Efficient computations for the expected value of information in health economic evaluations

Gianluca Baio

University College London

Department of Statistical Science

g.baio@ucl.ac.uk

(Joint work with Anna Heath and Ioanna Manolopoulou)

The Fifth Workshop on Bayesian Inference for Latent Gaussian Models with Applications

Bath, 14 September 2016

Outline of the talk

Health economic evaluation

1. What is it?
2. How does it work?
3. Uncertainty analysis

Value of information

1. Expected Value of Perfect Information (simple)
2. Expected Value of Perfect *Partial* Information (complex)
3. EVPPI as a regression problem

How to make your life miserable to (eventually) have a better life...

1. EVPPI as a regression problem — but faster
2. Spatial structure + reduction dimensionality
3. Examples
4. Conclusions

Outline of the talk

Health economic evaluation

1. What is it?
2. How does it work?
3. Uncertainty analysis

Value of information

1. Expected Value of Perfect Information (simple)
2. Expected Value of Perfect *Partial* Information (complex)
3. EVPPI as a regression problem

How to make your life miserable to (eventually) have a better life...

1. EVPPI as a regression problem — but faster
2. Spatial structure + reduction dimensionality
3. Examples
4. Conclusions

Outline of the talk

Health economic evaluation

1. What is it?
2. How does it work?
3. Uncertainty analysis

Value of information

1. Expected Value of Perfect Information (simple)
2. Expected Value of Perfect *Partial* Information (complex)
3. EVPPI as a regression problem

How to make your life miserable to (eventually) have a better life...

1. EVPPI as a regression problem — but faster
2. Spatial structure + reduction dimensionality
3. Examples
4. Conclusions

Health economic evaluation — What is it?

- **Objective**: Combine costs & benefits of a given intervention into a rational scheme for allocating resources
  - Rational decision-making is effected through the comparison of expected utilities $\Rightarrow$ **monetary net benefit**

Health economic evaluation — What is it?

- **Objective**: Combine costs & benefits of a given intervention into a rational scheme for allocating resources
  - Rational decision-making is effected through the comparison of expected utilities $\Rightarrow$ **monetary net benefit**

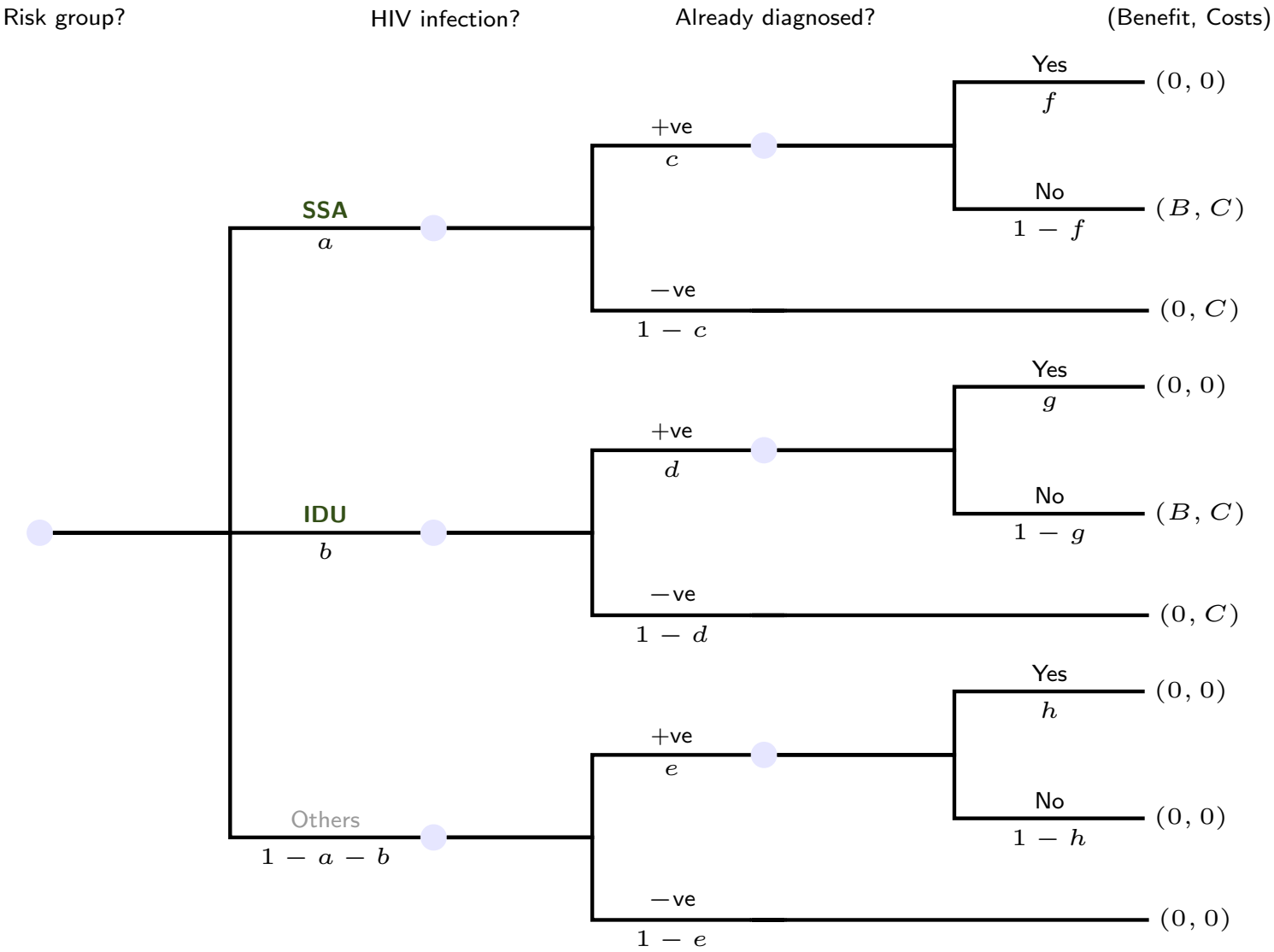
- Costs and benefits need to be modelled **jointly**
  - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research $\Rightarrow$ are associated with higher unit costs
  - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.

Health economic evaluation — What is it?

- **Objective**: Combine costs & benefits of a given intervention into a rational scheme for allocating resources
  - Rational decision-making is effected through the comparison of expected utilities $\Rightarrow$ **monetary net benefit**
- Costs and benefits need to be modelled **jointly**
  - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research $\Rightarrow$ are associated with higher unit costs
  - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
- Often needs to go "beyond RCTs"
  - Comparator(s) in the trial may not reflect standard of care
  - Limited follow up /small sample size / poor external validity
- Uses "**decision-analytic**" models instead
  - Describe full care management pathway
  - Can combine individual- and aggregate level data
  - Models include **many** relevant parameters

Decision-analytic model — HIV test (Welton et al 2012) [9]

$t = 0$: **Targeted** testing (high risk groups only)

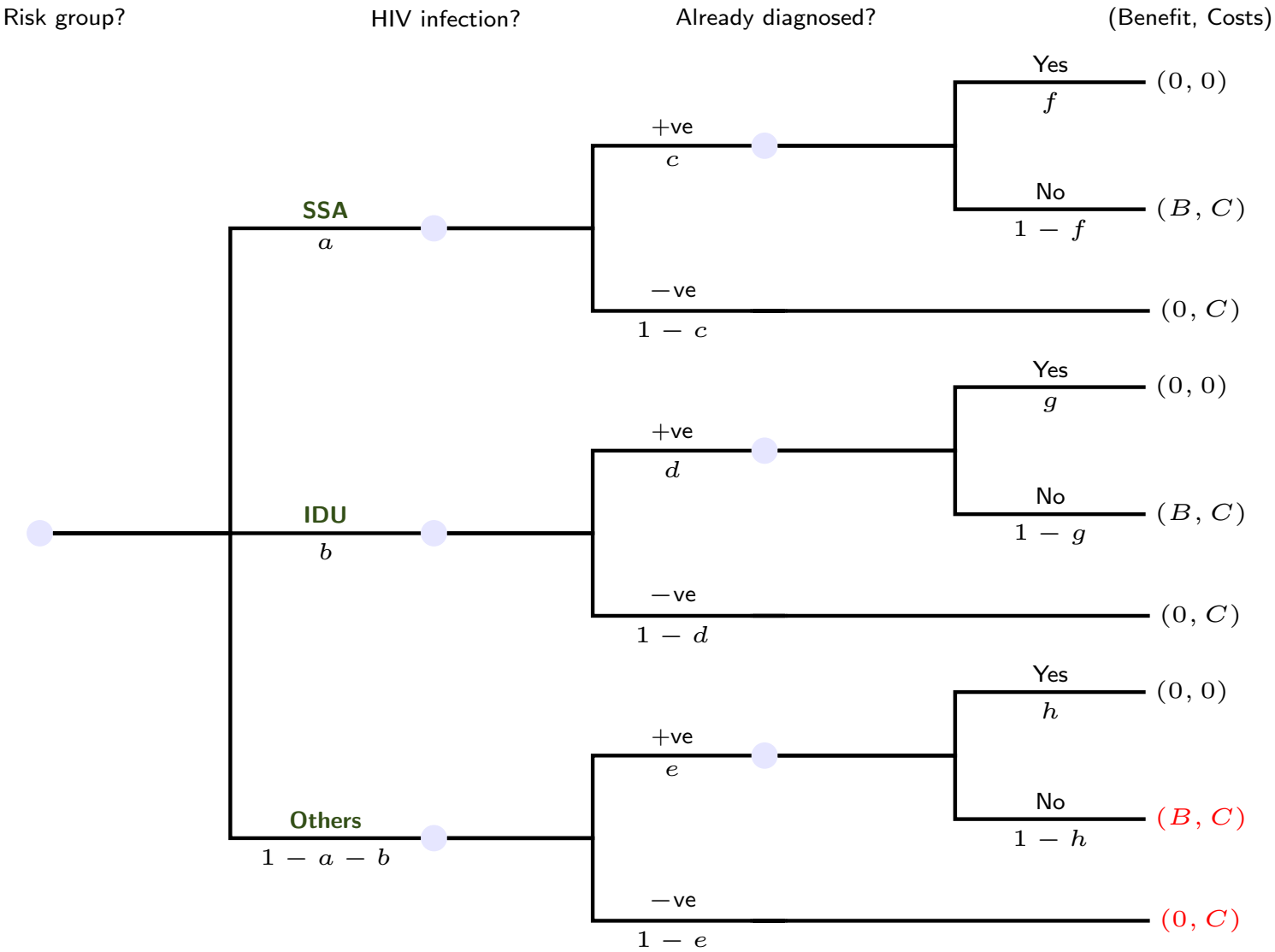


Parameters: $\boldsymbol{\theta} = (a, b, c, d, e, f, g, h)$

Utility: $\text{NB}_0(\boldsymbol{\theta}) = [ac(1-f) + bd(1-g)]\, B - [a(1-cf) + b(1-dg)]\, C$

Decision-analytic model — HIV test (Welton et al 2012) [9]

$t = 1$: **Universal** testing



Parameters: $\boldsymbol{\theta} = (a, b, c, d, e, f, g, h)$

Utility: $\text{NB}_1(\boldsymbol{\theta}) = [ac(1-f) + bd(1-g) + (1-a-b)e(1-h)]\, B - [a(1-cf) + b(1-dg) + (1-a-b)(1-eh)]\, C$

Health economic evaluation — How does it work?

	Parameters simulations			
Iter/n	a	b	...	h
1	0.365	0.076	...	0.162
2	0.421	0.024	...	0.134
3	0.125	0.017	...	0.149
4	0.117	0.073	...	0.120
5	0.481	0.008	...	0.191
6	0.163	0.127	...	0.004
...		...		
1000	0.354	0.067	...	0.117

- Characterise uncertainty in the model parameters
  - In a full Bayesian setting, these are draws from the joint posterior distribution of $\boldsymbol{\theta}$
  - In a frequentist setting, these are typically Monte Carlo draws from a set of univariate distributions that describe some level of uncertainty around MLEs (two-step/hybrid)

Health economic evaluation — How does it work?

	Parameters simulations				Expected utility	
Iter/n	a	b	...	h	$\text{NB}_0(\boldsymbol{\theta})$	$\text{NB}_1(\boldsymbol{\theta})$
1	0.365	0.076	...	0.162	19 214 751	*19 647 706*
2	0.421	0.024	...	0.134	*17 165 526*	17 163 407
3	0.125	0.017	...	0.149	18 710 928	16 458 433
4	0.117	0.073	...	0.120	16 991 321	*18 497 648*
5	0.481	0.008	...	0.191	*19 772 898*	18 662 329
6	0.163	0.127	...	0.004	17 106 136	*18 983 331*
...		...				...
1000	0.354	0.067	...	0.117	*18 043 921*	16 470 805
				Average	18 659 238	*19 515 004*

- Uncertainty in the parameters induces a distribution of decisions (based on the net benefits)
  - In each parameters configuration can identify the *optimal strategy*
- Averaging over the uncertainty in $\boldsymbol{\theta}$ provides the overall optimal decision, *given current uncertainty* (= choose the intervention associated with *highest expected utility*)

Health economic evaluation — How does it work?

	Parameters simulations				Expected utility		Maximum	Opportunity
Iter/n	a	b	...	h	$\text{NB}_0(\boldsymbol{\theta})$	$\text{NB}_1(\boldsymbol{\theta})$	net benefit	loss
1	0.365	0.076	...	0.162	19 214 751	*19 647 706*	19 647 706	—
2	0.421	0.024	...	0.134	*17 165 526*	17 163 407	17 165 526	2 119.3
3	0.125	0.017	...	0.149	18 710 928	16 458 433	18 710 928	2 252 495.5
4	0.117	0.073	...	0.120	16 991 321	*18 497 648*	18 497 648	—
5	0.481	0.008	...	0.191	*19 772 898*	18 662 329	19 772 898	1 110 569.3
6	0.163	0.127	...	0.004	17 106 136	*18 983 331*	18 983 331	—
...			...			...	...	
1000	0.354	0.067	...	0.117	*18 043 921*	16 470 805	18 043 921	1 573 116.0
				Average	18 659 238	*19 515 004*	**19 741 589**	**226 585**

- Summarise uncertainty in the decision, eg via the **Expected Value of "Perfect" Information** (EVPI)
  - Defined as the **average Opportunity Loss**
  - Can also be computed as the difference between the **average maximum expected utility under "perfect" information** and the **maximum expected utility overall** — in formula:

$$\text{EVPI} = \text{E}_{\boldsymbol{\theta}}\left[\max_t \text{NB}_t(\boldsymbol{\theta})\right] - \max_t \text{E}_{\boldsymbol{\theta}}\left[\text{NB}_t(\boldsymbol{\theta})\right]$$

Expected Value of Partial Information

- $\boldsymbol{\theta}$ = all the model parameters; can be split into two subsets
  - The "**parameters of interest**" ϕ, e.g. prevalence of a disease, HRQL measures, length of stay in hospital, ...
  - The "**remaining parameters**" ψ, e.g. cost of treatment with other established medications,
- We are interested in quantifying the value of gaining more information on ϕ, while leaving the current level of uncertainty on ψ unchanged

Expected Value of Partial Information

- $\boldsymbol{\theta}$ = all the model parameters; can be split into two subsets
  - The "**parameters of interest**" ϕ, e.g. prevalence of a disease, HRQL measures, length of stay in hospital, ...
  - The "**remaining parameters**" ψ, e.g. cost of treatment with other established medications,
- We are interested in quantifying the value of gaining more information on ϕ, while leaving the current level of uncertainty on ψ unchanged

- In formulæ:
  - First, consider the expected utility (EU) **if** we were able to learn ϕ but not ψ

$$\mathrm{E}_{\psi|\phi}\left[\mathrm{NB}_t(\boldsymbol{\theta})\right]$$

Expected Value of Partial Information

- $\boldsymbol{\theta}$ = all the model parameters; can be split into two subsets
  - The "**parameters of interest**" ϕ, e.g. prevalence of a disease, HRQL measures, length of stay in hospital, ...
  - The "**remaining parameters**" ψ, e.g. cost of treatment with other established medications,
- We are interested in quantifying the value of gaining more information on ϕ, while leaving the current level of uncertainty on ψ unchanged

- In formulæ:
  - First, consider the expected utility (EU) **if** we were able to learn ϕ but not ψ
  - **If** we knew ϕ perfectly, best decision = the maximum of this EU

$$\max_t \mathrm{E}_{\psi|\phi}\left[\mathrm{NB}_t(\boldsymbol{\theta})\right]$$

Expected Value of Partial Information

- $\boldsymbol{\theta}$ = all the model parameters; can be split into two subsets
  - The "**parameters of interest**" ϕ, e.g. prevalence of a disease, HRQL measures, length of stay in hospital, ...
  - The "**remaining parameters**" ψ, e.g. cost of treatment with other established medications,
- We are interested in quantifying the value of gaining more information on ϕ, while leaving the current level of uncertainty on ψ unchanged

- In formulæ:
  - First, consider the expected utility (EU) **if** we were able to learn ϕ but not ψ
  - **If** we knew ϕ perfectly, best decision = the maximum of this EU
  - Of course we cannot learn ϕ **perfectly**, so take the expected value

$$\mathsf{E}_{\phi}\left[\max_{t} \mathsf{E}_{\psi|\phi}\left[\mathsf{NB}_t(\boldsymbol{\theta})\right]\right]$$

Expected Value of Partial Information

- $\boldsymbol{\theta}$ = all the model parameters; can be split into two subsets
  - The "**parameters of interest**" ϕ, e.g. prevalence of a disease, HRQL measures, length of stay in hospital, ...
  - The "**remaining parameters**" ψ, e.g. cost of treatment with other established medications,
- We are interested in quantifying the value of gaining more information on ϕ, while leaving the current level of uncertainty on ψ unchanged

- In formulæ:
  - First, consider the expected utility (EU) **if** we were able to learn ϕ but not ψ
  - **If** we knew ϕ perfectly, best decision = the maximum of this EU
  - Of course we cannot learn ϕ **perfectly**, so take the expected value
  - And compare this with the **maximum expected utility overall**

$$\mathsf{E}_{\phi}\left[\max_{t}\mathsf{E}_{\psi|\phi}\left[\mathsf{NB}_{t}(\boldsymbol{\theta})\right]\right]-\max_{t}\mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_{t}(\boldsymbol{\theta})\right]$$

Expected Value of Partial Information

- $\boldsymbol{\theta}$ = all the model parameters; can be split into two subsets
  - The "**parameters of interest**" ϕ, e.g. prevalence of a disease, HRQL measures, length of stay in hospital, ...
  - The "**remaining parameters**" ψ, e.g. cost of treatment with other established medications,
- We are interested in quantifying the value of gaining more information on ϕ, while leaving the current level of uncertainty on ψ unchanged

- In formulæ:
  - First, consider the expected utility (EU) **if** we were able to learn ϕ but not ψ
  - **If** we knew ϕ perfectly, best decision = the maximum of this EU
  - Of course we cannot learn ϕ **perfectly**, so take the expected value
  - And compare this with the **maximum expected utility overall**
  - This is the EVPPI!

$$\text{EVPPI} = \mathsf{E}_{\phi}\left[\max_t \mathsf{E}_{\psi|\phi}\left[\text{NB}_t(\boldsymbol{\theta})\right]\right] - \max_t \mathsf{E}_{\boldsymbol{\theta}}\left[\text{NB}_t(\boldsymbol{\theta})\right]$$

Expected Value of Partial Information

- $\boldsymbol{\theta}$ = all the model parameters; can be split into two subsets
  - The "**parameters of interest**" ϕ, e.g. prevalence of a disease, HRQL measures, length of stay in hospital, ...
  - The "**remaining parameters**" ψ, e.g. cost of treatment with other established medications,
- We are interested in quantifying the value of gaining more information on ϕ, while leaving the current level of uncertainty on ψ unchanged

- In formulæ:
  - First, consider the expected utility (EU) **if** we were able to learn ϕ but not ψ
  - **If** we knew ϕ perfectly, best decision = the maximum of this EU
  - Of course we cannot learn ϕ **perfectly**, so take the expected value
  - And compare this with the **maximum expected utility overall**
  - This is the EVPPI!

$$\text{EVPPI} = \text{E}_{\phi}\left[\max_t \text{E}_{\psi|\phi}\left[\text{NB}_t(\boldsymbol{\theta})\right]\right] - \max_t \text{E}_{\boldsymbol{\theta}}\left[\text{NB}_t(\boldsymbol{\theta})\right]$$

  - That's the difficult part! Can do nested Monte Carlo, but takes for ever to get accurate results, so nobody bothers...

EVPPI as a regression problem

- Can model as a **regression** problem

$$\begin{aligned} \text{NB}_t(\boldsymbol{\theta}) &= \text{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[\text{NB}_t(\boldsymbol{\theta})\right] + \varepsilon, \qquad \text{with } \varepsilon \sim \text{Normal}(0, \sigma_\varepsilon^2) \\ &= g_t(\boldsymbol{\phi}) + \varepsilon \end{aligned}$$

"Data": **simulations** for $\text{NB}_t(\boldsymbol{\theta})$ as "response"
simulations for $\boldsymbol{\phi}$ as "covariates"

a	b	c	...	f	g	h	$\text{NB}_0(\boldsymbol{\theta})$	$\text{NB}_1(\boldsymbol{\theta})$
0.365	0.076	0.243	...	0.622	0.001	0.162	19 214 751	19 647 706
0.421	0.024	0.115	...	0.519	0.010	0.134	17 165 526	17 163 407
0.125	0.017	0.420	...	0.482	0.007	0.149	18 710 928	16 458 433
0.117	0.073	0.419	...	0.317	0.003	0.120	16 991 321	18 497 648
0.481	0.008	0.176	...	0.497	0.004	0.191	19 772 898	18 662 329
0.163	0.127	0.227	...	0.613	0.083	0.004	17 106 136	18 983 331
	...		...		...		...	
0.354	0.067	0.318	...	0.519	0.063	0.117	18 043 921	16 470 805

"covariates" "response" "response"

Strong and Oakley (2014) [7]

EVPPI as a regression problem

- Can model as a **regression** problem

$$\begin{aligned} \text{NB}_t(\boldsymbol{\theta}) &= \text{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[\text{NB}_t(\boldsymbol{\theta})\right] + \varepsilon, \qquad \text{with } \varepsilon \sim \text{Normal}(0, \sigma^2_\varepsilon) \\ &= g_t(\boldsymbol{\phi}) + \varepsilon \end{aligned}$$

"Data": **simulations** for $\text{NB}_t(\boldsymbol{\theta})$ as "response"
simulations for $\boldsymbol{\phi}$ as "covariates"

- Once the functions $g_t(\boldsymbol{\phi})$ are estimated, then can approximate

$$\begin{aligned} \text{EVPPI} &= \text{E}_{\boldsymbol{\phi}}\left[\max_t \text{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[\text{NB}_t(\boldsymbol{\theta})\right]\right] - \max_t \text{E}_{\boldsymbol{\theta}}\left[\text{NB}_t(\boldsymbol{\theta})\right] \\ &\approx \frac{1}{S}\sum_{s=1}^{S} \max_t \hat{g}_t(\boldsymbol{\phi}_s) - \max_t \frac{1}{S}\sum_{s=1}^{S} \hat{g}_t(\boldsymbol{\phi}_s) \end{aligned}$$

Strong and Oakley (2014) [7]

EVPPI as a regression problem

- Can model as a **regression** problem

$$\begin{aligned} \text{NB}_t(\boldsymbol{\theta}) &= \text{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[\text{NB}_t(\boldsymbol{\theta})\right] + \varepsilon, \qquad \text{with } \varepsilon \sim \text{Normal}(0, \sigma^2_\varepsilon) \\ &= g_t(\boldsymbol{\phi}) + \varepsilon \end{aligned}$$

"Data": **simulations** for $\text{NB}_t(\boldsymbol{\theta})$ as "response"
simulations for $\boldsymbol{\phi}$ as "covariates"

- Once the functions $g_t(\boldsymbol{\phi})$ are estimated, then can approximate

$$\begin{aligned} \text{EVPPI} &= \text{E}_{\boldsymbol{\phi}}\left[\max_t \text{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[\text{NB}_t(\boldsymbol{\theta})\right]\right] - \max_t \text{E}_{\boldsymbol{\theta}}\left[\text{NB}_t(\boldsymbol{\theta})\right] \\ &\approx \frac{1}{S}\sum_{s=1}^{S} \max_t \hat{g}_t(\boldsymbol{\phi}_s) - \max_t \frac{1}{S}\sum_{s=1}^{S} \hat{g}_t(\boldsymbol{\phi}_s) \end{aligned}$$

- **NB**: $g_t(\boldsymbol{\phi})$ can be complex, so need to use **flexible** regression methods
  - GAMs are very fast, but only work if number of important parameters $P \leq 5$
  - If $P > 5$, can use **Gaussian Process** regression

Strong and Oakley (2014) [7]

EVPPI via GP regression

Model

$$\begin{pmatrix} \text{NB}_t(\boldsymbol{\theta}_1) \\ \text{NB}_t(\boldsymbol{\theta}_2) \\ \vdots \\ \text{NB}_t(\boldsymbol{\theta}_S) \end{pmatrix} := \mathbf{NB}_t \sim \text{Normal}(\boldsymbol{H\beta}, \boldsymbol{\mathcal{C}}_{\text{Exp}} + \sigma_\varepsilon^2 \boldsymbol{I})$$

$$\boldsymbol{H} = \begin{pmatrix} 1 & \phi_{11} & \cdots & \phi_{1P} \\ 1 & \phi_{21} & \cdots & \phi_{2P} \\ \vdots & & \ddots & \\ 1 & \phi_{S1} & \cdots & \phi_{SP} \end{pmatrix} \quad \text{and} \quad \mathcal{C}_{\text{Exp}}(r,s) = \sigma^2 \exp\left[\sum_{p=1}^{P} \left(\frac{\phi_{rp} - \phi_{sp}}{\delta_p}\right)^2\right]$$

- Parameters: $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, σ^2, σ_ε^2
- Very flexible structure — good approximation level
- Can use conjugate priors + numerical optimisation, **but** can still be very slow — computational cost in the order of S^3 (involves inversion of a dense covariance matrix)

Strong and Oakley (2014) [7]

EVPPI via GP regression — **but faster**

Heath et al (2016) [3]

EVPPI via GP regression — **but faster**

1. Build from ideas in spatial statistics and use a Matérn covariance function

$$\mathcal{C}_{\mathrm{M}}(r,s) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa\|\boldsymbol{\phi}_r - \boldsymbol{\phi}_s\|)^{\nu}\mathrm{K}_{\nu}(\kappa\|\boldsymbol{\phi}_r - \boldsymbol{\phi}_s\|)$$

   - Fewer parameters, but still implies a dense covariance matrix
   - **But**: can use efficient algorithms to solve **Stochastic Partial Differential Equations** (SPDE) to approximate it — with computational cost $\propto S^{3/2}$!

Heath et al (2016) [3]; Lindgren et al (2011) [4]

EVPPI via GP regression — **but faster**

1. Build from ideas in spatial statistics and use a Matérn covariance function

$$\mathcal{C}_{\mathrm{M}}(r,s) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa\|\boldsymbol{\phi}_r - \boldsymbol{\phi}_s\|)^{\nu}\mathrm{K}_{\nu}(\kappa\|\boldsymbol{\phi}_r - \boldsymbol{\phi}_s\|)$$

   - Fewer parameters, but still implies a dense covariance matrix
   - **But**: can use efficient algorithms to solve **Stochastic Partial Differential Equations** (SPDE) to approximate it — with computational cost $\propto S^{3/2}$!

2. Re-formulate the model as

$$\begin{aligned} \mathbf{NB}_t &\sim \text{Normal}(\boldsymbol{H\beta}, \boldsymbol{\mathcal{C}}_{\mathrm{M}} + \sigma^2_{\varepsilon}\boldsymbol{I}) \\ &\sim \text{Normal}(\boldsymbol{H\beta} + f(\boldsymbol{\omega}), \sigma^2_{\varepsilon}\boldsymbol{I}) \end{aligned}$$

   - $f(\boldsymbol{\omega})$ are a set of "spatially structured" effects, with $\boldsymbol{\omega} \sim \text{Normal}\left(0, \boldsymbol{Q}^{-1}(\xi)\right)$
   - $\boldsymbol{Q}(\xi)$ is a **sparse** precision matrix determined by the SPDE solution

Heath et al (2016) [3]; Lindgren et al (2011) [4]

EVPPI via GP regression — **but faster**

1. Build from ideas in spatial statistics and use a Matérn covariance function

$$\mathcal{C}_{\mathrm{M}}(r,s) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa\|\boldsymbol{\phi}_r - \boldsymbol{\phi}_s\|)^{\nu}\mathrm{K}_{\nu}(\kappa\|\boldsymbol{\phi}_r - \boldsymbol{\phi}_s\|)$$

   - Fewer parameters, but still implies a dense covariance matrix
   - **But**: can use efficient algorithms to solve **Stochastic Partial Differential Equations** (SPDE) to approximate it — with computational cost $\propto S^{3/2}$!

2. Re-formulate the model as

$$\begin{aligned}\mathbf{NB}_t &\sim \text{Normal}(\boldsymbol{H\beta}, \boldsymbol{\mathcal{C}}_{\mathrm{M}} + \sigma^2_{\varepsilon}\boldsymbol{I}) \\ &\sim \text{Normal}(\boldsymbol{H\beta} + f(\boldsymbol{\omega}), \sigma^2_{\varepsilon}\boldsymbol{I})\end{aligned}$$

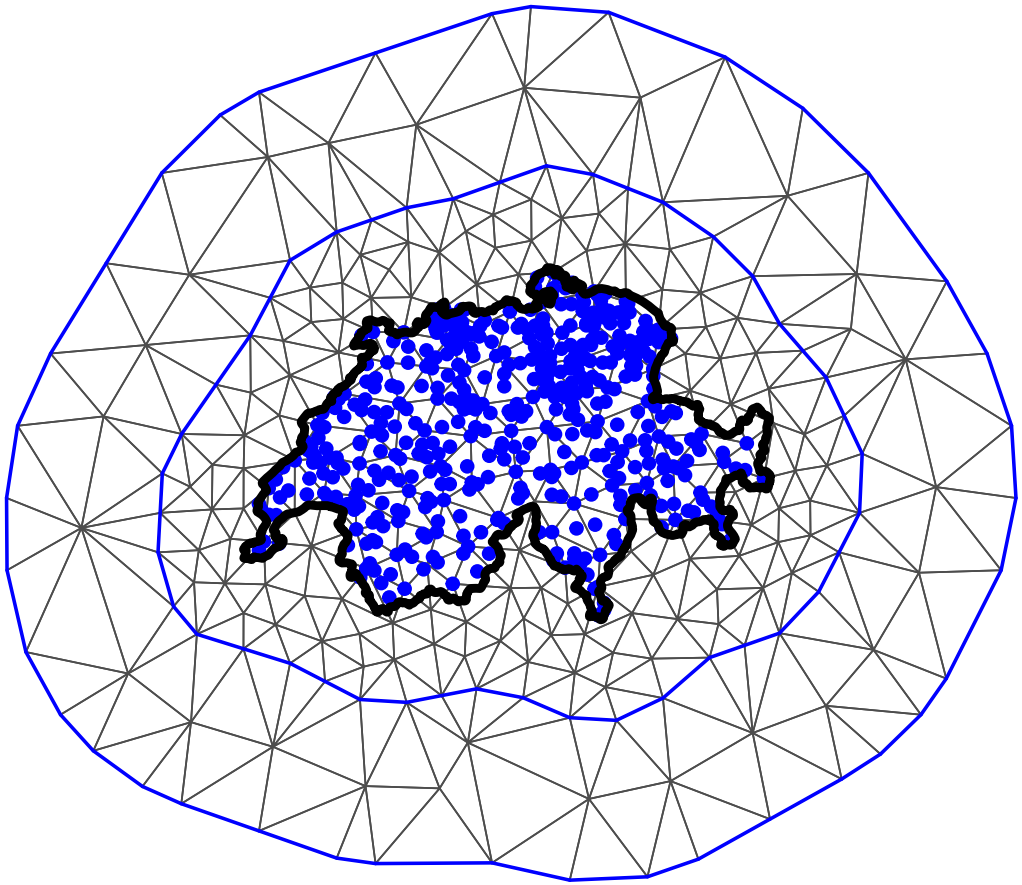
   - $f(\boldsymbol{\omega})$ are a set of "spatially structured" effects, with $\boldsymbol{\omega} \sim \text{Normal}\left(0, \boldsymbol{Q}^{-1}(\xi)\right)$
   - $\boldsymbol{Q}(\xi)$ is a **sparse** precision matrix determined by the SPDE solution

3. Crucially, if we set a sparse Gaussian prior on $\boldsymbol{\beta}$, this is a Latent Gaussian model $\Rightarrow$ can be estimated using super-fast **Integrated Nested Laplace Approximation** (INLA)

Heath et al (2016) [3]; Lindgren et al (2011) [4]; Rue et al (2009) [5]

Lost in space

- In a "proper" spatial problem, data are observed at a bivariate grid of points
  - Points that are closer tend to be more correlated than points further apart
  - The INLA-SPDE procedure builds a grid approximation of the underlying bidimensional space
  - Points not on the grid are estimated by interpolation — deriving a full surface



Lost in space

- In a "proper" spatial problem, data are observed at a bivariate grid of points
  - Points that are closer tend to be more correlated than points further apart
  - The INLA-SPDE procedure builds a grid approximation of the underlying bidimensional space
  - Points not on the grid are estimated by interpolation — deriving a full surface
- In our case, data are observed on a high-dimensional space, with no proper "spatial" interpretation!
- Need to use some form of **dimensionality reduction** to project the P-dimensional space of ϕ to a 2-dimensional space
  - Simple solution: use PCA to preserve Euclidean distances and thus capture the "spatial" correlation across the elements of ϕ
  - Even better, regression-based dimension reduction method: **Principal Fitted Components**

Lost in space

- In a "proper" spatial problem, data are observed at a bivariate grid of points
  - Points that are closer tend to be more correlated than points further apart
  - The INLA-SPDE procedure builds a grid approximation of the underlying bidimensional space
  - Points not on the grid are estimated by interpolation — deriving a full surface
- In our case, data are observed on a high-dimensional space, with no proper "spatial" interpretation!
- Need to use some form of **dimensionality reduction** to project the P-dimensional space of ϕ to a 2-dimensional space
  - Simple solution: use PCA to preserve Euclidean distances and thus capture the "spatial" correlation across the elements of ϕ
  - Even better, regression-based dimension reduction method: **Principal Fitted Components**

NB: All methods implemented in the `R` package `BCEA` (Bayesian Cost-Effectiveness Analysis: `http://www.statistica.it/gianluca/BCEA`)

Baio et al (2016) [1]

Principal Fitted Components

- Objective: find a **sufficient** dimensionality reduction
  - Estimate the function $R(\boldsymbol{\phi}) : P \rightarrow d$ so that $\mathbf{NB}_t \perp\!\!\!\perp \boldsymbol{\phi} \mid R(\boldsymbol{\phi})$
  - "**Project**" the $P-$dimensional information contained in $\boldsymbol{\phi}$ to the $d-$dimensional function $R(\cdot)$
  - Ideally, $d << P$ — in fact, would like $d \leq 2$

Principal Fitted Components

- Objective: find a **sufficient** dimensionality reduction
  - Estimate the function $R(\boldsymbol{\phi}) : P \rightarrow d$ so that $\mathbf{NB}_t \perp\!\!\!\perp \boldsymbol{\phi} \mid R(\boldsymbol{\phi})$
  - "**Project**" the $P-$dimensional information contained in $\boldsymbol{\phi}$ to the $d-$dimensional function $R(\cdot)$
  - Ideally, $d << P$ — in fact, would like $d \leq 2$

- "Inverse regression" model

$$\boldsymbol{\phi} = \boldsymbol{\mu} + \boldsymbol{\Upsilon} \boldsymbol{f}(\mathbf{NB}_t) + \boldsymbol{\epsilon}$$

  with
  - $\boldsymbol{\mu}$ = intercept
  - $\boldsymbol{\Upsilon} = P \times d$ dimensionality reduction matrix
  - $\boldsymbol{f}(\mathbf{NB}_t)$ = vector-valued function of the "response"
  - $\boldsymbol{\epsilon}$ = error term

Principal Fitted Components

- Objective: find a **sufficient** dimensionality reduction
  - Estimate the function $R(\boldsymbol{\phi}) : P \rightarrow d$ so that $\mathbf{NB}_t \perp\!\!\!\perp \boldsymbol{\phi} \mid R(\boldsymbol{\phi})$
  - "**Project**" the $P-$dimensional information contained in $\boldsymbol{\phi}$ to the $d-$dimensional function $R(\cdot)$
  - Ideally, $d << P$ — in fact, would like $d \leq 2$

- "Inverse regression" model

$$\boldsymbol{\phi} = \boldsymbol{\mu} + \boldsymbol{\Upsilon} \boldsymbol{f}(\mathbf{NB}_t) + \boldsymbol{\epsilon}$$

  with
  - $\boldsymbol{\mu}$ = intercept
  - $\boldsymbol{\Upsilon} = P \times d$ dimensionality reduction matrix
  - $\boldsymbol{f}(\mathbf{NB}_t)$ = vector-valued function of the "response"
  - $\boldsymbol{\epsilon}$ = error term

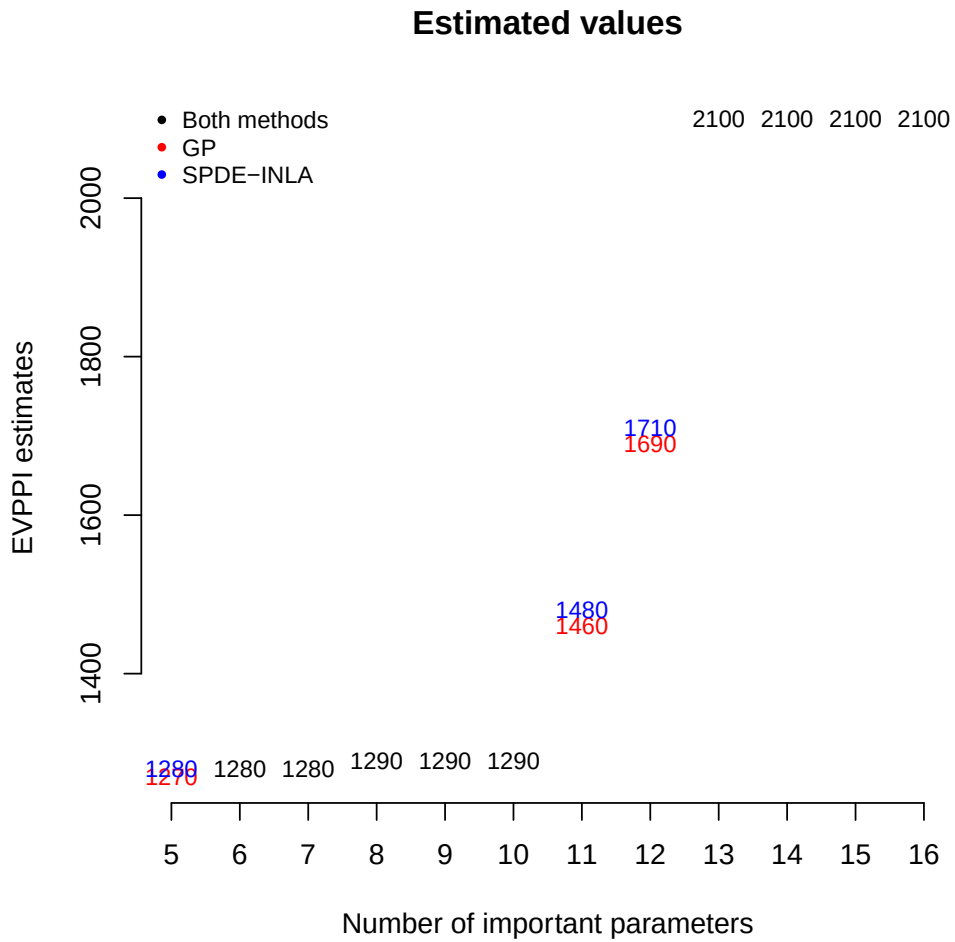
- Main advantages
  - Computational cost is negligible
  - Can use model-fitting statistics (eg AIC) to determine the "best" model for given choices of d $(= 2, 3, \ldots)$
  - **NB**: if the AIC suggests $d > 2$ then EVPPI estimates likely to be biased!

In a nutshell...

1. Make the model structure more complex, by assuming that the EVPPI is estimated by a **linear predictor** ("fixed" effects) + a **spatially structured** ("random" effect) component, accounting for the correlation among parameters
2. Find the best performing inverse regression model by AIC (as a failsafe measure) & compute the PFC model with 2 dimensions
3. Use the projections as the "spatial location" for the net benefit values and estimate the Matérn GP via SPDE
4. Use INLA to estimate the posterior distribution for the model parameters
5. Compute the fitted values $\hat{g}_t(\boldsymbol{\phi}_s)$
6. Use the fitted values to calculate the estimate of the EVPPI as

$$\widehat{\text{EVPPI}} = \frac{1}{S}\sum_{s=1}^{S} \max_t \hat{g}_t(\boldsymbol{\phi}_s) - \max_t \frac{1}{S}\sum_{s=1}^{S} \hat{g}_t(\boldsymbol{\phi}_s)$$

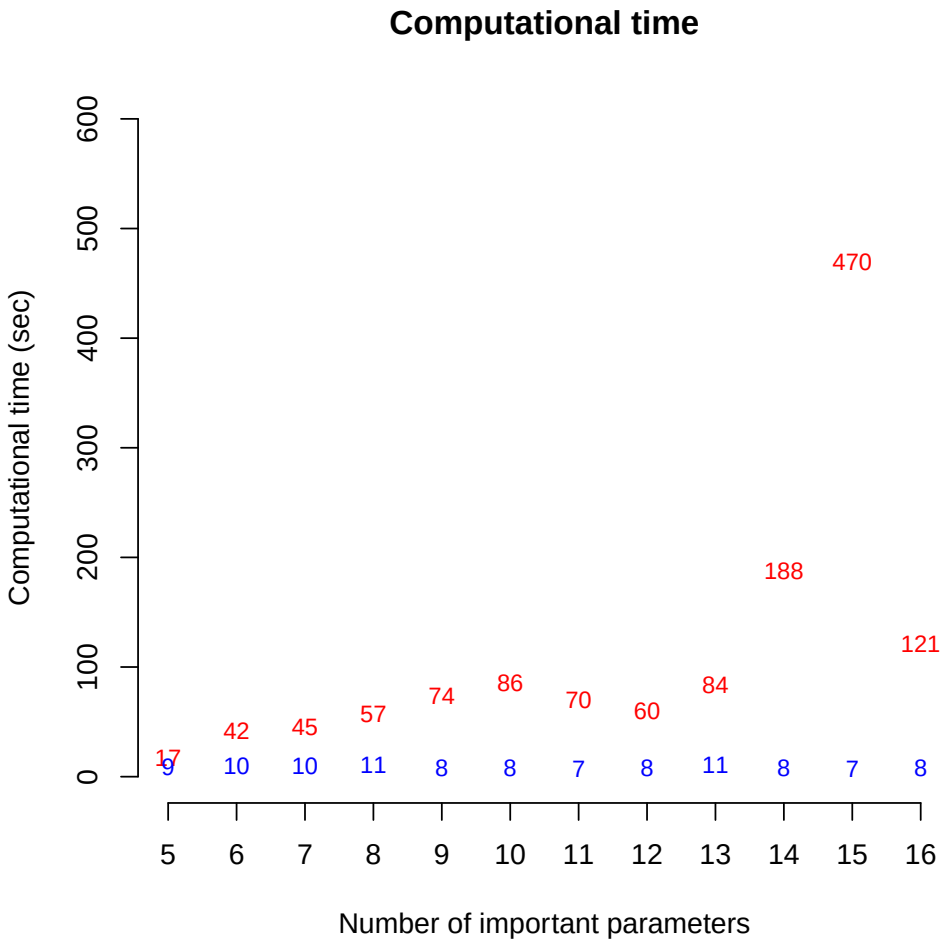
Examples — SAVI

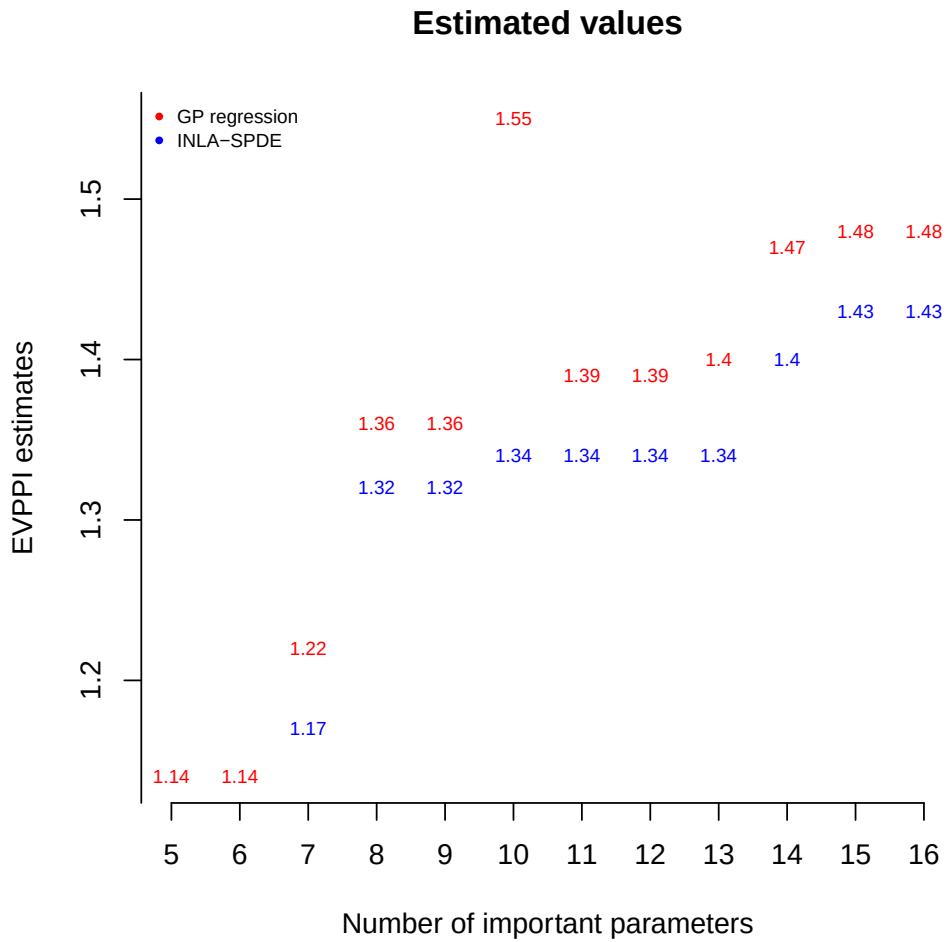


Running time for a single value of k

Sheffield Accelerated Value of Information [6]

Examples — Vaccine





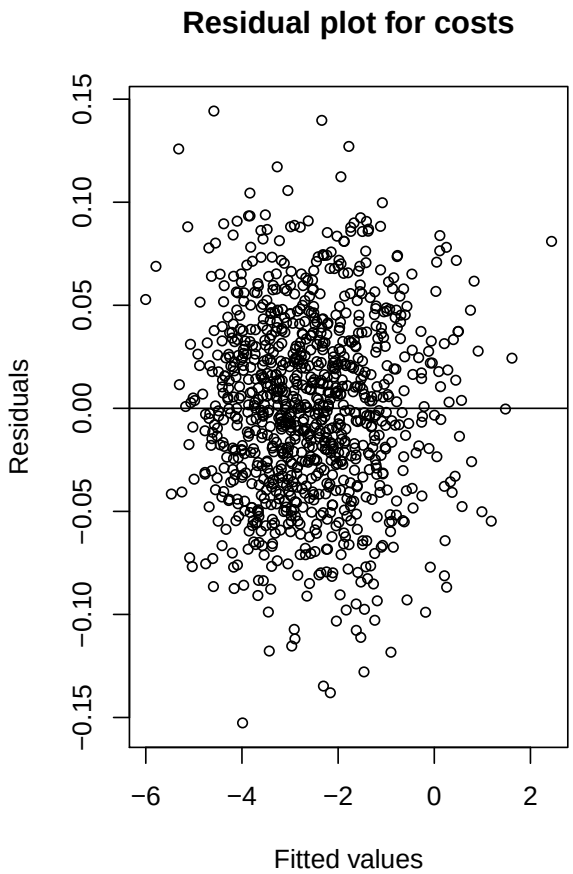
Running time for a single value of k

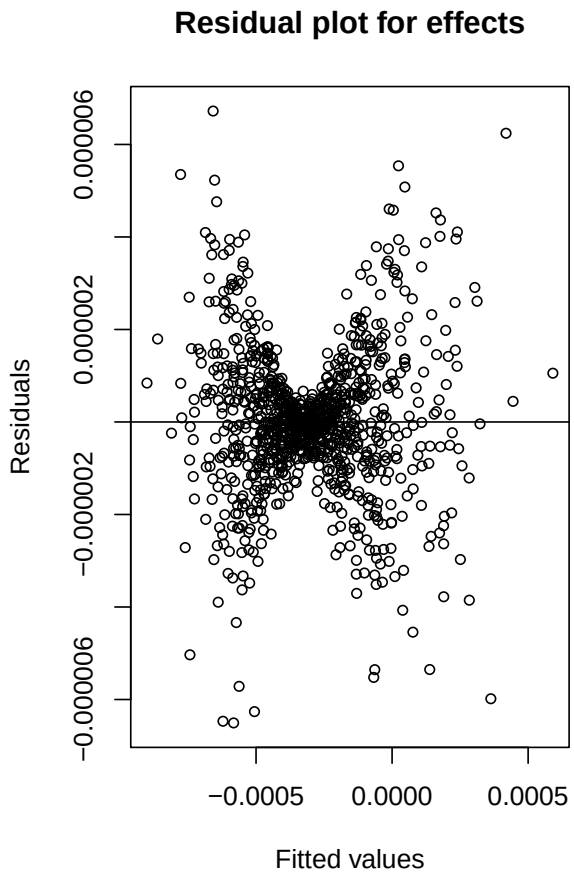
Baio and Dawid (2011) [2]

Breaking bad...

Breast cancer screening (Welton et al 2008) [8]

- Multi-decision model developed for the UK setting, with 4 interventions
- Complex evidence synthesis for 6 parameters — highly structured!

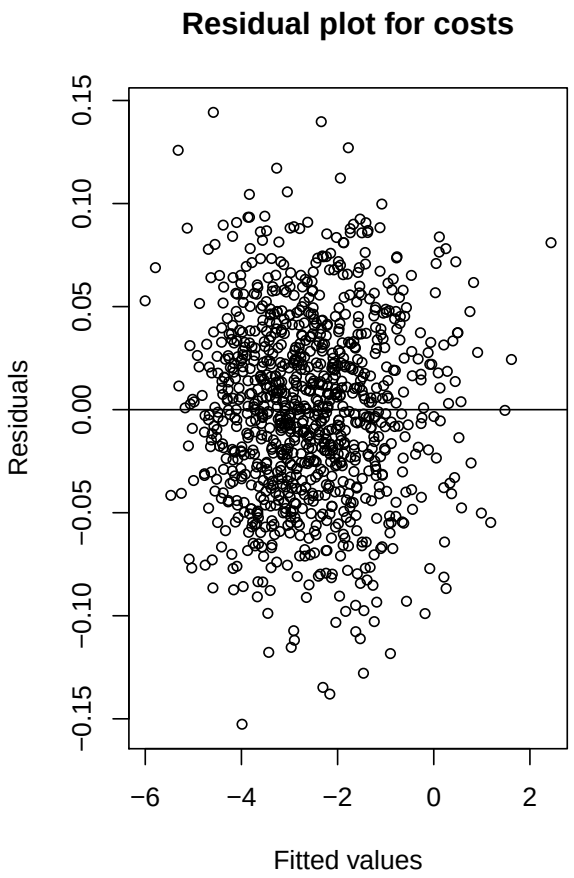


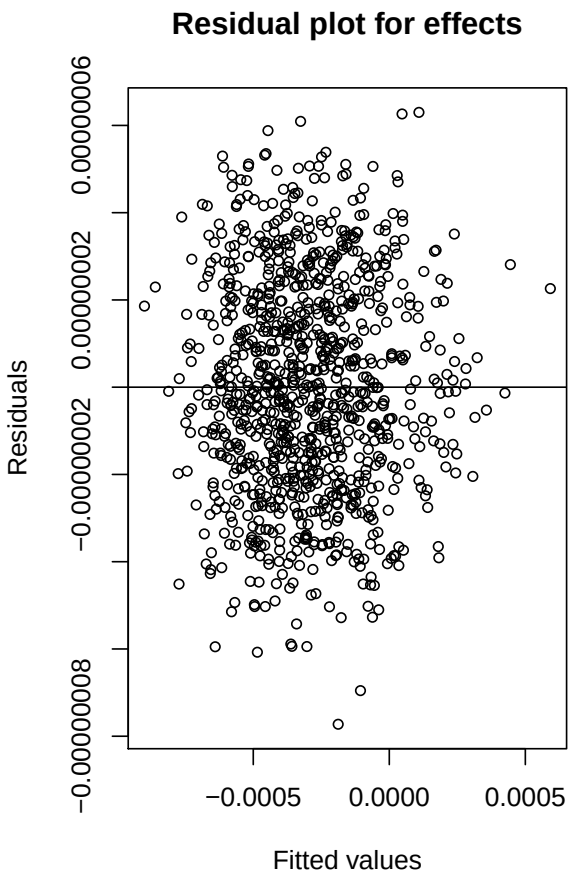


The fix!

- Can relatively easily modify the basic structure of the model, e.g. include interaction terms to make $\boldsymbol{H\beta}$ non-linear

$$\beta_0 + \beta_1\phi_{1s} + \beta_2\phi_{2s} + \beta_3\phi_{3s} + \beta_4\phi_{1s}\phi_{2s} + \beta_5\phi_{1s}\phi_{3s} + \beta_6\phi_{2s}\phi_{3s}$$





Baio et al (2016) [1]

Conclusions

- VoI methods are theoretically valid (**ideal**?) to quantify decision uncertainty
  - Directly related to research prioritisation
  - Address the issue of uncertainty vs consequences
- **But**: their application has been hindered by the computational cost involved in calculating the EVPPI

- Methods based on non-parametric regression to calculate the EVPPI are efficient, but in some cases still computationally expensive

- Can overcome these limitations by drawing on methods from spatial statistics
  - Efficient algorithm — around 10 seconds for 1000 PSA samples in the basic formulation
  - Relatively easy (and **not too expensive**!) to use more complex formulation to deal with more complex cases
  - Implemented in `BCEA` — practitioners can use them in a relatively straightforward way!

References

[1] G. Baio, A. Berardi, and A. Heath. *Bayesian Cost Effectiveness Analysis with the R package BCEA*. Springer, forthcoming.

[2] G. Baio and P. Dawid. Probabilistic sensitivity analysis in health economics. *Statistical methods in medical research*, Sep 2011.

[3] A Heath, I Manolopoulou, and G Baio. Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation. *Statistics in Medicine*, May 2016.

[4] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

[5] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal the Royal Statistical Society, Series B*, 71:319–392, 2009.

[6] M. Strong, P. Breeze, C. Thomas, and A. Brennan. *SAVI - Sheffield Accelerated Value of Information, Release version 1.013 (2014-12-11)*, 2014.

[7] M. Strong, J. Oakley, and A. Brennan. Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample A Nonparametric Regression Approach. *Medical Decision Making*, 34(3):311–326, 2014.

[8] N. Welton, A. Ades, D. Caldwell, and T. Peters. Research prioritization based on expected value of partial perfect information: a case-study on interventions to increase uptake of breast cancer screening. *Journal of the Royal Statistical Society Series A*, 171:807–841, 2008.

[9] N. Welton, A. Sutton, N. Cooper, K. Abrams, and A. Ades. *Evidence Synthesis for Decision Making in Healthcare*. John Wiley & Sons, Ltd., Chichester, United Kingdom, 2012.

Thank you!