Optimal Policy Perturbations

Regis Barnichon (FRB San Francisco)
Geert Mesters (Universitat Pompeu Fabra)

Context

Although contention about the appropriate model of the economy [...] continues, macroeconomic policy decisions have to be made.

Blanchard and Fisher (1989)

- Policy makers use heuristics to decide on policy actions:
  - Combine insights from multiple models
  - Rely heavily on instinct and judgement calls
- How to know whether the policy choice is the best one?

Illustrative example

The COVID crisis:

- Is the Fed doing enough?
- Should it be more aggressive with QE?
- ...

This paper

- Start from high-level loss function *given by* policy maker
- Propose a statistic —the Optimal Policy Perturbation (OPP)— to detect optimization failures in policy process
- OPP does not rely on specifying an underlying model
- OPP informs whether chosen policy is optimal and, if not, which improvements can be made

Two perspectives for the OPP

1. A researcher interested in assessing the historical performance of policy makers

2. A tool to help decision making in real time
3. A tool to articulate policy prescriptions around three concepts
    3.1 preferences
    3.2 economic outlook
    3.3 effects of policy

Idea of policy perturbation

- Idea similar to Sufficient Statistic approach, but in a macro stabilization setting
- Explore whether a perturbation to the policy choice can lower the loss function
- Exploits idea that *at the optimum* perturbations should have no first-order effect on loss function

Idea of policy perturbation

- Idea similar to Sufficient Statistic approach, but in a macro stabilization setting
- Explore whether a perturbation to the policy choice can lower the loss function
- Exploits idea that *at the optimum* perturbations should have no first-order effect on loss function
- OPP is a *well-chosen* perturbation that only requires
    1. Forecasts for the policy objectives given the policy choice
    2. Impulse response of policy objectives to changes in the policy instruments

Applicability

- OPP can be applied to a broad range of macro policy problems
  - Monetary policy
  - Fiscal policy with stabilization vs budget deficit concerns
  - Exchange rate management
  - Foreign exchange reserve management
  - ...
- Today we illustrate the OPP for US monetary policy decisions

This talk

1. Problem description
2. Optimal Policy Perturbation
3. Inference
4. US monetary policy

Problem description

Environment policy maker

Policy maker has

- $m = 1, \ldots, M$ mandates over $h = 0, \ldots, H$ horizons
- target variables $y_{m,t+h}$ with objective y_m^*
- K policy instruments $p_t = (p_{1,t}, \ldots, p_{K,t})'$
- preference parameters λ_m and discount factors β_h

Policy maker's problem

Policy maker (under discretion) aims to solve

$$\min_{p_t \in \mathcal{D}} \mathcal{L}_t$$

with loss function

$$\mathcal{L}_t = \mathbb{E}_t \sum_{h=0}^{H} \sum_{m=1}^{M} \lambda_m \beta_h \left(y_{m,t+h} - y_m^*\right)^2$$

where $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$.

Convenient static re-formulation (1)

- Stack targets

$$Y_{t:t+H} = [\sqrt{\lambda_j \beta_h}(y_{m,t+h} - y_m^*)]_{m=1,\dots,M,h=0,\dots,H}$$

- To ease on notations, refer to $Y_{t:t+H}$ as Y_t
- Postulate generic model

$$Y_t = f_t(p_t, X_t) + \xi_t \ ,$$

f_t differentiable wrt p_t, X_t is $\mathcal{F}_t$ measurable, ξ_t future shocks

Convenient static re-formulation (2)

- We obtain *static* description of the *dynamic* policy problem

$$\min_{p_t \in \mathcal{D}} \mathcal{L}_t, \qquad \mathcal{L}_t = \mathbb{E}_t \| Y_t \|^2$$

$$\text{s.t. } Y_t = f_t(p_t, X_t) + \xi_t$$

- K instruments to hit $(H+1)M$ targets

Policy maker's proposed solution

- Policy maker proposes p_t^0 as the solution to the policy problem
- Goal of the paper:

  Verify whether p_t^0 is optimal without knowing f_t

Policy maker's proposed solution

- Policy maker proposes p_t^0 as the solution to the policy problem
- Goal of the paper:

  Verify whether p_t^0 is optimal without knowing f_t
- Not the goal of the paper:

  Derive an optimal policy rule

Why would p_t^0 NOT be optimal?

- Mistake
- Mis-specification, if policy maker doesn't have access to true $f_t(.)$
- Complexity of $f_t(.)$ makes it very expensive to evaluate
  - Example: Board Fed production of Tealbook
  - Computing $f_t(.)$ involves many model iterations and judgment calls
  - Impossible to compute $f_t(.)$ for all possible p_t
  - An incomplete grid search

Optimal Policy Perturbation

Policy perturbations

Main idea:

- Perturbate p_t^0 with $\delta_t = (\delta_{1,t}, \ldots, \delta_{K,t})'$
- If $p_t^0 + \delta_t$ generates lower loss, conclude p_t^0 is not optimal
- Is there a smart choice for δ_t?

Optimal policy perturbations (OPP)

We consider the perturbation

$$\delta_t^* = -(\mathcal{R}_t'\mathcal{R}_t)^{-1}\mathcal{R}_t'\mathbb{E}_t Y_t^0$$

which depends on

- impulse responses

$$\mathcal{R}_t \equiv \mathcal{R}_t(p_t^0, X_t) = \left.\frac{\partial f_t(p_t, X_t)}{\partial p_t'}\right|_{p_t = p_t^0}$$

- forecasts

$$\mathbb{E}_t Y_t^0 \equiv \mathbb{E}_t f_t(p_t^0, X_t)$$

Optimal policy perturbations (OPP)

We consider the perturbation

$$\delta_t^* = -(\mathcal{R}_t'\mathcal{R}_t)^{-1}\mathcal{R}_t'\mathbb{E}_t Y_t^0$$

which depends on

- impulse responses

$$\mathcal{R}_t \equiv \mathcal{R}_t(p_t^0, X_t) = \left.\frac{\partial f_t(p_t, X_t)}{\partial p_t'}\right|_{p_t = p_t^0}$$

- forecasts

$$\mathbb{E}_t Y_t^0 \equiv \mathbb{E}_t f_t(p_t^0, X_t)$$

Note: δ_t^* is a function of $\mathcal{F}_t$ not a shock

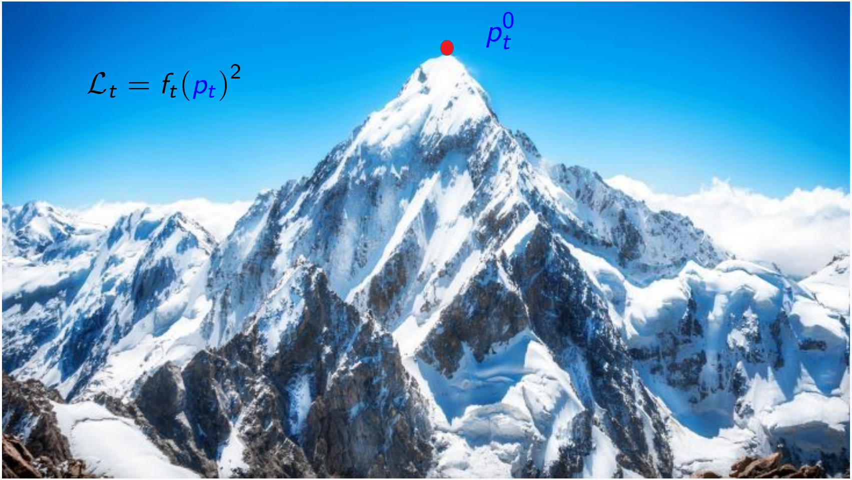
Properties of OPP

What can we learn from OPP δ_t^*?

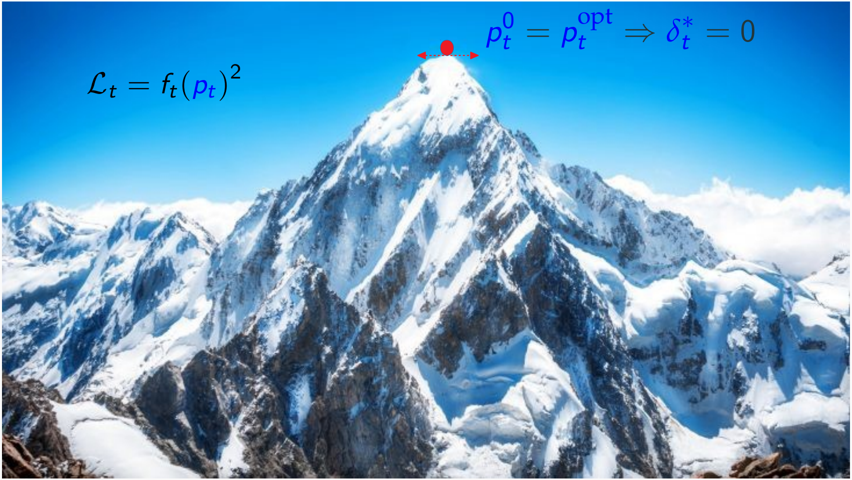
1. Discarding optimality: when p_t^0 is not optimal
2. Improving policy: when $p_t^0 + \delta_t^*$ improves p_t^0
3. Optimal policy: when $p_t^0 + \delta_t^*$ optimal

Intuition: Discarding Optimality

Intuition: Discarding Optimality



Intuition: Discarding Optimality



Intuition: Discarding Optimality

$$\delta_t^* = -(\mathcal{R}_t'\mathcal{R}_t)^{-1}\mathcal{R}_t'\mathbb{E}_t Y_t^0$$

- At the optimum,

$$\left.\frac{\partial \mathcal{L}}{\partial p_t}\right|_{p_t=p_t^0} = 2\mathcal{R}_t'\mathbb{E}_t Y_t^0 = 0 \qquad \Rightarrow \qquad \delta_t^* = 0$$

- Impulse responses (IR) should be orthogonal to forecasts $\Rightarrow$ There is no adjustment to the instruments, i.e., no combination of the IRs, that can lower the loss function

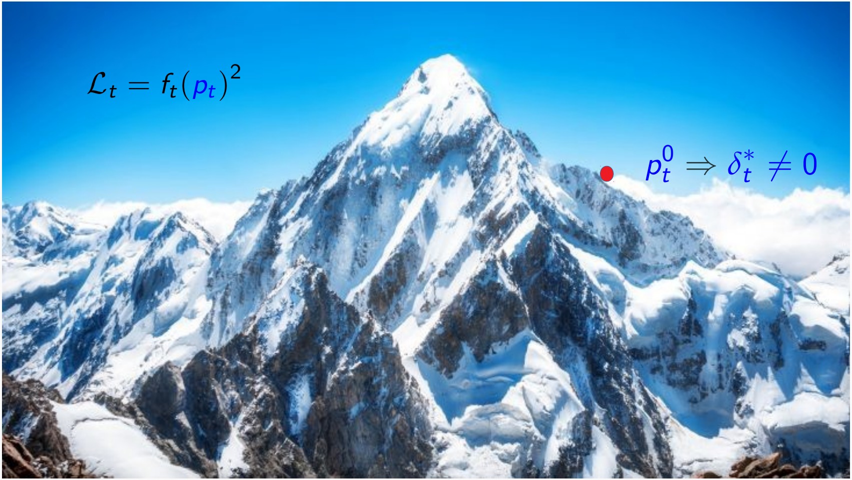
Intuition: Improving/Optimal Policy

$$\delta_t^* = -(\mathcal{R}_t'\mathcal{R}_t)^{-1}\mathcal{R}_t'\mathbb{E}_t Y_t^0$$

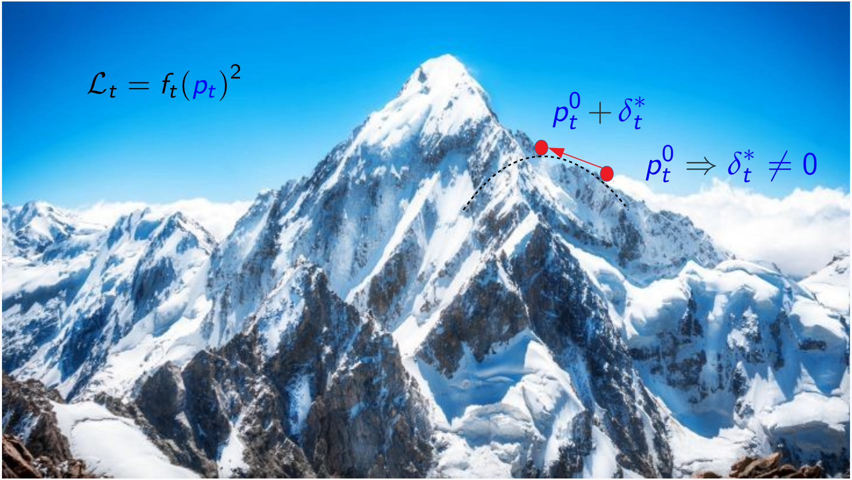
- Optimization perspective:
  $\rightarrow$ OPP as the first-step of a Gauss-Newton algorithm

Intuition: Improving Policy

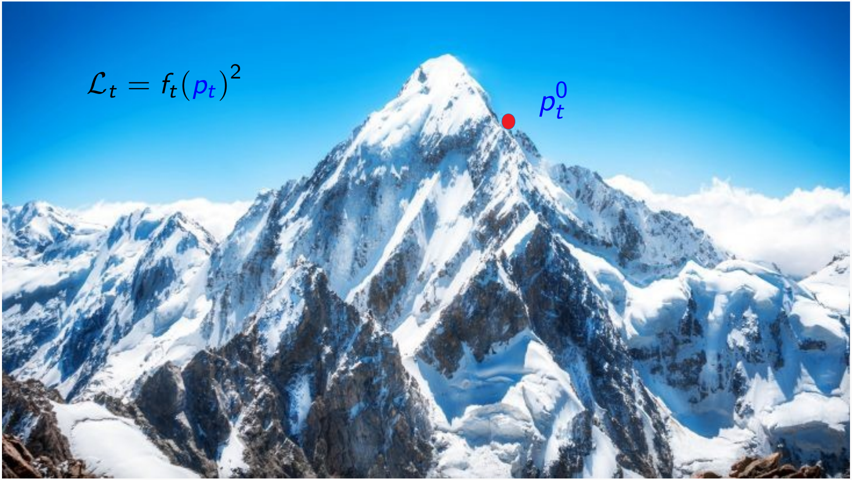
Intuition: Improving Policy



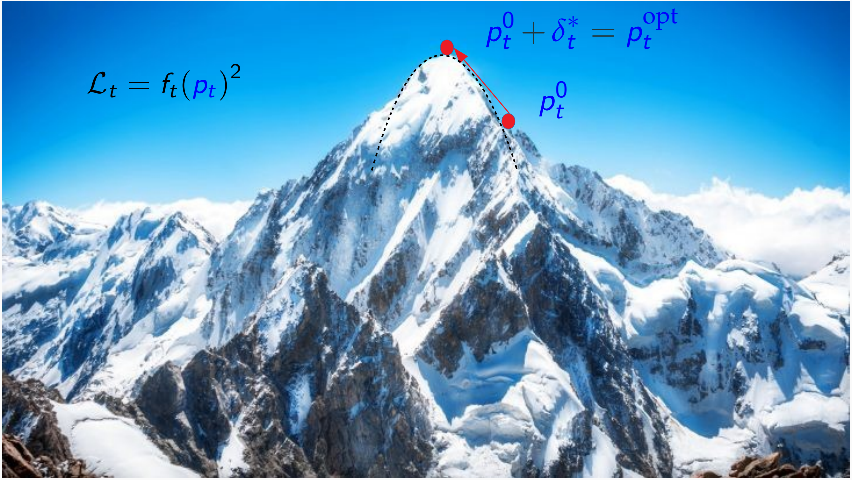
Intuition: Improving Policy



Intuition: Optimal Policy



Intuition: Optimal Policy



Intuition: Optimization perspective

$p_t^0 + \delta_t^*$ is first step of Gauss-Newton (GN) optimization algorithm

- Approximate linearity ($Y_t \approx \mathcal{R}_t \delta_t + Y_t^0$): GN converges
  $\Rightarrow$ OPP improves policy
- Linearity ($Y_t = \mathcal{R}_t \delta_t + Y_t^0$): GN converges in one step
  $\Rightarrow$ OPP gives optimal policy

Two comments related to Lucas critique

1. Assume we know $\mathcal{R}_t$
   - Detecting an optimization failure
     $\Rightarrow$ immune to Lucas critique ($\nabla_{p_t}\mathcal{R}_t \neq 0$)
   - Improve/optimal policy restricts $\mathcal{R}_t(p_t)$
     $\Rightarrow$ not fully robust to Lucas critique
2. Can we know/estimate $\mathcal{R}_t$?
   - Same issue in Sufficient Statistics literature
   - Need sufficient data/experiments relevant for period t

A different viewpoint

Why is Discarding Optimality immune to Lucas critique?

- Estimating the optimal policy is hard (Lucas critique)
- Discarding Optimality is easier, because assessing $\delta_t^* = 0$ is like a score test
    - You can impose the null $p_t^0 = p_t^{opt}$ and use the score at p_t^0
    - No need to estimate optimal policy as it is fixed under the null

Intuition: Econometrics perspective

- OPP formula looks like OLS regression of $\mathbb{E}_t Y_t^0$ on $\mathcal{R}_t$

$$\delta_t^* = -\left(\mathcal{R}_t' \mathcal{R}_t\right)^{-1} \mathcal{R}_t' \mathbb{E}_t Y_t^0.$$

- With linearity assumption, get

$$Y_t^0 = -\mathcal{R}_t \delta_t + Y_t$$

- Goal of δ_t^* is to use $\mathcal{R}_t$ to minimize $\mathbb{E}_t \left\| Y_t \right\|^2$: an OLS reg.!

Illustration: simple example I

- Suppose policy maker has 1 mandate and 1 instrument

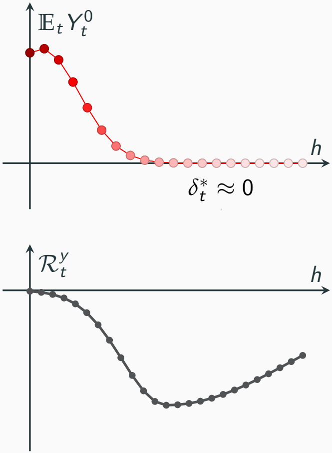
$$\min_{p_t} \mathcal{L}_t \qquad \text{with} \qquad \mathcal{L}_t = \mathbb{E}_t \| Y_t \|^2$$

  where $Y_t = (y_t - y^*, y_{t+1} - y^*, \ldots, y_{t+H} - y^*)'$

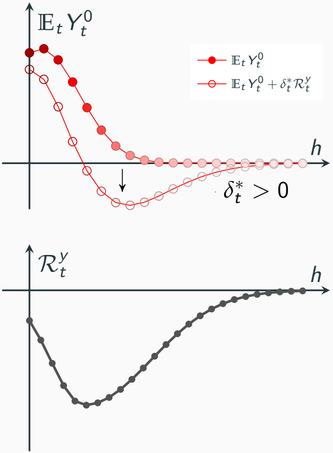
- Two scenarios:
  (a) Failing to disregard optimality
  (b) Discarding optimality

Illustration: simple example I

$$\delta_t^* = -\left(\mathcal{R}_t^{y\prime}\mathcal{R}_t^y\right)^{-1}\mathcal{R}_t^{y\prime}\mathbb{E}_t Y_t^0$$



(a) Failing to discard optimality



(b) Discarding optimality

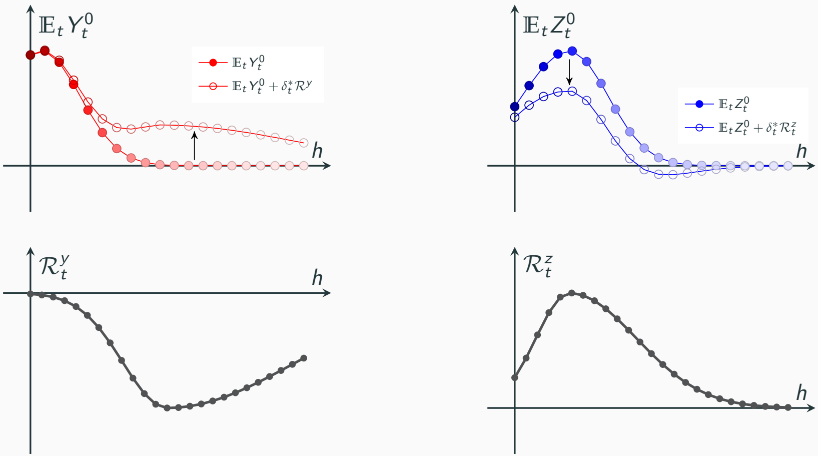
Illustration: simple example II

- Suppose policy maker has 2 mandates and 1 instrument

$$\min_{p_t} \mathcal{L}_t \qquad \text{with} \qquad \mathcal{L}_t = \mathbb{E}_t \| Y_t \|^2 + \mathbb{E}_t \| Z_t \|^2$$

- Now there are two types of trade-offs
  - Across horizons
  - Across mandates

Illustration: simple example II



$$\boldsymbol{\delta_t^* = \omega \delta_t^{y*} + (1-\omega)\delta_t^{z*} < 0}$$

Inference for OPP

Inference for OPP

- Computation of OPP requires two statistics
  - Impulse response $\mathcal{R}_t$
  - Conditional expectation $\mathbb{E}_t Y_t^0$
- In practice
  - $\mathcal{R}_t$ is estimated: **IR estimation uncertainty**
  - Policy maker does not know $\mathbb{E}_t Y_t^0$ (the optimal forecast) and only produces $\widehat{Y}_{t|t}$: **Model uncertainty**

Avoiding type-1 (false positive) error

- We do not want that researcher rejects optimality because of
  - noise in impulse responses
  - model mis-specification (incorrect forecasts)
- Therefore we compute confidence bands around the OPP
- Conservative inference: reject optimality if the confidence bands exclude zero

Inference

- IR estimation uncertainty

$$\hat{r}_t = \mathrm{vec}(\widehat{\mathcal{R}_t}) \sim N(r_t, \Omega_t)$$

- Conditional expectation uncertainty

$$\widehat{Y}_{t|t} \sim N\left(\mathbb{E}_t Y_t^0, \Sigma_{t|t}\right)$$

- Simulate/delta method to get distribution of

$$\delta_t^* = -(\mathcal{R}_t' \mathcal{R}_t)^{-1} \mathcal{R}_t' \mathbb{E}_t Y_t^0$$

A Brainard conservatism principle for the OPP

- Denote by $\widehat{\delta}_t$ the mean of the distribution of δ_t^*
- Can show

$$\widehat{\delta}_t = (\widehat{\mathcal{R}}_t'\widehat{\mathcal{R}}_t + \widetilde{\Omega}_t)^{-1}\widehat{\mathcal{R}}_t'\widehat{Y}_{t|t} \ ,$$

- $\widetilde{\Omega}_t$ captures an **attenuation bias** coming from measurement error in the IRs

Applications of OPP

Two applications of OPP

1. A retrospective analysis of Fed policy
2. A tool to help decision making in real time

Two data requirements

1. Forecasts for π and u
   $\rightarrow$ Survey of Economic Projections from FOMC (1980-2020)

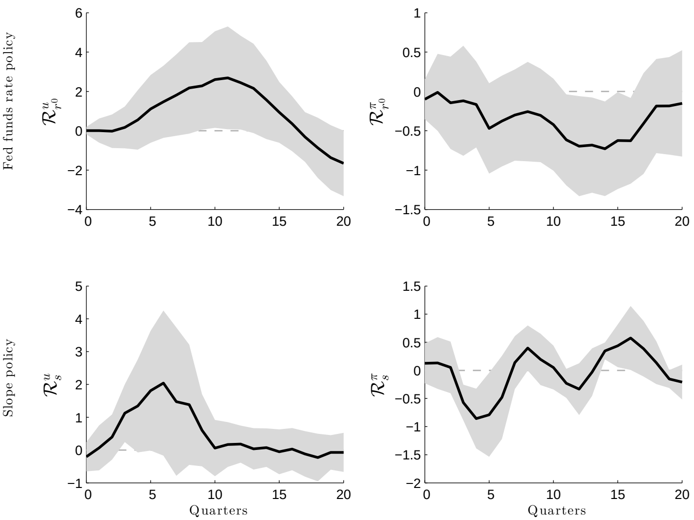
2. Impulse responses to monetary shocks
   - Fed instruments:
     2.1 Fed funds rate
     2.2 Slope of yield curve (LSAP, QE)

     Eberly, Stock and Wright (2019)

Estimation of $\mathcal{R}$

- Estimation by LP-IV
- IVs based on surprises to bond market during 30min window around FOMC announcements Kuttner(2001), Eberly, Stock and Wright (2019)
    - r_t^0 shock: difference between r_t^0 decision and current-month fed funds futures contract $FF1$
    - slope shock: Surprise to $r_t^{10yr} - r_t^0$ spread, holding r_t^0 shock constant

Estimation of $\mathcal{R}$

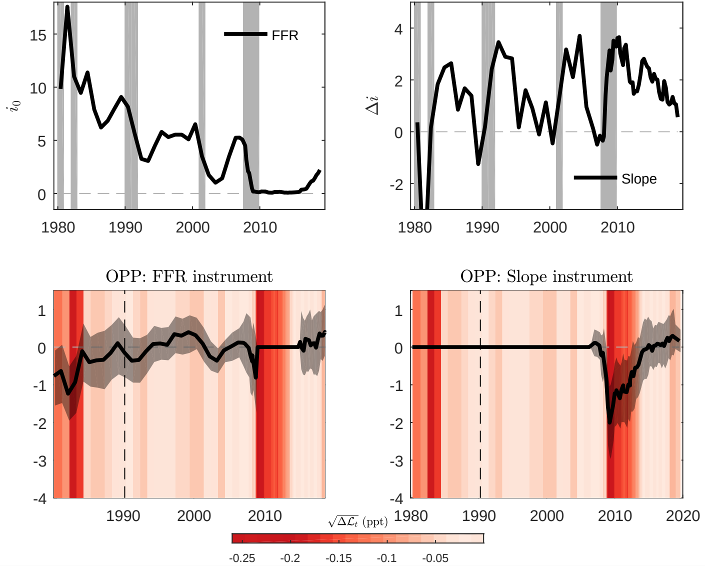


A first application of OPP

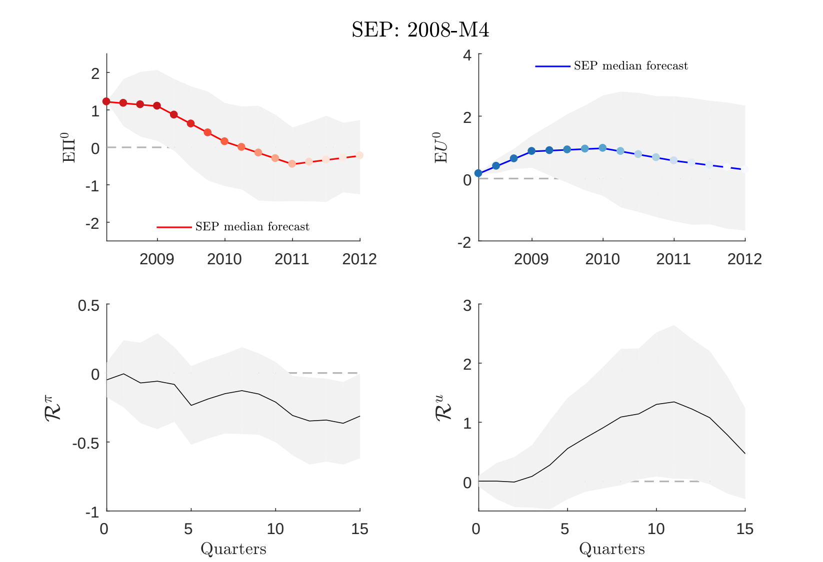
A retrospective analysis of Fed policy

- Fed balanced approach: $\lambda = 1$
- Instruments
  - FFR alone until 2007
  - Slope policy alone over 2008-2013
  - FFR and slope policy over 2014-2018

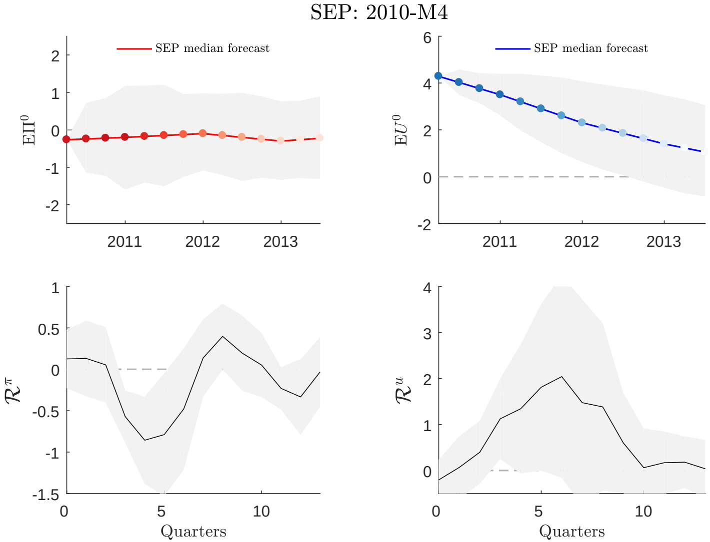
Retrospective analysis of Fed policy



Should the Fed have lowered FFR faster in 2008?



Should the Fed have used more slope policy in 2010?

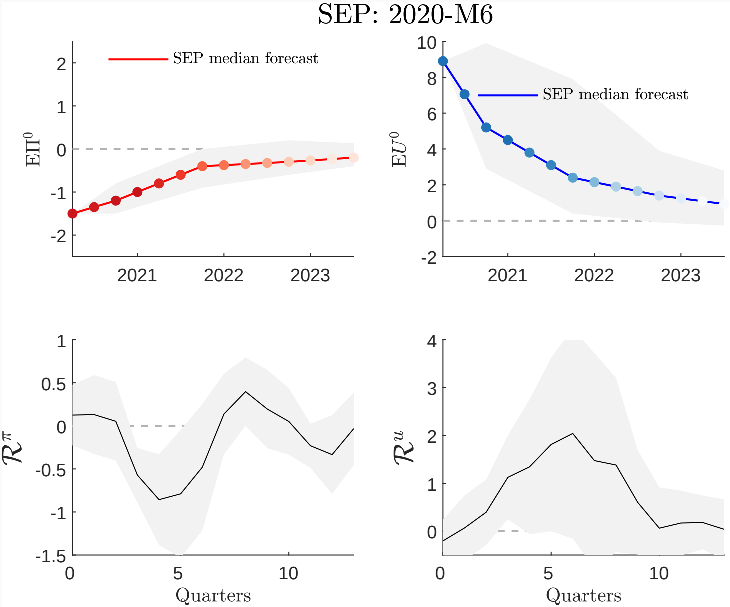


Another application of OPP

A tool for decision making in real time

- Take the FOMC as of June 2020
- Summarize SEP forecasts with two parameters:
  - Second COVID wave?: $\mathbb{E}_t u_{2020q4}$
  - Speed of recovery in 2021-2022 (half-life)
- Capture model uncertainty from SEP dispersion
- Question: Should Fed use its slope policy more aggressively?
- **Warning**: This is an illustration. Yield curve is already very close to flat

Slope policy in 2020-M6



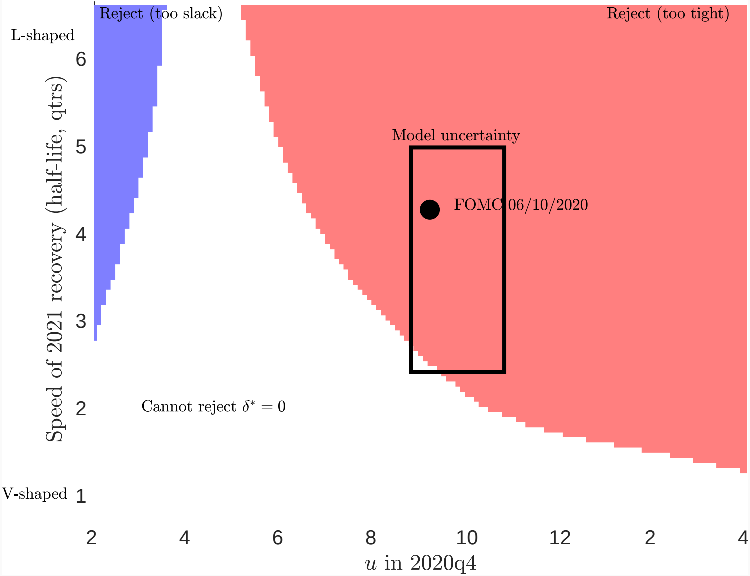
Illustration

A decision map:

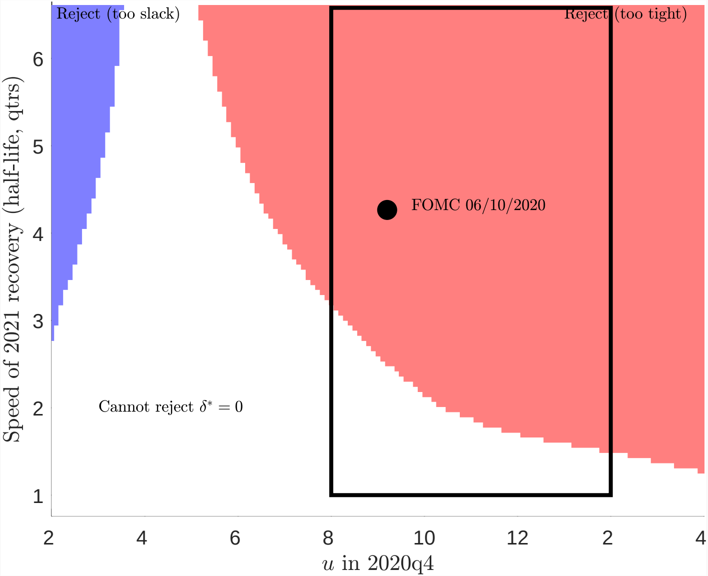
- Show results of "test" $\delta_t^* = 0$ over forecast space
  $\rightarrow$ captures effect of IR uncertainty on test

- Show uncertainty "rectangle" around SEP median forecast
  $\rightarrow$ captures effect of model uncertainty on "test"

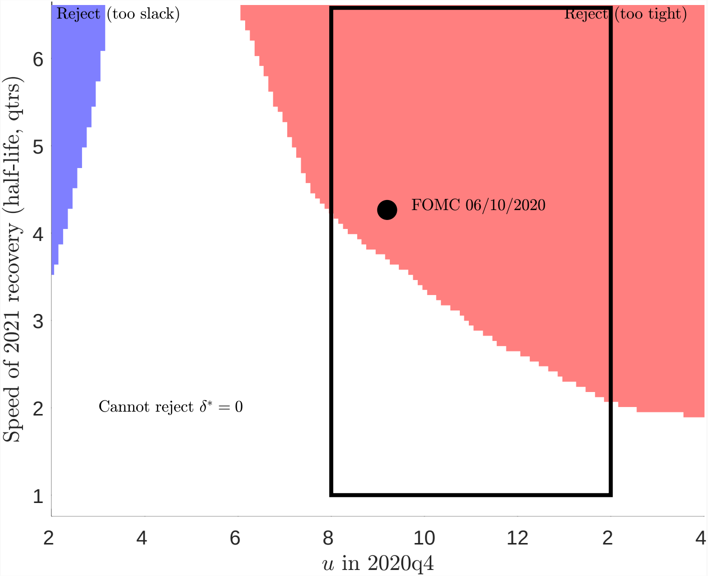
A decision map using SEP “central tendency” around forecast



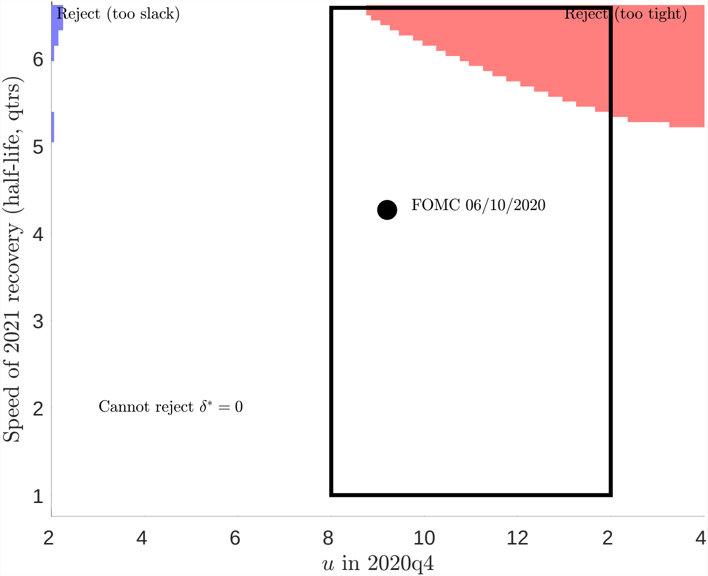
With higher model uncertainty: using range of SEP



With higher IR uncertainty



With even higher IR uncertainty



Take-away (1)

- Given the large uncertainty today, we cannot reject that Fed slope policy is appropriate

- How about the new instruments?
  (e.g., Fed lending to small/medium businesses)
- Uncertain effects (high IR uncertainty) makes OPP less informative

Take-away (2)

$$\delta_t^* = -(\mathcal{R}_t'\mathcal{R}_t)^{-1}\mathcal{R}_t'\mathbb{E}_t Y_t^0$$

- OPP helps policy makers articulate/communicate their prescription around three concepts:

  1. **preference** between different objectives
  2. assessment of the **economic outlook**
  3. views on the **effects of policy**

Conclusion

- A framework to help decision making in real life setting
- OPP helps detect optimization failures when
    - Underlying model is complex and costly to compute
    - Policy makers use heuristics to decide on policy

Two perspectives on literature (1)

- Identification of structural IRs (Ramey, 2016)
    - Impulse responses can be used to test optimality

Two perspectives on literature (2)

- **Forecast Targeting**: "*select a policy-rate path so that the forecasts of the target variables look good, meaning appears to best fulfill the mandates and return to their target at an appropriate pace*" Svensson (1999, 2017, 2019)
- Captures operational procedure of most π-targeting central banks
- Limitation: no quantitative criterion for "appropriate"
- $\Rightarrow$ OPP provides such a quantitative criterion

Extension

- OPP can be applied to broad range of policy problems
  - Monetary policy
  - Fiscal policy with stabilization vs budget deficit concerns
  - Exchange rate management
  - Foreign exchange reserve management
  - ...

A forecast-targeting rule

The Fed has a rule. The Fed's rule is that we will go for a 2percent inflation rate; we will go for the natural rate of unemployment; we put equal weight on those two things; we will give you information about our projections, our interest rate. That is a rule and that is a framework that should clarify exactly what the Fed is doing.

Bernanke (2015)

If the FOMC is going to have a forecast-based framework, it is not enough to say "eventually we will get back to 2 percent". The FOMC needs to talk about a time horizon over which it is planning to hit 2%.

Kocherlakota (2016)

Formalizing the OPP properties

Proposition

1. *Discarding optimality: Given smoothness of* f_t*, we have that* $\delta_t^* \neq 0$ *implies* $p_t^0 \neq p_t^{\text{opt}}$*, where* $p_t^{\text{opt}} = \arg\min_{p_t \in \mathcal{D}} \mathcal{L}_t$*.*

   ▸ details

2. *Improving policy: Given* $\mathcal{R}_t(p_t)$ *not too non-linear we have there exists* $\epsilon > 0$ *such that for all* $p_t^0 \in \mathcal{N}(p_t^{\text{opt}}, \epsilon)$*, and*

$$\|p_t^0 + \delta_t^* - p_t^{\text{opt}}\| \leq \|p_t^0 - p_t^{\text{opt}}\|$$

   ▸ details

3. *Optimal policy: Given* $\mathcal{R}_t$ *independent of* p_t *we have*

$$p_t^0 + \delta_t^* = p_t^{\text{opt}}$$

Assumption: Discarding optimality

Assumption

Let $X_t \in \mathcal{X}$ *and* $\xi_t \in \Xi$ *be random vectors and* $\mathcal{D}$ *an open convex subset of* $\mathbb{R}^K$. *We assume that*

1. $f_t : \mathcal{D} \times \mathcal{X} \to \mathbb{R}^{M(H+1)}$ *is continuous and there exists a random variable* Z_t *such that* $\|f_t\| \leq Z_t$ *uniformly with* $\mathbb{E}(Z_t) < \infty$.
2. *there exists a function* $R_t \equiv \partial f_t / \partial p_t$ *such that uniformly we have* R_t *has full column rank.*

Back

Assumption: Improving policy perturbation

Assumption

We assume that

1. $\|(R_t(p_t, X_t) - R_t(p_t^{\mathrm{opt}}, X_t))'\mathbb{E}_t\left(f_t(p_t^{\mathrm{opt}}, X_t) + \xi_t\right)\| \leq c\|p_t - p_t^{\mathrm{opt}}\|$, *with constant* $c < \mu_{\min}(R_t'R_t)$ *for all* $(p_t, X_t) \in \mathcal{D} \times \mathcal{X}$
2. R_t *is Lipschitz continuous with respect to* p_t *on* $\mathcal{D}$ *with parameter* γ.

▸ Back

Assumption: Optimal policy perturbation

Assumption

R_t is independent of p_t.

The Lucas critique as an omitted variable bias

- In our context, the Lucas critique can be understood as

$$\nabla \mathcal{R}_t = \left. \frac{\partial \mathcal{R}_t(p_t)}{\partial p_t'} \right|_{p_t} \neq 0$$

The Lucas critique as an omitted variable bias

- In our context, the Lucas critique can be understood as

$$\nabla \mathcal{R}_t = \left. \frac{\partial \mathcal{R}_t(p_t)}{\partial p_t'} \right|_{p_t} \neq 0$$

- If the data is generated according to

$$Y_t = Y^0 + \mathcal{R}_t \delta_t + \frac{1}{2} \nabla \mathcal{R}_t \delta_t^2$$

- Then the OPP should be

$$\delta_t^* = \arg\min \| Y^0 + \mathcal{R}_t \delta_t + \frac{1}{2} \nabla \mathcal{R} \delta_t^2 \|^2$$

but we calculate

$$\tilde{\delta}_t = \arg\min \| Y^0 + \mathcal{R}_t \delta_t \|^2$$

- $\tilde{\delta}_t$ is a biased estimate of δ_t^* unless $\nabla \mathcal{R}_t = 0$ or $\delta_t^* = 0$.

Backgound: optimization (1)

- Step of a Newton line search algorithm

$$\delta_t = -(\nabla^2 \mathcal{L}_t)^{-1} \nabla \mathcal{L}_t$$

- Gauss-Newton (GN) is a modification of Newton for problems of the form

$$\mathcal{L}_t = \frac{1}{2} \min_{p_t} Y_t^{'} Y_t$$

- GN step is

$$\delta_t = - \left(\mathcal{R}_t' \mathcal{R}_t \right)^{-1} \nabla \mathcal{L}_t$$

Backgound: optimization (2)

- GN approximates Hessian with first-derivatives

$$\nabla^2 \mathcal{L}_t = \mathcal{R}_t' \mathcal{R}_t + \underbrace{\nabla \mathcal{R}_t' Y_t^0}_{\simeq 0}$$

- Equivalent to minimizing the linear-quadratic model

$$\min_{\delta_t} \left(\mathcal{R}_t \delta_t + Y_t^0\right)' \left(\mathcal{R}_t \delta_t + Y_t^0\right)$$

- **GN step maximizes a quadratic approximation of the loss function** using

$$Y_t \approx \mathcal{R}_t \delta_t + Y_t^0$$