Wildlife population assessment methods

Takis Besbeas

CAPAM 2015, La Jolla, CA

October 2015

Athens University of
Economics and Business

ncse
National Centre for Statistical Ecology

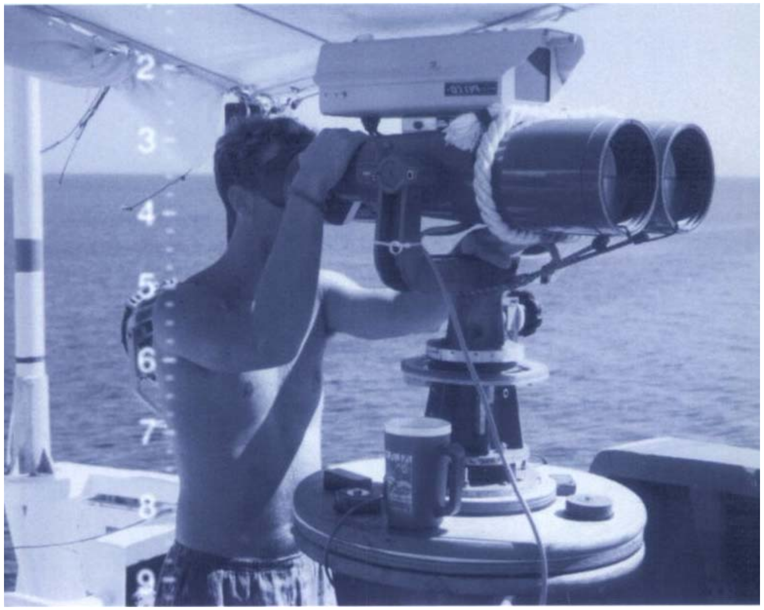
Outline

- IPM in ecology
- IPM aspects
  - Modelling data reflecting population size
  - Goodness of fit (with B.J.T. Morgan)
  - New methods of model-fitting (with R. Langrock)
- Conclusions

Motivation

- In ecology, aspects of the demography of wild animals are estimated through different studies, resulting in multiple data sets.

- Main types of study:
  - individual studies (MRR data)
  - population studies (abundance data)



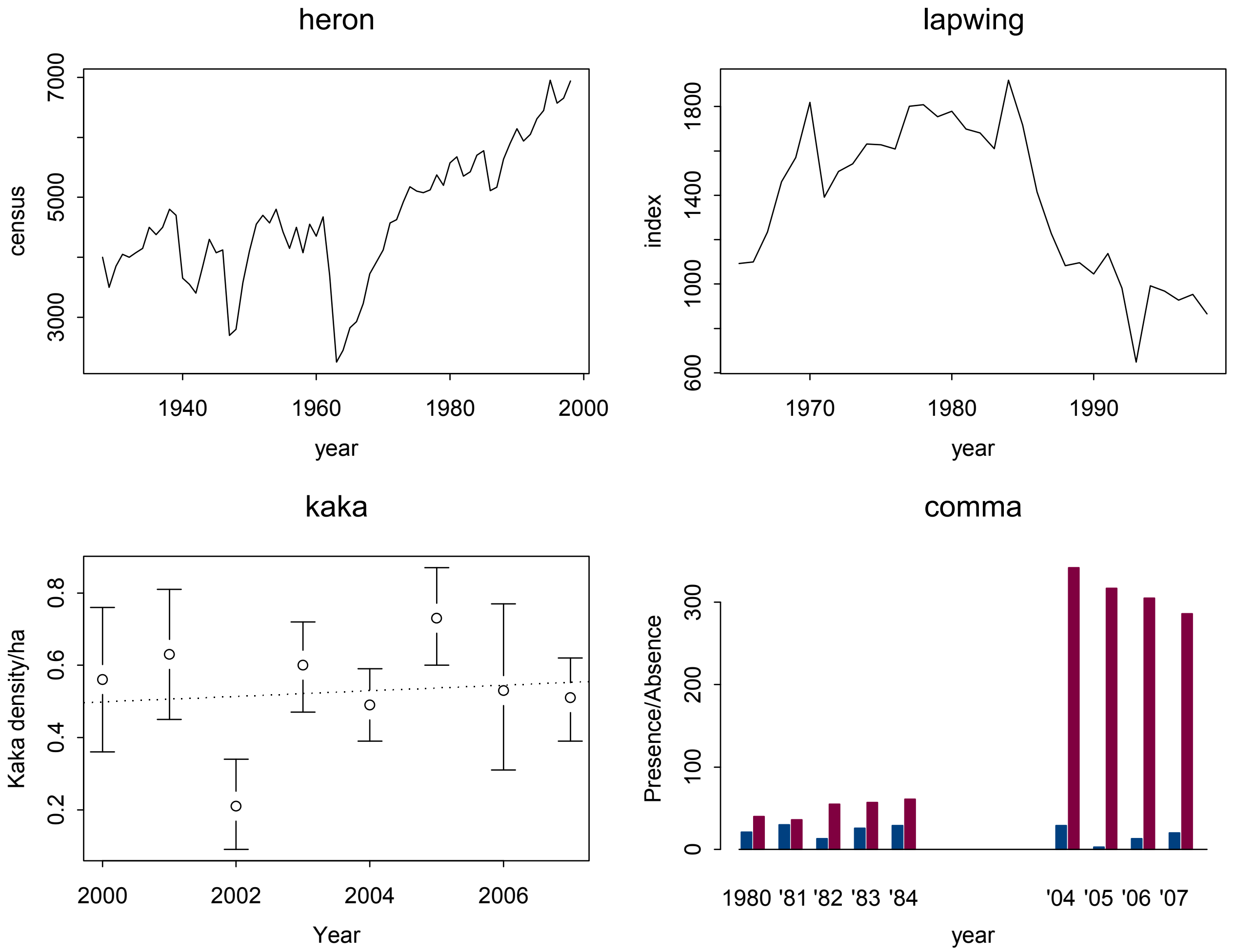
Individual studies (MRR data)

- These provide information on survival from studying previously marked animals.
- Animals may be observed again alive (recaptures) or dead (recoveries):

1963	1147	14	4	1	2	1
1964	1285		20	3	4	0
1965	1106			10	1	2
1966	1615				9	7
1967	1618					12

Population studies (abundance data)

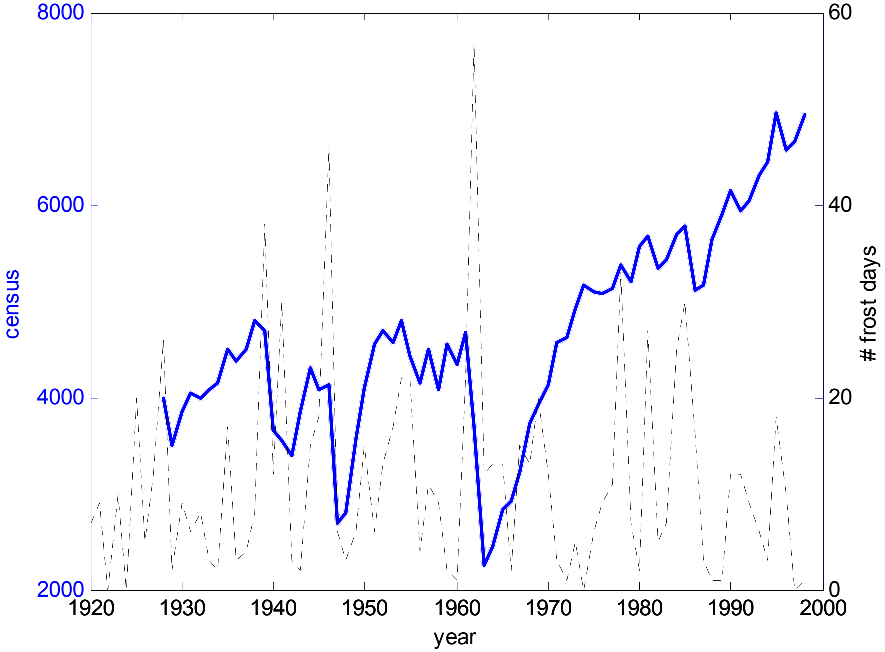
- These also arise in a variety of different forms:



Illustrating example I: British herons (*Ardea cinerea*)

- Mark-recovery and census data

Year Ringed	Number Ringed	Year Recovered					
		'56	'57	'58	…	'97	'98
1955	348	33	5	0	…	0	0
1956	141		17	6	…	0	0
1957	265			31	…	0	0
…	…				…	…	…
1996	4488					16	0
1997	4339						18



Methods and notation

- Use of capture-recapture methods or state-space models to form MRR likelihood(s) (Lebreton et al, 1992; Gimenez et al, 2007).

- Use of state-space models to form "census" likelihood:

$$N_{t+1} = T_t N_t + \eta_t$$

$$y_t = z_t N_t + \epsilon_t$$

- Use of integrated population modelling to estimate parameters.

- ϕ denotes annual survival probability,
  λ denotes reporting probability,
  p denotes productivity rate.

Relevant state-space models

$$\begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ N_{a-1} \\ N_{a+} \end{pmatrix}_{t+1} = \begin{pmatrix} 0 & p\phi_1 & \cdots & p\phi_1 & p\phi_1 \\ \phi_2 & 0 & 0 & 0 & 0 \\ & & \vdots & & \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \phi_a & \phi_a \end{pmatrix}_t \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ N_{a-1} \\ N_{a+} \end{pmatrix}_t + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{a-1} \\ \eta_{a+} \end{pmatrix}_t$$

$$y_t = \begin{pmatrix} 0 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} N_{1,t} & N_{2,t} & \cdots N_{a+,t} \end{pmatrix} + \epsilon_t$$

based on Poisson for recruitment and binomial for survival assumptions.

Model fitting method

Computational demand →

Parameter	KF (Besbeas et al 2002)		Bayes (Brooks et al 2004)		MCKL (de Valpine, 2012)	
	Est	**SE**	**Mean**	**SD**	**Est**	**SE**
ϕ_1 intercept	0.540	0.068	0.543	0.069	0.54	0.066
ϕ_1 slope	-0.196	0.059	-0.197	0.060	-0.19	0.059
ϕ_a intercept	1.542	0.070	1.550	0.071	1.54	0.069
ϕ_a slope	-0.241	0.039	-0.243	0.039	-0.24	0.038
λ intercept	-3.945	0.082	-3.910	0.087	-3.924	0.036
λ slope	-0.034	0.004	-0.034	0.004	-0.034	0.004
p intercept	-0.693	0.090	-0.668	0.095	-0644	0.087
p slope	-0.026	0.004	-0.027	0.005	-0.027	0.005
σ	159.47	21.87	169.11	25.47	159	22.70

- Minor differences but methods differ in ease of application

Model fitting method

Computational demand →

Parameter	KF (Besbeas et al 2002)		Bayes (Brooks et al 2004)		MCKL (de Valpine, 2012)	
	Est	**SE**	**Mean**	**SD**	**Est**	**SE**
ϕ_1 intercept	0.540	0.068	0.543	0.069	0.54	0.066
ϕ_1 slope	-0.196	0.059	-0.197	0.060	-0.19	0.059
ϕ_a intercept	1.542	0.070	1.550	0.071	1.54	0.069
ϕ_a slope	-0.241	0.039	-0.243	0.039	-0.24	0.038
λ intercept	-3.945	0.082	-3.910	0.087	-3.924	0.036
λ slope	-0.034	0.004	-0.034	0.004	-0.034	0.004
p intercept	-0.693	0.090	-0.668	0.095	-0644	0.087
p slope	-0.026	0.004	-0.027	0.005	-0.027	0.005
$\boldsymbol{\sigma}$	**159.47**	21.87	**169.11**	25.47	**159**	22.70

- Minor differences but methods differ in ease of application

Advantages of integration

- Simultaneous description of all the data
- Estimation of parameters not estimable from separate analyses
- Generally more precise parameter estimators
- Reduction in correlation between estimators

IPM aspects

- Assumption of independence (Besbeas et al, 2009; Abadi et al 2010; Chandler and Clark, 2014)
- Model selection (Besbeas et al, 2015)
- **Modelling data reflecting population size**
  - Pseudo replication
  - Penalised likelihood
- Goodness-of-fit
- New methods of model-fitting

Modelling data reflecting population size

- In many cases, the only type of data available are a time-series of population abundances y_t.
  - See, eg, the Global Population Dynamics Database

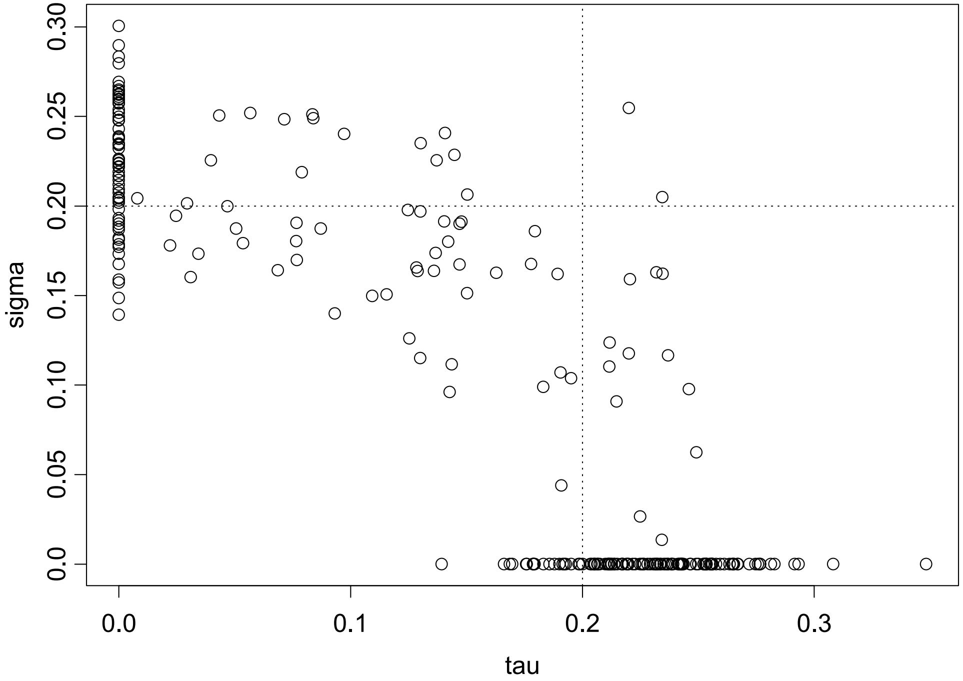
- A frequently used model for log-abundance is the Gompertz SSM

$$\begin{aligned} X_t &= a + cX_{t-1} + \eta_t, \qquad & \eta_t &\sim N(0, \tau^2) \\ Y_t &= X_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma^2) \end{aligned}$$

Modelling data reflecting population size

- However practical performance eg wrt partition variance is less than ideal:

Simulation $a=1$, $c= 0.3$, $\sigma=\tau = 0.2$, $n =25$



IPM protection

- Recall, eg for two age classes, $\boldsymbol{N}_t = (N_{1,t}, N_{2+,t})$

$$N_{1,t+1} | \boldsymbol{N}_t \sim \mathrm{Po}(p\phi_1 N_{2+,t})$$

$$N_{2+,t+1} | \boldsymbol{N}_t \sim \mathrm{bin}(N_{1,t} + N_{2+,t},\ \phi_a)$$

$$y_t | \boldsymbol{N}_t \sim \mathrm{N}(N_{2+,t}, \sigma^2)$$

- Thus process variances τ are fully specified in IPM.

IPM breakdowns: fully time-dependent parameters

- Model notation: $\phi_1, \phi_2, \phi_3, \phi_a / \lambda_t / p = \{c, t\}$

  Best Kth-order time-dependent model:

	Model	-ℓ	$\hat{\sigma}$	ΔAIC
K=0	$\mathrm{cccc}/\lambda_t/\mathrm{c}$	8942.41	651.97	347.0
1	$\mathrm{cccc}/\lambda_t/t$	8781.12	16.95	162.4
2	$\mathrm{ccct}/\lambda_t/t$	8652.30	0.00	42.8
3	$\mathrm{tcct}/\lambda_t/t$	8600.97	0.00	78.1
$\vdots$				94<value<216
5	$\mathrm{tttt}/\lambda_t/t$	8539.98	0.00	232.1
–	$\boldsymbol{\mathrm{vvvv}/\lambda_t/t}$	**8695.92**	**0.00**	**0**

- Several models, incl. best AIC, result in unrealistic findings, eg $\hat{\sigma} = 0$

IPM breakdowns: individual heterogeneity

- Several authors (King et al, 2008; Besbeas et al, 2009) have considered the use of more general process error structures as might result for example from individual heterogeneity.

- We replace the binomial number of survivors

$$N_{t+1} \sim \mathrm{bin}(n, \phi)$$

with beta-binomial

$$N_{t+1} \sim \mathrm{betabin}(\mathrm{mean} = n\mu, \mathrm{var} = n\mu(1-\mu)\{1 + \frac{n-1}{1+\theta}\theta\})$$

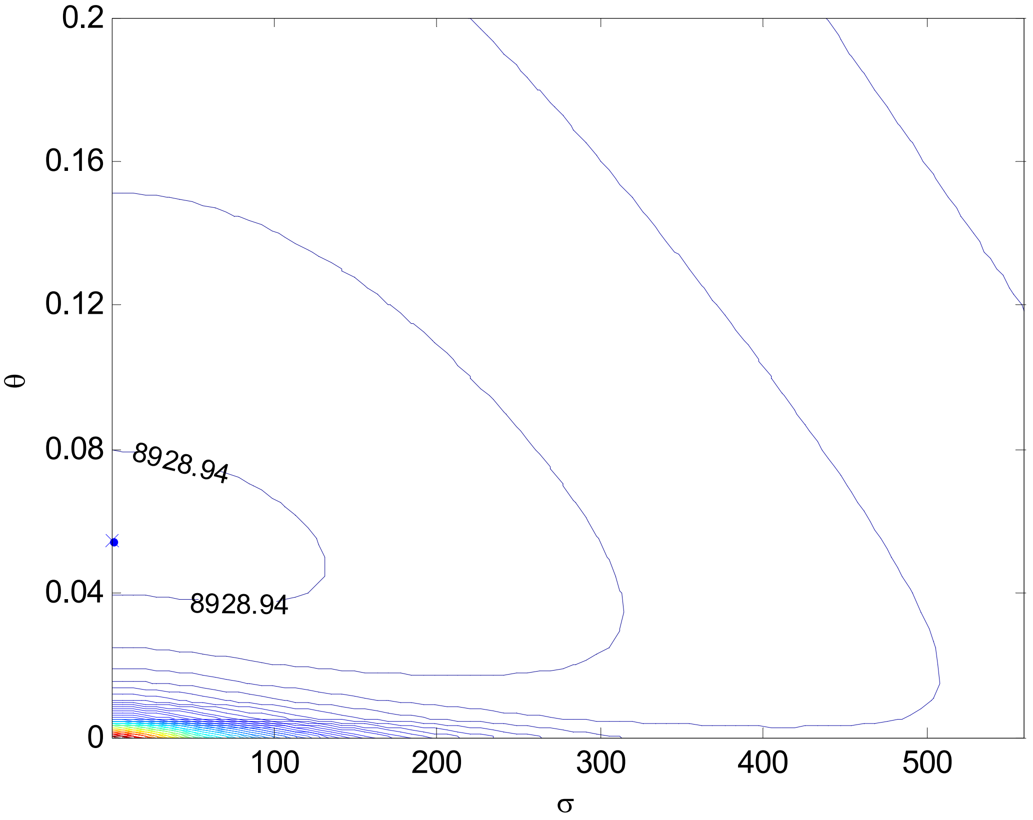
where $\theta \geq 0$ is an unknown overdispersion parameter

- Similarly for recruitment (negative binomial)

$$N_0 \sim \mathrm{NB}(\mathrm{mean} = pn, \mathrm{variance} = pn(1 + \theta pn))$$

- For illustration we concentrate on adult survival ϕ_a

Profile log-likelihoods from selected model

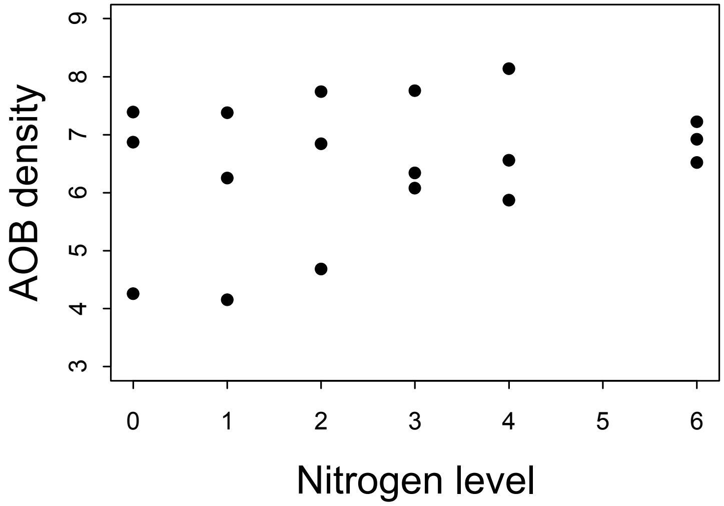


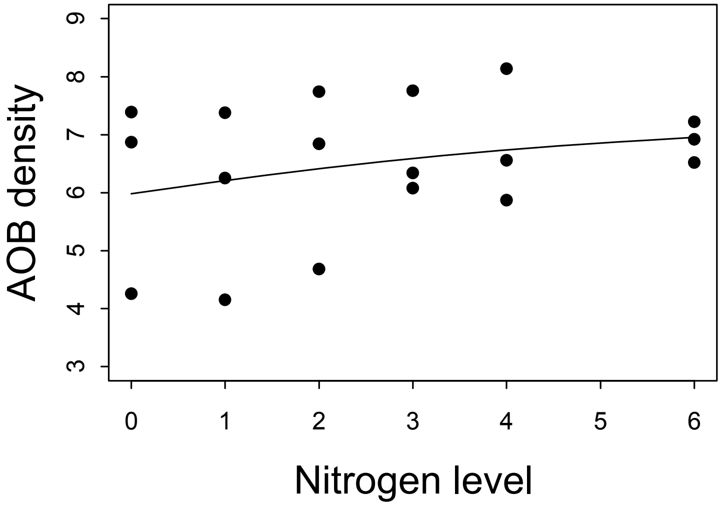
Two solutions: a) Pseudo replication

- In non-IPM, various authors (Dennis et al, 2010; Knape et al, 2012) have proposed the use of replicated sampling as a means to improve estimation:

$$y_{t,k} = X_t + \epsilon_{t,k}, \quad k = 1, \cdots, K$$

- Data from Robert and Casella (2010):

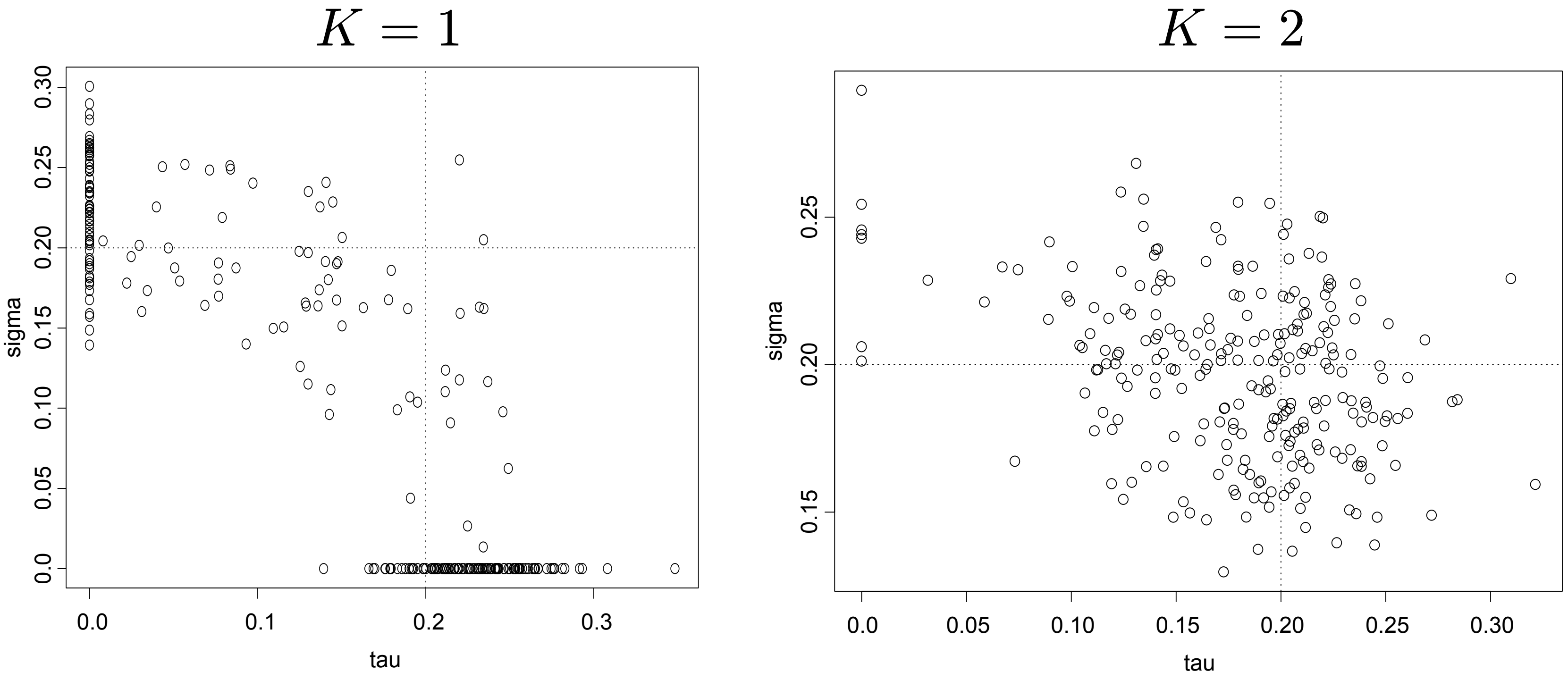




Replicated sampling

- Practical performance is vastly improved, even when K=2.

$$y_{t,k} = X_t + \epsilon_{t,k}, \quad k = 1, \cdots, K$$



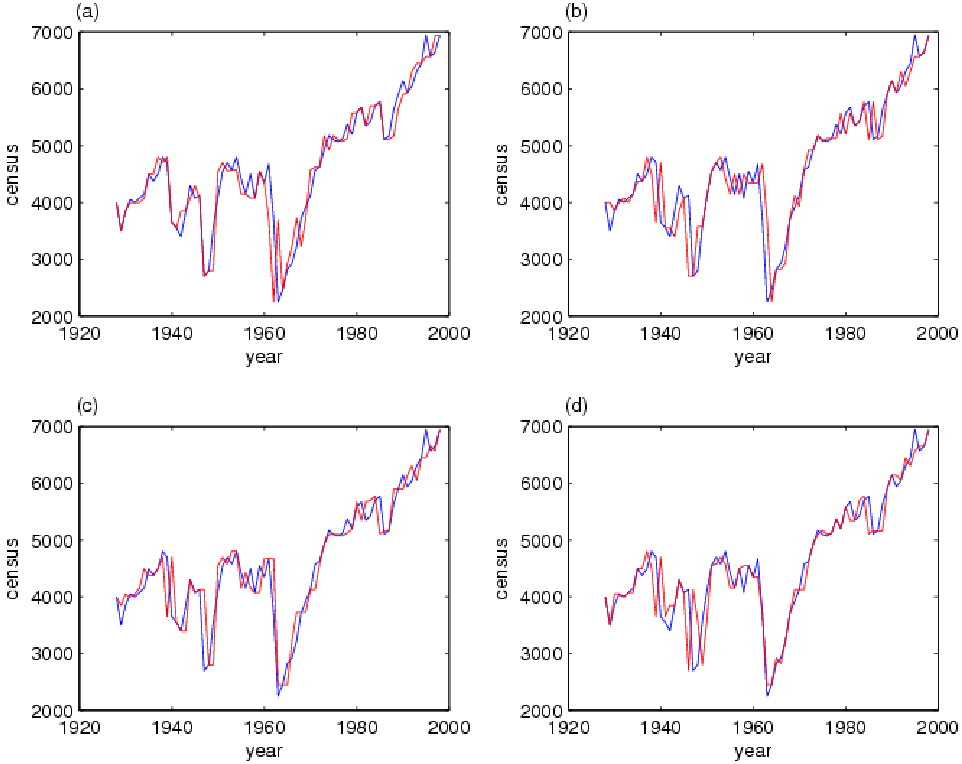
$K = 1$

$K = 2$

Four pseudo replicates

- In IPM, Besbeas and Morgan (2015) propose pseudo replication if no replicates are available:

$$\tilde{y}_t = \mathrm{random}(y_{t-1}, y_t, y_{t+1})$$



- Several variants, incl. suggestions by referee.

Parameter estimate comparison

No replication			Replication with 3 values		Replication with 5 values	
parameter	**Estimate**	**se**	**Av. Estimate**	**Av. se**	**Av. Estimate**	**Av. se**
1	-0.171	0.049	-0.174	0.049	-0.174	0.049
2	-0.021	0.004	-0.023	0.004	-0.022	0.005
3	0.411	0.074	0.410	0.074	0.409	0.075
4	-0.022	0.006	-0.018	0.006	-0.016	0.006
5	0.931	0.105	0.924	0.106	0.924	0.107
6	-0.023	0.009	-0.019	0.009	-0.017	0.009
7	1.325	0.100	1.341	0.100	1.340	0.102
8	-0.017	0.005	-0.013	0.005	-0.009	0.005
$\boldsymbol{\theta}$	0.053	0.009	0.043	0.009	0.042	0.011
10	-2.019	0.026	-2.021	0.026	-2.021	0.026
11	-0.831	0.046	-0.833	0.046	-0.835	0.046
12	-0.094	0.092	-0.110	0.093	-0.118	0.095
$\boldsymbol{\sigma}$	0.000	–	213.22	16.42	313.39	23.30

Two solutions: b) Penalised likelihood

- An alternative approach is to adopt a penalised likelihood approach (Wang & Lindsay, 2005) to bound σ away from 0.
- Thus

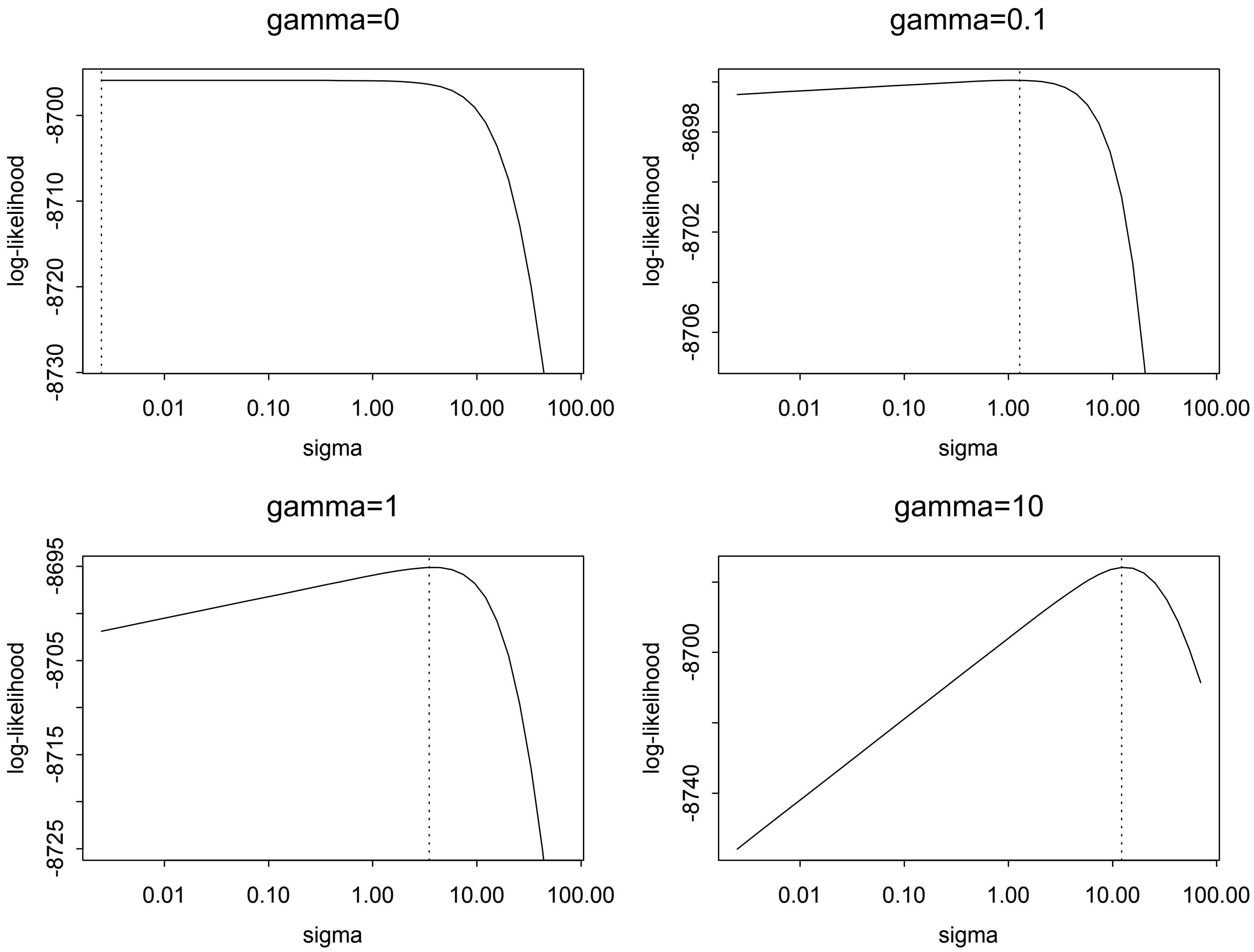
$$\log L_j^{\gamma} = \log L_j - \gamma h(\sigma)$$

where $\gamma > 0$ and $h(\sigma)$ becomes large when σ is small.

- We adopt $h(\sigma) = -\log\sigma$ and we observe performance by application to the heron data and simulation.

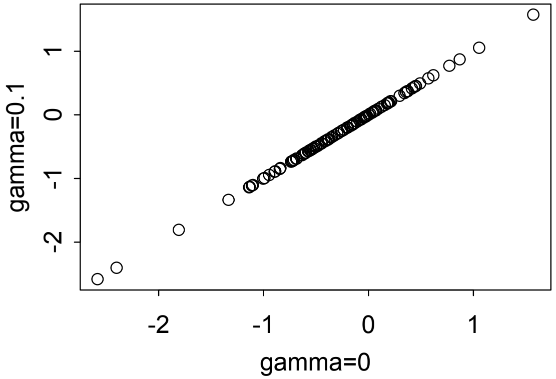
Profile log-likelihoods

- Model $\mathrm{vvvv}/\ \lambda_t/\mathrm{t}$

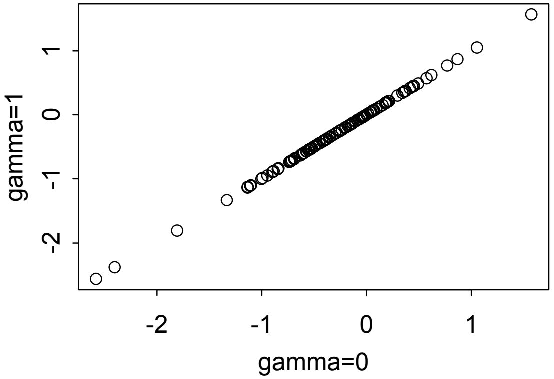


(Penalised) MLEs

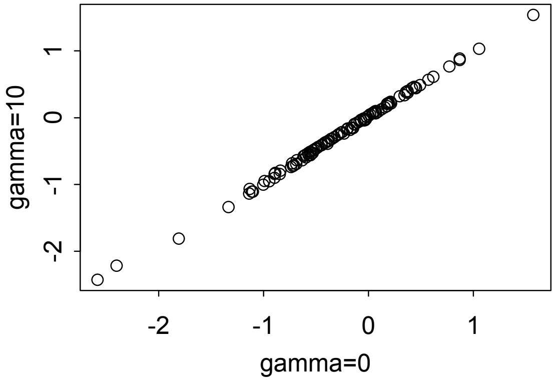
$\gamma = 0.1$ vs $\gamma = 0$



$\gamma = 1$ vs $\gamma = 0$



$\gamma = 10$ vs $\gamma = 0$

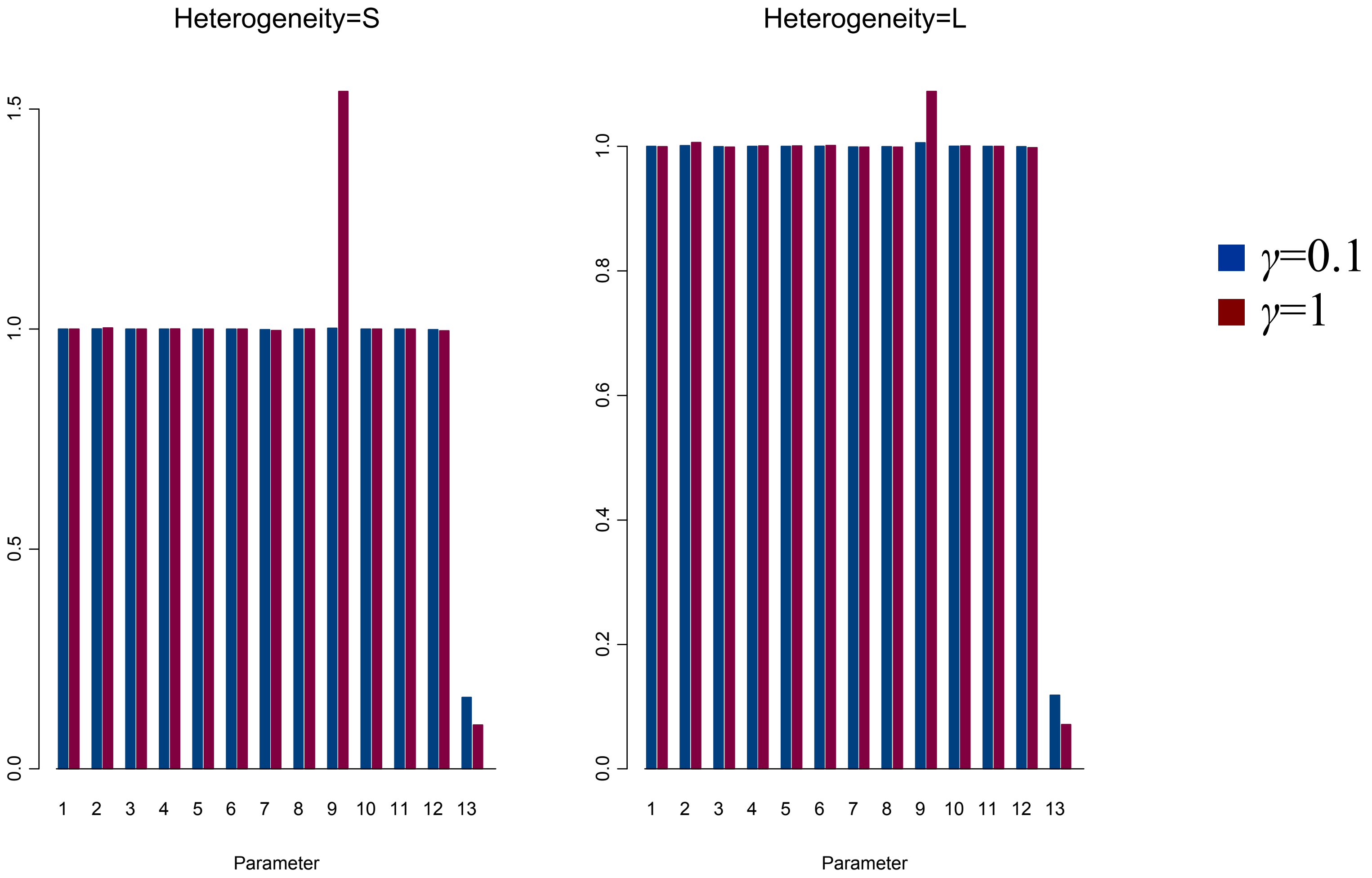


Simulation

- Three models (simple to complex)
- Four levels of heterogeneity (S/M/L/XL)
- Observed data sample size
- 500 replications

RMSE ratios: RMSE($\gamma=i$)/RMSE($\gamma=0$)

- Complex model

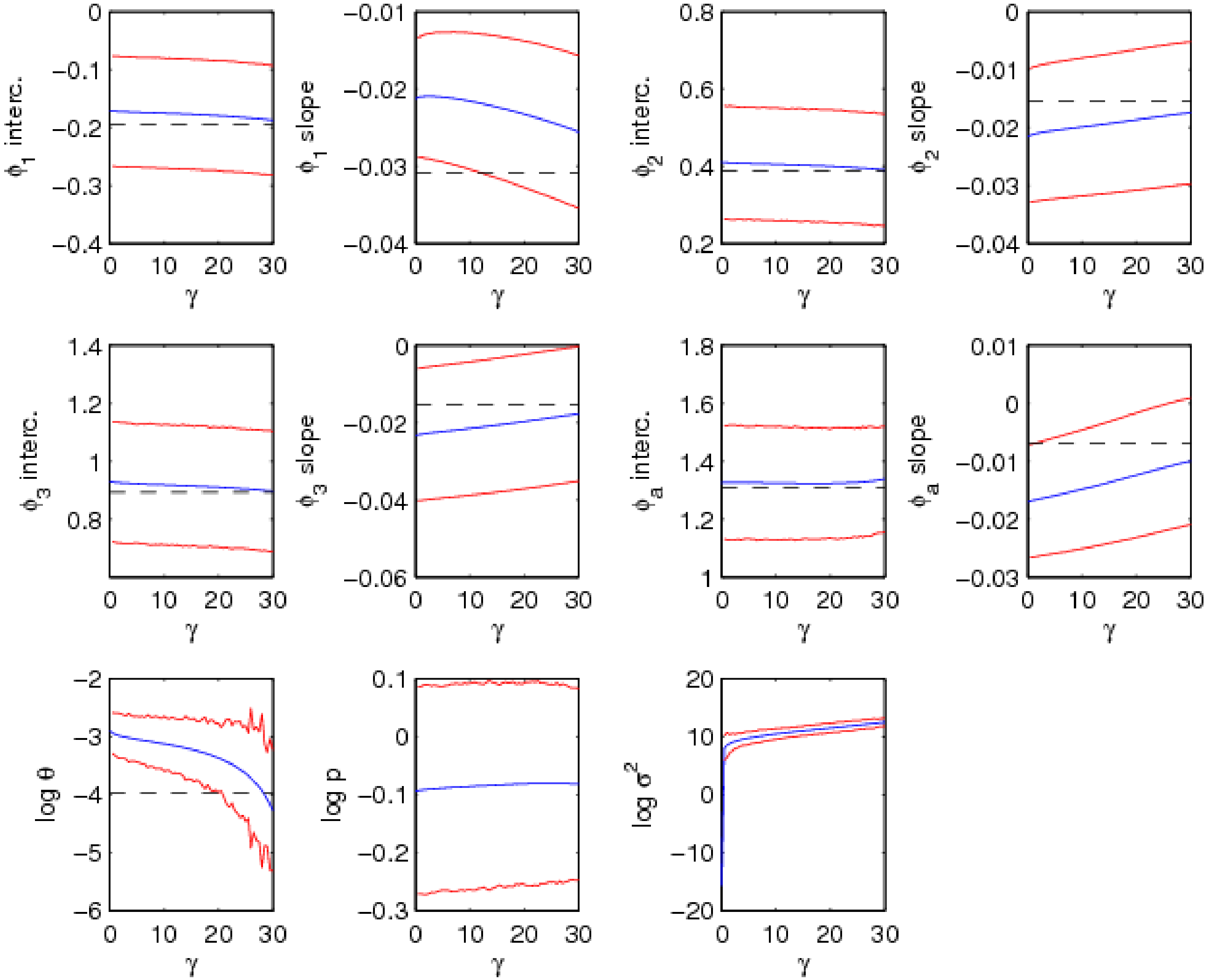


Data weighting and parameter stability

- Recall that observation std σ has multiple roles:
  1) Partitions variance in SSM
  2) Weights influence of different data sets
  3) Controls model-selection (anonymous referee)

- Penalising σ away from zero offers a natural way to investigate influence of population data and explore sensitivity of model results to different data weightings.

95% CIs for parameters as population data are downweighted

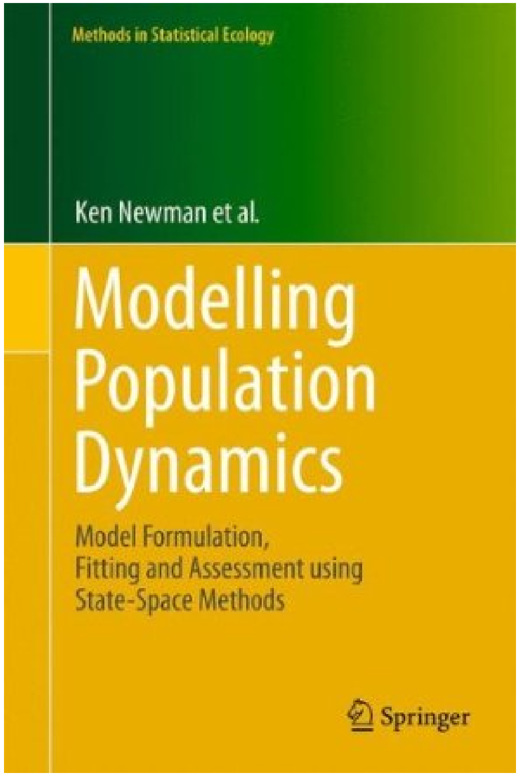


IPM aspects

- Assumption of independence
- Model selection
- Modelling data reflecting population size
  - Pseudo replicate
  - Penalised likelihood
- **Goodness-of-fit**
- New methods of model-fitting

Motivation

- State-space methods are receiving increasing attention in ecology.
- *"Constructing diagnostics for judging the goodness of fit of state-space models to data is complex"* (Newman et al, 2014, p117).



Methods in Statistical Ecology

Ken Newman et al.

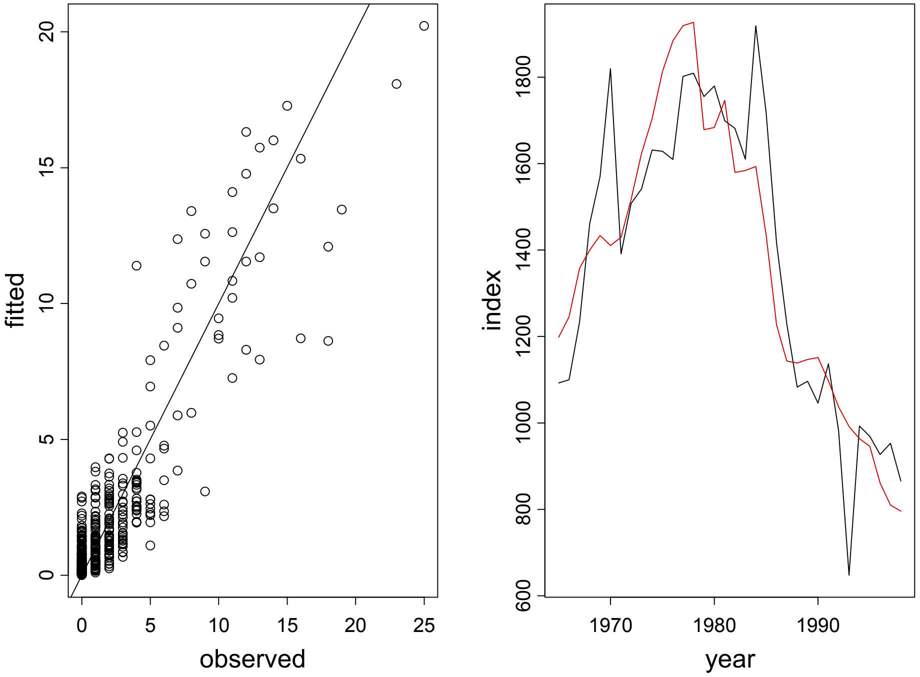
Modelling Population Dynamics

Model Formulation, Fitting and Assessment using State-Space Methods

Springer

IPM Goodness-of-fit

- Component-wise but informal:
  - eg comparing observed vs expected m-arrays for MRR components.
  - comparing observed vs fitted trends for census components.



- Bootstrap gof options available but not widely employed.

A new goodness-of-fit procedure

Besbeas and Morgan (2014) propose the use of calibrated simulation to assess component gof:

- Simulate $\theta_i \sim N(\hat{\theta}, \Sigma),\ i = 1, \cdots, s$
- For each θ_i
  - Work out fitted values $\hat{\boldsymbol{x}}_i$ and simulate new data $\boldsymbol{x}_i \sim f(\cdot|\theta_i)$ for each component model
  - Plot $D(\boldsymbol{x}; \hat{\boldsymbol{x}}_i)$ vs $D(\boldsymbol{x}_i; \hat{\boldsymbol{x}}_i)$ for some measure of discrepancy D.
- If the model fits the data well, one would expect approximately half of the points to lie above the diagonal (Johnson, 2004; Fletcher, 2009).
- Note:
  - Model is fitted only once.
  - The approach is akin to Bayesian p-values.

Choice of discrepancy measure

- There is complete freedom in the choice of distance measure
  - cf Millar and Meyer, 2000: *p*-value=0.69, 0.27, 0.50 and 0.42.
- MRR / GLM
  - Freeman-Tukey: $D(\boldsymbol{x};\, \boldsymbol{e}) = \sum_i \left(\sqrt{x_i} - \sqrt{e_i}\right)^2$
- Time series $\boldsymbol{x}_t$
  - MAPE

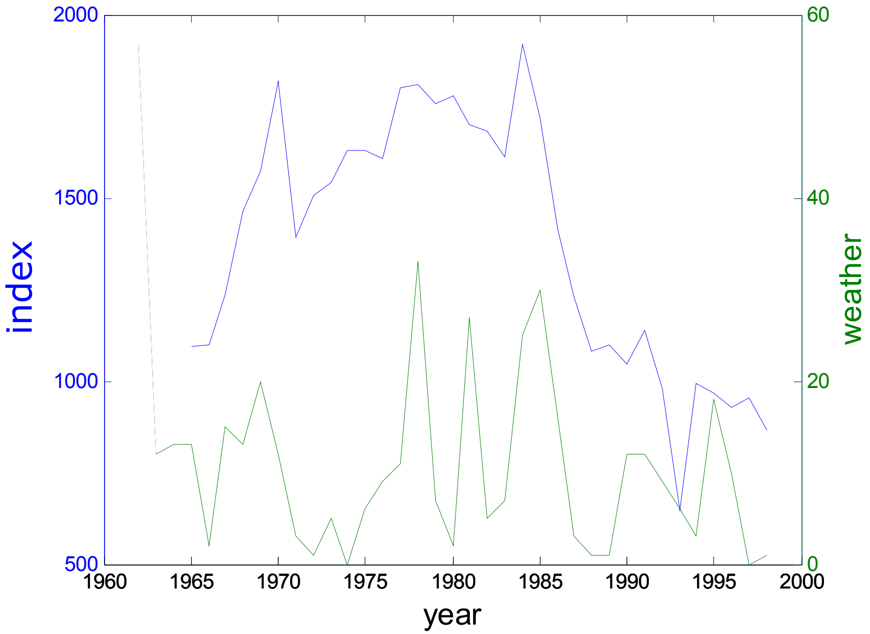
    $$D(\boldsymbol{x}; \hat{\boldsymbol{x}}) = \frac{100}{n} \sum_t |(x_t - \hat{x}_t)/x_t|$$
  - MPE

    $$D(\boldsymbol{x}; \hat{\boldsymbol{x}}) = 100 \max\{(x_t - \hat{x}_t)/x_t\}$$

Illustrating example II: lapwings (Vanellus vanellus)

- Mark-recovery and index of abundance data

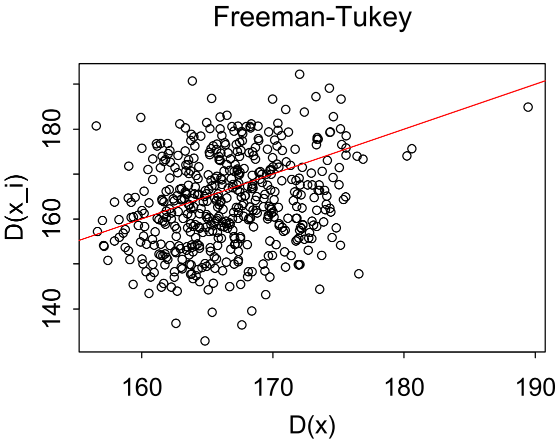
Year Ringed	Number Ringed	Year Recovered					
		‘56	‘57	‘58	…	‘97	‘98
1955	348	33	5	0	…	0	0
1956	141		17	6	…	0	0
1957	265			31	…	0	0
…	…				…	…	…
1996	4488					16	0
1997	4339						18

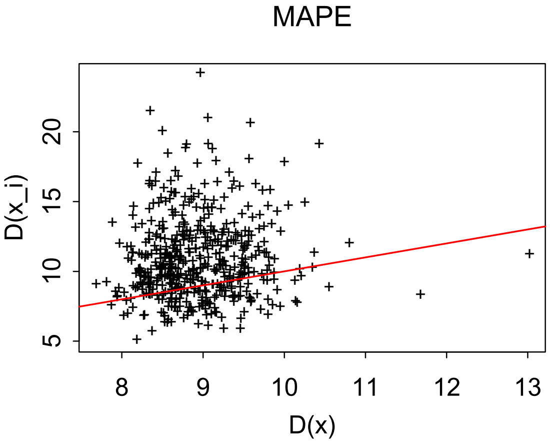


Motivating example: IPM

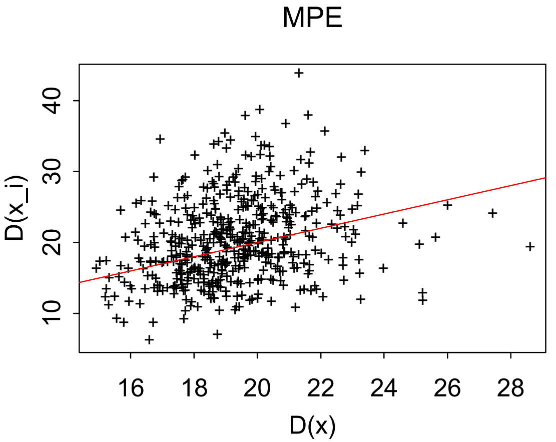
- Model $\phi_1(w_t), \phi_a(w_t)/\lambda(year)/p(year)$, following model-selection:

$p_c = 0.39$





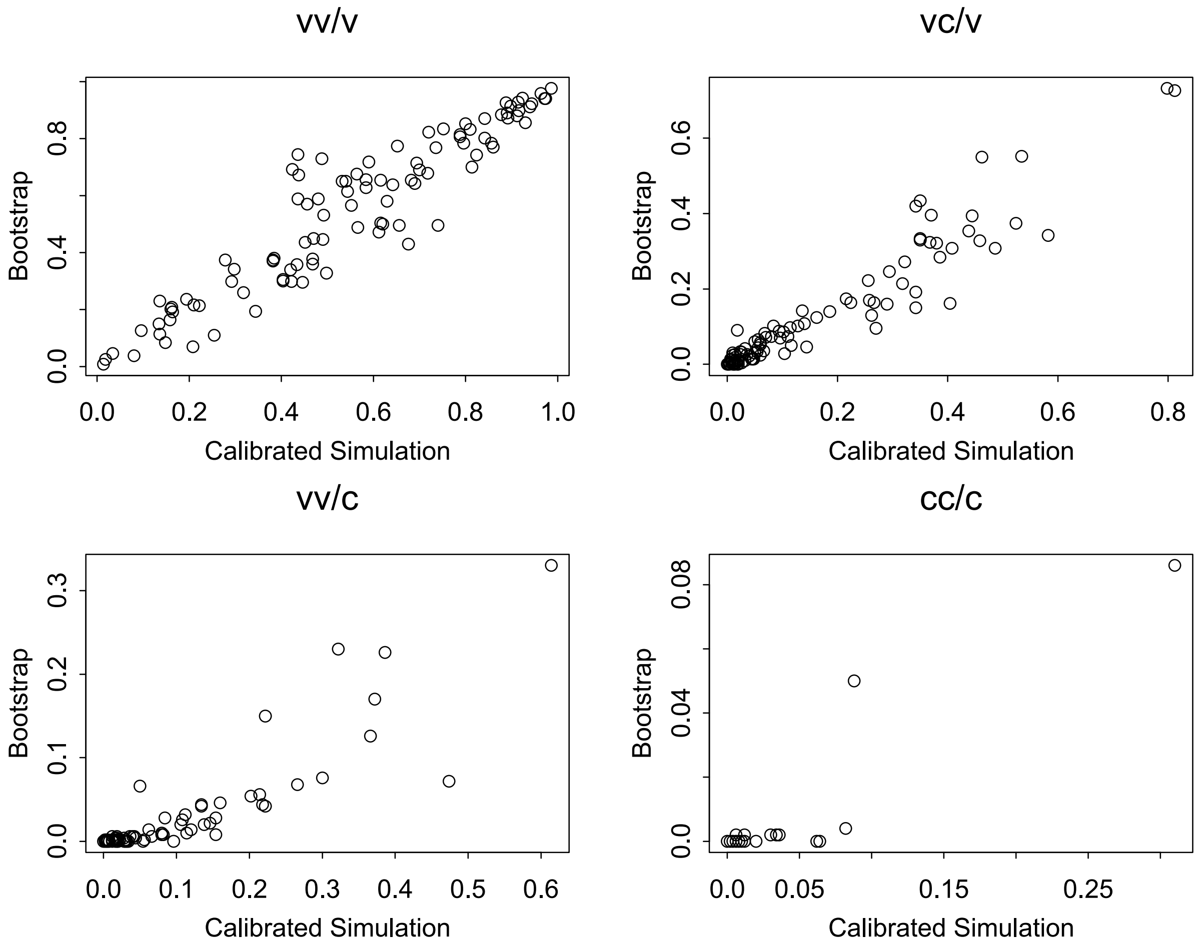
$p_c = 0.72$



$p_c = 0.52$

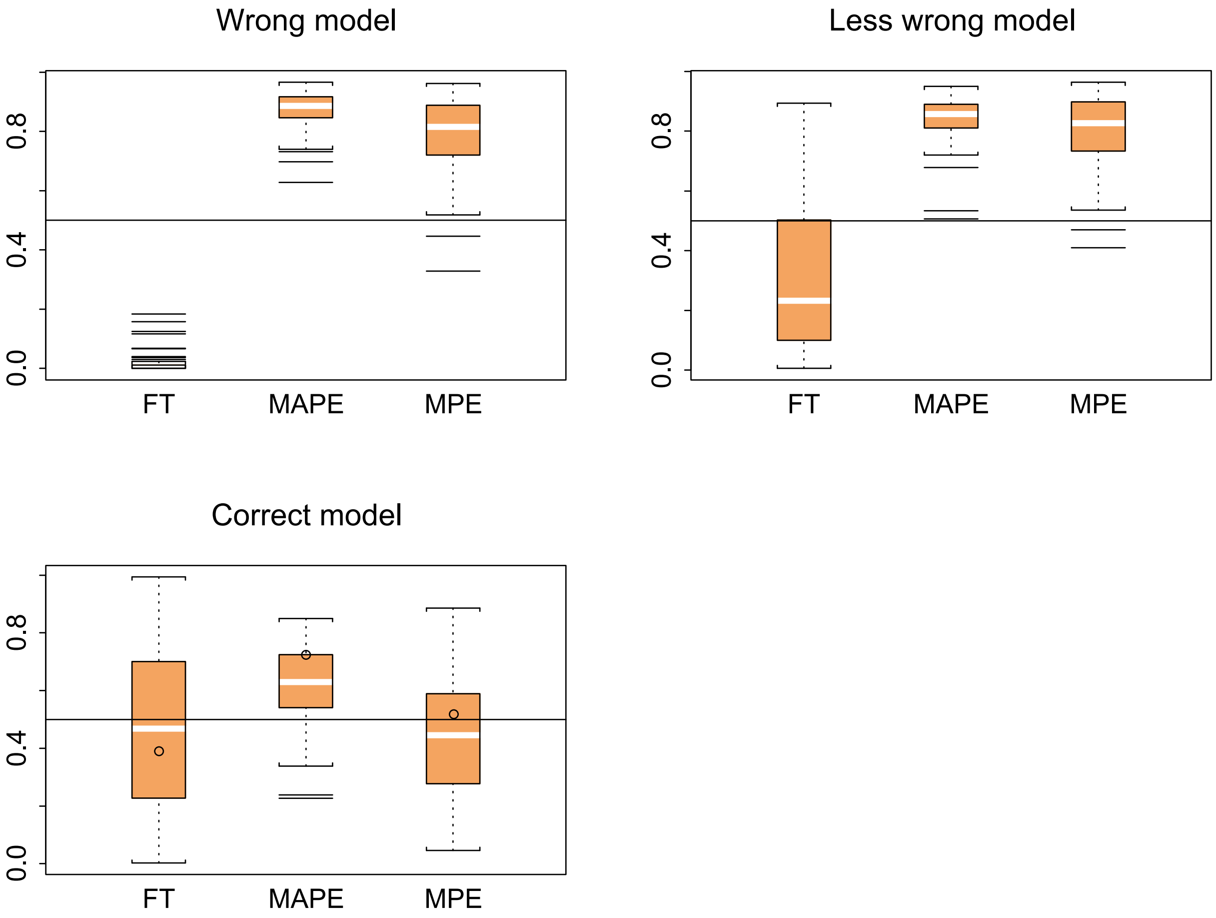
Simulation comparison with bootstrap: MRR component

- Calibrated simulation vs bootstrap p-values when true model (vv/v) is fitted by:



Calibration of test -statistic

- 100 simulations from selected model, fitted 3-ways



Wrong model

Less wrong model

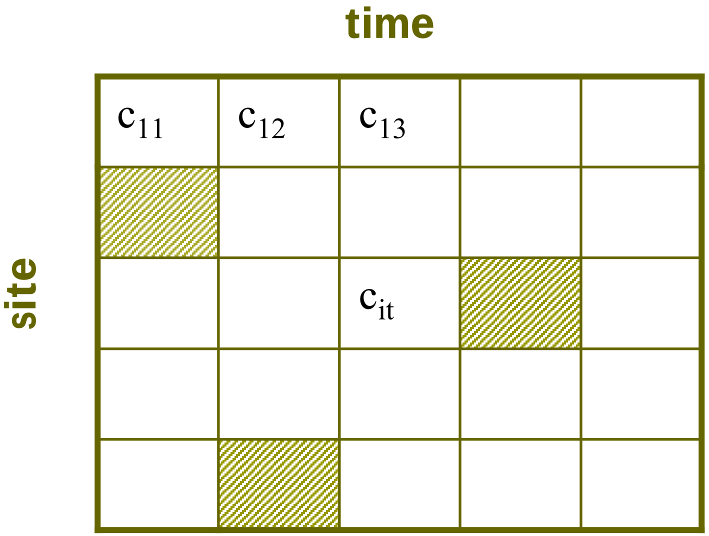
Correct model

Modelling site by year survey data

- In practice, census data are rare.
- Often we collect spatio-temporal observations, c_{it} (eg BBS, CBC).
- We typically model the c_{it} using a Poisson-based linear model:

$$c_{it} \sim Po(\mu_{it})$$

$$\log(\mu_{it}) = s_i + u_t$$



c_{11}	c_{12}	c_{13}		
▨				
		c_{it}	▨	
	▨			

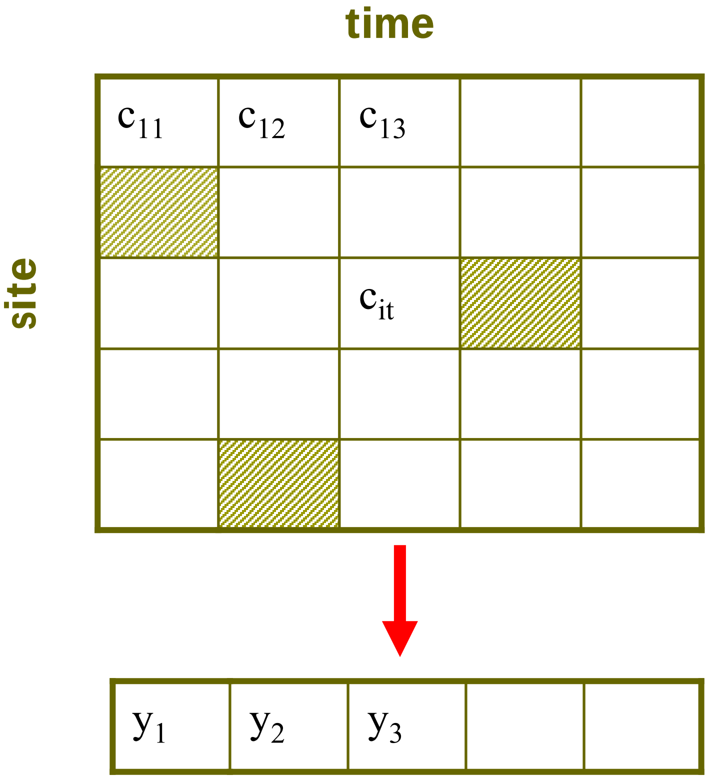
Modelling site by year survey data

- In practice, census data are rare.
- Often we collect spatio-temporal observations, c_{it} (eg BBS, CBC).
- We typically model the c_{it} using a Poisson-based linear model:

$$c_{it} \sim Po(\mu_{it})$$

$$\log(\mu_{it}) = s_i + u_t$$

- “Census” likelihood can be formed in two or one stages
  - Based on a derived index of abundance, $y_t \propto e^{u_t}$
  - Building a dynamic model directly on the u_t

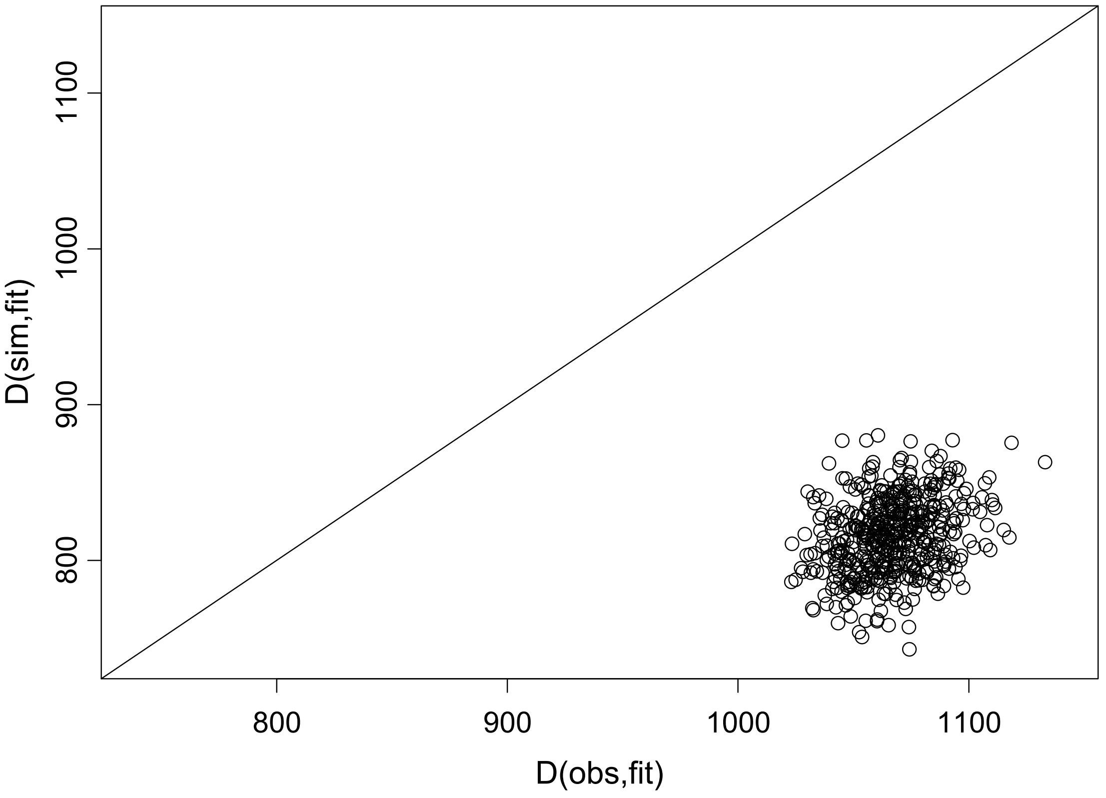


Pros and Cons of the two approaches

- Details in
  - Maunder (2001) — fisheries
  - Besbeas and Freeman (2006) — ecology
- Similar performance trading off speed vs efficiency
- However one-stage method offers greater flexibility not just for modelling counts but for incorporating data of an entirely different nature.
  - For example presence/absence data (Freeman and Besbeas, 2012).

Illustrating example II revisited

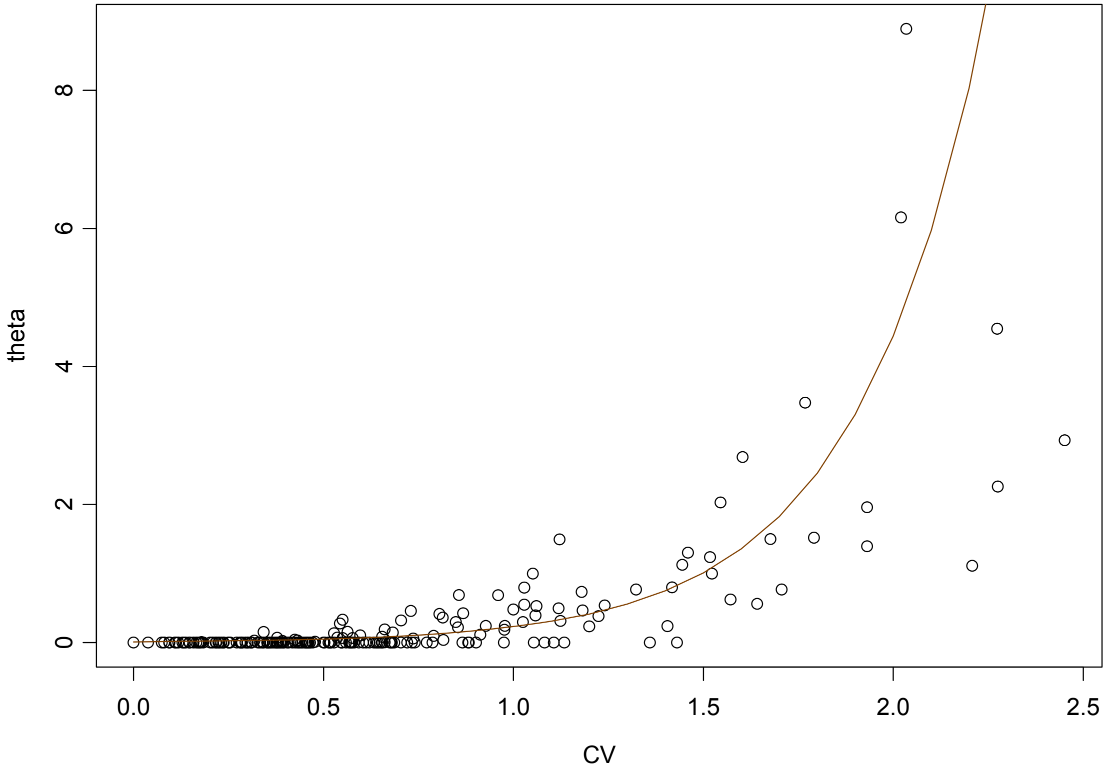
- Simulated calibration gof for Poisson distribution of c_{it}



Site by year survey data: model selection

Model	-ℓ	np	ΔAIC
$\text{Poisson}(\mu_{it})$	3629.8	266	246.8
$\text{ZIP}(\mu_{it}, \pi)$	3629.8	267	248.8
$\text{Negbin}(\mu_{it}, \theta)$	3594.0	267	177.2
$\text{Poisson-Tweedie}(\mu_{it}, D, a)$	3589.3	257	167.8
$\text{ZIP}(\mu_{it}, \pi_i)$	3520.4	497	490.0
$\text{Hurdle Po}(\mu_{it}, \pi_i)$	3478.4	497	406.0
$\text{Negbin}(\mu_{it}, \theta_i)$	3426.3	497	301.8
$\text{Negbin}(\mu_{it}, \theta(\text{CV}))$	3504.4	268	0

Individual site θ_t vs $\theta(CV)$



- We can then fit the (integrated) model exactly as we do for Poisson data.

From deterministic to stochastic transitions

- Recall

$$c_{it} \sim Po(\mu_{it})$$

$$\log(\mu_{it}) = s_i + u_t$$

$$e^{u_t} = (\phi_a e^{u_{t-1}} + p\phi_1\phi_a e^{u_{t-2}})$$

- We introduce lognormal “process error”

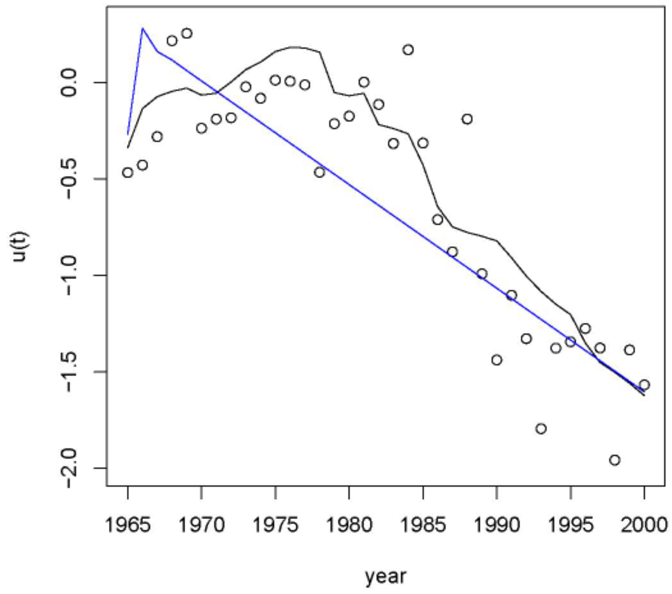
$$e^{u_t} = (\phi_a e^{u_{t-1}} + p\phi_1\phi_a e^{u_{t-2}})\varepsilon$$

and adopt a Bayesian approach.

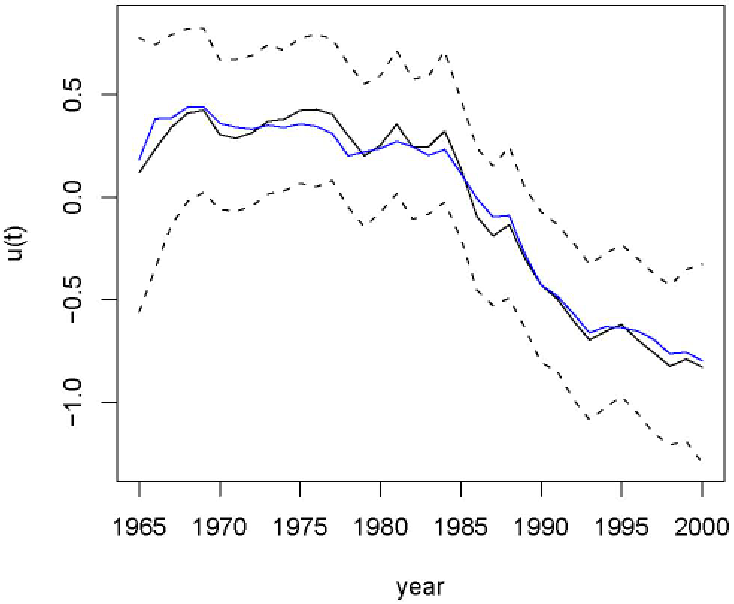
Deterministic vs stochastic transitions

- Comparison of performance between “best” and “worst” (constant) models from analysis of index:

Deterministic



Stochastic



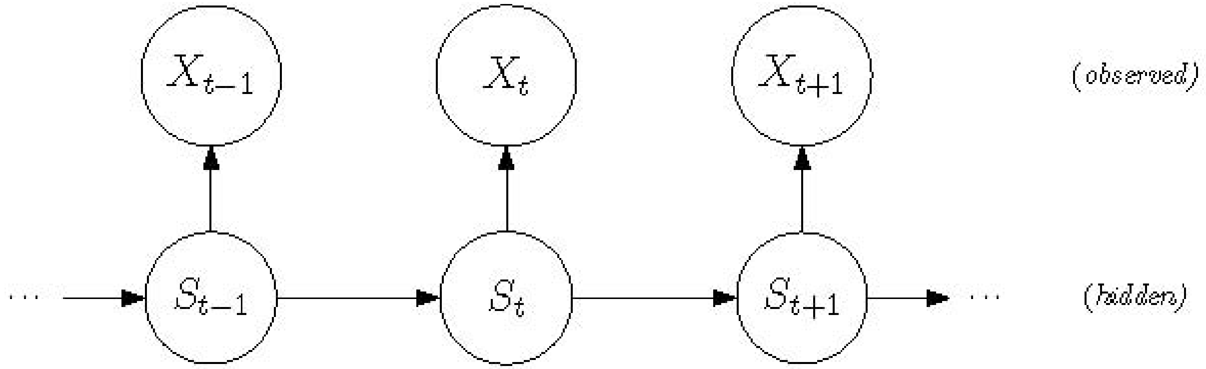
> *«Every time I think I know what’s going on, suddenly there’s another layer of complications. I just want this damn thing solved.»*
>
> *John Scalzi*

IPM aspects

- Assumption of independence
- Model selection
- Modelling data reflecting population size
  - Single replicate
  - Multiple replicate
- Goodness-of-fit
- **New methods of model-fitting**

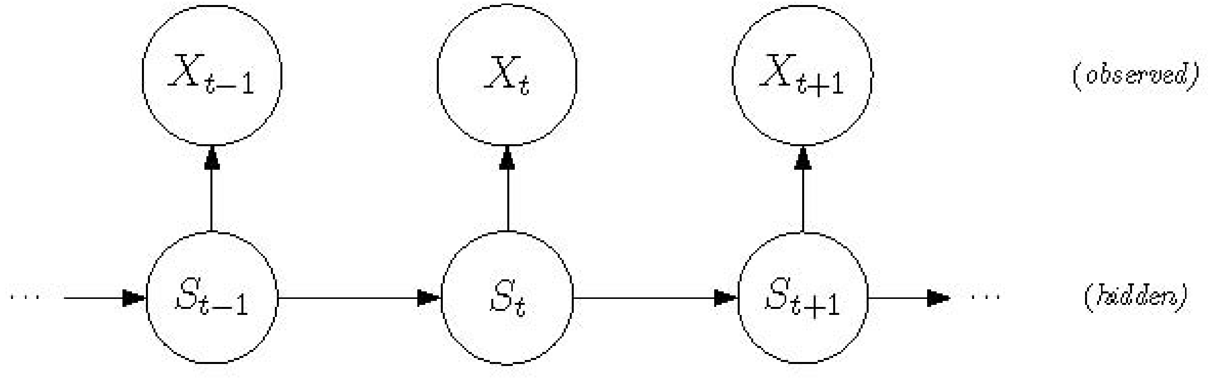
Hidden Markov Models - overview

- Flexible and mathematically tractable time series model.
- Two discrete-time stochastic processes: *observed* and *hidden.*
- Hidden process is an N-state Markov chain.
- Observed process determined by underlying hidden state.



- Applications, inter alia, in speech recognition, biology, medicine, psychology, economics and finance.

SSM estimation via hidden Markov models



- HMMs have the same two-process structure as SSMs.
- In SSMs: S_t continuous-valued.
- Discretizing S_t yields approximation by HMM.
- Benefit: HMM methodology becomes applicable.

Illustration of an HMM-based population model

- N_t: true (unknown) number of individuals at time t (the **states**).
- $N_t = N_{1t} + N_{at}$, where $N_{1t} \sim Po(pN_t)$ and $N_{at} \sim bin(N_t, \phi)$.
- Specifying some upper bound for N_t , this is a Markov chain (with a lot of states, but only two parameters).
- Observations, conditional on states: $X_t|N_t = j \sim bin(j, r_t)$
- HMM likelihood structure as usual, using a forward algorithm
- If the upper bound for N_t is in the thousands, computer memory can be an issue
  - consider coarser state-space, e.g. [0, 10), [10, 20), …,[4990, 5000) instead of 0, 4999.
- Integrated Population model can be formulated and fitted as usual.

Reference example: lapwings

- Model $\phi_1, \phi_a/\lambda/p$ (Homogeneous Markov chain)

Parameter	**HMM**					**KF**
	w=100	**w=40**	**w=20**	**w=10**	**w=8**	
φ_1	0.427	0.419	0.418(0.066)	0.418	0.418	0.415(0.066)
φ_a	1.448	1.401	1.397(0.063)	1.396	1.396	1.374(0.062)
λ intercept	-4.678	-4.680	-4.681(0.033)	-4.681	-4.681	-4.683(0.032)
p intercept	-0.978	-0.885	-0.879(0.077)	-0.875	-0.874	-0.942(0.079)
$\log \sigma^2$	10.977	11.693	11.764(0.277)	11.804	11.824	11.186(0.273)

Reference example: lapwings

- Model $\phi_1, \phi_a/\lambda/p$ (Homogeneous Markov chain)

Parameter	**HMM**					**KF**
	w=100	**w=40**	**w=20**	**w=10**	**w=8**	
φ_1	0.427	0.419	0.418(0.066)	0.418	0.418	0.415(0.066)
φ_a	1.448	1.401	1.397(0.063)	1.396	1.396	1.374(0.062)
λ intercept	-4.678	-4.680	-4.681(0.033)	-4.681	-4.681	-4.683(0.032)
p intercept	-0.978	-0.885	-0.879(0.077)	-0.875	-0.874	-0.942(0.079)
$\log \sigma^2$	10.977	11.693	11.764(0.277)	11.804	11.824	11.186(0.273)

Max % RAM	0.3	0.5	3.7	34.1	61.2

Conclusions

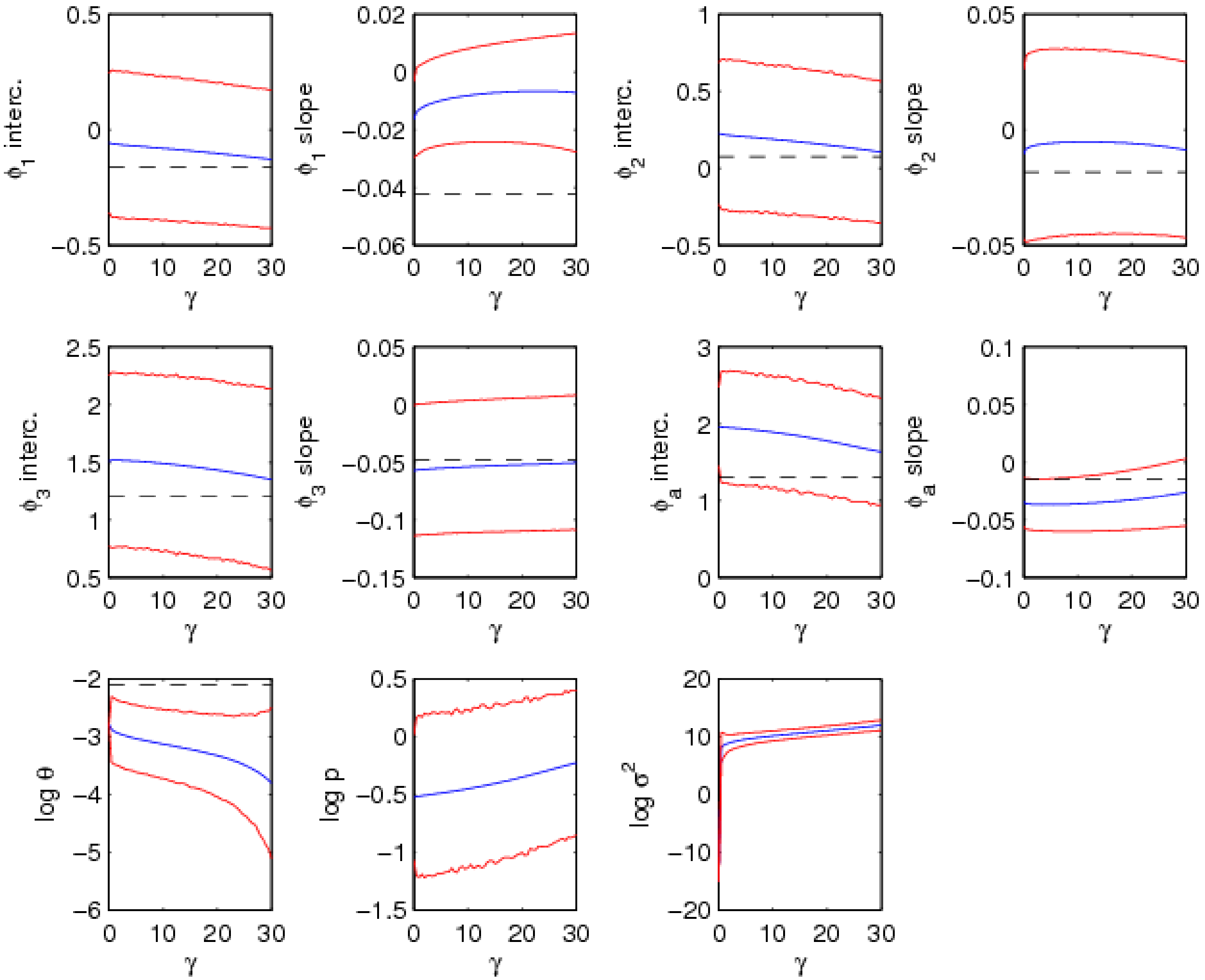
- Modelling data reflecting population size
  - Replication has advantages and permits more sophisticated models to be fitted.
  - Penalising σ natural way to investigate sensitivity of model results to different data weightings.
- Goodness-of-fit
  - For integrated models, gof preferably component-wise
  - Different distance measures highlight different aspects of fit.
  - Calibrated simulation appears to have reasonable potential.
- Model-fitting
  - HMMs provide a promising alternative.
  - Computer memory can be an issue but more development to follow.

References

- Abadi, Gimenez, Altettaz, Schaub (2010). An assessment of integrated population models. *Ecology*, 91: 7-14.
- Besbeas, Freeman, Morgan, Catchpole (2002). Integrating mark-recapture-recovery and census data. *Biometrics*, 58: 540-547.
- Besbeas, Morgan (2014). Goodness of fit of integrated population models. *Methods Ecol. Eval.*, 5: 1373-1382.
- Besbeas, Morgan (2015). Pseudo replication for integrated population models. In revision.
- Brooks, King, Morgan (2004). A Bayesian approach to combining abundance and demographic data. *Animal Biod Cons*, 27: 515-529.
- Dennis, Ponciano, Taper (2010). Replicated Sampling increases efficiency in population monitoring. *Ecology*, 91: 610-620.
- de Valpine (2012). Frequentist analysis of hierarchical models for population dynamics. *J. Ornith*. 152: 393-408.
- Gimenez, Rossi, Choquet, Dehais, Doris, Varella, Vila, Pradel (2007) State-space modelling of data on marked individuals. *Ecol Modell*, 206: 431-438.
- Maunder (2001). A general framework for integrating the standardization of CPUE into stock assessment models. *Can. J. Fish. Aquat. Sci* 58: 795-803.
- Newman et al (2014). *Modelling Population Dynamics*. Springer.

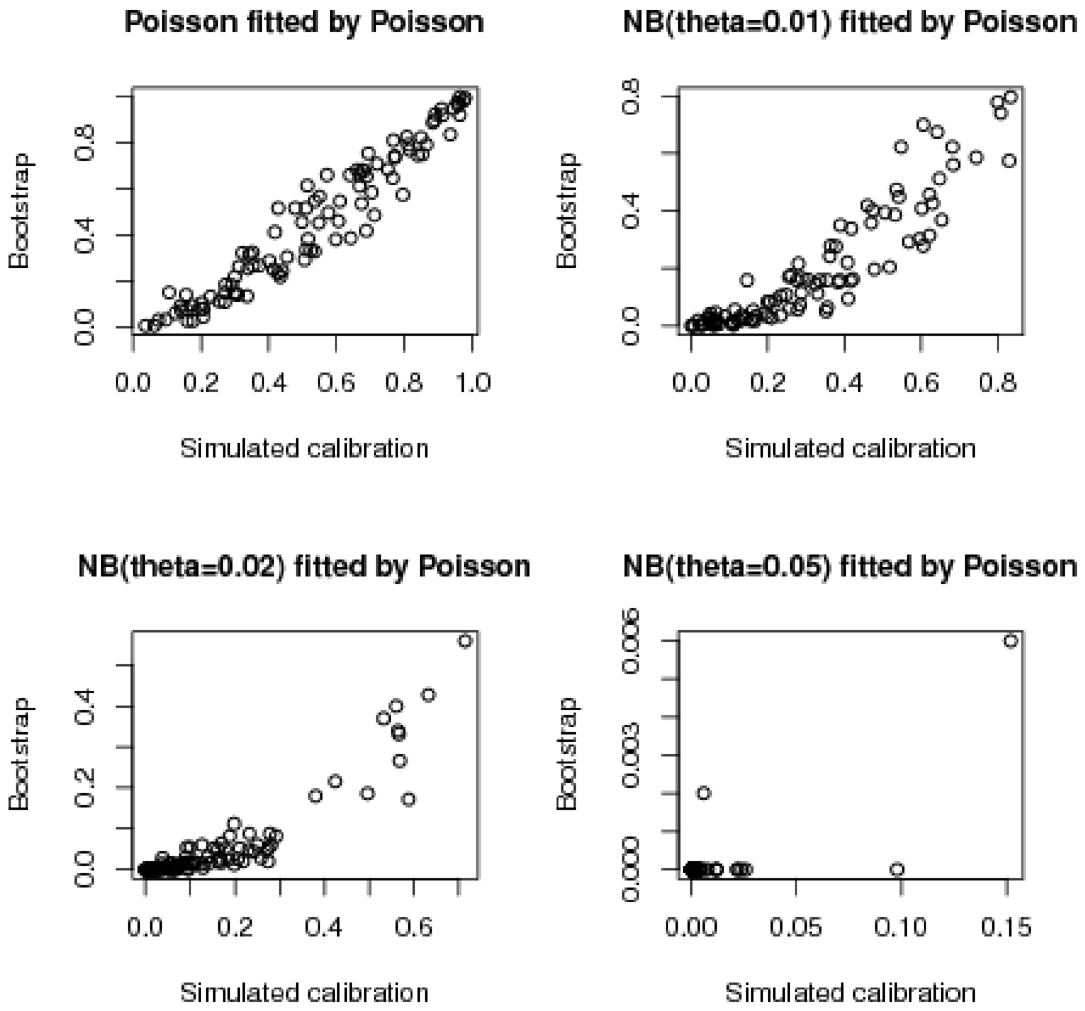
Thank you

10% MRR



φ1 interc.
0.5
0
−0.5
φ1 slope
0.02
0
−0.02
−0.04
−0.06
φ2 interc.
1
0.5
0
−0.5
φ2 slope
0.05
0
−0.05
φ3 interc.
2.5
2
1.5
1
0.5
φ3 slope
0.05
0
−0.05
−0.1
−0.15
φa interc.
3
2
1
0
φa slope
0.1
0.05
0
−0.05
−0.1
log θ
−2
−3
−4
−5
−6
log p
0.5
0
−0.5
−1
−1.5
log σ2
20
10
0
−10
−20
0
10
20
30
γ

Simulation comparison with bootstrap: GLM



Sampling distributions of parameter estimates