 percent Pr. Ration (FC)	4.99	1.08	Gross domestic product (GDP)	10617	762
Slaughter cow price (SCP)	42.44	4.49	First quarter (DQ1)	0.26	0.44
Cheese quantity (CHQ)	2.46e+08	5.10e+07	Second quarter (DQ2)	0.24	0.43
Whey quantity (WHQ)	1.38e+08	2.65e+07	Third quarter (DQ3)	0.24	0.43
NDM quantity (NDMQ)	2.51e+08	7.39e+07	Time trend (t)	24	13
Butter quantity (BTRQ)	4.89e+07	8.83e+06	Time trend2 (t2)	728	651
Food industry wage (WG)	8.57	0.69	Butter domestic supply price difference (PRSPBTR)	0.38	0.25
Energy index (EI)	1.56	0.41	Cheese domestic supply price difference (PRSPCH)	-0.64	0.30
T-Bill (TB)	2.86	1.85	Farm use of milk (FUSE)	0.30	0.02
Butter domestic Int. price difference (BTRPDI)	-0.21	0.85	NDM domestic Int. price difference (NDMPDI)	0.05	0.33

Source: USDA, Economic Research Service.

Table 3
Data sources and variable units of measurement for quarterly dairy forecasting models[1]

Variable (units)	Source (website)	Variable (units)	Source (website)
COW (thousands)	Milk Production (http://usda.mannlib.cornell.edu/usda/current/MilkProd)	STBFAT (ME)	Dairy at a Glance (http://www.ers.usda.gov/Publications/ldp/)
PPC (pound/ head)	Milk Production (http://usda.mannlib.cornell.edu/usda/current/MilkProd)	IMPFAT (ME)	Dairy at a Glance (http://www.ers.usda.gov/Publications/ldp/)
CHP ($/pound)	Dairy Product Prices (http://usda.mannlib.cornell.edu/usda/current/ DairProdPr/)	EXPFAT (ME)**	(http://www.fas.usda.gov/)
BTRP ($/pound)	Dairy Product Prices (http://usda.mannlib.cornell.edu/usda/current/Dair-ProdPr/)	GOVNRF (ME)	Dairy at a Glance (http://www.ers.usda.gov/Publications/ldp/)
NDMP ($/pound)	Dairy Product Prices (http://usda.mannlib.cornell.edu/usda/current/ DairProdPr/)	STBSS (ME)	Dairy at a Glance (http://www.ers.usda.gov/Publications/ldp/)
WHP ($/pound)	Dairy Product Prices (http://usda.mannlib.cornell.edu/usda/current/ DairProdPr/)	IMPSS (ME)	Dairy at a Glance (http://www.ers.usda.gov/Publications/ldp/)
AMP ($/CWT)	Agricultural Prices (http://usda.mannlib.cornell.edu/usda/current/AgriPric/)	EXPSS (ME)**	(http://www.fas.usda.gov/)
		GOVNRS (ME)	Dairy at a Glance (http://www.ers.usda.gov/Publications/ldp/)
TB (%) ***	ICEC	GDP Def. **	ICEC
MILC ($/cwt) *	Price Support—Milk Income Loss Contract (http://www.fsa.usda.gov/FSA/ webapp?area=home&subject=prsu&topic=mpp-mi)	BTRPP ($/mt) **	Dairy at a Glance (http://www.ers.usda.gov/Publications/ldp/)
FC ($/cwt)***	ICEC	NDMPP ($/mt.) **	Dairy at a Glance (http://www.ers.usda.gov/Publications/ldp/)
SCP ($/ cwt)***	ICEC	GDP2005 ($) ***	ICEC
CHQ (lb.) **	Dairy Product Prices (http://usda.mannlib.cornell.edu/usda/current/ DairProdPr/)	DQ1 (D. Var.)	N/A
WHQ (lb.) **	Dairy Product Prices (http://usda.mannlib.cornell.edu/usda/current/ DairProdPr/)	DQ2 (D. Var.)	N/A
NDMQ (lb.) **	Dairy Product Prices (http://usda.mannlib.cornell.edu/usda/current/ DairProdPr/)	DQ3 (D. Var)	N/A
BTRQ (lb.) **	Dairy Product Prices (http://usda.mannlib.cornell.edu/usda/current/ DairProdPr/)	t (trend)	N/A
WG ($/hour)**	Average Hourly Compensation Production Workers (www.bls.gov)	t2 (trend sq)	N/A
EI (N/A) **	ERS from data from www.eia.gov		

mt=Metric ton. ME=Milk equivalent.
* External forecast (documented in report). ** Internal forecast and/or calculation (documented in report). *** Forecast provided by the Inter-agency Commodity Estimates Committee.

Source: USDA, Economic Research Service.

Table 4

Specification of U.S. quarterly dairy model, Q4/1998-Q1/2009

Milk production and inverse supply for dairy products block system statistics

Method	Equation name	RMSE (in sample)	FSTAT/ CHI2	Probability
OLS 2SLS 3SLS SURE	1.1 COW	0.0020 0.0020 0.0018 0.0018	130.07 130.07 1191.41 1191.41	0.0000
“	1.2 PPC	0.012 0.012 0.011 0.011	122.45 122.45 791.28 791.28	“
“	1.3 CHP	0.12 0.12 0.11 0.11	5.14 5.14 51.35 51.35	“
“	1.4 BTRP	0.11 0.11 0.10 0.10	10.38 11.38 133.99 120.96	“
“	1.5 NDMP	0.13 0.13 0.12 0.12	8.98 8.98 91.44 90.70	“
“	1.6 WHP	0.17 0.17 0.15 0.15	16.29 16.29 219.80 218.56	“
“	1.7 AMP	0.027 0.028 0.027 0.024	110.17 103.39 1148.56 1163.64	“

RMSE=Root-mean-square error.
FSTAT/CHI2= F-Statistic (OLS, 2SLS)/Chi Squared (3SLS,SURE).
OLS=Ordinary least squares.
2SLS=Two-stage least squares.
3SLS=Three-stage least squares.
SURE=Seemingly unrelated regression.

Note: Rank condition is satisfied for each of the seven equations in this block (i.e. the system is identified and hence unique values of the structural parameters can be derived from the reduced form of the system).

Source: USDA, Economic Research Service estimates.

Continued—

The specification of the SM and the VARX models only employed data from Q4/1998 to Q1/2009. Model equations were specified in logarithmic form, permitting the determination of elasticities from the estimated parameters and a way to assess whether their signs made theoretical sense, especially in the case of the SM. Additionally, the log transformation of the data may have a variance stabilizing effect (Lutkepohl and Xu, 2009).[7] Once specified, the *entire* system of 15 equations grouped in 5 blocks was estimated recursively using 6 different estimation methods that generated 6 different point forecasts for each endogenous variable. The limited number of observations used in the estimations did not permit estimation of the entire structure simultaneously. Rank conditions were used to identify each block so that unique values of structural parameters could be derived from the reduced form of the system.

These estimates were compared across methods as well as with the point estimates agreed upon by ICEC. The final specification (estimation) of the SM was performed using SURE,[8] the method with the lowest root mean squared error (RMSE) compared with the three other model estimation techniques—OLS, 2SLS, and 3SLS. The VARX model was also estimated in five blocks parallel to the SM (i.e., the same endogenous and exogenous variables were employed to estimate each block). Moreover, the VARX model was used to estimate two models—one restricted[9] and one free. The forecasting performance of each method was subsequently determined.

Table 4 presents the SM specification of the model. The endogenous variables in this table constitute the monthly forecasting needs mentioned previously. The forecasts are quarterly estimates of these variables. The variables in the different equations were selected according to the role each equation plays within the overall model. Each block was estimated by OLS, 2SLS, 3SLS, and SURE. The results presented correspond to the method that minimized in-sample RMSE in a majority of cases, SURE, relative to other methods. Parameter estimates are also shown for comparison purposes across methods for equations 1.1 and 1.7. The parameters in these two important equations are fairly stable across methods.

Equation 1.1 represents cow inventory. Total milk production is a reflection of the number of cows and their productivity (defined as production per cow) (see table 4). As in any production system, inputs are transformed into outputs using a given technology, in a given environment, and within a specific set of constraints. Two important factors were modeled explicitly in this set of equations—feed costs and the all-milk price. Coefficients with a t-value greater than 1 were considered for specification purposes to decrease type II errors as suggested in Kennedy (2008, p. 90). Also, on theoretical grounds, some variables with t-values less than 1 were considered.

The number of cows is expected to be a function of lagged cow inventory (COW), lagged price of milk (AMP), and lagged feed costs (FC). Real values for these price and cost variables are determined by dividing current magnitudes by the GDP deflator. In the case of the AMP variable, direct payments, such as those from the Milk Income Loss Contract (MILC) program, were added before converting the all-milk price variable to a real value to model the farmer's incentive or effective price. The variable AMP is expected to have a positive lagged effect on the number of cows. It can be expected that

[7]Unit root test resulted in some variables being nonstationary. Typical characteristics of nonstationary data include: mean that varies widely, variances that explode, and shocks that appear permanent are not, with the implication that forecasts might be misleading. Improvements to the ERS forecasting model will evaluate the costs and benefits of addressing this issue by implementing nonstationary econometric methods.

[8]For variance structure, see Appendix C.

[9]Note that this exclusion could compromise the *ceteris paribus* context if the model were developed further and used for policy analysis.

as FC (or approximately 60 percent of dairy farm operating costs according to the Agricultural Resource Management Survey) increases, all other things being equal, there will be a lagged decrease in herd size. Both AMP and FC effects are related to biological lags, culling, and replacements apart from the other variables. In the double logarithm specification of the equations, the coefficients can be interpreted as elasticities. Most importantly, the coefficient for AMP is 0.01, meaning that a 1-percent increase in AMP will result in a 0.01-percent increase in cow numbers after two quarters. The coefficient FC is -0.007, implying that a 1-percent increase in FC will result in a 0.007-percent reduction in cow numbers after two quarters. Also, a 1-percent increase in slaughter cow price will result in a 0.0009-percent decrease in herd size after two quarters. Cow numbers then follow a cyclical behavior related to the AMP/FC price ratio (see fig. 2) using a *ceteris paribus* assumption and not just as a correlation. The supply response is quite inelastic, reflecting the changing structure of U.S. dairy farms. Larger farms with more capital costs need more milk to cover expenses and therefore have a lower supply response.

Equation 1.2 captures the combined effect of new technologies and factors like scale economies on cow productivity. Generally, larger operations have higher cow productivity, represented by the variable production per cow (PPC). The effect of AMP and FC on PPC is positive and negative, respectively. A 1-percent rise in AMP will result in a 0.03-percent increase in production per cow, and a 1-percent increase in FC leads to a 0.04-percent decrease in PPC. The variable DQ2 captures the effect of the spring flush on cow productivity. This seasonal effect means that spring productivity is higher by about 0.03 percent relative to the fourth quarter—the excluded "base" quarter—and holding other factors constant. The t term indicates that PPC has been increasing throughout the period at about 0.006 percent per year (see fig. 1).

Equations 1.3, 1.4, 1.5, and 1.6 represent inverse (price dependent) supply equations for cheese, NDM, butter, and whey, respectively. As such, they have some theoretical properties. An inverse supply equation is positive with respect to its own quantity and positive with respect to input prices.[10] The sign of related products will depend on whether they complement each other in production. A positive sign, as is the case in all the inverse supply equations in this study, reflects that the products are either byproducts or that there is some interrelation in their production as seen in the dairy industry. In the estimation and validation stage, a simultaneous system structure acknowledges explicitly that the production processes of these goods are related. The macro input prices—wages, interest rates, and energy—as well as the dairy commodity quantities have a strong positive effect on dairy commodity prices, though only the cheese quantity is significant.

The equation for AMP (equation 1.7) is an inverse-derived demand for farm milk. This type of demand results from the final demand (for cheese, NDM, etc.) that farm milk produces. Hence, a higher dairy product price indicates a short dairy product supply relative to demand, which in turn signals a need for more milk to make dairy products. Hence, commodity prices (inverse to dairy product quantity) are positive derived demand shifters. The effect of the price of cheese on AMP is the largest among all dairy products, followed by NDM and butter. Whey, a byproduct of cheese production, enters the equa-

[10]According to Varian (1992, p. 216), an inverse supply function "measures the price that must prevail in order for a firm to find it profitable to supply a given amount of output." In a one output and one input case, for example, if the price of the input increases, the price of the output needs to increase for the firm to keep providing the same level of output, *ceteris paribus*. Also, given that the output price is greater than the average variable cost, and that marginal cost is increasing, an increase in the level of output will need a higher output price, all other things equal. Hence, the output quantity and input price coefficients are positive in an inverse supply function.

Table 4

Specification of U.S. quarterly dairy model, Q4/1998-Q1/2009—continued

SURE Estimates of behavioral equations[1] block 1

1.1 Cow inventory (COW):

$$\ln COW_t = \beta_1 \ln COW_{t-1} + \beta_2 \ln COW_{t-2} + \beta_3 \ln AMP_{t-2} + \beta_4 \ln FC_{t-2} + \beta_5 \ln SCP_{t-2} + \beta_6 t + \beta_7 t^2 + \beta_o$$

OLS	1.18	-0.44	0.01	-0.007	-0.0009	-0.0007	0.00002	2.38
	(8.11)	(-2.95)	(2.54)	(-1.90)	(-2.16)	(-2.48)	(2.93)	(2.84)
2SLS	1.17	-0.44	0.01	-0.007	-0.009	-0.0007	0.00002	2.38
	(8.11)	(-2.95)	(2.54)	(-1.90)	(-2.16)	(-2.48)	(2.93)	(2.84)
3SLS	1.15	-0.44	0.01	-0.005	-0.01	-0.0007	0.00002	2.69
	(9.60)	(-3.67)	(2.81)	(-1.68)	(-2.71)	(-2.83)	(3.37)	(3.94)
SURE	1.17	-0.45	0.01	-0.005	-0.01	-0.0006	0.00002	2.56
	(9.80)	(-3.74)	(2.81)	(-1.70	(-2.69)	(-2.60)	(3.13)	(3.75)

1.2 Production per cow (PPC):

$$\ln PPC_t = \beta_1 \ln PPC_{t-2} + \beta_2 \ln AMP_{t-2} + \beta_3 \ln FC_{t-2} + \beta_4 DQ2 + \beta_5 t + \beta_o$$

-0.36	0.03	-0.04	0.03	0.006	11.39
(-4.82)	(1.58)	(-2.62)	(7.19)	(16.29)	(17.89)

1.3 Inverse cheese supply (CHP):

$$\ln CHP_t = \beta_1 \ln BTRP_{t-1} + \beta_2 \ln WG_{t-2} + \beta_3 \ln EI_{t-2} + \beta_4 \ln TB_{t-2} + \beta_5 \ln CHQ_{t-4} + \beta_6 DQ1 +$$

0.56	3.47	0.52	0.11	0.43	-0.07
(5.75)	(2.86)	(1.64)	(2.56)	(2.25)	(-2.22)

$$\beta_7 t + \beta_8 t^2 + \beta_o$$

-0.05	0.001	-15.60
(-2.86)	(3.31)	(-3.88)

1.4 Inverse supply butter (BTRP):

$$\ln BTRP_t = \beta_1 \ln CHP_t + \beta_2 \ln NDMP_t + \beta_3 \ln WG_t + \beta_4 \ln EI_t + \beta_5 \ln TB_t + \beta_6 \ln BTRQ_{t-4} +$$

1.22	-0.58	3.71	1.42	0.18	0.07
(8.55)	(-4.28)	(3.07)	(4.35)	(3.12)	0.72)

$$\beta_7 t + \beta_8 t^2 + \beta_o$$

-0.05	0.0009	-9.48
(-3.38)	(2.46)	(-2.90)

1.5 Inverse supply NDM (NDMP):

$$\ln NDMP_t = \beta_1 \ln CHP_{t-1} + \beta_2 \ln WG_{t-1} + \beta_3 \ln EI_{t-1} + \beta_4 \ln TB_{t-1} + \beta_5 \ln NDMQ_{t-4} + \beta_6 t + \beta_7 t^2 + \beta_o$$

0.47	6.09	0.73	0.39	0.13	-0.07	0.002	-16.22
(3.66)	(4.36)	(1.86)	(7.61)	(1.43)	(-4.02)	(5.23)	(-4.69)

1.6 Inverse supply whey (WHP):

$$\ln WHP_t = \beta_1 \ln WHP_{t-2} + \beta_2 \ln BTRP_{t-1} + \beta_3 \ln WG_{t-1} + \beta_4 \ln EI_{t-1} + \beta_5 \ln TB_{t-1} + \beta_6 WHQ_{t-1} +$$

-0.43	0.78	13.16	1.66	0.93	0.33
(-4.16)	(6.18)	(7.29)	(3.65)	(11.97)	(1.58)

$$\beta_7 t + \beta_8 t^2 + \beta_o$$

-0.07	0.003	-38.59
(-3.16)	(5.72)	(-7.53)

Continued—

Table 4

Specification of U.S. quarterly dairy model, Q4/1998-Q1/2009—continued

SURE Estimates of behavioral equations[1] block 1

1.7 All-milk price (AMP):

$$\ln AMP_t = \beta_1 \ln COW_{t-1} + \beta_2 \ln PPC_{t-1} + \beta_3 \ln NDMP_t + \beta_4 \ln CHP_t + \beta_5 \ln WHP_{t-1} +$$

	β_1	β_2	β_3	β_4	β_5
OLS	-2.37	-0.52	0.20	0.48	0.04
	(-2.59)	(-3.68)	(4.50)	(8.35)	(1.70)
2SLS	-2.23	-0.53	0.16	0.53	0.06
	(-2.44)	(-3.86)	(3.16)	(8.46)	(2.14)
3SLS	-1.46	-0.58	0.15	0.60	0.06
	(-1.98)	(-5.17)	(3.74)	(12.04)	(2.77)
SURE	-1.66	-0.56	0.18	0.52	0.05
	(-2.22)	(-4.84)	(5.17)	(11.22)	(2.54)

$$\beta_6 BTRP_t + \beta_7 \ln FC_{t-1} + \beta_o$$

	β_6	β_7	β_o
OLS	0.04	0.16	28.30
	(1.24)	(3.04)	(3.35)
2SLS	0.02	0.14	27.55
	(0.55)	(2.65)	(3.27)
3SLS	-0.009	0.10	20.91
	(-0.28)	(2.33)	(3.08)
SURE	0.03	0.13	22.30
	(1.15)	(2.97)	(3.22)

[1]For OLS and 2SLS, t statistics are shown in parentheses; for 3SLS and SURE, z statistics.

Note: Parameter estimates for other than preferred method are shown.

Beginning inventory, imports, and exports (fat basis) block
System statistics

Method	Equation name	RMSE (in sample)	FSTAT/ CHI2	Probability
	2.1 STBFAT	0.12	27.60	0.0000
OLS		0.12	27.60	0.0000
2SLS		0.11	96.30	0.0000
3SLS		0.11	96.30	0.0000
SURE				
	2.2 IMPFAT	0.16	7.45	
“		0.16	7.45	0.0009
		0.17	16.10	0.0003
		0.17	16.10	0.0001
				0.0001
“	2.3 EXPFAT	0.30	4.94	0.0090
		0.30	4.94	0.0090
		0.29	10.34	0.0060
		0.29	10.34	0.0060

RMSE=Root-mean-square error.
FSTAT/CHI2= F-Statistic (OLS, 2SLS)/Chi Squared (3SLS,SURE).
OLS=Ordinary least squares; 2SLS=Two-stage least squares; 3SLS=Three-stage least squares.
SURE=Seemingly unrelated regression.

Continued—

tion with a lag. The elasticities' magnitudes parallel the correlations shown in table 1 and figure 3 in a *non ceteris paribus* framework, a likely reflection of their share of use as well as the underlying market effects like the interaction of farm supply and processor demand for dairy products. This trend translates into an increase in derived demand for milk (rightward shift). The supply curve in this context comes from equations 1.1 and 1.2.

In equation 1.7, total farm milk processed is decomposed into PPC and number of cows, and their respective price (AMP) elasticities show the effect of productivity on AMP as being very inelastic and that of the number of cows as very elastic (i.e., an increase in cow productivity has a larger downward effect on AMP than a comparable percentage change increase in the dairy herd, all other things equal). Feed price increases, *ceteris paribus,* decrease supply (less milk supply at the same price—leftward supply shift), requiring a higher AMP in this equation to equilibrate supply with derived demand.

Blocks 2, 3, and 4 represent aggregate quantities of dairy products and, as such, discussing them in connection to figures 4 and 5 is useful. Equations 2.1, 2.2, and 2.3 have the variable BTRPDI in common, which is the real difference between domestic and international butter prices and which affects movements in stocks, imports, and exports (see fig. 4). An increase in the difference between domestic and international prices may result in a tightening of the domestic market relative to the international market, which will lead to a reduction in stocks on a fat basis. The same analysis could be done for imports, where an increase in BTRPDI will lead to an increase in imports and a decrease in exports. A parallel analysis could be done for the effect of NDMPDI—the real difference between domestic and international NDM prices—on stocks, imports, and exports when examining equations 3.1, 3.2, and 3.3. By using domestic prices for butter and NDM, an explicit connection is made to block 1 of the model.

Net removals for fat and skims are estimated using a Tobit regression with lower and upper bounds equal to the maximum positive observation and maximum negative observation respectively.[11] Government net removals could be viewed as a "corner solution" as are purchases, such as cars and houses, in which an individual's consumption might be positive one year and zero at other times. Since net Government purchases of dairy products can be positive or negative, a two-limit Tobit model is employed. The variables PRSPBTR and PRSPCH represent the differences between domestic and support butter and NDM prices. Equations for both these variables show that an increase in GDP leads to a decrease in removals, and that PRSPCH is significant in explaining a decrease in net removals of skim solids. Details of this estimation are presented in Appendix D. Results of an alternative way to model net removals—hurdle or two-part model—are presented as well.

Figure 5 summarizes the aggregate supply and demand system for the dairy sector. Net export, inventory, and Government purchases bring the system into balance in each period. If equilibrium is not desirable because of low average prices for dairy products and AMP, then a policy like the most recent herd retirement program from the National Milk Producers Federation, Cooperatives Working Together (CWT)[12] would shift aggregate supply to the left and dairy prices would increase. Similarly, an increase in net exports,

[11]The censoring values for the lower and upper limit are the minimum and maximum observed magnitudes in the data. This is STATA's default if no numerical limits are specified.

[12]CWT is a voluntary farmers' organization whose stated objective is to "strengthen and stabilize milk prices by balancing supply and demand." See http://www.cwt.coop for more information.

Table 4

Specification of U.S. quarterly dairy model, Q4/1998-Q1/2009—continued

SURE estimates of behavioral equations block 2

2.1 Beginning stocks (STBFAT):

$$\ln STBFAT_t = \beta_1 BTRPDI_{t-1} + \beta_2 DQ_1 + \beta_3 DQ_3 + \beta_o$$

-0.21	-0.22	0.20	2.41
(-5.73)	(-5.23)	(4.55)	(83.31)

2.2 Imports (IMPFAT):

$$\ln IMPFAT_t = \beta_1 \ln IMPFAT_{t-4} + \beta_2 BTRPDI_{t-1} + \beta_o$$

0.36	0.17	0.04
(1.95)	(3.05)	(0.73)

2.3 Exports (EXPFAT):

$$\ln EXPFAT_t = \beta_1 \ln EXPFAT_{t-1} + \beta_2 BTRPDI_{t-3} + \beta_o$$

0.31	-0.20	-0.27
(2.07)	(-2.02)	(-2.85)

Beginning inventory, imports, and exports (skim solids basis) block
System statistics

Method	Equation name	RMSE (in sample)	FSTAT/ CHI2	Probability
	3.1 STBSS	0.072	11.20	0.0000
OLS		0.072	11.20	0.0000
2SLS		0.069	25.21	0.0000
3SLS		0.069	25.21	0.0000
SURE				
	3.2 IMPSS	0.138	8.35	0.0004
“		0.138	8.35	0.0004
		0.132	19.75	0.0001
		0.132	19.75	0.0001
“	3.3 EXPSS	0.146	124.94	0.0000
		0.146	124.94	0.0090
		0.140	272.39	0.0000
		0.140	272.39	0.0000

RMSE=Root-mean-square error.
FSTAT/CHI2= F-Statistic (OLS,2SLS)/Chi Squared (3SLS,SURE).
OLS=Ordinary least squares.
2SLS=Two-stage least squares.
3SLS=Three-stage least squares.
SURE=Seemingly unrelated regression.
Note: Rank condition is satisfied for each of the three equations in this block.

Continued—

Table 4

Specification of U.S. quarterly dairy model, Q4/1998-Q1/2009—continued

SURE estimates of behavioral equations block 3

3.1 Beginning stocks (STBSS):

$$\ln STBSS_t = \beta_1 \ln NDMPDI_{t-1} + \beta_2 DQ3 + \quad \beta_o$$

-0.16	0.10	2.20
(-2.87)	(3.86)	(174.36)

3.2 Imports (IMPSS):

$$\ln IMPSS_t = \beta_1 \ln IMPSS_{t-4} + \beta_2 \ln NDMPDI_{t-1} + \beta_o$$

0.62	0.21	0.04
(4.20)	(1.92)	(1.27)

3.3 Exports (EXPSS):

$$\ln EXPSS_t = \beta_1 \ln EXPSS_{t-1} + \beta_2 \ln NDMPDI_{t-3} + \beta_o$$

0.79	-0.33	0.29
(8.51)	(-1.88)	(2.42)

Continued—

Table 4

Specification of U.S. quarterly dairy model, Q4/1998-Q1/2009—continued

Net removals fats and skim solids block 4

4.1 Net removals fat basis (GOVNRF):

$$GOVNRF_t = \beta_1 GOVNRF_{t-1} + \beta_2 GOVNRF_{t-2} + \beta_3 PRSPBTR_t + \beta_4 PRSPBTR_{t-1} + \beta_5 \ln GDP_{t-2} + \beta_o$$

0.68	-0.43	-0.09	-0.06	-0.00007	0.86
(4.21)	(-2.87)	(-1.40)	(-0.79)	(-2.92)	(3.08)

Censored: lower bound -0.1, upper bound 0.6
Log-likelihood (LL) = 39.22

Alternative two-part specification:

Part 1:
Decision: Decision to remove fats (DRF = 1) if GOVNRF ≠ 0

Probit regression

$$DRF_1 = \beta_1 GOVNRF_{t-1} + \beta_2 PRSPBTR_{t-1} + \beta_3 \ln GDP_{t-2} + \beta_o$$

12.22	2.50	-0.001	0.86
(1.45)	(1.26)	(-1.87)	(3.08)

LL= -12.12

Part 2:
Outcome level (given DRF = 1)

$$GOVNRF_t = \beta_1 GOVNRF_{t-1} + \beta_2 GOVNRF_{t-2} + \beta_3 PRSPBTR_t + \beta_o$$

0.76	-0.36	-0.09	0.10
(4.03)	(-1.97)	(-1.40)	(2.48)

Joint decision and outcome LL = 15.30.

Continued—

Table 4

Specification of U.S. quarterly dairy model, Q4/1998-Q1/2009—continued

Net removals fats and skim solids block 4

4.2 Net removals block skim solids basis (GOVNRS):

Tobit regression

$$GOVNRS_t = \beta_1 GOVNRS_{t-1} + \beta_2 GOVNRS_{t-2} + \beta_3 PRSPCH_t + \beta_4 \ln GDP_{t-2} + \beta_o$$

0.72	-0.48	-2.17	-0.0003	4.28
(5.00)	(-3.69)	(-4.08)	(-1.61)	(2.46)

Censored: lower bound -0.4, upper bound 3.5

LL = -38.23

Alternative two-part specification:

Part 1:

Decision: Decision to remove skim solids (DRS = 1) if GOVNRS ≠ 0

Probit regression

$$DRS_t = \beta_1 PRSPCH_{t-1} + \beta_2 \ln GDP_{t-2} + \beta_o$$

-3.27	-0.002	27.88
(2.01)	(-2.01)	(2.43)

LL= -6.53

Part 2:

Outcome level (given DRS = 1)

$$GOVNRS_t = \beta_1 GOVNRS_{t-2} + \beta_2 GOVNRS_{t-2} + \beta_3 PRSPCH_1 + \beta_o$$

0.66	-0.34	-2.52	1.57
(4.28)	(-2.47)	(-3.78)	(5.18)

Joint decision and outcome LL = -38.29

Continued—

Table 4

Specification of U.S. quarterly dairy model, Q4/1998-Q1/2009—continued

Milk production, marketing, total supply, domestic commercial disappearance and overall balance (fat and skim solid basis) identities block 5

5.1 Milk production (Milk):

$$MILK_t = COW_t * PPC_t$$

5.2 Marketing (MKT):

$$MKT_t = MILK_t - FUSE_t$$

5.3 Total supply (TS):

$$TS_{ti(i=FAT,SS)} = MKT_t + STB_{ti(i=FAT,SS)} + IMP_{ti(i=FAT,SS)}$$

5.4 Domestic commercial disappearance (DCD)

$$DCD_{ti(i=FAT,SS)} = TS_{ti(i=FAT,SS)} - STE_{ti(i=FAT,SS)} - EXP_{ti(i=FAT,SS)} - GOVNR_{ti(i=FAT,SS)}$$

5.5 Overall balance

$$0 = (MILK_t - FUSE_t - DCD_{ti(i=FAT,SS)} - GOVNR_{ti(i=FAT,SS)}) + (STB_{ti} - STE_{ti})_{(i=FAT,SS)} + (IMP_{ti} - EXP_{ti})_{(i=FAT,SS)}$$

Note: Total supply, domestic commercial disappearance, and overall balance are in both a fat and skim solids basis and are calculated independently.

Source: USDA, Economic Research Service estimates.

which programs like the current CWT's export promotion program are designed to encourage, would shift aggregate demand right. The sum of the production consumption balance, net inventory change, net exports, and net removals is zero in every period. The increase in domestic production over domestic consumption results in a combination of increases of product inventories, net exports, or net removals (fig. 5).

Estimation and Validation

The following methods were employed to estimate the dairy quarterly model: Ordinary least squares (OLS), two-stage least squares (2SLS), seemingly unrelated regression (SURE), three-stage least squares (3SLS), and constrained and unconstrained vector autoregression with exogenous variables (VARX). The OLS, 2SLS, SURE, and 3SLS approaches rely on the least-squares method. The VARX procedure is estimated by iterated seemingly unrelated regression. OLS and 2SLS are single equation methods. SURE, 3SLS, and VARX use a system of equations, VARX specifically uses a time-series method. The only approach that clearly dominates is VARX, although this superiority is not complete, and it might change. The robustness of the result has been maintained through several iterations of this report, where the data has been revised as more observations were added.

Single equation methods do not account for potential error correlations across equations, making them asymptotically less efficient than full information methods. Although single equation methods are not necessarily inferior in all situations, system methods are more sensitive to specification error since errors are carried across equations. Moreover, 2SLS may be preferred to 3SLS when a small number of observations is available. Because ERS's model is quarterly as opposed to monthly, this last consideration is especially important.

As mentioned above, the ex-post forecast evaluation is conducted by comparing projections for Q2/2009 to Q1/2010 endogenous variables to actual values for them. Forecasts are evaluated two ways: in terms of their accuracy and their ability to predict turns in the data. Root mean square error (RMSE) and Theil's *U* (another measure of forecasting quality) are two statistics employed to assess accuracy:[13]

$$RMSE = \sqrt{\frac{1}{n^o}\sum_t (y_t - \hat{y}_t)^2}$$

$$U = \sqrt{\frac{(1/n^o)\sum_t (y_t - \hat{y})^2}{(1/n^o)\sum_t y_t^2}}$$

The variables $n^o, y, \hat{y}$ correspond to the number of periods being forecasted, actual values, and forecast values, respectively. The formulation of Theil's *U* is equivalent to *U2* (Theil, 1966, chapter 2).

[13]Accuracy and turn equations, except percentage of correct turn measures, come from Greene (2008, pp. 101-2).

Turns in a given variable occur when the sign of the difference from quarter t to quarter t+1 changes. So if, for variable y, the signs of the difference $\Delta y_{t+1} = (y_{t+1} - y_t)$ and $\Delta y_t = (y_t - y_{t-1})$ are different, we have a turn in the data. A measure of percentage of correct turns is the rate of correct turn predictions to total turn predictions. In addition, a Theil *U* measure that tracks the turns in the data is:

$$U_\Delta = \sqrt{\frac{(1/n)\sum_t (\Delta y_t - \Delta \hat{y}_t)^2}{(1/n)\sum_t (\Delta y_t)^2}}$$

where $\Delta y_t = (y_t - y_{t-1}) / y_{t-1}$ and $\Delta \hat{y}_t = (\hat{y}_t - y_{t-1}) / y_{t-1}$.

Tables 5 and 6 present the validation results for the quarterly dairy model, illustrating the comparison of the forecasts for the various variables derived from the different models and the ICEC based on the accuracy and turn statistics presented previously. The accuracy and turning statistics are computed for four periods. A lower RMSE and a lower Theil's U point to a better—more accurate—forecast. A higher RMSE and a lower Theil's U indicate a better performance in predicting turning points in the data.

The various models are "coded" respectively as:

- A (restricted) and B (unrestricted): vector autoregression with exogenous variables (VARX);
- C: ordinary least squares (OLS);
- D: two-stage least squares (2SLS);
- E: seemingly unrelated regression (SURE);
- F: three-stage least squares (3SLS); and
- I: the ICEC forecasts.

Forecasts by the ICEC, as mentioned above, are a combination of expert judgment and econometric model results. Forecasts for A, B, C, D, E, and F are derived only from ERS's econometric analysis. AV represents the average of the three best forecasts. In this connection, Timmermann (2006, p.181) noted that simple combination schemes like arithmetic averages often do better "…than more sophisticated rules relying on estimating optimal weights."

Tables 5 and 6 show the forecast evaluation results for the various blocks of the dairy quarterly model. The three best forecasts, their average, and the ICEC combined forecast are ordered from best to worst for each variable. The unconstrained VARX tends to forecast prices and the number of cows better (both in terms of accuracy and correct turns predicted) than other models and combining forecasts, in some instances, improve forecasts as shown for whey (see table 5). Other methods (both single- and multiple-equation) outperform VARX when forecasting elements of the commodity balance sheets (see table 6).

Table 5
Block 1 forecast evaluation

Variable	Method[1]	RMSE	Theil U-accuracy	Method[1]	Percent correct turns	Theil U-turns
Number of cows	B	1238	0.000015	B	67	0.30
	I	1552	0.000019	I	67	0.38
	A	3826	0.000046	A	67	0.92
	AV	3918	0.000047	AV	33	0.95
	E	12254	0.000146	E	33	2.98
Production per cow	C	136	0.0000051	C	100	0.0072
	AV	175	0.0000066	AV	100	0.1958
	D	186	0.0000070	D	100	0.0099
	E	220	0.0000082	E	100	0.012
	I	3920	0.0001465	I	67	0.21
Cheese price	I	0.002	0.0012	I	100	0.03
	AV	0.010	0.0054	AV	67	0.12
	B	0.018	0.0099	B	33	0.28
	A	0.034	0.0182	A	33	0.50
	F	0.107	0.0578	F	67	1.47
Butter price	I	0.0002	0.0002	I	33	0.04
	AV	0.0036	0.0022	AV	33	0.41
	A	0.0124	0.0075	A	33	1.34
	B	0.0142	0.0086	B	33	1.53
	E	0.0154	0.0093	E	33	1.84
NDM price	AV	0.005	0.005	AV	100	0.30
	I	0.011	0.011	I	100	0.56
	B	0.019	0.019	B	33	0.94
	A	0.042	0.042	A	33	1.90
	F	0.057	0.057	F	67	2.51
Whey price	B	0.0020	0.020	B	67	0.41
	AV	0.0022	0.022	AV	33	0.64
	C	0.0043	0.042	C	33	1.12
	D	0.0045	0.044	D	33	1.15
	I	0.0072	0.071	I	67	1.27
All-milk price	I	1.03	0.005	I	33	0.31
	B	1.71	0.009	B	67	0.60
	A	2.08	0.011	A	33	0.67
	AV	2.81	0.012	AV	67	0.85
	E	7.32	0.038	E	67	2.46

NDM = Nonfat dry milk. RMSE = Root mean squared error. AV = Average of three best forecasts.
[1]Methods' key corresponds to A, constrained vector autoregression with exogenous variables (VARX); B, unconstrained VAR; C, ordinary least squares (OLS); D, two-stage least squares (2SLS); E, seemingly unrelated regression (SUR); F, three-stage least squares (3SLS); and I, forecasts by ICEC. The specific version of the VARX model employed here contains no exogenous variables; previous versions of this research, as well as ERS's regular forecasting activities, usually employ them.

Sources: USDA, World Agricultural Outlook Board and USDA, Economic Research Service estimates.

Table 6
Blocks 2, 3, and 4 forecast evaluation

Variable	Method[1]	RMSE	Theil U-accuracy	Method[1]	Percent correct turns	Theil U-turns
FATS						
Beginning Stocks	E	0.40	0.0023	E	100	0.16
	F	0.40	0.0023	F	100	0.16
	AV	0.41	0.0024	AV	100	0.17
	D	0.44	0.0025	D	100	0.18
	I	1.51	0.009	I	100	0.66
Imports	AV	0.0150	0.0146	AV	0	0.63
	E	0.0151	0.0152	E	0	0.63
	F	0.0151	0.0052	F	0	0.63
	C	0.0154	0.0155	C	0	0.64
	I	0.0161	0.016	I	67	0.70
Exports	I	0.03	0.02	I	67	1.0
	C	0.0628	0.051	C	0	2.25
	D	0.0628	0.051	D	0	2.25
	AV	0.0638	0.052	AV	0	2.28
	E	0.0659	0.052	E	0	2.35
Net Removals	B	0.039	0.40	B	33	0.466
	A	0.0407	0.42	A	33	0.488
	AV	0.0414	0.43	AV	33	0.493
	F	0.045	0.47	F	33	0.530
	I	0.048	0.50	I	33	0.577
SKIM SOLIDS						
Beginning Stocks	I	0.2	0.002	I	33	0.77
	B	0.6	0.004	A	33	2.4
	A	0.8	0.006	B	33	1.9
	AV	1.3	0.009	AV	33	3.9
	C	3.2	0.02	C	33	10
Imports	I	0.015	0.020	I	67	0.37
	A	0.018	0.024	A	67	0.44
	B	0.020	0.026	B	67	0.48
	AV	0.022	0.030	AV	33	0.52
	C	0.045	0.061	C	33	1.10
Exports	E	0.396	0.0143	E	33	0.91
	F	0.396	0.0143	F	33	0.91
	AV	0.398	0.0144	AV	33	0.92
	C	0.40	0.0145	C	33	0.93
	I	0.49	0.0178	I	33	1.14
Net Removals	C	0.027	0.114	C	67	0.8
	I	0.028	0.117	I	100	0.2
	D	0.03	0.14	D	67	0.9
	AV	0.04	0.19	AV	67	1.3
	E	0.1	0.42	E	67	2.7

NDM = Nonfat dry milk. RMSE = Root mean squared error. AV = Average of three best forecasts.
[1]Methods' key corresponds to A, constrained vector autoregression with exogenous variables (VARX); B, unconstrained VAR; C, ordinary least squares (OLS); D, two-stage least squares (2SLS); E, seemingly unrelated regression (SUR); F, three-stage least squares (3SLS); and I, forecasts by ICEC. The specific version of the VARX model employed here contains no exogenous variables; previous versions of this research, as well as ERS's regular forecasting activities, usually employ them.

Sources: USDA, World Agricultural Outlook Board and USDA, Economic Research Service estimates.

Table 7 presents results for the forecast evaluation both in terms of accuracy and percentage of correct turn prediction in the data. A Theil's U measure is used to measure accuracy. The labels Q2/2009, Q3/2009, Q4/2009, and Q1/2010 represent the quarters for which the measure is calculated. The best predictor by quarter is the one with the highest cumulative Theil's U. Turn predictors were calculated as the difference between the value of a given variable in a quarter versus the value in the next period and then determining if this difference was positive or negative. The value of the sign of the difference in the data was compared with the sign of the difference in the predictions. A correct turn is when the sign of the difference in the data is the same as the sign of the prediction. Correct signs of differences were added from one period to the next. The best predictors by quarter are the ones that have the highest cumulative percentage at that point in time.

When forecasting dairy product prices, the VARX unrestricted and restricted models tend to be the best among the ERS forecasting models, performing even better when forecasting the prices of whey and NDM than the consensus forecast by ICEC. Constrained or unconstrained VARX models have the highest percentage of correct turn predictions for butter, whey, or all-milk prices in Q2/2009-Q1/2010. Structural models are also good turn predictors for these prices. ICEC is better at predicting turns for cheese and NDM in this forecasting activity.

Forecasting performance can be very fragile, and good performance in one period does not imply good performance in the next. When the analysis was done using different years of data (Q4/1998 to Q4/2008, withholding the year 2008 for the ex-post forecasting exercise), however, forecasting reliability measures exhibited a similar performance structure in the time series models vis-à-vis single- and multiple-equation models. Moreover, simple arithmetic averages of forecasts did show some performance improvement but not in every instance. Expert judgment is important in this context to extrapolate from the observed trends and to complement the results of this model. As such, forecasting performance measures are useful tools for continually improving forecasting models.

Table 7

Robustness of the best performing forecasting methods[1]

Theil U and Share of Correct Turns for Q2/2009 to Q1/2010

	Most accurate quarterly cumulative predictor (Theil U)			
Variable/quarter	Q2/2009	Q3/2009	Q4/2009	Q1/2010
Number of cows	B	E	I	B
Production per cow	E	E	C	C
Cheese price	A	I	I	I
Butter price	I	E	I	I
NDM price	I	E	AV	AV
Whey price	F	B	B	B
All-milk price	I	I	I	I
Beginning stocks F	D	E	D	F
Imports F	F	E	F	F
Exports F	AV	E	I	I
Net removals F	I	I	B	B
Beginning stocks S	I	E	I	I
Imports S	B	E	AV	I
Exports S	D	E	F	E
Net removals S	AV	E	D	I

	Best cumulative turn predictor (correct turn/total predictions)		
Variable/quarter	Q2-3/2009	Q3-4/2009	Q4/2009-Q1/2010
Number of cows	A/B/D/AV/I	A/D/I	A/B/I
Production per cow	A/B/C/D/E/F/AV/I	A/B/C/D/E/F/AV/I	C/D/E/F/AV
Cheese price	A/B/C/D/E/F/AV/I	C/D/E/F/AV/I	AV/I
Butter price	C/D/E/F	C/D/E/F	C/D
NDM price	A/C/D/E/F/AV/I	C/D/E/F/AV/I	AV/I
Whey price	B	B	B/I
All-milk price	I	A/B/C/D/E/F/AV/I	B/C/D/E/F/AV
Beginning stocks F	C/D/E/F/AV/I	C/D/E/F/AV/I	C/D/E/F/AV/I
Imports F	A/B/C/D/E/F/AV/I	A/B/D/I	A/B/D/I
Exports F	A/B/AV	A/B/AV/I	A/B/I
Net removals F	A/B/C/D/E/F/AV/I	A/B/C/D/E/F/AV/I	A/B/C/D/E/F/AV/I
Beginning stocks S	A/B/C/D/E/F/AV/I	A/B/C/D/E/F/AV/I	A/B/C/D/E/F/AV/I
Imports S	C	A/B/C/D/E/F/AV/I	A/B/I
Exports S	A/B/C/D/E/F/AV/I	B/C/D/E/F/AV/I	B
Net removals S	B/C/D/E/F/AV/I	C/AV/I	I

NDM = Nonfat dry milk. RMSE = Root mean squared error. AV = Average of three best forecasts.
[1]Methods' key corresponds to A, constrained vector autoregression with exogenous variables (VARX); B, unconstrained VAR; C, ordinary least squares (OLS); D, two-stage least squares (2SLS); E, seemingly unrelated regression (SUR); F, three-stage least squares (3SLS); and I, forecasts by ICEC. The specific version of the VARX model employed here contains no exogenous variables; previous versions of this research, as well as ERS's regular forecasting activities, usually employ them.

Sources: USDA, World Agricultural Outlook Board and USDA, Economic Research Service estimates.

Summary and Conclusion

A model for the U.S. dairy industry using quarterly data for Q4/1998-Q1/2009 was specified and estimated as a tool to generate forecasts for USDA. The model was validated using data for Q2/2009-Q1/2010 and employs a variety of single-equation and system methods to estimate values for several variables. Clearly, different methods better forecast different variables. Vector autoregression with exogenous variables (VARX) outperforms structural regression models when forecasting prices, but single-equation and system estimations of structural models outperform time series models when forecasting some crucial items in farm supply and commodity balance sheets. This study has shown that it is better not to rely on any one particular estimating method. Careful econometric specification and data development should ensure that greater weight is given to a model-based dairy forecasting system relative to a judgment-based forecasting system.

References

Alston, J.M., J.V. Balagtas, H. Brunke, and D.A. Sumner. "Supply and demand for commodity components: Implications of Free Trade Versus the AUSFTA for the U.S. Dairy Industry," *The Australian Journal of Agricultural and Resource Economics* 50: pp. 131–52, 2006.

Bailey, K., J. Dunn, and M. Pajic. *Short Run Price Determination for the U.S. Dairy Industry*, Staff Paper #378, Penn State University, Department of Agricultural Economics and Rural Sociology, 2006.

Bailey, K. *Documentation of a Dynamic and Simultaneous Econometric Model of the U.S. Dairy Industry*, Staff Paper #381, Penn State University, Department of Economics and Rural Sociology, 2009.

Blayney, D. Personal interview, Washington, DC, November 9, 2010.

Cameron, A.C., and P.K. Trivedi. *Microeconometric Analysis With Stata*, College Station, TX: Stata Press, 2009.

Coelli, T.J., D.S.P. Rao, C.J. O'Donnell, and G.E. Battese. *An Introduction to Efficiency and Productivity Analysis*, Second Edition, New York: Springer, 2005.

Engle, R. "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics," *Journal of Economic Perspectives* 15(4): pp. 157-68, 2001.

Greene, W.H. E*conometric Analysis*, Sixth Edition, Upper Saddle River, NJ: Prentice Hall, 2008.

Jesse, E., and J. Schuelke. *Forecasting Class III and Class IV Milk Prices*, Staff Paper #453, University of Wisconsin-Madison, Department of Agricultural & Applied Economics, 2002.

Kennedy, P. *A Guide to Econometrics,* Sixth Edition, Wiley-Blackwell, 2008.

Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag, 2006.

Lütkepohl, H., and F. Xu, *The Role of the Log Transformation in Forecasting Economic Variables*, Working Paper Series No. 2591, Munich, Germany: CESifo, 2009.

Miller, J. "Problems with Aggregating Dairy Products," *Agricultural Outlook Magazine*, pp. 11-3, U.S. Department of Agriculture, Economic Research Service, September 1989.

Premakumar, V., and S. Chaudhary. *Support Prices as Policy Tools in Dairy Industry: Issues in Theoretical Modeling*, Working Paper 96-WP 151, Iowa State University, Center for Agricultural and Rural Development, February 1996.

Theil, H. *Applied Economic Forecasting*, Chicago: Rand McNally, 1966.

Timmermann, A. "Forecast Combination," *Handbook of Economic Forecasting*, G. Elliot, C.W.J. Granger, and A. Timmermann (eds.), Vol. 1, pp. 135-96, Amsterdam: Elsevier, 2006.

U.S. Department of Agriculture, Agricultural Marketing Service (AMS) Dairy Programs. *National Econometric Model Documentation: For Model Calibrated to USDA Agricultural Baseline Projections to 2015*, 2006.

Varian, H.R. *Microeconomic Analysis*, Third Edition, New York: W.W. Norton & Company, 1992.

Westcott, P. C. *A Quarterly Model of the U.S. Dairy Sector and Some of Its Policy Implications*, TB-1717, U.S. Department of Agriculture, Economic Research Service, 1986.

Appendix A: Constructing the Energy Index

The energy variable used in the ERS dairy forecasting model is constructed using a Törnqvist index (see Coelli et al., 2005, p. 90 for details) that can in general aggregate different types of commodities consistently across time, space or both. In particular, prices and quantities for industrial consumption for four types of energy—natural gas, propane, gasoline, and electricity—from the Energy Information Administration's Short-Term Energy Outlook databases are employed. The index compares the geometric mean of the four energy prices in time *s* with time *t*, and it is weighted by the *m* cost shares of the various energy components:

(A1) $$\ln P_{st}^{T} = \sum_{m=1}^{4}\left(\frac{\omega_{ms}+\omega_{mt}}{2}\right)(\ln p_{mt} - \ln p_{ms}),$$

where t= 1, 2, 46, and s= 1 is the base period which corresponds to Q4/1998,

m = 1, 2, 3, and 4 are the energy cost components, and

$\omega_{ms} = \dfrac{p_{ms}q_{ms}}{\sum_{i=1}^{M} p_{ms}q_{ms}}$ is the value of the *m*-th component for the base period.

Appendix B: The VARX Model

In general, and following Lütkepohl (2006, p. 387), a VARX model with p and s lags for the endogenous and exogenous variables, respectively, is written as:

(B1) $$y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{i=0}^{s} \Theta_i^* x_{t-i} + \varepsilon_t,$$

where $\delta = (\delta_1, \cdots, \delta_k)'$ are the constants for the *k* equations,

$y_t = (y_{1t}, \cdots, y_{Kt})'$ is the *K-dimensional vector of* endogenous variables,

Φ_i is a $K \times K$ matrix for the *p* order autoregressive parameters,

$x_t = (x_{1t}, \cdots, x_{Mt})'$ is the M-dimensional vector of exogenous variables,

Θ_i is a $K \times M$ matrix for the s order exogenous parameters,

$\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{kt})'$ is a vector of white noise

so that $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = \Sigma$ and $E(\varepsilon_t \varepsilon_s') = 0$ for $t \neq s$.

Appendix C: Estimate of Correlations Between Equations' Disturbances

The estimated $\hat{\Sigma}$ matrices (symmetric) for blocks 1, 2, and 3 are:

Block 1

Equation	1.1	1.2	1.3	1.4	1.5	1.6	1.7
1.1	3.15e-06						
1.2	6.57e-06	.00011					
1.3	-2.82e-06	-.00020	.011				
1.4	5.89e-06	-.00027	-.00039	.0097			
1.5	-4.53e-07	-.00010	.0058	.00098	.014		
1.6	-.000028	-.00014	.0084	.0049	.0096	.022	
1.7	-5.39e-07	-.000082	-.00042	.000036	.00027	-.00038	.00057

Source: USDA, Economic Research Service estimates.

Block 2

Equation	2.1	2.2	2.3
2.1	.012		
2.2	.0020	.025	
2.3	.0072	.0034	.08

Source: USDA, Economic Research Service estimates.

Block 3

Equation	3.1	3.2	3.3
3.1	.0048		
3.2	-.0033	.017	
3.3	.00015	.0012	.020

Source: USDA, Economic Research Service estimates.

Appendix D: Hurdle or Two-Part Model Versus Tobit Tests

The hurdle or two-part model is presented as an alternative to the Tobit model specification of net removals as it relaxes some of the stronger assumptions that the latter makes. Specifically, the hurdle model splits the selection from the outcome mechanisms, and it can be employed for different variables at these two different stages. Here, a Probit model is employed to model the censoring mechanism and then a regression equation is used to characterize the outcome. Both parts are assumed to be independent (see Cameron and Trivedi (2009, pp. 538-541) and Kennedy (2008, p. 269) for details).

Formally:

$$f(y/x) = \begin{Bmatrix} \Pr(d=0|x) & if\ \ y=0 \\ \Pr(d=1|x\, f(y|d=1, x & if\ \ y>0 \end{Bmatrix},$$

where d is an indicator variable, (If $d = 1$, $y > 0$ and if $d = 0$, y = 0) (Cameron and Trivedi, 2009).

The sample period (Q4/1998-Q1/2009) has 11 cases for fat and 7 for skim solids where net removals are zero. From the assumption of independence, estimates of the two-part model's joint likelihoods for fat and skim solid net removals are 15.28 and -38.29, respectively. In contrast, the Tobit model's likelihoods for fat and skim solids are 39.22 and -38.23, respectively. A larger number for the log likelihoods indicate that the model fits the data better. Hence, the Tobit model is better in this particular sample.