What is probability?

September 12, 2013

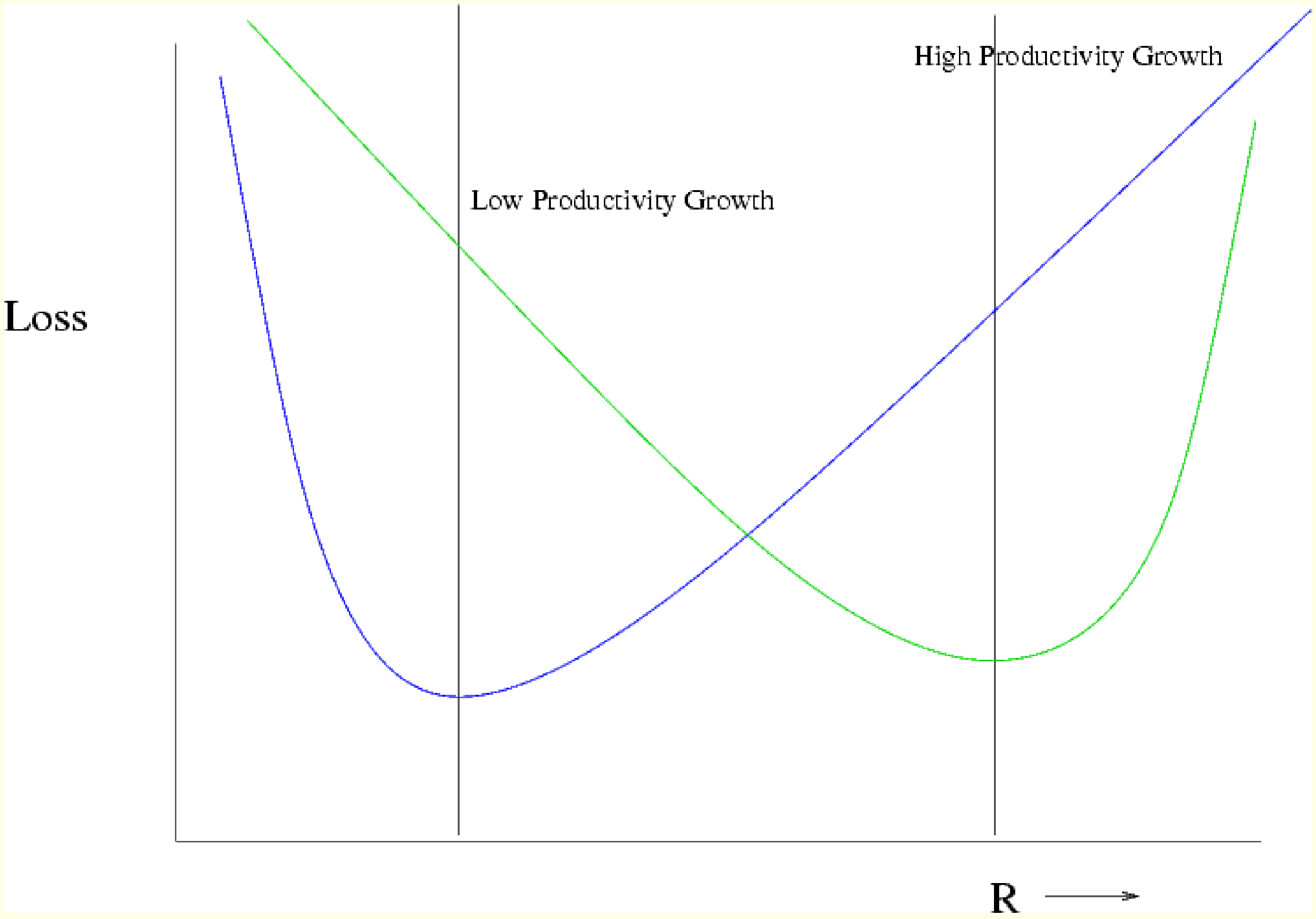
Monetary Policy with Uncertain Productivity Growth

State of the economy in the late 90's:

- unemployment falling, to levels associated with accelerating inflation
- no sign of rising inflation
- productivity growing more rapidly than usual

Policy views:

- Hawks
  - high productivity growth temporary
  - low unemployment makes inflation very likely
  - monetary policy should be restrictive, since its effects are delayed
- Doves
  - high productivity growth sustainable
  - real costs therefore declining
  - inflation therefore likely to remain modest
  - monetary policy should continue to accommodate rapid growth


High Productivity Growth
Low Productivity Growth
Loss
R

Common sense

- no setting of R outside the region bounded by the vertical lines is a good policy.

Common sense

- no setting of R outside the region bounded by the vertical lines is a good policy.
- For each R in outside the region, there is some other R within the region that delivers lower loss no matter which hypothesis is correct.
- policies within the region are "**admissible**", policies outside it "**inadmissible**".
- We should choose an R nearer the right-hand vertical line if we think the hawks are probably right, an R closer to the left-hand one otherwise.
- If we observe new data that tell us something about which hypothesis is true, we should adjust our beliefs about the relative likelihood of the two hypotheses and our choice of R accordingly.

Decision theory

- Formalizes the process of assessing likelihoods and adjusting decisions in the light of evidence.

- Statistical inference is the component that involves adjusting beliefs in the light of newly observed data.

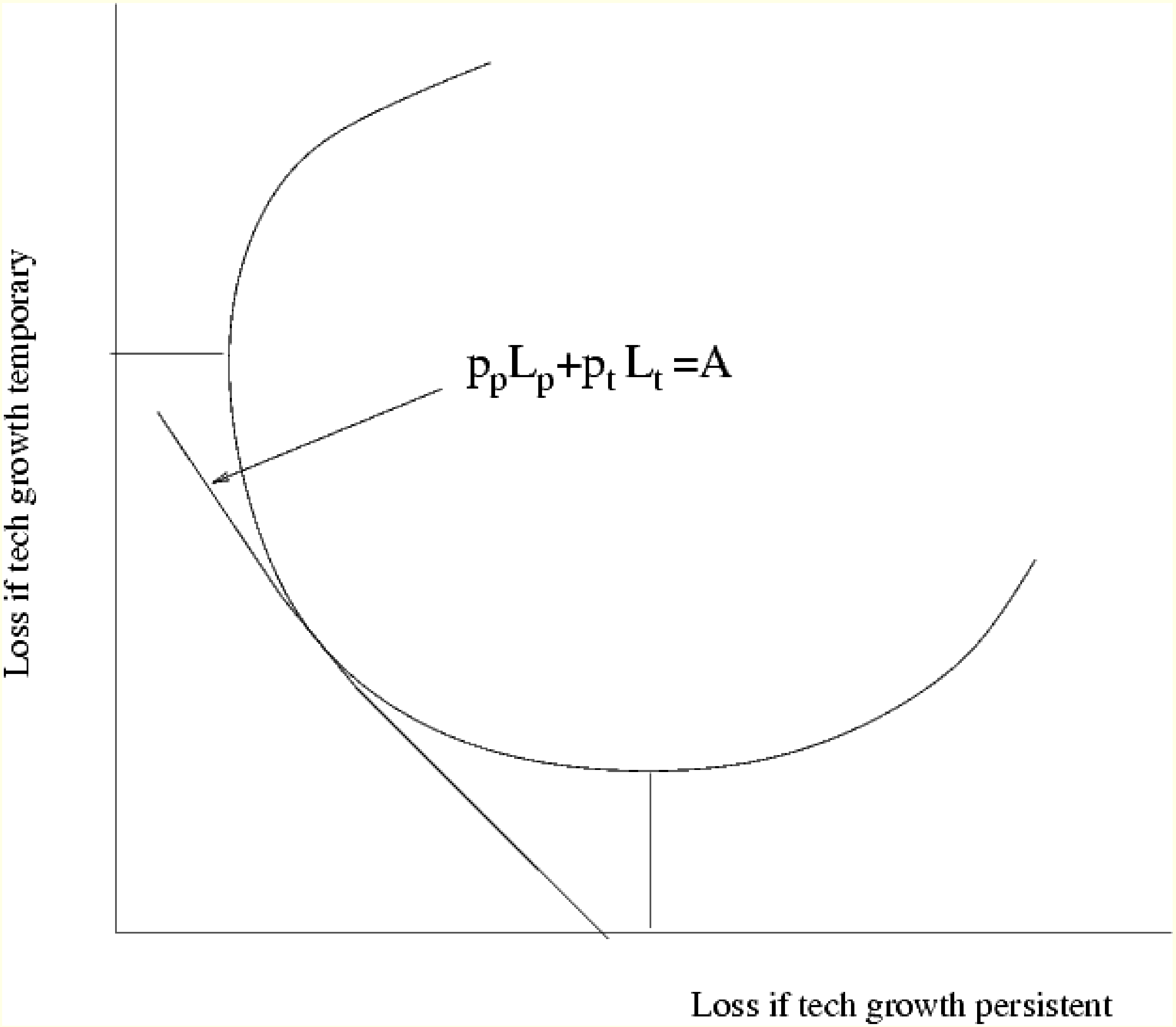
- Good decision making need not formally invoke the mathematics of decision theory.

- Formal decision theory can be helpful in organizing thinking about complex decisions or in facilitating communication in group decision making.

- Formal decision theory can help in criticizing or interpreting decisions that have been arrived at informally.

Probabilities and expectations in a decision problem

We now summarize the information in the graph another way: we plot all the pairs of loss function values available on the graph. Interest rates no longer appear explicitly, though each point on the graph still corresponds to a choice of R.


Loss if tech growth temporary
$p_p L_p + p_t L_t = A$
Loss if tech growth persistent

Probabilities as budget-line-like tangents

- Every one of the admissible points corresponds to a tangent line, like the one that has been drawn in on the figure.

- Such a line will be a linear function of the form $p_p L_p + p_t L_t = A$.

- We can normalize by requiring that $p_p + p_t = 1$.

- The point at which the line touches the curve clearly is the point that minimizes $p_p L_p + p_t L_t$ in the set of available loss pairs.

- p_p and p_t are **probabilities** on the "persistent" and "temporary" (dove and hawk) hypotheses.

- A is the **expectation** of losses for that choice of interest rate and the p_p, p_t probabilities.

- Every choice of R that is not dominated can be described as minimizing expected loss for some choice of probability weights.

- The result here, that admissible decisions can be represented as minimizing expected losses under some set of probability weights, is true under quite general conditions.

- This fact accounts for one main interpretation of probability theory, that probabilities are weights on uncertain prospects that underly optimal decisions.

Probability from arbitrage-free pricing in competitive markets

- S: possible states of the world, or contingencies
- Nsecurities with yield functions $y_j : S \to \mathbb{R}$, $j = 1, \ldots, N$.
- Securities sold on a competitive market.
- Any agent can create new securities with yield functions that are linear combinations of the y_j's. The set of such linear combinations is F.
- $Q : F \to \mathbb{R}$. $Q(z)$ is the market price of the security with yield z.

No-arbitrage conditions $\to$ market expectations

- Linearity: $z = \sum a_j y_j \Rightarrow Q(z) = \sum a_j Q(y_j)$.
- Positivity: $z(\omega) > 0$ for all $\omega \in S \Rightarrow Q(z) > 0$.
- If $z(\omega) \equiv c > 0$ (i.e., z is a risk-free security), we define $\Phi = Q(c)/c$.
- Φ is the risk-free discount factor; and Φ^{-1} the risk-free gross interest rate.
- $E[y] = Q(y)/\Phi$ then has most of the properties of mathematical expectation.

Market expectations $\rightarrow$ market probabilities

- If S is a finite space with M elements, and if $N \geq M$, with the y_j functions linearly independent, we can price, for each $\omega_i \in S$, a security with yield e_i defined as $e_i(\omega_i) = 1$, $e_i(\omega_j) = 0,\ j \neq i$.

- Set $p(\omega_i) = E[e_i]$.

- the M $p(\omega_i)$'s will be non-negative and sum to one, and E will be the expectation operator with respect to the probability defined by these weights. That is, $E[z] = \sum p(\omega_i) z(\omega_i)$.

Market probabilities are not frequencies

The market probability and expectation operator have all the mathematical properties of probability and expectation, but they do not connect to frequencies. That is, e.g., if the same market repeats at many dates t, it is not true that

$$\frac{1}{T}\sum_{t=1}^{T} z_t \to E[z_t]$$

Physical probability: symmetry

- One may examine a coin carefully, weighing it, balancing it, etc., and expect to reach a conclusion as to whether it is "fair" — that is, equally likely to come up heads or tails when flipped properly.

- Decision-makers should not disagree about this. Markets should price a security that delivers $1 for heads the same as one that delivers $1 for tails, if the coin is fair.

- Sometimes we can find a collection of non-overlapping sets for which symmetry arguments are compelling, and we can then construct from them probabilities of more complicated sets.

Physical probability: frequency

- Assume many replicas of the state space S, indexed by $t = 1, \ldots, \infty$.
- We will observe for each t the value of $z_t(\omega)$.
- Supose $\{f_j\}$, $T^{-1}\sum_1^T f_j(z_t)$ always converges to a limit as $T \to \infty$ for some collection $\{f_j\}$.
- Treat the mapping from f_j to this limit as an expectation operator $E[f_j]$.
- As before, we can use the E operator to define a probability function.

Few would disagree with the idea that when the conditions allowing building probability from physical symmetry considerations or from limiting frequencies are met, probabilities should be built that way. Furthermore, in characterizing scientific results it makes sense to maintain a clean distinction between such physically based probabilities and probabilities that do not have such a foundation. On the other hand, in most real-world decision problems most of the uncertainty has to be given weights without any possibility of appeal to such long-run frequencies or physical symmetry.

Some descriptions of the foundations of probability theory seem to imply that the interpretations we have given here are in conflict, as if they are mutually exclusive. In fact, there is no conflict at all between decision-theoretic interpretations of probability and the physical ones. Physical probabilities are, from the decision-theoretic perspective, a special case. Conflict only arises when physical probabilities are claimed to be the only legitimate type of probability.

Market probabilities are also understandable from a decision-theoretic perspective. In a competitive market with rational agents behaving according to the postulates of decision theory, it is possible to derive the form of the market probabilities from knowledge of the probabilities used by the individuals participating in the market (plus knowledge of their budget constraints and utility functions).

Formally defining expectations

- We can start from expectations and derive probabilities, or vice versa. First we start with E.
- S: The "state space" of possible states of the world.
- F: A set of functions $f : S \to W$, where W is a linear space. Such functions are called "random variables".
- Sometimes "random variable" is reserved for the case $W = \mathbb{R}$, "random vector" for $W = \mathbb{R}^n$, "stochastic process" for W consisting of countable sequences of real numbers ($\mathbb{R}^{\mathbb{Z}^+}$), and "continuous time stochastic process" for W consisting of functions over an interval of the real line.

Properties of $E: F \to W$

1. Ef is defined for all $f \in F$ and F is a linear space.

2. If f and g are each in F and a and b are real numbers, $E[af + bg] = aEf + bEg$.

3. If $f(\omega) \geq 0$ for all $\omega \in S$, $Ef \geq 0$.

4. If $\forall(\omega \in S) f(\omega) = c$, $Ef = c$. (Sometimes this is written loosely as "$E[c] = c$". We are also here assuming that the constant function is in F.)

5. If $f_n \in F$ for every n, $f_n(\omega) > 0$ for every n and ω, and if for each $\omega \in S$ $f_n(\omega) \downarrow 0$ monotonically as $n \to \infty$, then $E[f_n] \to 0$ as $n \to \infty$.

In order to connect expectation to probability we need one more condition, which we state here for the case where $W = \mathbb{R}$:

6. For any $f \in F$ the function f^+, defined by

$$f^+(\omega) = \begin{cases} f(\omega) & \text{if } f(\omega) > 0 \\ 0 & \text{if } f(\omega) \leq 0 \end{cases},$$

   is also in F.

Condition (5) is not needed if we restrict ourselves to S with finitely many elements.

Formally defining probability

- $\mathcal{F}$ is a collection of subsets of S that forms a σ-**field**. This means

  1. If A_i is in $\mathcal{F}$ for every $i = 1, \ldots, \infty$, then $\bigcup_i A_i \in \mathcal{F}$.
  2. $S \in \mathcal{F}$.
  3. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$. (A^c is the complement of A in S.)

  Then a probability on $\mathcal{F}$ is a function $P : \mathcal{F} \to [0, 1]$ satisfying

  1. For any $A \in \mathcal{F}$, $P[A] \geq 0$;
  2. $P[S] = 1$;
  3. For any disjoint sets $A, B \in \mathcal{F}$, $P[A \cup B] = P[A] + P[B]$; ($A$ and B disjoint means $A \cap B = \emptyset$.)
  4. If $A_j \in \mathcal{F}$ for every integer j, then

$$P\left[\bigcup_{j=1}^{n} A_j\right] \xrightarrow[n\to\infty]{} P\left[\bigcup_{j=1}^{\infty} A_j\right] .$$

Connections to monetary policy example

- S is "hawks are right" and "doves are right", which we can label as points 0 and 1, say, to save typing.
- Random variables are, e.g., the $\big(L_H(R),\, L_D(R)\big)$ pairs that we generate with different R's. That is, they are functions mapping $\{0,1\}$ to the real line, or in other words just pairs of real numbers.
- F might contain all such pairs (i.e. be $\mathbb{R}^2$) or it might be the linear subspace of $\mathbb{R}^2$ consisting of all pairs (x,y) such that $x=y$. (Can't be the space for which $x=2y$. Why?)
- The most interesting σ-field on S is $\{\emptyset, \{0\}, \{1\}, \{0,1\}\}$, i.e. the set of all subsets of S. There is one other: $\{\emptyset, \{0,1\}\}$.

- The "interesting" σ-field and the $F = \mathbb{R}^2$ case correspond to each other.

- For this case, there must be a $p \in [0, 1]$ such that we give probability $p \geq 0$ to $\{0\}$ and $1 - p$ to $\{1\}$, with the probabilities of other sets following from the rules.

- For this case, $Ef = pf(0) + (1 - p)f(1)$.

- The other case puts probability 1 on $\{0, 1\}$, 0 on $\emptyset$ and, since for $f(0) = f(1)$ for every $f \in F$, $Ef = f(0) = f(1)$.

Proof

- Here are some things related to the monetary policy example (actually any two-element S) that are easy to prove. We're not going to prove them, but making sure you can see how to prove them is a good way to check your own understanding.

- The only σ-fields on S are the two we've listed.

- The only linear spaces F of functions on S that meet the conditions we've placed on F are the two we've listed.

- E and P functions of the form we've described here have all the properties we've asserted for general P and E functions.

- Any P defined on one of the two σ-fields that has the properties we've asserted for a probability function has the form we've claimed.

- Any E defined on one of the two possible F's has the form we've claimed.

Connecting expectations to probabilities

- The general principle: If we have an E operator, we can generate a P by setting $P[A] = E[\mathbf{1}_A]$; if we have a P, we can generate an E from $Ef = \int f(\omega)\, dP(\omega)$

- But what is the $F \leftrightarrow \mathcal{F}$ connection, and what does $\int f(\omega)\, dP(\omega)$ mean? On this whole books have been written — e.g. the Pollard book listed as supplementary material on the reading list.

- In this course we stick to some relatively simple special cases. All the F's we consider will be either the linear space **generated by** the continuous, bounded functions on S, or subspaces of that space.

- All the $\mathcal{F}$'s we will consider will be either the σ-field **generated by** sets of the form $\{\omega \mid f(\omega) \leq a\}$ for some f in F and some real number a, or sub-sigma-fields of that one.

- The idea of a linear space generated by a collection of elements should be familiar. A σ-field generated by a collection of sets is similar — the smallest σ-field containing all the specified sets.

- Another way to say this is that we are always going to be talking about probabilities in the context of some collection of random variables with defined expectations, and that we always want to be able to put probabilities on events (subsets of S) of the form $a < f(\omega) \leq b$, i.e. on the event that a certain random variable lies in a certain interval on the real line.

Interpreting the integral

We will not try to develop a general interpretation of the integral (known as a Lebesgue integral). For now, note that when S has a finite or countable number of points, the j'th point in S has a probability p_j and

$$\int_S f(\omega)\, dP(\omega) = \sum_{\omega_j \in S} p_j f(\omega_j)\ , \tag{1}$$

while if S is $\mathbb{R}^n$, the most common special case is one in which the probability has a **density function** p and we can write

$$\int_S f(\omega)\, dP(\omega) = \int \int \ldots \int p(\omega) f(\omega)\, d\omega_1\, d\omega_2\, \ldots\, d\omega_n\ ,$$

where what appears on the right *is* an ordinary Riemann integral.

More generally, we can mix together probabilities spread smoothly over $\mathbb{R}^n$ and probabilities that concentrate on lower-dimensional subsets or subspaces of $\mathbb{R}^n$.

For example: Consider the joint distribution of the minimum wage in the state of employment and the actual wage for a person drawn at random from a sample of fast-food workers.