



COLLEGE OF
AGRICULTURE &
NATURAL RESOURCES

DEPARTMENT OF AGRICULTURAL AND
RESOURCE ECONOMICS

WP 16-02

Thinking About Agricultural Productivity Accounting in the Presence of By-Products

Robert G. Chambers

University of Maryland

Department of Agricultural and Resource Economics

Copyright © 2016 by Robert G. Chambers

All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Thinking About Agricultural Productivity Accounting in the Presence of By-Products

Robert G Chambers¹

December 30, 2016

¹Agricultural and Resource Economics, University of Maryland and School of Economics, University of Queensland. An earlier version of the paper was presented at the *OECD Expert Workshop: Measuring Environmentally-Adjusted TFP for Agriculture*, OECD, Paris, France, December 14-15, 2015. I would like to thank the program participants, particularly Sean Cahill, for suggestions that helped improve this paper.

Ever since Solow (1957) published his landmark study of technical change, the craft of calculating total factor productivity (TFP) has been evolving and changing to accommodate our growing understanding of technical progress. Just before Solow’s work appeared, Abramovitz (1956) had famously declared measured productivity growth to be “...a measure of our ignorance”. Abramovitz’s remark reflected his response to the “lopsided importance” that his calculations gave to productivity growth as a determinant of measured output growth.

The effort to “eliminate” the Solow residual in US agriculture had already started. Schultz (1956), attributing the concept to Zvi Griliches, suggested that an *ideal input-output formula* would have the input-output ratio stay close to one over time. If achieved that would suggest that a complete economic explanation of output growth had been accomplished. Following this line of thought, Griliches (1960, 1963) argued that much of the Solow residual appearing in early TFP calculations for US agriculture (Barton and Cooper 1948, Barton 1958, Barton and Durost 1960) could be traced to two sources: mismeasurement of agricultural inputs (in particular labor and capital) and the exploitation of economies of scale.

When Solow, Schultz, and Griliches were writing, the US “farm problem” was overproduction and chronically depressed prices. A key to its solution seemed to lay in eliminating this overproduction. That required understanding its source, but productivity calculations at that time suggested that agricultural input use had remained remarkably stable since before the First World War even as output had greatly expanded (Barton and Cooper 1948). The Solow residual, identified almost by default as technical progress, was the then unexplained source of expanding agricultural output.

Despite constant methodological refinements, the Solow residual in US agriculture remains robustly persistent. Calculated US agricultural TFP has grown at an average annual rate of slightly less than 1.5% over the postwar period. That growth remains the primary driver of US agricultural output growth as aggregate US agricultural input use has now remained remarkably stable, even declining at times, for the last 100 years (Barton and Cooper 1948, Ball, Wang, and Nehring 2015).¹ This growth pattern distinguishes US agriculture from most other industries where output growth is primarily driven by input growth.

Current agricultural productivity calculations do not consider the effect that by-products (typically bad outputs) have upon measured TFP. This is an important omission. By-products are an inherent aspect of many, if not most, agricultural technologies. Moreover, as agricultural production practices have evolved, they have become increasingly reliant upon material inputs that are likely associated with by-products (Wang, Heisey, Schimmelpfennig, and Ball 2015). One cannot imagine confinement agriculture without manure generation, odor contamination, and potential runoff any more than one can imagine row agriculture without nutrient and pesticide runoff polluting surrounding water systems. Ignoring by-production can distort measures of production-system performance, and that bias seems likely to grow as agriculture uses more and more material inputs that generate by-products. How by-production affects measured agricultural TFP is an empirical issue whose correct resolution requires an economic model of by-production that is both consistent with the physical laws of production and empirically tractable.

This paper considers the role that the presence of by-products plays in agricultural productivity. It starts with a simple model of the technology and a functional representation of that technology. Although simple, the model is general. Initially, only enough structure is imposed to ensure the existence of a function representation of the technology. The basic problem of productivity measurement is then spelt out. Familiar productivity-accounting methods for measuring TFP in the presence of by-production are quickly reviewed.

¹Although aggregate input use has changed remarkably little, the input mix has changed significantly shifting away from labor and land towards increased use of intermediate inputs (Wang, Heisey, Schimmelpfennig, Ball 2015).

This is done for both discrete and continuous approaches. An obvious result emerges. If the technology exhibits constant returns to scale, an easy correction for the presence of a by-product exists. The pragmatic issues associated with making that correction are then considered and certain difficulties with current imputation procedures are articulated. The next section discusses how recent developments in the theoretical literature on the representation and characterization of technologies involving by-products impinge upon measuring TFP. An illustration of how these developments manifest themselves in an empirical framework is then considered, and the paper then closes with some brief, general remarks.

1 The Model

1.1 The Technology and Optimal Behavior

The technology is defined at time t by

$$T(t) = \{(x, r, y, z) : (x, r) \text{ can produce } (y, z) \text{ at time } t\},$$

where $x \in \mathbb{R}_+^N$ denotes a vector of inputs that do not directly produce unintended by-products, $r \in \mathbb{R}_+^J$ denotes a vector of inputs that may have by-products associated with them, $y \in \mathbb{R}_+^M$ is a vector of good outputs, and $z \in \mathbb{R}_+^K$ is a vector of by-products.

Our primary technical assumptions on $T(t)$ are that it be closed, nonempty, and that

$$(x, r, y, z) \in T(t) \implies (\mu x, \mu r, y, z) \in T(t)$$

for $\mu > 1$ (weak input disposability). Associated with the technology is the input distance function²

$$I(x, r, y, z, t) = \sup \left\{ \lambda > 0 : \left(\frac{x}{\lambda}, \frac{r}{\lambda}, y, z, t \right) \in T(t) \right\}. \quad (1)$$

I is positively linearly homogeneous in x and r and provides a convenient function representation of T in the sense that

$$T(t) = \{(x, r, y, z, t) : I(x, r, y, z, t) \geq 1\}.$$

Output prices are denoted by $p \in \mathbb{R}_{++}^M$, the prices of the inputs not associated with a by-product are denoted by $w \in \mathbb{R}_{++}^N$ and the prices of the inputs with by-products are denoted by $v \in \mathbb{R}_{++}^K$. The cost function for the technology is defined

$$C(w, v, y, z, t) = \min_{x, r} \{wx + vr : I(x, r, y, z, t) \geq 1\}$$

if there exists (x, r) such that $I(x, r, y, z, t) \geq 1$ and ∞ otherwise. The variable profit function dual to $C(w, v, y, z, t)$ is defined

$$\pi(p, w, v, z, t) = \max_y \{py - C(w, v, y, z, t)\}$$

In what follows, for notational convenience, we treat x, r, y , and z as though they are scalars and typically assume (except where specifically noted) that both I and C are smoothly differentiable in all arguments.

²For the sake of conciseness, I rely solely upon input distance functions and input-based productivity indexes. It is straightforward to rework the same arguments to fit output distance functions and output-based productivity measures.

We can always justify the former by operating in terms of aggregate inputs and output, and the latter by practicality.³ The cost-minimization problem in Lagrangean form is

$$C(w, v, y, z, t) = \min_{x, r} \{wx + vr - \mu I(x, r, y, z, t)\},$$

where $\mu \geq 0$ is a Lagrange multiplier. The first-order conditions for x , r , and μ are

$$\begin{aligned} w - \mu \frac{\partial I}{\partial x} &= 0, \\ v - \mu \frac{\partial I}{\partial r} &= 0, \\ I(x, r, y, z, t) &= 1 \end{aligned} \tag{2}$$

while the envelope theorem implies (at the optimum).

$$\begin{aligned} \frac{\partial C}{\partial y} &= -\mu \frac{\partial I}{\partial y}, \text{ and} \\ \frac{\partial C}{\partial z} &= -\mu \frac{\partial I}{\partial z}. \end{aligned} \tag{3}$$

The first-order conditions imply

$$\begin{aligned} wx + rv &= \mu \left(\frac{\partial I}{\partial x} x + \frac{\partial I}{\partial r} r \right) \\ &= \mu I(x, r, y, z, t) \\ &= \mu, \end{aligned}$$

where the second equality follows from the positive linear homogeneity of I in (x, r) . Hence, at the optimum, $\mu = C(w, v, y, z, t)$.

The first-order condition for variable profit maximization requires

$$p - \frac{\partial C}{\partial y} = 0,$$

while the envelope theorem and (3) imply (at the optimum)

$$\begin{aligned} \frac{\partial \pi}{\partial z} &= -\frac{\partial C}{\partial z} \\ &= C \frac{\partial I}{\partial z}. \end{aligned} \tag{4}$$

$T(t)$ exhibits constant returns to scale in (x, r, y, z) if

$$(x, r, y, z) \in T(t) \implies (\mu x, \mu r, \mu y, \mu z) \in T(t),$$

for all $\mu > 0$. Hence,

$$\begin{aligned} I(\mu x, \mu r, \mu y, \mu z, t) &= \sup \left\{ \lambda > 0 : \left(\frac{\mu x}{\lambda}, \frac{\mu r}{\lambda}, \mu y, \mu z, t \right) \in T(t) \right\} \\ &= \sup \left\{ \lambda > 0 : \left(\frac{x}{\lambda}, \frac{r}{\lambda}, y, z, t \right) \in T(t) \right\} \\ &= I(x, r, y, z, t). \end{aligned}$$

³The main places where restricting attention to scalars will require qualifying arguments occurs when we invoke specific assumptions on returns to scale. It is also straightforward to rework all of the calculus-based arguments made here in terms of superdifferentials and subdifferentials in the nonsmooth case.

$T(t)$ exhibits constant returns to scale in (x, r, y) if

$$(x, r, y, z) \in T(t) \implies (\mu x, \mu r, \mu y, z) \in T(t)$$

for $\mu > 0$, and

$$I(\mu x, \mu r, \mu y, z, t) = I(x, r, y, z, t).$$

Constant returns in (x, r, y, z) implies

$$\begin{aligned} C(w, v, \mu y, \mu z, t) &= \min \{wx + vr : I(x, r, \mu y, \mu z, t) \geq 1\} \\ &= \mu \min \left\{ w \frac{x}{\mu} + v \frac{r}{\mu} : I\left(\frac{x}{\mu}, \frac{r}{\mu}, y, z, t\right) \geq 1 \right\} \\ &= \mu C(w, v, y, z, t), \quad \mu > 0, \end{aligned}$$

whence for $z > 0$

$$\begin{aligned} \pi(p, w, v, z, t) &= \max \{py - C(w, v, y, z, t)\} \\ &= z \max \left\{ pz - C\left(w, v, \frac{y}{z}, 1, t\right) \right\} \\ &= z\pi(p, w, v, 1, t). \end{aligned} \tag{5}$$

Constant returns in (x, r, y) implies for $y > 0$

$$\begin{aligned} C(w, v, y, z, t) &= \min \{wx + vr : I(x, r, y, z, t) \geq 1\} \\ &= y \min \left\{ w \frac{x}{y} + v \frac{r}{y} : I\left(\frac{x}{y}, \frac{r}{y}, 1, z, t\right) \geq 1 \right\} \\ &= yC(w, v, 1, z, t). \end{aligned}$$

1.2 The Basic Problem

The task is to explain output growth. That requires accounting for output growth attributable to input growth and to technical change. Both output growth and input growth are observable. Technical change, however, is not directly observable. And thus, if one is truly to be capable of explaining output growth, one must disentangle its input growth and technical change components. Thus, while the task is to explain output growth, the practical problem to be solved is to account for technical change. The practical approach to solving this problem is to account exhaustively for both input and output growth.

Technical change is associated analytically with how changes in t affect production possibilities holding the input-output bundle constant. In the continuous case, this effect is measured by

$$\frac{\partial \ln I(x, r, y, z, t)}{\partial t}. \tag{6}$$

In the discrete case, the analogous index measure is

$$\frac{I(x, r, y, z, t_1)}{I(x, r, y, z, t_0)}. \tag{7}$$

Technical progress occurs when the first measure is positive and the second is greater than one when $t_1 > t_0$. Intuitively, production isoquants shift towards the origin if there is technical progress. More formally, if $t_1 > t_0 \implies I(x, r, y, z, t_1) \geq I(x, r, y, z, t_0)$, then $T(t_0) \subset T(t_1)$. Technical progress makes a greater array of production outcomes technically feasible.

Our specific targets in what follows are either (6) or (7). The practical problem is that neither I nor C are directly observable. Hence, inferences must be made from observed production outcomes. That is, inferences must be made from observed prices and (x_t, r_t, y_t, z_t, t) and $(\dot{x}_t, \dot{r}_t, \dot{y}_t, \dot{z}_t, t)$ in the continuous case or $(x_1, r_1, y_1, z_1, t_1)$ and $(x_0, r_0, y_0, z_0, t_0)$ in the discrete case, where dots over variables denote percentage changes over time. There are several routes one can take.

One is to construct empirical approximations to $I(x, r, y, z, t)$, $C(w, v, y, z, t)$, or $\pi(p, w, v, z, t)$. Once done, either technical-change measure can be computed from the estimated structure. Another is to use productivity-accounting methods where possible, while relying upon further empirical work to accommodate needed extensions. By and large, most sectoral and national efforts at productivity measurement use the productivity-accounting approach. On the other hand, many firm-level efforts at productivity measurement use the structural approach. I will look at each in turn.

2 Productivity Accounting with By-Products

2.1 Continuous Case

Assuming technical efficiency, $I(x, r, y, z, t) = 1$, differentiation with respect to t establishes

$$\frac{\partial \ln I}{\partial t} + \frac{\partial \ln I}{\partial \ln x} \dot{x}_t + \frac{\partial \ln I}{\partial \ln r} \dot{r}_t + \frac{\partial \ln I}{\partial \ln y} \dot{y}_t + \frac{\partial \ln I}{\partial \ln z} \dot{z}_t = 0.$$

Thus,

$$\frac{\partial \ln I}{\partial t} = - \left(\frac{\partial \ln I}{\partial \ln x} \dot{x}_t + \frac{\partial \ln I}{\partial \ln r} \dot{r}_t + \frac{\partial \ln I}{\partial \ln y} \dot{y}_t + \frac{\partial \ln I}{\partial \ln z} \dot{z}_t \right). \quad (8)$$

Using (2) and (3) in (8) gives:

$$\frac{\partial \ln I}{\partial t} = \overbrace{\frac{\partial \ln C}{\partial \ln y} \dot{y}_t - \frac{wx}{C} \dot{x}_t - \frac{vr}{C} \dot{r}_t}^{TFP} + \frac{\partial \ln C}{\partial \ln z} \dot{z}_t. \quad (9)$$

The term falling under the upper brace in (9), labelled TFP , is the traditional measure of TFP growth, which roughly defined is aggregate (good) output growth less input growth. Under the assumption of constant returns to scale in (x, r, y) , the traditionally measured inputs and outputs, $\frac{\partial \ln C}{\partial \ln y} = 1$, and the bracketed term reduces to

$$TFP = \dot{y}_t - \frac{wx}{C} \dot{x}_t - \frac{vr}{C} \dot{r}_t$$

which is the continuous version of the Solow residual. If constant returns in (x, r, y) is implausible, the traditional correction is to replace the “elasticity of size”, $\frac{\partial \ln C}{\partial \ln y}$, with $\frac{py}{C}$, whence

$$\frac{\partial \ln I}{\partial t} = \overbrace{\frac{py}{C} \dot{y}_t - \frac{wx}{C} \dot{x}_t - \frac{vr}{C} \dot{r}_t}^{TFP} + \frac{\partial \ln C}{\partial \ln z} \dot{z}_t.$$

One excellent reason to question the plausibility of constant returns in (x, r, y) is to recall the traditional *physical* argument for it.⁴ It runs something like: If you can do something once you should be able to do it twice. And if you can do it twice, you should be able to do it thrice and so on. If one can truly measure all inputs and outputs (that means accounting for all possible differentials including quality, time, space, state

⁴I consider the economic argument below.

of Nature, etc.), that reasoning seems quite convincing when applied to all inputs and outputs, but not so appealing when only applied to (x, r, y) .

Recall that constant returns in (x, r, y) requires that

$$(x, r, y, z) \in T(t) \implies (\mu x, \mu r, \mu y, z) \in T(t), \quad \mu > 0.$$

A simple example illustrates why this is problematic. Consider a chicken-egg farmer producing one desirable output (eggs) using variable inputs (hens, shelter, feed) while generating one by-product (manure). Thus, y now represents eggs, x represents hens and coops, r represents feed, and z is manure. If this farmer produced 1500 eggs using 3000 hens, appropriate feed and shelter for 3000 hens, while generating a half ton of manure, constant returns in (x, r, y) would imply that he or she could also produce 3000 eggs using 6000 hens, appropriate feed and shelter for 6000 hens, *and a half ton of manure*. In other words, twice as many hens being fed the same amount produce exactly the same manure as the original 3000. On average, constant returns requires each hen to generate half as much manure as before. Implausible.

On the other hand, if $T(t)$ exhibits constant returns in (x, r, y, z) , a similar quandary does not arise because manure is allowed to increase proportionately with hens, feed, and shelter. Moreover, constant returns in (x, r, y, z) affords a straightforward means of developing a shadow price for z . By expression (5), the firm's variable profit, which is its economic rent for z , decomposes as

$$\pi(p, w, v, z, t) = z\pi(p, w, v, 1, t).$$

$\pi(p, w, v, 1, t)$ is the marginal rent to z and thus its proper shadow price. And if the firm operates efficiently

$$\begin{aligned} \pi(p, w, v, z, t) &= z\pi(p, w, v, 1, t) \\ &= py - C(w, v, y, z, t), \end{aligned}$$

whence

$$\pi(p, w, v, 1, t) = \frac{py - C(w, v, y, z, t)}{z}, \quad (10)$$

so that the shadow price for pollution can be imputed by dividing observed profit by z .

Expressions (3) and (4) now imply

$$\frac{\partial \ln I}{\partial z} = \frac{\pi(p, w, v, 1, t)}{C}, \quad (11)$$

which yields:

$$\frac{\partial \ln I}{\partial t} = \overbrace{\frac{py}{C} \dot{y}_t - \frac{wx}{C} \dot{x}_t - \frac{vr}{C} \dot{r}_t}^{TFP} + \frac{C - py}{C} \dot{z}_t$$

By-products carry economic rent. Thus, a profit-based decomposition of input and output growth based on only (x, r, y) misses a primary source of intertemporal profit growth, changes in rent associated with changing z . Correcting entails weighting \dot{z} by the negative of its marginal rent.⁵

2.2 Discrete Case

The task is now to convert observations on $(x_1, r_1, y_1, z_1, t_1)$ and $(x_0, r_0, y_0, z_0, t_0)$ into a measure of technical change. Following Caves, Christensen, and Diewert (1982), define an input-oriented productivity index as

⁵Here the specific argument hinges upon z being a scalar. However, it is straightforward though computationally burdensome in the current framework to extend the argument to the case where z is multidimensional.

the geometric average of two Malmquist input-oriented productivity indexes. One taken relative to the technology t_1 and the other relative to the technology at t_0

$$P_{t_1, t_0}(x_0, r_0, y_0, z_0, x_1, r_1, y_1, z_1) \equiv \left(\frac{I(x_0, r_0, y_0, z_0, t_0) I(x_0, r_0, y_0, z_0, t_1)}{I(x_1, r_1, y_1, z_1, t_0) I(x_1, r_1, y_1, z_1, t_1)} \right)^{\frac{1}{2}}.$$

This index decomposes as

$$P_{t_1, t_0}(x_0, r_0, y_0, z_0, x_1, r_1, y_1, z_1) = \frac{I(x_0, r_0, y_0, z_0, t_0)}{I(x_1, r_1, y_1, z_1, t_1)} \left(\frac{I(x_0, r_0, y_0, z_0, t_1) I(x_1, r_1, y_1, z_1, t_1)}{I(x_0, r_0, y_0, z_0, t_0) I(x_1, r_1, y_1, z_1, t_0)} \right)^{\frac{1}{2}}.$$

The first component, $\frac{I(x_0, r_0, y_0, z_0, t_0)}{I(x_1, r_1, y_1, z_1, t_1)}$, measures efficiency in period t_0 relative to efficiency in period t_1 . The second component,

$$\left(\frac{I(x_0, r_0, y_0, z_0, t_1) I(x_1, r_1, y_1, z_1, t_1)}{I(x_0, r_0, y_0, z_0, t_0) I(x_1, r_1, y_1, z_1, t_0)} \right)^{\frac{1}{2}} \quad (12)$$

is the geometric average of two indexes of technical change. The first measures technical change in terms of the inputs and outputs observed in period t_0 and the second in terms of the inputs and outputs observed at t_1 . If this geometric average is greater than one, there is said to be technical progress.

Assume that:

$$\begin{aligned} \ln I(x, r, y, z, t) &= \alpha_t + \beta_t^\top \ln x + \psi_t^\top \ln r + \delta_t^\top \ln y + \varepsilon_t^\top \ln z + \frac{1}{2} \ln x^\top B \ln x \\ &+ \frac{1}{2} \ln r^\top C \ln r + \frac{1}{2} \ln y^\top D \ln y + \frac{1}{2} \ln z^\top E \ln z + \ln x^\top F \ln r \\ &+ \ln x^\top G \ln y + \ln x^\top H \ln z + \ln r^\top I \ln y + \ln r^\top J \ln z + \ln y^\top K \ln z, \end{aligned}$$

with $B, C, D,$ and E all symmetric and sufficient parametric restrictions are imposed to ensure positive linear homogeneity in x and r . Thus, changes in the technology over time are assumed to be captured solely by the first-order time-specific parameters and all second-order terms are assumed time-invariant. Then assuming cost minimization, so that $\frac{I(x_0, r_0, y_0, z_0, t_0)}{I(x_1, r_1, y_1, z_1, t_1)}$ equals one, and using Diewert's Quadratic Identity, standard manipulations reveal that:

$$\begin{aligned} \ln P_{t_1, t_0}(x_0, r_0, y_0, z_0) &= \left. \begin{aligned} &\frac{1}{2} \left[\frac{\partial \ln C(w_1, v_1, y_1, z_1, t_1)}{\partial \ln y_1} + \frac{\partial \ln C(w_0, v_0, y_0, z_0, t_0)}{\partial \ln y_0} \right] (\ln y_1 - \ln y_0) \\ &-\frac{1}{2} \left[\frac{v_1 r_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{v_0 r_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln r_1 - \ln r_0) \\ &-\frac{1}{2} \left[\frac{w_1 x_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{w_0 x_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln x_1 - \ln x_0) \end{aligned} \right\} T\dot{F}P \\ &+ \frac{1}{2} \left[\frac{\partial \ln C(w_1, v_1, y_1, z_1, t_1)}{\partial \ln z_1} + \frac{\partial \ln C(w_0, v_0, y_0, z_0, t_0)}{\partial \ln z_0} \right] (\ln z_1 - \ln z_0). \quad (13) \end{aligned}$$

The first three rows of (13) are the Törnqvist measure of TFP growth. It's the basis for many sectoral or national TFP calculations.⁶ The final row is the correction needed to adjust for the presence of z . Assuming

⁶Productivity accounting is often based on flexible functional forms other than the transcendental logarithmic. However, computed versions of indexes for flexible functional forms typically closely approximate one another. Thus, while (14) may not be exactly relevant in many instances, it is a close approximation that has the distinct advantage of widespread familiarity. And we use it here for the purposes of illustration. It is an easy matter to develop an exactly parallel analysis for other flexible functional forms. The results extend directly.

profit maximization,

$$\ln P_{t_1, t_0}(x_0, r_0, y_0, z_0) = \left. \begin{aligned} & \frac{1}{2} \left[\frac{p_1 y_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{p_0 y_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln y_1 - \ln y_0) \\ & - \frac{1}{2} \left[\frac{v_1 r_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{v_0 r_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln r_1 - \ln r_0) \\ & - \frac{1}{2} \left[\frac{w_1 x_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{w_0 x_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln x_1 - \ln x_0) \end{aligned} \right\} T\dot{F}P \\ + \frac{1}{2} \left[\frac{\partial \ln C(w_1, v_1, y_1, z_1, t_1)}{\partial \ln z_1} + \frac{\partial \ln C(w_0, v_0, y_0, z_0, t_0)}{\partial \ln z_0} \right] (\ln z_1 - \ln z_0), \quad (14)$$

If constant returns to scale in (x, r, y, z) is assumed,

$$\ln P_{t_1, t_0}(x_0, r_0, y_0, z_0) = \left. \begin{aligned} & \frac{1}{2} \left[\frac{p_1 y_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{p_0 y_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln y_1 - \ln y_0) \\ & - \frac{1}{2} \left[\frac{v_1 r_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{v_0 r_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln r_1 - \ln r_0) \\ & - \frac{1}{2} \left[\frac{w_1 x_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{w_0 x_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln x_1 - \ln x_0) \end{aligned} \right\} T\dot{F}P \\ + \frac{1}{2} \left[\frac{C_1 - p_1 y_1}{C_1} + \frac{C_0 - p_0 y_0}{C_0} \right] (\ln z_1 - \ln z_0).$$

2.3 A Pragmatic Issue

Under constant returns in (x, r, y, z) , a straightforward method exists to correct commonly computed TFP measures for the presence of by-products. Treat variable profit as the economic rent to z to obtain an implicit "rental rate". Then use that rental rate to adjust standard computed measures for the presence of z . Unfortunately, if one applied that correction to existing aggregate data used in *TFP* calculations, one may end where she started. *The adjusted TFP measure would equal the unadjusted measure.* The reason is simple and lies in methods used to calculate implicit price and quantity indexes.

Many existing TFP measures assume constant returns to scale in (x, r, y) and use an accounting identity analogous to (10) to impute either an implicit price for a fixed factor or to impute an implicit quantity.⁷ Under constant returns in (x, r, y) , a standard argument shows that variable profit can assume either one of two possible values, $\pi(p, w, r, z, t) = 0$ or $\pi(p, w, r, z, t) = \infty$. The practical procedure is to assume the former and then use it as the basis for imputing the price for a fixed factor. Notice, however, that this assumption also implies that by-products carry no rents. Hence, the imputed shadow price for z must be zero.

To illustrate, suppose that x is treated as fixed at \bar{x} , and that the goal is to impute a shadow price for it. Then profit at (\bar{x}, z) is

$$R(p, v, \bar{x}, z, t) - w\bar{x},$$

where

$$R(p, v, x, z, t) \equiv \max_y \{py - vr : (x, r, y, z) \in T(t)\},$$

is quasi-rent to (x, z) . If the imputation procedure assumes that profit is zero and imputes w as

$$w = \frac{R(p, v, \bar{x}, z, t)}{\bar{x}},$$

rents attributable to z are awarded to \bar{x} . If this imputed w is used without correction in later calculations, the rents properly attributable to z will already have been exhausted. For this reason, it seems apparent that the "easy" correction may require a recalculation of existing aggregate price and quantity measures

⁷To conserve space, I only consider the problem of imputing an unobserved factor price. A completely parallel argument shows that, if information is available upon w but not upon x and this same logic is used to impute the value of x , the imputed x value elements of x embedded in it.

The presence of z in the technology raises other issues with this approach to imputation. The imputation procedure described only truly works if constant returns in (x, r, y) is plausible. If, as I argued above, constant returns in (x, r, y) is implausible as a *structural* assumption, there is no *structural* reason to believe

$$\pi(p, w, v, z, t) = 0.$$

Thus, it's hard to accept this residual-based imputation as reflecting the *structure* of $T(t)$. I qualify these statements with the adjective *structural*. That's because theory suggests that long-run profit should equal zero. This result is often invoked to justify the structural assumption of constant returns as a relatively harmless approximation under the additional assumption of an equilibrium.

That logic, however, raises its own problems. The long-run argument requires *all* factors and outputs, including z , to be *freely* variable. If z is not priced, it should be chosen to exhaust marginal rents. That is,

$$\frac{\partial \pi(p, w, r, z, t)}{\partial z} = -\frac{\partial C(w, v, y, z, t)}{\partial z} \leq 0. \quad (15)$$

If z is unrestricted, (15) holds as an equality, then

$$\begin{aligned} \frac{\partial \ln I}{\partial t} &= \overbrace{\frac{py}{C} \dot{y}_t - \frac{wx}{C} \dot{x}_t - \frac{vr}{C} \dot{r}_t}^{TFP} + \frac{\partial \ln C}{\partial \ln z} \dot{z}_t \\ &= \overbrace{\frac{py}{C} \dot{y}_t - \frac{wx}{C} \dot{x}_t - \frac{vr}{C} \dot{r}_t}^{TFP}. \end{aligned}$$

The standard TFP measure exactly captures technical change. No adjustment for the presence of by products is necessary. The market cures all!

In the discrete case, similar reasoning establishes

$$\begin{aligned} \ln P_{t_1, t_0}(x_0, r_0, y_0, z_0) &= \frac{1}{2} \left[\frac{p_1 y_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{p_0 y_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln y_1 - \ln y_0) \\ &\quad - \frac{1}{2} \left[\frac{v_1 r_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{v_0 r_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln r_1 - \ln r_0) \\ &\quad - \frac{1}{2} \left[\frac{w_1 x_1}{C(w_1, v_1, y_1, z_1, t_1)} + \frac{w_0 x_0}{C(w_0, v_0, y_0, z_0, t_0)} \right] (\ln x_1 - \ln x_0), \end{aligned}$$

which yields a similar conclusion.

One is reminded of a dog chasing her tail. If standard arguments for constant returns in (x, r, y) are invoked, z should carry no marginal rents. As such, it can be *freely ignored in productivity calculations*. Italics are used to emphasize that the conclusion only applies for productivity calculations. That's because productivity calculations hinge upon the shadow value of z to producers and not to Society at large. Thus, the same argument does not extend to a Social Accounting Framework (for example, a green GNP).

As Ayres and Kneese (1969) noted long ago, environmental by-products present no special problem so long as the environment is priced zero, or equivalently the environment's absorptive capacity is infinite. In that case, standard economic theory would suggest that private optimizers and Society price z at zero at the margin. Is that case truly of interest? It would seem obvious that most agree that the environment is not priced at zero and absorptive capacity is not finite. Therefore, it would seem obvious that Society does not price z at zero at the margin.

But what about producers? It's their shadow prices we seek for proper productivity accounting. Their behavior is going to be determined by the institutional setting they face. If it's a competitive and unrestricted

market, theory suggests either that they will price z at either zero or negatively. Zero is, of course, easy to understand, but what is the sense of (15) implying a negative shadow price. Again that's easy to understand with some simple reasoning. First though, one must be careful to disabuse oneself of any notion that rational firms pollute for any other reason than economic gain. Then, it's clear that given a particular level of good output, rational behavior always requires firms to choose z to make costs as low as possible. And so if costs can be lowered, $\frac{\partial C}{\partial z} < 0$, in an unconstrained market setting by pouring pollutants into a stream, one would expect that to happen and to continue to happen until no more gain could be had. It would only be when costs cannot be lowered, $\frac{\partial C}{\partial z} \geq 0$, any further that one would expect this behavior to stop. At the margin, z and y should be weak production *substitutes* in an unregulated setting.

Substitutability between z and y , however, violates a stylized fact of broad swathes of research in environmental and resource economics. The next section uses this observation as the starting point to reprise the consequences of more modern approaches to by-product modelling in production systems for productivity accounting.

2.4 Shadow Pricing By-products

It seems inevitable that adjusting TFP measures to accommodate unpriced by-products requires measuring their shadow prices. If so, accounting methods will need to be augmented by some empirical procedure that can calculate these shadow prices. Depending upon the empirical procedure and the representation chosen, this can be done via $\frac{\partial \ln I}{\partial \ln z}$, $\frac{\partial \ln C}{\partial \ln z}$, or $\frac{\partial \ln \pi}{\partial \ln z}$. Doing so properly raises potential complications. Up to now, the analysis has imposed minimal structural assumptions on $T(t)$. That ensured as great a generality as possible. But it carried a cost: we cannot yet infer a sign or magnitude of $\frac{\partial \ln I}{\partial \ln z}$, $\frac{\partial \ln C}{\partial \ln z}$, or $\frac{\partial \ln \pi}{\partial \ln z}$.

This section raises conceptual questions that should be answered either before or during the selection of an appropriate empirical procedure for estimating these shadow prices. These issues arise regardless if the empirical approach is econometric or via mathematical programming. Because it is presumed that most readers are intimately familiar with an econometric approach, the next section considers the consequences in a programming framework.

Until quite recently, modelling by products followed one of two approaches. One, more prevalent in production economics, treats by-products as *weakly disposable outputs*. The basic idea, due to Shephard (1970), is that by-production reduction requires surrendering (proportionately) some good output. Färe and a constellation of his coauthors have been particularly influential in implementing this idea empirically (Färe et al. 1993, 1998).

The other, more prevalent in environmental economics, is to treat by-products *as though* they were "inputs" (for example, Barbera and McConnell 1990; Cropper and Oates 1992; Brandt, Schreyer, and Zipperer 2014).⁸ The basic idea is simple enough. Assuming technical efficiency, solve $I(x, r, y, z, t) = 1$ to obtain z as a function of x, r, y, t . Then by the implicit function theorem:

$$\frac{\partial z}{\partial y} = -\frac{I_y(x, r, y, z, t)}{I_z(x, r, y, z, t)}.$$

If z is treated as though it were a good output, its marginal effect on I would likely have the same sign as y , whence $\frac{\partial z}{\partial y} < 0$. Reducing bad output, if it is unpriced, involves increasing good output which benefits the producer. Hence, producers seemingly have no economic incentive to pollute, and even in the absence of regulation, pollution would never be a problem at the margin. On the other hand, it is treated as an input, just the reverse happens and good production and pollution become complements, again at the margin.

⁸See particularly their section entitled "A framework for thinking about bad outputs and its shadow prices".

A growing literature has noted conceptual problems with both approaches. If taken literally, they can require asserting the existence of production technologies that violate fundamental laws of nature (Førsund 1998, 2009; Pethig 2006; Murty and Russell 2002, 2006; Murty, Russell, and Levkoff 2011; Hoang and Coelli 2011; Kuosmanen and Kuosmanen 2013). The key problem is a failure to recognize that *material balance* imposes restrictions on physical technologies.

Material balance is not a new idea. Ayres and Kneese (1969) recognized its importance almost half a century ago. Briefly, material balance requires recognizing that certain material inputs, our r , are only partially converted or captured in final goods y . The remainder exists as a residual, part of our z , that must ultimately be absorbed by the environment. With this in mind, consider the by-product as input specification. Formally, this requires $z' \geq z \implies I(x, r, y, z', t) \geq I(x, r, y, z, t)$ (see, for example, Brandt et al. 2014).⁹ Intuitively, increasing z shifts production isoquants for x and r towards the origin allowing potentially cheaper input bundles to be associated with producing y . This specification requires that

$$\frac{\partial C(w, v, y, z, t)}{\partial z} \leq 0,$$

which we note is the opposite of (15). Here's the formal argument. By definition

$$C(w, v, y, z, t) = \min \{wx + vr : I(x, r, y, z, t) \geq 1\}.$$

Let (x, r) represent a solution. It must be true that $I(x, r, y, z, t) \geq 1$. The assumed monotonicity condition guarantees that $z' \geq z \implies I(x, r, y, z', t) \geq I(x, r, y, z, t) \geq 1$ which implies that (x, r) remains a feasible choice for $C(w, v, y, z', t)$. But that means $C(w, v, y, z', t)$ can never be larger than $C(w, v, y, z, t)$, variable cost is non-increasing in z . Economically, this means that a producer *always has a marginal incentive to produce more z* .

Our chicken farmer illustrates what this particular assumption means for a physical technology. Recalling that example, suppose that current (x, r, y, z) as associated with 1500 eggs, 3000 hens, appropriate feed and shelter for 3000 hens, and a half ton of manure satisfy

$$I(x, r, y, z, t) \geq 1.$$

The assumption requires that for z' equals 4 tons of manure

$$I(x, r, y, z', t) \geq I(x, r, y, z) \geq 1,$$

so that (x, r, y, z') is also technically feasible. *Producing four tons of manure is possible from the exact same level of feed and other inputs that produced one half ton.* The obvious question (which is the opposite of my earlier query): If eggs, hens, shelter and feed are all held constant, where does the extra manure come from? The assertion is silly.

Now consider pollution from applying pesticides to crops. In this case, r represents pesticides applied and z represents pesticides escaping into the environment. Call the crop being produced strawberries. The standard damage-control production model (Lichtenberg and Zilberman 1986) suggests that the strawberry technology be modelled as something like

$$T^S(t) = \{(x, r, y, z) : y \leq f(x, t) \exp(g(r - z, t)) \text{ at time } t\},$$

⁹We employ the input terminology because of its familiarity. More formally, the problem is not whether by-products are inputs or outputs. Obviously, most non-economists would view them as "outputs" because they result from the production process. The real issue is what production economists think of as their disposability property within the technology. The so-called input formulation imposes the same disposability properties upon by-products as are usually imposed upon inputs.

where $f(x, t)$ represents maximal amount of strawberries obtainable from x and $\exp(g(r - z, t))$ represents *damage control* as a nondecreasing function of the pesticide staying on the plant, $r - z$. The function g has embedded in it such factors as pest pressure. Allowing z to increase while holding r constant results in less active ingredient on the plant, less damage control, and ultimately less production. Here again, the by-product as input or complementary output specification is not consistent with physical reality. Pethig (2006) contains an extended discussion of such issues in a more general framework. Nutrient-balance approaches (for example, Hoang and Coelli 2011, Kuosmanen and Kuosmanen 2013, Serra, Chambers, and Oude Lansink 2014, Chambers, Serra, and Oude Lansink 2014) invoke similar reasoning for agricultural nutrients.

The primary message from both examples is that *by-production in agricultural systems often involves wasting valuable material inputs*. If the producer could, he or she would prefer that this wastage not occur. Unfortunately, physical reality often intervenes to ensure that some must occur. Anyone who has fed an animal in containment knows this first hand. Moreover, anyone familiar with US land-grant institutions knows that vast amounts of public and private research funds have been devoted to improving the efficiency with which animals and plants convert material inputs, such as feed, fertilizer, phosphates, etc. into output. But conversion rarely reaches 100%. By-products are the result. And given a fixed level of material input applied the more by-product that emerges, the less is imbedded in physical production.

A quick example from the poultry industry illustrates. The phytate-bound phosphorus in many cereal feeds is not digestible by chickens. As a consequence, the phosphorus needs for chickens must be met from other sources. This has two consequences. One because cereal feeds remain economically attractive, they continue to be used. But because the phytate-bound phosphorus cannot be absorbed by the chicken, a by product is created. Second, the supplementation of diets with phosphorus presents another opportunity for by production. As the *pfisteria-hysteria* crisis, which was linked to phosphate contamination, that confronted the Delmarva poultry industry illustrates many benefits might emerge from scientific advances that either lowered the production of phytate-bound phosphorus in cereal feed or that enabled chickens to digest phytate-bound phosphorus.

At this juncture, it seems a bit clearer why an economist might believe that $\frac{\partial C(w, v, y, z, t)}{\partial z} \geq 0$. Recall from (15) that's what theory suggests in the unregulated case, and not $\frac{\partial C(w, v, y, z, t)}{\partial z} < 0$ as the input formulation dictates. But the natural bias is to assume that (15) holds as an equality in the unregulated case, and $\frac{\partial C(w, v, y, z, t)}{\partial z} < 0$ in the regulated case. It's easy to see why an economist might think that way.¹⁰ And in many cases, it's perfectly plausible. But it is also seems plausible that when material inputs are involved, material balance requirements can limit the manner in which z can be either increased or decreased. If you want to produce eggs, you must also produce manure, and that brings an economic cost, even if it is only requiring you to be careful where you step. Physical limits exist to how far by production can be eliminated. Once these limits are reached, the firm may truly want to proceed even further because it would involve embedding more of its market costly inputs, r , in the physical product rather than waste them in the form of z . But science, technology, or both may prevent that from happening.

That brings us to the competing alternative, weak output disposability. Formally, it requires that $I(x, r, \mu y, \mu z, t) \geq I(x, r, y, z, t)$ for $\mu < 1$. Intuitively, shrinking the by-product requires shrinking some of the good product proportionately. This, too, can conflict with material balance concerns (see Murty and Russell 2002, 2006) in the presence of material inputs that are not fully absorbed into y .¹¹

A more nuanced view emerges from the work of Frisch (1965), Førsund (1998, 2009), Murty and Russell

¹⁰However, one need only recall the standard economic example of a Pigouvian externality, ostensibly drawn from agriculture, the apiary and the orchard, to see how far-fetched economic intuition can be when it comes to real-world agricultural systems.

¹¹Similar reasoning also leads us to question our maintained assumption of weak disposability of inputs.

(2002, 2006), Murty, Russell, and Levkoff (2011), and Hoang and Coelli (2011). The essential idea is that production systems involving material inputs that generate by-products be modelled by interdependent production processes. Practically speaking, that means that more than one distance function (transformation function, production function, etc.) may be required. Førsund (2008), Murty and Russell (2002, 2006) and Murty, Russell, and Levkoff (2011) provide detailed arguments for thinking in these terms, and I refer the interested reader to them for further insight into the merits of this approach.

What I want to do here is to consider what their arguments imply for TFP measurement in the presence of by-production. For concreteness, I use a simple example, while noting that actual systems likely require deeper thought. What follows is more a parable about what might happen rather than a description of what does happen in any particular setting.

With that in mind, let's return to the strawberry production example. There, the essential idea is that pesticides do not promote plant growth. Instead, they prevent plant damage. Applying chemical pesticides typically involves pesticides escaping into the environment. You can picture this in terms of spray application. Some of the pesticide sprayed ends on the plant, some on the (human) applicator, and some in the surrounding environment. So here, r is the amount of pesticide sprayed and z is the amount that finds its way into the surrounding environment, on to the applicator, and not on the plant.

The amount that escapes into the environment depends upon how it is applied. If the spraying equipment is leaky and the applicator is inattentive, less is likely to reach the plant than if the equipment were not leaky and the applicator were attentive. Thus, the amount of the pesticide reaching the plant might be viewed as an intermediate output whose production depends upon other inputs such as labor and capital. The key idea behind the Frisch-Førsund-Murty-Russell (FFMR) approach is to model the production of this intermediate output and the production of the final output, y , as *separate, but not necessarily separable, production processes*. To that end, write

$$T^Z(t) = \{(x, r, y, z) : (x, r) \text{ can produce } r - z \text{ at time } t\},$$

for this production process. The overall production technology combines this process with $T^S(t)$. More formally,

$$T(t) = T^S(t) \cap T^Z(t).$$

$T(t)$ is the intersection of two production sets, each of which describes a physical production process. One models the process between y , x , and the amount of r that reaches the plant. The other describes the process of delivering r to the plant. This specification ensures that $T(t)$ is consistent with material balance by ensuring that all of the physical mass of r sprayed is accounted for.

This specification carries complications. Note first that there are now two distance functions to consider because (McFadden 1978)

$$\begin{aligned} I(x, r, y, z, t) &= \sup \left\{ \lambda > 0 : \left(\frac{x}{\lambda}, \frac{r}{\lambda}, y, z, t \right) \in T^S(t) \cap T^Z(t) \right\} \\ &= \max \left\{ I^S(x, r - z, y, t), I^Z(x, r, r - z, t) \right\}, \end{aligned}$$

where

$$I^S(x, r - z, y, t) = \sup \left\{ \lambda > 0 : y \leq f\left(\frac{x}{\lambda}, t\right) \exp\left(g\left(\frac{r}{\lambda} - z\right), t\right) \text{ at time } t \right\},$$

and

$$I^Z(x, r, r - z, t) = \sup \left\{ \lambda > 0 : \left(\frac{x}{\lambda}, \frac{r}{\lambda} \right) \text{ can produce } \frac{r}{\lambda} - z \text{ at time } t \right\}.$$

By construction $z' \geq z^o \implies I^S(x, r - z', y, t) \leq I^S(x, r - z^o, y, t)$. In words, increasing pesticide runoff, all else equal, decreases the percentage of applied r reaching the crop. That implies greater pest damage and less final production. One can visualize this as the production isoquant for process S shifting out as z increases. But, on the other hand, it seems clear that $z' \geq z^o \implies I^Z(x, r, r - z', t) \geq I^Z(x, r, r - z^o, t)$. Increasing z' now implies manufacturing less of the intermediate output $r - z$ so that the input isoquant shifts back towards the origin.

Figure 1 illustrates. There are two isoquants for two variable inputs, capital and labor. One is for process Z and the other for S . The isoquant for process S is labelled S and the isoquant for Z is labelled Z . As drawn, the Z process is more labor intensive than the S process. The production isoquant for $T(t)$, is the lower boundary of the cross-hatched area that falls above both isoquants. As the figure illustrates, which process isoquant is relevant for a capital-labor ratio can vary as the ratio varies.

The associated cost function is

$$\begin{aligned} C(w, v, y, z, t) &= \inf \{wx + vr : I(x, r, y, z, t) \geq 1\} \\ &= \inf \{wx + vr : \max \{I^S(x, r - z, y, t), I^Z(x, r, r - z, t)\} \geq 1\} \\ &\geq \max \{C^S(w, v, y, z, t), C^Z(w, v, r - z, t)\}, \end{aligned}$$

which under weak restrictions reduces to

$$C(w, v, y, z, t) = \max \{C^S(w, v, y, z, t), C^Z(w, v, r - z, t)\}. \quad (16)$$

Both the input distance function and the cost function can be analytically inconvenient. In particular, both $I(x, r, y, z, t)$ and $C(w, v, y, z, t)$ are non-differentiable at certain points. This is a technical issue, typically restricted to sets of measure zero, that we happily brush under the rug. Even so, other issues remain.

The damage-control specification, which has proven crucial in obtaining plausible empirical results for pesticide-demand behavior, imposes structural restrictions that would ideally be recognized in empirical work. By Bellman's Principle

$$\begin{aligned} C^S(w, v, y, z, t) &= \min_{x, r} \{wx + vr : y \leq f(x, t) \exp(g(r - z, t))\} \\ &= \min_r \left\{ vr + \min_x \{wx : y \leq \exp(g(r - z, t)) f(x, t)\} \right\} \\ &= \min_r \left\{ vr + \min_x \left\{ wx : \frac{y}{f(x, t)} \leq \exp(g(r - z, t)) \right\} \right\} \\ &= \min_r \left\{ vr + C^f \left(w, \frac{y}{\exp(g(r - z, t))}, t \right) \right\} \end{aligned}$$

where C^f denotes the variable cost function associated with $f(x, t)$. Ideally, one would wish that the econometric procedure would capture the structure of C^f . Methods for doing just that exist (Chambers, Karagiannis, Tzouvelekas 2010, 2014), but unfortunately another empirical problem intrudes.

Econometrically, expression (16) defines a switching regression model for the cost function. Moreover, which regime prevails is endogenously determined within the system. Estimation would be complicated even if the cost structures were linear. (Note, a similar situation arises if one instead attempts to estimate I for this structure.)

We're still not done. Suppose that

$$C^S(w, v, y, z, t) > C^Z(w, v, r - z, t).$$

Then

$$\begin{aligned} \frac{\partial C(w, v, y, z, t)}{\partial z} &= \frac{\partial C^S(w, v, y, z, t)}{\partial z} \\ &= \frac{\partial C^f\left(w, \frac{y}{\exp(g(r-z, t))}, t\right)}{\partial y} \frac{g'(r-z, t)}{\exp(g(r-z, t))} \geq 0, \end{aligned}$$

where the second equality follows from the envelope theorem. Conversely, if

$$C^S(w, v, y, z, t) < C^Z(w, v, r-z, t),$$

then

$$\frac{\partial C(w, v, y, z, t)}{\partial z} = -\frac{\partial C^Z(w, v, r-z, t)}{\partial(r-z)} \leq 0.$$

Figure 2 illustrates. The solid directional arrows emerging from the effective isoquants illustrate the direction in which we anticipate the isoquant moving as z increases. As discussed above, one expects an increase in z to move the isoquant for S outwards and the isoquant for Z towards the origin. Therefore, if estimated cost functions (similar issues emerge with a profit function) are used to derive shadow prices for z , whether those shadow prices are positive or negative will depend crucially upon which production isoquant relative input prices place the producer. Moreover, as those prices changes, so too may the relevant isoquants. Sometimes the shadow price for z may be positive and sometimes negative. It depends, and *a priori* it is inappropriate to assign a sign to marginal cost.¹²

This structure also involves an identification issue. We aim to measure

$$\frac{\partial \ln I}{\partial t}.$$

As Figure 3 illustrates, two distinct measures may emerge. Depending upon the capital-labor ratio involved,¹³ one captures either $\frac{\partial \ln I^S}{\partial t}$ or $\frac{\partial \ln I^Z}{\partial t}$, but generally not both. Moreover, as the capital-labor ratio changes, one might capture $\frac{\partial \ln I^S}{\partial t}$ for some observations and $\frac{\partial \ln I^Z}{\partial t}$ for others. This raises the question of which technical change is being measured. Is it that for the intermediate output (controlling residual released), or is it for producing strawberries? Does it matter? I would argue that it does.

2.5 Structural Productivity Accounting and By-Products

Another way to appreciate the difficulties that can emerge is to consider the main alternative to econometric estimation of either I or C . That involves using a mathematical programming approach and observed data to construct a conservative approximation to $T(t)$. Thus, if at time t one had K observations on observed inputs and outputs,

$$(x_t^k, r_t^k, y_t^k, z_t^k), \quad k = 1, 2, \dots, K$$

either an activity-analysis approach or a data-envelopment analysis (DEA) approach could be used to approximate $T(t)$. The DEA approach has proven especially popular in recent years, and so I will use it here. Examples of its application to problems involving by-products can be found in Murty, Russell, and Levkoff

¹²Using a programming formulation of similar production process, Chambers, Serra, and Oude Lansink (2014) report statistical evidence that the state-contingent shadow prices of by-products for a group of Spanish farmers were not appreciably different from zero.

¹³Here it seems particularly useful to recall expression (12). It's easy to visualize instances where the input-output bundles land one on different isoquants.

(2011), Hoang and Coelli (2011), and Chambers, Serra, and Oude Lansink (2014), and Serra, Chambers, and Oude Lansink (2014).

A primary reason for DEA's popularity is that it is often seen as requiring no prior restrictions on functional form. That often leads to it being referred to as *non-parametric*. Another is practical. It virtually always yields results even in instances where data-specific problems may prevent econometric estimation of functional structures.

Oftentimes, the primary issue in integrating by-products into DEA formulations is which form linear constraints should take. A generic DEA model illustrates. For the given data set at time t , the pollution as input DEA approximation to the technology, assuming constant returns, might look something like

$$T^{PI}(t; K) = \left\{ \begin{array}{l} (x, r, y, z) : x \geq \sum_k \lambda_k x_t^k, r \geq \sum_k \lambda_k r_t^k, y \leq \sum_k \lambda_k y_t^k, z \geq \sum_k \lambda_k z_t^k, \\ \lambda_k \geq 0, k = 1, 2, \dots, K \end{array} \right\},$$

while one involving weak disposability of (y, z) might look something like

$$T^{WD}(t; K) = \left\{ \begin{array}{l} (x, r, y, z) : x \geq \sum_k \lambda_k x_t^k, r \geq \sum_k \lambda_k r_t^k, y = \sum_k \lambda_k y_t^k, z = \sum_k \lambda_k z_t^k, \\ \lambda_k \geq 0, k = 1, 2, \dots, K \end{array} \right\},$$

so that the difference between the two specifications basically boils down to how one writes the output-based constraints.

Given either version of the DEA specification, one can compute

$$I^n(x, r, y, z; t, K)^{-1} = \min \left\{ \begin{array}{l} \beta > 0 : \beta x \geq \sum_k \lambda_k x_t^k, \beta r \geq \sum_k \lambda_k r_t^k, y^{??} \sum_k \lambda_k y_t^k, z^{??} \sum_k \lambda_k z_t^k, \\ \lambda_k \geq 0, k = 1, 2, \dots, K \end{array} \right\},$$

where ?? stands for the binary operator appropriate for specification $n = WD, PI$. And then, one obtains the following measured version¹⁴ of

$$\hat{P}_{t_1, t_0}^n(x_0, r_0, y_0, z_0, x_1, r_1, y_1, z_1) = \frac{I^n(x_0, r_0, y_0, z_0; t_0, K)}{I^n(x_1, r_1, y_1, z_1; t_1, K)} \left(\frac{I^n(x_0, r_0, y_0, z_0; t_1, K)}{I^n(x_0, r_0, y_0, z_0; t_0, K)} \frac{I^n(x_1, r_1, y_1, z_1; t_1, K)}{I^n(x_1, r_1, y_1, z_1; t_0, K)} \right)^{\frac{1}{2}}.$$

And thus, the productivity measure incorporates both an efficiency component and a measure of technical change.

Such specifications are particularly common for firm-level productivity comparisons, because researchers can relax the assumption of economic efficiency typically associated with the productivity-accounting approach. In making *firm performance comparisons*, this can be very important. It's easy to see, of course, that neither accounts for material-balance concerns. Thus, for these technologies, manure can appear out of thin air on the one hand, or manure and chickens can be costlessly decreased while maintaining feed levels constant on the other.

Suppose that one decided to account for material balance using the FFMR two-process parable developed above. If these same observed data were used to construct a DEA approximation to process Z , it might look something like:

$$T^Z(t; K) = \left\{ \begin{array}{l} (x, r, r - z) : x \geq \sum_k \lambda_k x_t^k, r \geq \sum_k \lambda_k r_t^k, r - z \leq \sum_k \lambda_k (r_t^k - z_t^k), \\ \lambda_k \geq 0, k = 1, 2, \dots, K \end{array} \right\}.$$

¹⁴Note each *DEA* approximation is constructed using only data observed from a common time period. If there is a natural way to order t_1 and t_0 , one would construct different representations of the technology under the presumptions of either technical progress or regress.

The associated empirical approximation to the input distance function is:

$$I^Z(x, r, r - z, t; K)^{-1} = \min \left\{ \begin{array}{l} \beta : \beta x \geq \sum_k \lambda_k x_t^k, \beta r \geq \sum_k \lambda_k r_t^k, \beta r - z \leq \sum_k \lambda_k (r_t^k - z_t^k), \\ \lambda_k \geq 0, k = 1, 2, \dots, K \end{array} \right\}.$$

A *standard* DEA approximation to the S process would be

$$T^S(t; K) = \left\{ \begin{array}{l} (x, r - z, y) : x \geq \sum_k \gamma_k x_t^k, r - z \geq \sum_k \gamma_k (r_t^k - z_t^k), y \leq \sum_k \gamma_k y_t^k, \\ \gamma_k \geq 0, k = 1, 2, \dots, K \end{array} \right\},$$

and the standard DEA approximation to $T(t)$ would be

$$T(t; K) = T^S(t; K) \cap T^Z(t; K)$$

This is a *standard DEA approximation* to the two-process version of $T(t)$. Specific examples of using related models, for a different problem, are Serra, Chambers, and Oude Lansink (2014) and Chambers, Serra, and Oude Lansink (2014). This specification does not account for the *damage-control nature of pesticides*. Not surprising. The standard DEA approximation is non-parametric and approximates $T(t)$ polyhedrally. But the essence of the damage-control specification is that the biology of the system dictates a multiplicatively separable structure for the S process. That biological structure, although economically inconvenient, is inherited from the biological literature, whose study of yield response often predates econometric production analysis. In short, if we base productivity measurement on similarly misspecified *DEA* models, we risk ignoring lessons learnt by others to accommodate shortcomings built into our tools.

As with econometric estimation of the cost function, this difficulty can be surmounted. But the lesson remains. Standard techniques need to be adapted to the physical structure of the problem. Both cases illustrate the same practical problem. Accommodating the presence of z likely requires empirically approximating $I(x, r, y, z, t)$, $C(w, r, y, z, t)$, or $\pi(p, w, v, z, t)$. If that's done appropriately, one can infer a measure of $\frac{\partial \ln I}{\partial t}$ directly from the approximated technology or by accounting measures. But specifying an empirical representation of the technology that is consistent with physical reality and the nature of z carries complications. And those complications are not only conceptual, they require nonstandard empirical representations. And the nonstandard empirical representations may well require methodological innovations. The problem is not one of simply specifying a generic cost or profit function, estimating it, and taking a partial derivative. Neither is it one of building a standard DEA model and calculating the measure of technical progress directly.

3 Some Final Remarks

It has been the tradition in some areas to assume that $\frac{\partial \ln I}{\partial \ln z} \geq 0$. One easily sees why. By-products released into the environment carry a social cost that may or may not be equated to private marginal costs by firms. Thus, if one were to attempt to construct a measure of, say, *green* GNP, those costs should be subtracted from value-added if *green* GNP is to be a representative measure of an economy's welfare.

That's not the problem at hand. And, that intuition should not bias our thinking, especially if it requires a conceptual model that violates physical realities. The problem is not social welfare measurement or calculation of a *green* productivity. Rather, it's measuring as accurately as possible technical progress in the presence of by-products. That requires measuring the private marginal cost of by-products to the producer. Whether that measure's sign is positive or negative seemingly is an empirical issue. Common approaches to imposing structure on the physical model can predetermine the empirical outcome. But they can also do so

at the cost of endowing empirical technologies with magical powers that physical technologies cannot hope to replicate. To avoid such outcomes, productivity accounting in agricultural systems needs to be consonant with the physical realities of the underlying biological systems. That likely requires modification of existing procedures and retabulation of existing measures.

4 References

- Abramovitz, M. "Resource and Output Trends in the United States Since 1870." *American Economic Review* 46 (1956): 5-23.
- Ayres, R. U., and A. W. Kneese. "Production, Consumption, and Externalities." *American Economic Review* 59 (1969): 282-97.
- Ball, E., S. L. Wang, and R. Nehring. "Productivity and Growth in US Agriculture." Economic Research Service, United States Department of Agriculture, 2015.
- Barbera, A., and V. McConnell. "The Impact of Environmental Regulations on Industry Productivity: Direct and Indirect Effects." *Journal of Environmental Economics and Management* 18 (1990): 50-65.
- Barton, G., and M. R. Cooper. "Relation of Agricultural Production to Inputs." *Review of Economics and Statistics* 30 (1948): 117-26.
- Barton, G. T. "Technical Change, Food Needs, and Aggregate Resource Adjustment." *Journal of Farm Economics* 40 (1958): 1429-37.
- Barton, G. T., and D. Durost. "The New USDA Index of Inputs." *Journal of Farm Economics* 42 (1960): 1398-410.
- Brandt, N., P Schreyer, and V Zipperer. "Productivity Measurement With Natural Capital and Bad Outputs." OECD Working Paper, No. 1154, 2014.
- Caves, D. W., L. R. Christensen, and W. E. Diewert. "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Indexes." *The Economic Journal* 92 (1982b): 73-86.
- Chambers, R. G., G. Karagiannis, and V. Tzouvelekas. "Another Look at Pesticide Productivity and Pest Damage." *American Journal of Agricultural Economics* 92 (2010): 1401-19.
- . "Productivity Accounting for Separable Technologies." *Journal of Productivity Analysis* 41 (2014): 41-50.
- Chambers, R. G., T. Serra, and A. Oude Lansink. "On the Pricing of Undesireable Outputs." *European Review of Agricultural Economics* 41 (2014): 485-509.
- Cropper, M., and W. Oates. "Environmental Economics: A Survey." *Journal of Economic Literature* XXX (1992): 675-740.
- Fare, R., Y. Chung, and S. Grosskopf. "Productivity and Undesirable Outputs: A Directional Distance Function Approach." *Journal of Environmental Management* (1997): 229-40.
- Färe, R., S. Grosskopf, C. A. Knox Lovell, and S. Yaisawarng. "Derivation of Shadow Prices for Undesirable Outputs: A Distance Function Approach." *Review of Economics and Statistics* (1993): 374-85.
- Forsund, F. "Good Modeling of Bad Outputs; Pollution and Multi-Output Production." *International Review of Environmental and Resource Economics* 3 (2009): 1-38.
- . "Good Modelling of Bad Outputs." Working Paper, 1998.
- Frisch, R. *Theory of Production*. Dordrecht: D. Reidel Publishing Company, 1965.
- Griliches, Z. "Measuring Inputs in Agriculture: A Critical Survey." *Journal of Farm Economics* 42 (1960): 1411-27.
- . "The Sources of Measured Productivity Growth: United States Agriculture, 1940-60." *Journal of Political Economy* 71 (1963): 331-46.
- Hoang, V.-N., and T. Coelli. "Measurement of Agricultural Total Factor Productivity Incorporating Environmental Factors: A Nutrient-Balance Approach." *Journal of Environmental Economics and Management* 62 (2011): 462-74.
- Kuosmanen, N., and T. Kuosmanen. "Modeling Cumulative Effects of Nutrient Surpluses in Agriculture: A

- Dynamic Approach to Material-Balance Accounting." *Ecological Economics* 90 (2013): 159-67.
- Lichtenberg, E., and D. Zilberman. "The Econometrics of Damage Control: Why Specification Matters." *American Journal of Agricultural Economics* 68 (1986): 261-73.
- McFadden, D. "Cost, Revenue, and Profit Functions." *Production Economics: A Dual Approach to Theory and Applications*. eds. M. Fuss, and D. McFadden. Amsterdam: North-Holland, 1978.
- Murty, S., and R. R. Russell. "On Modeling Pollution-Generating Technologies." Working Paper, University of California, Riverside, 2006.
- , On Modeling Pollution Generating Technologies. " Working Paper, University of California, Riverside, 2002.
- Murty, S., R. R. Russell, and S. Levkoff. "On Modeling Pollution-Generating Technologies." *Journal of Environmental Economics and Management* 2012, no. 64: 117-35.
- Pethig, R. "Nonlinear Production, Pollution, Abatement and Materials-Balance Reconsidered." *Journal of Environmental Economics and Management* 51 (2006): 185-204.
- Schultz, T. W. "Reflections on Agricultural Production, Output, and Supply." *Journal of Farm Economics* 48 (1956): 748-62.
- Serra, T., R. G. Chambers, and A. Oude Lansink. "Measuring Technical and Environmental Efficiency in a State-Contingent Technology." *European Journal of Operational Research* 236 (2014): 706-17.
- Shephard, R. W. *Theory of Cost and Production Functions*. Princeton, NJ: Princeton University Press, 1970.
- Solow, R. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39 (1957): 312-20.
- Wang, S. L., P. Heisey, D. Schimmelfennig, and E. Ball. *Agricultural Productivity Growth in the United States: Measurement, Trend, and Drivers*, Economic Research Report 189. Economic Research Service, United States Department of Agriculture, Washington DC.

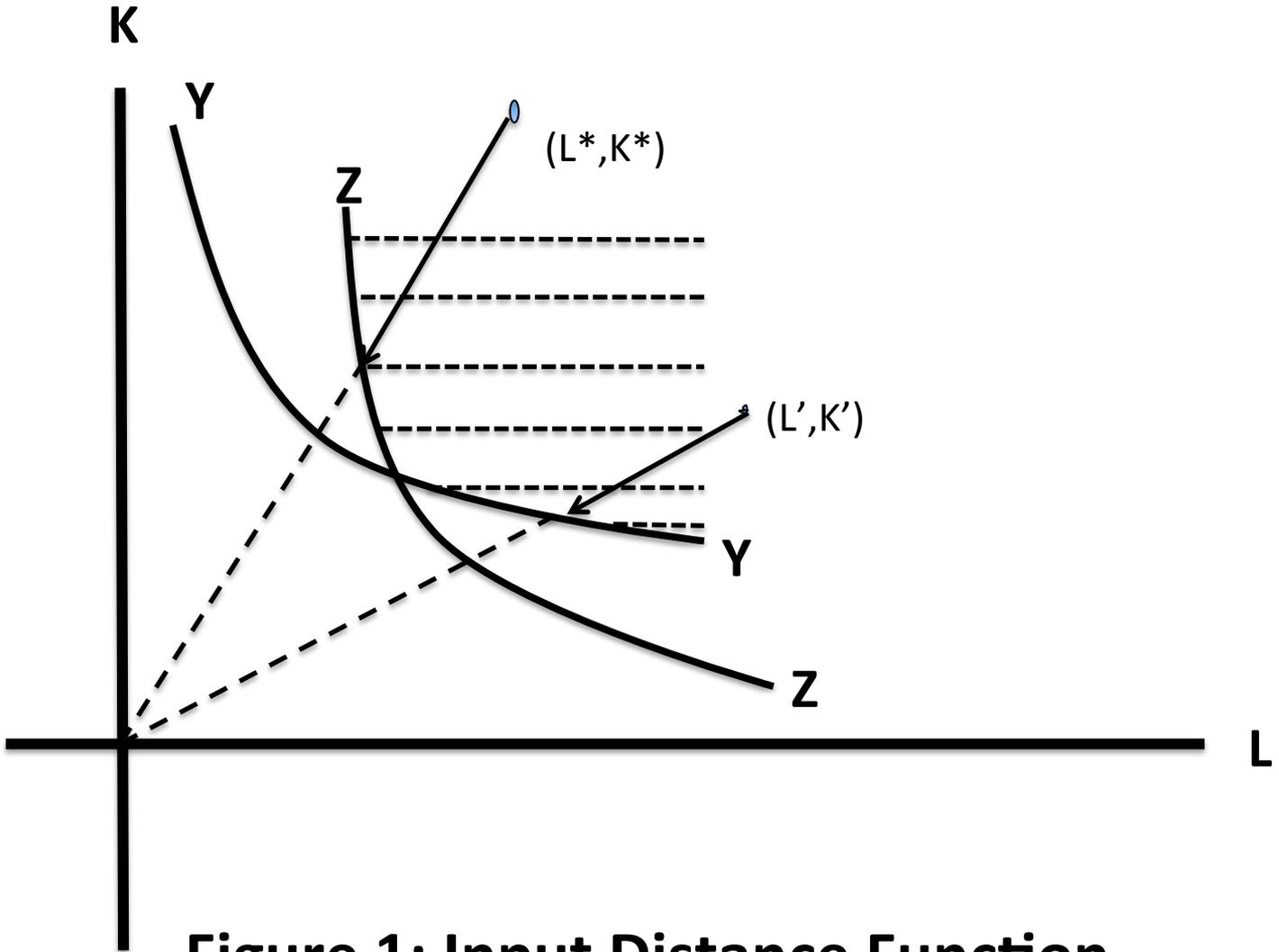


Figure 1: Input Distance Function

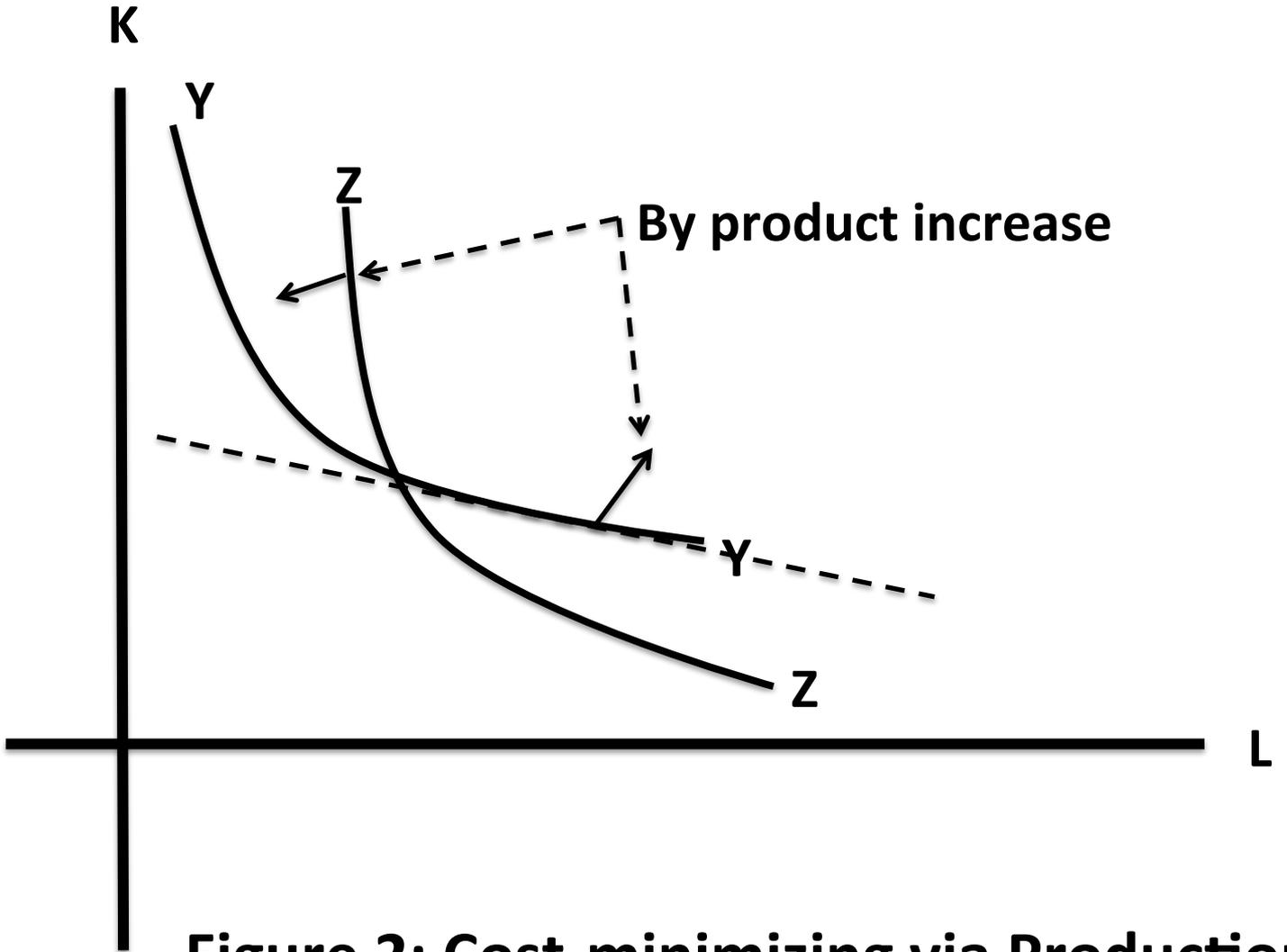


Figure 2: Cost-minimizing via Production

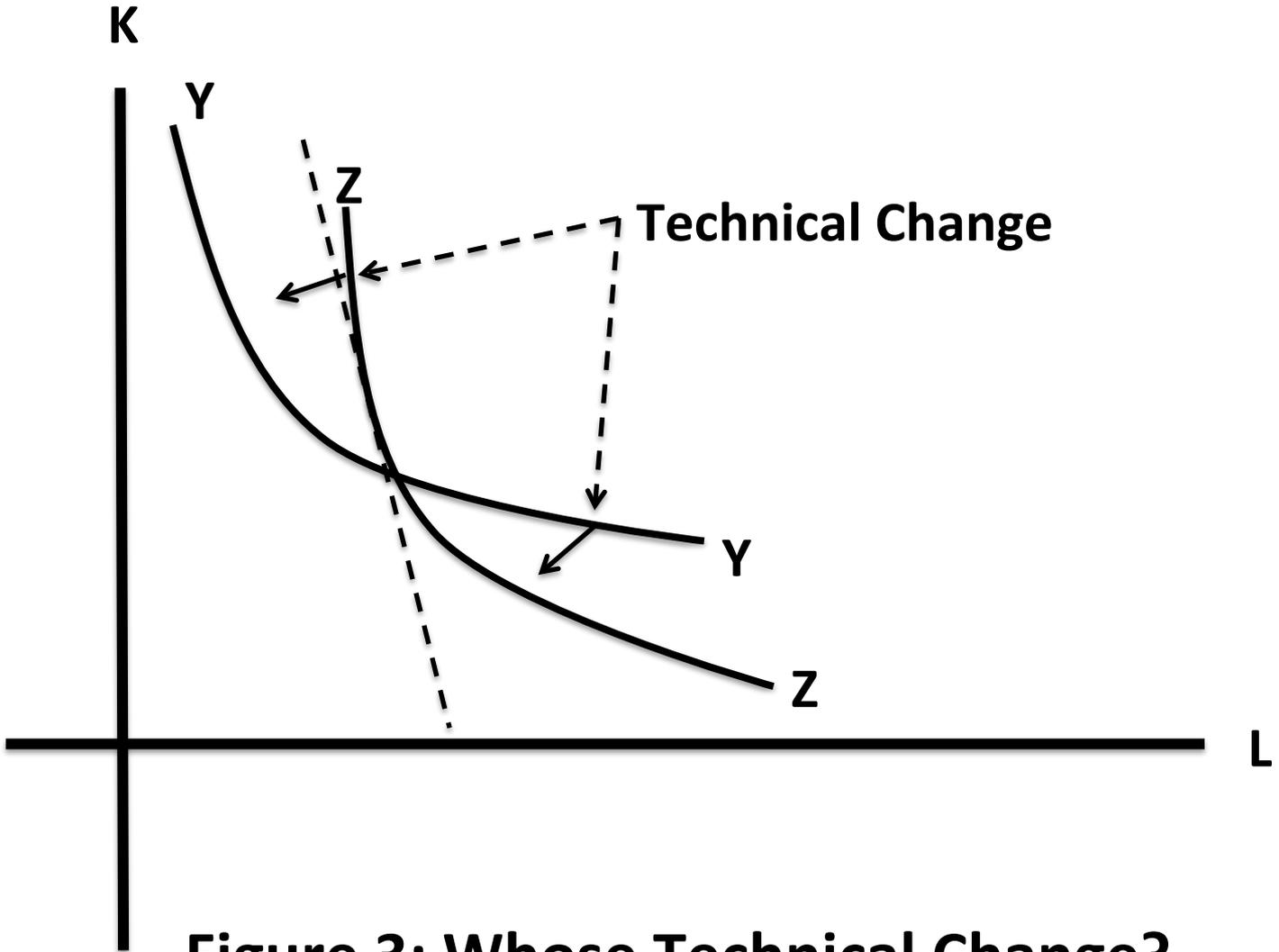


Figure 3: Whose Technical Change?