Toward a Normative Theory of Crop Yield Skewness

David A. Hennessy

Working Paper 08-WP 473
June 2008

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070
www.card.iastate.edu

David Hennessy is a professor in the Department of Economics and Center for Agricultural and Rural Development at Iowa State University.

Questions or comments about the contents of this paper should be directed to David A. Hennessy, 578C Heady Hall, Iowa State University, Ames, IA 50011-1070; Ph. (515) 294-6171; Fax: (515) 294-6336; E-mail: hennessy@iastate.edu.

Abstract

While the preponderance of empirical studies point to negative crop yield skewness in a wide variety of contexts, the literature provides few clear insights on why this is so. The purpose of this paper is to make three points on the matter. We show formally that statistical laws on aggregates do not suggest a normal yield distribution. We explain that whenever the weather-conditioned mean yield has diminishing marginal product, then there is a disposition toward negative skewness in aggregate yields. This is because a high marginal product in bad weather states stretches out the left tail of the yield distribution relative to that of the weather distribution. Turning to disaggregated yields, we decompose unconditional skewness into weather-conditioned skewness plus two other terms and study each in turn.

Keywords: conditional distribution, crop insurance, negative skewness, spatial heterogeneity, statistical laws.

Introduction

Crop yield distributions are used to model risk exposures, and also as an input in informed rating-making when designing and marketing crop insurance. As a reading of recent literature on crop yield modeling should confirm, it is a controversial topic (Just and Weninger 1999; Ker and Goodwin 2000; Atwood, Shaik, and Watts 2002, 2003; Sherrick *et al.* 2004). One difficulty is with appropriate conditioning of empirical data. Plots differ across space because of climate and soil variation, while different technologies may also be used. So concerns about the effects of spatial aggregation naturally arise. There may also be temporal and spatial heterogeneities in scrutinized data due to random weather events. Complete control over these differences is practically impossible.

While cases in which skewness is likely positive have been identified in Day (1965) and also Ramirez, Misra, and Field (2003), the majority of studies have estimated negative skewness (Nelson and Preckel 1989; Swinton and King 1991; Moss and Shonkwiler 1993; Ramirez, Misra, and Field 2003; Atwood, Shaik, and Watts 2002, 2003). This is at variance with the suggestion that the crop yield distribution should be normal and so should have zero skewness, in light of statistical limit considerations. Referring to a version of the Central Limit Theorem (CLT), Just and Weninger (p. 302) state "Central limit theory thus calls into question the nonnormal empirical results found to date with aggregate time-series data." On the other hand, Goodwin and Mahul (2004, pp. 13-14) and others hold that spatial dependence and systemic risk factors mean a straightforward application of CLT is not appropriate.[1] We see a need for clarification.

Apart from Just and Weninger (1999), two other papers have sought to provide foundations for the origin of observed yield skewness. Ker and Goodwin (2000, p. 465 and Fig. 1) reason

[1] Wang and Zhang (2003) use dependent versions of the CLT, but for the purpose of estimating a crop insurance company's portfolio risk as the number of risks grows, not for identifying yield distributions. As we will explain, it is more meaningful to use CLT when studying the distribution of a sum than when studying the distribution of an average.

that aggregating over some weather-conditioned yield distribution should predispose unconditional yield toward negative skewness whenever the weather variable is itself negatively skewed. Focusing on the small plot level of analysis, Hennessy (2008) has shown that the law of the minimum technology together with stochastic resource availabilities can support positive or negative skewness. Efforts to tightly control the left tail of a resource availability can create a negative skewness tendency. Irrigation, however, may increase skewness by completely eliminating a source of yield shortfall.

Assuming a general technology and dealing with aggregation issues, this paper makes three main points. We invoke the systemic-idiosyncratic risk decomposition that is widely used in financial risk theory (LeRoy and Werner 2001).[2] The first point is that statistical large number and limit laws assert more about the distribution of an average than just convergence to normality. In specifications relevant to modeling mean crop yields, the laws also provide a limiting value of zero for the distribution variance. In short, in the aggregate only the distribution of systemic risk sources remains as a determinant of the crop yield distribution. The relevance of statistical large number and limit laws to crop yield distributions has long been recognized (Hirshleifer 1961). Yet the state of the crop yield distribution modeling literature suggests that what they imply for that context needs to be parsed.

Our second main point is constructive. It concerns the role of weather conditioning and so can be seen as a formalization of and development on the aforementioned graphical argument in Ker and Goodwin (2000). We clarify that, because of a result in van Zwet (1964), a beneficial random systematic weather factor with decreasing marginal product ensures that the yield distribution will have smaller (i.e., more negative) skewness than that of the underlying weather factor. Intuitively this is because, in the change of random variables from weather

[2] Portfolio theory could, and sometimes does, seek to model spatial dimensions to systemic risk. There of course it is not geographic distance but rather attribute distance that is modeled, as when food company shares are grouped to be more similar to each other than to defense company shares.

factor to yield, the production function’s concavity stretches out the left tail of the weather factor distribution. This negative bent is consistent with the empirical literature on yield skewness.

Our third main point is to relate skewness of aggregate yields to skewness of local yields, where local idiosyncrasies play a role. We decompose the relationship into three terms and study each in turn. One is the systemic effect on aggregate yields, studied in point two. Another is the skewness of the local idiosyncrasies, where the impact on unconditional skewness is clear. Finally there is the covariance between systematic factor conditioned mean yield and conditioned variance as the systematic factor varies. This covariance could plausibly be of either sign, depending on the nature of the local idiosyncratic randomness.

The paper’s layout is as follows. We first provide an analysis under the assumption that production plots are homogeneous. One deduction is then probed for implications on yield skewness. The case of plot heterogeneity is then addressed. Upon introducing a disaggregated idiosyncratic risk component, an expression relating conditional skewness with unconditional skewness is presented and analyzed. A brief discussion concludes.

Analysis under Yield Aggregation

For any plot of land, we consider three contributions to the yield realization. I): The land could differ innately from other plots because of endowments. This contribution is not random, and we will refer to it as spatial heterogeneity or land heterogeneity. II): There could be a systemic factor, such as weather, that is common to all plots, but realizations could differ from year to year. Then there would be a common yield across plots each year, but this yield would vary from year to year. III): There could be a local idiosyncratic, or LI, factor. This contribution involves different yield realizations for otherwise identical plots in the same year, perhaps because of local weather, pest, or infrastructure problems. In this section it will be assumed that the yield observation is at an aggregated level, perhaps at the townland or county level.

We will first jointly consider systemic (contribution II) and idiosyncratic sources of temporal randomness (contribution III). We will then allow for contribution I, spatial heterogeneities.

Homogeneous Land

Our statistical model of crop yields is as follows. There are n land units of equal area in a defined geographic region. The land units are homogeneous in that they are agronomically identical, face the same systemic weather events, and are farmed in the same way. Yield is random in two senses. There is systemic randomness but also local idiosyncrasies. Firstly, conditional on some 'weather favorability' index $w \in W = [0,1]$, yield on the *i*th land unit is the random realization $y_i(w)$ with support on some interval Y. This is contribution III).

These conditional draws $y_i(w), i \in \{1, \ldots, n\}$, are independent and from the identical conditional distribution $F^{|w}(y_i): Y \to [0,1]$ with mean $\mu(w)$. Here the weather favorability index is given content by the assumption that $\mu(w)$ is a strictly increasing function. Define the inverse function as $w = \mu^{-1}(y)$. Secondly, contribution II is modeled by letting nature draw a w from distribution $G(w)$ where bounded density $g(w)$ exists everywhere on $w \in W$. The following ex ante, or weather unknown, yield distribution is identified in the Appendix.

Proposition 1. Suppose *i*) land is homogeneous with yields given by independently distributed and identically drawn weather-conditioned draws from $F^{|w}(y_i)$, and *ii*) the weather-conditioned mean yield $\mu(w)$ exists and is strictly increasing. Let weather be distributed according to $G(w)$ on $w \in W$ where the density exists everywhere on the support. Then the region's limiting ex ante distribution, as $n \to \infty$, of mean crop yield is represented by $T(y)$ where $T(y) = G[\mu^{-1}(y)]$.

The proposition shows that LI are averaged out under the Strong Law of Large Numbers, leaving only the systemic component. The point here is that non-systemic risks will be drowned out in the face of area-wide systemic weather risks so that the distribution of the systemic component dominates and there is no special role for the normal distribution. One may wonder how robust this conclusion is when crop yields are clearly positively correlated over space. Note first when thinking this through that much if not most of this dependence may not persist after conditioning on a systemic weather factor.

Suppose for the moment one can control for a systemic factor and that there only remain dependent LI. CLT results allow spatial dependence among LI to exist. Indeed, Wang and Zhang (2003) in their study of crop insurance risk pooling opportunities have allowed for finite range dependence, whereby yields are assumed to be independent beyond a defined spatial distance. The variance of mean yield still converges to zero. In their empirical study, Wang and Zhang (2003) used county data for wheat, soybeans, and corn across 2,000 to 2,640 counties in the United States over 1972-1997. They found the distance for any positive dependence to be at most 570 miles and in many cases much less. They did not control for weather, so their distribution is largely due to weather and not to weather-conditioned idiosyncratic risks. Their range conclusions are similar to those found in Goodwin (2001) for Illinois, Indiana, and Iowa corn production, where spatial correlation declines to 0.1 by 200 miles in typical years and by 400 miles in drought years.

Of what relevance is all this when interpreting Proposition 1? That depends on how data are being aggregated. If yield data are from a small area then one can reasonably assume a common w realization for all land at issue. If data are from a larger area then a spatial stochastic process is appropriate. Let $w(i,j)$ and $y(i,j)$ be weather and yield random variables at map coordinate (i,j), $(i,j)\in\{1,\dots,I\}\times\{1,\dots,J\}$, where weather follows the arbitrary joint distribution $G[w(1,1),w(1,2),\dots,w(I,J-1),w(I,J)]$. Then $w(i,j)=\mu^{-1}(i,j)$, and yield follows the joint distribution $G[\mu^{-1}(1,1), \mu^{-1}(1,2), \ldots, \mu^{-1}(I, J-1), \mu^{-1}(I, J)]$. Introducing the spatial dependencies adds a layer of notation to the finding in Proposition 1 but does not change the underlying result.

Local weather variations are a large component of local idiosyncrasies. It is difficult to disentangle the two when presenting a formal model of spatio-temporal yield randomness risk, and even more so as a practical matter. But no matter how one looks at it, the normal distribution should have no special role in understanding how averages are distributed. Suppose that *i*) all practical conditioning problems have been sorted out, including the thorny issue of spatial correlations in weather, and that *ii*) a CLT applies to what is left. Then mean yield will have zero variance at the limit. More positively, while we may not know how to condition well or have the requisite data, the observed distribution should be seen to reflect what has not been conditioned on and not to have been formed by CLT.

Upon reflection, one may conclude that Proposition 1 is an almost trivial (if formal) application of statistical laws. We do not disagree but see its merits on two fronts. Firstly, yield random variables are realizations of an unknown spatio-temporal process where yields at each point in space need further conditioning to account for relevant agronomic attributes. Given the complexity of the context, quite what to condition on can be lost in the mix so that how statistical laws apply warrants formal delineation. In addition the procedure of formal delineation reveals a relationship that is important in understanding the yield distribution, namely, $T(y) = G[\mu^{-1}(y)]$.

Notice that the process of arriving at $T(y)$ in Proposition 1 is an example of that ubiquitous statistical technique, the change of variables. Our paper's second main point arises from a consideration of what effect this change of variables, as the weather variable changes to the yield variable, has on the skewness. For some random variable η, and with $\mathbb{E}[\cdot]$ as the expectation operator, recall that the skewness statistic is

(1) $$\Upsilon(\eta) = \frac{\mathbb{E}\left[\left(\eta - \mathbb{E}[\eta]\right)^3\right]}{\left\{\mathbb{E}\left[\left(\eta - \mathbb{E}[\eta]\right)^3\right]\right\}^{3/2}}.$$

In Theorem 2.2.1 on p. 10 and later remarks on p. 16 of his seminal work on convex transformations of random variables, van Zwet (1964) established the following. Provided the skewness statistics exist, then any transformed random variable $\xi = L(\eta)$ has a smaller skewness statistic than η whenever $L(\eta)$ is increasing and concave. Thus we have

Proposition 2. Make the assumptions of Proposition 1, and also that $\mu(w)$ is concave while the skewness statistics exist. Then yield skewness is smaller than weather index skewness, or

(2) $$\Upsilon[\mu(w)] = \frac{\mathbb{E}\left[\left(\mu(w) - \mathbb{E}[\mu(w)]\right)^3\right]}{\left\{\mathbb{E}\left[\left(\mu(w) - \mathbb{E}[\mu(w)]\right)^2\right]\right\}^{3/2}} \le \frac{\mathbb{E}\left[\left(w - \mathbb{E}[w]\right)^3\right]}{\left\{\mathbb{E}\left[\left(w - \mathbb{E}[w]\right)^2\right]\right\}^{3/2}} = \Upsilon(w).$$

So if mean yield is an increasing and concave function of, say, growing degree days, then mean crop yield is more negatively skewed than is growing degree days. In particular, if growing degree days has zero skewness, then mean crop yield will have negative skewness. To develop some intuition on this, consider the probability mass given to each of weather index interval $[w^l, w^l + \delta]$ and index interval $[w^h, w^h + \delta]$ where $w^h > w^l$ and $\delta > 0$. For $j \in \{l, h\}$, the probability weighting on interval $[w^j, w^j + \delta]$ is transformed to apply over yield interval $[\mu(w^j), \mu(w^j + \delta)]$, where concavity ensures that $\mu(w^l + \delta) - \mu(w^l) > \mu(w^h + \delta) - \mu(w^h)$. The transformation stretches the left tail density weightings over a comparatively larger interval in the yield variable and contracts the right tail weightings. This creates a tendency toward the sort of left-tail to right-tail asymmetry associated with negative skewness. Fig. 1 illustrates.

An alternative means of making the same observation is through differentiating $T(y)$ to obtain density

(3) $$t(y) = g[\mu^{-1}(y)]\frac{d\mu^{-1}(y)}{dy} = \left.\frac{g(w)}{\partial\mu(w)/\partial w}\right|_{w=\mu^{-1}(y)}.$$

As the numerator $\partial\mu(w)/\partial w$ is declining in w, the yield density at the same quantile, or fixing $w = \mu^{-1}(y)$, is increasing in w. In other words, the yield distribution's left tail's density is stretched longer and thinner upon mapping from weather domain to yield domain.

Note that a concave $\mu(w)$ would suggest a cardinal interpretation of the weather favorability index in that expected yield has decreasing marginal product in the index. The weather favorability indices that come to mind, such as growing degree days or the moisture stress index, are cardinal representations of scientific measures. As in Schlenker (2006) and Schlenker, Hanemann, and Fisher (2006), it may be necessary to include a separate index for harmfully high degree days. A second but related issue is whether concavity and monotonicity properties apply. Weather extremes are seldom good for crops suited to a given climatic region. While yield will likely not be monotone in precipitation or temperature, it should be monotone increasing in a well-designed weather favorability index. There remains the issue of concavity. Although using disparate approaches and contexts and although not seeking evidence on concavity, Porter and Seminov (1999, 2005), Schlenker (2006), Almaraz *et al.* (2008) and others find that an increase in inter-annual weather variance reduces average crop yield. This suggests an overall concave shape.

Example 1. Consider a single-input Mitscherlich-Baule production function relationship between mean yield and weather:

(4) $$\mu(w) = \lambda_0 - \lambda_1 e^{-\lambda_2 w},$$

where $w > 0, \lambda_0 \geq \lambda_1 > 0, \lambda_2 > 0$ to ensure a positive, increasing, concave relation. Let w be normally distributed with mean θ and variance σ^2 so that skewness for the weather distribution is zero.[3] As for yield skewness, use of moment generation functions for the normal

[3] As is typical when working with the normal distribution, the left tail problem of w such that

distribution shows that

$$(5) \qquad \Upsilon[\mu(w)] = -\left(e^{\lambda_2^2\sigma^2} - 1\right)^{0.5}\left(e^{\lambda_2^2\sigma^2} + 2\right) < 0,$$

where details are provided in the Appendix. The function is decreasing in both the relative curvature parameter λ_2 and the weather index variance parameter σ^2.

Heterogeneous Land

In this sub-section the model is extended so that land is no longer homogeneous. In particular, land attribute $z \in Z = [0,1]$ follows continuously differentiable distribution $M(z)$. Conditional on any given value of z, there are n land units of equal area in a defined geographic region. Yield on the *i*th land unit is $y_i(w;z)$. These conditional draws are independent and from the identical conditional distribution $F^{|w,z}(y)$ with mean $\mu(w,z)$, which is assumed to be increasing and differentiable in both arguments. Define the attribute-conditioned inverse as $w = h(y;z)$ and the weather-conditioned inverse as $z = r(y;w)$.

Proposition 3. Let *a*) land be heterogeneous with yields given by i.i.d. weather- and yield-conditioned draws from $F^{|w,z}(y)$, and let *b*) the weather and land attribute conditional mean yield $\mu(w,z)$ exist. Let weather and land attributes be distributed according to $G(w): W \to [0,1]$ and $M(z): Z \to [0,1]$, respectively, in each case with densities defined over the entire support. Then, as $n \to \infty$, the region's

i) limiting mean crop yield distribution, ex post for given w, is represented by $L^{|w}(y)$ where $L^{|w}(y) = M[r(y;w)]$.

ii) limiting mean crop yield distribution for a given land attribute value z is represented by $T^{|z}(y)$ where $T^{|z}(y) = G[h(y;z)]$.

$\mu(w) < 0$ is assumed to be negligible and is ignored.

iii) unconditional limiting mean crop yield distribution is represented by $T(y)$ where $T(y) =$

$\int_Z G[h(y;z)]dF(z) = \int_W M[r(y;w)]dG(w)$.

Again, in each case the limiting distribution is that in which the idiosyncratic randomness is ignored. The usual left-tail problem aside, the normal distribution may or may not result. But it has no special place in the analysis.

Relating Conditional and Unconditional Moments

In this section we introduce an LI that is not averaged away, so the scale of analysis is at the small plot level. The marginal distribution for yield at some arbitrarily chosen *i*th location with fixed z attributes is $F(y): Y \rightarrow [0,1]$, where we will henceforth drop the location and z indicators in order to avoid notational clutter. The unconditional mean yield is $\mu =$ $\int_W \mu(w)dG(w)$ and the conditioned deviation in mean yield is $\delta(w) = \mu(w) - \mu$. The residual between *i*th location yield and the conditional mean is $\varepsilon(w) = y - \mu(w)$. This difference arises from LI. Of course, $\int_Y \varepsilon(w)dF^{|w}(y) = 0$. We are interested in comparing the conditional and unconditional expectations of some mean-normalized function

(6) $$S(y-\mu) = S\left[\varepsilon(w) + \delta(w)\right],$$

where the functions of interest will all be of *k*th moment form $\left[\varepsilon(w) + \delta(w)\right]^k$. With $\partial^j S(\cdot)/\partial y^j$ as the *j*th derivative of function $S(\cdot)$, a third-order Taylor's series expansion around $\varepsilon(w) = 0$ identifies

(7) $$S\left[\varepsilon(w)+\delta(w)\right] \approx S(\cdot)\Big|_{\varepsilon(w)=0} + \varepsilon(w)\frac{\partial S(\cdot)}{\partial y}\bigg|_{\varepsilon(w)=0} + \frac{\left[\varepsilon(w)\right]^2}{2}\frac{\partial^2 S(\cdot)}{\partial y^2}\bigg|_{\varepsilon(w)=0} + \frac{\left[\varepsilon(w)\right]^3}{6}\frac{\partial^3 S(\cdot)}{\partial y^3}\bigg|_{\varepsilon(w)=0}.$$

Use the notation for *k*th conditional and unconditional moments:

$$\text{(8)} \quad \begin{aligned} &\text{Conditional:} \quad && C^{(k)}(y \mid w) = \int_Y \left[y - \mu(w)\right]^k dF^{|w}(y); \\ &\text{Unconditional:} \quad && C^{(k)}(y) = \int_W \int_Y \left[y - \mu\right]^k dF^{|w}(y) dG(w) = \int_Y \left[y - \mu\right]^k dF(y). \end{aligned}$$

Noting that $C^{(1)}(y \mid w) = 0$, insert (7) as $[y - \mu]^k$ into the $C^{(k)}(y)$ equation in (8) to establish

$$\text{(9)} \quad \begin{aligned} C^{(k)}(y) \approx & \int_W S\left[\delta(w)\right] dG(w) + \frac{1}{2}\int_W \left.\frac{\partial^2 S(\cdot)}{\partial y^2}\right|_{\varepsilon(w)=0} C^{(2)}(y \mid w) dG(w) \\ & + \frac{1}{6}\int_W \left.\frac{\partial^3 S(\cdot)}{\partial y^3}\right|_{\varepsilon(w)=0} C^{(3)}(y \mid \theta) dG(w). \end{aligned}$$

Now let[4]

$$\text{(10)} \qquad S(y - \mu) = \left[y - \mu(w) + \delta(w)\right]^3,$$

so that the third-order approximation in (7) is exact. Relation (9) becomes

$$\text{(11)} \qquad C^{(3)}(y) = \int_W \left[\delta(w)\right]^3 dG(w) + 3\int_W \delta(w) C^{(2)}(y \mid w) dG(w) + \int_W C^{(3)}(y \mid w) dG(w),$$

as identified in Hennessy (2008). We are interested in better understanding the signs of these three right-hand terms.

Suppose that, conditional on w, there can be only two states of nature where w is assumed to be some beneficial growth factor such as growing degree days. There can be good overall growing conditions in the locality with probability π and yield $y^g(w)$, and also bad overall growing conditions with probability $1-\pi$ and yield $y^b(w)$. Of course, $y^g(w) \geq y^b(w)\ \forall w \in W$ while $\partial y^j(w)/\partial w \geq 0\ \forall w \in W, j \in \{b, g\}$. Some algebra confirms:

$$\text{(12)} \quad \begin{aligned} &\mu(w) = \pi y^g(w) + (1-\pi) y^b(w); \\ &C^{(2)}(y \mid w) = \pi(1-\pi)\left[y^g(w) - y^b(w)\right]^2; \\ &C^{(3)}(y \mid w) = \pi(1-\pi)(1-2\pi)\left[y^g(w) - y^b(w)\right]^3. \end{aligned}$$

---

[4] Conditional and unconditional kurtosis and non-moment statistics can be compared in a similar manner.

Further observations about eqn. (11) are that $\int_W \delta(w)dG(w)=0$, meaning that the well-known covariance decomposition leads to

$$
\begin{aligned}
&\int_W \delta(w)C^{(2)}(y\,|\,w)dG(w) \\
(13)\qquad &= \int_W \delta(w)dG(w)\int_W C^{(2)}(y\,|\,w)dG(w) + \mathrm{Cov}\left[\delta(w), C^{(2)}(y\,|\,w)\right] \\
&= \mathrm{Cov}\left[\mu(w)-\mu, C^{(2)}(y\,|\,w)\right] = \mathrm{Cov}\left[\mu(w), C^{(2)}(y\,|\,w)\right],
\end{aligned}
$$

since constant μ has zero covariance with any other variable.

In light of the points made above, eqn. (11) may be written as

$$
\begin{aligned}
&C^{(3)}(y) = \overbrace{\int_W \left[\delta(w)\right]^3 dG(w)}^{\text{Term I}} \\
(14)\qquad &+\overbrace{3\pi(1-\pi)\mathrm{Cov}\left[\pi y^g(w)+(1-\pi)y^b(w), \left[y^g(w)-y^b(w)\right]^2\right]}^{\text{Term II}} \\
&+\overbrace{\pi(1-\pi)(1-2\pi)\int_W \left[y^g(w)-y^b(w)\right]^3 dG(w)}^{\text{Term III}}.
\end{aligned}
$$

We will look at each term in turn.

<u>Term I</u>. This expression has already been studied in Proposition 2 and the discussions surrounding it. It is the systemic component of the unconditional third central moment.

<u>Term II</u>. By the covariance inequality, Term II is negative whenever one expression is increasing in w while the other is decreasing in w. Now $\partial\mu(w)/\partial w \geq 0$ as $\partial y^j(w)/\partial w \geq 0\ \forall w \in W, \forall j \in \{b,g\}$, and $\pi \in [0,1]$. The second term in the covariance expression has derivative $2\left[y^g(w)-y^b(w)\right]\left\{\partial y^g(w)/\partial w - \partial y^b(w)/\partial w\right\}$. This is negative if the LI and systemic risk factors substitute but positive if they complement. Our present knowledge on crop production theory suggests that both are possible where the crop in question, location, and production practices will factor into determining the sign. Some examples illustrate.

CASE A: If $y^g(w) = y^b(w) + \lambda \; \forall w \in W$ where $\lambda > 0$ then $\partial y^g(w)/\partial w \equiv \partial y^b(w)/\partial w$ and Term II has value zero.

CASE B: If instead $y^g(w) = y^b(w+\lambda) \; \forall w \in W$ where $\lambda > 0$ and $y^b(\cdot)$ is strictly concave then $\partial y^g(w)/\partial w < \partial y^b(w)/\partial w \; \forall w \in W$ and Term II is negative.

CASE C: Finally, consider a special instance of the von Liebig technology as specified in, e.g., Berck and Helfand (1990). Let $y(w,\varepsilon) = \min[w,\varepsilon]$ where $\varepsilon^g > \varepsilon^b$, $\varepsilon = \varepsilon^g$ with probability π, and $\varepsilon = \varepsilon^b$ with probability $1-\pi$. Then[5]

$$y^g(w) = \min[w,\varepsilon^g]; \qquad y^b(w) = \min[w,\varepsilon^b];$$

$$\frac{\partial y^g(w)}{\partial w}\begin{cases} =1 & \text{if } w < \varepsilon^g \\ \in [0,1] & \text{if } w = \varepsilon^g \\ =0 & \text{if } w > \varepsilon^g \end{cases}; \qquad \frac{\partial y^b(w)}{\partial w}\begin{cases} =1 & \text{if } w < \varepsilon^b \\ \in [0,1] & \text{if } w = \varepsilon^b \\ =0 & \text{if } w > \varepsilon^b \end{cases}. \tag{15}$$

Clearly, $\partial y^g(w)/\partial w - \partial y^b(w)/\partial w$ is nowhere strictly negative on $w \in W$ while $\partial y^g(w)/\partial w > \partial y^b(w)/\partial w \; \forall w \in (\varepsilon^b, \varepsilon^g)$ so that Term II is non-negative.

Term III. If good growing conditions are more common, or $\pi > 0.5$, then monotonicity condition $y^g(w) \geq y^b(w) \; \forall w \in W$ ensures that Term III is negative. A comparatively heavy weighting on the good state disposes the LI component of yield toward negative skewness while $\pi < 0.5$ ensures positive skewness.

Figure 2 illustrates Case B above. In it the production functions are the same except that $y^b(w)$ is a rightward translation of $y^g(w)$, parallel to the w axis. Because of concavity, the vertical gap between outputs under the two states declines as weather improves from any index

[5] The $\in [0,1]$ component merely provides bounds on the derivative at the kink point, where the derivative from above is 0 and the derivative from below is 1. Any line through the kink point with slope in $[0,1]$ does not intersect the production function.

value w^l to another value w^h such that $w^h > w^l$. Weather-conditioned mean yield increases with w while conditional variance decreases with w. Somewhat similar to the analysis in Proposition 2 and Fig. 1, introducing LI creates more dispersion on the left tail of the yield distribution, and this is what generates the tendency toward negative skewness. For Case C on the other hand, the LI creates more dispersion on the right tail. In spreading out the right tail realizations, it promotes positive skewness.

A final comment about decomposition (14) is that Term II disappears whenever the conditional variance is independent of the conditioner. This is the case in Figure 1 of Ker and Goodwin (2000), where a mixture of zero skew weather-conditioned yield distributions is represented. The conditioner itself is negatively skewed. This negative skewness, our Term I, ensures negative unconditional yield skewness because Ker and Goodwin do not include an idiosyncratic component (our Term III) in their graphical model.

Conclusion

With the ultimate goal of moving the literature toward developing and testing hypotheses on contexts in which positive or negative yield skewness should be observed, this paper has sought to do three things. Firstly, it has pointed out that while both the Strong Law of Large Numbers and the Central Limit Theorem should indeed apply given adequate approximation of the relevant statistical assumptions, there is a catch. Yield variance, when appropriately conditioned, should recede to zero and the limiting distribution is degenerate. So if systemic heterogeneities exist in the data under consideration, these will dominate to determine the shape of the yield distribution. Secondly, it has identified an effect that tilts the skewness of aggregate yield to be more negative than weather factor skewness whenever the weather factor expresses a positive but diminishing marginal impact on aggregated mean yield. Thirdly, moving to disaggregated yields generates two further effects. The paper shows that both could act to increase or decrease yield skewness at the small plot level of analysis.

The paramount question when conditioning yield distributions is to what end the results will be put. For the purpose of crop insurance, conditioners should be what can be observed at economically low cost by the insurer before contractual agreement. Ideally that will include clay content, water holding capacity, and other agronomic characteristics, as well as pre-season weather variables such as soil moisture or El Niño. The growing literature on yield distributions seeking to account for spatial effects has yet to control for land heterogeneity beyond the use of location indicators, e.g., Ozaki *et al.* (2008). This is understandable because the focus is on other aspects of systemic variation. But improvements in geographical information systems and efforts by yield modelers in the climate change literature suggest that such conditioning should be feasible.

In order to better explain what remains, more attention needs to be paid to how weather variables shape yield distributions. This matters because weather derivative markets may be of use in reassuring crop insurance products (Woodard and Garcia 2008). The connections also matter because, as all our propositions develop upon, these connections can help us better understand yield distribution tails.

Appendix

Relevant Large Number and Limit Laws

Let $P(\cdot)$ be a probability measure on some probability space $(\Omega, \mathcal{S}, P(\cdot))$, where $P(A)$ is the probability attached to condition set $A \subseteq \mathcal{S}$.[6] Our analysis depends crucially upon large number theory, where we choose to emphasize the standard Strong Law. Let $x_1, x_2,\ \ldots$ be independent random variables having the same probability distribution, i.e., they are independent drawn from identical distributions (i.i.d.).

Strong Law of Large Numbers (SLLN) (Lamperti 1996, p. 50). For i.i.d. random variables $x_1, x_2,\ \ldots$ with distribution $F(x_i)$, assume their expected value, $\mu = \int x_i dF(x_i)$, exists. Then

(A1) $$P\left(\mathrm{Lim}_{n\to\infty} \frac{x_1 +\ \ldots\ + x_n}{n} = \mu \right) = 1.$$

Condition (A1) is also commonly referred to as almost sure convergence. Essentially, this condition asserts that the probability that the sample average $n^{-1}\sum_{i=1}^{n} x_i$ does not converge to true mean μ is negligible. The implications are very strong; the law allows for Monte Carlo methods to identify asymptotically exact estimates of the true expectation of some function $f(x)$, i.e., $\mathbb{E}[f(x)] = \mathrm{Lim}_{n\to\infty} n^{-1}\sum_{i=1}^{n} f(x_i)$ where the x_i are independent draws from a common distribution and $\mathbb{E}[\cdot]$ is the expectation operator with respect to that distribution.

While the law establishes what the mean converges to (almost surely), it provides very limited information on how convergence occurs. That is left to the Central Limit Theorem, which describes the distribution a sample mean is likely to follow when the sample size is

[6] See p. 46 of Capiński and Kopp (2004) on probability spaces. The technical details here are not relevant to our analysis but are needed in order to be precise and avoid confusion when invoking widely cited large number results.

large. Distribution being a function, to characterize a Central Limit Theorem one must have a concept of how to measure convergence between functions.

Convergence in Distribution (Bain and Engelhardt 1992, p. 71). Probability measure $\upsilon_n(x)$ is said to converge in distribution to measure $\upsilon(x)$, written as $\upsilon_n \xrightarrow{dist} \upsilon$, if

(A2) $$\text{Lim}_{n\to\infty}\upsilon_n(x) = \upsilon(x)$$

for all values on the domain of x at which $\upsilon(x)$ is continuous.

The normal distribution is continuous everywhere so the issue of continuity is moot.

Central Limit Theorem (CLT) (Lamperti 1996, p. 95). Let $x_1, x_2, \ ...$ be i.i.d. random variables having probability distribution $F(x_i)$. Assume that their expected value, μ, and variance, σ^2, are both finite. Define $\overline{x}_n \equiv (x_1 + \ ... \ + x_n)/n$. Then

(A3) $$P\left(\frac{(\overline{x}_n - \mu)\sqrt{n}}{\sigma} \le x\right) \xrightarrow{dist} \text{Nor}_{(0,1)}(x),$$

where $\text{Nor}_{(\mu,\sigma^2)}(x)$ is the cumulative normal distribution with mean μ and variance σ^2.

Comparing the law and the theorem, they may appear to be inconsistent. The SLLN asserts that the distribution of the mean converges to a distribution with discontinuity at μ, or

(A4) $$P\left(\text{Lim}_{n\to\infty}\overline{x}_n \le x\right) = \begin{cases} 0, & x < \mu; \\ 1, & x \ge \mu. \end{cases}$$

However, divide through by $\sqrt{n}$ and use scaling properties of the normal distribution to consider an alternative version of (A3) as

(A5) $$P\left(\frac{\overline{x}_n - \mu}{\sigma} \le x\right) \xrightarrow{dist} \text{Nor}_{(0,n^{-1})}(x).$$

In this case, as $n \to \infty$ then the variance of the mean contracts toward 0, or $\text{Lim}_{n\to\infty}\sigma^2/n = 0$, and the distribution does indeed become degenerate with a discontinuity at $x = \mu$. Factor $\sqrt{n}$ in (A3) is required to ensure a non-degenerate limiting distribution. It is often called a 'stabilizing' or 'blow-up' transformation (Greene 2003, p. 908).[7]

Proof of Proposition 1. From the SLLN we know that the weather-conditioned mean yield over the region converges to $\mu(w)$ with probability 1, or

(A6) $$P\left(\mathrm{Lim}_{n\to\infty}\hat{y}_n(w) = \mu(w)\right) = 1; \qquad \hat{y}_n(w) = n^{-1}\sum_{i=1}^{n} y_i(w).$$

We will proceed in two steps. The first uses the SLLN to get a weather-conditioned limiting distribution. The second uses the Bounded Convergence Theorem to extend an implication for the unconditional distribution.

Step 1: (Using SLLN). Consider function sequence

(A7) $$u_n(w, y) = \int_Y \int_Y \cdots \int_Y I\left[\hat{y}_n(w) \le y\right] \prod_{i=1}^{n} dF^{|w}(y_i),$$

where $I[C]$ is the indicator function with value 1 if condition C applies and value 0 otherwise. From almost-sure convergence of $\hat{y}_n(w)$ to $\mu(w)$ under SLLN, it follows that the function limit exists and is given by

(A8) $$u(w, y) \equiv \mathrm{Lim}_{n\to\infty} u_n(w, y) = \begin{cases} 1, & \mu(w) \le y; \\ 0, & \mu(w) > y. \end{cases}$$

Step 2: (Using the Bounded Convergence Theorem). This theorem, a corollary of the Dominated Convergence Theorem, gives conditions under which the $\mathrm{Lim}_{n\to\infty}$ and $\int$ operations commute.

Theorem: (Lewin 1987). For a bounded interval $A \subset \mathbb{R}$, suppose the sequence of functions $f_n(x): A \to \mathbb{R}$ satisfies $\int_A |f_n(x)| v(x) dx < \infty$ for $v(x)$ a density function and for all $n \in \mathbb{N}_0$,

[7] In econometric theory, CLT can be used to transform low variance estimator distributions so that standard significance tests can be employed.

$\mathbb{N}_0$ the natural numbers other than 0. Suppose *a*) the sequence converges almost everywhere to a limit function $f(x): A \to \mathbb{R}$, and *b*) there exists a number $\Gamma \in \mathbb{R}$ such that for every $n \in \mathbb{N}_0$, $|f_n(x)|v(x) \le M$ for almost all $x \in A$.[8] Then $\int_A |f(x)|v(x)dx < \infty$, and

(A9) $$\mathrm{Lim}_{n\to\infty} \int_A f_n(x)v(x)dx = \int_A \mathrm{Lim}_{n\to\infty} f_n(x)v(x)dx = \int_A f(x)v(x)dx.$$

In particular, our use of the theorem involves the limiting behavior of $\int u_n(w,y)g(w)dw$ since that provides a description of the limiting yield distribution. Note that

(A10) $$\int_W u_n(w,y)g(w)dw = \int_W \int_Y \int_Y \cdots \int_Y I\left[\hat{y}_n(w) \le y\right] \prod_{i=1}^n dF^{|w}(y_i)g(w)dw = P\left(\bar{y}_n \le y\right),$$

or the unconditional probability that the mean upon drawing from n plots is no more than yield value y. Taking the limit, noting that $|u_n|g(w) \le \Gamma$ to satisfy the theorem's condition *b*), and using step 1 to satisfy the theorem's condition *a*), the Bounded Convergence Theorem implies

(A11) $$\mathrm{Lim}_{n\to\infty} \int_W u_n(w,y)g(w)dw = \int_W u(w,y)g(w)dw = \int_{\mu(w)\le y} g(w)dw = G[\mu^{-1}(y)].$$

But $T(y) = G[\mu^{-1}(y)]$. ■

Details for Example 1. When w is normally distributed with mean μ and variance σ^2, then

(A12) $$\mathbb{E}\left[e^{tw}\right] = e^{t\mu + 0.5t^2\sigma^2}.$$

Therefore,

[8] Here, almost all means almost everywhere or with probability 1. The technical assumption is innocuous in our context.

$$\mathbb{E}\left[\left(\mu(w)-\mathbb{E}[\mu(w)]\right)^3\right]=\mathbb{E}\left[\left(\lambda_0-\lambda_1 e^{-\lambda_2 w}-\mathbb{E}\left[\lambda_0-\lambda_1 e^{-\lambda_2 w}\right]\right)^3\right]$$

$$=-\lambda_1^3\mathbb{E}\left[\left(e^{-\lambda_2 w}-\mathbb{E}\left[e^{-\lambda_2 w}\right]\right)^3\right]$$

(A13) $$=3\lambda_1^3\mathbb{E}\left[e^{-2\lambda_2 w}\right]\mathbb{E}\left[e^{-\lambda_2 w}\right]-2\lambda_1^3\left(\mathbb{E}\left[e^{-\lambda_2 w}\right]\right)^3-\lambda_1^3\mathbb{E}\left[e^{-3\lambda_2 w}\right]$$

$$=3\lambda_1^3 e^{-2\lambda_2\theta+2\lambda_2^2\sigma^2}e^{-\lambda_2\theta+0.5\lambda_2^2\sigma^2}-2\lambda_1^3 e^{-3\lambda_2\theta+1.5\lambda_2^2\sigma^2}-\lambda_1^3 e^{-3\lambda_2\theta+4.5\lambda_2^2\sigma^2}$$

$$=\lambda_1^3 e^{-3\lambda_2\theta+1.5\lambda_2^2\sigma^2}\left(3e^{\lambda_2^2\sigma^2}-2-e^{3\lambda_2^2\sigma^2}\right).$$

Also,

$$\mathbb{E}\left[\left(\mu(w)-\mathbb{E}[\mu(w)]\right)^2\right]=\mathbb{E}\left[\left(\lambda_0-\lambda_1 e^{-\lambda_2 w}-\mathbb{E}\left[\lambda_0-\lambda_1 e^{-\lambda_2 w}\right]\right)^2\right]$$

(A14) $$=\lambda_1^2\mathbb{E}\left[\left(e^{-\lambda_2 w}-\mathbb{E}\left[e^{-\lambda_2 w}\right]\right)^2\right]=\lambda_1^2\mathbb{E}\left[e^{-2\lambda_2 w}\right]-\lambda_1^2\left(\mathbb{E}\left[e^{-\lambda_2 w}\right]\right)^2$$

$$=\lambda_1^2 e^{-2\lambda_2\theta+2\lambda_2^2\sigma^2}-\lambda_1^2 e^{-2\lambda_2\theta+\lambda_2^2\sigma^2}=\lambda_1^2 e^{-2\lambda_2\theta+\lambda_2^2\sigma^2}\left(e^{\lambda_2^2\sigma^2}-1\right).$$

So

(A15) $$\Upsilon[\mu(w)]=\frac{\lambda_1^3 e^{-3\lambda_2\theta+1.5\lambda_2^2\sigma^2}\left(3e^{\lambda_2^2\sigma^2}-2-e^{3\lambda_2^2\sigma^2}\right)}{\left\{\lambda_1^2 e^{-2\lambda_2\theta+\lambda_2^2\sigma^2}\left(e^{\lambda_2^2\sigma^2}-1\right)\right\}^{3/2}}=\frac{3e^{\lambda_2^2\sigma^2}-2-e^{3\lambda_2^2\sigma^2}}{\left(e^{\lambda_2^2\sigma^2}-1\right)^{3/2}}.$$

Set $p=e^{\lambda_2^2\sigma^2}$ and write this as

(A16) $$\Upsilon[\mu(w)]=-\frac{\left(p^3-3p+2\right)}{\left(p-1\right)^{3/2}}=-\frac{\left(p-1\right)^2\left(p+2\right)}{\left(p-1\right)^{3/2}}=-\left(p-1\right)^{0.5}\left(p+2\right)$$

$$=-\left(e^{\lambda_2^2\sigma^2}-1\right)^{0.5}\left(e^{\lambda_2^2\sigma^2}+2\right)<0.$$

Proof of Proposition 3. For parts *i*) and *ii*), simply extend the proof of Proposition 1 to allow for conditioning on z also. For part *iii*), note that

(A17) $$L^{|w}(y)=M[r(y;w)]=\text{Prob}\left(\text{yield}\le y\,|\,w\right),$$

where w has distribution $G(w)$. Then integrate through to obtain the unconditional yield distribution

(A18) $$\int_W M[r(y;w)]dG(w) = \int_W \text{Prob}\left(\text{yield} \le y \mid w\right) dG(w) = \text{Prob}\left(\text{yield} \le y\right).$$

Similarly, $T^{|z}(y) = G[h(y;z)] = \text{Prob}\left(\text{yield} \le y \mid z\right)$ and

(A19) $$\int_Z G[h(y;z)]dF(z) = \int_Z \text{Prob}\left(\text{yield} \le y \mid z\right) dF(z) = \text{Prob}\left(\text{yield} \le y\right).$$

References

Almaraz, J.J., F. Mabood, X. Zhou, E.G. Gregorich, and D.L. Smith. "Climate Change, Weather Variability and Corn Yield at a Higher Latitude Locale: Southwestern Quebec?" *Climatic Change* 88(May, 2008):187–197.

Atwood, J., S. Shaik, and M. Watts. "Can Normality of Yields be Assumed for Crop Insurance?" *Canadian Journal of Agricultural Economics* 50(July, 2002):171–184.

Atwood, J., S. Shaik, and M. Watts. "Are Crop Yields Normally Distributed? A Reexamination." *American Journal of Agricultural Economics* 85(November, 2003):888–901.

Bain, L.J., and M. Engelhardt. *Introduction to Probability and Mathematical Statistics*, 2nd ed. Belmont, CA: Duxbury Press, 1992.

Berck, P., and G. Helfand. "Reconciling the von Liebig and Differentiable Crop Production Functions." *American Journal of Agricultural Economics* 72(November, 1990):985–996.

Capiński, M., and E. Kopp. *Measure, Integral and Probability*, 2nd ed. London: Springer, 2004.

Day, R.H. "Probability Distributions of Field Crop Yields." *Journal of Farm Economics* 47(August, 1965):713–741.

Goodwin, B.K. "Problems with Market Insurance in Agriculture." *American Journal of Agricultural Economics* 83(August, 2001):643–649.

Goodwin, B.K., and O. Mahul. "Risk Modeling Concepts Relating to the Design and Rating of Agricultural Insurance Contracts." Washington DC: World Bank Policy Research Working Paper 3392, September 2004.

Greene, W.H. *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall, 2003.

Hennessy, D.A. "Crop Yield Skewness Under Law of the Minimum Technology." *American Journal of Agricultural Economics*, forthcoming 2008.

Hirshleifer, J. "Risk, the Discount Rate, and Investment Decisions." *American Economic Review* 51(May, 1961):112–120.

Just, R.E., and Q. Weninger. "Are Crop Yields Normally Distributed?" *American Journal of Agricultural Economics* 81(May, 1999):287–304.

Ker, A.P., and B.K. Goodwin. "Nonparametric Estimation of Crop Insurance Rates Revisited." *American Journal of Agricultural Economics* 83(May, 2000):463–478.

Lamperti, John W. *Probability: A Survey of Mathematical Theory*, 2nd ed. New York: John Wiley, 1996.

LeRoy, S.J., and J. Werner. *Principles of Financial Economics*. Cambridge, UK: Cambridge University Press, 2001.

Lewin, J.W. "Some Applications of the Bounded Convergence Theorem for an Introductory Course in Analysis." *American Mathematical Monthly* 94(December, 1987):988–993.

Moss, C.B., and J.S. Shonkwiler. "Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors." *American Journal of Agricultural Economics* 75(November, 1993):1056–1062.

Nelson, C.H., and P.V. Preckel. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71(May, 1989):370–378.

Ozaki, V.A., S.K. Ghosh, B.K. Goodwin, and R. Shirota. "Spatio-Temopral Modeling of Agricultural Yield Data with an Application to Pricing Crop Insurance Contracts." *American Journal of Agricultural Economics*, forthcoming 2008.

Porter, J.R., and M.A. Semenov. "Climate Variability and Crop Yields in Europe." *Nature* 400(August 19, 1999):724.

Porter, J.R., and M.A. Semenov. "Crop Response to Climatic Variation." *Philosophical Transactions of the Royal Society, B-Biological Sciences* 360(November 29, 2005):2021–2035.

Ramirez, O.A., S. Misra, and J. Field. "Crop Yield Distributions Revisited." *American*

Journal of Agricultural Economics 85(February, 2003):108–120.

Schlenker, W. “Inter-annual Weather Variation and Crop Yields.” New York NY: Unpublished Working Paper, Dept. of Economics and School of International and Public Affairs, Columbia University, July 2006.

Schlenker, W., W.M. Hanemann, and A.C. Fisher. “The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions.” *Review of Economics and Statistics* 88(February, 2006):113–125.

Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, and S.H. Irwin. “Crop Insurance Valuation under Alternative Yield Distributions.” *American Journal of Agricultural Economics* 71(May, 2004):406–419.

Swinton, S.M., and R.P. King. “Evaluating Robust Regression Techniques for Detrending Crop Yield Data with Nonnormal Errors.” *American Journal of Agricultural Economics* 73(May, 1991):446–451.

van Zwet, W.R. *Convex Transformations of Random Variables*. Amsterdam: Matematisch Centrum, 1964.Wang, H.H., and H. Zhang. “On the Possibility of a Private Crop Insurance Market: A Spatial Statistics Approach.” *Journal of Risk and Insurance* 70(March, 2003):111–124.

Woodard, J.D., and P. Garcia. “Weather Derivatives, Spatial Aggregation, and Systemic Risk: Implications for Reinsurance Hedging.” *Journal of Agricultural and Resource Economics* 33(April, 2008):34–51.

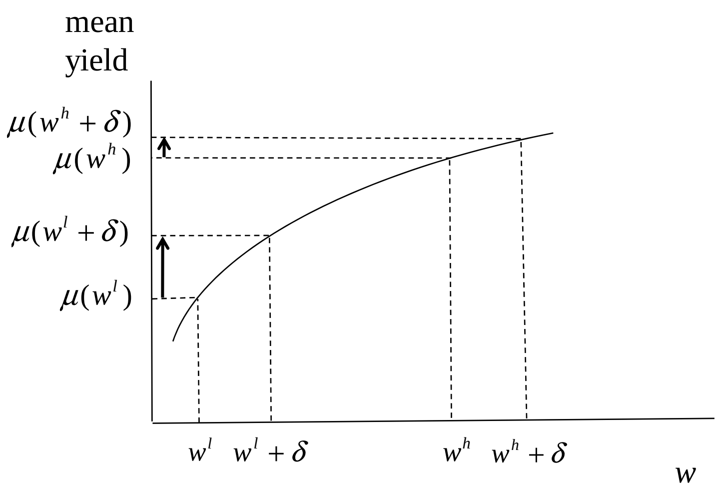

Figure 1. Mapping weather variation to yield variation

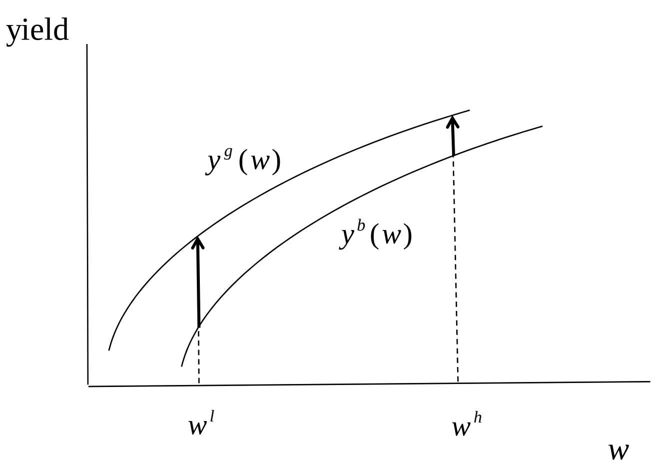

Figure 2. Yield under good and bad local idiosyncrasies