Correction of Measurement Error in Monthly USDA Pig Crop: Generating Alternative Data Series

In-Seck Kim, Ronald L. Plain, J. Bruce Bullock, and Sang Young Jei

The imputed pig death loss contained in the reported monthly U.S. Department of Agriculture (USDA) pig crop data over the December 1995–June 2006 period ranged from −4.93% to 12.75%. Clearly, there are substantial measurement errors in the USDA monthly pig crop data. In this paper, we present alternative monthly U.S. pig crop data using the biological production process, which is compatible with prior knowledge of the U.S. hog industry. Alternative pig crop data are applied to a slaughter hog model and tested comparatively to USDA pig crop. Test results reject the validity of USDA pig crop data in favor of the alternative data.

Key Words: biological production process, measurement error, monthly USDA pig crop data, pig death loss

JEL Classifications: Q11, Q13, C12

Most livestock supply analyses have concentrated on the incorporation of farmers' expectation of future prices into the supply function to explain cyclical patterns of livestock prices and quantities. The analyses also examine supply elasticities of slaughter animals, or meat products, in response to price changes. Aadland and Bailey well documented the previous research of these approaches in the U.S. beef cattle industry. In cases of the pork hog market, Chavas investigated the nature of the expectation formation within the U.S. pork market with the use of annual data from 1960 to 1996.

However, unlike annual or quarterly models, monthly changes in slaughter hogs (barrow and gilt) are rarely affected by economic variables, once the size of the breeding herd is determined by the hog producers within the monthly framework. Fluctuations of monthly slaughter hogs are determined by seasonal variations of production variables in the biological production process, rather than by producer expected price changes.

Hence, the challenge of modeling supply response for monthly slaughter hog forecasting is not to specify the nature of monthly changes in producer expectations about slaughter hog prices or pork prices in month t; rather, the challenge is to model the biological production process so as to predict the number of animals slaughtered in month t

In-Seck Kim is a former postdoctoral fellow in Food and Agricultural Policy Research Institute at the University of Missouri and is now with the Department of Agricultural and Food Economics, Agri-Food & Biosciences Institute, Belfast, Northern Ireland. Ronald L. Plain, J. Bruce Bullock, and Sang Young Jei are professor, former professor, and graduate student, respectively, in the Department of Agricultural Economics, University of Missouri.

given the number of pigs farrowed in month (t − 6), which was determined by the number of females bred in (t − 10), which, in turn, was determined by the size of the breeding herd in (t − 11).[1]

Thus, monthly estimates of the size of the U.S. pig crop are necessary to develop an estimate of monthly slaughter hogs in future months. Before 1995, only quarterly estimates of the size of the pig crop were available. In December 1995, the U.S. Department of Agriculture (USDA) began reporting estimates of sows farrowing and pig crop on a monthly basis.

However, these monthly estimates of farrowing and the size of the pig crop lead to some logical and biological inconsistencies in historical data. For example, comparison of the calculated monthly slaughter hog produced in month t,[2] as well as the reported pig crop 6 months before, leads to imputed death loss during the growing/finishing phase of pork production over the December 1995– June 2006 period. During that period, it ranged from −4.93% of reported pigs farrowed to 12.75%.

This observation suggests a need for an alternative pig crop projection system that reflects (is constrained/generated by the use of) widely observed and accepted norms regarding trends and seasonal patterns in conception rates, pigs per litter, and pig and sow death losses during the hog reproduction and growing process.

The objective of this study is to present an alternative monthly U.S. pig crop data series that is compatible with prior knowledge of the U.S. hog industry. A second objective is to test alternative pig crop data comparatively to the USDA pig crop. In the balance of this paper, we discuss potential problems of monthly USDA pig crop data, describe the alternative data generation process, compare alternative pig crop data with reported USDA numbers, develop monthly slaughter hog model, and test the appropriateness of two different data series on the basis of the developed monthly slaughter hog model.

Problems of Monthly USDA Pig Crop Data

Monthly USDA pig crop data are obtained from quarterly surveys of pork producers. Producers are surveyed quarterly regarding the number of sows farrowed and the size of the pig crop during each of the previous 3 months. Data for each of the months are published in the USDA Hogs and Pigs Report. In some quarters, previously reported monthly data regarding farrowings and the pig crop are adjusted on the basis of recently observed slaughter numbers.

The USDA Hogs and Pigs Report production data system treats pig death loss during the growing process as the residual required to balance hog slaughter numbers in month t and the reported pig crop in month (t − 6). However, this simple comparison between observed slaughter in month t and the reported pig crop in month (t − 6) might misrepresent actual pig death loss contained in the USDA pig crop data because the pig crop is used not only for slaughter but also for breeding, and some of the observed slaughter hogs in month t are not raised from the pig crop in month (t − 6).

Therefore, this study obtained monthly USDA pig death loss by comparing the reported pig crop in month (t − 6) with the calculated monthly slaughter hog produced in month t, which takes into account other factors as well as slaughter number. Figure 1 shows the imputed pig death loss contained in the monthly USDA pig crop data for the December 1995–June 2006 period.

The imputed death loss data range from a low of −4.93% in May 2003 to a high of 12.75% in February 1996. The highly volatile and sometimes negative death losses imputed in the USDA pig crop over the historical period implies measurement error in the

---

[1] Even though there is definitely some flexibility in the timing of slaughter and gestation, we believe that this characterization of biological production process for slaughter hog is the best way to portray short-run hog production dynamics in the monthly framework.

[2] There is no reported data for slaughter hogs produced. The data generation process for slaughter hogs produced will be discussed in a later part of the study.
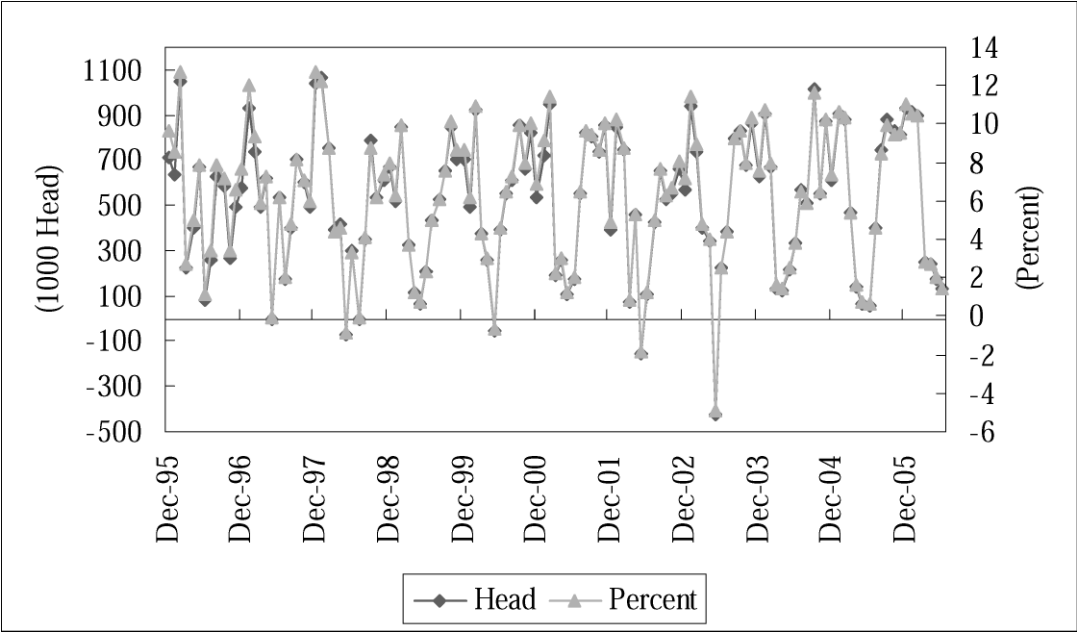
Figure 1. Imputed Pig Death Loss Contained in the Monthly USDA Pig Crop Data

USDA pig crop data. As mentioned before, in the monthly framework, it is extremely important to model biological production process appropriately from the consistent production data so as to correctly represent dynamics of the short-run hog supply.

Mismeasured pig crop data possibly provide erroneous information about the nature of short-run slaughter hog dynamics and lead to inconsistent conclusions in the economic analysis of the U.S. hog industry. More importantly, it would adversely affect the decision making of market participants and policy makers.

Consistent production data series concerning all dimensions of the swine sector is a necessary condition for economic analysis of the U.S. swine industry. Regression analysis that uses mismeasured variables could produce statistically inconsistent estimates. Furthermore, test results based on the ordinary least squares (OLS) residuals might be misleading if the OLS estimator is inconsistent because of the measurement error problem. Large amounts of literature have been published on this topic. In particular, Griliches and Hausman well illus-

trate the problems of mismeasured variables in econometric analysis in a variety of situations.

In addition to the econometric problems, mismeasured pig crop data would affect economic activities of market participants and policy decisions of policy makers in the U.S. swine industry. For example, mismeasured pig crop data would misrepresent the biological relationship between production variables. They would provide erroneous information of previous months' breeding herd size and hence producers' decision-making process with respect to price changes. Besides, it would also mislead slaughter hog number and pork production in future months.

The effects of mismeasured pig crop data are not limited on the production side of the hog pork market. Erroneous slaughter hog and pork production forecasting results would send incorrect signals to market participants in projecting price of hog and pork. These potentially inconsistent forecasting results of production and prices of hog pork would adversely affect investment decisions of hog producers and pork packers and, hence, policy

decisions of policy makers who monitor market activities of the U.S. swine industry.

Data Generation Process

Unlike the USDA pig crop data projection process, this study imposes pig death loss calculations as one step in the process of constructing pig crop, sows farrowing, and females bred that are consistent with observed slaughter numbers and with our current knowledge of the biological production function for hogs.

The daily number of hogs slaughtered is the item within hog production data that is measured with a high level of accuracy. However, several aspects of the biological process in the production slaughter of hogs are known with a high degree of certainty that make it possible to impute the number of sows that must have been bred in month $(t - 10)$ and farrowed in month $(t - 6)$ to have produced the number of hogs slaughtered in month t. For example:

- For every domestic slaughter hog slaughtered in month t there were $(1 + k_1)$ pigs farrowed in month $(t - 6)$, where $0 < k_1 < 1$ is the proportion of pigs that die during the growing process. k_1 is made of two components in the data generation process: One is the annual pig death loss percentage as reported in Table 1, and the other is the pig death loss seasonal index as a percentage of the annual average, which is reported in Table 2. The seasonal index of pig death loss is incorporated into the annual pig death loss percentage to produce the monthly pig death loss percentage.
- For every sow that farrowed in month $(t - 6)$ there were $[(1/k_2)(1 + k_3)]$ sows and gilts bred in month $(t - 10)$, where $0 < k_2 < 1$ is the proportion of sows and gilts bred that become pregnant and $0 < k_3 < 1$ is the proportion of bred sows that die during the gestation period. k_2 is made up of two components in the data generation process: One is the sow conception factor and the other is the gilt conception rate. Both percentages are reported in Table 2. k_3 is made up of two components in the data generation process: One is annual sow death loss percentage, which is reported in Table 1, and the other is

Table 1. Annual Pig and Sow Death Loss (%)

Year	Pig Death Loss[a]	Feeder Pig Death Loss[b]	Sow Death Loss
1990	6.3	3.1	2.5
1991	6.2	3.1	2.5
1992	6.1	3.1	2.65
1993	6.0	3.1	2.9
1994	5.9	3.1	3.15
1995	5.8	3.0	3.4
1996	5.7	3.0	3.65
1997	5.6	2.9	3.9
1998	5.5	2.9	4.15
1999	5.4	2.9	4.4
2000	5.3	2.8	4.65
2001	5.2	2.8	4.9
2002	5.1	2.7	5.15
2003	5.0	2.7	5.4
2004	4.9	2.7	5.6
2005	4.8	2.6	5.7
2006	4.8	2.6	5.8

Note: These numbers are slightly higher than Iowa State University reports because Iowa reports are based on small farms. Numbers over 2000–2006 are obtained from Ronald L. Plain, Professor, University of Missouri–Columbia.
Source: Iowa Livestock Enterprise Summaries; 1990–1999.
[a] Birth to market death loss.
[b] Feeder to market death loss.

the sow death loss seasonal index as a percentage of annual average, which is reported in Table 2. The seasonal index of sow death loss is incorporated into the annual sow death loss percentage to produce the monthly sow death loss percentage.
- The number of pigs per litter has trended steadily upward and follows a seasonal pattern.

In the case of the annual death loss percentage, we obtained specific percentages from the Iowa State University (ISU) report regarding pig and sow death loss. Our numbers are slightly higher than ISU reports, which are based on the small farm.

In addition to the published reports regarding annual trend of animal death loss, we developed the set of seasonal indices on the basis of our experience in economic analyses of the U.S. swine industry. Even though management and technology has developed over the years, swine production from breeding to finishing is still significantly subject to weather; hence, distinct seasonal patterns for

Table 2. Seasonal Patterns for Hog Death Loss and Conception Rate

Month	Pig Death Loss[a]	Gilt Conception Rate (%)[b]	Sow Conception Factor (%)[c]	Sow Death Loss[d]
Jan	0.99	90	9	0.95
Feb	1.00	86	10	0.92
Mar	0.99	84	12	0.88
Apr	0.99	82	15	0.92
May	1.01	79	16	0.98
Jun	1.03	75	16	1.05
Jul	1.05	77	12	1.14
Aug	1.03	82	9	1.13
Sep	0.99	85	8	1.08
Oct	0.98	87	6	1.02
Nov	0.97	88	6	0.98
Dec	0.97	88	7	0.95

[a] Seasonal index as percentage of annual average.
[b] Percentage of gilts bred from 1-month lagged gilts added in the breeding herd. Thus in June, $(1/0.75)100 = 133$ gilts must be added to the breeding herd to obtain 100 bred gilts in July. In January, only $(1/0.9)100 = 111$ gilts must be added to obtain 100 bred gilts in February. Therefore, there must be corresponding seasonal variation in gilt additions.
[c] Proportion of sows farrowed that are not settled on first rebreeding attempt but settle on second rebreeding attempt after nursing. Therefore, sow seasonal conception factor 9 in January means that 91% of sows farrowed in January rebred in March, and the remaining 9% of sows farrowed in January rebred in April.
[d] Seasonal index as percentage of annual average.

hog death loss and conception rate are well-observed phenomena in the swine industry.

Consequently, it is possible to use values of k_1, k_2, and k_3 to develop estimates of unobserved numbers of the following that must have occurred to produce the observed number of hogs slaughtered in month t:

(1) sows farrowed, pigs/litter, and the size of the pig crop in $(t - 6)$,
(2) number of sows and gilts bred and settled in month $(t - 10)$, and
(3) gilts added to the breeding herd in month $(t - 11)$.

The data generation process works in the reverse of observed commercial slaughter hogs. The steps and procedures of the data generation process are described below. Mnemonic descriptions of all variables are summarized in Table 3. Subscripts of mnemonics refer to month; for example, SF_{t-6} is sows farrowed in month $(t - 6)$.

Sows Farrowed

SF_{t-6} is defined as PC_{t-6} divided by PPL_{t-6}. However, the alternative PC is not a given number but it is determined in the alternative data generation process. Thus, PC_{t-6} is unknown at this point of data generation process. This study substituted PC_{t-6} with $CDBGS_t{}^3$ adjusted by the WDA_t in the calculation for SF_{t-6} in that most PC are fed to be slaughtered and it takes 6 months for PC to reach slaughter weight. WDA is equal to the average number of slaughter work days during the month over the sample period (i.e., 21.25 days) divided by the WD.

Slaughter plant work days adjustment factor. Even though most of PC are either slaughtered or retained to the breeding herd at 6 months, there is definitely some flexibility in the timing of slaughter and breeding herd retention. Therefore, $CDBGS_t$ is determined not only by PC_{t-6} but also by WD_t. However, unlike the slaughter number, PC_{t-6} and SF_{t-6} are not affected by WD_t. Therefore, PC_{t-6} and SF_{t-6} will be biased if we deduce it from actual $CDBGS_t$ numbers without consideration of WD_t.

---

³ $CDBGS_t = CBGS_t - NISWBG_t$, where $CBGS_t = (CHS_t/FIHS_t) \times FIBGS_t$, $NISWBG_t = (FPI_{t-4} - DLIFP_{t-4}) \times 0.5 + (FPI_{t-5} - DLIFP_{t-5}) \times 0.5 + HIM_t - FPI_t - SBI_t - HEX_t$, and $DLIFP_t = FPI_t \times (APFPDL_t/100)$. $APFPDL_t$ is reported in Table 1.

Table 3. Variable Definitions

Observed		
APFPDL	Annual percentage of feeder pig death loss	(ISU)
APPDL	Annual percentage of pig death loss	(ISU)
APSDL	Annual percentage of sow death loss	(ISU)
CHS	Commercial hogs slaughtered	(LMIC)
FIBGS	Federally inspected barrows and gilts slaughtered	(LMIC)
FIHS	Federally inspected hogs slaughtered	(LMIC)
FPI	Feeder pig imports	(ERS)
HEX	Hog exports	(ERS)
HIM	Hog imports	(ERS)
SBI	Sow and boar imports	(AMN)
PC	Pig crop	(NASS)
PPL	Pigs per litter	(NASS)
SF	Sows farrowed	(NASS)
WD	Slaughter plant work days in the month	(LMIC)
Generated		
APBRBG	Annual percentage of boars slaughtered relative to the barrows and gilts slaughtered	
BA	Boar additions	
BHA	Breeding herd additions	
CBGS	Commercial barrows and gilts slaughtered	
CDBGS	Commercial domestic barrows and gilts slaughtered	
CDBGP	Commercial domestic barrows and gilts produced	
CDSS	Commercial domestic sows slaughtered	
DLIFP	Death loss of imported feeder pigs	
FB	Females bred	
GA	Gilt additions	
GB	Gilts bred	
MPPDL	Monthly percentage of pig death loss	
NBGS	Nonbreeding gilts slaughtered	
NISWBG	Net imports of slaughter weight barrows and gilts	
PDL	Pig death loss	
SDL	Sow death loss	
SRB	Sows rebred	
WDA	Slaughter plant work days adjustment factor	
Index		
CRG	Conception rate of gilt	
PDLSI	Pig death loss seasonal index	
SCF	Sow conception factor	
SDLSI	Sow death loss seasonal index	

Note: Data sources for the observed variables are in parentheses. PC, PPL, and SF are observed variables in the USDA data system, but they are generated variables in the alternative data system.

[a] AMN is Agricultural Marketing News, ERS is Economic Research Service, ISU is Iowa State University, LMIC is Livestock Marketing Information Center, NASS is National Agricultural Statistics Service.

For example, CDBGS was 8,037.59 thousand head in October 1996. It was 7,140.37 thousand head in November 1996. CDBGS in October 1996 was 12.6% higher than the number slaughtered in November 1996. However, this does not mean that the PC in April 1996 was 12.6% higher than the May 1996 PC.

October 1996 had 23 slaughter work days and November 1996 only 20 slaughter work days. During October 1996, the hog slaughter industry operated at a slaughter rate of

(8,037.59/23) = 349.5 thousand head per day. The slaughter industry operated at a rate of (7,140.37/20) = 357.0 thousand head per day during November 1996. The 12.6% reduction in total number of slaughter hogs from October to November 1996 was due completely to differences in the number of days that slaughter plants were operated during the 2 months.

The study adjusts the number of $CDBGS_t$ by multiplying WDA_t to remove the effect of WD_t on $CDBGS_t$ in generating PC_{t-6} so that we can obtain unbiased actual PC_{t-6} and SF_{t-6} numbers. After adjusting for the differences in the number of slaughter work days for October and November 1996, we conclude that the April 1996 PC was 8,315.45 thousand head. We also found that the May 1996 PC was 8,496.92 thousand head. The size of the PC and, hence, the number of SF in April and May 1996 (FB in December 1995 and January 1996) were quite similar, although 12.6% more hogs were slaughtered in October 1996 than during November of the same year.

Females Bred

SF_{t-6}, which was deduced from $CDBGS_t$, makes it possible to generate estimates of the FB_{t-10}. FB_t consists of SRB_t and GB_t.

Sows rebred. The number of SRB_t is determined by the biological lags related to the sow farrowing cycle. Most SF_{t-6} are rebred and settled in month $(t - 4)$ then refarrowed in month t because the gestation period is 114 days. Therefore, SRB_t would be specified as a function of SF_{t-2}, SDL_{t-2},[4] and $CDSS_t$[5] if sows get just one breeding attempt after nursing or most sows are settled in the first breeding attempt. However, in reality, most sows get two breeding attempts after nursing and noticeable portion of sows are settled in the second breeding attempt. Therefore, SRB_t must be specified both by SF_{t-2} being settled in the first breeding attempt and SF_{t-3} being settled in second breeding attempt.

This study developed a seasonal index (i.e., SCF) to incorporate proportions of the sows settled in the first breeding attempt and the second breeding attempt in a given month in the data generation process. SCF is defined as the proportion of sows farrowed that are not settled on the first breeding attempt but settled on the second breeding attempt after nursing. Table 2 shows seasonal patterns of SCF.

By definition of SCF, $SF \times SCF/100$ can be interpreted as sows that failed to be settled in the first breeding attempt but settled in the second attempt. Therefore, $SF_{t-2} - (SF_{t-2} \times SCF_{t-2}/100)$ is sows settled in the first breeding attempt in month $(t - 2)$, and $SF_{t-3} \times SCF_{t-3}/100$ is sows failed to be settled in the first breeding attempt in month $(t - 3)$ but being settled in the second breeding attempt in month $(t - 2)$. Therefore, the sows rebred that become pregnant in month t without consideration of death loss and slaughter of sow is: $[SF_{t-2} - (SF_{t-2} \times SCF_{t-2}/100)] + (SF_{t-3} \times SCF_{t-3}/100)$.

On the basis of the above discussion, the identity for SRB_t, which takes into account sow death loss and sow slaughter in the data generation process, should be specified as follows: $SRB_t = [SF_{t-2} - (SF_{t-2} \times SCF_{t-2}/100)] + (SF_{t-3} \times SCF_{t-3}/100) - SDL_{t-2} - CDSS_t \times WDA_t$.

Gilts bred. The number of SF and SRB generated in the previous sections make it possible to infer the number of GB each month. The number of SF_t is equal to the number of FB_{t-4} which is, in turn, equal to SRB_{t-4} plus GB_{t-4}; therefore, $SF_t = SRB_{t-4} + GB_{t-4}$ and the number of gilts bred in month t is $GB_t = SF_{t+4} - SRB_t$.

Breeding Herd Additions

Gilt additions. The number of GA_{t-11} can be inferred from GB_{t-10} and CRG_{t-11} because most gilts are bred and settled after 1 month of retention. Thus, GA_t is: $GA_t = GB_{t+1}/(CRG_t/100)$. CRG are shown in Table 2.

Boar additions. The number of BA accounts for a small portion of total BHA, in

---

[4] $SDL_t = SF_t \times (APSDL_t/100) \times SDLSI_t$, where $APSDL_t$ and $SDLSI_t$ are reported in Tables 1 and 2, respectively.
[5] CDSS is obtained in the same way that CDBGS is generated.

that a small number of boars can breed a large number of females because of the wide use of artificial insemination. The number of BA in the breeding herd is deduced from the boars slaughtered relative to barrows and gilt slaughtered.[6]

Nonbreeding Gilts Slaughtered

Most GA_t are bred in month $(t + 1)$. But gilts that fail to conceive are assumed to be culled from the breeding herd and slaughtered in month $(t + 3)$. Thus, the number of $NBGS_t$ is defined as $NBGS_t = GA_{t-3} \times [(1 - CRG_{t-3})/100]$.

Pig Crop

Alternative monthly PC data are developed by calculating the size of PC_{t-6} that is consistent with the sum of $CDBGP_t$ and PDL_t from PC_{t-6}: $PC_{t-6} = CDBGP_t + PDL_t$.

Commercial domestic barrows and gilts produced. $CDBGP_t$ from PC_{t-6} is calculated by the $CDBGS_t$ adjusted by the WDA_t minus the imputed $NBGS_t$ contained in the $CDBGS_t$ plus imputed BHA_t in the PC_{t-6}. $NBGS_t$ and BHA_t are not adjusted by WDA_t in the calculation for $CDBGP_t$ because both variables are already adjusted by WDA_t in the previous data generation process: $CDBGP_t = (CDBGS_t \times WDA_t - NBGS_t + BHA_t)$.

$NBGS_t$ is subtracted from $CDBGS_t$ adjusted by WDA_t because the $NBGS_t$ are not produced from PC_{t-6} but PC_{t-9}. As mentioned earlier, most GA_t are settled in month $(t + 1)$, but gilts that fail to conceive are culled from the breeding herd and slaughtered in month $(t + 3)$. Thus, $NBGS_t$ are assumed to have been retained in $(t - 3)$ and in turn were born in $(t - 9)$.

Pig death loss. According to the formula for the PC_{t-6} developed earlier, PDL_t is equal to PC_{t-6} less $CDBGP_t$. Even though this definition of PDL in this study is identical in both the USDA and alternative data systems (i.e., both PDL_ts are residuals between PC_{t-6} and $CDBGP_t$), the procedure to arrive at this formula is different in both systems because of the nature of the PC data in two systems is different.

The USDA PC is a given reported number, whereas the alternative PC is determined in the data generation process. Thus, PDL_t in the alternative data series is determined simultaneously with the alternative PC_{t-6} in the data generation process on the basis of the known/observed annual trend and seasonal patterns of pig death loss,

$$(1) \quad PDL_t = PC_{t-6} \times (MPPDL_t/100)$$

$$(2) \quad PC_{t-6} = CDBGP_t + PDL_t,$$

where $MPPDL_t$ is the product of $APPDL_t$ and $PDLSI_t$, which are reported in Tables 1 and 2, respectively. As one can see in the above Equations (1) and (2), both PC_{t-6} and PDL_t—unlike in the USDA data system, in which PC is a given exogenous variable—are endogenous variables in the alternative data system. If we substitute PC_{t-6} in Equation (2) into Equation (1), then PDL_t in the alternative data system is obtained as follows:

$$(3) \quad PDL_t = [(CDBGP_t) \times (MPPDL_t/100)] \div [1 - (MPPDL_t/100)].$$

It is important to note that the above formulas of Equations (1), (2), and (3) for PDL_t is the same in both the USDA and alternative data system, but a different procedure to arrive at this formula produces different numbers and percentages of PDL_t in the USDA and alternative data system. In the alternative data system, $MPPDL_t$ is determined first, and then PDL_t is determined later on the basis of this $MPPDL_t$ number simultaneously with PC_{t-6}. On the other hand, PDL_t is determined first on the basis of the reported PC_{t-6}, and then $MPPDL_t$ is determined later in the USDA system.

On the basis of the above formula, this study obtained PDL contained in the alterna-

---

[6] $BA_t = (APBRBG_t)/100 \times CDBGS_t \times WDA_t$, where APBRBG decreases from 1% in 1990 to 0.35% in 2006.

tive monthly PC over the December 1995–June 2006 period. During that period, it ranged from 2.08% of the alternative PC farrowed to 6.92%.

Limitations of the Data Generation Process and Their Remedies

Four-Year Moving Average of Gilt Additions

The data generation process described above generates estimates of SF_{t-6} consistent with observed $CDBGS_t$ adjusted by the WDA_t. These estimates of SF are used to generate other unobserved historical data series of FB, BHA, and NBGS. On the basis of these generated data series, which are consistent with observed slaughter numbers, we were able to generate estimates of PC_{t-6}. They are consistent with the sum of $CDBGP_t$ and PDL_t.

However, a limitation of this procedure is that alternative SF numbers (and, therefore, other generated data series from SF) could not reflect imputed BHA, PDL, and NBGS that must have been taken into account in generating PC_{t-6} in the calculation for SF_{t-6}. Therefore, the alternative SF and other generated data series from SF must have been underestimated. They were overestimated if NBGS is greater than the sum of imputed BHA and PDL.

That is, SF_{t-6} in the previous section was not defined as PC_{t-6} divided by PPL_{t-6}. Rather, it is defined as $CDBGS_t$ adjusted by WDA_t divided by PPL_{t-6}. The study was not able to use the alternative PC numbers in estimating SF_{t-6} because the process of generating PC data requires having the SF number, which causes circular referencing problems. Variables of BHA_t, PDL_t, and $NBGS_t$ in the calculation for PC_{t-6} were determined by GA numbers, which were in turn deduced from SF in the previous data generation process.

To solve underestimation problems without circular reference, SF were estimated with the use of PC numbers obtained from a 4-year moving average of GA. The components of PC_{t-6} (i.e., BHA_t, PDL_t, and $NBGS_t$) are obtained from a 4-year moving average of GA

rather than GA numbers generated directly from SF numbers in the calculation for SF_{t-6}.

Use of the 4-year moving average of GA in generating PC_{t-6} in the calculation for SF_{t-6} resolves both underestimation and circular problems now that SF_{t-6} is not a simple function of the hog slaughter number in month t, but a function of PC_{t-6} and PC_{t-6}, which, in turn, is not a direct function of SF but is a function of the 4-year moving average of GA.

The study also regenerated other data series of FB, GA, etc. on the basis of this renewed SF_{t-6}, which is consistent with PC_{t-6}. These regenerated data series not only take observed slaughter numbers into account but also reflect imputed BHA, imputed NBGS, and imputed PDL in the data estimation process.

Weighting Procedure for Alternative Pig Crop Data over the January 2006–June 2006 Period

In this study, we developed the alternative PC on the basis of the observed slaughter hog numbers and known biological function for slaughter hogs, which is ignored in the USDA PC data projection system. Even though this approach would provide more biologically and logically consistent historical PC data than the USDA approach, the alternative PC data are not available in real time because they are generated from 6 months ahead of slaughter number.

Therefore, the latest available alternative PC data always lagged 6 months behind the USDA PC data. In fact, the latest month of alternative PC data generated on the basis of the previous data generation process was December 2005, whereas the latest USDA PC was June 2006 (end month of the sample period of PC).

In this study, we followed the weighting procedure of using the ratio of the alternative PC to the USDA PC (alternative PC/USDA PC) over the December 1995–December 2005 period to generate an alternative PC over the January 2006–June 2006 period. The study obtained mean, median, standard deviation, skewness, and kurtosis for the ratio of

alternative PC to USDA PC in order to examine the distribution of the ratio. They are 0.99, 0.99, 0.04, 0.31, and 2.18, respectively, which indicates that the ratio over the December 1995–December 2005 period follows a normal distribution. Furthermore, the *p*-value of the Jarque and Bera normality test is 0.07,[7] which supports the normality of this ratio.

The study multiplied 0.99, which is the same mean and median of the ratio, with the USDA PC over the January 2006–June 2006 period because the ratio is normally distributed. Additionally, this study also multiplied monthly average of the ratio (MA) over the January 1996–December 2005 period with the USDA PC over the January 2006–June 2006 period to preserve seasonality. That is, alternative PC_t = USDA $PC_t \times 0.99 \times MA_t$, where *t* is January 2006–June 2006.

Formulas for the Major Generated Variables

The foregoing section of the study described the alternative data generation process of PC and SF coupled with unknown production variables such as FB, BHA, CDBGP, etc. on the basis of the alternative SF data. In addition to the alternative data set, we also generated FB, BHA, and NBGS in this study consistent with the USDA SF data to obtain $CDBGP_t$ from USDA PC_{t-6}. As mentioned earlier, this study compares USDA PC_{t-6} with the calculated $CDBGP_t$ from USDA PC_{t-6} rather than $CDBGS_t$ in calculating imputed pig death loss contained in the USDA PC so that we can obtain a more accurate USDA PDL.

Formulas for the unknown USDA hog production variables of FB, BHA, NBGS, and CDBGP are the same as those of the alternative data generation processes, except we did not use a 4-year moving average of GA. Instead, we used GA generated directly from USDA SF because there is no circular

problem in generating FB, BHA, and NBGS from the given USDA SF number.

Table 4 documents formulas for the major generated variables and also compares the mean and standard deviation of variables between the USDA and alternative data systems. It is important to note that all variables have the same formula, but PC_{t-6} and SF_{t-6} are reported as variables in USDA system. All variables except PC_{t-6} and PDL_t are obtained from each SF_{t-6}, and SF_{t-6} is generated from PC_{t-6}. Therefore, comparison of each PC and PDL between the USDA and alternative systems is sufficient for comparisons of all variables in the two data system.

The mean of the USDA PC (8,451.4 thousand heads) is 0.97% higher than the mean of the alternative PC (8,369.9 thousand heads). In contrast, the standard deviation of the USDA PC (340.64 thousand heads) is 38.78% lower than the standard deviation of the alternative PC (556.41 thousand heads).

Relatively constant USDA PC numbers compared with alternative PC data lead to a considerably volatile USDA imputed PDL in Figure 1. The standard deviation of the USDA PDL (300.43 thousand head) is 468.46% higher than the standard deviation of the alternative PDL (52.850 thousand head), whereas the mean of the USDA PDL (520.42 thousand head) is only 19.27% higher than the mean of the alternative PDL (436.34 thousand head).

Flow Chart of the Data Generation Process

The discussion of data generation processes is very detailed and complex. It is quite important to capture every detailed interrelationship between data components in order to generate hog production data that are consistent with the known biological production function for hogs. It is also extremely important to incorporate comprehensive death losses and conception rates of animals into data components in the data generation process to correctly reflect the U.S. swine industry.

---

[7] We failed to reject the null hypothesis of normality at the 5% level of significance.

Table 4. Formulas and Descriptive Statistics of Major Generated Variables

Formula[a]	Mean		SD	
	USDA	Alternative	USDA	Alternative
$SF_{t-6} = (CDBGS_t \times WDA_t)/PPL_{t-6} = PC_{t-6}/PPL_{t-6}$	958.88	949.62	31.130	60.566
$FB_t = SRB_t + GB_t$	960.06	950.47	29.869	60.528
$SRB_t = [SF_{t-2} - (SF_{t-2} \times SCF_{t-2}/100)] +$				
$(SF_{t-3} \times SCF_{t-3}/100) - SDL_{t-2} - CDSS_t \times WDA_t$	683.65	675.17	35.505	57.605
$GB_t = SF_{t+4} - SRB_t$	276.41	275.31	38.328	98.671
$BHA_t = BA_t + GA_t$	361.13	355.72	41.768	106.28
$GA_t = GB_{t+1}/(CRG_t/100)$	330.52	325.11	40.830	104.76
$NBGS_t = GA_{t-3} \times [(1 - CRG_{t-3}/100]$	54.110	49.810	16.028	11.115
$PC_{t-6} = CDBGP_t + PDL_t$	8,451.4	8,369.9	340.64	556.41
$CDBGP_t = CDBGS_t \times WDA_t - NBGS_t + BHA_t$	7,931.0	7,933.5	497.97	542.42
$PDL_t = PC_{t-6} - CDBGP_t = PC_{t-6} \times (MPPDL_t/100)$				
$= [(CDBGP_t) \times (MPPDL_t/100)]/[1 - (MPPDL_t/100)]$	520.42	436.34	300.43	52.850

Note: Unit for mean and standard deviation is 1,000 heads.
[a] SF_{t-6} and PC_{t-6} are reported variables in the USDA data system.

However, the detail is difficult to follow, but easier in an overall flow chart for those not familiar with the U.S. swine industry. Figure 2 summarizes the data generation process from the slaughter hog numbers to the size of breeding herd. The flow chart explaining the data generation process does not attempt to replicate every detailed component; rather, it tries to simplify the entire process for easier understanding.

Monthly U.S. Slaughter Hog Model

In addition to the pure comparison between USDA data and the alternative data series by basic descriptive statistics in Table 4, we compare the relative empirical performance of each PC data set by putting it into a monthly slaughter hog model.

Monthly variations in slaughter hog supply in month t are governed by the biological production process once the size of the breeding herd is determined in $(t - 11)$.[8] However, as mentioned earlier, there is some flexibility in the timing of slaughter. There-

fore, the domestic hogs slaughtered commercially in month t should be specified not only by PC_{t-6} but also WD_t,

$$(4) \quad CDBGS_t = \alpha_0 + \alpha_1 PC_{t-6} + \alpha_2 WD_t + \varepsilon_t,$$

where ε_t is the error term. Definitions for other variables are reported in Table 3.

On the basis of the slaughter hog model in Equation (4), two empirical models are specified to evaluate empirical performances of USDA and alternative PC data series.

$$(5) \quad CDBGS_t = a_0 + a_1 PC_{t-6}^U + a_2 WD_t + a_3 TU_{t-6} + V_{t1}$$

$$(6) \quad CDBGS_t = b_0 + b_1 PC_{t-6}^A + b_2 WD_t + b_3 TA_{t-6} + V_{t2}$$

where PC_{t-6}^U is the size of the USDA PC_{t-6}; TU_{t-6} is the product of the annual time trend and size of USDA PC_{t-6}; PC_{t-6}^A is the size of the alternative PC_{t-6}; TA_{t-6} is the product of the annual time trend and size of the alternative PC_{t-6}; and V_{t1}, V_{t2} are error terms for Equations (5) and (6). Definitions for other variables are the same as Equation (4).

TU_{t-6}, TA_{t-6} are included to capture the increasing growing/finishing period of pig from birth to slaughter because of the increasing slaughter weights in recent years. We expect negative signs in both TU_{t-6} and TA_{t-6}.

---

[8] The decision process for determining the breeding herd size in month $(t - 11)$ is not the objective of the study. We limit our focus only to the slaughter hog equation in order to test the appropriateness of alternative PC relative to USDA PC data.
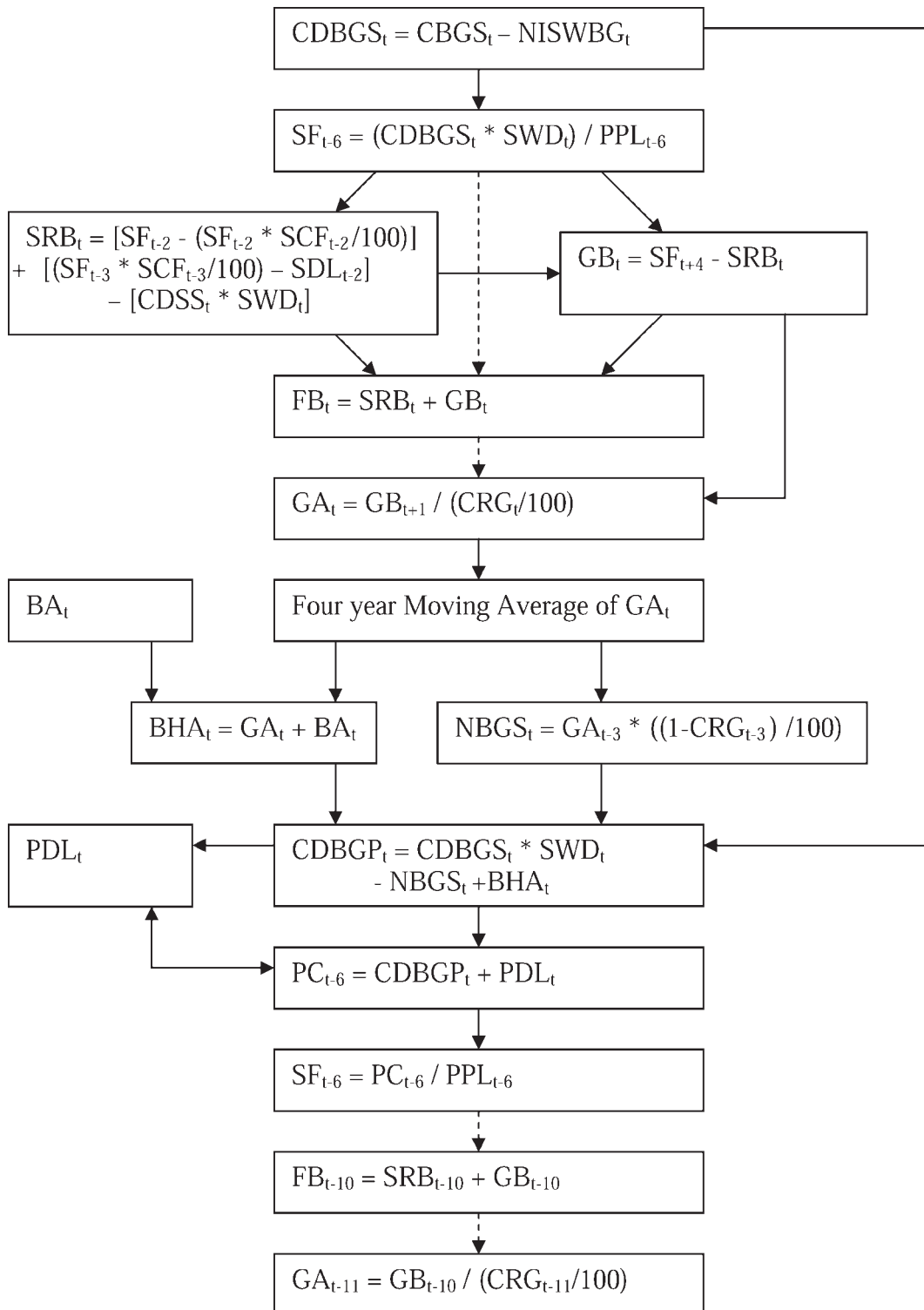
Figure 2. Flow Chart of Data Generation Process

Table 5. Estimation Results of Monthly U.S. Slaughter Hog Model with the Use of USDA and Alternative Pig Crop Data

	USDA Model—Equation (5)	Alternative Model—Equation (6)
Intercept	11.508 (24.14)[a] [25.33][b]	3.8039 (5.480)
PC^U_{t-6}	0.7433* (0.119) [0.164]	—
PC^A_{t-6}	—	0.9046* (0.020)
TU_{t-6}	−0.0078 (0.010) [0.009]	—
TA_{t-6}	—	−0.0034 (0.002)
WD_t	302.16* (13.22) [15.66]	376.16* (3.186)
R^2	0.8359	0.9915
Log likelihood	−877.85	−691.13
B–G LM[c]	0.0438	0.1076
J-test[d]	0.0000	0.3388
Ex post forecasting performance test results (July 2005–December 2006)		
RMSE	247.82	72.474
MAE	205.50	54.258
MAPE	58.660	18.278

Notes: RMSE is root mean squared error, MAE is mean absolute error, MAPE is mean absolute percentage error.
[a] Ordinary least squares standard error.
[b] Newey and West standard error.
[c] The *p*-values of the Breusch–Godfrey serial correlation Lagrange multiplier test with the minimized Akaike Information Criterion (AIC). Using the minimized AIC, we identified a fourth-order serial correlation to be tested in the Breusch–Godfrey Lagrange multiplier test.
[d] The *p*-values of Davidson and Mackinnon's nonnested specification test (pairwise *J*-test).
* $p \leq .01$.

Empirical Results

Equations were estimated with CDBGS, WD for the June 1996–December 2006 period (127 months), 11 years annual time trend, and both USDA and alternative PC for the December 1995–June 2006 period (127 months).

We conducted the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) stationary test for all variables before estimation.[9] The null hypothesis of stationary was rejected for all variables except PC^A and WD. All variables were first differenced for consistency in evaluating the empirical performance of PC^U and PC^A in Equations (5) and (6), even though PC^A and WD are stationary and the first differenced variables were found to be stationary. Table 5 reports estimation results of

Equations (5) and (6) by OLS coupled with various test results.

All estimated variables in Equations (5) and (6) had expected signs, but the parameter estimates in the two sets of data exhibit quite distinct differences between them, as expected.

Parameter estimates of Equation (5) in the first column indicate that, on average, 74.3% of pigs born in the 6 months previous are slaughtered, and about 302,000 domestic slaughter hogs are slaughtered commercially per day on the basis of the reported USDA pig crop data series over the sample period. This is not quite consistent with our observations of the U.S. swine industry. Moreover, a TU_{t-6} term with a value of −.0078 modifies the average hog slaughtered number per pig crop in the 6 months previous to .66 in year 2006.

On the other hand, estimation results for Equation (6) in the second column shows that, on average, 90.4% of pigs born in the 6 months previous are slaughtered, and about 376,000 domestic slaughter hogs are slaugh-

[9] The KPSS test statistic for CDBGS, PC^U, PC^A, WD, TU, and TA are 0.4001, 0.5046, 0.1399, 0.1195, 1.3768, and 1.3726, respectively, and the asymptotic critical value at 1%, 5%, and 10% are 0.739, 0.463, and 0.347, respectively.

tered commercially per day on the basis of the generated alternative pig crop data series during the same period. A TA_{t-6} term with a value of $-.0034$ reduces average hog slaughtered number per pig crop in the 6 months previous to about .87 in year 2006. This is quite consistent with current trends in the U.S. swine industry.

On the basis of the estimation results, we conducted Davidson and Mackinnon's non-nested specification tests because neither Equation (5) nor (6) is nested by other equations. Davidson and Mackinnon recommend the *J*-test among many nonnested tests when null hypothesis is linear.

Table 5 reports a pairwise nonnested *J*-test result for Equations (5) and (6). The *p*-value of the *J*-test statistics for Equation (5) indicates strong rejection of the null hypothesis that it is correctly specified against the alternative hypothesis that Equation (6) is correctly specified. This implies that the alternative pig crop data series provides significant additional information for monthly variations of commercial domestic hogs slaughtered that was not accounted for by the USDA pig crop data.

On the other hand, the *p*-value of the *J*-test statistic for Equation (6) shows strong non-rejection of the null hypothesis that it is correctly specified against the alternative hypothesis that Equation (5) is correctly specified. This suggests that the USDA pig crop data series cannot supply additional information for monthly variations of commercial domestic hogs slaughtered that has not already been explained by the alternative pig crop data.

Even though estimation and specification test results were consistent with our prior expectation, the highly volatile and systematic death loss contained in the USDA pig crop over the sample period strongly suggests the existence of serial correlation in the residual of Equation (5). Table 5 reports the Breusch–Godfrey serial correlation Lagrange multiplier (B–G LM) test. As expected, the *p*-values of the B–G LM test indicate the existence of serial correlation in Equation (5)

but no serial correlation in the residual of Equation (6).[10]

In the presence of serial correlation, the usual OLS standard errors and, subsequently, test statistics are not valid, even asymptotically. So, this study provides Newey and West standard errors which are consistent in the presence of both heteroskedasticity and serial correlation of unknown form. The Newey and West standard errors were not significantly different from the OLS standard errors, which suggests no major effect from the serial correlation on the estimated OLS standard errors and, hence, test results in Equation (5).

We also compared *ex post* forecasting performance of two data series over the July 2005–December 2006 period (18 months) on the basis of the estimated parameters we discussed earlier. As reported in Table 5, all three forecasting performance statistics indicate that the alternative model performs better than the USDA model in predicting CDBGS; that is, the alternative model exhibits smaller errors than the USDA model in predicting CDBGS over the July 2005–December 2006 period.

Conclusions

Previous studies of the monthly market for slaughter hogs and pork have been hampered by data limitations. The USDA began reporting monthly estimates of the size of the U.S. pig crop and sows farrowing in December 1995. However, there are significant measurement errors in the USDA monthly pig crop data series. We found that the imputed pig death loss contained in the reported monthly pig crop data over the December 1995–June 2006 period ranged from -4.93% to 12.75%.

With this study, we generated historical monthly pig crop data along with other unobserved production data series. The data

---

[10] We rejected null hypothesis of no serial correlation at the 5% level of significance in Equation (5) but failed to reject the null hypothesis even at the 10% level in Equation (6).

generation process used monthly USDA hog slaughter data as the anchor and imposed known information about gestation and growing periods, death loss, and conception rates of the animal.

The alternative monthly pig crop data series is logically and biologically consistent over the historical period, in contrast to the USDA pig crop data. Furthermore, the specification and *ex post* forecasting tests, on the basis of the same slaughter hog model, indicate that the alternative pig crop is a better estimate than the reported monthly USDA pig crop data in forecasting future months' slaughter hog numbers.

These generated monthly production data series will provide consistent data for further studies on the U.S. hog industry. They will allow researchers to incorporate the biological lag process of pork production into their specifications of short-term (month or quarter) supply models. Furthermore, the newly generated gilt addition and females bred data will supply additional important information for future studies on U.S. hog producers' decision-making process.

References

Aadland, D., and D. Bailey. "Short-Run Supply Responses in the U.S. Beef-Cattle Industry." *American Journal of Agricultural Economics* 83(November 2001):826–39.

Chavas, J.-P. "On the Economic Rationality of Market Participants: The Case of Expectations in the U.S. Pork Market." *Journal of Agricultural and Resource Economics* 24(1999):19–37.

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

Griliches, Z. "Economic Data Issues." *Handbook of Econometrics*, Volume 3. Z. Griliches and M.D. Intriligator, eds. Amsterdam, New York, and Oxford: North-Holland, 1986.

Hausman, J.A. "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left." *Journal of Economic Perspectives* 15(Autumn 2001):57–67.

ISU (Iowa State University), University Extension. *Iowa Livestock Enterprise Summaries*. Ames, Iowa: Various issues, 1990–1999.

Livestock Marketing Information Center. *Monthly Commercial Hog Slaughter, Monthly Federally Inspected Hog Slaughter by Classification (Hogs, Barrow & Gilts, Sows, and Boars)*. Lakewood, Colorado, Internet site: http://www.lmic.info (Accessed July 2007).

U.S. Department of Agriculture, Agricultural Marketing News. *Livestock and Grain Market News, Canadian Live Animal Imports by States of Entry*. Internet site: http://www.ams.usda.gov (Accessed July 2007).

U.S. Department of Agriculture, Economic Research Service. *Livestock and Poultry Situation and Outlook Report, Hog Imports and Exports*. Various issues.

U.S. Department of Agriculture, NASS (National Agricultural Statistics Service). US & State—Hogs & Pigs, Number of Pigs per Litter, Pig Crop, and Sows Farrowed. Internet site: http://www.nass.usda.gov/QuickStats/ (Accessed July 2007).