Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts

Konstantin Kashin, Gary King, Samir Soneji

Institute for Quantitative Social Science
Harvard University

References (all forthcoming)
References (all forthcoming)

- Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  
  *Journal of Economic Perspectives*
References (all forthcoming)

- 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 (all forthcoming)

- **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*

- **Online Appendix: Systematic Bias and Nontransparency in US Social Security Administration Forecasts**
  
  *Journal of Economic Perspectives*
References (all forthcoming)

- 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*

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

- Replication Data for: Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *Journal of Economic Perspectives Archive, and Dataverse*
References (all forthcoming)

- 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*

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

- Replication Data for: Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *Journal of Economic Perspectives Archive, and Dataverse*

- Replication Data for: Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *The Political Analysis Dataverse*
References (all forthcoming)

- 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*

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

- Replication Data for: Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *Journal of Economic Perspectives Archive, and Dataverse*

- Replication Data for: Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts
  *The Political Analysis Dataverse*

- (Results and data shared with SSA Technical Panel: 11/2014)
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

Social Security is the single largest U.S. government program, representing 37% of federal outlays ($1.3T in 2013 expenditures). It brings 20% of elderly Americans above the poverty level. Proposals for change are 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
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
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.
  - 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”
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)
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:
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
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 $\rightsquigarrow$ 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
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
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.
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 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

The standard is not whether OCACT thinks they’ve shared enough; it’s whether they have made it easy enough for others to contribute

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

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

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - 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

The standard is not whether OCACT thinks they’ve shared enough; it’s whether they have made it easy enough for others to contribute.

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

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

- Who has been able to fully replicate OCACT’s forecasts?
  - 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

The standard is not whether OCACT thinks they’ve shared enough; it’s whether they have made it easy enough for others to contribute

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

The standard is not whether OCACT thinks they’ve shared enough; it’s whether they have made it easy enough for others to contribute

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

- Who forecasts independently of SSA’s Office of the Chief Actuary?
  - No one
- Who has been able to fully replicate OCACIT’s forecasts?
  - No one
  - Some data shared: in difficult, disorganized, non-automated formats

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

The standard is not whether OCACT thinks they’ve shared enough; it’s whether they have made it easy enough for others to contribute

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

The standard is not whether OCACT thinks they’ve shared enough; it’s whether they have made it easy enough for others to contribute
Nontransparency in 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

- The standard is not whether OCACT thinks they’ve shared enough;
  it’s whether they have made it easy enough for others to contribute

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

- The standard is not whether OCACT thinks they’ve shared enough; it’s whether they have made it easy enough for others to contribute

- 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

⇝ Policy scores: mostly indistinguishable from random noise

Straightforward solutions exist for all problems discovered
Evaluating SSA Forecasts

- The history of all systematic evaluations of SSA forecasts:
  - None by SSA
  - None by others: (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 ⇝ Policy scores: mostly indistinguishable from random noise

Straightforward solutions exist for all problems discovered
Evaluating SSA Forecasts

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

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
    ⇝ Policy scores: mostly indistinguishable from random noise

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:

  - 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
    - Policy scores: mostly indistinguishable from random noise

  - Straightforward solutions exist for all problems discovered
Evaluating SSA Forecasts

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

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

⇝ Policy scores: mostly indistinguishable from random noise

Straightforward solutions exist for all problems discovered
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

How big is the bias?
- Larger than almost all of OCACT's policy scores
  ⇒ Policy scores: mostly indistinguishable from random noise

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:
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,
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

How big is the bias?

Larger than almost all of OCACT's policy scores

→ Policy scores: mostly indistinguishable from random noise

Straightforward solutions exist for all problems discovered
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
    ~ Policy scores: mostly indistinguishable from random noise
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 OC ACT’s policy scores
  - Policy scores: mostly indistinguishable from random noise
  - Straightforward solutions exist for all problems discovered
How OCACT Forecasts

SSA methods: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades—a period with breathtaking advances in big data, data science, statistics, and social psychology.

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.
- \( t + 26 \) to \( t + 75 \): change linearly from “historical” to “ultimate”.

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:**

  - 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.
How OCACT Forecasts

- **SSA methods**: Jerry rigged,

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**: Jerry rigged, suboptimal,
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc,
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable,
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology
- **Example**: Mortality Forecasts
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

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

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

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

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology
- **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**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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 $t + 26$
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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:** Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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,
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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 $0.75 \times \text{“historical”}$ if negative
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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”
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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
How OCACT Forecasts

- **SSA methods**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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 $0.75 \times \text{"historical"}$ 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**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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:** Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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**: Jerry rigged, suboptimal, ad hoc, not replicable, and little changed in decades — a period with breathtaking advances in big data, data science, statistics, and social psychology

- **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
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: \(\approx\) linear, different slopes, different variances,
Actual Mortality Time Profiles are Complex

Patterns: ≈ 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
Actual Mortality Age Profiles are also Complex

**Patterns:**

- Characteristic shape,
- Partially linear,
- Different time x 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
OCACT Qualitative Choices: Violate Known Information

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

Diabetes Death Rate, Males

Ages 80–84
Ages 85–89

Cancer Death Rate, Females

Ages 80–84
Ages 85–89

 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

Death Rate (per 100,000)

Year

Ages 50–54
Ages 45–49

Heart Disease Death Rate, Males

Death Rate (Log)

Year

2050
2060
2070
2080
2090
2100

Age

0
25
50
75

Unrealistic patterns:
Crossing age plots,
notch for 50-75 year olds
Unrealistic patterns:

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

Unrealistic patterns: Crossing age plots,
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

Patterns:
Vertical: Later Trustees Reports are overconfident
Not horizontal: Shorter term forecasts should be better, but aren’t
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
Trust Fund Ratio Forecasting Errors: 1 Year Ahead
Trust Fund Ratio Forecasting Errors: 2 Years Ahead

![Graph showing forecast errors over years](image-url)
Trust Fund Ratio Forecasting Errors: 3 Years Ahead

![Graph showing forecast vs. truth across years from 1980 to 2010.](image)
Trust Fund Ratio Forecasting Errors: 4 Years Ahead

![Graph showing forecast vs. truth for trust fund ratios over the years 1980 to 2010. The graph displays the difference between forecasted and actual values.](image-url)
Trust Fund Ratio Forecasting Errors: 5 Years Ahead

![Graph showing the relationship between forecast and truth values over years. The graph plots the forecast minus truth values against the year of the Trustees Report, with a trend line and shaded area indicating the range of error.](image-url)
Trust Fund Ratio Forecasting Errors: 6 Years Ahead
Trust Fund Ratio Forecasting Errors: 7 Years Ahead

-50 0 50 100 150

1980 1990 2000 2010

Year of Trustees Report

Forecast − Truth
Trust Fund Ratio Forecasting Errors: 8 Years Ahead
Trust Fund Ratio Forecasting Errors: 9 Years Ahead
Trust Fund Ratio Forecasting Errors: Summary

Year of Trustees Report vs. Forecast Error (Percentage Points)

1. 1980-2010
2. 1980-2010
3. 1980-2010
4. 1980-2010
5. 1980-2010
6. 1980-2010
7. 1980-2010
8. 1980-2010
9. 1980-2010
10. 1980-2010
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;

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 $\alpha \leq 0.05$?

These are extremely optimistic assumptions
Uncertainty Estimates for OCACT Policy Scores

- Who scores SSA Policy Proposals?
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

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 $\alpha \leq 0.05$?

These are extremely optimistic 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:

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 $\alpha \leq 0.05$?

These are extremely optimistic 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;

- 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 $\alpha \leq 0.05$?

These are extremely optimistic 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

- 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 $\alpha \leq 0.05$?

  - These are extremely optimistic 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:

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 $\alpha \leq 0.05$?

These are extremely optimistic 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;

- **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 \( \alpha \leq 0.05 \)?

- These are extremely optimistic 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 > 95th percentile i.e., with $\alpha \leq 0.05$?
- These are extremely optimistic 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;

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 $\alpha \leq 0.05$?

These are extremely optimistic 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**: 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 $\alpha \leq 0.05$?

$\Rightarrow$ These are extremely optimistic 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:
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
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
  - 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;
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 $\alpha \leq 0.05$? $\leadsto$ These are extremely optimistic 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

10-Year

75-Year
Social Psychological Conditions that make Bias Possible

Bias: Systematic errors, regardless of intention or direction

The social psychological literature suggests that 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 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:

 [Content continues on the next page]
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,

“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)
“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,
“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-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
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 situation & procedures

"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, 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 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-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 situation & procedures
- **Qualitative uncertainty estimates are also likely biased**
  - “Experts” are usually overconfident.
“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 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 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 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 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-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 — exactly OCACT’s 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.)
     - 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
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.)
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

   - 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
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

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 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:

- 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”
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
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
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,
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,
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
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”
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:**

  - [Details about consistency bias are not visible in the image]
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
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
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
  - 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 → 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 \(\leadsto\) 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 ⨆ 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 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:
    - 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,
- 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:
      - 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**
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:
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

![Graph showing Ultimate Rate of Decline of Mortality (Year 2000 to 2014)](image)

- Trustees Report Assumptions
- Higher LE & Higher Cost
- Technical Advisory Panel Recommendations
- Lower LE & Lower Cost

20/23
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.)
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

要求透明和数据共享

- 大程度更复杂的政治（细节在我们的论文中）

- 具体来说，从2000年开始，美国人开始活得更长（由于他汀类药物、早期癌症检测等）
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
The Problem

- Informal forecasting methods $\rightarrow$ the potential for bias

Conclusions

- The Problem
  - Informal forecasting methods $\rightarrow$ the potential for bias
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
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

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

For more information: GaryKing.org
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
- 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
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
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
  - 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
  - Require transparency and data sharing to catch errors that slip through

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
  - 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
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

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
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
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