Discovering and Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts

Gary King

Institute for Quantitative Social Science
Harvard University

(University of Florida, 3/18/2016)

---

Based on joint work with Konstantin Kashin and Samir Soneji
GaryKing.org
Systematic Bias and Nontransparency in US Social Security Administration Forecasts

*Journal of Economic Perspectives*
References

- Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *Journal of Economic Perspectives*

- Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *Political Analysis*
References

- Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *Journal of Economic Perspectives*

- Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *Political Analysis*

- Articles, data, code available at GaryKing.org/ssa
The Essential Role of Forecasting in the US Government

Social Security
- Single largest U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings 20% of elderly Americans above poverty level
- Enormously popular
- Proposals for change: highly controversial, partisan, cross-cutting, and personal — the "third rail of American politics"

Payroll taxes ➔ Trust Funds (now ≈ $2.8T) ➔ beneficiaries

SSA demographic and financial forecasts:
- under factual conditions, used to evaluate solvency
- under counterfactual conditions, used to score policy proposals

Other Programs that Rely on SSA Forecasts
- Medicare & Medicaid Trust Funds; CBO evaluations, etc.
- Programs comprising >50% of US government expenditures
Social Security

- Single largest U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings 20% of elderly Americans above poverty level
- Enormously popular
- Proposals for change: highly controversial, partisan, cross-cutting, and personal — the "third rail of American politics"

Payroll taxes $\Rightarrow$ Trust Funds (now $\approx$ $2.8T) \Rightarrow$ beneficiaries

SSA demographic and financial forecasts:
- Under factual conditions, used to evaluate solvency
- Under counterfactual conditions, used to score policy proposals

Other Programs that Rely on SSA Forecasts:
- Medicare & Medicaid Trust Funds; CBO evaluations, etc.

- Programs comprising > 50% of US government expenditures
Social Security

- Single largest U.S. government program
The Essential Role of Forecasting in the US Government

- **Social Security**
  - Single largest U.S. government program
  - 37% of federal outlays ($1.3T in 2013 expenditures)
The Essential Role of Forecasting in the US Government

**Social Security**
- Single **largest** U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings **20%** of elderly Americans above poverty level
Social Security

- Single largest U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings 20% of elderly Americans above poverty level
- Enormously popular

Proposals for change: highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”

Payroll taxes → Trust Funds (now ≈ $2.8T) → beneficiaries

SSA demographic and financial forecasts:
- Under factual conditions, used to evaluate solvency
- Under counterfactual conditions, used to score policy proposals

Other Programs that Rely on SSA Forecasts

Medicare & Medicaid Trust Funds; CBO evaluations, etc.

Programs comprising > 50% of US government expenditures
The Essential Role of Forecasting in the US Government

- **Social Security**
  - Single largest U.S. government program
  - 37% of federal outlays ($1.3T in 2013 expenditures)
  - Brings 20% of elderly Americans above poverty level
  - Enormously popular
  - Proposals for change: highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”

- Payroll taxes ➜ Trust Funds (now ≈ $2.8T) ➜ beneficiaries
- SSA demographic and financial forecasts:
  - Under factual conditions, used to evaluate solvency
  - Under counterfactual conditions, used to score policy proposals

- Other Programs that Rely on SSA Forecasts
  - Medicare & Medicaid Trust Funds; CBO evaluations, etc.
  - Programs comprising > 50% of US government expenditures
**The Essential Role of Forecasting in the US Government**

- **Social Security**
  - Single **largest** U.S. government program
  - 37% of federal outlays ($1.3T in 2013 expenditures)
  - Brings **20%** of elderly Americans above poverty level
  - Enormously **popular**
  - **Proposals for change:** highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”
  - Payroll taxes $\leadsto$ **Trust Funds** (now $\approx$2.8T) $\leadsto$ beneficiaries
The Essential Role of Forecasting in the US Government

- **Social Security**
  - Single *largest* U.S. government program
  - 37% of federal outlays ($1.3T in 2013 expenditures)
  - Brings 20% of elderly Americans above poverty level
  - Enormously popular
  - Proposals for change: highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”
  - Payroll taxes \(\leadsto\) Trust Funds (now \(\approx\)$2.8T) \(\leadsto\) beneficiaries
  - SSA demographic and financial forecasts:
Social Security

- Single largest U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings 20% of elderly Americans above poverty level
- Enormously popular
- Proposals for change: highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”
- Payroll taxes $\rightarrow$ Trust Funds (now $\approx$ $2.8T) $\rightarrow$ beneficiaries
- SSA demographic and financial forecasts:
  - under factual conditions, used to evaluate solvency
The Essential Role of Forecasting in the US Government

Social Security

- Single largest U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings 20% of elderly Americans above poverty level
- Enormously popular
- Proposals for change: highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”
- Payroll taxes ↦ Trust Funds (now ≈$2.8T) ↦ beneficiaries
- SSA demographic and financial forecasts:
  - under factual conditions, used to evaluate solvency
  - under counterfactual conditions, used to score policy proposals
Social Security

- Single largest U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings 20% of elderly Americans above poverty level
- Enormously popular
- Proposals for change: highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”
- Payroll taxes ⇝ Trust Funds (now ≈$2.8T) ⇝ beneficiaries
- SSA demographic and financial forecasts:
  - under factual conditions, used to evaluate solvency
  - under counterfactual conditions, used to score policy proposals

Other Programs that Rely on SSA Forecasts
Social Security

- Single **largest** U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings **20%** of elderly Americans above poverty level
- Enormously **popular**
- **Proposals for change:** highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”
- Payroll taxes \(\rightsquigarrow\) Trust Funds (now \(\approx\)$2.8T) \(\rightsquigarrow\) beneficiaries
- SSA demographic and financial forecasts:
  - under factual conditions, used to evaluate solvency
  - under counterfactual conditions, used to score policy proposals

Other Programs that Rely on SSA Forecasts

- Medicare & Medicaid Trust Funds; CBO evaluations, etc.
The Essential Role of Forecasting in the US Government

**Social Security**
- Single largest U.S. government program
- 37% of federal outlays ($1.3T in 2013 expenditures)
- Brings 20% of elderly Americans above poverty level
- Enormously popular
- Proposals for change: highly controversial, partisan, cross-cutting, and personal — the “third rail of American politics”
- Payroll taxes $\rightarrow$ Trust Funds (now $\approx$ $2.8T) $\rightarrow$ beneficiaries
- SSA demographic and financial forecasts:
  - under factual conditions, used to evaluate solvency
  - under counterfactual conditions, used to score policy proposals

**Other Programs that Rely on SSA Forecasts**
- Medicare & Medicaid Trust Funds; CBO evaluations, etc.
- $\rightarrow$ Programs comprising $\geq 50\%$ of US government expenditures
Nontransparency in Social Security Forecasting

Who forecasts independently of SSA's Office of the Chief Actuary?
No one

Who has been able to fully replicate OCACT's forecasts?
No one

Some data shared: in difficult, disorganized, non-automated formats
Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters

Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

Nontransparency and lack of data sharing violates:
repeated, emphatic calls from SSA's Technical Advisory Panels
Executive Orders requiring "a presumption in favor of openness," data that's "accessible, discoverable, and usable by the public"
the data sharing revolution in academia

Enormous missed opportunity: for the scientific community and others to check and improve SSA forecasts (for free); but easy to fix!
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

Nontransparency and lack of data sharing violates:
- repeated, emphatic calls from SSA’s Technical Advisory Panels
- Executive Orders requiring “a presumption in favor of openness,” data that’s “accessible, discoverable, and usable by the public”
- the data sharing revolution in academia

Enormous missed opportunity: for the scientific community and others to check and improve SSA forecasts (for free); but easy to fix!
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one

Some data shared: in difficult, disorganized, non-automated formats
Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

Nontransparency and lack of data sharing violates:
- repeated, emphatic calls from SSA’s Technical Advisory Panels
- Executive Orders requiring “a presumption in favor of openness,” data that’s “accessible, discoverable, and usable by the public”
- the data sharing revolution in academia

Enormous missed opportunity: for the scientific community and others to check and improve SSA forecasts (for free); but easy to fix!
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one
  - Some data shared: in difficult, disorganized, non-automated formats
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one
  - Some data shared: in difficult, disorganized, non-automated formats
  - Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one
  - Some data shared: in difficult, disorganized, non-automated formats
  - Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
  - Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

Nontransparency and lack of data sharing violates:

- Repeated, emphatic calls from SSA’s Technical Advisory Panels
- Executive Orders requiring “a presumption in favor of openness,” data that’s “accessible, discoverable, and usable by the public”
- The data sharing revolution in academia

Enormous missed opportunity: for the scientific community and others to check and improve SSA forecasts (for free); but easy to fix!
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one
  - Some data shared: in difficult, disorganized, non-automated formats
  - Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
  - Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

- Nontransparency and lack of data sharing violates:
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one
  - Some data shared: in difficult, disorganized, non-automated formats
  - Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
  - Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

- Nontransparency and lack of data sharing violates:
  - repeated, emphatic calls from SSA’s Technical Advisory Panels
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one
    - Some data shared: in difficult, disorganized, non-automated formats
    - Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
    - Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

- Nontransparency and lack of data sharing violates:
  - repeated, emphatic calls from SSA’s Technical Advisory Panels
  - Executive Orders requiring “a presumption in favor of openness,” data that’s “accessible, discoverable, and usable by the public”
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one
  - Some data shared: in difficult, disorganized, non-automated formats
  - Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
  - Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

- Nontransparency and lack of data sharing violates:
  - repeated, emphatic calls from SSA’s Technical Advisory Panels
  - Executive Orders requiring “a presumption in favor of openness,” data that’s “accessible, discoverable, and usable by the public”
  - the data sharing revolution in academia
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one

- Who has been able to fully replicate OCACT’s forecasts?
  - No one
  - Some data shared: in difficult, disorganized, non-automated formats
  - Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
  - Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

- Nontransparency and lack of data sharing violates:
  - repeated, emphatic calls from SSA’s Technical Advisory Panels
  - Executive Orders requiring “a presumption in favor of openness,” data that’s “accessible, discoverable, and usable by the public”
  - the data sharing revolution in academia

- Enormous missed opportunity: for the scientific community and others to check and improve SSA forecasts (for free);
Nontransparency in Social Security Forecasting

- Who forecasts independently of SSA’s Office of the Chief Actuary? No one
- Who has been able to fully replicate OCACT’s forecasts? No one
  - Some data shared: in difficult, disorganized, non-automated formats
  - Some impossible to share: informal, qualitative methods; e.g., committees choosing huge numbers of adjustable parameters
  - Much could be shared but is not (with the public, the scientific community, US government agencies, or even other parts of SSA)

- Nontransparency and lack of data sharing violates:
  - repeated, emphatic calls from SSA’s Technical Advisory Panels
  - Executive Orders requiring “a presumption in favor of openness,” data that’s “accessible, discoverable, and usable by the public”
  - the data sharing revolution in academia

- Enormous missed opportunity: for the scientific community and others to check and improve SSA forecasts (for free); but easy to fix!
Evaluating SSA Forecasts

The history of all systematic evaluations of SSA forecasts:

by SSA: None

by others: None

(A few highly selected numbers discussed in speeches)

Great opportunity for science and policy:

SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

Our methods:

Systematically compared each SSA forecast to the truth

Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

Preview of Results:

Before c. 2000:

Approximately unbiased forecasts

After 2000:

Systematically biased forecasts, increasingly so over time, all in the same direction

— making the Trust Funds consistently appear healthier than they actually are

How big is the bias?

Larger than almost all of OCACT’s policy scores

Straightforward solutions exist for all problems discovered
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA:

  - None
  - (A few highly selected numbers discussed in speeches)

Great opportunity for science and policy: SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

Our methods:

- Systematically compared each SSA forecast to the truth
- Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

Preview of Results:

Before c. 2000:

- Approximately unbiased forecasts

After 2000:

- Systematically biased forecasts, increasingly so over time, making the Trust Funds consistently appear healthier than they actually are

How big is the bias?

- Larger than almost all of OCACT's policy scores

Straightforward solutions exist for all problems discovered
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others:
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
Evaluating SSA Forecasts

The history of all systematic evaluations of SSA forecasts:
- by SSA: None
- by others: None
- (A few highly selected numbers discussed in speeches)
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)
- Great opportunity for science and policy:
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve
  - Our methods:
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

- Our methods:
  - Systematically compared each SSA forecast to the truth
Evaluating SSA Forecasts

The history of all systematic evaluations of SSA forecasts:
- by SSA: None
- by others: None
- (A few highly selected numbers discussed in speeches)

Great opportunity for science and policy:
- SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

Our methods:
- Systematically compared each SSA forecast to the truth
- Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve
- Our methods:
  - Systematically compared each SSA forecast to the truth
  - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

  **Our methods:**
  - Systematically compared each SSA forecast to the truth
  - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000:
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve
  - Our methods:
    - Systematically compared each SSA forecast to the truth
    - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000: Approximately unbiased forecasts
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve
  - Our methods:
    - Systematically compared each SSA forecast to the truth
    - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000: Approximately unbiased forecasts
  - After 2000:
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

- Our methods:
  - Systematically compared each SSA forecast to the truth
  - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000: Approximately unbiased forecasts
  - After 2000: Systematically biased forecasts,
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve
  - Our methods:
    - Systematically compared each SSA forecast to the truth
    - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000: Approximately unbiased forecasts
  - After 2000: Systematically biased forecasts, increasingly so over time,
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

- Our methods:
  - Systematically compared each SSA forecast to the truth
  - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000: Approximately unbiased forecasts
  - After 2000: Systematically biased forecasts, increasingly so over time, all in the same direction
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

- Our methods:
  - Systematically compared each SSA forecast to the truth
  - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000: Approximately unbiased forecasts
  - After 2000: Systematically biased forecasts, increasingly so over time, all in the same direction — making the Trust Funds consistently appear healthier than they actually are
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

- Our methods:
  - Systematically compared each SSA forecast to the truth
  - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000: Approximately unbiased forecasts
  - After 2000: Systematically biased forecasts, increasingly so over time, all in the same direction — making the Trust Funds consistently appear healthier than they actually are
  - How big is the bias?
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - by SSA: None
  - by others: None
  - (A few highly selected numbers discussed in speeches)

- Great opportunity for science and policy:
  - SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve
  - Our methods:
    - Systematically compared each SSA forecast to the truth
    - Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

- Preview of Results:
  - Before c. 2000: Approximately unbiased forecasts
  - After 2000: Systematically biased forecasts, increasingly so over time, all in the same direction — making the Trust Funds consistently appear healthier than they actually are
  - How big is the bias? Larger than almost all of OCACT’s policy scores
Evaluating SSA Forecasts

The history of all systematic evaluations of SSA forecasts:
- by SSA: None
- by others: None
- (A few highly selected numbers discussed in speeches)

Great opportunity for science and policy:
- SSA has been forecasting for so long, we can make truly out-of-sample evaluations, & use errors to improve

Our methods:
- Systematically compared each SSA forecast to the truth
- Conducted large number of detailed, semi-structured interviews with participants at every level of the policy and forecasting process

Preview of Results:
- Before c. 2000: Approximately unbiased forecasts
- After 2000: Systematically biased forecasts, increasingly so over time, all in the same direction — making the Trust Funds consistently appear healthier than they actually are
- How big is the bias? Larger than almost all of OCACT’s policy scores
- Straightforward solutions exist for all problems discovered
How OCACT Forecasts

Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science.

Example: Mortality Forecasts

Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity.

Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year t + 26 by committee in private.

Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:

- t + 1 to t + 26: “historical” rate; or 0.75 × “historical” if negative
- t + 3 to t + 25: change linearly from “historical” to “ultimate”
- t + 26 to t + 75: “ultimate” rate assumed constant for 50 years

Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups).

A committee in private evaluates forecasts, adjusts “ultimate” rates, and repeatedly reruns algorithm until consistent with their views.
How OCACT Forecasts

- SSA methods:

  - Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science.

  - Example: Mortality Forecasts
    - Estimate 294 "historical rates of decline" (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity.
    - Choose 210 "ultimate annual rates of mortality decline" (5 age groups × 2 sexes × 3 cost scenarios × 7 or 5 causes) for year \( t + 26 \) by committee in private.
    - Define future "annual rates of mortality decline" for each of the 294 groups, assuming constancy within each age group:
      - \( t + 1 \) to \( t + 25 \): "historical" rate; or 0.75 × "historical" if negative.
      - \( t + 26 \) to \( t + 75 \): "ultimate" rate assumed constant for 50 years.
    - Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups).
    - A committee in private evaluates forecasts, adjusts "ultimate" rates, and repeatedly reruns algorithm until consistent with their views.
How OCACT Forecasts

- **SSA methods**: Little changed in decades
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science.

Example: Mortality Forecasts

- Estimate 294 "historical rates of decline" (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity.
- Choose 210 "ultimate annual rates of mortality decline" (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year \(t + 26\) by committee in private.
- Define future "annual rates of mortality decline" for each of the 294 groups, assuming constancy within each age group:
  - \(t + 1\) to \(t + 2\): "historical" rate; or 0.75 × "historical" if negative.
  - \(t + 3\) to \(t + 25\): change linearly from "historical" to "ultimate".
  - \(t + 26\) to \(t + 75\): "ultimate" rate assumed constant for 50 years.
- Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups).
- A committee in private evaluates forecasts, adjusts "ultimate" rates, and repeatedly reruns algorithm until consistent with their views.
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example**: Mortality Forecasts

**Example: Mortality Forecasts**

- Estimate 294 "historical rates of decline" (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
- Choose 210 "ultimate annual rates of mortality decline" (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year t + 26 by committee in private
- Define future "annual rates of mortality decline" for each of the 294 groups, assuming constancy within each age group:
  - t + 1 to t + 2: "historical" rate
  - t + 3 to t + 25: change linearly from "historical" to "ultimate"
  - t + 26 to t + 75: "ultimate" rate assumed constant for 50 years
- Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups)
- A committee in private evaluates forecasts, adjusts "ultimate" rates, and repeatedly reruns algorithm until consistent with their views
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes)
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science

- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
How OCACT Forecasts

**SSA methods:** Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science

**Example: Mortality Forecasts**
- Estimate 294 “historical rates of decline” (21 ages $\times$ 2 sexes $\times$ 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
- Choose 210 “ultimate annual rates of mortality decline” (5 age groups $\times$ 2 sexes $\times$ 3 cost scenarios $\times$ 7 (or 5) causes) for year $t + 26$
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science

- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year $t + 26$ by committee
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year $t + 26$ by committee in private
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example: Mortality Forecasts**
  - Estimate **294** “historical rates of decline” (21 ages $\times$ 2 sexes $\times$ 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose **210** “ultimate annual rates of mortality decline” (5 age groups $\times$ 2 sexes $\times$ 3 cost scenarios $\times$ 7 (or 5) causes) for year $t + 26$ by committee in private
  - Define future “annual rates of mortality decline” for each of the 294 groups,
How OCACT Forecasts

**SSA methods:** Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science

**Example: Mortality Forecasts**
- Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
- Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year $t + 26$ by committee in private
- Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science.

- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity.
  - Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year \( t + 26 \) by committee in private.
  - Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
    - \( t + 1 \) to \( t + 2 \): “historical” rate; or \( 0.75 \times \) “historical” if negative.

A committee in private evaluates forecasts, adjusts “ultimate” rates, and repeatedly reruns algorithm until consistent with their views.
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year \( t + 26 \) by committee in private
  - Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
    - \( t + 1 \) to \( t + 2 \): “historical” rate; or 0.75×“historical” if negative
    - \( t + 3 \) to \( t + 25 \): change linearly from “historical” to “ultimate”
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science

- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year $t + 26$ by committee in private
  - Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
    - $t + 1$ to $t + 2$: “historical” rate; or $0.75 \times “historical”$ if negative
    - $t + 3$ to $t + 25$: change linearly from “historical” to “ultimate”
    - $t + 26$ to $t + 75$: “ultimate” rate assumed constant for 50 years

Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups). A committee in private evaluates forecasts, adjusts “ultimate” rates, and repeatedly reruns algorithm until consistent with their views.
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science

- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year $t + 26$ by committee in private
  - Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
    - $t + 1$ to $t + 2$: “historical” rate; or $0.75 \times$ “historical” if negative
    - $t + 3$ to $t + 25$: change linearly from “historical” to “ultimate”
    - $t + 26$ to $t + 75$: “ultimate” rate assumed constant for 50 years
  - Iteratively multiply 210 (or 150) mortality rates by the future annual rates;
How OCACT Forecasts

**SSA methods:** Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science

**Example: Mortality Forecasts**

- Estimate **294** “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
- Choose **210** “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year $t + 26$ by committee in private
- Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
  - $t + 1$ to $t + 2$: “historical” rate; or $0.75 \times$ “historical” if negative
  - $t + 3$ to $t + 25$: change linearly from “historical” to “ultimate”
  - $t + 26$ to $t + 75$: “ultimate” rate assumed constant for 50 years
- Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups)
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year $t + 26$ by committee in private
  - Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
    - $t + 1$ to $t + 2$: “historical” rate; or $0.75 \times$ “historical” if negative
    - $t + 3$ to $t + 25$: change linearly from “historical” to “ultimate”
    - $t + 26$ to $t + 75$: “ultimate” rate assumed constant for 50 years
  - Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups)
  - A committee
How OCACT Forecasts

- **SSA methods:** Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example: Mortality Forecasts**
  - Estimate 294 “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose 210 “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year \( t + 26 \) by committee in private
  - Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
    - \( t + 1 \) to \( t + 2 \): “historical” rate; or \( 0.75 \times \) “historical” if negative
    - \( t + 3 \) to \( t + 25 \): change linearly from “historical” to “ultimate”
    - \( t + 26 \) to \( t + 75 \): “ultimate” rate assumed constant for 50 years
  - Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups)
  - A committee in private
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science

- **Example: Mortality Forecasts**
  - Estimate **294** “historical rates of decline” (21 ages × 2 sexes × 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose **210** “ultimate annual rates of mortality decline” (5 age groups × 2 sexes × 3 cost scenarios × 7 (or 5) causes) for year \( t + 26 \) by committee in private
  - Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
    - \( t + 1 \) to \( t + 2 \): “historical” rate; or \( 0.75 \times \) “historical” if negative
    - \( t + 3 \) to \( t + 25 \): change linearly from “historical” to “ultimate”
    - \( t + 26 \) to \( t + 75 \): “ultimate” rate assumed constant for 50 years
  - Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups)
  - A committee in private evaluates forecasts, adjusts “ultimate” rates,
How OCACT Forecasts

- **SSA methods**: Little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social science
- **Example: Mortality Forecasts**
  - Estimate **294** “historical rates of decline” (21 ages $\times$ 2 sexes $\times$ 7 causes) by independent linear regressions on time, ignoring known risk factors, like smoking & obesity
  - Choose **210** “ultimate annual rates of mortality decline” (5 age groups $\times$ 2 sexes $\times$ 3 cost scenarios $\times$ 7 (or 5) causes) for year $t + 26$ by committee in private
  - Define future “annual rates of mortality decline” for each of the 294 groups, assuming constancy within each age group:
    - $t + 1$ to $t + 2$: “historical” rate; or $0.75 \times$ “historical” if negative
    - $t + 3$ to $t + 25$: change linearly from “historical” to “ultimate”
    - $t + 26$ to $t + 75$: “ultimate” rate assumed constant for 50 years
  - Iteratively multiply 210 (or 150) mortality rates by the future annual rates; sum across (7 or 5) causes (within age-sex-cost groups)
  - A committee in private evaluates forecasts, adjusts “ultimate” rates, and repeatedly reruns algorithm until consistent with their views
Actual Mortality Time Profiles are Complex
Actual Mortality Time Profiles are Complex

Patterns:
- linear,
- different slopes,
- different variances,
- diagonal ripples
Actual Mortality Time Profiles are Complex

Patterns:

- ≈ linear,
- different slopes,
- different variances,
- diagonal ripples
Actual Mortality Time Profiles are Complex

Patterns: \( \approx \) linear,
Actual Mortality Time Profiles are Complex

Patterns: \( \approx \) linear, different slopes,
Actual Mortality Time Profiles are Complex

Patterns: ≈ linear, different slopes, different variances,
Actual Mortality Time Profiles are Complex

Patterns: \( \approx \) linear, different slopes, different variances, diagonal ripples
Actual Mortality Age Profiles are also Complex
Actual Mortality Age Profiles are also Complex

Patterns:
- Characteristic shape
- Partly linear
- Different time-age trends

8/23
Actual Mortality Age Profiles are also Complex

Patterns:

Characteristic shape, partly linear, different time × age trends
Actual Mortality Age Profiles are also Complex

Patterns: Characteristic shape,
Actual Mortality Age Profiles are also Complex

Patterns: Characteristic shape, partly linear,
Actual Mortality Age Profiles are also Complex

Patterns: Characteristic shape, partly linear, different time×age trends
OCACT Qualitative Choices

Unrealistic patterns:
- change of directions,
- change of differences
OCACT Qualitative Choices: Violate Known Information
OCACT Qualitative Choices: Violate Known Information

Diabetes Death Rate, Males

Cancer Death Rate, Females

Ages 80−84

Ages 85−89

Death Rate (per 100,000)

Year

Death Rate (per 100,000)

Year

Unrealistic patterns: change of directions, change of differences
OCACT Qualitative Choices: Violate Known Information

Unrealistic patterns:
OCACT Qualitative Choices: Violate Known Information

Unrealistic patterns: change of directions,
OCACT Qualitative Choices: Violate Known Information

Unrealistic patterns: change of directions, change of differences
OCACT Qualitative Choices: Violate Known Information

**Heart Disease Death Rate, Males**

- Ages 45–49
- Ages 50–54

**Death Rate (per 100,000)**

- Year 1975 to 2100

**Heart Disease Death Rate, Males**

- Death Rate (Log)
- Year: 2050 to 2100
- Age: 0 to 75

Unrealistic patterns:
- Crossing age plots
- Notch for 50-75 year olds
OCACT Qualitative Choices: Violate Known Information

Unrealistic patterns:
OCACT Qualitative Choices: Violate Known Information

Unrealistic patterns: Crossing age plots,
OCACT Qualitative Choices: Violate Known Information

Unrealistic patterns: Crossing age plots, notch for 50-75 year olds
SSA Life Expectancy Forecasts: Increasing Bias Since 2000
SSA Life Expectancy Forecasts: Increasing Bias Since 2000
(LE at 65; 1-5 year SSA forecasts)
SSA Life Expectancy Forecasts: Increasing Bias Since 2000
(LE at 65; 1-5 year SSA forecasts)
Life Expectancy “Uncertainty Interval” Coverage
Life Expectancy “Uncertainty Interval” Coverage
Systematic overconfidence since at least 2000
Life Expectancy “Uncertainty Interval” Coverage
Systematic overconfidence since at least 2000
Life Expectancy “Uncertainty Interval” Coverage
Systematic overconfidence since at least 2000

Patterns:
Life Expectancy “Uncertainty Interval” Coverage
Systematic overconfidence since at least 2000

Patterns:
- **Vertical**: Later Trustees Reports are overconfident
Life Expectancy “Uncertainty Interval” Coverage
Systematic overconfidence since at least 2000

Patterns:
- Vertical: Later Trustees Reports are overconfident
- Not horizontal: Shorter term forecasts should be better, but aren’t
Trust Fund Ratio Forecasting Errors: 1 Year Ahead

![Graph showing trust fund ratio forecasting errors from 1980 to 2010. The x-axis represents the year of the Trustees Report, ranging from 1980 to 2010. The y-axis shows the forecast minus truth values, ranging from -15 to 15. The data points are scattered across the graph, with a trend line indicating an increasing error over time. The shaded area represents the range of errors, with a notable increase in error after 2000.]
Trust Fund Ratio Forecasting Errors: 2 Years Ahead

![Graph showing forecast minus truth over years]

- X-axis: Year of Trustees Report
- Y-axis: Forecast - Truth
- Data points and trend lines indicating forecasting errors over time.
Trust Fund Ratio Forecasting Errors: 3 Years Ahead
Trust Fund Ratio Forecasting Errors: 4 Years Ahead

![Chart showing forecast errors over the years, with data points and a trend line indicating increasing divergence over time.]
Trust Fund Ratio Forecasting Errors: 5 Years Ahead
Trust Fund Ratio Forecasting Errors: 6 Years Ahead

The diagram shows the forecasted trust fund ratios compared to the actual (truth) values over a period from 1980 to 2010. The data points indicate the difference between the forecast and the truth, with blue and red dots representing different forecasts with the forecast error shaded in gray. The trend lines and shaded areas illustrate the trend and variability in the forecast errors over the years.
Trust Fund Ratio Forecasting Errors: 7 Years Ahead

![Graph showing trust fund ratio forecasting errors from 1980 to 2010. The graph plots forecast values against actual truth values, with error bars indicating the range of errors. The x-axis represents the year of the Trustees Report, ranging from 1980 to 2010, and the y-axis represents the forecast minus truth ratio, ranging from -150 to 150.]
Trust Fund Ratio Forecasting Errors: 8 Years Ahead
Trust Fund Ratio Forecasting Errors: 9 Years Ahead

![Graph showing forecast vs. truth for the year of trustees' report, with data points and error trends from 1980 to 2010.](image-url)
Trust Fund Ratio Forecasting Errors: Summary

![Graphs showing forecast error (percentage points) over years](image)
Uncertainty Estimates for OCACT Policy Scores

Who scores SSA Policy Proposals?

OCACT: the monopoly supplier for every major proposal (105 since 1993);

- Lack of data sharing makes it impossible for others

Advantages:

- Both parties can negotiate to one point;
- Being in OCACT is more exciting

Disadvantages:

- The one point the parties are negotiating to may be wrong;
- No one can check;
- Hard to improve anything in isolation;
- The scientific community can't contribute

OCACT's reported uncertainty estimates:

None.

Actual uncertainty: two components

1. Forecasting under factual conditions
2. Intervening under counterfactual conditions

We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95th percentile i.e., with $\alpha \leq 0.05$?

Optimistic assumptions, but still assumptions.
Who scores SSA Policy Proposals?

OCACT: the monopoly supplier for every major proposal (105 since 1993);

Advantages:
- Both parties can negotiate to one point;
- being in OCACT is more exciting

Disadvantages:
- The one point the parties are negotiating to may be wrong;
- no one can check;
- hard to improve anything in isolation;
- the scientific community can't contribute

OCACT's reported uncertainty estimates: none.

Actual uncertainty: two components
1. Forecasting under factual conditions
2. Intervening under counterfactual conditions

We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95th percentile i.e., with $\alpha \leq 0.05$?

Optimistic assumptions, but still assumptions
Uncertainty Estimates for OCACT Policy Scores

Who scores SSA Policy Proposals?

- OCACT: the monopoly supplier for every major proposal (105 since 1993);
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - OCACT: the *monopoly supplier* for every major proposal (105 since 1993); lack of data sharing makes it impossible for others.
Uncertainty Estimates for OCACT Policy Scores

- Who scores SSA Policy Proposals?
  - OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - Advantages:
    - Both parties can negotiate to one point;
    - being in OCACT is more exciting
    - Disadvantages:
      - The one point the parties are negotiating to may be wrong;
      - no one can check;
      - hard to improve anything in isolation;
      - the scientific community can't contribute
    - OCACT's reported uncertainty estimates: none.
    - Actual uncertainty: two components
      - 1 Forecasting under factual conditions
      - 2 Intervening under counterfactual conditions
    - We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95th percentile i.e., with $\alpha \leq 0.05$?
      - Optimistic assumptions, but still assumptions
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - OCACT: the *monopoly supplier* for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - **Advantages:** Both parties can negotiate to one point;

- OCACT's reported uncertainty estimates: none.
- Actual uncertainty: two components
  - Forecasting under factual conditions
  - Intervening under counterfactual conditions

- We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95th percentile i.e., with $\alpha \leq 0.05$?

- Optimistic assumptions, but still assumptions
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - **OCACT**: the *monopoly supplier* for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - **Advantages**: Both parties can negotiate to one point; being in OCACT is more exciting
Who scores SSA Policy Proposals?

OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others.

Advantages: Both parties can negotiate to one point; being in OCACT is more exciting.

Disadvantages:
Uncertainty Estimates for OCACT Policy Scores

Who scores SSA Policy Proposals?

- OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
- Advantages: Both parties can negotiate to one point; being in OCACT is more exciting
- Disadvantages: The one point the parties are negotiating to may be wrong;

Actual uncertainty: two components

1. Forecasting under factual conditions
2. Intervening under counterfactual conditions

We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95th percentile i.e., with $\alpha \leq 0.05$?

Optimistic assumptions, but still assumptions
Uncertainty Estimates for OCACT Policy Scores

Who scores SSA Policy Proposals?

OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others

Advantages: Both parties can negotiate to one point; being in OCACT is more exciting

Disadvantages: The one point the parties are negotiating to may be wrong; no one can check;

OCACT's reported uncertainty estimates: none.

Actual uncertainty: two components

1. Forecasting under factual conditions
2. Intervening under counterfactual conditions

We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95\text{th} percentile i.e., with $\alpha \leq 0.05$?

Optimistic assumptions, but still assumptions
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - **OCACT:** the *monopoly supplier* for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - **Advantages:** Both parties can negotiate to one point; being in OCACT is more exciting
  - **Disadvantages:** The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation;

**Forecasting under factual conditions**

**Intervening under counterfactual conditions**

We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are >95th percentile i.e., with $\alpha \leq 0.05$?

Optimistic assumptions, but still assumptions
Who scores SSA Policy Proposals?

- **OCACT**: the *monopoly supplier* for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
- **Advantages**: Both parties can negotiate to one point; being in OCACT is more exciting
- **Disadvantages**: The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

---

Uncertainty Estimates for OCACT Policy Scores

- **Forecasting under factual conditions**
- **Intervening under counterfactual conditions**

We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95th percentile i.e., with $\alpha \leq 0.05$? 

 CRA Optimistic assumptions, but still assumptions
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - OCACT: the *monopoly supplier* for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - **Advantages:** Both parties can negotiate to one point; being in OCACT is more exciting
  - **Disadvantages:** The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

- **OCACT’s reported uncertainty estimates:**
Uncertainty Estimates for OCACT Policy Scores

- Who scores SSA Policy Proposals?
  - OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - Advantages: Both parties can negotiate to one point; being in OCACT is more exciting
  - Disadvantages: The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

- OCACT’s reported uncertainty estimates: none.
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - OCACT: the **monopoly supplier** for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - **Advantages:** Both parties can negotiate to one point; being in OCACT is more exciting
  - **Disadvantages:** The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

- **OCACT’s reported uncertainty estimates:** none.
- **Actual uncertainty:** two components

1. Forecasting under factual conditions
2. Intervening under counterfactual conditions

We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are >95th percentile i.e., with $\alpha \leq 0.05$?

$\Rightarrow$ Optimistic assumptions, but still assumptions
Who scores SSA Policy Proposals?

- OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
- Advantages: Both parties can negotiate to one point; being in OCACT is more exciting
- Disadvantages: The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

OCACT’s reported uncertainty estimates: none.

Actual uncertainty: two components

1. Forecasting under factual conditions
Uncertainty Estimates for OCACT Policy Scores

- Who scores SSA Policy Proposals?
  - OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - Advantages: Both parties can negotiate to one point; being in OCACT is more exciting
  - Disadvantages: The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

- OCACT’s reported uncertainty estimates: none.

- Actual uncertainty: two components
  1. Forecasting under factual conditions
  2. Intervening under counterfactual conditions
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - OCACT: the *monopoly supplier* for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - **Advantages:** Both parties can negotiate to one point; being in OCACT is more exciting
  - **Disadvantages:** The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

- **OCACT’s reported uncertainty estimates:** none.

- **Actual uncertainty:** two components
  1. Forecasting under factual conditions
  2. Intervening under counterfactual conditions

- **We estimate actual uncertainty:** use 1st only (as a lower bound);
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - OCACT: the *monopoly supplier* for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - **Advantages:** Both parties can negotiate to one point; being in OCACT is more exciting
  - **Disadvantages:** The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

- **OCACT’s reported uncertainty estimates:** none.

- **Actual uncertainty:** two components
  1. Forecasting under factual conditions
  2. Intervening under counterfactual conditions

- **We estimate actual uncertainty:** use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears;
Uncertainty Estimates for OCACT Policy Scores

- Who scores SSA Policy Proposals?
  - OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - Advantages: Both parties can negotiate to one point; being in OCACT is more exciting
  - Disadvantages: The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

- OCACT’s reported uncertainty estimates: none.

- Actual uncertainty: two components
  1. Forecasting under factual conditions
  2. Intervening under counterfactual conditions

- We estimate actual uncertainty: use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95th percentile i.e., with $\alpha \leq 0.05$?
Uncertainty Estimates for OCACT Policy Scores

- **Who scores SSA Policy Proposals?**
  - OCACT: the monopoly supplier for every major proposal (105 since 1993); lack of data sharing makes it impossible for others
  - **Advantages:** Both parties can negotiate to one point; being in OCACT is more exciting
  - **Disadvantages:** The one point the parties are negotiating to may be wrong; no one can check; hard to improve anything in isolation; the scientific community can’t contribute

- **OCACT’s reported uncertainty estimates:** none.

- **Actual uncertainty:** two components
  1. Forecasting under factual conditions
  2. Intervening under counterfactual conditions

- **We estimate actual uncertainty:** use 1st only (as a lower bound); compute percentile of error (among all forecast errors, 1-10 years out) where each score appears; how many are > 95\(^{th}\) percentile i.e., with \(\alpha \leq 0.05\)? \(\sim\) Optimistic assumptions, but still assumptions
SSA Policy Scoring: Mostly Random Noise
SSA Policy Scoring: Mostly Random Noise

Estimated policy effect size overwhelmed by (forecasting) uncertainty

Estimated policy effect size larger than (forecasting) uncertainty

Percentile of Forecast Error

Balance

Cost

10-Year 75-Year 10-Year 75-Year

0% 20% 40% 60% 80% 100%
Social Psychological Conditions that make Bias Possible

Bias: Systematic errors, regardless of intention or direction

The soc-psych literature:
- Bias is likely when human beings perform complex tasks,
- with high discretion,
- many decisions,
- little feedback on whether they made the right choice the last time,
- high external pressure,
- in a group,
- and few external checks—exactly OCACT's difficult situation & procedures.

Qualitative uncertainty estimates are also likely biased
- "Experts" are usually overconfident.
- "Do not trust anyone—including yourself—to tell you how much you should trust their judgment" (Kahneman 2011)
- The more prominent or central a forecaster, the more overconfident their statements (Tetlock 2005)
- and as the sole supplier of forecasts and policy evaluations, OCACT could hardly be more central.

It's not about the person: "Trying harder," or replacing one person with another, usually has no effect (Banaji and Greenwald 2013).

It can't be learned: "Teaching psychology is mostly a waste of time" (Kahneman 2011).
Social Psychological Conditions that make Bias Possible

- “Bias”: Systematic errors, regardless of intention or direction
Social Psychological Conditions that make Bias Possible

- **“Bias”:** Systematic errors, regardless of intention or direction
- **The soc-psych literature:**

  - **Bias** is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks—exactly OCACT's difficult situation & procedures.

  - Qualitative uncertainty estimates are also likely biased.

  - "Experts" are usually overconfident.

  - "Do not trust anyone—including yourself—to tell you how much you should trust their judgment" (Kahneman 2011).

  - The more prominent or central a forecaster, the more overconfident their statements (Tetlock 2005).

  - "and as the sole supplier of forecasts and policy evaluations, OCACT could hardly be more central.

  - It's not about the person: "Trying harder," or replacing one person with another, usually has no effect (Banaji and Greenwald 2013).

  - It can't be learned: "Teaching psychology is mostly a waste of time" (Kahneman 2011).
Social Psychological Conditions that make Bias Possible

- “Bias”: Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks,
Social Psychological Conditions that make Bias Possible

- “Bias”: Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks, with high discretion,
Social Psychological Conditions that make Bias Possible

- **“Bias”:** Systematic errors, regardless of intention or direction
- **The soc-psych literature:** Bias is likely when human beings perform complex tasks, with high discretion, many decisions,
Social Psychological Conditions that make Bias Possible

- “Bias”: Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time,
Social Psychological Conditions that make Bias Possible

- "Bias": Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure,
Social Psychological Conditions that make Bias Possible

- “Bias”: Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group,
Social Psychological Conditions that make Bias Possible

- "Bias": Systematic errors, regardless of intention or direction
- The soc-psy literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks
Social Psychological Conditions that make Bias Possible

- “Bias”: Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks — exactly OCACT’s difficult situation & procedures
Social Psychological Conditions that make Bias *Possible*

- "Bias": Systematic errors, regardless of intention or direction
- The soc-psyche literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks — exactly OCACT’s difficult situation & procedures
- Qualitative uncertainty estimates are also likely biased
Social Psychological Conditions that make Bias *Possible*

- **“Bias”:** Systematic errors, regardless of intention or direction
- **The soc-psyh literature:** Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks — exactly OCACT’s difficult situation & procedures
- **Qualitative uncertainty estimates are also likely biased**
  - “Experts” are usually overconfident.
Social Psychological Conditions that make Bias Possible

- “Bias”: Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks — exactly OCACT’s difficult situation & procedures
- Qualitative uncertainty estimates are also likely biased
  - “Experts” are usually overconfident.
  - “Do not trust anyone — including yourself — to tell you how much you should trust their judgment” (Kahneman 2011)
Social Psychological Conditions that make Bias Possible

- "Bias": Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks — exactly OCACT’s difficult situation & procedures
- Qualitative uncertainty estimates are also likely biased
  - “Experts” are usually overconfident.
  - “Do not trust anyone — including yourself — to tell you how much you should trust their judgment” (Kahneman 2011)
  - The more prominent or central a forecaster, the more overconfident their statements (Tetlock 2005)
Social Psychological Conditions that make Bias *Possible*

- **“Bias”:** Systematic errors, regardless of intention or direction
- **The soc-psych literature:** Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks — exactly OCACT’s difficult situation & procedures
- **Qualitative uncertainty estimates are also likely biased**
  - “Experts” are usually overconfident.
  - “Do not trust anyone — including yourself — to tell you how much you should trust their judgment” (Kahneman 2011)
  - The more prominent or central a forecaster, the more overconfident their statements (Tetlock 2005) — and as the sole supplier of forecasts and policy evaluations, OCACT could hardly be more central
Social Psychological Conditions that make Bias Possible

- "Bias": Systematic errors, regardless of intention or direction
- The soc-psych literature: Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks — exactly OCACT’s difficult situation & procedures
- Qualitative uncertainty estimates are also likely biased
  - “Experts” are usually overconfident.
  - “Do not trust anyone — including yourself — to tell you how much you should trust their judgment” (Kahneman 2011)
  - The more prominent or central a forecaster, the more overconfident their statements (Tetlock 2005) — and as the sole supplier of forecasts and policy evaluations, OCACT could hardly be more central
- It’s not about the person: “Trying harder,” or replacing one person with another, usually has no effect (Banaji and Greenwald 2013)
Social Psychological Conditions that make Bias *Possible*

- **“Bias”:** Systematic errors, regardless of intention or direction
- **The soc-psych literature:** Bias is likely when human beings perform complex tasks, with high discretion, many decisions, little feedback on whether they made the right choice the last time, high external pressure, in a group, and few external checks — exactly OCACT’s difficult situation & procedures
- Qualitative uncertainty estimates are also likely biased
  - “Experts” are usually overconfident.
  - “Do not trust anyone — including yourself — to tell you how much you should trust their judgment” (Kahneman 2011)
  - The more prominent or central a forecaster, the more overconfident their statements (Tetlock 2005) — and as the sole supplier of forecasts and policy evaluations, OCACT could hardly be more central
- It’s not about the person: “Trying harder,” or replacing one person with another, usually has no effect (Banaji and Greenwald 2013)
- It can’t be learned: “Teaching psychology is mostly a waste of time” (Kahneman 2011)
A Three-Part Solution, from Three Revolutions

1. Remove human judgment where possible, via formal statistical methods — automate what can be automated
   Evidence: The revolution in data science (big data, statistics, etc.)
   Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. Institute formal structural procedures when human judgment is required — focus experts on what they’re expert at
   Evidence: The revolution in social psychology
   Double-blind experiments, or peer review
   Violin competitions behind a curtain, without shoes

3. Require transparency and data sharing, to catch errors that slip through — bring the advantages of science to government
   Evidence: The revolution in data sharing in academia and government, (and even to some extent industry)
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible, via formal statistical methods.**

   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures when human judgment is required, focus experts on what they’re expert at.**

   - Evidence: The revolution in social psychology
     - Double-blind experiments, or peer review
     - Violin competitions behind a curtain, without shoes

3. **Require transparency and data sharing, to catch errors that slip through.**

   - Evidence: The revolution in data sharing in academia and government, (and even to some extent industry)
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible, via formal statistical methods — automate what can be automated**
Remove human judgment where possible, via formal statistical methods — automate what can be automated

Evidence: The revolution in data science (big data, statistics, etc.)
A Three-Part Solution, from Three Revolutions

1 Remove human judgment where possible, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible**, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures** when human judgment is required
A Three-Part Solution, from Three Revolutions

1. Remove human judgment where possible, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. Institute formal structural procedures when human judgment is required — focus experts on what they’re expert at
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible**, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures** when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in social psychology
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible, via formal statistical methods — automate what can be automated**
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures when human judgment is required — focus experts on what they’re expert at**
   - Evidence: The revolution in social psychology
   - Double-blind experiments, or peer review
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible,** via formal statistical methods — automate what can be automated
   - Evidence: The revolution in *data science* (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures** when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in *social psychology*
   - Double-blind experiments, or peer review
   - Violin competitions
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible**, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures** when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in social psychology
   - Double-blind experiments, or peer review
   - Violin competitions behind a curtain,
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible**, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures** when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in social psychology
   - Double-blind experiments, or peer review
   - Violin competitions behind a curtain, without shoes
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible**, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in *data science* (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures** when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in *social psychology*
   - Double-blind experiments, or peer review
   - Violin competitions behind a curtain, without shoes

3. **Require transparency and data sharing**, to catch errors that slip through
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible**, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures** when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in social psychology
   - Double-blind experiments, or peer review
   - Violin competitions behind a curtain, without shoes

3. **Require transparency and data sharing**, to catch errors that slip through — bring the advantages of science to government
A Three-Part Solution, from Three Revolutions

1. Remove human judgment where possible, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. Institute formal structural procedures when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in social psychology
   - Double-blind experiments, or peer review
   - Violin competitions behind a curtain, without shoes

3. Require transparency and data sharing, to catch errors that slip through — bring the advantages of science to government
   - Evidence: The revolution in data sharing in academia and government, (and even to some extent industry)
A Three-Part Solution, from Three Revolutions

1. **Remove human judgment where possible**, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. **Institute formal structural procedures** when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in social psychology
   - Double-blind experiments, or peer review
   - Violin competitions behind a curtain, without shoes

3. **Require transparency and data sharing**, to catch errors that slip through — bring the advantages of science to government
   - Evidence: The revolution in data sharing in academia and government, (and even to some extent industry)
A Three-Part Solution, from Three Revolutions

1. Remove human judgment where possible, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. Institute formal structural procedures when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in social psychology
   - Double-blind experiments, or peer review
   - Violin competitions behind a curtain, without shoes

3. Require transparency and data sharing, to catch errors that slip through — bring the advantages of science to government
   - Evidence: The revolution in data sharing in academia and government, (and even to some extent industry)
A Three-Part Solution, from Three Revolutions

1. Remove human judgment where possible, via formal statistical methods — automate what can be automated
   - Evidence: The revolution in data science (big data, statistics, etc.)
   - Commercial models: Netflix Challenge, Kaggle, TopCoder, Xprize

2. Institute formal structural procedures when human judgment is required — focus experts on what they’re expert at
   - Evidence: The revolution in social psychology
   - Double-blind experiments, or peer review
   - Violin competitions behind a curtain, without shoes

3. Require transparency and data sharing, to catch errors that slip through — bring the advantages of science to government
   - Evidence: The revolution in data sharing in academia and government, (and even to some extent industry)
Without Protections, Internal Pressures Make Bias Likely
Without Protections, Internal Pressures Make Bias Likely

- OCACT’s Stance as the Lone Island of Fairness
Without Protections, Internal Pressures Make Bias Likely

- **OCACT’s Stance as the Lone Island of Fairness**
  - Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
OCACT’s Stance as the Lone Island of Fairness

- Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
- We agree: no evidence of OCACT bending to political pressure
Without Protections, Internal Pressures Make Bias Likely

**OCACT’s Stance as the Lone Island of Fairness**
- Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
- We agree: no evidence of OCACT bending to political pressure
- But OCACT acts as if it has a monopoly on fairness
OCACT’s Stance as the Lone Island of Fairness

- Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
- We agree: no evidence of OCACT bending to political pressure
- But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals,
Without Protections, Internal Pressures Make Bias Likely

- OCACT’s Stance as the Lone Island of Fairness
  - Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
  - We agree: no evidence of OCACT bending to political pressure
  - But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts,
OCACT’s Stance as the Lone Island of Fairness

- Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
- We agree: no evidence of OCACT bending to political pressure
- But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
OCACT’s Stance as the Lone Island of Fairness

- Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
- We agree: no evidence of OCACT bending to political pressure
- But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
- Several said: “Goss is intellectually biased, not politically biased”
Without Protections, Internal Pressures Make Bias Likely

- **OCACT’s Stance as the Lone Island of Fairness**
  - Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
  - We agree: no evidence of OCACT bending to political pressure
  - But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
  - Several said: “Goss is intellectually biased, not politically biased”

- **Consistency Bias:**
OCACT’s Stance as the Lone Island of Fairness
- Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
- We agree: no evidence of OCACT bending to political pressure
- But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
- Several said: “Goss is intellectually biased, not politically biased”

Consistency Bias:
- Degrading accuracy to maintain central role in policy debate
Without Protections, Internal Pressures Make Bias Likely

- **OCACT’s Stance as the Lone Island of Fairness**
  - Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
  - We agree: no evidence of OCACT bending to political pressure
  - But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
  - Several said: “Goss is intellectually biased, not politically biased”

- **Consistency Bias:**
  - Degrading accuracy to maintain central role in policy debate
  - Intentionally biasing today’s forecast towards yesterday’s ⋆ much smoother over time than related forecasts
OCACT’s Stance as the Lone Island of Fairness

- Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
- We agree: no evidence of OCACT bending to political pressure
- But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
- Several said: “Goss is intellectually biased, not politically biased”

Consistency Bias:

- Degrading accuracy to maintain central role in policy debate
- Intentionally biasing today’s forecast towards yesterday’s ~ much smoother over time than related forecasts
- When the Technical Panel recommends a change in a parameter:
Without Protections, Internal Pressures Make Bias Likely

- **OCACT’s Stance as the Lone Island of Fairness**
  - Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
  - We agree: no evidence of OCACT bending to political pressure
  - But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
  - Several said: “Goss is intellectually biased, not politically biased”

- **Consistency Bias:**
  - Degrading accuracy to maintain central role in policy debate
  - Intentionally biasing today’s forecast towards yesterday’s \( \rightsquigarrow \) much smoother over time than related forecasts
  - When the Technical Panel recommends a change in a parameter:
    - If Goss has good evidence: he engages the Panel and convinces them
Without Protections, Internal Pressures Make Bias Likely

- **OCACT’s Stance as the Lone Island of Fairness**
  - Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
  - We agree: no evidence of OCACT bending to political pressure
  - But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
  - Several said: “Goss is intellectually biased, not politically biased”

- **Consistency Bias:**
  - Degrading accuracy to maintain central role in policy debate
  - Intentionally biasing today’s forecast towards yesterday’s $\rightarrow$ much smoother over time than related forecasts
  - When the Technical Panel recommends a change in a parameter:
    - If Goss has good evidence: he engages the Panel and convinces them
    - If the Panel has good evidence: he ignores the panel
Without Protections, Internal Pressures Make Bias Likely

- **OCACT’s Stance as the Lone Island of Fairness**
  - Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
  - We agree: no evidence of OCACT bending to political pressure
  - But OCACT acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
  - Several said: “Goss is intellectually biased, not politically biased”

- **Consistency Bias:**
  - Degrading accuracy to maintain central role in policy debate
  - Intentionally biasing today’s forecast towards yesterday’s $\implies$ much smoother over time than related forecasts
  - When the Technical Panel recommends a change in a parameter:
    - If Goss has good evidence: he engages the Panel and convinces them
    - If the Panel has good evidence: he ignores the panel
    - If the Panel has very strong evidence: he adjusts the parameter part way, and adjusts another so the forecast is unchanged
Without Protections, Internal Pressures Make Bias Likely

- **OCACT’s Stance as the Lone Island of Fairness**
  - Many extreme statements: E.g., Steve Goss: “I’ll take a bullet before I modify anything under any kind of political pressure”
  - We agree: no evidence of OCAST bending to political pressure
  - But OCAST acts as if it has a monopoly on fairness, letting no one else score proposals, make forecasts, or decide what’s evaluated
  - Several said: “Goss is intellectually biased, not politically biased”

- **Consistency Bias:**
  - Degrading accuracy to maintain central role in policy debate
  - Intentionally biasing today’s forecast towards yesterday’s \(\rightsquigarrow\) much smoother over time than related forecasts
  - When the Technical Panel recommends a change in a parameter:
    - If Goss has good evidence: he engages the Panel and convinces them
    - If the Panel has good evidence: he ignores the panel
    - If the Panel has very strong evidence: he adjusts the parameter part way, and adjusts another so the forecast is unchanged
  - Many quotes; e.g. Goss: “The hard part is trying to balance the need to change on the basis of new ideas and understanding with the desire for consistency and stability over time”
Ignoring Technical Panel Recommendations

Process:
- OCACT is extremely responsive in providing information
- "Steve Goss has a seat at every table" when policy is made

Technical Panel Methodological Recommendations
- Little evidence of serious engagement: After each Panel, for the last 15 years:
  - OCACT adopts a few recommendations,
  - ignores many,
  - and does not come close to the achievable ideal

Little progress on most important issues:
- Adopting formal statistical procedures,
- formal uncertainty estimates,
- transparency,
- data sharing,
- and routine systematic forecast evaluations

Technical Panel Substantive Recommendations
- For some: token dismissals in the Trustees Report
- For others: the Trustees Report contradicts the Panel, repeats identically worded assertions year after year, without engaging the Panel or the crucial issues raised

The Trustees and Technical Panel agree on many issues too, but the lack of engagement or mutual understanding is obvious
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - "Steve Goss has a seat at every table" when policy is made
  - Little evidence of serious engagement: After each Panel, for the last 15 years:
    - OCACT adopts a few recommendations,
    - ignores many,
    - and does not come close to the achievable ideal
  - Little progress on most important issues:
    - Adopting formal statistical procedures,
    - formal uncertainty estimates,
    - transparency,
    - data sharing,
    - and routine systematic forecast evaluations
  - Technical Panel Substantive Recommendations
    - For some: token dismissals in the Trustees Report
    - For others: the Trustees Report contradicts the Panel, repeats identically worded assertions year after year, without engaging the Panel or the crucial issues raised
    - The Trustees and Technical Panel agree on many issues too, but the lack of engagement or mutual understanding is obvious
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
Ignoring Technical Panel Recommendations

**Process:**

- OCACT is extremely responsive in providing information
- “Steve Goss has a seat at every table” when policy is made
Ignoring Technical Panel Recommendations

Process:
- OCACT is extremely responsive in providing information
- “Steve Goss has a seat at every table” when policy is made

Technical Panel Methodological Recommendations

Little evidence of serious engagement: After each Panel, for the last 15 years:
- OCACT adopts a few recommendations,
- ignores many,
- and does not come close to the achievable ideal

Little progress on most important issues:
- Adopting formal statistical procedures,
- formal uncertainty estimates,
- transparency,
- data sharing,
- and routine systematic forecast evaluations

Technical Panel Substantive Recommendations

For some: token dismissals in the Trustees Report
For others: the Trustees Report contradicts the Panel, repeats identically worded assertions year after year, without engaging the Panel or the crucial issues raised

The Trustees and Technical Panel agree on many issues too, but the lack of engagement or mutual understanding is obvious
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years:
    - Adopting formal statistical procedures,
    - Formal uncertainty estimates,
    - Transparency,
    - Data sharing,
    - Routine systematic forecast evaluations
Ignoring Technical Panel Recommendations

**Process:**
- OCACT is extremely responsive in providing information
- “Steve Goss has a seat at every table” when policy is made

**Technical Panel Methodological Recommendations**
- Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations,
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made
- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many,
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - Little progress on most important issues:
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - Little progress on most important issues: Adopting formal statistical procedures,
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - Little progress on most important issues: Adopting formal statistical procedures, formal uncertainty estimates,
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - **Little evidence of serious engagement:** After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - **Little progress on most important issues:** Adopting formal statistical procedures, formal uncertainty estimates, transparency,
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - Little progress on most important issues: Adopting formal statistical procedures, formal uncertainty estimates, transparency, data sharing,
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - Little progress on most important issues: Adopting formal statistical procedures, formal uncertainty estimates, transparency, data sharing, and routine systematic forecast evaluations
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - Little progress on most important issues: Adopting formal statistical procedures, formal uncertainty estimates, transparency, data sharing, and routine systematic forecast evaluations

- **Technical Panel Substantive Recommendations**
Ignoring Technical Panel Recommendations

- **Process:**
  - OCAct is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - Little evidence of serious engagement: After each Panel, for the last 15 years: OCAct adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - Little progress on most important issues: Adopting formal statistical procedures, formal uncertainty estimates, transparency, data sharing, and routine systematic forecast evaluations

- **Technical Panel Substantive Recommendations**
  - For some: token dismissals in the Trustees Report
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - *Little evidence of serious engagement:* After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - *Little progress on most important issues:* Adopting formal statistical procedures, formal uncertainty estimates, transparency, data sharing, and routine systematic forecast evaluations

- **Technical Panel Substantive Recommendations**
  - For some: *token dismissals* in the Trustees Report
  - For others: the Trustees Report contradicts the Panel, repeats *identically worded assertions* year after year, without engaging the Panel or the crucial issues raised
Ignoring Technical Panel Recommendations

- **Process:**
  - OCACT is extremely responsive in providing information
  - “Steve Goss has a seat at every table” when policy is made

- **Technical Panel Methodological Recommendations**
  - **Little evidence of serious engagement:** After each Panel, for the last 15 years: OCACT adopts a few recommendations, ignores many, and does not come close to the achievable ideal
  - **Little progress on most important issues:** Adopting formal statistical procedures, formal uncertainty estimates, transparency, data sharing, and routine systematic forecast evaluations

- **Technical Panel Substantive Recommendations**
  - For some: **token dismissals** in the Trustees Report
  - For others: the Trustees Report contradicts the Panel, repeats identically worded assertions year after year, without engaging the Panel or the crucial issues raised
  - The Trustees and Technical Panel agree on many issues too, but the lack of engagement or mutual understanding is obvious
Ignoring Technical Panel Recommendations
Ignoring Technical Panel Recommendations
E.g., Ultimate Rates of (All-Cause) Mortality Decline Assumptions
Ignoring Technical Panel Recommendations

E.g., Ultimate Rates of (All-Cause) Mortality Decline Assumptions

- Trustees Report Assumptions
- Technical Advisory Panel Recommendations

<table>
<thead>
<tr>
<th>Year</th>
<th>Ultimate Rate of Decline of Mortality (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.5</td>
</tr>
<tr>
<td>2002</td>
<td>0.7</td>
</tr>
<tr>
<td>2004</td>
<td>0.9</td>
</tr>
<tr>
<td>2006</td>
<td>1.1</td>
</tr>
<tr>
<td>2008</td>
<td>1.3</td>
</tr>
<tr>
<td>2010</td>
<td>Higher LE &amp; Higher Cost</td>
</tr>
<tr>
<td>2012</td>
<td>Lower LE &amp; Lower Cost</td>
</tr>
<tr>
<td>2014</td>
<td></td>
</tr>
</tbody>
</table>
So what explains the bias?
So what explains the bias?

- OCACT is vulnerable to bias, unprotected because they haven’t:
  - Removed human judgment where possible
  - Instituted formal structural procedures, when judgment is required
  - Required transparency and data sharing
  - Massively more intense & complicated politics than ever (details in our paper)
  - Actuaries hunkered down, insulated themselves, refused to budge

In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
So what explains the bias?

- OCACT is vulnerable to bias, unprotected because they haven’t:
  - Removed human judgment where possible

 Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
So what explains the bias?

- **OCACT** is vulnerable to bias, unprotected because they haven’t:
  - Removed human judgment where possible
  - Instituted formal structural procedures, when judgment is required
So what explains the bias?

- **OCACT is vulnerable to bias, unprotected because they haven’t:**
  - Removed human judgment where possible
  - Instituted formal structural procedures, **when judgment is required**
  - Required transparency and data sharing

Massively more intense & complicated politics than ever (details in our paper) Actuaries hunkered down, insulated themselves, refused to budge In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
So what explains the bias?

- OCACT is vulnerable to bias, unprotected because they haven’t:
  - Removed human judgment where possible
  - Instituted formal structural procedures, when judgment is required
  - Required transparency and data sharing

- Massively more intense & complicated politics than ever (details in our paper)
So what explains the bias?

- **OCACT** is vulnerable to bias, unprotected because they haven’t:
  - Removed human judgment where possible
  - Instituted formal structural procedures, when judgment is required
  - Required transparency and data sharing

- Massively more intense & complicated politics than ever (details in our paper)

- Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
So what explains the bias?

- **OCACT** is vulnerable to bias, unprotected because they haven’t:
  - Removed human judgment where possible
  - Instituted formal structural procedures, when judgment is required
  - Required transparency and data sharing

- Massively more intense & complicated politics than ever (details in our paper)

- Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes

- In the process, they also insulated themselves from the facts:
So what explains the bias?

- OCACT is vulnerable to bias, unprotected because they haven’t:
  - Removed human judgment where possible
  - Instituted formal structural procedures, when judgment is required
  - Required transparency and data sharing
- Massively more intense & complicated politics than ever (details in our paper)
- Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
- In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives
So what explains the bias?

- OCACT is vulnerable to bias, unprotected because they haven’t:
  - Removed human judgment where possible
  - Instituted formal structural procedures, when judgment is required
  - Required transparency and data sharing

- Massively more intense & complicated politics than ever (details in our paper)

- Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes

- In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
E.g.: Surprisingly Large Mortality Declines Since 2000
E.g.: Surprisingly Large Mortality Declines Since 2000
(Slopes from regression of log(mortality) on time from previous 10 years)
E.g.: Surprisingly Large Mortality Declines Since 2000
(Slopes from regression of log(mortality) on time from previous 10 years)
Conclusions

The Problem

Informal forecasting methods ⇝ the potential for bias
Civil servants working hard to resist intense pressure ⇝ insulation from the data as well
Nontransparency, little data sharing ⇝ no course corrections
Systematically & increasingly biased forecasts since 2000
Without better procedures, you or I could not do better

The Solution: Professionalize

Remove human judgment where possible, via formal statistical methods — via the data science revolution
Institute formal structural procedures when human judgment is required — via the social psychological revolution
Require transparency and data sharing to catch errors that slip through — via the scientific revolution

For more information:
GaryKing.org
Conclusions

- The Problem
Conclusions

- The Problem
  - Informal forecasting methods \( \leadsto \) the potential for bias
Conclusions

- **The Problem**
  - Informal forecasting methods \(\rightsquigarrow\) the potential for bias
  - Civil servants working hard to resist intense pressure \(\rightsquigarrow\) insulation from the data as well
Conclusions

The Problem

- Informal forecasting methods $\rightsquigarrow$ the potential for bias
- Civil servants working hard to resist intense pressure $\rightsquigarrow$ insulation from the data as well
- Nontransparency, little data sharing $\rightsquigarrow$ no course corrections

The Solution: Professionalize

- Remove human judgment where possible, via formal statistical methods — via the data science revolution
- Institute formal structural procedures when human judgment is required — via the social psychological revolution
- Require transparency and data sharing to catch errors that slip through — via the scientific revolution

For more information: GaryKing.org
Conclusions

The Problem

- Informal forecasting methods ⇒ the potential for bias
- Civil servants working hard to resist intense pressure ⇒ insulation from the data as well
- Nontransparency, little data sharing ⇒ no course corrections
- Systematically & increasingly biased forecasts since 2000
Conclusions

- **The Problem**
  - Informal forecasting methods $\leadsto$ the potential for bias
  - Civil servants working hard to resist intense pressure $\leadsto$ insulation from the data as well
  - Nontransparency, little data sharing $\leadsto$ no course corrections
  - Systematically & increasingly biased forecasts since 2000
  - Without better procedures, you or I could not do better

For more information: GaryKing.org
Conclusions

- **The Problem**
  - Informal forecasting methods $\leadsto$ the potential for bias
  - Civil servants working hard to resist intense pressure $\leadsto$ insulation from the data as well
  - Nontransparency, little data sharing $\leadsto$ no course corrections
  - Systematically & increasingly biased forecasts since 2000
  - Without better procedures, you or I could not do better

- **The Solution: Professionalize**
  - Remove human judgment where possible, via formal statistical methods — via the data science revolution
  - Institute formal structural procedures when human judgment is required — via the social psychological revolution
  - Require transparency and data sharing to catch errors that slip through — via the scientific revolution

For more information: GaryKing.org
Conclusions

- **The Problem**
  - Informal forecasting methods $\leadsto$ the potential for bias
  - Civil servants working hard to resist intense pressure $\leadsto$ insulation from the data as well
  - Nontransparency, little data sharing $\leadsto$ no course corrections
  - Systematically & increasingly biased forecasts since 2000
  - Without better procedures, you or I could not do better
- **The Solution: Professionalize**
  - Remove human judgment where possible, via formal statistical methods

For more information: GaryKing.org
Conclusions

The Problem
- Informal forecasting methods \(\rightarrow\) the potential for bias
- Civil servants working hard to resist intense pressure \(\rightarrow\) insulation from the data as well
- Nontransparency, little data sharing \(\rightarrow\) no course corrections
- Systematically & increasingly biased forecasts since 2000
- Without better procedures, you or I could not do better

The Solution: Professionalize
- Remove human judgment where possible, via formal statistical methods
  — via the data science revolution
Conclusions

- **The Problem**
  - Informal forecasting methods $\rightarrow$ the potential for bias
  - Civil servants working hard to resist intense pressure $\rightarrow$ insulation from the data as well
  - Nontransparency, little data sharing $\rightarrow$ no course corrections
  - Systematically & increasingly biased forecasts since 2000
  - Without better procedures, you or I could not do better

- **The Solution: Professionalize**
  - Remove human judgment where possible, via formal statistical methods — via the data science revolution
  - Institute formal structural procedures when human judgment is required

For more information: GaryKing.org
Conclusions

- **The Problem**
  - Informal forecasting methods $\leadsto$ the potential for bias
  - Civil servants working hard to resist intense pressure $\leadsto$ insulation from the data as well
  - Nontransparency, little data sharing $\leadsto$ no course corrections
  - Systematically & increasingly biased forecasts since 2000
  - Without better procedures, you or I could not do better

- **The Solution: Professionalize**
  - Remove human judgment where possible, via formal statistical methods — via the data science revolution
  - Institute formal structural procedures when human judgment is required — via the social psychological revolution

For more information: GaryKing.org
Conclusions

- The Problem
  - Informal forecasting methods \(\rightarrow\) the potential for bias
  - Civil servants working hard to resist intense pressure \(\rightarrow\) insulation from the data as well
  - Nontransparency, little data sharing \(\rightarrow\) no course corrections
  - Systematically & increasingly biased forecasts since 2000
  - Without better procedures, you or I could not do better

- The Solution: Professionalize
  - Remove human judgment where possible, via formal statistical methods — via the data science revolution
  - Institute formal structural procedures when human judgment is required — via the social psychological revolution
  - Require transparency and data sharing to catch errors that slip through
Conclusions

The Problem
- Informal forecasting methods \( \Rightarrow \) the potential for bias
- Civil servants working hard to resist intense pressure \( \Rightarrow \) insulation from the data as well
- Nontransparency, little data sharing \( \Rightarrow \) no course corrections
- Systematically & increasingly biased forecasts since 2000
- Without better procedures, you or I could not do better

The Solution: Professionalize
- Remove human judgment where possible, via formal statistical methods — via the data science revolution
- Institute formal structural procedures when human judgment is required — via the social psychological revolution
- Require transparency and data sharing to catch errors that slip through — via the scientific revolution
Conclusions

- **The Problem**
  - Informal forecasting methods $\Rightarrow$ the potential for bias
  - Civil servants working hard to resist intense pressure $\Rightarrow$ insulation from the data as well
  - Nontransparency, little data sharing $\Rightarrow$ no course corrections
  - Systematically & increasingly biased forecasts since 2000
  - Without better procedures, you or I could not do better

- **The Solution: Professionalize**
  - Remove human judgment where possible, via formal statistical methods — via the data science revolution
  - Institute formal structural procedures when human judgment is required — via the social psychological revolution
  - Require transparency and data sharing to catch errors that slip through — via the scientific revolution

For more information:

GaryKing.org