Uncertainty Quantification
Lecture 2: UQ in Public Policy

l'école doctorale--Winter School 2017
5--8 February 2017, Les Diablerets, Switzerland

Philip B. Stark, www.stat.berkeley.edu/~stark
Department of Statistics, University of California, Berkeley

Quantifauxcation

Assign a meaningless number, then pretend that since it's quantitative, it's meaningful.

Quantifauxcation

Assign a meaningless number, then pretend that since it's quantitative, it's meaningful.

Usually involves some combination of data, pure invention, ad hoc models, inappropriate statistics, and logical lacunae.

Quantifauxcation

Assign a meaningless number, then pretend that since it's quantitative, it's meaningful.

Usually involves some combination of data, pure invention, ad hoc models, inappropriate statistics, and logical lacunae.

In my opinion, *most* "probabilities" in policy and most cost-benefit analyses involve quantifauxcation.

Cost-Benefit analyses

Widely touted as the only rational basis for decisions: must quantify costs/risks/benefits.

But if there's no rational basis for quantitative inputs, can it be rational to insist on the analysis?

Not all "costs" can be put on a common scale. Some are incommensurable. Multidimensional scales cannot always be well ordered.

The cost of most policy cost-benefit analyses is high: lost rationality

Risk = probability × consequences?

Catchy slogan, but:

- What if "probability" doesn't apply to the phenomenon? (more below)
- What if consequences cannot be quantified on a one-dimensional scale?

Insisting on quantifying risk and on quantitative cost-benefit analyses requires putting a price on human life, on biodiversity, on relics, …

How do you incorporate uncertainty in probability (if it applies at all) and uncertainty in consequences?

What is Probability?

Axiomatic aspect and philosophical aspect.

- Kolmogorov's axioms:
  - "just math"
  - triple (S, Ω, P)
    - S a set
    - Ω a sigma-algebra on S
    - P a non-negative countably additive measure with total mass 1

What is Probability?

Axiomatic aspect and philosophical aspect.

- Kolmogorov's axioms:
    - "just math"
    - triple (S, Ω, P)
        - S a set
        - Ω a sigma-algebra on S
        - P a non-negative countably additive measure with total mass 1
- Philosophical theory that ties the math to the world
    - What does probability *mean*?
    - Standard theories
        - Equally likely outcomes
        - Frequency theory
        - Subjective theory
    - Probability models as empirical commitments
    - Probability as metaphor

How does probability enter a scientific problem?

- underlying phenomenon is essentially random (radioactive decay, thermodynamics)

- deliberate randomization (randomized experiments, random sampling)

- subjective probability
  - Constraints versus priors
  - No posterior distributions without prior distributions
  - Prior generally matters
  - elicitation issues
  - arguments from consistency, "Dutch book," ...
  - why should I care about your subjective probability

- invented model that's supposed to describe the phenomenon
  - in what sense?
  - to what level of accuracy?
  - description v. prediction v. predicting effect of intervention
  - testable to desired level of accuracy?

- metaphor: phenomenon behaves "as if random"

Two very different situations:

- Scientist creates randomness by taking a random sample, assigning subjects at random to treatment or control, etc.
- Scientist invents (assumes) a probability model for data the world gives.

(1) allows sound inferences.

Inferences drawn in (2) are only as good as the assumptions.

Gotta check the assumptions against the world: Empirical support? Plausible? Iffy? Absurd?

Making sense of probabilities in applied problems is hard

- Probability often applied without thinking
- Reflexive way to try to represent uncertainty
- Not all uncertainty can be represented by a probability
- "Aleatory" versus "Epistemic"

- Aleatory
    - Canonical examples: coin toss, die roll, lotto, roulette
    - under some circumstances, behave "as if" random (but not perfectly)

- Epistemic: stuff we don't know

- Standard way to combining aleatory variability epistemic uncertainty puts beliefs on a par with an unbiased physical measurement w/ known uncertainty.

    - Claims by introspection, can estimate without bias, with known accuracy, just as if one's brain were unbiased instrument with known accuracy
    - Bacon's triumph over Aristotle should put this to rest, but empirically:
        - people are bad at making even rough quantitative estimates
        - quantitative estimates are usually biased
        - bias can be manipulated by anchoring, priming, etc.
        - people are bad at judging weights *in their hands*: biased by shape & density
        - people are bad at judging when something is random
        - people are overconfident in their estimates and predictions
        - confidence unconnected to actual accuracy.
        - anchoring effects entire disciplines (e.g., Millikan, c, Fe in spinach)

- what if I don't trust your internal scale, or your assessment of its accuracy?

- same observations that are factored in as "data" are also used to form beliefs: the "measurements" made by introspection are not independent of the data

Rates versus probabilities

- In a series of trials, if each trial has the same probability p of success, and if the trials are independent, then the rate of successes converges (in probability) to p. Law of Large Numbers

- If a finite series of trials has an empirical rate p of success, that says nothing about whether the trials are random.

- If the trials are random *and* have the same chance of success, the empirical rate is an estimate of the chance of success.

- If the trials are random *and* have the same chance of success *and* the dependence of the trials is known (e.g., the trials are independent), can quantify the uncertainty of the estimate.

Thought experiments

You are one of a group of 100 people. You learn that one will die in the next year.

What's the chance it is you?

Thought experiments

You are one of a group of 100 people. You learn that one will die in the next year.

What's the chance it is you?

You are one of a group of 100 people. You learn that one is named "Philip."

What's the chance it is you?

Thought experiments

You are one of a group of 100 people. You learn that one will die in the next year.

What's the chance it is you?

You are one of a group of 100 people. You learn that one is named "Philip."

What's the chance it is you?

Why does the first invite an answer, and the second not?

Ignorance ≠ Randomness

Cargo Cult Confidence Intervals

- Have a collection of numbers, e.g., MME climate model predictions of warming
- Take mean and standard deviation.
- Report mean as the estimate; construct a confidence interval or "probability" statement from the results, generally using Gaussian critical values
- IPCC does this
- Case study: Accelerating extinction paper

Cargo Cult Confidence Intervals

- Have a collection of numbers, e.g., MME climate model predictions of warming
- Take mean and standard deviation.
- Report mean as the estimate; construct a confidence interval or "probability" statement from the results, generally using Gaussian critical values
- IPCC does this
- Case study: Accelerating extinction paper

What's wrong with it?

- No random sample; no stochastic errors.
- Even if there were a random sample, what justifies using normal theory?
- Even if random and normal, misinterprets confidence as probability. Garbled; something like Fisher's fiducial inference
- Ultimately, quantifauxcation.

Random versus haphazard/unpredictable

- Consider taking a sample of soup to tell whether it is too salty.
  - Stirring the soup, then taking a tablespoon, gives a random sample
  - Sticking in a tablespoon without looking gives a haphazard sample

Random versus haphazard/unpredictable

- Consider taking a sample of soup to tell whether it is too salty.
    - Stirring the soup, then taking a tablespoon, gives a random sample
    - Sticking in a tablespoon without looking gives a haphazard sample

- Tendency to treat haphazard as random
    - random requires deliberate, precise action
    - haphazard is sloppy/ignorant

Random versus haphazard/unpredictable

- Consider taking a sample of soup to tell whether it is too salty.
  - Stirring the soup, then taking a tablespoon, gives a random sample
  - Sticking in a tablespoon without looking gives a haphazard sample

- Tendency to treat haphazard as random
  - random requires deliberate, precise action
  - haphazard is sloppy/ignorant

- Notions like probability, p-value, confidence intervals, etc., *apply only if the sample is random* (or for some kinds of measurement errors)

  - Do not apply to samples of convenience, haphazard samples, etc.

  - Do not apply to populations.

Some brief examples

- Avian / wind-turbine interactions
- Earthquake probabilities
- Climate models and climate change probabilities

Wind power: "avian / wind-turbine interactions"

Wind turbines kill birds, notably raptors.

- how many, and of what species?
- how concerned should we be?
- what design and siting features matter?
- how do you build/site less lethal turbines?

Measurements

Periodic on-the-ground surveys, subject to:

- censoring
- shrinkage/scavenging
- background mortality
- is this pieces of two birds, or two pieces of one bird?
- how far from the point of injury does a bird land? attribution...

Measurements

Periodic on-the-ground surveys, subject to:

- censoring
- shrinkage/scavenging
- background mortality
- is this pieces of two birds, or two pieces of one bird?
- how far from the point of injury does a bird land? attribution...

Is it possible to ...

- make an unbiased estimate of mortality?
- quantify the uncertainty of the estimate?
- reliably relate the mortality to individual turbines in wind farms?

Stochastic model

Common: Mixture of a point mass at zero and some distribution on the positive axis.
E.g., "Zero-inflated Poisson"

Countless alternatives, e.g.:

- observe $\max\{0, \text{Poisson}(\lambda_j) - b_j\}$, $b_j > 0$
- observe $b_j \times \text{Poisson}(\lambda_j)$, $b_j \in (0, 1)$.
- observe true count in area j with error ϵ_j, where $\{\epsilon_j\}$ are dependent, not identically distributed, nonzero mean

Consultant

- bird collisions random, Poisson distributed
- same for all birds
- independent across birds
- rates follow hierarchical Bayesian model that depends on covariates: properties of site and turbine design

Consultant

- bird collisions random, Poisson distributed
- same for all birds
- independent across birds
- rates follow hierarchical Bayesian model that depends on covariates: properties of site and turbine design

What does this mean?

- when a bird approaches a turbine, it tosses a coin to decide whether to throw itself on the blades
- chance coin lands heads depends on site and turbine design
- all birds use the same coin for each site/design
- birds toss their coins independently

Where do the models come from?

- Why random?
- Why Poisson?
- Why independent from site to site? From period to period? From bird to bird? From encounter to encounter?
- Why doesn't chance of detection depend on size, coloration, groundcover, …?
- Why do different observers miss carcasses at the same rate?
- What about background mortality?

Complications at Altamont

- Why is randomness a good model? Random is not the same as haphazard or unpredictable.
- Why is Poisson in particular reasonable? Do birds in effect toss coins, independently, with same chance of heads, every encounter with a turbine? Is `#encounters` $\times P(\text{heads})$ constant?
- Why estimate the parameter of a contrived model rather than actual mortality?
- Do we want to know how many birds die, or the value of λ in an implausible stochastic model?
- Background mortality—varies by time, species, etc.
- Are all birds equally likely to be missed? Smaller more likely than larger? Does coloration matter?
- Nonstationarity (seasonal effects—migration, nesting, etc.; weather; variations in bird populations)
- Spatial and seasonal variation in shrinkage due to groundcover, coloration, illumination, etc.
- Interactions and dependence.
- Variations in scavenging. (Dependence on kill rates? Satiation? Food preferences? Groundcover?)
- Birds killed earlier in the monitoring interval have longer time on trial for scavengers.
- Differences or absolute numbers? (Often easier to estimate differences accurately.)
- Same-site comparisons across time, or comparisons across sites?

The Rabbit Axioms

1. For the number of rabbits in a closed system to increase, the system must contain at least two rabbits.
2. No negative rabbits.

The Rabbit Axioms

1. For the number of rabbits in a closed system to increase, the system must contain at least two rabbits.
2. No negative rabbits.

Freedman's Rabbit-Hat Theorem

You cannot pull a rabbit from a hat unless at least one rabbit has previously been placed in the hat.

The Rabbit Axioms

1. For the number of rabbits in a closed system to increase, the system must contain at least two rabbits.

2. No negative rabbits.

Freedman's Rabbit-Hat Theorem

You cannot pull a rabbit from a hat unless at least one rabbit has previously been placed in the hat.

Corollary

You cannot "borrow" a rabbit from an empty hat, even with a binding promise to return the rabbit later.

Applications of the Rabbit-Hat Theorem

- Can't turn a rate into a probability without assuming the phenomenon is random in the first place.
- Can't conclude that a process is random without making assumptions that amount to assuming that the process is random. (Something has to put the randomness rabbit into the hat.)
- Testing whether the process appears to be random using the *assumption* that it is random cannot prove that it is random. (You can't borrow a rabbit from an empty hat.)

Earthquake probabilities

- Probabilistic seismic hazard analysis (PSHA) is the basis for seismic building codes in many countries; basis for siting nuclear power plants
- Models earthquakes as random in space, time, magnitude; independent magnitudes
- Models ground motion as random, given the occurrence of an event. Distribution in a particular place depends on the location and magnitude of the event.
- Claim to estimate "exceedance probabilities": chance acceleration exceeds some threshold in some number of years
- In U.S.A., codes generally require design to withstand accelerations w probability ≥2% in 50y.
- PSHA arose from probabilistic risk assessment (PRA) in aerospace and nuclear power.
  Those are engineered systems whose inner workings are known but for some system parameters and inputs.
- Inner workings of earthquakes are almost entirely unknown: PSHA is based on metaphors and heuristics, not physics.

The PSHA equation

Model earthquake occurrence as a marked stochastic process with known parameters.

Model ground motion in a given place as a stochastic process, given the quake location and magnitude.

Then,

> probability of a given level of ground movement in a given place is the integral (over space and magnitude) of the conditional probability of that level of movement given that there's an event of a particular magnitude in a particular place, times the probability that there's an event of a particular magnitude in that place

- That earthquakes occur at random is an *assumption* not based in theory or observation.
- involves taking rates as probabilities
  - Standard argument:
    - M = 8 events happen about once a century.
    - Therefore, the chance is about 1% per year.

Earthquake casinos

- Models amount to saying there's an "earthquake deck"
- Turn over one card per period. If the card has a number, that's the size quake you get.
- Journals and journals full of arguments about how many "8"s in the deck, whether the deck is fully shuffled, whether cards are replaced and re-shuffled after dealing, etc.

Earthquake casinos

- Models amount to saying there's an "earthquake deck"
- Turn over one card per period. If the card has a number, that's the size quake you get.
- Journals and journals full of arguments about how many "8"s in the deck, whether the deck is fully shuffled, whether cards are replaced and re-shuffled after dealing, etc.
- but this is just a metaphor!

Earthquake terrorism

- Why not say earthquakes are like terrorist bombings?
  - don't know where or when
  - know they will be large enough to kill
  - know some places are "likely targets"
  - but no probabilities
- What advantage is there to the casino metaphor?

Rabbits and Earthquake Casinos

What would make the casino metaphor apt?

1. the physics of earthquakes might be stochastic. But it isn't.
2. stochastic models might provide a compact, accurate description of earthquake phenomenology. But it doesn't.
3. stochastic models might be useful for predicting future seismicity. But it isn't (Poisson, Gamma renewal, ETAS)

3 of the most destructive recent earthquakes were in regions seismic hazard maps showed to be relatively safe (2008 Wenchuan M7.9, 2010 Haiti M7.1, & 2011 Tohoku M9) Stein, Geller, & Liu, 2012

What good are the numbers?

Climate models

IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties, 2010

https://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf (at p.2)

> … quantified measures of uncertainty in a finding expressed probabilistically (based on statistical analysis of observations or model results, or expert judgment).
>
> … Depending on the nature of the evidence evaluated, teams have the option to quantify the uncertainty in the finding probabilistically. In most cases, author teams will present either a quantified measure of uncertainty or an assigned level of confidence.
>
> … Because risk is a function of probability and consequence, information on the tails of the distribution of outcomes can be especially important. … Author teams are therefore encouraged to provide information on the tails of distributions of key variables …

Cargo-cult confidence common in IPCC work:

As mentioned above

- have a list of numbers, not a sample from anything and certainly not a random sample
- take mean and SD
- treat as if random sample from Normal distribution
- confuse confidence with probability
- garble interpretation, using something like Fisher's fiducial inference

Result is gibberish.

Do Monte Carlo simulations estimate real-world probabilities?

Monte Carlo is a way to substitute computing for calculation.

It does not reveal anything that was not already an assumption in the calculation.

The distribution of the output results from the assumptions in the input.

The randomness in the formulation is an assumption, not a conclusion; the distribution of that randomness is an assumption, not a conclusion.

Does "expert judgment" reveal probability?

Recall from above:

- Claims by introspection, can estimate without bias, with known accuracy, just as if one's brain were unbiased instrument with known accuracy.
- But empirically,
    - people are bad at making even rough quantitative estimates
    - quantitative estimates are usually biased
    - bias can be manipulated by anchoring, priming, etc.
    - people are bad at judging weights *in their hands*: biased by shape & density
    - people are bad at judging when something is random.
    - people are overconfident in their estimates and predictions
    - confidence unconnected to actual accuracy.
    - anchoring effects entire disciplines (e.g., Millikan, c, Fe in spinach)
- what if I don't trust your internal scale, or your assessment of its accuracy?
- same observations that are factored in as "data" are also used to form beliefs: the "measurements" made by introspection are not independent of the data.

Rhodium Group Climate Prospectus

"Risky Business" project co-chaired by Michael R. Bloomberg, Henry Paulson, Tom Steyer.

Funded by Bloomberg Philanthropies, the Paulson Institute, and TomKat Charitable Trust. Also Skoll Global Threats Fund, Rockefeller Family Fund, McKnight Foundation, Joyce Foundation.

"While our understanding of climate change has improved dramatically in recent years, predicting the severity & timing of future impacts remains a challenge. Uncertainty surrounding the level of GHG emissions going forward & the sensitivity of the climate system to those emissions makes it difficult to know exactly how much warming will occur, & when. Tipping points, beyond which abrupt & irreversible changes to the climate occur, could exist. Due to the complexity of the Earth's climate system, we do not know exactly how changes in global average temperatures will manifest at a regional level. There is considerable uncertainty about how a given change in temperature, precipitation, or sea level will impact different sectors of the economy, & how these impacts will interact."

& yet, …

"In this American Climate Prospectus, we aim to provide decision-makers in business & in government with the facts about the economic risks & opportunities climate change poses in the United States."

They estimate the effect of changes in temperature and humidity on mortality, a variety of crops, energy use, labor force, and crime, **at the county level** through 2099.

"evidence-based approach."

Figure 8.3: Climate impact on heat and cold-related mortality

RCP 8.5 median

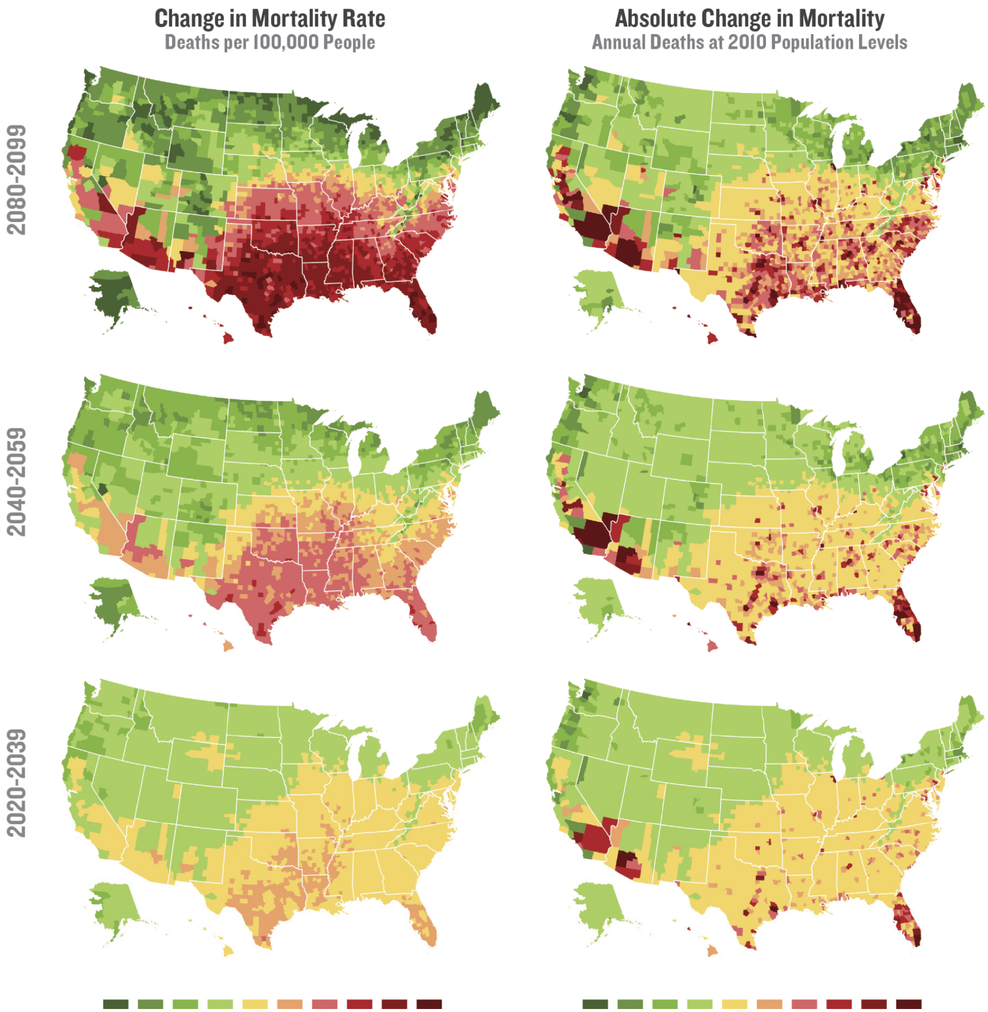


Figure 6.4: Projected change in grain, oilseed, and cotton yields by county

RCP 8.5 median projection, grey counties are those where no grain, oilseed, or cotton production currently occurs

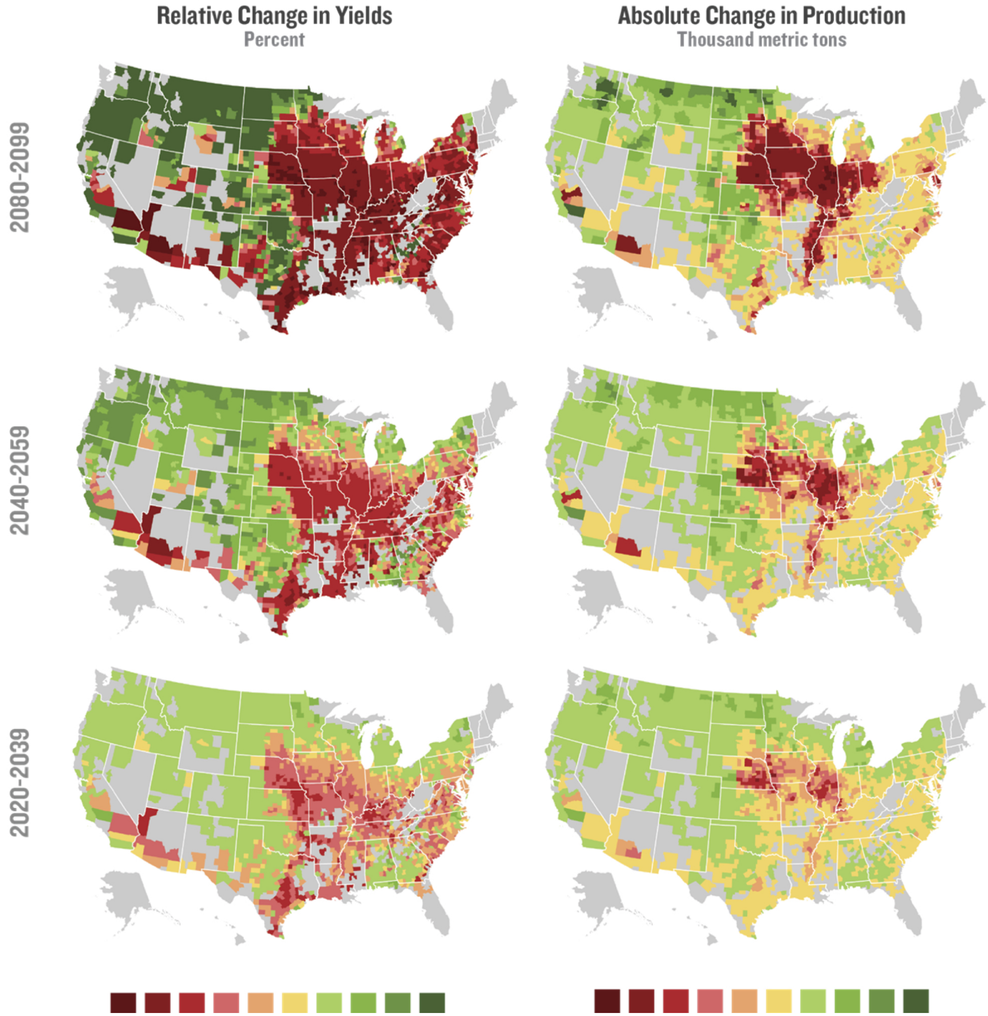


Figure 7.5: Change in labor productivity in “high-risk” sectors of the economy

RCP 8.5 median

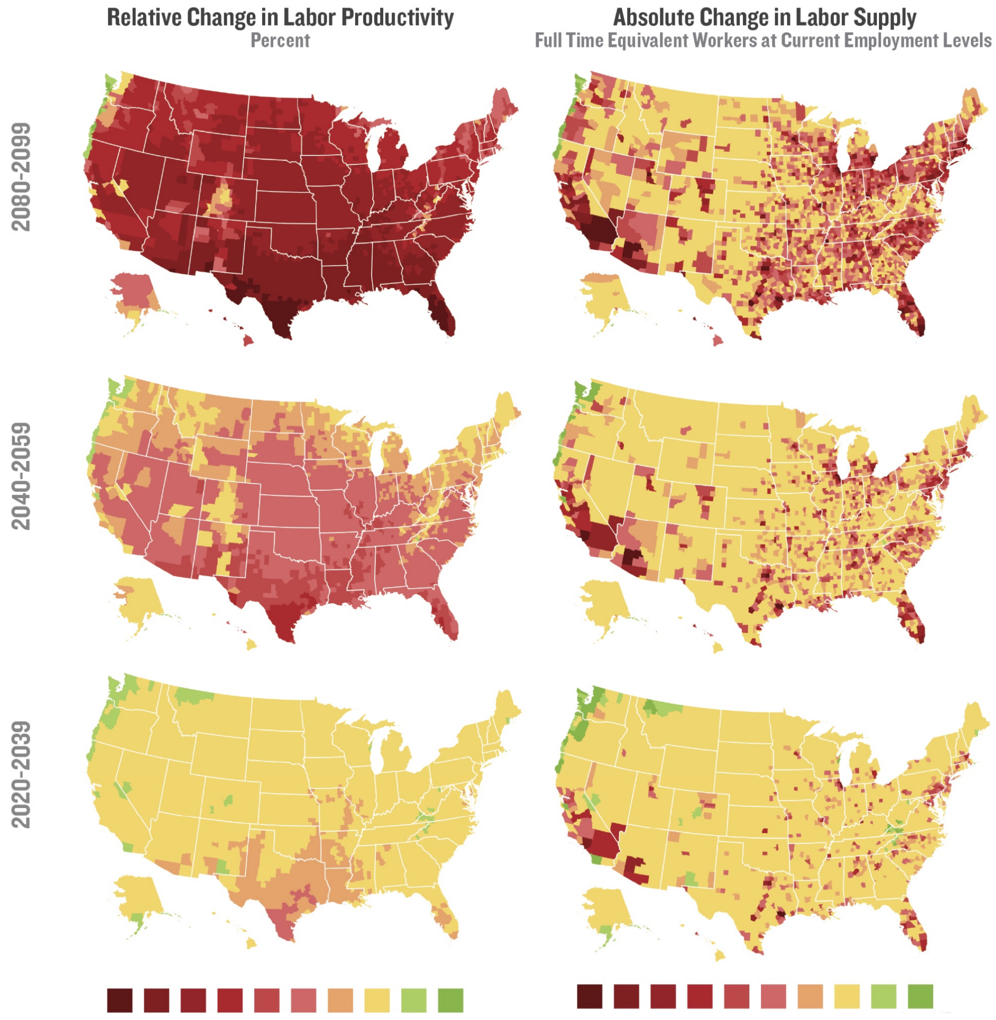
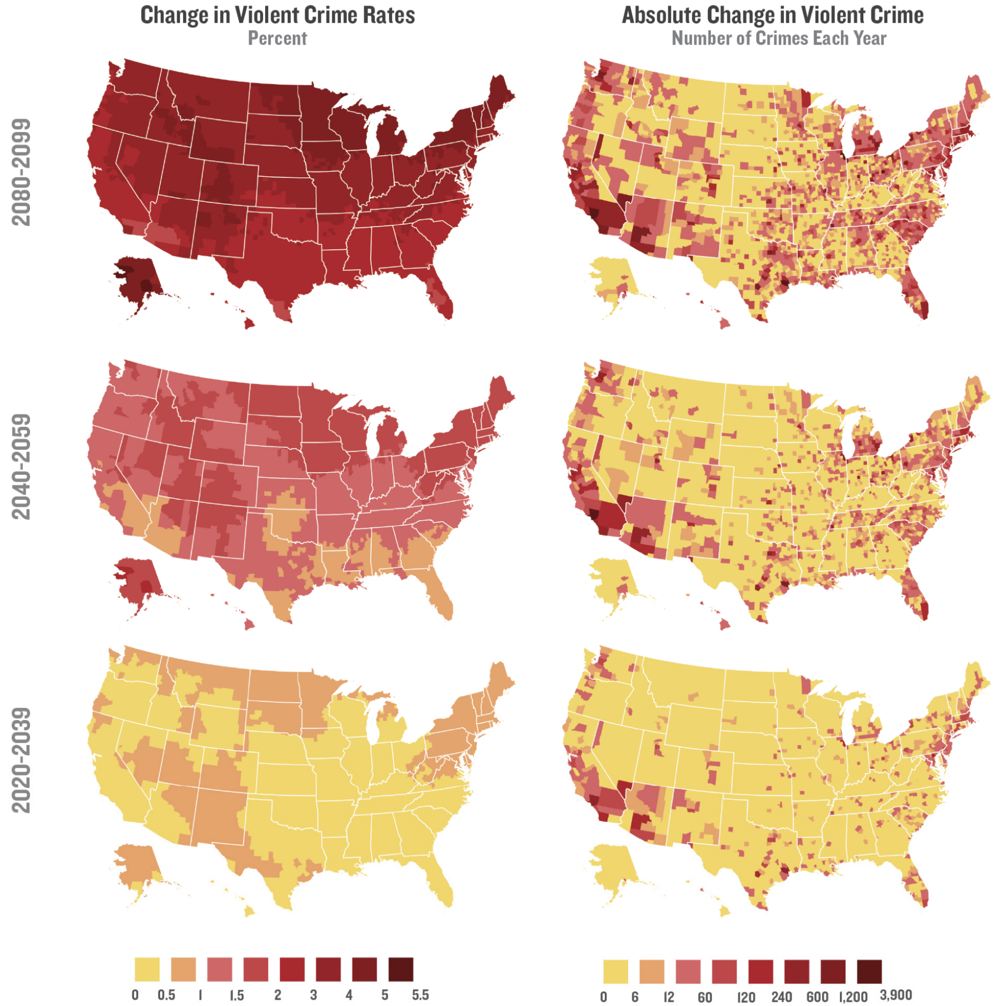


Figure 9.5: Change in violent crime

RCP 8.5, median



Sanity check

Even the notion that if you *knew exactly* what the hourly temperature and humidity will be in every square meter of the globe for the next hundred years, you could accurately predict the effect on violent crime—on any timescale or any spatial scale—is patently absurd.

Adding the uncertainty in temperature and humidity obviously doesn't make the problem easier.

Sanity check

Even the notion that if you *knew exactly* what the hourly temperature and humidity will be in every square meter of the globe for the next hundred years, you could accurately predict the effect on violent crime—on any timescale or any spatial scale—is patently absurd.

Adding the uncertainty in temperature and humidity obviously doesn't make the problem easier.

Ditto for labor, crop yields, mortality, etc.

Sanity check

Even the notion that if you *knew exactly* what the hourly temperature and humidity will be in every square meter of the globe for the next hundred years, you could accurately predict the effect on violent crime—on any timescale or any spatial scale—is patently absurd.

Adding the uncertainty in temperature and humidity obviously doesn't make the problem easier.

Ditto for labor, crop yields, mortality, etc.

And that's even if *ceteris* were *paribus*—which they will not be.

Sanity check

Even the notion that if you *knew exactly* what the hourly temperature and humidity will be in every square meter of the globe for the next hundred years, you could accurately predict the effect on violent crime—on any timescale or any spatial scale—is patently absurd.

Adding the uncertainty in temperature and humidity obviously doesn't make the problem easier.

Ditto for labor, crop yields, mortality, etc.

And that's even if *ceteris* were *paribus*—which they will not be.

And that's even for next year, much less for 80 years from now.

Sanity check

Even the notion that if you *knew exactly* what the hourly temperature and humidity will be in every square meter of the globe for the next hundred years, you could accurately predict the effect on violent crime—on any timescale or any spatial scale—is patently absurd.

Adding the uncertainty in temperature and humidity obviously doesn't make the problem easier.

Ditto for labor, crop yields, mortality, etc.

And that's even if *ceteris* were *paribus*—which they will not be.

And that's even for next year, much less for 80 years from now.

And that's even for predicting a global average, not county-level predictions.

Sanity check

Even the notion that if you *knew exactly* what the hourly temperature and humidity will be in every square meter of the globe for the next hundred years, you could accurately predict the effect on violent crime—on any timescale or any spatial scale—is patently absurd.

Adding the uncertainty in temperature and humidity obviously doesn't make the problem easier.

Ditto for labor, crop yields, mortality, etc.

And that's even if *ceteris* were *paribus*—which they will not be.

And that's even for next year, much less for 80 years from now.

And that's even for predicting a global average, not county-level predictions.

Clearly, the uncertainty is total: this is not science.