Chapter 2

Supply

In this section we will begin by studying models of production. Second, we will use duality to describe production technology in models where prices as dependent variables. Third, we will look into how to estimate supply response models. Finally, we will study the competitive storage model and livestock supply model.

Useful references for the material in this chapter are Chambers (1988) and Beattie et al. (2009). The chapter by Mundlack (2001) in the Handbook of Agricultural Economics also gives a good review but is quite difficult to read. You can also check Fuss and McFadden (1978a) and Fuss and McFadden (1978b). These are older texts but are available for free online. Henningsen (2018) is a good reference about applied work in production analysis. For the material about the competitive storage model, Miranda and Fackler (2002) is a great reference to learn how to solve the model using collocation.

2.1 Primal estimation of production

The objective in estimating a production function is quite obviously to describe production technology in function that explains how inputs combine to produce one or many outputs, how inputs substitute or complement each other and verify whether there are economies or scale or scope.

Why estimate a production function? Reasons include:

- Test theory that $p\frac{\partial f(\mathbf{x})}{\partial x_i} = w_i$;
- Examining the contribution of factors of production;
- Testing for technical change;
- Obtaining estimates of:
  - Production elasticities;
  - Returns to scales;
  - The effect of input quality on production;
  - The effect of management;
  - Technical efficiency.

The direct estimation of production functions is not very common anymore as the dual approach has mostly taken over. There are nonetheless some cases where estimating a production is a sensible approach, provided that the use of a proper method. At first glance, estimating a production function seems simple as a linear regression of the log of output on the log of inputs may appear to perform well. We will see that there are a number of important problems with that approach.

2.1.1 General issues

To help identify the issues regarding the estimation of production function, we will take a look in this section at two papers from great economists. You can also find a survey of these issues in Mundlack (2001).

2.1.1.1 Schultz (1956)

Theodore W. Schultz is maybe the most recognized agricultural economist of the 20th century. A former member and Chair of the department of economics at Iowa State University, Schultz received the Nobel Memorial Prize in Economics in 1979 as a member of the department of economics at the University of Chicago.[1]

Schultz (1956) is interested in explaining long-run growth in agricultural output: "Growth in output cannot be explained by an analysis which is based in conventional inputs." The following summarizes the main points.

Schultz (1956) make several interesting observations regarding why it is more difficult to estimate supply than demand. The first is that we know a lot more about demand than supply. Estimating elasticities of demand is relatively easy. Simple estimates of demand elasticities (e.g. log-log model) in general yield estimates close to estimates from models consistent with economic theory (e.g. AIDS). Much less is known on the supply side. For example, there were no meaningful estimates of elasticity of supply at the time of Schultz wrote more than 50 years ago, and this might still be true today. One issue is that supply elasticities depend on the length of run. Schultz (1956) also raised the issue that estimating a production function will not work because it is based on conventional inputs. Over time, the quality of inputs changes. For example, it is easy to imagine how machinery can improve in quality in a 10-year period. Think for example of the introduction of computers and GPS recently. Another issue in estimating a production function is that for estimates of a production function to be useful, it must either be stable or change in a predictable way. Tastes do not change over time. However, on the production side, innovations are unpredictable which makes the production technology unpredictable as well.

Schultz (1956) summarizes the issues by noting that demand rests on a stable foundation while supply stands on *shifting sand*. The two main reasons are:

- Tastes are constant while production technology and input quality change;
- Economies of size, scale and the free entry of firms.

These two points imply that estimating a production function requires including dynamic and modeling the competitive environment where firms operate.

[1] Schultz left the department of economics at Iowa State University following a dispute on academic freedom.

Schultz (1956) claims that an analysis based on conventional inputs does not have predictive power. Table 2.1 shows table 1 in Schultz (1956). Observe that there is little consistency between the growth in output and the growth in input. Additional inputs account for less than one-fifth of the growth in output (1870-1950). Given the disconnection between the change in output and the change in inputs, measuring productivity growth is a measure of our ignorance (Abramovitz, 1956).

TABLE 1. RECENT CHANGES IN OUTPUT AND INPUT IN AGRICULTURE IN THE UNITED STATES

Period	Output (percent)	Input (percent)	Proportion of additional output accounted for by additional input[e] (percent)
(1)	(2)	(3)	(4)
1900–1920[a]	25	18 to 21	72 to 84
1910–1920[b]	15	17	100[f]
1923–1929			
I[a]	7	4	57
II[b]	7	3.7	53
1930–1940[b]	16	−5	0
1940–1948			
I[a]	25	6.5[d]	26
II[c]	25	5	20
1930–1950[c]	39	1	4
1950–1955	12	n.a.[g]	n.a.[g]

[a] D. G. Johnson's estimates in "The Nature of the Supply Function for Agricultural Products," *American Economic Review*, Vol. 51, September 1950. For 1900–1920, the output is based on "agricultural production for sale and home use on farms"; for the other periods the recent estimates of "farm output" are presented here.

[b] My estimates based on data obtained in preparing Chapter 7 of *Economic Organization of Agriculture*, McGraw-Hill, 1953, using, however, 1935–39 prices for the weights. The input series based on these weights has not been published.

[c] Same source as in footnote b above, using 1946–48 input prices as weights. This input series appears in Table 7-5 of my book.

[d] Johnson's input estimates for 1940–48 were based on the old labor input series, which substantially understated the decline of this input in agriculture and which for 1940–48 indicated an increase in input of 10 percent or less. Using census estimates of farm labor that are now available, however, the increase in input becomes 6.5 percent, as is shown here.

[e] With "constant" returns for additional inputs.

[f] Implies diminishing returns for the additional inputs.

[g] Not available.

Figure 2.1: Changes in Output and Input in Agriculture in the United States: table 1 in Schultz (1956)

Schultz (1956) offers possible explanations:

i. Greater division of labor (specialization);
ii. Improvement in the quality of people (knowledge/education);
iii. Inventions (new technology);
iv. Scale economies (either pre-existing or due to i., ii. or iii.).

The first explanation suggests that the production function is not stable. The second and third explanations, together, suggest that some inputs are neglected in conventional analysis of production. Schultz (1956) did not try to explain the importance of these factors in output growth.

For Schultz (1956), a strong and satisfying linkage between input and output is one where the ratio of output and input equals one:

$$\frac{\text{Output}}{\text{Input}} = 1,$$

where the measure of input can be thought as an index that includes change in quality. The closer the ratio is to one, the more complete is the explanation of production. The measure of input must consider neglected inputs that are new technologies and improvements in labor force.

Overall, inputs account for a variable and declining fraction of additional outputs. Schultz (1956) did not attempt to explain output growth but suggested that human capital and new technologies are the main drivers.

2.1.1.2 Griliches (1963)

Zvi Griliches was a student a T. Schultz and made important contributions in the economics to technology adoption, R&D, patents and econometrics.

A typical production function can be represented as

$$y_t = g(\mathbf{x}_t, u_t, T_t),$$

where y_t is an index of physical output, $\mathbf{x}_t$ is a set of *measurable* inputs, u_t is a random variable, T_t is the *level of technology*, often approximated by a trend variable. For Griliches (1963), this approach is not fruitful and therefore not useful if it is not a stable function because there remains large unexplained shifts. A measure of technical change from the coefficient for the trend variable is a measure of our ignorance. Griliches (1963) proposes to attribute growth in output to changes in quantities and qualities of inputs and to economies of scale.

For Griliches (1963), the main sources of growth that are unexplained by conventional analysis are:

i. Improvement in quality of labor;
ii. Improvement in quality of machinery;
iii. Wrong factor weights - underestimation of the contribution of capital and overestimation of the contribution of labor;
iv. Economies of scale.

The third source of growth denies equilibrium assumptions that $w_K < MVP_K$ and $w_L > MVP_L$, where MVP is marginal value product. These inequalities imply a flow of capital toward agriculture and a flow of labor outside of agriculture.

Measured productivity growth is therefore a specification or measurement error. Griliches (1963) suggests to include new variables in the estimation of production functions that include:

- R&D capital;
- Education;
- Government and private sector research and extension work;
- Infrastructure.

These variables account for knowledge and quality of inputs.

Griliches (1963) shows the importance of including those variables in a production function by estimating a production function using 1949 data for sixty-eight regions of the United States. The model is a Cobb-Douglas production function

$$\begin{aligned}\ln(\text{Value of production}) &= \beta_1 \ln(\text{Livestock expense}) \\ &\quad + \beta_2 \ln(\text{Other expense}) + \beta_3 \ln(\text{Machinery}) \\ &\quad + \beta_4 \ln(\text{Land}) + \beta_5 \ln(\text{Buildings}) + \beta_6 \ln(\text{Man-years}) \\ &\quad + \beta_7 \ln(\text{Education}) + u_t.\end{aligned}$$

Table 2.2 shows the results obtained by Griliches (1963). The two rows in the table shows the outcomes of two regressions. The columns represent regressors. Observe that the coefficient are very different when the variable for education is included and note that the coefficient for education is very significant.

TABLE 1

AGGREGATE AGRICULTURAL PRODUCTION FUNCTION, UNITED STATES, 1949, SIXTY-EIGHT REGIONS*

REGRESSION	COEFFICIENTS								
	X_1 (Livestock Expense)	X_2 (Other Current Expense)	X_3 (Machinery)	X_4 (Land)	X_5 (Buildings)	X_6 (Man-Years)	E (Education)	R^2	Sum of Coefficients (Excluding E)
($U17$)	.169 (.023)	.121 (.032)	.359 (.048)	.170 (.033)	.094 (.044)	.449 (.072)	....	.977	1.362 .....
($R6$)	.140 (.025)	.111 (.031)	.325 (.049)	.167 (.032)	.085 (.042)	.524 (.076)	.431 (.181)	.979	1.352 .....

* All variables are logarithms of original values unless otherwise specified; units = averages per commercial farm. Numbers in parentheses are the calculated standard errors of the respective coefficients.

X_1, log of (purchases of livestock and feed and interest on livestock investment)
X_2, log of (purchases of seed and plants, fertilizer and lime, and cost of irrigation water purchased)
X_3, log of (purchases of gasoline and other petroleum fuel, repairs of tractors and other machinery, machine hire, and depreciation and interest on machinery investment)
X_4, log of interest on value of land
X_5, log of (building depreciation and interest)
X_6, log of average full-time equivalent number of workers per commercial farm
E, logarithm of the average education of the rural farm population weighted by total U.S. income by education class weights; not per commercial farm but per man.
Dependent variable—log of value of farm production per commercial farm

Figure 2.2: Aggregate agricultural production function: table 1 in Griliches (1963)

2.1.2 The Cobb-Douglas model of production

The Cobb-Douglas is a functional form that is commonly used when estimating a production function. Mundlack (2001) offers a good summary of what he calls the *Cobb-Douglas culture*.

The Cobb-Douglas production function for the case of two inputs is simply

$$y_t = A(t) x_{1t}^{\beta_1} x_{2t}^{\beta_2} e^{u_t},$$

where $A(t)$ can vary with time to capture *technical change*. The Cobb-Douglas production function is usually estimated in log-linear form

$$\ln y_t = \ln A(t) + \beta_1 \ln x_{1t} + \beta_2 \ln x_{2t} + u_t.$$

There are several advantages of estimating a Cobb-Douglas production function. First, it is easy to implement as it only requires estimating a linear model. Second, it does not require much data as observations for output quantity and input quantities are sufficient. Third, it allows for the measurement of return to scale. For example, in the two-input example above,

we observe increasing return to scale if $\beta_1 + \beta_2 > 1$. Fourth, the Cobb-Douglas allows to verify whether $w_L = MVP_L$. Finally, the Cobb-Douglas model does not assume perfect competition.

Many reasons explain why the Cobb-Douglas model of production, or other primal models of production, are not often used. In the case of the Cobb-Douglas model, it imposes unitary elasticities of substitution. The model also assumes that input quantities are exogenous, which is incorrect, causing OLS estimates to be inconsistent. Mundlack (2001) suggests using covariance analysis with panel data to correct the bias from endogeneity. It is also possible to use prices, which are exogenous, as instruments. It is difficult to estimate a model of production with micro data (cross section) because there is not much variation in prices and therefore little variation in input ratios. There are also aggregation problems with the Cobb-Douglas production function. Estimates that assume Cobb-Douglas technology lack robustness across studies. One difficulty is how to handle capital. Should capital be included as a stock or a flow (buildings and tractors are not used all year long)? Finally, one important shortcoming of the Cobb-Douglas model is how to handle multiple outputs. For instance, in agriculture, most farms produce more than one output. In Iowa, it is common for farms to produce corn and soybeans.

There might be no endogeneity problem in the primal model of production if production is not deterministic and producers know about the stochastic nature of production. If production is indeed not deterministic, then it is possible that inputs are not determined simultaneously with the output (Zellner et al., 1966; Henderson, 2015).

For the reasons described above, the primal model of production has fallen out of favor. Economists now rely most often on the duality of production in their empirical approach.

Note that there are many functional forms other than the Cobb Douglas to model production. An adaptation of the translog model that we saw in the chapter on demand can be used to estimate a production function. Flexible functional forms, that we will cover later in this chapter, are less restrictive and allow to consider multiple outputs.

2.1.3 Technical efficiency

One motive for estimating a production function is to measure the efficiency of firms. If firms are not efficient, it means that they are not using resources in an optimal manner, implying an allocation of resources that is not Pareto optimal. Inefficiency may be caused by technical change as firms do not adopt new, better, technology all at the same time.

Technical efficiency measures the ability of firms in combining inputs to produce a maximum output (Koopmans, 1951). It can be measured as the ratio between the observed output and the maximum output under the assumption of fixed input. Alternatively, technical efficiency is measured as the ratio between the observed input and the minimum output under the assumption of fixed output.

This section focuses on technical efficiency. We will ignore *allocative efficiency*, which is the combination of inputs and outputs in optimal proportions given prevailing prices, and is better estimated using a cost function (see Kopp and Diewert, 1982).

To look at measures of technical efficiency, we must first define the production technology. Consider that production involves m outputs $\mathbf{y} = (y_1, y_2, ..., y_m) \in \Re^m_+$ and n inputs $\mathbf{x} =$

$(x_1, x_2, ..., x_m) \in \Re^m_+$. We can summarize outputs and inputs through a netput vector $\mathbf{z} \equiv (-\mathbf{x}, \mathbf{y})$. The production technology is given by the feasible set F, such that $\mathbf{z} = (-\mathbf{x}, \mathbf{y}) \in F$, where we will assume that F is closed.

Measuring technical efficiency requires defining a metric that captures the efficiency of firms in combining inputs. To this end, define

$$Eff(-\mathbf{x}, \mathbf{y}) \equiv \{(-\mathbf{x}, \mathbf{y}), (-\mathbf{x}, \mathbf{y}) \in F, (-\mathbf{x}', \mathbf{y}) \notin F, \forall \mathbf{x}' \leq \mathbf{x}\}$$

as the efficient production. This is the frontier of the production set.

2.1.3.1 Shephard's input distance function

Shephard (1953) defines the input distance function as

$$\begin{aligned} D^S_I(-\mathbf{x}, \mathbf{y}) &= \max_{\alpha}\{\alpha : (-\frac{\mathbf{x}}{\alpha}, \mathbf{y}) \in F\} \text{ if } (-\frac{\mathbf{x}}{\alpha}, \mathbf{y}) \in F \text{ is feasible} \\ &= -\infty \text{ otherwise.} \end{aligned}$$

For $D^S_I(-\mathbf{x}, \mathbf{y}) > 1$, a firm is inefficient as it is possible to scale down its inputs and produce the same output given the existing technology. Letting $D^S_I(-\mathbf{x}, \mathbf{y}) = 1$ provides the multi-input, multi-output production frontier.

2.1.3.2 Farrell's input distance function

Farrell (1957) defines the input distance function as

$$\begin{aligned} D^F_I(-\mathbf{x}, \mathbf{y}) &= \min_{\beta}\{\beta : (-\beta\mathbf{x}, \mathbf{y}) \in F\} \text{ if } (-\beta\mathbf{x}, \mathbf{y}) \in F \text{ is feasible} \\ &= \infty \text{ otherwise.} \end{aligned}$$

Observe that Farrell's input distance function is the inverse of Shephard's input distance function. Thus, for $D^F_I(-\mathbf{x}, \mathbf{y}) < 1$, a firm is inefficient as it is possible to scale down its inputs and produce the same output given the existing technology. Letting $D^F_I(-\mathbf{x}, \mathbf{y}) = 1$ provides the multi-input, multi-output production frontier.

2.1.3.3 Shephard's output distance function

Shephard (1953) defines the output distance function as

$$\begin{aligned} D^S_O(-\mathbf{x}, \mathbf{y}) &= \min_{\theta}\{\theta : (-\mathbf{x}, \frac{\mathbf{y}}{\theta}) \in F\} \text{ if } (-\mathbf{x}, \frac{\mathbf{y}}{\theta}) \in F \text{ is feasible} \\ &= -\infty \text{ otherwise.} \end{aligned}$$

For $D^S_O(-\mathbf{x}, \mathbf{y}) < 1$, a firm is inefficient as it is possible to increase its outputs with the same inputs given the existing technology. Letting $D^S_O(-\mathbf{x}, \mathbf{y}) = 1$ provides the multi-input, multi-output production frontier.

2.1.3.4 Econometric approach to measure efficiency

Measuring efficiency requires an empirical description of technology. This section covers, very briefly, the basic parametric estimation of technical efficiency using a production function. It is possible to preform similar analysis using a cost function. I do not discuss non-parametric methods such as Data Envelopment Analysis which is based on linear programming.

Write a Cobb-Douglas production function for a single output and n inputs as

$$y = \alpha \prod_{i}^{n} \left(\frac{x_i}{\upsilon} \right)^{\beta_i},$$

where υ is a measure of technical efficiency as defined by Shephard's input distance function. Taking logs, and writing the production function as an empirical model for a cross-section of firms yields

$$\ln y_j = \ln \alpha + \Sigma_i^n \beta_i \ln x_{ij} + u_j + e_j, \tag{2.1}$$

where the subscript j identifies a firm and $u_j = -\ln \upsilon_j$ when assuming constant return to scale, and e_j is an error term. In (2.1), the error term has two components: the stochastic term e_j captures random shocks while the systematic component $u_j \leq 0$ measures firm efficiency. Identification of u_j is difficult with cross-section data but possible when making distributional assumptions for the two error terms.

A panel of firms helps in identifying efficiency parameters. Rewrite the empirical model for the production function, assuming constant return to scale, for a panel of firms as

$$\ln y_{jt} = \ln \alpha + \Sigma_i^n \beta_i \ln x_{ijt} + u_j + e_{jt},$$

where t identifies time. Observe now that the stochastic term e_{jt} varies over producers and time while the parameter u_j for technical efficiency varies only over producers. It is thus possible to estimate u_j using a fixed effect or a random effect model. It is then possible to compare firms' efficiency by comparing their values for u_j. For an example of such approach, see Battese and Coelli (1992). This approach however may not guarantee that the distributional assumptions on u_j are met.

In some applications, measuring efficiency is not sufficient and someone might want to explain efficiency. This can be done by specifying a functional form for u_j and explain technical efficiency simultaneously with production technology by writing

$$u_j = g(\mathbf{z_j} : \delta),$$

where $\mathbf{z_j} \neq \mathbf{x_j}$ is a vector of explanatory variables for efficiency. It is common to assume that u_j either follows a truncated normal distribution or a half-normal distribution. You can find an example of this type of model in Coelli and Battese (1996).[2] An alternative is to do this in two stages and regress $\hat{u_j}$ on a vector of explanatory variables.

The estimation of technical efficiency is not limited to the Cobb-Douglas model. Many other models of production can be used, including the translog production function and non-parametric econometric models. If you wish to learn more about technical efficiency, I suggest that you look at the work by William Greene and the material on his webpage (http://people.stern.nyu.edu/wgreene/).

Endogeneity is a possible problem when estimating a production frontier. There are however several approaches for dealing with the endogeneity problem. See for example Guan et al. (2009), Shee and Stefanou (2014), Amsler et al. (2016) or Latruffe et al. (2017).

[2] Tim Coelli makes available a software called Frontier 4.1 to estimate production frontier using maximum likelihood (http://www.uq.edu.au/economics/cepa/frontier.php). An implementation of the software is available through the package Frontier in R (http://cran.r-project.org/web/packages/frontier/index.html).

2.2 Dual estimation of production

In the dual estimation of production, technology is summarized in equations for profits, costs or revenue. We will cover the estimation of a cost function. The estimation techniques are similar for profit functions and revenue functions.

There are several reasons for duality in estimating production technology. Duality of production implies that prices appear on the right hand side rather than quantities. For competitive firms, prices are exogenous and therefore there should be no endogeneity problem using the dual approach. Duality also allows to consider multiple outputs.

Perfect competition is a maintained assumption in what we will cover. See Morrison-Paul (2001a) or Morrison-Paul (2001b) for examples of estimation of cost functions under imperfect competition.

2.2.1 Desirable restrictions from the theory of production

Consider a representative firm that minimizes its cost to produce a given output y. This example applies to a firm that produces a single output but can easily be generalized to multi-output production. A firm minimizes its cost by choosing a set of inputs such that the cost function for a single output firm is given by

$$C(\mathbf{w}, y) = \min_{\mathbf{x}} \mathbf{w} \cdot \mathbf{x} \text{ s.t. } f(\mathbf{x}) \geq y,$$

where $\mathbf{x}$ is a vector of inputs, $\mathbf{w}$ is a vector of input prices, $f(\mathbf{x})$ is the production function and y is the output.

When estimating a cost function, we will impose or test properties derive from the theory of production.

Properties:

i. Increasing in $\mathbf{w}$;
ii. Homogeneous of degree one in $\mathbf{w}$;
iii. Concave in $\mathbf{w}$;
iv. Strictly increasing in y.
v. Convex in y.

Using Shephard's lemma yields the input demand equations

$$\frac{\partial C(\mathbf{w}, y)}{\partial w_i} = x_i(\mathbf{w}, y). \tag{2.2}$$

2.2.2 Flexible functional forms

We will look at how to estimate cost functions using flexible functional forms. We used flexible function forms in the section about the translog model of demand. Before covering models of cost function, lets define in more details what is the meaning of a flexible functional form.

The concept of functional form that we will consider was introduced by Diewert (1971) and Diewert (1976). There are other definitions of flexibility. For instance, Gallant and Golub

(1984) introduce locally flexible functional forms using Fourier expansions while Diewert and Wales (1987) define globally flexible function forms.

A functional form is flexible if it has enough parameters to approximate a twice continuously differentiable function. A flexible function $f(\mathbf{x})$ of m variables has enough free parameters to satisfy the following $m^2 + m + 1$ equations

$$\begin{aligned} &f(\mathbf{x}) \to 1; \\ &\bigtriangledown f(\mathbf{x}) \to m; \\ &\bigtriangledown^2 f(\mathbf{x}) \to m^2. \end{aligned}$$

A flexible function form does not require $m^2 + m + 1$ parameters to satisfy these $m^2 + m + 1$ equations. By Young's theorem, $\bigtriangledown^2 f(\mathbf{x})$ is symmetric and therefore contains $\frac{m(m+1)}{2}$ equations (try to verify this). Thus, for a functional form to be called flexible, it needs only to satisfy one equation for the value function, m equations for the Jacobian and $\frac{m(m+1)}{2}$ equations for the Hessian for a total of $\frac{(m+1)(m+2)}{2}$ free parameters.

Note that flexible functional forms do not have the ability to closely approximate arbitrary technology. Parametric estimation of production technology always requires the specification of some functional form that has its own limitations. The advantage of flexible functional forms is not in their capacity to truly approximate technology but in their capacity to place fewer restrictions on technology.

2.2.3 Examples of flexible cost functions

We will see examples of functional forms that are linear in parameters. In particular we will look at functions of the *general linear form*

$$f(\mathbf{z}) = \sum_{i=1}^{N} \alpha_i b_i(\mathbf{z}),$$

where α_i is a parameter and $b_i(\mathbf{x})$ is a twice-continuously differentiable function. Mores specifically, we will consider *generalized quadratic* functional forms

$$f(\mathbf{z}) = \beta_0 + \sum_{i=1}^{N} \beta_i g_i(z_i) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_{ij} g_i(z_i) g_j(z_j),$$

where $\beta_{ij,} = \beta_{ji}$ and each $g_i(z_i)$ is a twice-continuously differentiable function. The difference between a general linear form and a generalized quadratic is that the generalized quadratic imposes more parametric restrictions.

2.2.3.1 Normalized quadratic cost function

The normalized quadratic cost function was introduced by Diewert and Wales (1987) but was called the *Symmetric Generalized McFadden* functional form.[3] The quadratic cost function with N inputs, M outputs and a time trend is given by

$$\begin{aligned} C^Q(\mathbf{W},\mathbf{Y},T) = \alpha_0 &+ \sum_{i=1}^{N}\beta_i W_i + \sum_{k=1}^{M}\delta_k Y_k + \tau_T T \\ &+ \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{i=1}^{N}\beta_{ij}W_iW_j + \sum_{k=1}^{M}\sum_{l=1}^{M}\delta_{kl}Y_lY_k + 2\tau_{TT}T^2\right) \\ &+ \frac{1}{2}\left(\begin{array}{l} \sum_{i=1}^{N}\sum_{k=1}^{M}\gamma_{ik}W_iY_k + \sum_{i=1}^{N}\rho_{iT}W_iT + \sum_{k=1}^{M}\pi_{kT}Y_kT \\ +\sum_{k=1}^{M}\sum_{i=1}^{N}\gamma_{ki}Y_kW_i + \sum_{i=1}^{N}\rho_{Ti}TW_i + \sum_{k=1}^{M}\pi_{Tk}TY_k \end{array}\right), \end{aligned}$$

where W_i is the price of input i, Y_k is the quantity of output k and T indicates time. You can verify that

$$\begin{aligned} &\bigtriangledown C^Q(\mathbf{W},\mathbf{Y},T) \rightarrow n+m+1 \text{ equations;} \\ &\bigtriangledown^2 C^Q(\mathbf{W},\mathbf{Y},T) \rightarrow (n+m+1)^2 \text{ equations.} \end{aligned}$$

We can impose symmetry by letting $\beta_{ik} = \beta_{ki}$, $\delta_{ik} = \delta_{ki}$, $\gamma_{ik} = \gamma_{ki}$, $\rho_{iT} = \rho_{Ti}$ and $\pi_{Tk} = \pi_{kT}$:[4]

$$\begin{aligned} C^Q(\mathbf{W},\mathbf{Y},T) = \alpha_0 &+ \sum_{i=1}^{N}\beta_i W_i + \sum_{k=1}^{M}\delta_k Y_k + \tau_T T + \frac{1}{2}\sum_{i=1}^{N}\beta_{ii}W_i^2 \\ &+ \sum_{i=1}^{N}\sum_{j=i+1}^{N}\beta_{ij}W_iW_j + \frac{1}{2}\sum_{k=1}^{M}\delta_{kk}Y_k^2 + \sum_{k=1}^{M}\sum_{l=k+1}^{M}\delta_{kl}Y_lY_k + \tau_{TT}T^2 \\ &+ \sum_{i=1}^{N}\sum_{k=1}^{M}\gamma_{ik}W_iY_k + \sum_{i=1}^{N}\rho_{iT}W_iT + \sum_{k=1}^{M}\pi_{kT}Y_kT. \end{aligned} \tag{2.3}$$

Symmetry of the Hessian means that we can reduce the number of equations from $(n+m+1)^2$ to $\frac{(n+m+1)(n+m+2)}{2}$. Thus, for the quadratic cost function to be flexible, it must have a total of $\frac{(n+m+2)(n+m+3)}{2}$ free parameters, which it does.

Recall that the cost function is homogeneous of degree one, i.e. $\lambda C^Q(\mathbf{W},\mathbf{Y},T) = C^Q(\lambda\mathbf{W},\mathbf{Y},T)$. We can impose homogeneity by dividing, that is normalizing, the cost and the input prices by one input price that represents a numeraire. Thus, using input 1 as numeraire, we can write that the *normalized* cost function is

$$\begin{aligned} \tilde{C^Q}(\mathbf{W},\mathbf{Y},T) = \alpha_0 &+ \beta_1 + \sum_{j=2}^{N}\beta_i\tilde{W}_i + \sum_{k=1}^{M}\delta_k Y_k + \tau_T T + \frac{1}{2}\beta_{11} + \frac{1}{2}\sum_{i=2}^{N}\beta_{ii}\tilde{W}_i^2 \\ &+ \sum_{j=2}^{N}\beta_{1j}\tilde{W}_j + \sum_{i=2}^{N}\sum_{j=i+1}^{N}\beta_{ij}\tilde{W}_i\tilde{W}_j + \frac{1}{2}\sum_{k=1}^{M}\delta_{kk}Y_k^2 + \sum_{k=1}^{M}\sum_{l=k+1}^{M}\delta_{kl}Y_lY_k + \tau_{TT}T^2 \\ &+ \sum_{k=1}^{M}\gamma_{1k}Y_k + \sum_{i=2}^{N}\sum_{k=1}^{M}\gamma_{ik}\tilde{W}_iY_k + \rho_{1T}T + \sum_{i=2}^{N}\rho_{iT}\tilde{W}_iT + \sum_{k=1}^{M}\pi_{kT}Y_kT. \end{aligned}$$

[3] Lau (1976) introduced the normalized quadratic profit function.

[4] It is possible write the expression in (2.3) more compactly. I write the cost function this way is because it is easier to show and impose parametric restrictions.

where the tildes ($\sim$) identify variables divided by the price of input 1. We can simplify the expression for the cost function to

$$\tilde{C^Q}(\mathbf{W},\mathbf{Y},T) = \tilde{\alpha}_0 + \sum_{j=2}^{N}\tilde{\beta}_i\tilde{W}_i + \sum_{k=1}^{M}\tilde{\delta}_k Y_k + \tilde{\tau}_T T + \frac{1}{2}\sum_{i=2}^{N}\beta_{ii}\tilde{W}_i^2 + \sum_{i=2}^{N}\sum_{j=i+1}^{N}\beta_{ij}\tilde{W}_i\tilde{W}_j$$
$$+ \frac{1}{2}\sum_{k=1}^{M}\delta_{kk}Y_k^2 + \sum_{k=1}^{M}\sum_{l=k+1}^{M}\delta_{kl}Y_lY_k + \tau_{TT}T^2 + \sum_{i=2}^{N}\sum_{k=1}^{M}\gamma_{ik}\tilde{W}_iY_k + \sum_{i=2}^{N}\rho_{iT}\tilde{W}_iT$$
$$+ \sum_{k=1}^{M}\pi_{kT}Y_kT.$$

where $\tilde{\alpha}_0 = \alpha_0 + \beta_1 + \frac{1}{2}\beta_{11}$, $\tilde{\beta}_i = \beta_i + \beta_{1j}$, $\tilde{\delta}_k = \delta_k + \gamma_{1k}$ and $\tilde{\tau}_T = \tau_T + \rho_{1T}$. When imposing homogeneity, identification of the parameters for the numeraire input is not possible. However, we will see below that for the other inputs the normalization has no effect.

It is common practice to estimate the cost function with the input demands simultaneously to increase precision of estimates. Adding the input demand functions adds information and therefore yields more precise estimates. Also, the input prices in the cost function may be highly collinear and the input demand equations help identify the parameters of the cost function. The input demand functions can be obtained by using the Shephard's lemma

$$\frac{\partial C^Q(\mathbf{W},\mathbf{Y},T)}{\partial W_i} = X_i^* = \beta_i + \sum_{j=1}^{N}\beta_{ij}W_j + \sum_{k=1}^{M}\gamma_{ik}Y_k + \rho_{iT}T, \forall i > 1. \tag{2.4}$$

If we instead consider the normalized cost function, applying Shephard's lemma yields

$$\frac{\partial \tilde{C^Q}(\mathbf{W},\mathbf{Y},T)}{\partial \tilde{W}_i} = X_i^* = \tilde{\beta}_i + \sum_{j=2}^{N}\beta_{ij}\tilde{W}_j + \sum_{k=1}^{M}\gamma_{ik}Y_k + \rho_{iT}T, \forall i > 1.$$

We can multiply the input price in the previous expression (input demand is homogeneous of degree zero) and recover the input demand in (2.4). Thus the normalization has no effect on the input demand. The demand for the numeraire can be recovered by writing

$$X_1^* = \tilde{C^Q}(\mathbf{W},\mathbf{Y},T) - \sum_{i=2}^{N}\tilde{W}_iX_i^*$$
$$= \tilde{\alpha}_0 + \sum_{k=1}^{M}\tilde{\delta}_kY_k + \tilde{\tau}_TT - \frac{1}{2}\sum_{i=2}^{N}\sum_{j=2}^{N}\beta_{ij}\tilde{W}_i\tilde{W}_j + \frac{1}{2}\sum_{k=1}^{M}\sum_{l=1}^{M}\delta_{kl}Y_lY_k + \tau_{TT}T^2$$
$$+ \sum_{k=1}^{M}\pi_{kT}Y_kT.$$

Recall that the cost function is concave in the input price (input demands slope down) and convex with respect to the outputs (output supplies slope up). Thus, the Hessian matrix for input price is negative semi-definite while the Hessian with respect to the output is positive semi-definite. Curvature restrictions on the cost function can easily be tested but imposing restrictions can be a little tedious when there are many outputs and many inputs.

Lau (1978) and Diewert and Wales (1987) show how to impose curvature restrictions on a cost function. See Featherstone and Moss (1994) for an application. The strategy involves taking a Cholesky decomposition of the Hessian matrix. Let B be the Hessian matrix

$$B = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1N} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \cdots & \beta_{NN} \end{pmatrix}.$$

Recall that every symmetric positive semi-definite matrix has a Cholesky decomposition. Thus, we can take a Cholesky decomposition of B and write that

$$B = -AA'.$$

From this decomposition we can impose restrictions on parameters such that estimates are consistent with curvature restrictions. However, such a method may impose stronger restrictions than needed...

2.2.3.2 Translog cost function

The translog cost function was proposed by Christensen et al. (1971). The translog cost function for N inputs and M outputs is given by

$$\begin{aligned}
\ln C^T(\mathbf{W}, \mathbf{Y}, T) = \alpha_0 &+ \sum_{i=1}^{N} \beta_i \ln W_i + \sum_{k=1}^{M} \delta_k \ln Y_k + \tau_T T \\
&+ \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{i=1}^{N} \beta_{ij} \ln W_i \ln W_j + \sum_{k=1}^{M}\sum_{k=1}^{M} \delta_{kl} \ln Y_l \ln Y_k + 2\tau_{TT} T^2 \right) \\
&+ \frac{1}{2}\left(\begin{array}{l} \sum_{i=1}^{N}\sum_{k=1}^{M} \gamma_{ik} \ln W_i \ln Y_k + \sum_{i=1}^{N} \rho_{iT} \ln W_i T + \sum_{k=1}^{M} \pi_{kT} \ln Y_k T \\ + \sum_{k=1}^{M}\sum_{i=1}^{N} \gamma_{ki} \ln Y_k \ln W_i + \sum_{i=1}^{N} \rho_{Ti} T \ln W_i + \sum_{k=1}^{M} \pi_{Tk} T \ln Y_k \end{array} \right).
\end{aligned}$$

Imposing symmetry, we can write

$$\begin{aligned}
\ln C^T(\mathbf{W}, \mathbf{Y}, T) = \alpha_0 &+ \sum_{i=1}^{N} \beta_i \ln W_i + \sum_{k=1}^{M} \delta_k \ln Y_k + \tau_T T \\
&+ \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \beta_{ij} \ln W_i \ln W_j + \sum_{k=1}^{M}\sum_{k=1}^{M} \delta_{kl} \ln Y_l \ln Y_k + 2\tau_{TT} T^2 \right) \\
&+ \sum_{i=1}^{N}\sum_{k=1}^{M} \gamma_{ik} \ln W_i \ln Y_k + \sum_{i=1}^{N} \rho_{iT} \ln W_i T + \sum_{k=1}^{M} \pi_{kT} \ln Y_k T.
\end{aligned}$$

We can impose homogeneity of the cost function by letting $\sum_{i=1}^{N} \beta_i = 1$, $\sum_{j=1}^{N} \beta_{ij} = 0$, $\sum_{i=1}^{N} \gamma_{ik} = 0$ and $\sum_{i=1}^{N} \rho_{iT} = 0$. Again, it is more efficient to estimate the cost function in a system with the input demand equations. The input demand functions can be obtained by using the Shephard's lemma: $\frac{\partial lnC}{\partial \ln W_i} = s_i$. Note that you must drop one demand equation because of singularity.

One problem with the translog cost function is the inability to place restrictions to assure that it is globally concave (e.g. Chambers, 1988, p. 181). The Hessian of the translog cost function is data dependent. Thus, parametric restrictions cannot assure the global concavity of the translog cost function. Ryan and Wales (2000) show how to impose concavity at a single point. The idea is to normalize to one the variables of the cost function at a reference point. Suppose for example normalization at time τ. For a variable z_{it}, the normalization implies that

$\tilde{z}_{it} = z_{it}/z_{i\tau}$. Thus, at $t = \tau$, the variable equals one and its log equals zero. This means that at the point of reference that the elements of the Hessian depend only on parameters of the model. Thus, it is possible to impose restrictions on the parameters of the model at the reference point using a Cholesky decomposition like we did in the normalized quadratic cost function. This technique does not however assure concavity at every points.

Another problem with the translog function is how to handle zero production. For example, if you estimate a production from a survey of farms, it is possible that not all farms produce all outputs. One option is to replace the zero outputs by a small positive value such that the logs are all analytic. Weninger (2003) introduces a methodology to deal with zero using different specifications for the cost function depending on the output bundle produced. For example, for similar firms produce Y_1 and/or Y_2, there would be three cost functions estimated in a SUR setting: 1) $Y_1 > 0$ and $Y_2 > 0$, 2) $Y_1 > 0$ and $Y_2 = 0$ and 3) $Y_1 = 0$ and $Y_2 > 0$. This Z-log cost function allows for testing whether the production technology is the same for multi-ouput firms and firms that do not produce the whole range of output. Another advantage of this procedure is that cost function remains linear in parameters.

2.2.3.3 Other linear in parameters flexible cost functions

The list below shows examples of other flexible cost functions for single-output production without a time trend.[5] The *Generalized Leontief*, the normalized quadratic and the translog are the most commonly used functional form for the cost function estimation.

- Generalized Leontief (Diewert, 1971):
  - $C(\mathbf{W}, Y) = \sum_{i=1}^{N} \alpha_i \sqrt{W_i} + Y \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_{ij} \sqrt{W_i W_j} + Y^2 \sum_{i=1}^{N} \gamma_i \sqrt{W_i}$
  - Restrictions: $\alpha_i = \gamma_i = 0$ for homogeneity.
- Generalized Cobb-Douglas (Diewert, 1973; Magnus, 1979):
  - $\ln C(\mathbf{W}, Y) = \alpha_0 + \alpha_y \ln Y + \frac{1}{2}\alpha_{yy}(\ln Y)^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{ij} \ln(\beta_i W_i + \beta_j W_j) + \sum_{i=1}^{N} \gamma_{yi} \ln Y \ln W_i$
  - Restrictions: $\gamma_{yi} = 0$ and $\sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{ij} = 1$ for homogeneity.
- Generalized Box-Cox (Berndt and Khaled, 1979):
  - $C(\mathbf{W}, Y) = \left(\frac{2}{\lambda} \sum_{i=1}^{N} \sum_{j=1}^{N} W_i^{\lambda/2} W_i^{\lambda/2}\right)^{\frac{1}{\lambda}} Q^{\beta(Y,\mathbf{W})}$ where $\beta(Y, \mathbf{W}) = \beta + \frac{\theta}{2} \ln Y + \sum_{i=1}^{N} \ln W_i$
  - Restrictions: Homogeneity is satisfied by construction.
- CES-translog (Pollak et al., 1984)
- Miniflex-Laurent (Barnett and Lee, 1985)
- Fourier (Gallant, 1981)
- Normalized quadratic semiflexible (Diewert and Wales, 1988)

2.2.4 Application: Ollinger et al. (2005)

The objective of Ollinger et al. (2005) is to analyze the drivers of structural changes in the poultry processing industry. The authors focus on scale economies and product and process innovations.

Two opposite cost effects are discussed. Table 2.3 shows table 1 and 2 in Ollinger et al. (2005). Observe that US poultry processing had been becoming more and more concentrated with larger plants. Thus, if there are economies of scale in the processing of chicken, this drives the

[5] The list is based on lecture notes from Richard J. Sexton at the University of California, Davis.

costs down. Second, firms have developed new products such as chicken nuggets or deboned breasts. Table 2 in Ollinger et al. (2005) shows the increase in the shipments of processed poultry products. These product innovations increase the cost of processing chicken. Another innovation is the rise of the turkey market which is no longer seasonal market (see table 3 in Ollinger et al. (2005)).

Table 1. Structural Change in Poultry Processing

	1967	1972	1977	1982	1987	1992
			Number of plants			
Chickens	140	194	179	134	125	144
Turkeys	75	59	50	36	31	30
			Animal slaughter (millions)			
Chickens	2,489	3,122	3,256	4,270	5,170	6,602
Turkeys	114	121	128	160	231	281
			Share of shipments from plants with over 400 employees			
Chickens	29	34	45	65	76	88
Turkeys	16	15	29	35	64	83
			Share of shipments from the largest four firms			
Chickens	23	18	22	32	42	41
Turkeys	28	41	41	40	38	35

Source: Slaughter data are from USDA (1997). Other data are drawn from Longitudinal Research Database, U.S. Bureau of the Census.

Table 2. Processed Poultry, Including Chicken Traypacks and Turkey Parts, Become a Major Component of Slaughter Plant Output

	Chicken			Turkey	
Year	Traypacks	Lunch Meat, Sausage, etc.	Parts[a]	Lunch Meat, Sausage, etc.	Parts[a]
		Share of industry shipments			
1963	n.a.	n.a.	12.5	n.a.	3.3
1972	11.0	2.6	28.4	8.5	15.7
1982	15.4	3.1	47.6	13.1	29.1
1992	18.9	3.1	78.5	16.8	55.1

[a]ERS, U.S. Egg and Poultry Statistical Series, 1960–90 (1991) for 1963–87 and ERS estimates for 1992.
Source: Longitudinal Research Database, U.S. Census Bureau and other sources as noted; n.a. is defined as not available.

Figure 2.3: Structural change in poultry processing: table 1 and 2 in Ollinger (2005)

2.2.4.1 Empirical strategy

Ollinger et al. (2005) estimate a cost function to identify the drivers of structural change in the US poultry using data from the Census of Manufactures. The cost function estimated take the form

$$C = f(Q, \mathbf{P}, \mathbf{Z}, \tau)$$

where Q is the output, $\mathbf{P}$ is a vector of factor (input) prices, $\mathbf{Z}$ are firm-level variables and τ is a vector of innovation measures.

The poultry processing industry has adopted cost lowering process innovations that include:

- Automated dressing;
- Larger chill bath;
- Larger and more uniform birds;
- Changes in production schedule.

The first two innovations are changes of processing plants production technology. Larger and more uniform birds are from better control of production practices in part driven by more vertical integration. Changes in production schedule are from removal of seasonality, especially for turkeys.

Some product innovations raise the cost of production:

- Deboned parts;
- Packaging by parts.

Chicken are now available by parts in grocery stores in the United States. This was not always the case and it used to be that consumers could only buy whole chicken.

Process innovations are difficult to observe and Ollinger et al. (2005) use a time trend as a proxy. Product innovations are captured by plant characteristics:

- Seasonality (↑ or ↓);
- Share of chicken in all poultry (↓);
- Share of bulk-packed products (↓);
- Share of chicken parts (↑);
- Share of whole bird in output for turkey plants.

The arrows indicate the expected sign of plant characteristics on the costs.

Ollinger et al. (2005) estimate a translog cost function using data from the Census of Manufacture between 1972 and 1992 for chicken and between 1967 and 1992 for turkey. The authors impose symmetry and homogeneity and select the preferred model using Likelihood Ratio (LR) tests. The preferred model for chicken includes: 1) factor prices, 2) output, 3) share of chicken in all poultry, 4) share of bulk-packed products and 5) share of chicken part in output. The preferred model for turkey include: 1) factor prices, 2) output, 3) share of bulk-packed products, 4) season and 5) share of whole bird in output.

Given the translog cost function estimated, the authors measure scale economies as

$$\epsilon_{CQ} = \frac{\partial \ln C}{\partial \ln Q}.$$

There are economies of scale if $\epsilon_{CQ} < 1$. That is, there are economies of scales of the percentage change in the cost is smaller than the percentage change in the output.

2.2.4.2 Results

The authors find substantial economies of scale: 0.883 for chicken and 0.808 for turkey for the largest plants. Given those economies of scale, why aren't plants even larger? First, there are high transportation costs for feed, chicks and mature poultry (e.g. death during transportation). Second, environmental laws restrict litter disposal and therefore raise costs. Thus, even though there are economies of scale within the plant, these economies of scale may not exit when considering production and transportation of live poultry.

Product differentiation (product innovations) raised plant costs by 6.3% for chicken and raised the costs by 18.6% for turkey. The authors also find that when not controlling for product mix, economies of scale are near one.

2.3 Supply response models

We can obtain a supply function from estimating a production function or a cost function as we saw in the previous section. However, with those approaches, it is more difficult to take into account dynamics that are important to firms' behavior and in practice it is not always

necessary to estimate production to obtain robust estimates of supply for the task at hand. Here, rather than explicitly investigating the technology of production, we will look at the dynamics of production and the relationship between the supply and prices. In particular, we will look at price expectations and production decisions.

2.3.1 Key points

We will begin by discussing two papers that identify the main difficulties in estimating supply curves. The first is an early discussion about the estimation of supply curve. The second a review of the "arts" of estimating supply curves.

2.3.1.1 Cassels (1933)

Cassels (1933) considers two approaches to estimate supply:

- Analysis of production: Fundamentally sound but of little practical value;
- Time series data: How does output depend on current or lagged prices?

Regarding the second approach, Cassels (1933) notes that there are three types of Marshallian curves that can be estimated:

i. Market supply curve: Holders' or dealers' curve rather than producers' curve;
ii. Long-run normal curve: Almost impossible to estimate. Too much short-term variation;
iii. Short-run normal curve: Supply curve constrained by current production equipments. Cassels (1933) argues that we always estimate short-run supply curves.

It is not obvious why Cassels (1933) thought it was almost impossible to estimate the long-run supply curve. The market supply curve can be estimated with the correct measurement of prices.

Cassels (1933) emphasizes on the time character of the supply function. There is not one supply curve but rather a fan-like system of supply curves where in the short run the supply is very inelastic and the very long term the supply is very elastic. Figure 2.4 illustrates this point.

Cassels (1933) defines the supply price as ... *the price, the expectation of which would call forth a given supply.* Thus, it is not the current price that affects production but rather the expected price at the moment of delivery. Cassels (1933) suggests that past prices might be alright to use.

Supply analysis is more difficult than demand analysis because of theoretical rather than empirical reasons:

- Time character of supply curve;
- One-way nature of supply curve (see figure 2.5) or asymmetric supply response.

The time character of the supply curve refers again to the lags between production decision and output delivery. The one-way nature of the supply curve refers to the difference in slope for an increase in production versus a decrease in production. For many products, it is easier to decrease production than to increase production. For example, in the production of eggs, the output can be easily reduced by culling laying hens. However, increasing production takes

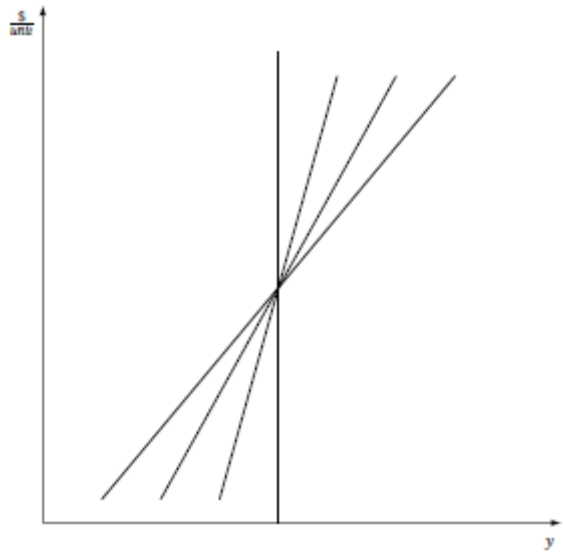


Figure 2.4: Fan-like system of supply curves

a long time because it may require building new facilities and raising hens, which takes about half a year. This means that the supply of eggs is much more elastic for price decrease than for price increase. Figure 2.5 an example of one-way supply curve.

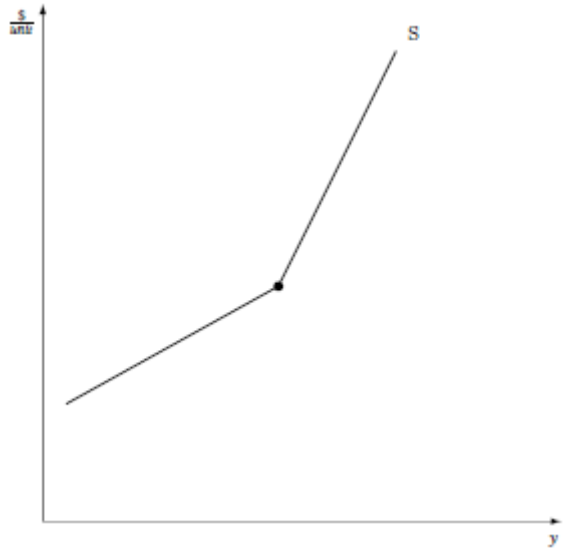


Figure 2.5: One-way nature of supply curve

In summary, Cassels (1933) identifies a number of issues with the estimation of supply including

- Methodology (synthetic vs time series);
- Dependent variable (output vs acreage);
- Length of run;
- Asymmetric response;
- Price expectations;
- Weather;
- Technical change;

- Competing product prices;
- Entry and exit of firms;
- Governmental support.

The second of these issues refers to the measurement of output. This is especially important in agriculture when exogenous shock such as weather can affect production. Cassels (1933) does not offer solutions for these issues. Cassels (1933)'s main points are still relevant today but better understood.

2.3.1.2 Colman (1983)

Colman (1983) offers a review of supply response models for agricultural products. Just (1993) offers a similar review. The author notes that the meaning of supply response is often confusing and rather uses the expression *output response*.

Colman (1983) suggests four types of approaches for analyzing supply response with the last three being different econometric approaches:

1. Programming;
2. Two-stage procedures;
3. Direct estimation of systems of supply;
4. Direct estimation of single commodity models.

Each of these approach has its advantages and each must be evaluated according to its theoretical correctness and predictive capacity. In each approach, an economist must make choices and use economic theory to determine:

- The form of the system of equations;
- The form of the equations;
- Choice of parameter values (restrictions).

1. Programming models

Programming models have been used for a long time for policy analysis. The FAPRI model, which used to be partly maintained here at Iowa State University by CARD, is an example of programming model. The GTAP model is another example. In programming models, by *knowing* technology, you can make behavioral assumptions and get supply equations. For example, linear programming models:

- Assume linear additive technology;
- Assume factors are fixed or perfectly elastically supplied;
- Generate parametric results;
- Aggregate across firms.

Of course, many of these assumptions are relaxed in models that allow for nonlinear technology. Powerful computers, and robust numerical optimization methods, now allow for the use of more complex nonlinear models.

The strengths of programming models are:

- Comprehensive;
- Can do dynamics;
- Optimal output related to all factors and product prices;
- Spatial equilibrium;
- Can deal with radical changes.

I would add to that list that programming models allow to calculate the general equilibrium supply curve. That is, programming models allow to find the supply elasticity considering induced price changes in factors of production.

Linear programming model however have (had) poor predictive performance. Why?

- Aggregation;
- Profit maximization assumption;
- Exogenous prices.

Since Colman (1983), the predictive power of programming models has very much improved.

2. Two-stage procedures

The two-stage procedure consists in 1) estimating a production function, cost function or production function and 2) derive the output relationship. That approach is subject to the criticisms that we have seen in the section on estimating a production function.

3. Supply systems

Estimating a supply system is analogous to estimate a demand system like we previously covered. It yields consistent estimates of supply equations and input demand equations. Still estimates from this approach are sensitive to aggregation and the endogeneity of prices. Supply systems are derived from static theory and most often assume certainty.

4. Direct estimation of single commodity models

Most studies of supply response are limited to a single commodity where an *ad hoc* equation is specified. The role of economic theory is limited and the emphasis is put on dynamics, uncertainty or asymmetry of price response. An advantage of the direct estimation of a single commodity model is that it is simple. As production data on agricultural products (for example) are often only available yearly, estimation is sensitive to structural changes, multicollinearity and small number of degrees of freedom.

General observations

The objective of supply response analysis is always conditional projections or forecasts. Depending on the context, different procedures might make better conditional forecast. To be verifiable and useful, a forecast must be dated (time series structure). However, in policy analysis, time-dating may be of secondary importance.

2.3.2 Single-commodity supply response models

In estimating supply response we must take into account the dynamic of production. For example, for corn and soybeans planting decisions are made in middle-late winter when farmers contract crop insurance. But farmers are paid only when they deliver their harvest in the Fall. Thus, planting decision is not based on the price observed in the Fall but on the expected harvest price at the time of making planting decisions. That expected price may be the base price for crop insurance, the futures price or a forecast by the USDA.

Note that the price is not the only factor explaining supply response. In the production of corn, rotation with soybeans is an important element to the decision of which crop to plant in the next crop year (e.g. Hendricks et al., 2014b).

Consider the following supply response model

$$Q_t = a + bE_{t-1}P_t + \mathbf{cZ_t} + u_t,$$

where Q_t is the output, $E_{t-1}P_t$ is the expected price at $t-1$, $\mathbf{Z_t}$ is a vector of control variables. We do not observe farmer's expectations. We will rely on models of price expectations and we will cover the following models in what follows:

1. Right expectations;
2. Naive expectations;
3. Adaptive expectations;
4. Rational expectations;
5. Futures prices.

You can find a summary of supply response models in Nerlove and Bessler (2001).

2.3.2.1 Right expectations

In a right expectations model, the expected price is simply given by the current price

$$E_{t-1}P_t = P_t.$$

The right expectations model assumes unbiased expectations regarding prices. This is of course implausible. Producers do not know all the parameters of demand and supply and cannot perfectly forecast all shocks to the system. There is also an endogeneity problem in the supply response equation as the price is determined simultaneously with the quantity.

2.3.2.2 Naive expectations

In a naive expectations model, the expected price is given by the price observed in the previous period

$$E_{t-1}P_t = P_{t-1}.$$

That type of model is often used. The number of lags may be different than one depending on the type of product considered. The naive expectations model has the advantage of being simple and that the price in the previous period is exogenous.[6] Those reasons alone explain why the naive expectations model is often used. There are however some problems with the

[6] This is not always true. If there is serial correlation in the price, lag prices are correlated contemporaneously with the error term. Remember that endogeneity is a statistical problem and that an economic argument does not guarantee exogeneity.

naive expectations model. It is implausible for most applications because producers use more information than the price in the previous period to determine their production.

2.3.2.3 Adaptive expectations

The adaptive expectations model was proposed by Nerlove (1958) and is often used in the literature. The model considers that producers revise their notion of *normal price* in proportion to the difference between the then current price and their previous idea of a normal price:

$$E_{t-1}P_t - E_{t-2}P_{t-1} = \beta(P_{t-1} - E_{t-2}P_{t-1}).$$

This means that producers systematically make errors in predicting the price in the next period. It therefore relates to the cobweb model.

We must make a few algebraic manipulations for the model to be more instructive regarding how farmers form their expectations. After a few manipulations (Koych lag transformation) we find:

$$\begin{aligned} E_{t-1}P_t &= \beta P_{t-1} + (1-\beta)E_{t-2}P_{t-1}; \\ &= \beta \sum_{k=0}^{\infty}(1-\beta)^k P_{t-k-1}. \end{aligned}$$

Thus, the expected price in a given period depends on all the previously observed prices and a proportion factors that weighs past observations at a geometrically declining rate.

Using this expected price in a supply response model we can write

$$\begin{aligned} Q_t &= a + b\beta \sum_{k=0}^{\infty}(1-\beta)^k P_{t-k-1} + \mathbf{cZ_t} + u_t \\ &= a + b\beta P_{t-1} + b(1-\beta)\beta \sum_{k=0}^{\infty}(1-\beta)^k P_{t-k-2} + \mathbf{cZ_t} + u_t. \end{aligned} \tag{2.5}$$

Similarly, the output in the previous period is given by

$$Q_{t-1} = a + b\beta \sum_{k=0}^{\infty}(1-\beta)^k P_{t-k-2} + \mathbf{cZ_{t-1}} + u_{t-1}.$$

This last expression implies that we can write that

$$b\beta \sum_{k=0}^{\infty}(1-\beta)^k P_{t-k-2} = Q_{t-1} - a - \mathbf{cZ_{t-1}} - u_{t-1}.$$

Using this last expression into (2.5)

$$Q_t = a\beta + b\beta P_{t-1} + (1-\beta)Q_{t-1} + \mathbf{cZ_t} - (1-\beta)\mathbf{cZ_{t-1}} + u_t - (1-\beta)u_{t-1}.$$

Note the following about the adaptive expectations model:

- Implausible. One reason is that it does not allow for full adjustment to a shock;
- Lag structure consistent with other theories of price expectation;
- The same lag structure for $\mathbf{Z}$ and P may not be consistent with production process;
- The model specification implies serial correlation of the error term.

2.3.2.4 Futures prices

Gardner (1976) and others (see a review in Nerlove and Bessler (2001)) propose to use the futures price as the expected price. Futures contract specify the delivery at a future date of a standardized quantity and quality of a commodity at a specific location at a specified price. Futures contracts are tradeable on central exchanges where thousands of traders meet. Expectations using futures mean that

$$E_{t-1}P_t = F_{t-1}P_t.$$

The argument for using futures as price expectations is based on the hypothesis that markets are efficient and that futures prices represent the expected intersection of supply and demand at a given date. If markets are efficient, futures prices should contain all the current information about prices in the future. As such, futures are a way to obtain rational price expectations, which we cover next. Futures might not be an unbiased predictor of the price because of speculation, futures contract might not be available on the right commodity, at the right place and at the right time.

Are futures prices really unbiased measure of expected prices? Some claim that yield shocks are partially predictable such that they should be instrumented. See Hendricks et al. (2014a) for a recent contribution on the topic.

2.3.2.5 Rational expectations

The rational expectations model is due to Muth (1961). The model considers that the formation of expectations is an economic process where all information available about demand and supply is used to make an optimum economic decision. Consider the following model

$$\begin{aligned} E_{t-1}P_t &= E_{t-1}(P_t|\text{Supply,Demand}) \\ \text{Supply}: Q_t &= a + bE_{t-1}P_t + \mathbf{cZ_t} + u_t \\ \text{Demand}: P_t &= e - fQ_t + \mathbf{gI_t} + v_t, \end{aligned}$$

where $\mathbf{I_t}$ is vector of demand shifters. We assume that $u_t \sim (0, \sigma_u^2)$ and $v_t \sim (0, \sigma_v^2)$ are independent. Solving the model for P_t

$$P_t = e - f(a + bE_{t-1}P_t + \mathbf{cZ_t} + u_t) + \mathbf{gI_t} + v_t.$$

Taking expectations and solving

$$E_{t-1}P_t = \frac{1}{1+fb}(e - f(a + \mathbf{c}E_{t-1}\mathbf{Z_t} + E_{t-1}u_t) + \mathbf{g}E_{t-1}\mathbf{I_t} + E_{t-1}v_t).$$

Assuming no serial correlation for the error terms, we find that

$$\begin{aligned} E_{t-1}P_t &= \frac{1}{1+fb}(e - fa + f\mathbf{c}E_{t-1}\mathbf{Z_t} + \mathbf{g}E_{t-1}\mathbf{I_t}) \\ Q_t &= \frac{a+be}{1+fb} + \frac{b\mathbf{g}E_{t-1}\mathbf{I_t} - bf\mathbf{c}E_{t-1}\mathbf{Z_t}}{1+fb} + \mathbf{cZ_t} + u_t. \end{aligned}$$

Thus, the expected price depends on the expected value of the exogenous variables. Other specifications yield different expressions for the expected price and the output. For instance, one can assume serial correlations and other model of expectations for the demand and supply shifters.

2.3.3 Dynamic adjustment to price changes

We will see in this section models where supply responds gradually to a change in the expected price because of biological lags or maybe because production requires important capital investment. These models also attempt to explain why supply responds asymmetrically to increases and decreases in prices.

2.3.3.1 Partial adjustment model

The partial adjustment model is due to Nerlove (1958) and a voluminous literature uses that model. The model considers a desired quantity $\hat{Q}_t$ that depends on an expected price

$$\hat{Q}_t = a + bE_{t-1}P_t + \mathbf{cZ_t} + u_t. \tag{2.6}$$

However, the desired quantity is not the quantity observed, as, for example, capital may limit expansion of a firm in the short run. The model assumes that the actual quantity adjusts in proportion to the desired quantity

$$Q_t - Q_{t-1} = \gamma(\hat{Q}_t - Q_{t-1}),$$

or

$$Q_t = \gamma\hat{Q}_t + (1-\gamma)Q_{t-1}.$$

Using this expression in (2.6) yields

$$Q_t = \gamma a + \gamma bE_{t-1}P_t + \gamma\mathbf{cZ_t} + (1-\gamma)Q_{t-1} + \gamma u_t.$$

Model fit is typically very good with the dependent lag variable.

We can combine the partial adjustment model to a price expectation model. For instance, considering naive expectations in the partial adjustment model yields

$$Q_t = \gamma a + \gamma bP_{t-1} + \gamma\mathbf{cZ_t} + (1-\gamma)Q_{t-1} + \gamma u_t.$$

This is similar to the adaptive expectations model but differs with respect to the absence of dependent variables and the error structure. Combining the partial adjustment model to the adaptive expectation model yields Nerlove's model of supply (Nerlove and Bessler, 2001).

2.3.3.2 Perennial crops

Perennial crops like many types of fruits and nuts are only productive many years after plantation and produce for many years afterwards (more than 50 years for many varieties of grapes and olives). Dynamics of production and the expected returns in the long term are therefore a very important factor in plantation decisions. See for example French et al. (1985).

- Returns over many years motivate planting;
- Yields vary over age of trees;
- Production cycles.

2.3.3.3 Government programs

Government programs can affect price expectations and may constraint supply response (quota). Are the current programs going to be available in the future? Are future programs expected to be more or less generous? Do farmers respond contemporaneously to the announcement of a program change or do they anticipate the change or react only with lags (see e.g. Hendricks and Sumner, 2014)?

2.3.4 Other issues

- Risk;
- Storage;
- ...

2.3.5 Application of supply response model: Roberts and Schlenker (2013)

Roberts and Schlenker (2013) estimate world demand and supply elasticities for calories from wheat, rice, corn and soybeans. The authors instrument prices (futures) exploiting yield shocks and past weather shocks.

2.3.5.1 A model of Supply and Demand

Roberts and Schlenker (2013) aggregate wheat, rice, corn and soybeans using their calorie contents, thus aggregating over the most important attribute of food staples. An advantage of this aggregation is that prices for these commodities tend to vary together making identification of supply and demand parameters for individual commodities more difficult. As the four commodities are substitutes both in demand and production, aggregating them still characterizes food markets reasonably well.

Wheat, rice, corn and soybeans are storable commodities. One role of storage is to smooth prices such that current demand and supply conditions do not solely determine prices and quantities. Roberts and Schlenker (2013) show, but do not solve, a simple model of storage. Write that c_t is the consumption in the current period, z_t is the amount of food available at the beginning of the period and x_t is the amount of food stored for the next period. Thus, consumption is the disappearance of the quantity of food available during a period:

$$c_t = z_t - x_t.$$

Storage has convex cost $\phi(x_t)$ and consumption gives consumers utility $u(z_t - x_t)$. Production is stochastic as unknown weather shock w_{t+1} at time of planting affects production. Farmers can impact production through their effort λ_t which has convex cost $g(\lambda_t)$. Total production in the coming harvest is $\lambda_t w_{t+1}$.

The Bellman equation for social maximization problem is

$$\begin{aligned} \upsilon(z_t) = \max_{x_t \lambda_t} \{ & u(z_t - x_t) - \phi(x_t) - g(\lambda_t) + \delta E[\upsilon(z_{t+1})]\} \qquad \text{subject to} \qquad (2.7) \\ & z_{t+1} = x_t + \lambda_t w_{t+1}, \\ & x_t \geq 0, z_t - x_t \geq 0, \lambda_t \geq 0, \end{aligned}$$

where δ is a discount factor. Why consider the maximization of the total surplus? Remember that perfect competition maximizes the total surplus (first theorem of welfare economics). Thus, equilibrium quantities that maximizes social surplus are the same in the outcome of perfect competition. Equilibrium prices can be found along the supply or the demand and through intertemporal arbitrage. Roberts and Schlenker (2013) note that the most important point of the model is that exogenous shocks affect both current consumption and inventory adjustment.

The empirical model is

$$\text{Supply: } q_{st} = \alpha_s + \beta_s p_{st} + \gamma_s w_t + f_{s2}(t) + u_t \tag{2.8}$$

$$p_{st} = \delta_s + \mu_{s0} w_t + \mu_{s1} w_{t-1} + f_{s1}(t) + \epsilon_t \tag{2.9}$$

$$\text{Demand: } q_{dt} = \alpha_d + \beta_d p_{dt} + f_{d2}(t) + \upsilon_t \tag{2.10}$$

$$p_{dt} = \delta_d + \mu_{d0} w_t + f_{d1}(t) + \eta_t, \tag{2.11}$$

where $q_{st} = \ln(\lambda_{t-1} w_t)$ and $q_{dt} = \ln(\lambda_{t-1} w_t + x_{t-1} - x_t)$, which is the new production minus the change in inventory. In the supply equation, the price is measured as the log of future price $p_{st} = \ln(p_t|_{t-1})$ while in the demand equation the price is measured as the price during the month of delivery $p_{dt} = \ln(p_t)$. The functions $f()$ denote trend functions.

Both the supply and the demand price are instrumented. Typically, only the current price in the demand curve is instrumented while the price in the supply equation use a prediction such the futures price. Roberts and Schlenker (2013) argue that futures prices reflect both anticipated and unanticipated shocks. The anticipation of certain shocks causes endogeneity of futures prices in the supply function and thus should be instrumented.

The baseline model uses yield deviations ξ_{cit}, for crop c in country i. That is, the yield shock is the difference between trend yield and observed yield for each country and each crop as given by $y_{cit} = g_{cit}(t) + \xi_{cit}$ where the function $g()$ is specific to a crop c and a country i and depends on a time trend. The production shocks are calculated as the weighted average of all shocks for all crops and all countries. The weights ρ_{cit} are the predicted yields times the growing area times the caloric content. That is, the production shocks are weighted by the expected caloric output for each country and each crop. The production shocks are thus calculated as

$$w_t = \sum_c \sum_i \hat{\xi}_{cit} \rho_{cit}.$$

This instrumentation strategy to estimate supply of course assumes that yields are exogenous and that yield does not affect demand. The authors claim that this is possible because of inventory. Do farmers have incentives to increase yields when prices are high?

To address the possibility that yields are endogenous, Roberts and Schlenker (2013) use weather variables in another set of regressions. The average weather variables (average temperature and total precipitation) are aggregated using production quantities for all countries and crops.

2.3.5.2 The data

The data on production and storage are from the Food and Agricultural Organization (FAO) of the United Nations (http://faostat.fao.org/) for 1961-2007. The data include production, area harvested, yields and stock variation for the four crops. Consumption is calculated as production minus the net change in inventories. All quantities are converted in terms of their calorie content. Roberts and Schlenker (2013) do a sensitivity check using data from the Foreign Agricultural Service (FAS) of the United States Department of Agriculture (http://www.fas.usda.gov).

Weather data are from the Center for Climatic Research at the University of Delaware.

Futures prices are from the Chicago Board of Trade with a delivery month of December for corn and wheat and a delivery month of November for soybeans and rice.[7] The demand price

[7] You can obtain for free all futures from http://www.quandl.com/.

p_{dt} is the log of the average futures price during the month when delivery occurs. The supply price $p_{st} = p_t|_{t-1}$ is the log of the average futures price in December one year prior to delivery.[8] All prices are deflated by the Consumer Price Index. All prices are converted to the caloric equivalent with the world price of calorie equal to the production-weighted averages for the four commodities.

A second price series is built using prices received by US farmers, as available from the USDA.

2.3.5.3 Empirical results

Figure 2.6 summarizes the main results for different specifications of the models. Overall, elasticity estimates are not too sensitive to the model specifications. In table 2.6, the multiplier $\frac{1}{\widehat{\beta_s}-\widehat{\beta_d}}$ captures the change in the equilibrium price from a shift on the demand Δq. Thus, if US ethanol mandates increase demand for calories by 5%, this means an increase in the price of calories 30-35%.

Figure 2.7 reports results that use weather variables as instruments. Elasticity estimates are not too different than those in figure 2.7 but are less efficient than those that use yield shocks.

Roberts and Schlenker (2013) go through several robustness checks and explanations of their results. They find that supply response is explained at the extensive margin (i.e. increase in cropland) but that is a result that is debatable because of double-cropping. They find that the traditional method where prices are not instrumented yield a more inelastic supply and a more inelastic demand. The authors explain these results by `anticipated shocks` included in futures and lagged prices.

Roberts and Schlenker (2013) estimate supply and demand using prices observed in the United States (Chicago Board of Trade). Are these prices appropriate measures of prices everywhere else in the world? Agricultural markets are well integrated as arbitrageur exploit for profit price differences that exceed transportation costs. If this is true, then using prices in the United States is a valid approach. The authors find that a production shock in the United States has the same effects on prices as an equal size production shock elsewhere in the world. Thus, using US prices is sensible approach.

Further, Roberts and Schlenker (2013) show that their estimates of supply elasticity are actually for the short-run or the long-run. Adding two lags to the expected price, the authors find that the long-run supply is very similar to their first set of estimated.

The authors present several robustness check, including many in appendix. Roberts and Schlenker (2013) verify whether their results are sensitive to the aggregation of commodities. The reason for this robustness check is that the ethanol mandate is on ethanol from corn and the four commodities not necessarily perfect substitutes in demand and production. They create two commodities: corn and an aggregate of rice, wheat and soybeans. The supply elasticities are not too different than in the first set of estimates in tables 2.6 and 2.7. However, the demand elasticities for the disaggregated demand are more elastic. Accounting for cross-price elasticities, the multiplier is smaller but once weighting the estimator according to production, the average

[8] My understanding of this price is that it is the futures price as observed in the prior year. Using the notation earlier in this chapter, this would be $p_{st} = E_{t-1}P_t = F_{t-1}P_t$.

	Instrumental Variables			Three Stage Least Squares		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
	Panel A: Supply Equation					
Supply Elast. β_s	0.102***	0.096***	0.087***	0.116***	0.112***	0.097***
	(0.025)	(0.025)	(0.020)	(0.019)	(0.020)	(0.019)
Shock ω_t	1.184***	1.229***	1.211***	1.249***	1.279***	1.241***
	(0.146)	(0.138)	(0.105)	(0.111)	(0.101)	(0.091)
First Stage ω_{t-1}	-3.901***	-3.628***	-3.824***	-3.546***	-3.113***	-3.226***
	(1.145)	(0.945)	(0.910)	(0.800)	(0.704)	(0.731)
First Stage ω_t	-2.918*	-2.276*	-2.372*	-2.885***	-2.350***	-2.420***
	(1.647)	(1.294)	(1.279)	(0.967)	(0.815)	(0.819)
	Panel B: Demand Equation					
Demand Elast. β_d	-0.028	-0.055**	-0.054**	-0.034	-0.062***	-0.066***
	(0.021)	(0.024)	(0.022)	(0.023)	(0.022)	(0.021)
First Stage ω_t	-5.564***	-4.655***	-4.770***	-5.354***	-4.445***	-4.332***
	(1.489)	(1.300)	(1.249)	(1.384)	(1.210)	(1.186)
	Panel C: Effect of Demand Shift					
Multiplier $\frac{1}{\widehat{\beta}_s - \widehat{\beta}_d}$	7.73	6.63	7.06	6.65	5.75	6.12
Exp. Multiplier	8.39	7.08	7.42	6.90	5.92	6.31
(95% Conf. Int.)	(5.2,15.3)	(4.6,12.2)	(5.0,12.0)	(4.9,10.4)	(4.3,8.5)	(4.6,9.1)
$F_{1st\text{-}stage}$ Supply	11.61	14.73	17.66			
$F_{1st\text{-}stage}$ Demand	13.97	12.81	14.60			
Observations	46	46	46	46	46	46
Spline Knots	3	4	5	3	4	5

Notes: Tables show regression results for the supply and demand of calories. The first three columns (1a)-(1b) use instrumental variables, while columns (2a)-(2c) use three stages least squares. Columns (a), (b), and (c) include restricted cubic splines in time with 3, 4, and 5 knots, respectively. Panel A gives results for the supply equations (1) and (2), i.e., coefficients above the vertical line give the results for log quantity, while coefficients below the line give first stage results of log price. Similarly, panel B gives results for demand equations (3) and (4). Coefficients on time trends are suppressed. Panel C gives the effect of a demand shift on commodity prices: multipliers translate percentage changes in demand into percentage changes in equilibrium price. Stars indicate significance levels: *** : 1%; ** : 5%; * : 10%.

Figure 2.6: Supply and Demand Elasticity (FAO Data): table 1 in Roberts and Schlenker (2013)

	(1a) log Q	(1b) log P	(2a) log Q	(2b) log P	(3a) log Q	(3b) log P
			Panel A: Supply Equation			
Supply Elast. β_s	0.085*		0.089		0.091*	
	(0.048)		(0.055)		(0.053)	
Temperature T_{t-1}		3.252***		3.080***		3.050***
		(0.863)		(0.738)		(0.768)
Temperature T^2_{t-1}		-0.084***		-0.080***		-0.078***
		(0.022)		(0.019)		(0.019)
Precipitation P_{t-1}		2.937		2.597		3.174
		(2.249)		(1.923)		(2.097)
Precipitation P^2_{t-1}		0.130		-0.193		-0.497
		(1.072)		(0.922)		(0.982)
Temperature T_t	0.037	1.776**	-0.015	1.413**	-0.041	1.259*
	(0.184)	(0.750)	(0.201)	(0.669)	(0.195)	(0.691)
Temperature T^2_t	-0.002	-0.045**	-0.000	-0.035**	0.000	-0.031*
	(0.005)	(0.019)	(0.005)	(0.017)	(0.005)	(0.018)
Precipitation P_t	0.547	2.274	0.498	1.952	0.610*	2.847
	(0.332)	(1.805)	(0.363)	(1.611)	(0.371)	(1.742)
Precipitation P^2_t	-0.426***	-0.694	-0.431***	-0.952	-0.464***	-1.233
	(0.143)	(0.829)	(0.156)	(0.737)	(0.157)	(0.763)
			Panel B: Demand Equation			
Demand Elast. β_d	-0.014		-0.056**		-0.047*	
	(0.025)		(0.028)		(0.025)	
Temperature T_t		1.282		1.099		1.305
		(1.209)		(1.003)		(1.038)
Temperature T^2_t		-0.034		-0.030		-0.035
		(0.030)		(0.025)		(0.026)
Precipitation P_t		1.991		1.636		0.759
		(3.227)		(2.675)		(2.853)
Precipitation P^2_t		1.075		0.705		1.015
		(1.460)		(1.211)		(1.266)
			Panel C: Effect of Demand Shift			
Multiplier $\frac{1}{\beta_s-\beta_d}$	10.07		6.91		7.25	
Exp. Multiplier	14.13		9.35		7.71	
(95% Conf. Int.)	(4.8,52.9)		(3.8,25.7)		(4.0,27.4)	
Observations	46	46	46	46	46	46
Spline Knots	3	3	4	4	5	5

Notes: Table replicates the three-stage least square results in Table 1 except that prices are instrumented with weather (quadratic in average temperature and precipitation) instead of yield shocks. Column pairs (a) and (b) present results from one joint regression, where columns (a) give results for the quantity regression, and columns (b) the results for the price regressions. Columns (1a-b), (2a-b), and (3a-b) include restricted cubic splines in time with 3, 4, and 5 knots, respectively. Stars indicate significance levels: *** : 1%; ** : 5%; * : 10%.

Figure 2.7: Supply and Demand Elasticity - Weather as Instrument (FAO Data): table 2 in Roberts and Schlenker (2013)

multiplier is comparable to the pool estimates. The production-weighted is also similar to the pooled estimate when disaggregating over the four crops. Estimates from disaggregated data show that the effect of the mandate on ethanol is twice as large on the price of corn as in the overall price increase for calories.

2.3.5.4 Shortcomings?

Of course, like any empirical study, the work by Roberts and Schlenker (2013) is subject to criticisms. Believing econometric results requires making a certain leap of faith. Depending on whether the shortcomings are judged too significant, the econometric analysis might then be judged not credible. Here are a few possible shortcomings:

- Instrumentation of prices: are there other exogenous variables that explain production? What about technological changes? What does the time trend capture?
- Why instrument futures prices? What is the meaning of an anticipated shock?
- How can w_t be used to instrument futures price given that w_t should happen after $p_{t|t-1} = F_{t-1}P_t$?
- There are only 46 annual observations;
- Seasonality: the timing of production is different in the southern and northern hemisphere;
- Are US prices appropriate measures of prices all across the world (robustness check for this);
- Corn and soybeans are grown largely as feeds for animals. Is it correct to aggregate these commodities with rice and wheat and that are mostly grown as human food?
- FAO data - not always reliable.
- Identification depends only on a few data points as the trends capture almost everything.

2.4 Competitive storage model

Storage plays an important role in commodity markets. In any given year, 10 to 15 percent of the corn crop is still in storage for the next crop year. The stock to supply ratio tends to be smaller for soybeans.

The role of commodity storage is to stabilize markets. Without storage, market prices depend only on the size of the current crop. Storage allows to carry volumes from one year to the next and hence smooth prices over time. The economic of commodity storage has long been a topic of interest to economists (e.g. Working, 1948, 1949; Gustafson, 1958; Johnson and Sumner, 1976). It is really in the 1980's and 1990's that the economic theory of commodity storage was developped (e.g. Newberry and Stiglitz, 1981; Wright and Williams, 1982; Scheinkman and Schechtman, 1983; Wright and Williams, 1984, 1988; Deaton and Laroque, 1992, 1996). We will cover the competitive storage model, like in most studies, assuming that storers form rationale expectations about the return of their activities.

In explaining the competitive storage model, I will follow Wright and Williams (1982), Miranda (1997) and Gouel (2013). The role of a competitive storer is to store a quantity s_t from period t to period $t+1$. The competitive storer purchases the commodity at equilibrium price p_t and stores it at a cost $c(s_t)$. At period $t+1$, the revenue of the storer is $p_{t+1}s_t(1-\delta)/(1+r)$, where δ is shrinkage and r is the interest rate. At time t, the storer does not know the price in period $t+1$ and hence the storer must form expectations about what that price will be the next year.

We will assume in what follows that the demand is deterministic but that supply is stochastic.[9] Hence, the expected price depends on distribution of shocks on the supply and the resulting supply and demand equilibria. We denote acreage as A_t and stochastic yield as y_t. Production at time t is the product of acreage at time $t-1$, because production decisions are made in the previous period, and yield realization in the current period such that $q_t = A_{t-1} * y_t$.

Write that the total quantity available at time t is $x_t = A_{t-1}y_t + (1-\delta)s_{t-1}$. That is, the total quantity available equals the production in the current period and the stock that carries over from the last period. Total disappearance is the sum of the quantity stored into the next period and consumption denoted as z_t. At equilibrium, total disappearance equals the total quantity available such that we can write

$$s_t + z_t = A_{t-1}y_t + (1-\delta)s_{t-1}.$$

Alternatively, we can write that consumption equals production in the current period plus the change in storage:

$$z_t = A_{t-1}y_t + (1-\delta)s_{t-1} - s_t.$$

A storer profits by storing a quantity if it expects to receive a higher price in the next period. We will consider rational expectation under which storers form expectations for the price at $t+1$ utilising all the information available. We cannot solve the rational expectation model because this is not a typical problem where it is possible to solve for the model's variables. It is not possible to solve for all possible future contingencies and obtain consistent expectations. This is a nontrivial functional equation problem where we will solve for a price function.[10]

The expected profit of a storer is

$$E\Pi_{t+1} = \left(\frac{1-\delta}{1+r}E_t p_{t+1} - p_t\right) s_t - c(s_t),$$

where $c(s_t)$ is the cost of storage for one period. The first order condition for profit maximization by the storer is

$$\frac{1-\delta}{1+r}E_t p_{t+1} - p_t - c'(s_t) = 0,$$

for $s_t > 0$. Depending on the functional form for the marginal cost of storage $c'(s_t)$, $s_t = 0$, i.e. stockout situation, is a possibility, in which case

$$\frac{1-\delta}{1+r}E_t p_{t+1} - p_t - c'(s_t) < 0.$$

We will focus on the case where $s_t > 0$ as we will assume that $c'(s_t)$ takes a very large negative value for s_t small. The negative cost is a convenience yield for storing a quantity greater than zero. Convenience yields are observed in practice and there are several explanations for its existence. Note that $s_T \geq 0$ because it is not possible to bring some quantities from period $t+1$ to period t.

At period t, given that the supply is pre-determined by the acreage decision in the previous period, A_{t-1}, and the exogenous yield realization, y_t, the supply is perfectly inelastic, the price

[9] We could consider that the demand is also stochastic. Having a single source of stochasticity simplifies finding a solution to the model as otherwise we would have to take a double integral.

[10] It is possible in formulate the model considering storage as the response variable (see Miranda and Fackler, 2002, pages 215-216).

is determined by the demand curve. Write the inverse demand as $P_t(z_t)$. Note that z_t is variable determined contemporaneously and that is will be more useful to write that the inverse demand as a function of the total quantity available. Thus, we write that $P_t(z_t) = P_t(x_t - s_t)$.

The competitive storage model summarizes to three equations:

$$p_t = P_t(x_t - s_t); \tag{2.12}$$

$$\frac{1-\delta}{1+r} E_t p_{t+1} - p_t - c'(s_t) = 0; \tag{2.13}$$

$$A_t = f\left(\frac{1}{1+r} E_t p_{t+1} y_{t+1}\right). \tag{2.14}$$

Equation (2.12) says that contemporaneously the equilibrium price equals the price along the demand curve. Equation (2.13) is the first order condition for profit maximization by a storer and equation (2.14) is the supply curve for acreage as a function of the expected revenue per acre.

Scheinkman and Schechtman (1983) and Deaton and Laroque (1992) show that with a perfectly inelastic stochastic supply, a continuous and strictly decreasing demand and for $(1-\delta)/(1+r) < 1$, there is a unique stationary rational expectations equilibrium. That is, there is a function that describes the current price p_t as a function of the total quantity available x_t. The function p_t is defined as

$$p_t = \max[\frac{(1-\delta)}{1+r} E_t p_{t+1} - c'(s_t), P(x_t)],$$

where recall that $P()$ is the inverse demand. There might be a proof for the case where the supply is not perfectly inelastic but I am not aware of it. My guess is that if the supply function is continuous and strictly increasing, then there exists an equilibrium to the competitive storage model.

To model the expected price under rational expectations, let us write a function $\lambda()$ which produces the equilibrium price. Given this, we can write under rational expectation that

$$E_t p_{t+1} = E_y \lambda(A_t y + (1-\delta) s_t); \tag{2.15}$$

$$E_t p_{t+1} y_{t+1} = E_y \lambda(A_t y + (1-\delta) s_t) y. \tag{2.16}$$

Equation (2.15) is the expected price paid to a storer. Equation (2.16) is the expected revenue per acre to a producer. With rational expectation, the equations of the model become:

$$\lambda(x_t) = P_t(x_t - s_t); \tag{2.17}$$

$$\frac{1-\delta}{1+r} E_y \lambda(A_t y + (1-\delta) s_t) - P_t(x_t - s_t) - c'(s_t) = 0; \tag{2.18}$$

$$A_t = f\left(\frac{1}{1+r} E_y \lambda(A_t y + (1-\delta) s_t) y\right). \tag{2.19}$$

A solution to the competitive storage model yields for every realizable supply x_t a solution for $\lambda()$, s_t and A_t. Note that $\lambda()$ is a function, thus making this a functional equation problem. We will see below how to solve the competitive storage model using collocation.

Figure 2.8 shows the solution for the competitive storage model. The top panel compares prices with and without storage. If the total quantity available (x_t) is small, then prices with and without storage are nearly identical. This is because, as the second panel shows, that

the quantity stored is very close to zero. As the total quantity available increases, storage increases, which causes the gap between the prices with and without storage to increase. That is because the current price decreases, hence making the prospect of selling some quantity the next period comparatively more profitable. Figure 2.8 shows that storage does not increase the price when the total quantity available is small, but does increase the price when the total quantity available is large.

The third panel of figure 2.8 shows acreage decision. This is for the case where the supply is not deterministic and responds to the expected price. As the total supply increases, producers respond by decreasing acreage. When storage is near zero, acreage tends toward is maximum value as the expected price for the next period depends almost only on the expected production in the next period.

2.4.1 Solving the competitive storage model using collocation

We have not discussed how to solve the competitive storage. As mentioned above, the competitive storage model with rational expectations is a functional equation problem whose solution takes the form of a function rather than finite-sized vector of prices and quantities (Miranda, 1997). If the model is non-linear, there is no closed form solution and the model can only be solved numerically. There are different methods that can be used to solve the model. Miranda (1997) and Gouel (2013) compare different solution methods. In this section, we will see how to solve the competitive storage model using collocation method.

Miranda (1997) describes numerical solution strategies for the competitive storage model. The solution method involves numerically approximating (using a computer) the infinite-dimensional functional problem into a finite dimensional problem. More specifically, we will use a function to approximate $\lambda()$. We will evaluate that function at multiple values for the state variable x_t over an interval that covers the range of values that x_t can take. We will find the coefficients for that function such that the system of equations in 2.17 to 2.19 hold at each of selected values for x_t.

The first step is to specify a function to approximate $\lambda()$. We will do this by using a linear combination of basis functions $\phi_0, \phi_1, ..., \phi_n$:

$$\lambda(x) \approx \sum_{j=0}^{n} c_j \phi_j(x),$$

where c_j are the coefficients that we must estimate. Equations (2.17)-(2.19) are required to hold at selected number of collocation nodes $x_0, x_1, ..., x_n$. The collocation nodes must cover the range of possible values that the total supply can take and they do not need to be equidistant.

Finding the expected price requires numerically integrating. We can perform this by Monte-Carlo simulations but this would be computationally intensive. An alternative is to use Gaussian quadrature. The continuous yield distribution is approximated by using a finite number of points. The integration routine assigns weights $w_1, w_2, ..., w_m$ to yield values $y_1, y_2, ..., y_m$. The discrete values for the yield do not need to be equidistant and must be chosen strategically to obtain an efficient and unbiased estimate of the integral.

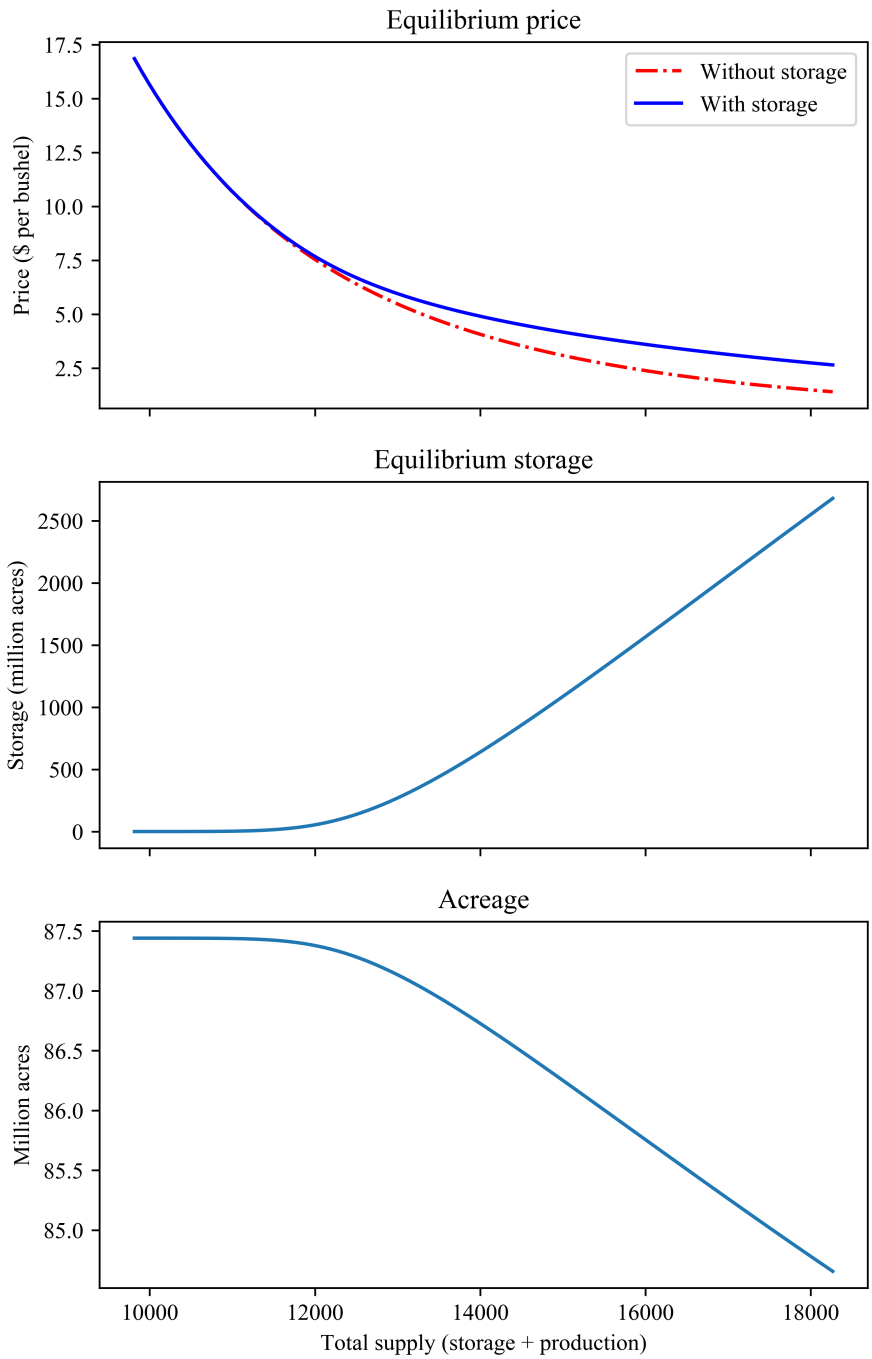


Figure 2.8: Solutions to the competitive storage model

Given these steps, we can thus write the model as

$$\sum_{j=0}^{n} c_j \phi_j(x_i) = P_t(x_i - s_t); \tag{2.20}$$

$$\frac{1-\delta}{1+r} \sum_{k=1}^{m} \sum_{j=0}^{n} w_k c_j \phi_j((1-\delta)s_t + A_t y_k) - P_t(x_i - s_t) - c'(s_t) = 0; \tag{2.21}$$

$$A_t = f(\frac{1-\delta}{1+r} \sum_{k=1}^{m} \sum_{j=0}^{n} w_k c_j \phi_j((1-\delta)s_t + A_t y_k) y_k). \tag{2.22}$$

The model is evaluated at all $n+1$ collocation nodes and for each of them the optimal solutions for s and A are found. This means that in total we must solve for $3(n+1)$ coefficients from $3(n+1)$ equations. The discretization of the model means that instead of having to solve for a function and two variables, the model reduces to solving for $3(n+1)$ coefficients.

We must specify the form of the basis function. Here, I cover approximation using Chebyshev polynomials, which usually perform quite well. An alternative is using (cubic) splines interpolation. In storage models without convenience yield, the price function has a kink because of the non-negativity constraint on storage, making the function continuous but not differentiable everywhere. Miranda (1997) shows that cubic splines interpolation is preferable in that particular case because the discontinuity affects the interpolation only locally while for Chebyshez polynomials the discontinuity affects the entire interpolation interval (Miranda and Fackler, 2002; Gouel, 2013).

The Chebyshez collocation nodes are specified as $x_j = 0.5(\overline{x} + \underline{x}) + 0.5(\overline{x} - \underline{x}) \cos(\frac{n-j+0.5}{n}\pi)$ for $j = 1, ..., n$ (Miranda and Fackler, 2002, p. 119), where $\overline{x}$ and $\underline{x}$ are the upper and lower bounds of the interval for x. The nodes are not evenly spaced and are more concentrated toward the boundaries of the interval. Before introducing the nodes in the Chebushev polynomial, they must be normalized on the $[-1, 1]$ interval such that $z_j = 2(x_j - \underline{x})/(\overline{x} - \underline{x}) - 1$. Chebyshev polynomials specify the basis functions using a recurrence relation

$$\begin{aligned} T_0(z) &= 1; \\ T_1(z) &= z; \\ T_{n+1}(z) &= 2zT_n(z) - T_{n-1}(z). \end{aligned}$$

Alternatively, the Chebyshez polynomial can be specified using a trigonometric definition $T_n(z) = \cos(n \arccos(z))$.

We will use the algorithm from Miranda and Glauber (1995) and Miranda (1997) but note that there are other algorithms that might work better in certain cases (Gouel, 2013). The algorithm from Miranda (1997) specifies the following steps:

1. Select the degree of approximation n, the form of the basis functions, the collocation nodes and guess values for $c_0, c_1, ..., c_n$.
2. Given the values for $c_0, c_1, ..., c_n$, find for each $i = 0, 1, ..., n$ the optimal values for s_i and A_i from the equations (2.21) and (2.22).
3. Given the solutions for s_i and A_i, update the values for $c_0, c_1, ..., c_n$ solving the system of n equations from (2.20).
4. Repeat steps 2 and 3 until the values for $c_0, c_1, ..., c_n$ converge.

2.5 Livestock response model

Section 2.3.3 discusses dynamic adjustments to prices. One such case that has gathered quite a bit of attention is the livestock sector where supply response occurs according to biological lags. For cattle, it takes between 18 and 24 months between birth and slaughter. Any significant increase in production takes however more time as it requires increasing the size of the breeding herd, which means sending fewer cattle to slaughter in the short run.

Cattle have two functions to producers. First, a cattle is a consumption good and therefore an output to a farm. Second, a cattle is also capital because the current stock of cattle is used to produce future cattle (Jarvis, 1974). Breeding animals come at the cost of marketing animals in the current period.

Figure 2.9 shows the January inventory of cattle in the United States. Notice the recurring cycles in the inventory. These cycles are due to the lags between the time production decisions take place and the marketing of cattle from biological lags in production.

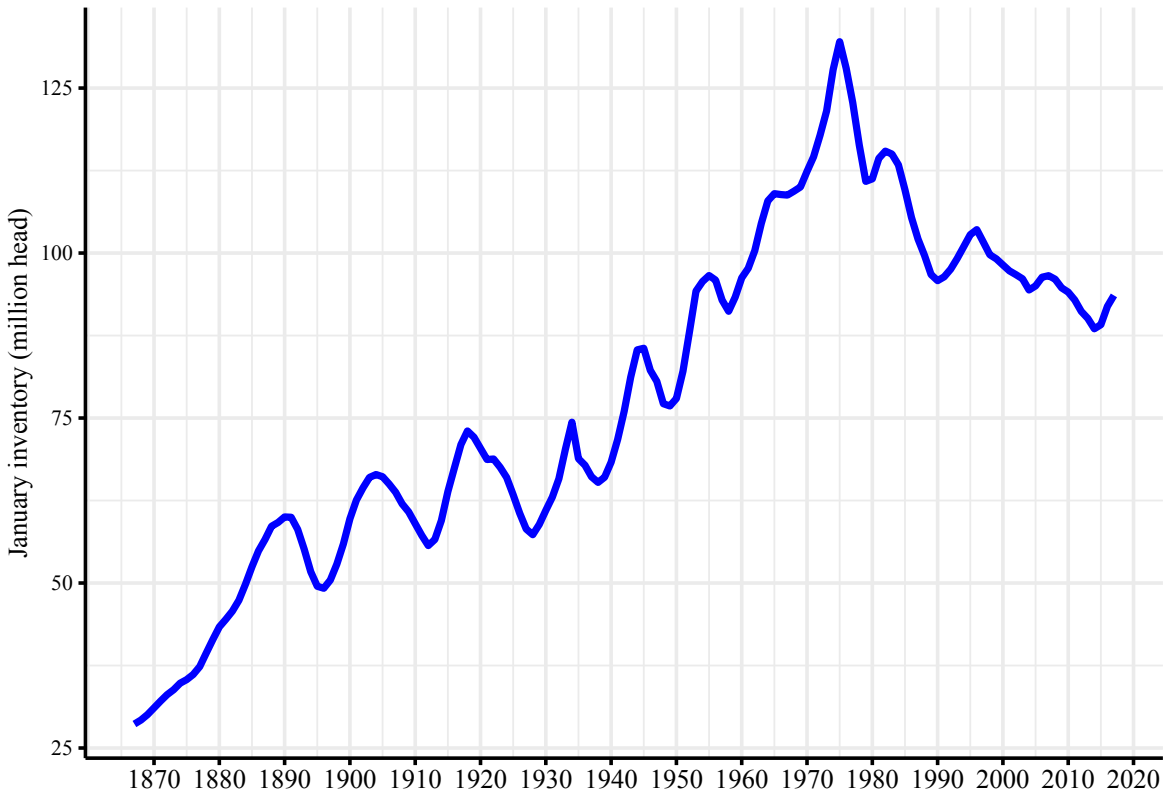


Figure 2.9: January cattle inventory (including calves) in the United States

Note: The data were obtained from USDA - National Agricultural Statistics Service (2013).

There are several models to explain cycles in livestock production. Jarvis (1974) was the first to model livestock as capital. Below, we will see the main elements of Rosen et al. (1994) model.

The model of Rosen et al. (1994) assumes:

- 1-year gestation-birth;
- 2-year maturation;
- Steers are treated as heifers sent to slaughter;
- Adults are homogeneous, undifferentiated, independent of age or prior fertility;
- Quality invariant of age.

This is of course a simplified exposition of the dynamics of cattle production but it is reasonable and it captures essential features. Figure 2.10 shows the population dynamics.

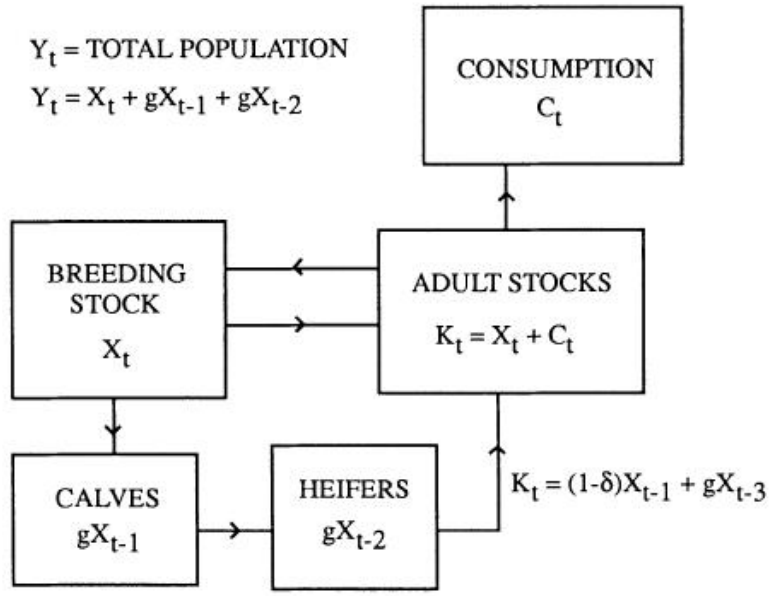


FIG. 2.—Population dynamics

Figure 2.10: Cattle population dynamics: figure 2 in Rosen et al. (1994)

The first two equations of the model count the total number of heads and the number of animals for breeding:

$$\text{Total head count: } y_t = x_t + gx_{t-1} + gx_{t-2};$$
$$\text{Breeding stock: } x_t = (1-\delta)x_{t-1} + gx_{t-3} - c_t,$$

where $g < 1$ is the breeding rate and δ is the death rate.

Production decisions are based on the return of holding a cattle for one more year vs. sending that cattle for slaughter in the current year. Rosen et al. (1994) write that the cost of holding a cattle for one more year is

$$z_t = h_t + \beta g \gamma_0 h_{t+1} + \beta^2 g \gamma_1 h_{t+2},$$

where h_t is a unit holding cost, β is a discount factor and γ_0 and γ_1 are proportion factors for the cost of holding calve (1-year old) and steers (2-year old). Thus the cost of holding one head of cattle includes the cost of the unit of cattle plus the cost of holding it's expected progeniture for two years. On the revenue side, let p_t be the price of one cattle and m_t be the feeding costs for that cattle (finishing to make the cattle ready for slaughter). Thus, we can write that $q_t = p_t - m_t$ is the net return from selling an animal.

At equilibrium, the value of slaughter equals the value of breeding

$$q_t = E_t[\beta(1-\delta)q_{t+1} + g\beta^3 q_{t+3} - z_t].$$

where $g\beta^3 q_{t+3}$ is the return from the calve born during the year that the cattle is held off the market.

Let the demand be given by

$$c_t = \alpha_0 - \alpha p_t + d_t,$$

where d_t is a demand shifter. With a few manipulations, the system reduces to two equations:

$$x_t - (1-\delta)x_{t-1} - gx_{t-3} = -\alpha_0 + \alpha p_t - d_t,$$

which link the current consumption to past stocks of cattle, and

$$E_t[q_t - \beta(1-\delta)q_{t+1} - g\beta^3 q_{t+3} + h_t + \beta g\gamma_0 h_{t+1} + \beta^2 g\gamma_1 h_{t+2}] = 0,$$

which describes the intertemporal arbitrage relationship.

Rosen et al. (1994) then go on solving the model assuming that h_t and m_t are exogenous, that q_t is independent of x_t, consumption is demand determined at price p_t and that the stochastic variables of the model d_t, m_t and h_t each follow an AR(1) process. We will not solve the model here. On page 476, the authors make the following observation:

> Cycles arise from demographic *echo effects* because the current age distribution constrains future reproductive capacity. Demand and supply shocks cause ranchers to change their breeding stock inventories, and these decisions percolate through all future birth cohorts. The percolations oscillate unless initial stocks are accidentally lined up just right.

2.6 Appendix: Review of production theory

The following is a brief review of production theory for competitive firms. For thorough treatments of production theory I refer you to Varian (1992), Mas-Colell et al. (1995) or Jehle and Reny (2001). You will find the review in this section useful for empirical models of production and in many topics that we will address later in this course.

2.6.1 Production function

A production function describes the technology that combines inputs into outputs. Figure 2.11 shows an example of production function where an input x is defined as a positive quantity and the output is given by $y = f(x)$. Note that Mas-Colell et al. (1995) define input as a negative quantity while output is defined as a positive quantity.

Properties:

- Continuous;
- Strictly increasing;
- Strictly quasiconcave.

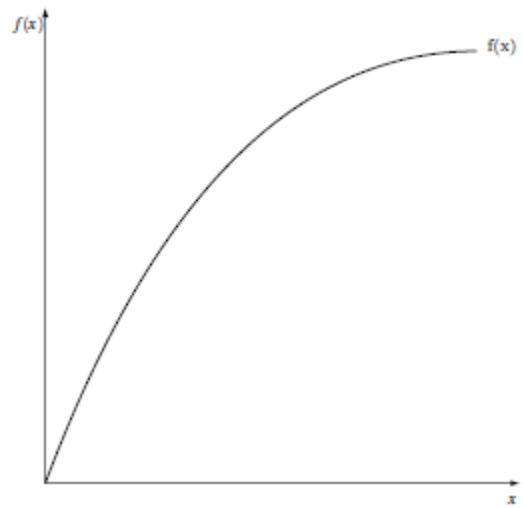


Figure 2.11: Example of production function

2.6.2 Marginal rate of technical substitution (MRTS)

The MRTS is the trade-off between two inputs such that production remains constant. Figure 2.12 shows that the MRTS is given by the slope of an isoquant.

$$MRTS_{ij} = -\frac{\partial f(\mathbf{x})/\partial x_i}{\partial f(\mathbf{x})/\partial x_j}. \tag{2.23}$$

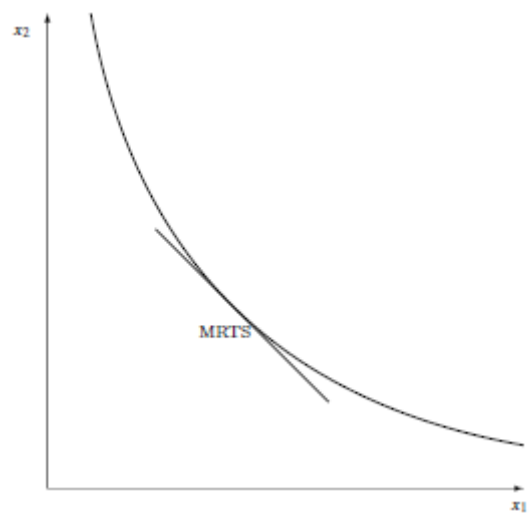
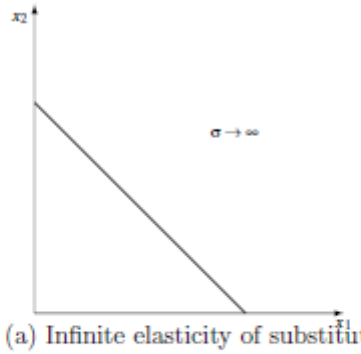


Figure 2.12: Marginal rate of technical substitution

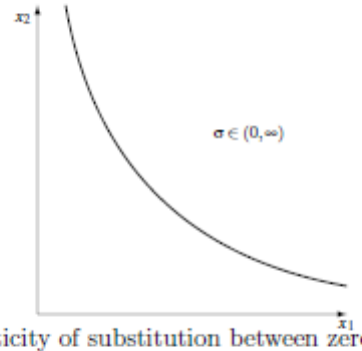
2.6.3 Elasticity of substitution

The elasticity of substitution (or the Allen elasticity of substitution) represents the ease with which it is possible to substitute an input with another while keeping the production constant. Figure 2.13 shows isoquants for production function of different shapes and their elasticity of substitution. The expression for the elasticity of substitution is given by

$$\sigma_{ij} = \frac{d\ln\left(\frac{x_j}{x_i}\right)}{d\ln\left(\frac{f_i}{f_j}\right)} = \frac{d\left(\frac{x_j}{x_i}\right)}{\frac{x_j}{x_i}} \frac{\frac{f_i}{f_j}}{d\left(\frac{f_i}{f_j}\right)}. \tag{2.24}$$



(a) Infinite elasticity of substitution



(b) Elasticity of substitution between zero and infinity

(c) Elasticity of substitution equal to zero

Figure 2.13: Elasticity of substitution for production functions of different shapes

2.6.4 Return to scale

Return to scale tells what happens to the output when the same increase in percentage is applied to all inputs. That is, does the output double when all inputs are doubled?

- Constant return to scale: $f(t\mathbf{x}) = tf(\mathbf{x})$ where $t > 0$;
- Increasing return to scale: $f(t\mathbf{x}) > tf(\mathbf{x})$ where $t > 1$;
- Decreasing return to scale: $f(t\mathbf{x}) < tf(\mathbf{x})$ where $t > 1$.

2.6.5 Profit maximization - single output

Consider a vector of n inputs given by $\mathbf{x}$ and a corresponding vector of prices given by $\mathbf{w}$. Let the production function be given by $y = f(\mathbf{x})$. A firm maximizes its profit by choosing a set of inputs:

$$\arg\max_{\mathbf{x}} pf(\mathbf{x}) - \mathbf{w} \cdot \mathbf{x},$$

where p is the price of the output. The first order condition for an interior solution for an input i is given by

$$p\frac{\partial f(\mathbf{x})}{\partial x_i} = w_i.$$

That is, a firm maximizes its profit by equating the marginal product of x_i, $p\frac{\partial f(\mathbf{x})}{\partial x_i}$, to its wage w_i. The arguments that maximize the profit function yield a vector of input demand $\mathbf{x}(p, \mathbf{w})$ and the profit maximizing output $y(p, \mathbf{w}) = f(\mathbf{x}(p, \mathbf{w}))$.

The input demand $\mathbf{x}(p, \mathbf{w})$ is homogeneous of degree zero in $\mathbf{w}$. The substitution matrix

$$\sigma^*(p, \mathbf{w}) = \begin{pmatrix} \frac{\partial x_1}{\partial w_1} & \cdots & \frac{\partial x_1}{\partial w_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial w_1} & \cdots & \frac{\partial x_n}{\partial w_n} \end{pmatrix}.$$

is symmetric and semi negative definite.

2.6.6 Profit function

The profit function of a firm is given by

$$\Pi(p, \mathbf{w}) = \max_{\mathbf{x}} pf(\mathbf{x}) - \mathbf{w} \cdot \mathbf{x}.$$

The profit function has the following properties:

i. Increasing in p;
ii. Decreasing in $\mathbf{w}$;
iii. Homogeneous of degree one in p and $\mathbf{w}$.

Also, a useful equality is given by Hotelling's lemma

$$\frac{\partial \Pi(p, \mathbf{w})}{\partial p} = y(p, \mathbf{w}) \text{ and } -\frac{\partial \Pi(p, \mathbf{w})}{\partial w_i} = x_i(p, \mathbf{w}). \tag{2.25}$$

2.6.7 Cost minimization - dual problem

Equivalent to the profit maximization problem, we can consider that a firm minimizes its cost to produce a given output y. A firm minimizes its cost by choosing a set of inputs:

$$\arg\min_{\mathbf{x}} \mathbf{w} \cdot \mathbf{x} \text{ s.t. } f(\mathbf{x}) \geq y.$$

The first order condition for interior solution, letting λ be the Lagrangian multiplier, is given by

$$w_i = \lambda f_i(\mathbf{x}).$$

For two inputs i and j, we can write that

$$\frac{w_i}{w_j} = \frac{\frac{\partial f(\mathbf{x})}{\partial x_i}}{\frac{\partial f(\mathbf{x})}{\partial x_j}}.$$

That is, a firm maximizes its profit when the ratio of wages equals $MRTS_{ij}$.

2.6.8 Cost function

The cost function of a firm is given by

$$C(\mathbf{w}, y) = \min_{\mathbf{x}} \mathbf{w} \cdot \mathbf{x} \text{ s.t. } f(\mathbf{x}) \geq y.$$

The cost function has the following properties:

i. Increasing in $\mathbf{w}$;
ii. Homogeneous of degree one in $\mathbf{w}$;
iii. Concave in $\mathbf{w}$;
iv. Strictly increasing in y;
v. Convex in y.

In addition, Shephard's lemma is very useful

$$\frac{\partial C(\mathbf{w}, y)}{\partial w_i} = x_i(\mathbf{w}, y). \tag{2.26}$$

2.6.9 Profit maximization - single output (2)

Using the cost function, we can write that the profit function is given by

$$\Pi(p, \mathbf{w}) = \max_y py - C(\mathbf{w}, y).$$

The first order condition for profit maximization is

$$p = \frac{\partial C(\mathbf{w}, y)}{\partial y} = mc(\mathbf{w}, y),$$

where $mc(\mathbf{w}, y)$ is the marginal cost. The maximization problem yields the profit maximizing output $y(p, \mathbf{w})$. Free entry implies that the profit of competitive firms equals zero such that

$$py(p, \mathbf{w}) = C(\mathbf{w}, y).$$

We can write that

$$p = \frac{C(\mathbf{w}, y)}{y} = AC(\mathbf{w}, y),$$

where $AC(\mathbf{w}, y)$ is the average cost.

2.6.10 Economies of scale

The concept of economies of scale tells us whether the cost of producing one more unit of output is larger than the average cost of infra-marginal units. Thus, it says whether the average cost increases or decreases when one more unit of output is produced. The concept of economies of scale is related but different to the concept of return to scale.

- Economies of scale: $\frac{AC}{mc} > 1$;
- Constant economies of scale: $\frac{AC}{mc} = 1$;
- Diseconomies of scale: $\frac{AC}{mc} < 1$.

The definitions above apply for the production of a single output. For the case of multiple outputs, we must first define the incremental cost (IC) (Featherstone and Moss, 1994):

$$IC_i = C(\mathbf{w}, \mathbf{y}) - C(\mathbf{w}, \mathbf{y}_{m-i}),$$

where $\mathbf{y}_{m-i} = (y_1, ..., y_{i-1}, 0, y_{i+1}, ..., y_m)$. Thus, for multiple output, the product specific economies of scale is defined as

$$S_i(\mathbf{y}) = \frac{IC_i}{y_i} \Big/ \frac{\partial C(\mathbf{w}, \mathbf{y})}{\partial y_i}. \tag{2.27}$$

We observe product-specific economies of scale if $S_i(\mathbf{y}) > 1$, constant economies of scale if $S_i(\mathbf{y}) = 1$ and diseconomies of scale if $S_i(\mathbf{y}) < 1$.

2.6.11 Economies of scope

Economies of scope tells us whether it is cheaper to produce multiple products into a single firm or whether it is cheaper to produce the differing products in individual firms. For two outputs y_1 and y_2, we observe economies of scope if

$$C(\mathbf{w}, y_1, y_2) < C(\mathbf{w}, y_1, 0) + C(\mathbf{w}, 0, y_2).$$

More generally, we can define to following measure

$$SC_m(y_1, y_2) = \frac{C(\mathbf{w}, y_1, 0) + C(\mathbf{w}, 0, y_2) - C(\mathbf{w}, y_1, y_2)}{C(\mathbf{w}, y_1, y_2)}.$$

We observe economies of scope if $SC_m(y_1, y_2) > 0$. We can also define a similar measure for more than two outputs.

2.6.12 Economies of scale - multiple outputs

We used expression (2.27) to define product-specific economies of scale. Here, we use the concepts of economies of scope and product-specific economies of scale to define a measure of economies of scale for an individual firm. For the case of two outputs, lets define economies of scale as in Featherstone and Moss (1994):

$$S_m(\mathbf{y}) = \frac{\alpha_1 S_1(\mathbf{y}) + (1 - \alpha_1) S_2(\mathbf{y})}{1 - SC_m(\mathbf{y})},$$

where $\alpha_1 = y_1 \frac{\partial C(\mathbf{w},\mathbf{y})}{\partial y_1} \Big/ \left(y_1 \frac{\partial C(\mathbf{w},\mathbf{y})}{\partial y_1} + y_2 \frac{\partial C(\mathbf{w},\mathbf{y})}{\partial y_2} \right)$. We observe economies of scale if $S_m(\mathbf{y}) > 1$, constant economies of scale if $S_m(\mathbf{y}) = 1$ and diseconomies of scale if $S_m(\mathbf{y}) < 1$. The definition for the case of more than two outputs is analogous to the one for one output.

Bibliography

Abramovitz, M. (1956). Resources and output trends in the United States since 1870. *American Economic Review*, 46:5–23.

Amsler, C., Prokhorov, A., and Schmidt, P. (2016). Endogeneity in stochastic frontier models. *Journal of Econometrics*, 190(2):280 – 288. Endogeneity Problems in Econometrics.

Barnett, W. A. and Lee, Y. W. (1985). The global properties of the miniflex laurent, generalized leontief, and translog flexible functional forms. *Econometrica*, 53(6):1421–37.

Battese, G. and Coelli, T. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of productivity analysis*, 3(1):153–169.

Beattie, B. R., Taylor, C. R., and Watts, M. J. (2009). *The Economics of Production*. Krieger, 2nd edition.

Berndt, E. R. and Khaled, M. S. (1979). Parametric productivity measurement and choice among flexible functional forms. *The Journal of Political Economy*, 87(6):1220–1245.

Cassels, J. M. (1933). The nature of statistical supply curves. *Journal of Farm Economics*, 15(2):378–387.

Chambers, R. G. (1988). *Applied Production Analysis: A Dual Approach*. Cambridge University Press.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1971). Conjugate duality and the transcendental logarithmic functions. *Econometrica*, 39:255–256.

Coelli, T. and Battese, G. (1996). Identification of factors which influence the technical inefficiency of Indian farmers. *Australian Journal of Agricultural and Resource Economics*, 40(2):103–128.

Colman, D. (1983). A review of the arts of supply response analysis. *Review of Marketing and Agricultural Economics*, 51(03).

Deaton, A. and Laroque, G. (1992). On the behaviour of commodity prices. *The Review of Economic Studies*, 59(1):1–23.

Deaton, A. and Laroque, G. (1996). Competitive storage and commodity price dynamics. *Journal of Political Economy*, 104(5):896–923.

Diewert, W. E. (1971). An application of the shephard duality theorem: A generalized leontief production function. *Journal of Political Economy*, 79(3):481–507.

Diewert, W. E. (1973). Separability and a generalization of the cobb-douglas cost, production and indirect utility functions. Technical report, Research Branch, Department of Manpower and Immigration, Ottawa, Canada. Available at: http://faculty.arts.ubc.ca/ediewert/cobb.pdf.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, 4(2):115–145.

Diewert, W. E. and Wales, T. J. (1987). Flexible functional forms and global curvature conditions. *Econometrica*, 55(1):43–68.

Diewert, W. E. and Wales, T. J. (1988). A normalized quadratic semiflexible functional form. *Journal of Econometrics*, 37(3):327–342.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3):253–290.

Featherstone, A. M. and Moss, C. B. (1994). Measuring economies of scale and scope in agricultural banking. *American Journal of Agricultural Economics*, 76(3):655–661.

French, B. C., King, G. A., and Minami, D. D. (1985). Planting and removal relationships for perennial crops: An application to cling peaches. *American Journal of Agricultural Economics*, 67(2):215–223.

Fuss, M. and McFadden, D. L., editors (1978a). *Production Economics: A Dual Approach to Theory and Applications - Volume I: The Theory of Production*. North-Holland. Available online at: http://elsa.berkeley.edu/users/mcfadden/prodecon1.html.

Fuss, M. and McFadden, D. L., editors (1978b). *Production Economics: A Dual Approach to Theory and Applications -Volume II: Applications of the Theory of Production*. North-Holland. Available online at: http://elsa.berkeley.edu/users/mcfadden/prodecon2.html.

Gallant, A. R. (1981). On the bias in flexible functional forms and an essentially unbiased form : The fourier flexible form. *Journal of Econometrics*, 15(2):211–245.

Gallant, A. R. and Golub, G. H. (1984). Imposing curvature restrictions on flexible functional forms. *Journal of Econometrics*, 26(3):295–321.

Gardner, B. L. (1976). Futures prices in supply analysis. *American Journal of Agricultural Economics*, 58(1):81–84.

Gouel, C. (2013). Comparing numerical methods for solving the competitive storage model. *Computational Economics*, pages 1–29.

Griliches, Z. (1963). The sources of measured productivity growth: United States agriculture, 1940-60. *The Journal of Political Economy*, 71(4):331–346.

Guan, Z., Kumbhakar, S. C., Myers, R. J., and Lansink, A. O. (2009). Measuring excess capital capacity in agricultural production. *American Journal of Agricultural Economics*, 91(3):765–776.

Gustafson, R. L. (1958). Carryover levels for grains.

Henderson, H. (2015). Considering technical and allocative efficiency in the inverse farm size-productivity relationship. *Journal of Agricultural Economics*, 66(2):442–469.

Hendricks, N. P., Janzen, J. P., and Smith, A. (2014a). Futures prices in supply analysis: Are instrumental variables necessary? *American Journal of Agricultural Economics*.

Hendricks, N. P., Smith, A., and Sumner, D. A. (2014b). Crop supply dynamics and the illusion of partial adjustment. *American Journal of Agricultural Economics*.

Hendricks, N. P. and Sumner, D. A. (2014). The effects of policy expectations on crop supply, with an application to base updating. *American Journal of Agricultural Economics*, 96(3):903–923.

Henningsen, A. (2018). *Introduction to Econometric Production Analysis with R*. Leanpub. Available at https://leanpub.com/ProdEconR/.

Jarvis, L. S. (1974). Cattle as capital goods and ranchers as portfolio managers: An application to the argentine cattle sector. *The Journal of Political Economy*, 82(3):489–520.

Jehle, G. A. and Reny, P. J. (2001). *Advanced Microeconomic Theory*. Addison-Wesley.

Johnson, D. G. and Sumner, D. (1976). Optimization approach to grain reserves for developing countries. Technical report, ERS US Department of Agricultural.

Just, R. E. (1993). Discovering production and supply relationships: Present status and future opportunities. *Review of Marketing and Agricultural Economics*, 61(01).

Koopmans, T. C., editor (1951). *Activity Analysis of Production and Allocation*. John Wiley & Sons, Inc. Available at http://cowles.econ.yale.edu/P/cm/m13/.

Kopp, R. J. and Diewert, W. E. (1982). The decomposition of frontier cost function deviations into measures of technical and allocative efficiency. *Journal of Econometrics*, 19(2-3):319–331.

Latruffe, L., Bravo-Ureta, B. E., Carpentier, A., Desjeux, Y., and Moreira, V. H. (2017). Subsidies and technical efficiency in agriculture: Evidence from european dairy farms. *American Journal of Agricultural Economics*, 99(3):783–799.

Lau, L. J. (1976). A characterization of the normalized restricted profit function. *Journal of Economic Theory*, 12(1):131–163.

Lau, L. J. (1978). *Production Economics: A Dual Approach to Theory and Applications - Volume I: The Theory of Production*, chapter A.4: Testing and imposing monoticity, convexity and quasi-convexity constraints, pages 409–453. North-Holland. Available at: http://elsa.berkeley.edu/users/mcfadden/prodecon1.html.

Magnus, J. R. (1979). Substitution between energy and non-energy inputs in the Netherlands 1950-1976. *International Economic Review*, 20(2):465–484.

Mas-Colell, A., Whinston, M., and Green, J. (1995). *Microeconomic theory*. Oxford University Press New York.

Miranda, M. J. (1997). Numerical strategies for solving the nonlinear rational expectations commodity market model. *Computational Economics*, 11(1-2):71–87.

Miranda, M. J. and Fackler, P. L. (2002). *Applied computational economics and finance*. MIT press.

Miranda, M. J. and Glauber, J. W. (1995). Solving stochastic models of competitive storage and trade by chebychev collocation methods. *Agricultural and Resource Economics Review*, 24(1):70–77.

Morrison-Paul, C. J. (2001a). Cost economies and market power: The case of the u.s. meat packing industry. *The Review of Economics and Statistics*, 83(3):531–540.

Morrison-Paul, C. J. (2001b). Market and cost structure in the u.s. beef packing industry: A plant-level analysis. *American Journal of Agricultural Economics*, 83:64–76.

Mundlack, Y. (2001). *Handbook of Agricultural Economics*, volume 1, chapter 1. Production and Supply, pages 3–85. Elsevier, Amsterdam.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29(3):315–335.

Nerlove, M. (1958). Adaptive expectations and cobweb phenomena. *The Quarterly Journal of Economics*, 72(2):227–240.

Nerlove, M. and Bessler, D. A. (2001). Expectations, information and dynamics. In Gardner, B. G. and Rausser, G., editors, *Handbook of Agricultural Economics*, volume 1, chapter 3, pages 155–206. Elsevier Science B.V.

Newberry, D. and Stiglitz, J. (1981). *The theory of commodity price stabilization*.

Ollinger, M., MacDonald, J. M., and Madison, M. (2005). Technological change and economies of scale in u.s. poultry processing. *American Journal of Agricultural Economics*, 87(1):116–129.

Pollak, R. A., Sickles, R. C., and Wales, T. J. (1984). The ces-translog: Specification and estimation of a new cost function. *The Review of Economics and Statistics*, 66(4):602–607.

Roberts, M. J. and Schlenker, W. (2013). Identifying supply and demand elasticities of agricultural commodities: Implications for the us ethanol mandate. *American Economic Review*, 103(6):2265–2295.

Rosen, S., Murphy, K. M., and Scheinkman, J. A. (1994). Cattle cycles. *The Journal of Political Economy*, 102(3):468–492.

Ryan, D. L. and Wales, T. J. (2000). Imposing local concavity in the translog and generalized leontief cost functions. *Economics Letters*, 67(3):253–260.

Scheinkman, J. A. and Schechtman, J. (1983). A simple competitive model with production and storage. *The Review of Economic Studies*, 50(3):427–441.

Schultz, T. W. (1956). Reflections on agricultural production, output and supply. *Journal of Farm Economics*, 38(3):748–762.

Shee, A. and Stefanou, S. E. (2014). Endogeneity corrected stochastic production frontier and technical efficiency. *American Journal of Agricultural Economics*.

Shephard, R. W. (1953). *Cost and production functions*. Princeton University Press.

USDA - National Agricultural Statistics Service (2013). Quick stats. Available online at: http://quickstats.nass.usda.gov/.

Varian, H. R. (1992). *Microeconomic Analysis*. W. W. Norton & Company, 3rd edition.

Weninger, Q. (2003). Estimating multiproduct costs when some outputs are not produced. *Empirical Economics*, 28:753–765.

Working, H. (1948). Theory of the inverse carrying charge in futures markets. *Journal of Farm Economics*, 30(1):1–28.

Working, H. (1949). The theory of price of storage. *The American Economic Review*, 39(6):1254–1262.

Wright, B. D. and Williams, J. C. (1982). The economic role of commodity storage. *The Economic Journal*, 92(367):596–614.

Wright, B. D. and Williams, J. C. (1984). The welfare effects of the introduction of storage. *The Quarterly Journal of Economics*, 99(1):169–192.

Wright, B. D. and Williams, J. C. (1988). The incidence of market-stabilising price support schemes. *The Economic Journal*, 98(393):1183–1198.

Zellner, A., Kmenta, J., and Dreze, J. (1966). Specification and estimation of cobb-douglas production function models. *Econometrica: Journal of the Econometric Society*, pages 784–795.