Coherently using AI forecasting in statistical estimation

Glenn Shafer

July 9, 2026

ITQM, Rouen

1. Test forecaster (or forecasting strategy) by betting against it.
2. If it passes, use its forecasts (or strategies) to make predictions.
3. Use predictions to make inferences about the world.

1. Test forecaster (or forecasting strategy) by betting against it.
2. If it passes, use its forecasts (or strategies) to make predictions.
3. Use predictions to make inferences about the world.

Unifies AI prediction with statistical inference.

- Forecaster may be an artificial intelligence.
- Strategy may be a probability measure.

ERCIM News 145
July 2026

https://ercim-news.ercim.eu

Includes my article "On testing by imaginary betting".

Cournot’s principle

An event with very small probability will not happen.

An event with probability close to one will happen.

Cournot's principle does a lot of work.

1. We predict events that have very high probability.

2. We use the happening of events with very small probabilities to discredit theories.

An event with very small probability will not happen.

Antoine Augustin Cournot
1801-1877

The only way probability theory makes contact with the world.

Testing by betting allows us to

- discipline Cournot's principle,
- extend it to AI forecasts.

The diversity of probability forecasting

Numerical forecasts can be produced by…

- statistical models with estimated parameters
- physical models (hurricane forecasting)
- neural networks
- financial market
- seat of pants or whatever (financial analyst)

The diversity of probability forecasting

Forecast may be

- a probability (e.g. for rain)
- an estimate (e.g. for size of dividend)

Each forecast may be on a different topic.

We can always test by betting.

Forecaster can be theory, person, or AI.

He may…

- state probabilities (offering to sell payoffs for their expected values) or
- make more limited betting offers.

You choose which offers to accept.

The bets are imaginary!

Factor by which you multiply initial unit capital

=

betting score or *e-value*.

You can make successive bets, using the capital from one bet to make the next.

Capital after multiple rounds

=

cumulative betting score.

Replace Cournot’s principle with five principles about testing by betting.

1. Principle for testing by betting.
2. Principle for multiple discredit.
3. Principle for event prediction.
4. Principle for point prediction.
5. Principle for statistical estimation.

1. Principle for testing by betting

If you make successive bets against Forecaster,

- beginning with unit capital and
- never risking additional capital,

and you obtain a large cumulative betting score,
then you discredit Forecaster.

2. Principle for multiple discredit

If you simultaneously test multiple forecasters who forecast the same outcome, obtaining a large betting score against each forecaster you test, then you have discredited all the forecasters.

3. Principle for event prediction

You may predict that a forecaster who has consistently withstood certain test strategies in the past will similarly withstand a similar test strategy in the future.

4. Principle for point prediction

If you can replicate a payoff using bets Forecaster has withstood in the past, the cost of replication predicts the payoff.

5. Principle for statistical estimation

A warranted prediction about outcomes not observed may further warrant the inference that related unknowns are consistent with the prediction.

Example

You make predictions about errors of a measuring instrument based on experience testing the instrument to measure known quantities.

You do not observe the errors you are predicting.

But your prediction that the average error is within certain bounds warrants the inference that the quantity you are measuring is within certain bounds.

Let’s look at the math

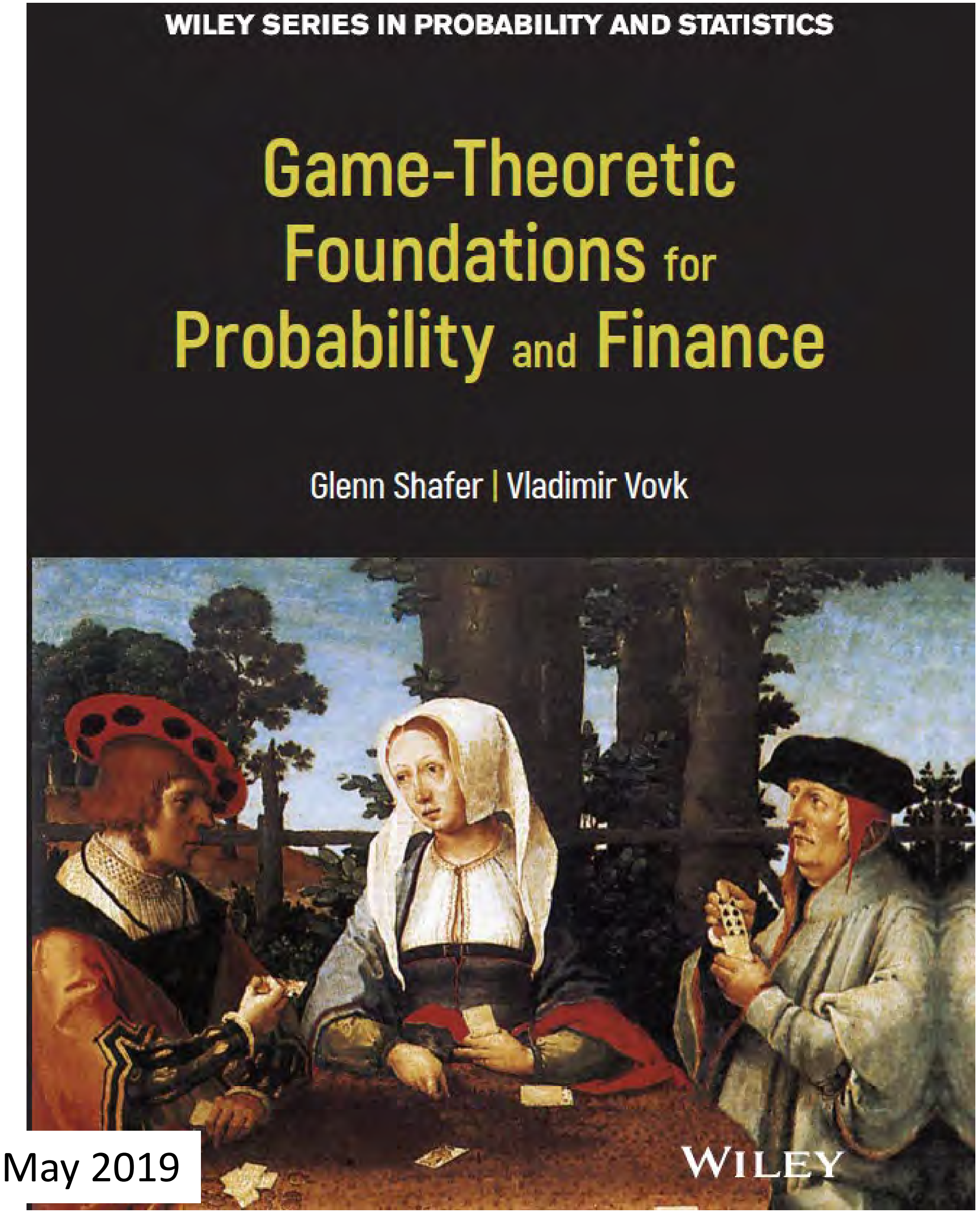
Game-Theoretic Statistics

Glenn Shafer

Statistical modelling is the art of connecting abstract models with scientific or practical questions, so that the models' probabilities throw light on these questions. Standard statistical theory summarizes the results with p-values. Unfortunately, p-values are widely misunderstood. Game-theoretic statistics replaces them with e-values, also called betting scores. The model makes a forecast, the statistician makes an imaginary bet against it, and the e-value is the factor by which she multiplies her stake.

Written at secondary-school level

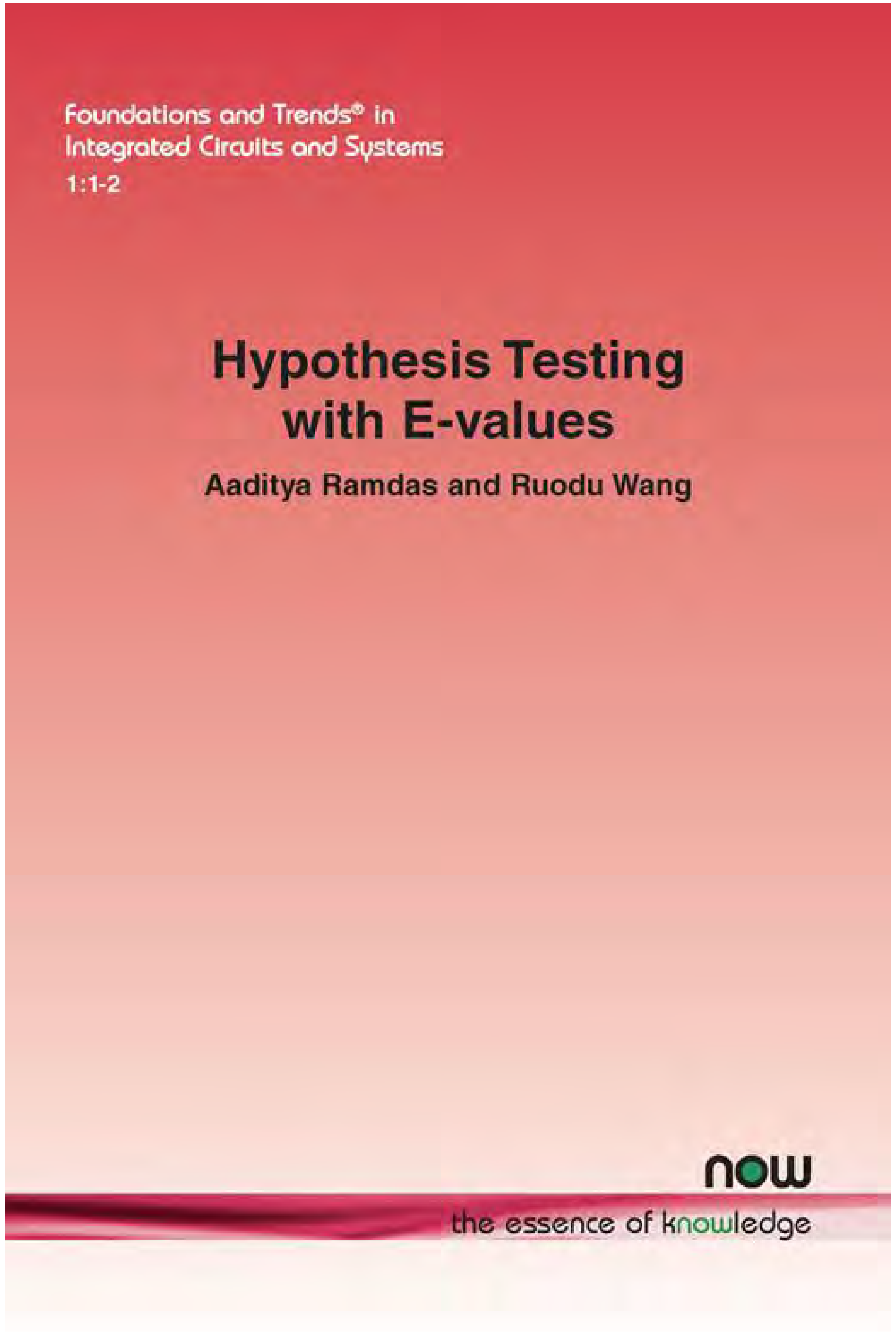
https://www.probabilityandfinance.com/misc/snapshot.pdf



May 2019

Bases **mathematical probability** on testing by betting.

Working papers at
www.probabilityandfinance.com



Available on arXiv

E-value

=

factor by which you multiply your stake betting against a forecaster.

Working papers at

www.probabilityandfinance.com

https://alrw.net/e/

https://alrw.net/

The Game

The game

Initial capital

PLAYERS: Skeptic, Forecaster, Reality
PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$

$s_0 := 1$.
For $n = 1, \ldots, N$:
Forecaster announces probability distribution P_n on $\mathcal{Y}$.
Skeptic announces nonnegative variable Z_n on $\mathcal{Y}$
such that $\mathbf{E}_{P_n}(Z_n) = s_{n-1}$.
Reality announces $y_n \in \mathcal{Y}$.
$s_n := Z_n(y_n)$.

Final capital

Skeptic bets with play money.

Betting score = s/s_0

PLAYERS: Skeptic, Forecaster, Reality
PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$

$s_0 := 1$.
For $n = 1, \ldots, N$:
Forecaster announces probability distribution P_n on $\mathcal{Y}$.
Skeptic announces nonnegative variable Z_n on $\mathcal{Y}$
such that $\mathbf{E}_{P_n}(Z_n) = s_{n-1}$.
Reality announces $y_n \in \mathcal{Y}$.
$s_n := Z_n(y_n)$.

The game is in Statistician's imagination.
She tells Skeptic how to bet.
Sometimes she tells Forecaster how to forecast.

The statistician can work inside the game.

PLAYERS: Skeptic, Experimenter, Forecaster, Reality

Skeptic announces $s_0 \in \mathbb{R}$.

For $n = 1, 2, \ldots$:

Experimenter announces probability space $\mathcal{Y}_n$.

Forecaster announces probability distribution P_n on $\mathcal{Y}_n$.

Skeptic announces variable Z_n on $\mathcal{Y}_n$ with finite $\mathbf{E}_{P_n}(Z_n)$.

Reality announces $y_n \in \mathcal{Y}_n$.

$s_n := s_{n-1} + Z_n(y_n) - \mathbf{E}_{P_n}(Z_n)$.

Regression

PLAYERS: Skeptic, Informant, Reality, Forecaster
PARAMETERS: Set $\mathcal{X}$, probability space $\mathcal{Y}$

Skeptic announces $s_0 \in \mathbb{R}$.
Informant announces $x \in \mathcal{X}$.
Forecaster announces probability distribution P on $\mathcal{Y}$.
Skeptic announces variable Z on $\mathcal{Y}$ with finite $\mathbf{E}_P(Z)$.
Reality announces $y \in \mathcal{Y}$.
$s := s_0 + Z(y) - \mathbf{E}_P(Z)$.

To test, Skeptic must keep his capital non-negative.

PLAYERS: Skeptic, Forecaster, Reality
PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$

$s_0 := 1$.
For $n = 1, \ldots, N$:
Forecaster announces probability distribution P_n on $\mathcal{Y}$.
Skeptic announces nonnegative variable Z_n on $\mathcal{Y}$
such that $\mathbf{E}_{P_n}(Z_n) = s_{n-1}$.
Reality announces $y_n \in \mathcal{Y}$.
$s_n := Z_n(y_n)$.

Betting score

A betting score against a probability distribution is a likelihood ratio.

- A bet S is a function of Y satisfying $S \geq 0$ and $\sum_y S(y)P(y) = 1$.
- So SP is also a probability distribution. Call it the alternative Q.
- But $Q(y) = S(y)P(y)$ implies $S(y) = Q(y)/P(y)$.
- A bet against P defines an alternative Q and the betting score $S(y)$ is the likelihood ratio $Q(y)/P(y)$.

You say P describes Y.

I want to bet against you.

I think Q describes Y.

Should I use Q/P as my bet?

$S = Q/P$ maximizes $\mathbf{E}_Q(\ln S)$.

$$\mathbf{E}_Q\left(\ln \frac{Q(Y)}{P(Y)}\right) \geq \mathbf{E}_Q\left(\ln \frac{R(Y)}{P(Y)}\right) \forall R$$

Gibbs's inequality

Why maximize $\mathbf{E}_Q(\ln S)$? Why not $\mathbf{E}_Q(S)$? Or $Q(S \geq 20)$?

Neyman-Pearson lemma

When S is the product of successive factors, $\mathbf{E}(\ln S)$ measures the rate of growth (Kelly, 1956). This has been used in gambling theory, information theory, finance theory, and machine learning. Here it opens the way to a theory of multiple testing and meta-analysis.

Three examples of hypothesis testing

Example 1.

Result statistically and practically significant but hopelessly contaminated with noise.

$$P\colon\ Y \sim \mathcal{N}(0, 10)$$
$$Q\colon\ Y \sim \mathcal{N}(1, 10)$$
$$y = 30$$

$$P\colon\ Y \sim \mathcal{N}(0, 10)$$

$$Q\colon\ Y \sim \mathcal{N}(1, 10)$$

$$y = 30$$

- p-value: $P(Y \geq 30) \approx 0.00135$.
- 5% test rejects when $y \geq 16.445$. Power 6%.
- Bet Q/P has implied target 1.005. Betting score is $S(30) \approx 1.34$.

- Power and implied target agree: study is worthless.
- But Neyman-Pearson rejects with low p-value, while betting score sees that evidence is slight.

Example 2.

Test with α = 5% and high power rejects with borderline outcome even though likelihood ratio favors null.

$$P\colon\ Y \sim \mathcal{N}(0, 10)$$

$$Q\colon\ Y \sim \mathcal{N}(37, 10)$$

$$y = 16.5$$

P: $Y \sim \mathcal{N}(0, 10)$

Q: $Y \sim \mathcal{N}(37, 10)$

$y = 16.5$

- p-value: $P(Y \geq 16.5) \approx 0.0495$.
- 5% test rejects when $y \geq 16.445$.
  Power 98%.
- Bet Q/P has implied target 939.
  Betting score is $S(16.5) \approx 0.477$.

- Power and implied target agree: study is good.
- Neyman-Pearson rejects.
  Betting score says evidence slightly favors null.

Example 3.

High p-value is interpreted as evidence for null.

$$P\colon Y \sim \mathcal{N}(0, 10)$$

$$Q\colon Y \sim \mathcal{N}(20, 10)$$

$$y = 5$$

$$P\colon\ Y \sim \mathcal{N}(0, 10)$$
$$Q\colon\ Y \sim \mathcal{N}(20, 10)$$
$$y = 5$$

- p-value: $P(Y \geq 5) \approx 0.3085$.
- 5% test rejects when $y \geq 16.445$. Power 64%.
- Bet Q/P has implied target 7.39. Betting score is $S(5) \approx 0.368$.

- Power and implied target agree: study is marginal.
- Neyman-Pearson simply does not reject. Betting score says evidence slightly favors null.

Coherently using AI forecasting in statistical estimation. The traditional framework for mathematical statistics, launched by Laplace and Gauss in the early 19th century and generalized by the British school of statistics in the early 20th century, makes forecasting (a.k.a. prediction) subsequent and secondary to estimation. You create a model, parametric, nonparametric, or semiparametric. Then you estimate the parameters from random samples. Then you might (or might not) use the estimated model for forecasting. The forecasting step does not work well for complicated models, and during the past half century, the advent of better performing machine learning models has made this painfully obvious. Statisticians are now realizing that they should reverse the order of estimation and forecasting: AI forecasting should be used in statistical estimation. We see this philosophy in conformal prediction (especially in the 2022 edition of Vovk et al.'s *Algorithmic Learning in a Random World*) and in Angelopoulos et al.'s prediction-powered inference (*Science*, 2023). These frameworks remain traditional, insofar as (X,Y) pairs are assumed to be probabilistically independent or exchangeable. Game-theoretic statistics replaces this traditional assumption with the notion of testing by betting.