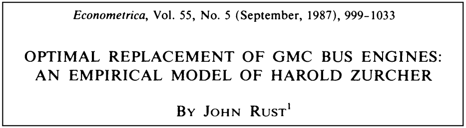
Sensitivity and counterfactual analysis of structural models

Tamara Broderick, Ryan Giordano,
Jan-Christian Huetter, Yaroslav Mukhin

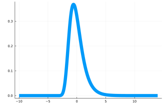
DSE2021 @ UBonn

19 Aug 2021


Econometrica, Vol. 55, No. 5 (September, 1987), 999-1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]

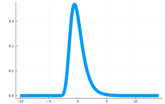


Analytic solution requires the Gumbel assumption.

Econometrica, Vol. 55, No. 5 (September, 1987), 999-1033

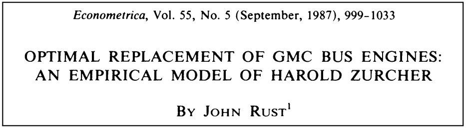
OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



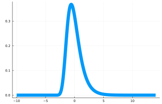
Analytic solution requires the Gumbel assumption.

Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.


Econometrica, Vol. 55, No. 5 (September, 1987), 999-1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



Analytic solution requires the Gumbel assumption.

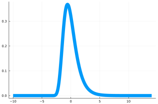
Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.

Real world definitely does not follow Gumbel. Is this a problem? We don't know!

Econometrica, Vol. 55, No. 5 (September, 1987), 999–1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



Analytic solution requires the Gumbel assumption.

Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.

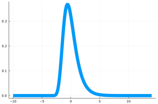
Real world definitely does not follow Gumbel. Is this a problem? We don't know!

So we should check how sensitive counterfactuals are to the Gumbel assumption.

Econometrica, Vol. 55, No. 5 (September, 1987), 999–1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



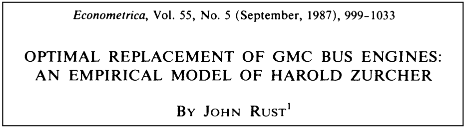
Analytic solution requires the Gumbel assumption.

Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.

Real world definitely does not follow Gumbel. Is this a problem? We don't know!

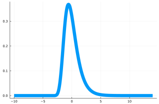
So we should check how sensitive counterfactuals are to the Gumbel assumption.

Try distributions other than Gumbel? Problem: computationally expensive, can write down only a few, no hope of guessing the truth.


Econometrica, Vol. 55, No. 5 (September, 1987), 999-1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



Analytic solution requires the Gumbel assumption.

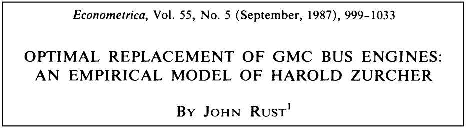
Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.

Real world definitely does not follow Gumbel. Is this a problem? We don't know!

So we should check how sensitive counterfactuals are to the Gumbel assumption.

Try distributions other than Gumbel? Problem: computationally expensive, can write down only a few, no hope of guessing the truth.

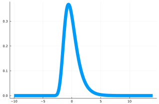
Try estimating the identified set of model parameters and counterfactuals? Problem: hard, expensive, probably a lot more work than necessary.



Econometrica, Vol. 55, No. 5 (September, 1987), 999-1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



Analytic solution requires the Gumbel assumption.

Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.

Real world definitely does not follow Gumbel. Is this a problem? We don't know!

So we should check how sensitive counterfactuals are to the Gumbel assumption.

Try distributions other than Gumbel? Problem: computationally expensive, can write down only a few, no hope of guessing the truth.

Try estimating the identified set of model parameters and counterfactuals? Problem: hard, expensive, probably a lot more work than necessary.

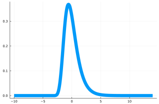
Christensen and Connault (2019) solve important special case, but limited to models with constraints that are linear in the distribution and sensitivity over neighborhoods of ϕ-divergence.

We propose a framework to characterize sensitivity in structural models with (i) general smooth dependence on the distributional assumptions and

Econometrica, Vol. 55, No. 5 (September, 1987), 999–1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



Analytic solution requires the Gumbel assumption.

Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.

Real world definitely does not follow Gumbel. Is this a problem? We don't know!

So we should check how sensitive counterfactuals are to the Gumbel assumption.

Try distributions other than Gumbel? Problem: computationally expensive, can write down only a few, no hope of guessing the truth.

Try estimating the identified set of model parameters and counterfactuals? Problem: hard, expensive, probably a lot more work than necessary.

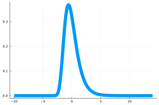
Christensen and Connault (2019) solve important special case, but limited to models with constraints that are linear in the distribution and sensitivity over neighborhoods of ϕ-divergence.

We propose a framework to characterize sensitivity in structural models with
(i) general smooth dependence on the distributional assumptions and
(ii) perturbations over neighborhoods of a general metric, that promises

Econometrica, Vol. 55, No. 5 (September, 1987), 999-1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



Analytic solution requires the Gumbel assumption.

Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.

Real world definitely does not follow Gumbel. Is this a problem? We don't know!

So we should check how sensitive counterfactuals are to the Gumbel assumption.

Try distributions other than Gumbel? Problem: computationally expensive, can write down only a few, no hope of guessing the truth.

Try estimating the identified set of model parameters and counterfactuals? Problem: hard, expensive, probably a lot more work than necessary.

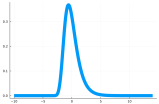
Christensen and Connault (2019) solve important special case, but limited to models with constraints that are linear in the distribution and sensitivity over neighborhoods of ϕ-divergence.

We propose a framework to characterize sensitivity in structural models with
(i) general smooth dependence on the distributional assumptions and
(ii) perturbations over neighborhoods of a general metric, that promises
(iii) to be computationally appealing and

Econometrica, Vol. 55, No. 5 (September, 1987), 999-1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]



Analytic solution requires the Gumbel assumption.

Solution depends on the Gumbel assumption, therefore counterfactuals and policy evaluations depend on the Gumbel assumption.

Real world definitely does not follow Gumbel. Is this a problem? We don't know!

So we should check how sensitive counterfactuals are to the Gumbel assumption.

Try distributions other than Gumbel? Problem: computationally expensive, can write down only a few, no hope of guessing the truth.

Try estimating the identified set of model parameters and counterfactuals? Problem: hard, expensive, probably a lot more work than necessary.

Christensen and Connault (2019) solve important special case, but limited to models with constraints that are linear in the distribution and sensitivity over neighborhoods of ϕ-divergence.

We propose a framework to characterize sensitivity in structural models with
(i) general smooth dependence on the distributional assumptions and
(ii) perturbations over neighborhoods of a general metric, that promises
(iii) to be computationally appealing and (iv*) automatable.

Example: Forecast demand for a new product, an electric bus, with Rust87.

Example: Forecast demand for a new product, an electric bus, with Rust87.

Observation: forecast depends on the distributional assumption F.

Example: Forecast demand for a new product, an electric bus, with Rust87.

Observation: forecast depends on the distributional assumption F.

Flow costs:

$$\pi_{d,x}(\theta) = \begin{cases} -\theta_0 \cdot x & \text{if } d = 0 \\ -\theta_1 & \text{if } d = 1 \end{cases}$$

parameter $\theta = (\theta_0, \theta_1)$, mileage x, decision $d \in \{0\text{:maintain}, 1\text{:replace}\}$.

Example: Forecast demand for a new product, an electric bus, with Rust87.

Observation: forecast depends on the distributional assumption F.

Flow costs:

$$\pi_{d,x}(\theta) = \begin{cases} -\theta_0 \cdot x & \text{if } d = 0 \\ -\theta_1 & \text{if } d = 1 \end{cases}$$

parameter $\theta = (\theta_0, \theta_1)$, mileage x, decision $d \in \{0\text{:maintain, } 1\text{:replace}\}$.

Assumption: latent state $U_d \overset{\text{iid}}{\sim} F$; solve for surplus:

$$V_x(\theta, F) = \mathrm{E}^F \max_{d \in \{0,1\}} \left\{ \pi_{d,x}(\theta) + U_d + \beta \mathrm{E}\left[V_{x'} | x, d \right] \right\}$$

and choice probabilities $\mathrm{P}(d = 0 | x; \theta, F)$.

Example: Forecast demand for a new product, an electric bus, with Rust87.

Observation: forecast depends on the distributional assumption F.

Flow costs:

$$\pi_{d,x}(\theta) = \begin{cases} -\theta_0 \cdot x & \text{if } d = 0 \\ -\theta_1 & \text{if } d = 1 \end{cases}$$

parameter $\theta = (\theta_0, \theta_1)$, mileage x, decision $d \in \{0\text{:maintain}, 1\text{:replace}\}$.

Assumption: latent state $U_d \overset{\text{iid}}{\sim} F$; solve for surplus:

$$V_x(\theta, F) = \mathrm{E}^F \max_{d \in \{0,1\}} \Big\{ \pi_{d,x}(\theta) + U_d + \beta \mathrm{E}\Big[V_{x'} | x, d \Big] \Big\}$$

and choice probabilities $\mathrm{P}(d = 0 | x; \theta, F)$.

$$\hat{\theta}^{\text{MLE}}(F) = \arg\max_{\theta} \prod_t \mathrm{P}(d_t | x_t; \theta, F).$$

Example: Forecast demand for a new product, an electric bus, with Rust87.

Observation: forecast depends on the distributional assumption F.

Flow costs:

$$\pi_{d,x}(\theta) = \begin{cases} -\theta_0 \cdot x & \text{if } d = 0 \\ -\theta_1 & \text{if } d = 1 \end{cases}$$

parameter $\theta = (\theta_0, \theta_1)$, mileage x, decision $d \in \{0\text{:maintain, } 1\text{:replace}\}$.

Assumption: latent state $U_d \overset{\text{iid}}{\sim} F$; solve for surplus:

$$V_x(\theta, F) = \mathrm{E}^F \max_{d \in \{0,1\}} \left\{ \pi_{d,x}(\theta) + U_d + \beta \mathrm{E}\left[V_{x'} | x, d \right] \right\}$$

and choice probabilities $\mathrm{P}(d = 0 | x; \theta, F)$.

$$\hat{\theta}^{\mathsf{MLE}}(F) = \arg\max_{\theta} \prod_t \mathrm{P}(d_t | x_t; \theta, F).$$

Estimate of surplus: $\hat{V} = V(\hat{\theta}(F), F)$.

Example: Forecast demand for a new product, an electric bus, with Rust87.

Observation: forecast depends on the distributional assumption F.

Flow costs:

$$\pi_{d,x}(\theta) = \begin{cases} -\theta_0 \cdot x & \text{if } d = 0 \\ -\theta_1 & \text{if } d = 1 \end{cases}$$

Counterfactual flow costs:

$$\tilde{\pi}_{d,x}(\theta) = \begin{cases} -0.5 \cdot \theta_0 \cdot x & \text{if } d = 0 \\ -2 \cdot \theta_1 & \text{if } d = 1 \end{cases}$$

parameter $\theta = (\theta_0, \theta_1)$, mileage x, decision $d \in \{0\text{:maintain, } 1\text{:replace}\}$.

Assumption: latent state $U_d \overset{\text{iid}}{\sim} F$; solve for surplus:

$$V_x(\theta, F) = \mathrm{E}^F \max_{d\in\{0,1\}} \Big\{ \pi_{d,x}(\theta) + U_d + \beta \mathrm{E}\Big[V_{x'} | x, d\Big] \Big\}$$

and choice probabilities $\mathrm{P}(d = 0|x; \theta, F)$.

$$\hat{\theta}^{\text{MLE}}(F) = \arg\max_{\theta} \prod_t \mathrm{P}(d_t | x_t; \theta, F).$$

Estimate of surplus: $\hat{V} = V(\hat{\theta}(F), F)$.

Example: Forecast demand for a new product, an electric bus, with Rust87.

Observation: forecast depends on the distributional assumption F.

Flow costs:

$$\pi_{d,x}(\theta) = \begin{cases} -\theta_0 \cdot x & \text{if } d = 0 \\ -\theta_1 & \text{if } d = 1 \end{cases}$$

Counterfactual flow costs:

$$\tilde{\pi}_{d,x}(\theta) = \begin{cases} -0.5 \cdot \theta_0 \cdot x & \text{if } d = 0 \\ -2 \cdot \theta_1 & \text{if } d = 1 \end{cases}$$

parameter $\theta = (\theta_0, \theta_1)$, mileage x, decision $d \in \{0\text{:maintain}, 1\text{:replace}\}$.

Assumption: latent state $U_d \overset{\text{iid}}{\sim} F$; solve for surplus:

$$V_x(\theta, F) = \mathrm{E}^F \max_{d \in \{0,1\}} \left\{ \pi_{d,x}(\theta) + U_d + \beta \mathrm{E}\Big[V_{x'} | x, d \Big] \right\}$$

and choice probabilities $\mathrm{P}(d = 0 | x; \theta, F)$.

$$\hat{\theta}^{\text{MLE}}(F) = \arg\max_{\theta} \prod_t \mathrm{P}(d_t | x_t; \theta, F).$$

Estimate of surplus: $\hat{V} = V(\hat{\theta}(F), F)$.

Counterfactual change in surplus:

$$k = w^\top \left(\tilde{V}\left(\hat{\theta}(F), F\right) - V\left(\hat{\theta}(F), F\right) \right)$$

depends on distributional assumption F.

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0 \leq \delta \leq a}$:

$$F_\delta = \underset{F \in \mathsf{Ball}_\rho(F_0, \delta)}{\arg\max}\ k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0 \le \delta \le a}$:

$$F_\delta = \underset{F \in \mathsf{Ball}_\rho(F_0, \delta)}{\arg\max}\; k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0 \le \delta \le a}$:

$$F_\delta = \underset{F \in \mathsf{Ball}_\rho(F_0, \delta)}{\arg\max}\ k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;

(ii) update F_δ;

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0\leq\delta\leq a}$:

$$F_\delta = \underset{F\in \mathsf{Ball}_\rho(F_0,\delta)}{\arg\max}\ k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;
(ii) update F_δ;
(iii) update model solution
e.g. $\theta(F_\delta)$, $V(F_\delta)$ and $\mathrm{P}_{d|x}(F_\delta)$
without resolving the model;

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0 \leq \delta \leq a}$:

$$F_\delta = \underset{F \in \mathsf{Ball}_\rho(F_0, \delta)}{\arg\max} \; k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;
(ii) update F_δ;
(iii) update model solution
e.g. $\theta(F_\delta)$, $V(F_\delta)$ and $\mathrm{P}_{d|x}(F_\delta)$
without resolving the model;
(iv) repeat.

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0\leq\delta\leq a}$:

$$F_\delta = \underset{F \in \mathsf{Ball}_\rho(F_0,\delta)}{\arg\max}\ k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;
(ii) update F_δ;
(iii) update model solution
e.g. $\theta(F_\delta)$, $V(F_\delta)$ and $\mathrm{P}_{d|x}(F_\delta)$
without resolving the model;
(iv) repeat.

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0 \le \delta \le a}$:

$$F_\delta = \underset{F \in \mathsf{Ball}_\rho(F_0, \delta)}{\arg\max} \; k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;
(ii) update F_δ;
(iii) update model solution e.g. $\theta(F_\delta)$, $V(F_\delta)$ and $\mathrm{P}_{d|x}(F_\delta)$ without resolving the model;
(iv) repeat.

How to pose and solve the local problem?

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0\leq\delta\leq a}$:

$$F_\delta = \arg\max_{F \in \mathsf{Ball}_\rho(F_0,\delta)} k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;
(ii) update F_δ;
(iii) update model solution e.g. $\theta(F_\delta)$, $V(F_\delta)$ and $\mathrm{P}_{d|x}(F_\delta)$ without resolving the model;
(iv) repeat.

How to pose and solve the local problem?
How to integrate the local solutions?

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0\leq\delta\leq a}$:

$$F_\delta = \underset{F \in \mathsf{Ball}_\rho(F_0,\delta)}{\arg\max}\ k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;
(ii) update F_δ;
(iii) update model solution e.g. $\theta(F_\delta)$, $V(F_\delta)$ and $\mathrm{P}_{d|x}(F_\delta)$ without resolving the model;
(iv) repeat.

How to pose and solve the local problem?
How to integrate the local solutions?
How to update F_δ?

Our framework

F_0 assumed distribution of unobserved variables e.g. Gumbel

$\theta(F)$ vector of structural parameters e.g. maintain/replace costs

$k(\theta(F), F)$ scalar counterfactual e.g. change in surplus

To characterize the sensitivity of k to F_0 we propose a greedy algorithm to construct the least favorable path $(F_\delta)_{0 \leq \delta \leq a}$:

$$F_\delta = \arg\max_{F \in \mathsf{Ball}_\rho(F_0, \delta)} k\big(\theta(F), F\big), \qquad \text{where} \tag{1}$$

$\delta > 0$ small number e.g. 0.1,

$\rho(F, F_0)$ metric on the space of probability distributions e.g. Wasserstein.

Greedy algorithm solves and iterates the local version of problem (1):

(i) solve (1) locally;
(ii) update F_δ;
(iii) update model solution e.g. $\theta(F_\delta)$, $V(F_\delta)$ and $\mathrm{P}_{d|x}(F_\delta)$ without resolving the model;
(iv) repeat.

How to pose and solve the local problem?
How to integrate the local solutions?
How to update F_δ?
How to update $\theta(F), V(F)$, etc?

Local problem I

Question: what is the local version of $\max\limits_{\rho(F,F_0)\leq\delta} k\big(\theta(F), F\big)$?

Local problem I

Question: what is the local version of $\max_{\rho(F,F_0)\leq\delta} k\big(\theta(F), F\big)$?

1. replace $k(F)$ with its derivative Dk with respect to distribution F

Local problem I

Question: what is the local version of $\max_{\rho(F,F_0)\leq\delta} k\big(\theta(F), F\big)$?

1. replace $k(F)$ with its derivative Dk with respect to distribution F
2. replace least favorable F_δ with direction of most rapid change in k

Local problem I

Question: what is the local version of $\max_{\rho(F,F_0)\leq\delta} k\big(\theta(F), F\big)$?

1. replace $k(F)$ with its derivative Dk with respect to distribution F
2. replace least favorable F_δ with direction of most rapid change in k

Challenge: space of probability distributions F is not linear.

Local problem I

Question: what is the local version of $\max_{\rho(F,F_0)\leq\delta} k\big(\theta(F),F\big)$?

1. replace $k(F)$ with its derivative Dk with respect to distribution F
2. replace least favorable F_δ with direction of most rapid change in k

Challenge: space of probability distributions F is not linear.

Solution: scores and influence functions!

Local problem I

Question: what is the local version of $\max_{\rho(F,F_0)\leq\delta} k(\theta(F), F)$?

1. replace $k(F)$ with its derivative Dk with respect to distribution F
2. replace least favorable F_δ with direction of most rapid change in k

Challenge: space of probability distributions F is not linear.

Solution: scores and influence functions!

- Pathwise derivative: consider one-dimensional path F_h with *score*

$$g(u) = \partial_h \log f_h(u), \quad \text{where } f \text{ is density of } F.$$

Local problem I

Question: what is the local version of $\max_{\rho(F,F_0)\leq\delta} k(\theta(F), F)$?

1. replace $k(F)$ with its derivative Dk with respect to distribution F
2. replace least favorable F_δ with direction of most rapid change in k

Challenge: space of probability distributions F is not linear.

Solution: scores and influence functions!

- Pathwise derivative: consider one-dimensional path F_h with *score*

$$g(u) = \partial_h \log f_h(u), \quad \text{where } f \text{ is density of } F.$$

- Differential $Dk_F[\cdot]$ is an operator on the linear space of scores $L_0^2(F)$.

Local problem I

Question: what is the local version of $\max_{\rho(F,F_0)\leq\delta} k(\theta(F),F)$?

1. replace $k(F)$ with its derivative Dk with respect to distribution F
2. replace least favorable F_δ with direction of most rapid change in k

Challenge: space of probability distributions F is not linear.

Solution: scores and influence functions!

- Pathwise derivative: consider one-dimensional path F_h with *score*

$$g(u) = \partial_h \log f_h(u), \quad \text{where } f \text{ is density of } F.$$

- Differential $Dk_F[\cdot]$ is an operator on the linear space of scores $L_0^2(F)$.
- Influence function is the Riesz representation:

$$Dk_F[g] = \left\langle g, \mathcal{I}_k(F) \right\rangle_\rho.$$

Local problem I

Question: what is the local version of $\max_{\rho(F,F_0)\leq\delta} k(\theta(F),F)$?

1. replace $k(F)$ with its derivative Dk with respect to distribution F
2. replace least favorable F_δ with direction of most rapid change in k

Challenge: space of probability distributions F is not linear.

Solution: scores and influence functions!

- Pathwise derivative: consider one-dimensional path F_h with *score*
$$g(u) = \partial_h \log f_h(u), \quad \text{where } f \text{ is density of } F.$$
- Differential $Dk_F[\cdot]$ is an operator on the linear space of scores $L_0^2(F)$.
- Influence function is the Riesz representation:
$$Dk_F[g] = \langle g, \mathcal{I}_k(F)\rangle_\rho .$$
- Influence function solves the local problem:
$$\text{const} \cdot \mathcal{I}_k(F) = \arg\max_{\|g\|_\rho\leq\delta} Dk_F[g]. \tag{1*}$$

Local problem II

Question: how to integrate local solutions into the least favorable path?

Local problem II

Question: how to integrate local solutions into the least favorable path?

Answer:

Solve the gradient flow equation for the least favorable path $(F_h)_{0\le h\le\delta}$:

$$\underbrace{\partial_h \log f(h)}_{\text{score}} = \underbrace{\text{const} \cdot \mathcal{I}_k(F)}_{\text{most rapid direction}}, \qquad F(0) \text{ is given by } F_0. \tag{2}$$

Local problem II

Question: how to integrate local solutions into the least favorable path?

Answer:

Solve the gradient flow equation for the least favorable path $(F_h)_{0 \le h \le \delta}$:

$$\underbrace{\partial_h \log f(h)}_{\text{score}} = \underbrace{\text{const} \cdot \mathcal{I}_k(F)}_{\text{most rapid direction}}, \qquad F(0) \text{ is given by } F_0. \tag{2}$$

Challenge: space of probability distributions F is not linear.

Local problem II

Question: how to integrate local solutions into the least favorable path?

Answer:

Solve the gradient flow equation for the least favorable path $(F_h)_{0 \le h \le \delta}$:

$$\underbrace{\partial_h \log f(h)}_{\text{score}} = \underbrace{\text{const} \cdot \mathcal{I}_k(F)}_{\text{most rapid direction}}, \qquad F(0) \text{ is given by } F_0. \tag{2}$$

Challenge: space of probability distributions F is not linear.

Solution: manifold parametrization!

- To pose and solve the ordinary differential equation (2) in variable F that is a probability distribution, we need a manifold parametrization of the space of probability distributions by a linear space.

Local problem II

Question: how to integrate local solutions into the least favorable path?

Answer:

Solve the gradient flow equation for the least favorable path $(F_h)_{0 \le h \le \delta}$:

$$\underbrace{\partial_h \log f(h)}_{\text{score}} = \underbrace{\text{const} \cdot \mathcal{I}_k(F)}_{\text{most rapid direction}}, \qquad F(0) \text{ is given by } F_0. \tag{2}$$

Challenge: space of probability distributions F is not linear.

Solution: manifold parametrization!

- To pose and solve the ordinary differential equation (2) in variable F that is a probability distribution, we need a manifold parametrization of the space of probability distributions by a linear space.
- We handle this with exponential Orlicz spaces.
  Reference: Pistone and Sempi (1995), Annals of Statistics.

Local problem II

Question: how to integrate local solutions into the least favorable path?

Answer:

Solve the gradient flow equation for the least favorable path $(F_h)_{0 \leq h \leq \delta}$:

$$\underbrace{\partial_h \log f(h)}_{\text{score}} = \underbrace{\text{const} \cdot \mathcal{I}_k(F)}_{\text{most rapid direction}}, \qquad F(0) \text{ is given by } F_0. \tag{2}$$

Challenge: space of probability distributions F is not linear.

Solution: manifold parametrization!

- To pose and solve the ordinary differential equation (2) in variable F that is a probability distribution, we need a manifold parametrization of the space of probability distributions by a linear space.
- We handle this with exponential Orlicz spaces.
  Reference: Pistone and Sempi (1995), Annals of Statistics.
- Can impose normalizations on (F_δ), e.g. mean is 0, variance is 1, via normalizations on the right-hand side of equation (2).

Greedy algorithm I: exponential tilting

Question: how to do one step of the greedy algorithm?

Greedy algorithm I: exponential tilting

Question: how to do one step of the greedy algorithm?

One step of greedy is a linear approximation to

$$\partial_h \log f(h) = \mathcal{I}_k(F). \tag{2}$$

Greedy algorithm I: exponential tilting

Question: how to do one step of the greedy algorithm?

One step of greedy is a linear approximation to

$$\partial_h \log f(h) = \mathcal{I}_k(F). \tag{2}$$

Given f_h, can try updating linearly: $f_{h+\epsilon}(u) = \big(1 + \epsilon \cdot \mathcal{I}_k(u)\big) \cdot f_h(u)$.

Greedy algorithm I: exponential tilting

Question: how to do one step of the greedy algorithm?

One step of greedy is a linear approximation to

$$\partial_h \log f(h) = \mathcal{I}_k(F). \qquad (2)$$

Given f_h, can try updating linearly: $f_{h+\epsilon}(u) = \big(1 + \epsilon \cdot \mathcal{I}_k(u)\big) \cdot f_h(u)$.

Problem: if the influence function is unbounded, then this does not work for any $\epsilon > 0$ because densities must be non-negative!

Greedy algorithm I: exponential tilting

Question: how to do one step of the greedy algorithm?

One step of greedy is a linear approximation to

$$\partial_h \log f(h) = \mathcal{I}_k(F). \tag{2}$$

Given f_h, can try updating linearly: $f_{h+\epsilon}(u) = \big(1 + \epsilon \cdot \mathcal{I}_k(u)\big) \cdot f_h(u)$.

Problem: if the influence function is unbounded, then this does not work for any $\epsilon > 0$ because densities must be non-negative!

Solution: exponential tilting

$$f_{h+\epsilon}(u) = \underbrace{\exp\Big(\epsilon \cdot \mathcal{I}_k(u)\Big)}_{\approx 1+\epsilon\cdot\mathcal{I}_k} \cdot f_h \;\Big/\; \mathrm{E}^{f_h}\Big[\exp\Big(\epsilon \cdot \mathcal{I}_k(U)\Big)\Big] \tag{3}$$

Greedy algorithm I: exponential tilting

Question: how to do one step of the greedy algorithm?

One step of greedy is a linear approximation to

$$\partial_h \log f(h) = \mathcal{I}_k(F). \tag{2}$$

Given f_h, can try updating linearly: $f_{h+\epsilon}(u) = \big(1 + \epsilon \cdot \mathcal{I}_k(u)\big) \cdot f_h(u)$.

Problem: if the influence function is unbounded, then this does not work for any $\epsilon > 0$ because densities must be non-negative!

Solution: exponential tilting

$$f_{h+\epsilon}(u) = \underbrace{\exp\Big(\epsilon \cdot \mathcal{I}_k(u)\Big)}_{\approx 1+\epsilon\cdot\mathcal{I}_k} \cdot f_h \Big/ \mathrm{E}^{f_h}\Big[\exp\Big(\epsilon \cdot \mathcal{I}_k(U)\Big)\Big] \tag{3}$$

is the linear update in the manifold parameterization of the space.

Greedy algorithm II: linear approximations

Question: how to use greedy without resolving model with updated F_h?

Greedy algorithm II: linear approximations

Question: how to use greedy without resolving model with updated F_h?

Assume θ is scalar (otherwise consider each coordinate θ_j).

Greedy algorithm II: linear approximations

Question: how to use greedy without resolving model with updated F_h?

Assume θ is scalar (otherwise consider each coordinate θ_j).

Solution: use influence functions $\mathcal{I}_k$ and $\mathcal{I}_\theta$ to form linear Taylor approximations to changes along the least favorable path:

$$\begin{aligned} \theta(F_{h+\epsilon}) &\approx \theta(F_h) + \epsilon \cdot D\theta_{F_h}\left[\mathcal{I}_k(F_h)\right] \\ &\approx \theta(F_h) + \epsilon \cdot \left\langle \mathcal{I}_k(F_h), \mathcal{I}_\theta(F_h) \right\rangle_{F_h}. \end{aligned}$$

Greedy algorithm II: linear approximations

Question: how to use greedy without resolving model with updated F_h?

Assume θ is scalar (otherwise consider each coordinate θ_j).

Solution: use influence functions $\mathcal{I}_k$ and $\mathcal{I}_\theta$ to form linear Taylor approximations to changes along the least favorable path:

$$\begin{aligned}\theta(F_{h+\epsilon}) &\approx \theta(F_h) + \epsilon \cdot D\theta_{F_h}\left[\mathcal{I}_k(F_h)\right] \\ &\approx \theta(F_h) + \epsilon \cdot \Big\langle \mathcal{I}_k(F_h), \mathcal{I}_\theta(F_h) \Big\rangle_{F_h}.\end{aligned}$$

Compute exact change on $(F_t)_{0\le t\le h}$ via fundamental theorem of calculus:

$$\begin{aligned}\theta(F_h) - \theta(F_0) &= \int_0^h \Big\langle \mathcal{I}_k(F_t), \mathcal{I}_\theta(F_t) \Big\rangle_{F_t} \, \mathrm{d}t \\ &\approx \sum_j \Big\langle \mathcal{I}_\theta(F_{j\cdot\epsilon}), \mathcal{I}_\theta(F_{j\cdot\epsilon}) \Big\rangle_{F_{j\cdot\epsilon}}.\end{aligned}$$

Greedy algorithm III: influence function calculation

Question: how to compute the influence function?

Recall that score of a path F_h is the derivative of log-density of the path:

$$g(u) = \frac{d}{dh}_{|h=0} \log f_h(u).$$

Greedy algorithm III: influence function calculation

Question: how to compute the influence function?

Recall that score of a path F_h is the derivative of log-density of the path:

$$g(u) = \frac{d}{dh}_{|h=0} \log f_h(u).$$

Example: suppose $k(F) = \int \psi(u)\,\mathrm{d}F(u)$, then

$$\begin{aligned} \frac{d}{dh}_{|h=0} k(F_h) &= \frac{d}{dh}_{|h=0} \int \psi(u)\,\mathrm{d}F_h(u) \\ &= \int \frac{d}{dh}_{|h=0} \psi(u) f_h(u)\,\mathrm{d}u \\ &= \int \psi(u) \frac{\frac{d}{dh}_{|h=0} f_h(u)}{f_0(u)}\,\mathrm{d}F_0(u) \\ &= \int \psi(u) g(u)\,\mathrm{d}F_0(u). \end{aligned}$$

Greedy algorithm III: influence function calculation

Question: how to compute the influence function?

Recall that score of a path F_h is the derivative of log-density of the path:

$$g(u) = \frac{d}{dh}_{|h=0} \log f_h(u).$$

Example: suppose $k(F) = \int \psi(u)\, \mathrm{d}F(u)$, then

$$\begin{aligned} \frac{d}{dh}_{|h=0} k(F_h) &= \frac{d}{dh}_{|h=0} \int \psi(u)\, \mathrm{d}F_h(u) \\ &= \int \frac{d}{dh}_{|h=0} \psi(u) f_h(u)\, \mathrm{d}u \\ &= \int \psi(u) \frac{\frac{d}{dh}_{|h=0} f_h(u)}{f_0(u)}\, \mathrm{d}F_0(u) \\ &= \int \psi(u) g(u)\, \mathrm{d}F_0(u). \end{aligned}$$

True for every path F_h, by Riesz representation theorem $Dk = \langle \cdot, \psi \rangle_{F_0}$.

Greedy algorithm III: influence function calculation

Question: how to compute the influence function?

Recall that score of a path F_h is the derivative of log-density of the path:

$$g(u) = \frac{d}{dh}_{|h=0} \log f_h(u).$$

Example: suppose $k(F) = \int \psi(u)\,\mathrm{d}F(u)$, then

$$\begin{aligned}
\frac{d}{dh}_{|h=0} k(F_h) &= \frac{d}{dh}_{|h=0} \int \psi(u)\,\mathrm{d}F_h(u) \\
&= \int \frac{d}{dh}_{|h=0} \psi(u) f_h(u)\,\mathrm{d}u \\
&= \int \psi(u) \frac{\frac{d}{dh}_{|h=0} f_h(u)}{f_0(u)}\,\mathrm{d}F_0(u) \\
&= \int \psi(u) g(u)\,\mathrm{d}F_0(u).
\end{aligned}$$

True for every path F_h, by Riesz representation theorem $Dk = \langle \cdot, \psi \rangle_{F_0}$.

Conclude that $\mathcal{I}_k(F) = \psi$, the Riesz representer in Fisher-Rao metric.

Greedy algorithm III: influence function calculation

Question: how to compute the influence function?

Recall that score of a path F_h is the derivative of log-density of the path:

$$g(u) = \frac{d}{dh}_{|h=0} \log f_h(u).$$

Example: suppose $k(F) = \int \psi(u)\, \mathrm{d}F(u)$, then

$$\begin{aligned}
\frac{d}{dh}_{|h=0} k(F_h) &= \frac{d}{dh}_{|h=0} \int \psi(u)\, \mathrm{d}F_h(u) \\
&= \int \frac{d}{dh}_{|h=0} \psi(u) f_h(u)\, \mathrm{d}u \\
&= \int \psi(u) \frac{\frac{d}{dh}_{|h=0} f_h(u)}{f_0(u)}\, \mathrm{d}F_0(u) \\
&= \int \psi(u) g(u)\, \mathrm{d}F_0(u).
\end{aligned}$$

True for every path F_h, by Riesz representation theorem $Dk = \langle \cdot, \psi \rangle_{F_0}$.

Conclude that $\mathcal{I}_k(F) = \psi$, the Riesz representer in Fisher-Rao metric.

Calc with general k is similar.

Greedy algorithm III: influence function calculation

Question: how to compute the influence function?

Recall that score of a path F_h is the derivative of log-density of the path:

$$g(u) = \frac{d}{dh}_{|h=0} \log f_h(u).$$

Example: suppose $k(F) = \int \psi(u)\, \mathrm{d}F(u)$, then

$$\begin{aligned}
\frac{d}{dh}_{|h=0} k(F_h) &= \frac{d}{dh}_{|h=0} \int \psi(u)\, \mathrm{d}F_h(u) \\
&= \int \frac{d}{dh}_{|h=0} \psi(u) f_h(u)\, \mathrm{d}u \\
&= \int \psi(u) \frac{\frac{d}{dh}_{|h=0} f_h(u)}{f_0(u)}\, \mathrm{d}F_0(u) \\
&= \int \psi(u) g(u)\, \mathrm{d}F_0(u).
\end{aligned}$$

True for every path F_h, by Riesz representation theorem $Dk = \langle \cdot, \psi \rangle_{F_0}$.

Conclude that $\mathcal{I}_k(F) = \psi$, the Riesz representer in Fisher-Rao metric.

Calc with general k is similar. Riesz rep in another norms is a tranform.

Greedy algorithm III: influence function calculation

Question: how to compute the influence function?

Recall that score of a path F_h is the derivative of log-density of the path:

$$g(u) = \frac{d}{dh}_{|h=0} \log f_h(u).$$

Example: suppose $k(F) = \int \psi(u)\,\mathrm{d}F(u)$, then

$$\begin{aligned}
\frac{d}{dh}_{|h=0} k(F_h) &= \frac{d}{dh}_{|h=0} \int \psi(u)\,\mathrm{d}F_h(u) \\
&= \int \frac{d}{dh}_{|h=0} \psi(u) f_h(u)\,\mathrm{d}u \\
&= \int \psi(u) \frac{\frac{d}{dh}_{|h=0} f_h(u)}{f_0(u)}\,\mathrm{d}F_0(u) \\
&= \int \psi(u) g(u)\,\mathrm{d}F_0(u). \qquad {}^{\dagger}\text{(automate?)}
\end{aligned}$$

True for every path F_h, by Riesz representation theorem $Dk = \langle \cdot, \psi \rangle_{F_0}$.

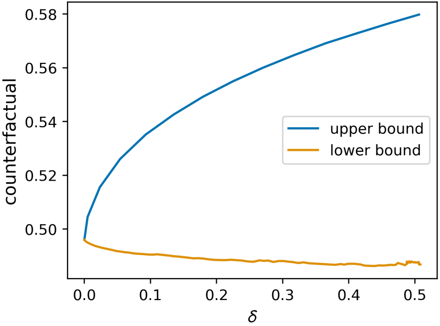
Conclude that $\mathcal{I}_k(F) = \psi$, the Riesz representer in Fisher-Rao metric.

Calc with general k is similar. Riesz rep in another norms is a tranform.

Rust87 example: numerical experiment I

Sensitivity analysis to the Gumbel assumption for counterfactual change in surplus (cost) of a switch from GMC diesel bus to electric bus.

Bounds on the counterfactual change in surplus over a δ ball of probability distributions around the Gumbel distribution of cost shocks in Rust model.



Note: preliminary numerical results using partial influence functions that hold θ fixed and allow $F \mapsto V(\theta, F)$ to vary.

Rust87 example: numerical experiment II

Least favorable distributions:

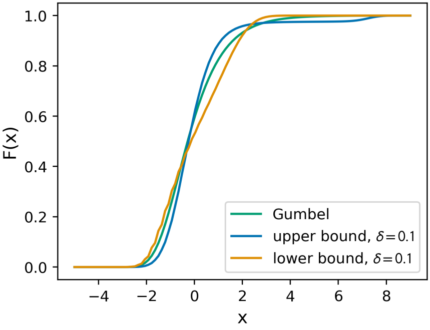


Figure: $\delta = 0.1$

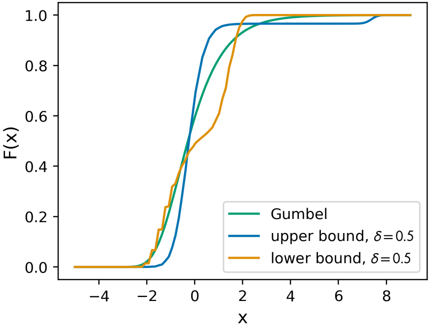


Figure: $\delta = 0.5$

I proposed a framework for sensitivity analysis of counterfactuals in structural models with respect to distributional assumptions of latent states.

I proposed a framework for sensitivity analysis of counterfactuals in structural models with respect to distributional assumptions of latent states.

(i) structural models with general smooth dependence on F

(ii) perturbations over neighborhoods of a general metric, e.g. Wasserstein

(iii) not solving model under any F other than the one you have assumed

(iv) automatable†

(v) Limitation: no guarantees to find the global max (as with any greedy). Probably a bigger problem for estimation than sensitivity analysis.

Questions, comments, feedback are very welcome and highly appreciated!

This presentation is based on:

Y. Mukhin. "Sensitivity of regular estimators". In: *arXiv:1805.08883* (2018)
Y. Mukhin. "Counterfactual analysis of differentiable functionals". In: *MIT Thesis* (2019)

Joint work in progress with Tamara Broderick, Ryan Giordano and Jan-Christian Huetter.