A practical test for the choice of mixing distribution in discrete choice models

Mogens Fosgerau, mf@dtf.dk

paper written with Michel Bierlaire

DANMARKS TRANSPORTFORSKNING

Outline

1. Brief motivation
2. Tricks with densities
3. Monte Carlo Study
4. Application to real data
5. Conclusion

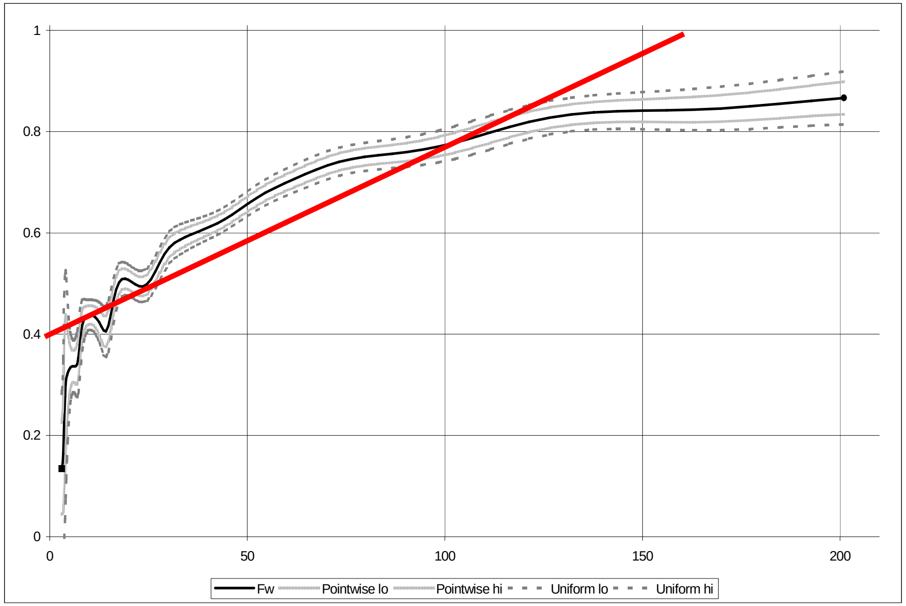
Motivation 1

Consider a model for discrete choice data such as a mixed multinomial logit model

The distribution for a random parameter is generally just assumed

Sometimes with reference to "behavioural realism" in order to get the desired result

But the chosen distribution may not fit the data



Consequences may be dramatic for the estimation of characteristics of the distribution

Motivation 2

This paper proposes a test of the fit of a given one-dimensional mixing distribution against an arbitrarily general alternative

The technique may also be used just to generate a flexible class of densities around some base density

And it may be used to identify an appropriate distributional assumption

Tricks with densities 1

Consider the space of functions on the unit interval, $L_1[0:1]$, with norm: $||f||=\int|f(x)|dx$, inner product $<f,g> = \int fg$

Let $\{L_k\}$ be an orthonormal base for $L_1[0:1]$

We use Legendre polynomials, could also be sines, cosines, etc.

Then any density in this space may be approximated arbitrarily well by:

$$q_K(\eta)=\frac{1}{1+\sum_{k<K}\gamma_k^2}\left(1+\sum_{k<K}\gamma_k L_k(\eta)\right)^2$$

Square to ensure positivity

Integrate to 1

This approximates any function in L_1, also the squareroot of a density

DANMARKS TRANSPORTFORSKNING

Tricks with densities 2

$$q_K(\eta) = \frac{1}{1 + \sum_{k<K} \gamma_k^2} \left(1 + \sum_{k<K} \gamma_k L_k(\eta)\right)^2$$

- This is a density
- Parametrised by $\gamma=(\gamma_1,\ldots,\gamma_{K-1})$
- $q_K=1$ when $\gamma=0$, so reduces to the density of a uniform distribution
- Flexibility and ability to approximate arbitrary densities on [0:1] determined by K.

Tricks with densities 3
Turning one density into another

- Let G be a **true** unknown distribution with density g
- Let F be a **base** distribution with density f, our null hypothesis
  - Assume $\mathrm{supp}(g) \subseteq \mathrm{supp}(f)$
- Write the equation $G(w) = Q(F(w))$ (or $Q(\eta) = G(F^{-1}(\eta))$)
- Then Q is a CDF on [0;1] with density q

- The true density can be written in terms of f and q

$$g(w) = q(F(w))f(w)$$

- Finding Q is equivalent to finding G
  - Problem of G is reduced to finding a density on the unit interval

Tricks with densities 4
Modifying a discrete choice model

- Consider a model, $P(y|w)$, for the distribution of a discrete y conditional on a random latent parameter w
- The latent parameter has true density **g**
- We think the density may be **f**, our null hypothesis

Tricks with densities 5
The likelihood function

Unconditional model with unknown true g:

$$P(y) = \int g(w) P(y \mid w) dw$$

Substitute with base distribution η=F(w)

$$P(y) = \int q(\eta) P(y \mid F^{-1}(\eta)) d\eta$$

Unknown density on the unit interval enters as a weight

Random draws like in usual simulation of the likelihood function

Tricks with densities 6
Extending the model

- Insert approximation of q into the model
- Estimate the model including parameters γ for q
- This nests the standard model using f
- If f = g then q = 1, such that $\gamma = 0$
  - This is a straight-forward test of a parameter restriction.
  - Use LR test
  - If accept, then the model reduces to the standard model

$$P(y) \approx \int q_K(\eta : \gamma_K) P(y \mid F^{-1}(\eta)) d\eta$$

$$P(y) = \int P(y \mid F^{-1}(\eta)) d\eta$$

DANMARKS TRANSPORTFORSKNING

Monte Carlo study – MC1

- Quasi-simulated panel data
- Choice experiment: 1000 subjects, each 8 binary choices
- Route choice: time and cost only
  - One alternative is slow and cheap, the other is fast and expensive
- Responses generated according to known model
  - Mixed logit
  - Parameter for cost is fixed
  - Parameter for time follows g, which is either normal or lognormal
  - So WTP for time is either normal or lognormal
  - Estimate models assuming f is normal or lognormal
  - Generate 100 datasets for each case
  - with 0, 1 or 2 SNP γ terms in approximation of q
  - Measure power and size of test that f = g

MC2 - results

			H_0 (f)	
	True (g)		Normal	Lognormal
1 SNP term	Normal	95%	9	99
	Lognormal	95%	100	5
	Normal	99%	1	78
	Lognormal	99%	88	0
2 SNP terms	Normal	95%	9	100
	Lognormal	95%	100	3
	Normal	99%	4	99
	Lognormal	99%	100	0

Power is high!

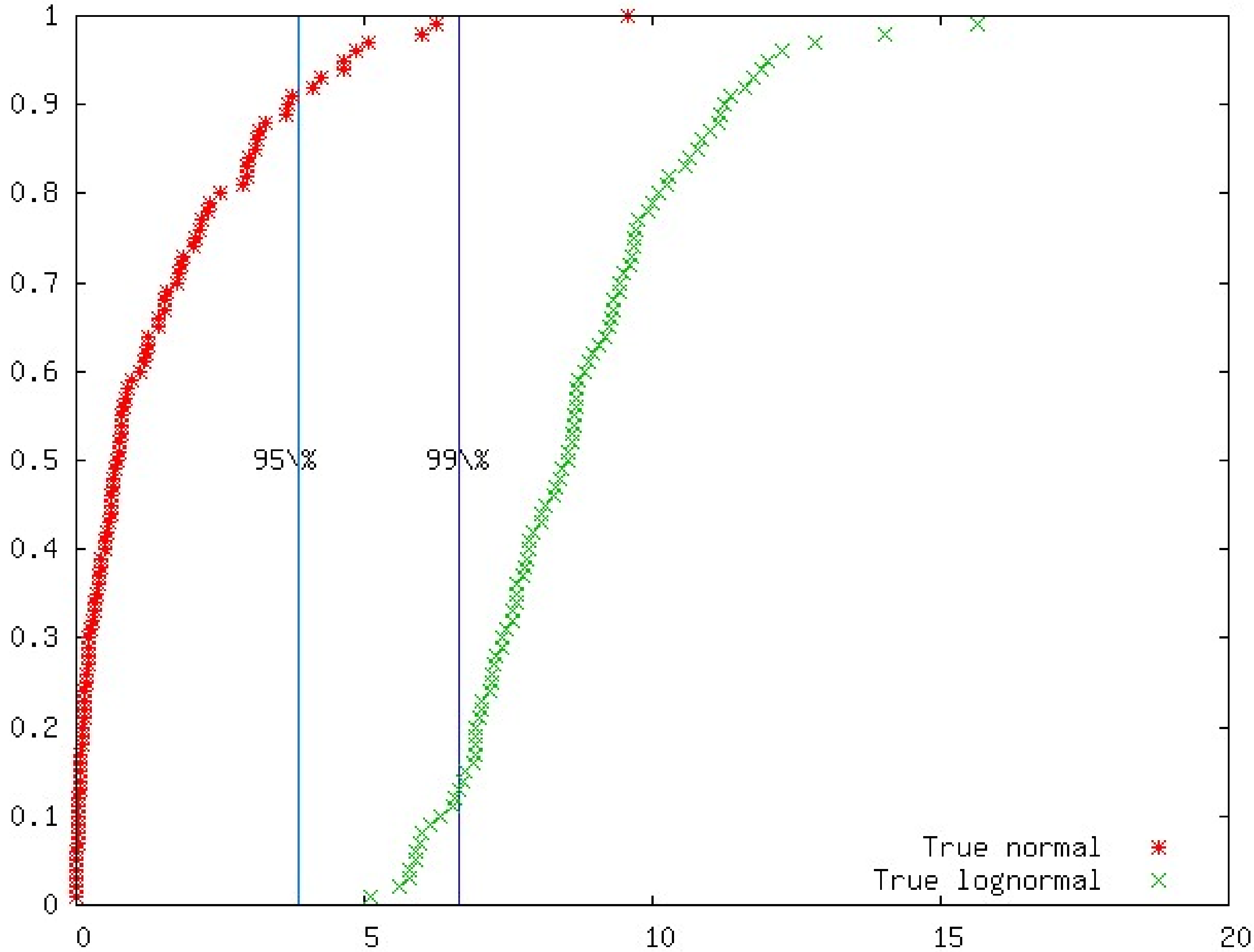
Higher with 2 SNP terms

In this case we can almost always reject a wrong hypothesis

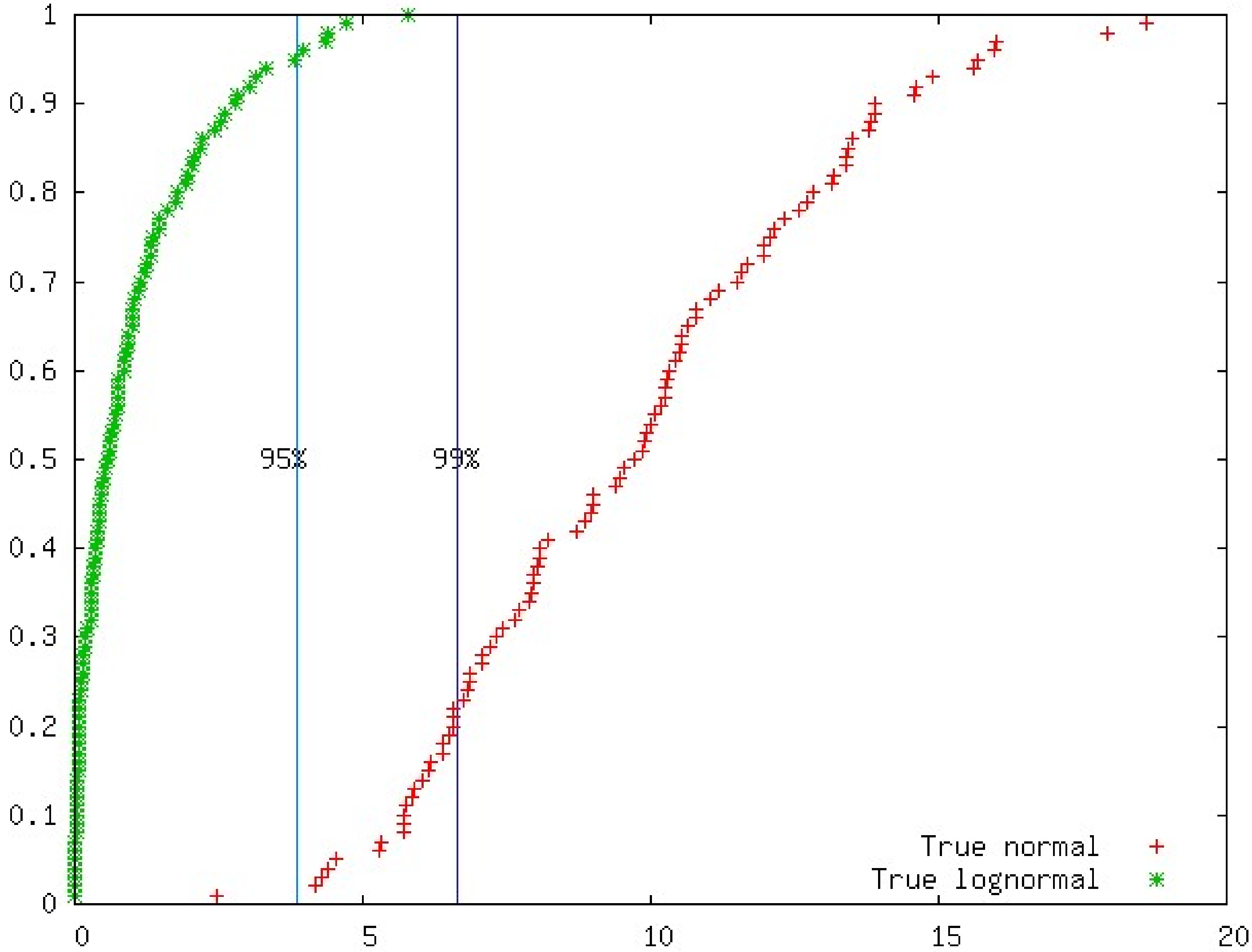
Size is good but sometimes high.

We sometimes reject a true hypothesis, but (mostly) not too often

Likelihood ratio under H_0="true normal"



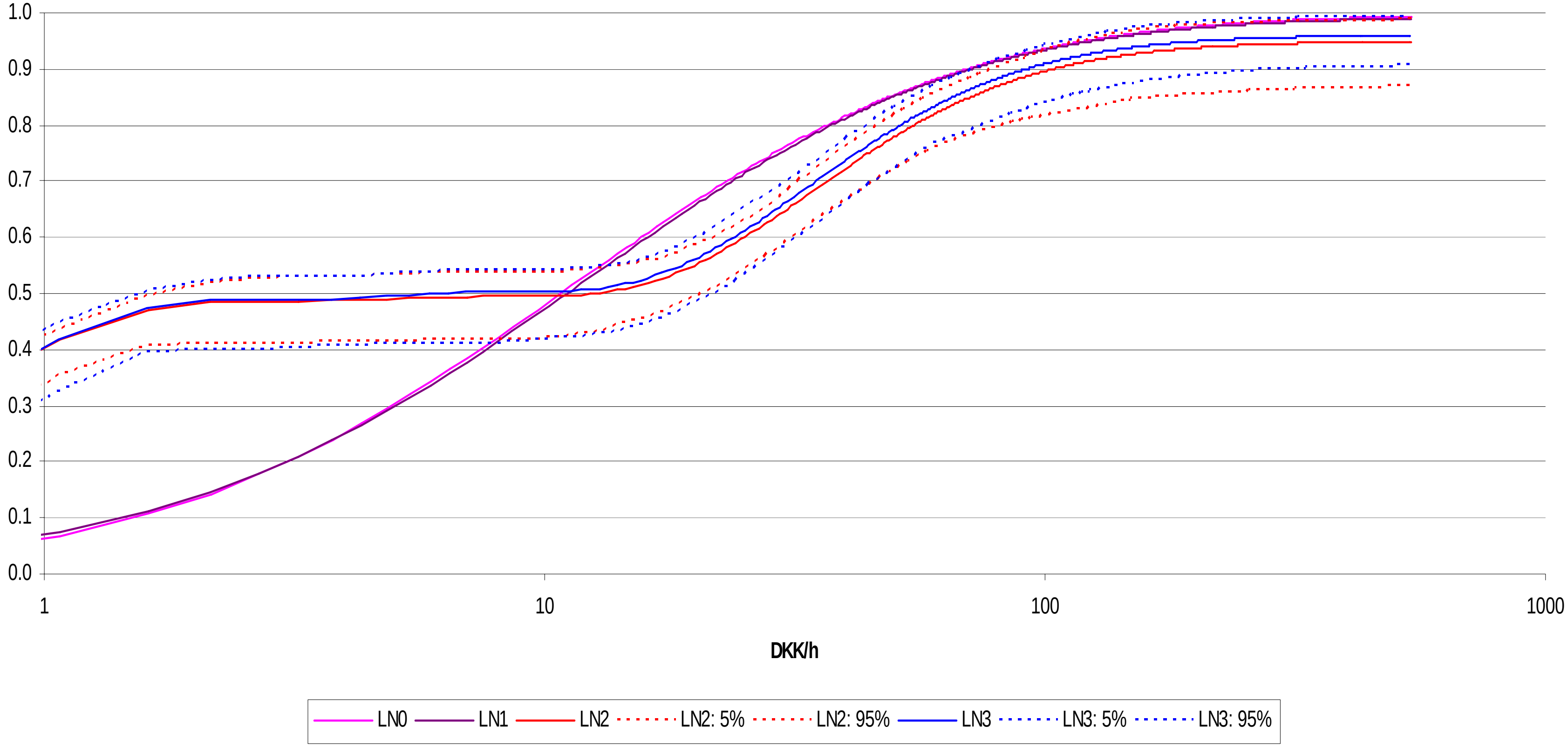
Likelihood ratio under H_0="true lognormal"



Application to real data 1

- Model 1: $Y = 1\{W\,\Delta time_t + \alpha\Delta cost_t + \varepsilon_t < 0\}$
  - Assume W is normal
  - Test with 1 SNP term: LR stat = 6.86. **REJECT**
- Model 2: same
  - Assume W is lognormal.
  - 1 SNP term: LR stat = 2.76. **ACCEPT**
  - 2 SNP terms: LR stat = 81.50. **REJECT**
  - 3 SNP terms: LR stat = 82.06. **REJECT**
  - So 1 SNP term may not be enough

Real data 2
Lognormals with 0 to 3 SNP terms



Real data 3
Go to logWTP space

- Models 1 and 2 predict that P = ½, when $\Delta time_t + \alpha \Delta cost_t$ is small
  - If this is not the case, then model is misspecified
  - Graph suggests that something like this may be going on
  - See also Fosgerau (2005: Specification of a model to measure the value of travel time savings, forthcoming TR-A, maybe title will change)

- Model 3: logWTP space
  - $Y = 1\{\log W < \log(-\Delta cost_t/\Delta time_t) + \varepsilon_t\}$.
  - Likelihood improves by more than 400!
  - 1 SNP term: LR stat = 4.06. **WEAK REJECT (96%)**
  - 2 SNP terms: LR stat = 6.42. **WEAK REJECT (96%)**
  - 3 SNP terms: LR stat = 8.50. **WEAK REJECT (97%)**

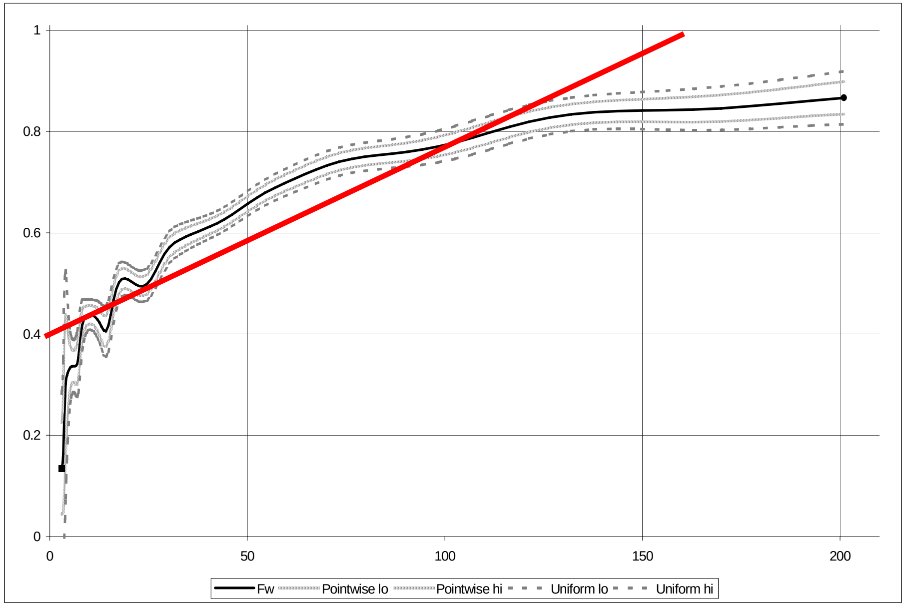
- Real size may be higher than nominal size, so we could accept normal in this case

Conclusions 1

- Representing the distribution of a random parameter in a discrete choice model is crucial. Otherwise
  - The mean (WTP) may be off
  - Predictions may be off
- Proposed a method to deal with this issue in the case of a single random parameter with discrete (panel) data
- Method is easy to work with - implemented in BioGeme
  - Simulation study shows that model works well as a test of parametric assumption
  - Can be used to yield very flexible distributions
  - Can be used to inform distributional assumptions

Conclusions 2 – a warning

- A distribution may fit without being identified
- if a large support condition is not satisfied (Fosgerau 2006 TR-B)



- So it is still necessary to check identification when e.g. the mean of the random parameter is of interest (this could be a mean WTP)

Conclusions 3 - Extensions

- Random parameters of several dimensions
  - Seems feasible, but how to avoid the curse of dimensionality?
- Consistency of the estimator is very likely
  - See Fosgerau & Nielsen for a general case with unknown error distribution
  - Proof in the binary logit case with random constant is likely (almost done)
  - Proof with more general models?
- Convergence rates and asymptotic distribution of estimates
  - Euclidian parameters: Asymptotically normal?
  - Estimate of g: How fast is convergence?

Mogens Fosgerau

mf@dtf.dk