Forecasting numbers of claims.

Tom Wright MA CStat FIA

ASTIN 2007

Context: Collective risk model for pricing or reserving

- Model for individual claim severity, Z (eg Log-Normal)
- Model for number of claims, N (eg Negative Binomial)
- Assume claim amounts are iid and independent of number of claims
- Calculate aggregate loss distribution: $T = Z_1 + Z_2 + \ldots + Z_N$ (eg Monte Carlo, Recursion, Fourier Transform)
- Apply pricing/risk-load principles to the aggregate loss distribution

Here we are concerned with the second point only: modelling the number of claims N.

Focus is not on forecasting for individual risks, as in credibility theory, but on forecasting aggregate claim numbers for a collective

What constitutes a good claim number forecast?

For a particular proposed reinsurance contract, two actuaries A and B have each produced a forecast for the number of claims N:

A says: expected number is 36, standard deviation is 6

B says: expected number is 36, standard deviation is 9

Which is the better forecast?

What constitutes a good claim number forecast?

For a particular proposed reinsurance contract, two actuaries A and B have each produced a forecast for the number of claims N:

A says: expected number is 36, standard deviation is 6

B says: expected number is 36, standard deviation is 9

Which is the better forecast?

A is more precise, but is a poor forecast if this level of precision is not warranted by the available information

B is the better forecast if the larger standard deviation comes from recognising a genuine source of uncertainty that was ignored by A

Key message: good forecasts are not unrealistically precise

- Good forecasts are as precise as possible given the available information.
- “As precise as possible” is often not very precise.
- Good forecasts recognise this.
- Insurance is about accepting risk for an appropriate price.
- If uncertainty of forecasts is understated, premiums will be too low for the true risk.
- An insurer that consistently understates uncertainty will lose money!

Example: Pricing a reinsurance treaty

- Number of claims last year was 17
- No other information is available
- What is your forecast for the number of claims next year?

Example: Pricing a reinsurance treaty

- Number of claims last year was 17
- No other information is available
- What is your forecast for the number of claims next year?

- Mean: 17 ?

Example: Pricing a reinsurance treaty

- Number of claims last year was 17
- No other information is available
- What is your forecast for the number of claims next year?

- Mean: 17 ?
- Variance (or ‘mean square prediction error’): 34 ?
- 34 is variance of the difference of two independent Poisson variables with mean of 17: can be viewed as process variance of 17 (in respect of next year’s count) plus parameter estimation variance of 17 (from last year’s count).

Example: Pricing a reinsurance treaty

- Suppose number of claims last year was zero
- What is your forecast for the number of claims next year?

Example: Pricing a reinsurance treaty

- Suppose number of claims last year was zero
- What is your forecast for the number of claims next year?
- Clearly nonsense to say the mean is zero for next year
  - number of claims cannot be negative so zero mean implies certainty that there will no claims next year – this does not follow from having observed zero claims last year!
- Forecast mean should be greater than the observed mean in this case
- This illustrates the impracticality of frequentist “unbiasedness”

Example: Pricing a reinsurance treaty

- Suppose number of claims last year was zero
- What is your forecast for the number of claims next year?
- Clearly nonsense to say the mean is zero for next year
  - number of claims cannot be negative so zero mean implies certainty that there will no claims next year – this does not follow from having observed zero claims last year!
- Forecast mean should be greater than the observed mean in this case
- This illustrates the impracticality of frequentist "unbiasedness"
- Solution: Bayesian methods. If process is Poisson with uninformative prior, posterior distribution for Poisson parameter is Gamma, so forecasting distribution is a Gamma mixture of Poissons, ie Negative Binomial. Forecast mean = $n+1$, variance = $2*(n+1)$ (where n is number of claims observed last year).

Example: Pricing a reinsurance treaty – Bayesian method

- Poisson process:

$$\Pr(n_1 \mid \lambda) = e^{-\lambda}.\lambda^{n_1} / n_1!$$

- Viewed as a distribution for $\lambda|n_1$ this is a Gamma pdf, $f(\lambda|n_1)$ say.
- Forecasting distribution for $n_2|n_1$ is therefore a Gamma mixture of Poissons:

$$\Pr(n_2 \mid n_1) = \int \Pr(n_2 \mid \lambda, n_1).f(\lambda \mid n_1).d\lambda$$

$$= \int e^{-\lambda} \frac{\lambda^{n_2}}{n_2!}.e^{-\lambda} \frac{\lambda^{n_1}}{n_1!}.d\lambda$$

$$= \frac{(n_1 + n_2)!}{n_1!n_2!} (1/2)^{n_2+n_1+1}$$

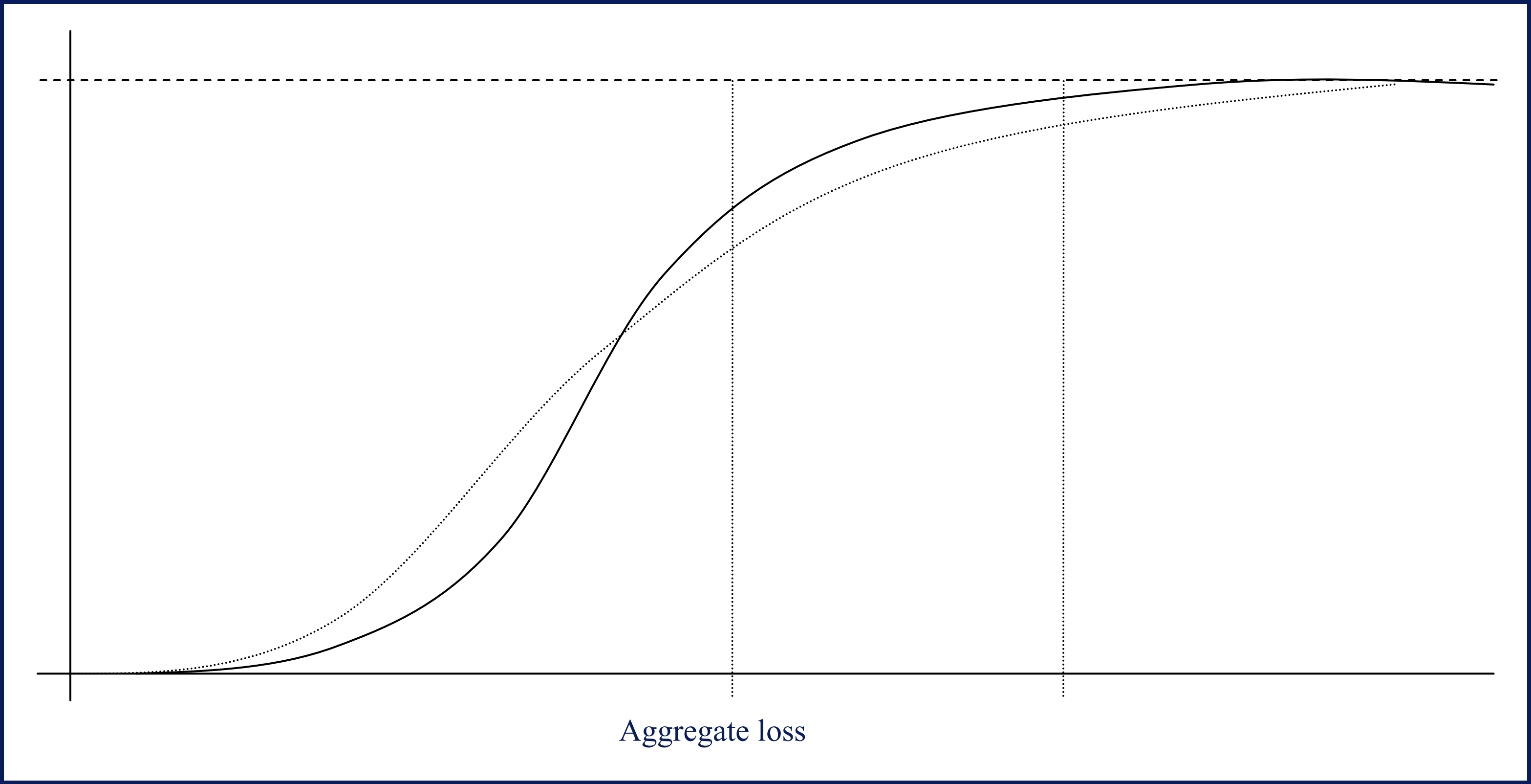
- This is a Negative Binomial distribution for n_2, with $p=1/2$, $r = n_1+1$ so:
- $E(n_2) = r.(1-p)/p = n_1+1$
- $Var(n_2) = E(n_2)/p = 2.(n_1+1)$

Example: Pricing a reinsurance treaty – Other sources of uncertainty

- In our forecast of claim numbers, we took account of:
  - process uncertainty (assumed Poisson)
  - parameter estimation uncertainty (via Bayesian argument)
- But we assumed the Poisson parameter to be the same in both years. Poisson parameter = exposure * claim frequency, so it may change due to any or all of the following:
  - Increase or decrease in exposure
  - Increase or decrease in claim frequency caused by:
    - trend change in frequency (including cycles)
    - short-term random change in frequency ("contagion")
    - changing mix of exposure from heterogeneous population
- These possibilities all act to increase the uncertainty in claim number forecasts: if not adequately allowed for, risk margins (in premiums and/or reserves) will be understated.
- In some circumstances, "best estimates" may also be understated
- Further uncertainty arises from potential model error: process might not be Poisson at all.

How can understanding claim number uncertainty lead to biased "best estimates" of premiums and/or reserves?

- This occurs when pricing treaties with aggregate limits (eg stop-loss or back-up treaties):



Objective of the ASTIN colloquium paper

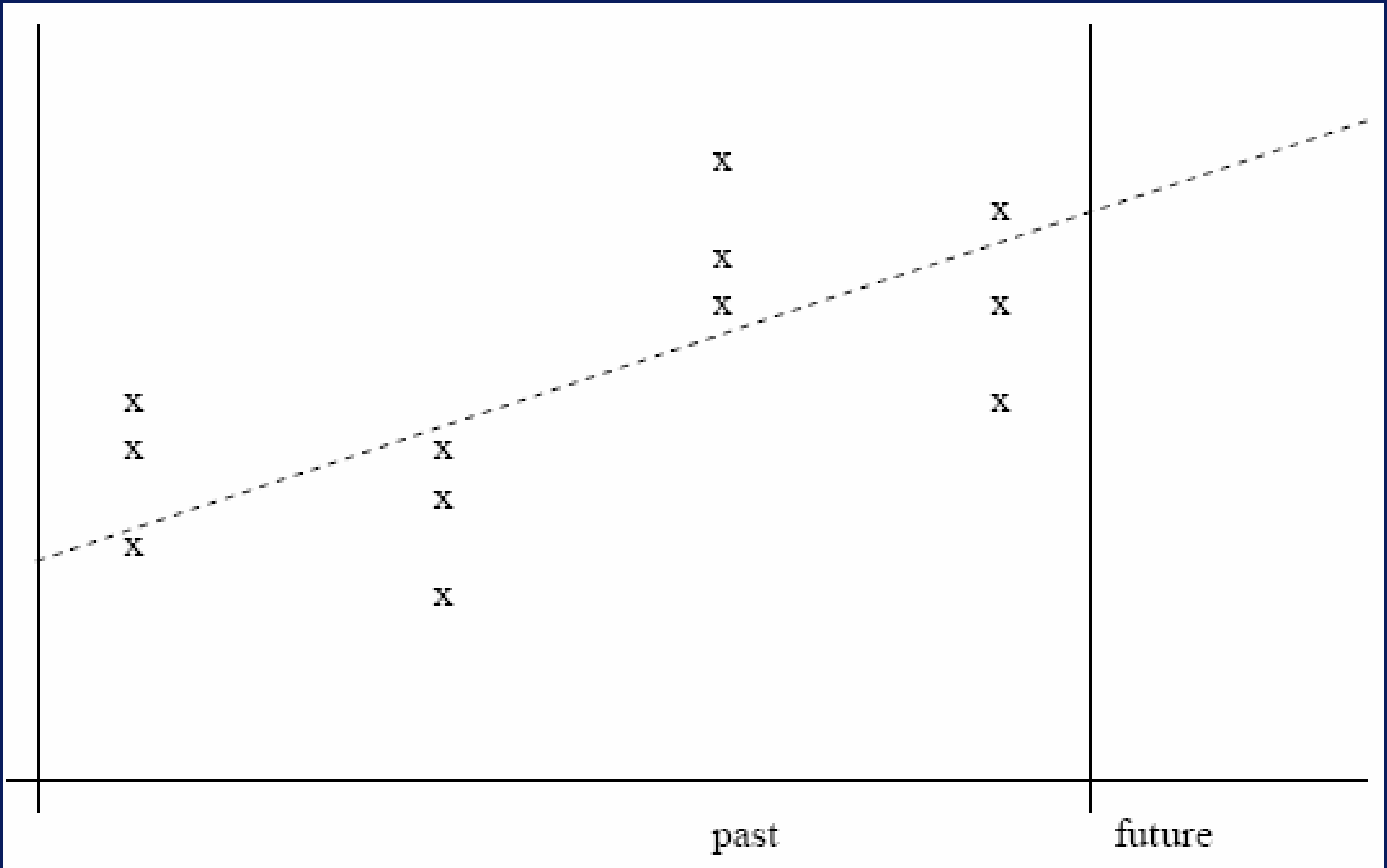
- To construct a general claim number forecasting model that takes account of all the main sources of uncertainty:
  - Process (or aleatoric) uncertainty
  - Parameter estimation uncertainty
  - Trends and cycles in claim frequency
  - Short-term random fluctuations in claim frequency (“contagion”)
  - Changing mix of underlying risk units from heterogeneous population
  - Uncertainty in quantum of future exposure

Objective of the ASTIN colloquium paper

- To construct a general claim number forecasting model that takes account of all the main sources of uncertainty:
  - Process (or aleatoric) uncertainty
  - Parameter estimation uncertainty
  - Trends and cycles in claim frequency
  - Short-term random fluctuations in claim frequency (“contagion”)
  - Changing mix of underlying risk units from heterogeneous population
  - Uncertainty in quantum of future exposure

- As always with modelling, seek the best compromise between realism and simplicity. Eg assume time effects can be adequately partitioned into trends/cycles and stochastically independent contagion effects. Unlikely this is adequate in all situations: an alternative would be to allow serial correlation in contagion effects. (Model does have serial correlation, but only that induced by persistency of risk-units.)

Process variation, heterogeneity, contagion and trends

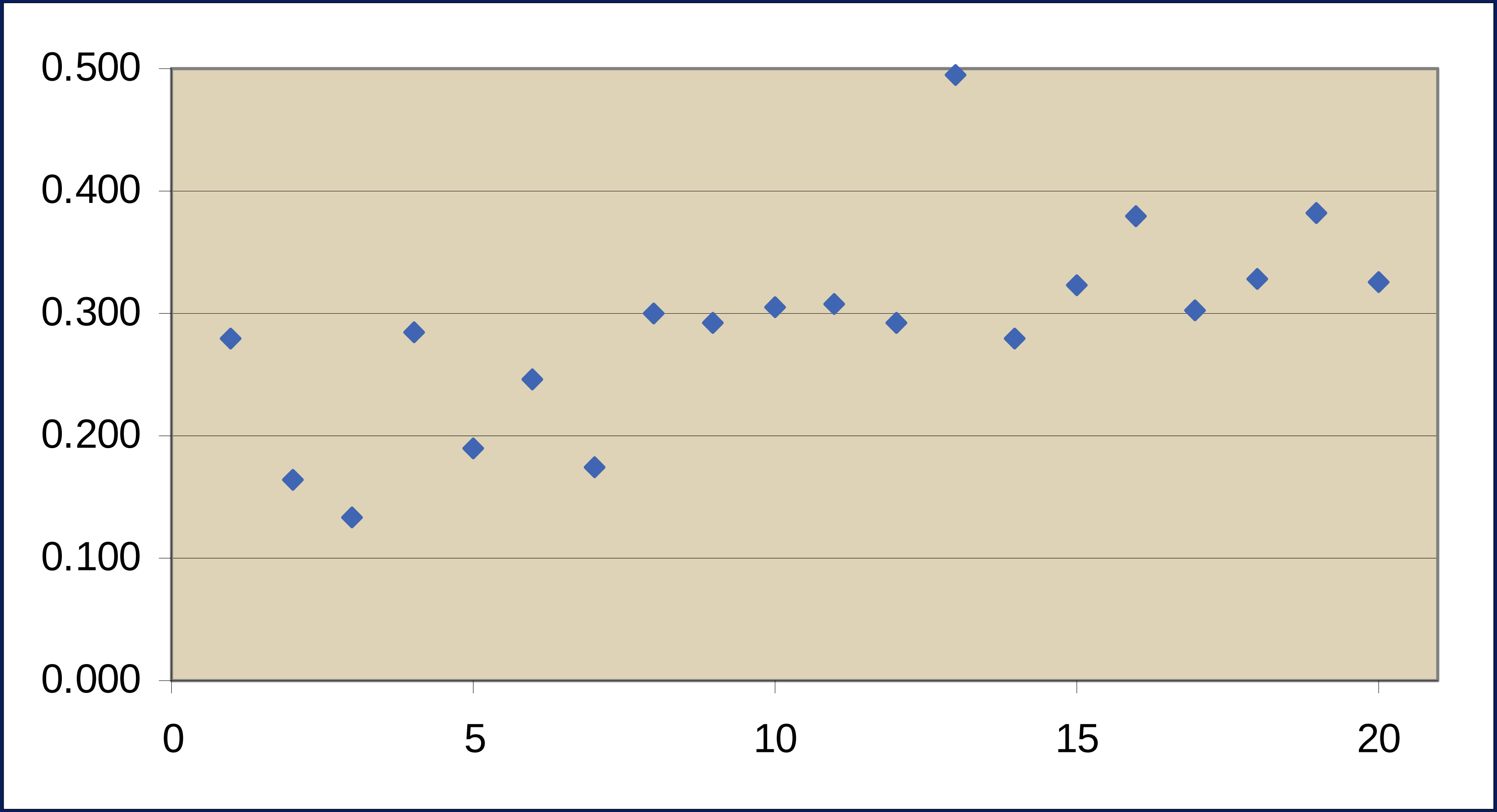


Example data-set – case of one observation in each epoch

Epoch (t)	Exposure (x_t)	Number of claims (n_t)
-20	100	28
-19	97	16
-18	106	14
-17	112	32
-16	111	21
-15	102	25
-14	120	21
-13	117	35
-12	113	33
-11	118	36
-10	124	38
-9	113	33
-8	103	51
-7	93	26
-6	87	28
-5	95	36
-4	89	27
-3	79	26
-2	84	32
-1	77	25
0	74 to 90 with 95% confidence	N_0 (to be forecast)

Example data-set – case of one observation in each epoch

Observed frequency against time:



Alternative approaches

- Approach 1
  - second moment assumptions for each source of uncertainty
  - derive mean and variance for N
  - use Negative Binomial distribution with the required mean and variance
  - (or Poisson–Inverse-Gauss: heavier-tailed than Neg Bin: see Willmot 1986)
- Approach 2
  - full distribution assumptions for each source of uncertainty
  - derive full predictive distribution for N (ideally using Bayesian methods)

Practical approach – Hybrid of Approaches 1 and 2

- Consider Approach 1 initially: derive formula for first two moments of forecast from assumptions on first two moments of each source.
- To calibrate the model make full distributional assumptions for some components (borrowing from Approach 2).
- Revert to Approach 1 for final stage: use Negative Binomial forecasting distribution with first two moments as required.
- Negative Binomial forecasting distribution provides useful compromise between realism and simplicity in many situations: certainly more realistic than Poisson. However, it may be inadequate when expected number of claims is large.

Notation

- x_t = number of risk-units exposed in epoch t
- μ_t = mean claim frequency in epoch t, allowing for trends and cycles but not including contagion effects: eg $\mu_t = \exp(\beta_0 + \beta_1.t + \beta_2.t^2)$
- λ_t = mean claim frequency in epoch t, allowing for trends/cycles and contagion effects: $\lambda_t = \mu_t + \delta_t$ say
- q_t = proportion of risk-units in epoch t not previously observed
- λ_{it} = expected number of claims from risk-unit i in epoch t (not necessarily equal to λ_t because of heterogeneity)
- N_{it} = actual number of claims from risk-unit i in epoch t

Approach 1: Second moment assumptions

Process uncertainty	$E(N_{it}) = \lambda_{it}$	$Var(N_{it}) = \lambda_{it}.(1+\theta.\lambda_{it})$
Heterogeneity	$E(\lambda_{it}) = \lambda_t$	$Var(\lambda_{it}) = (\rho_h.\lambda_t)^2$
Contagion	$E(\lambda_t) = \mu_t$	$Var(\lambda_t) = (\rho_c.\mu_t)^2$
Parameter estimation uncertainty (note the Bayesian ethos)	$E(\mu_t) = \mu_t'$	$Var(\mu_t) = (\rho_e.\mu_t')^2$
Future exposure uncertainty	$E(x_t) = x_t'$	$Var(x_t) = (\rho_x.x_t')^2$

Derivation of predictive mean and variance (Section 2 of ASTIN paper)

General formulas:

- $E(X) = E(E(X|Y))$
- $Var(X) = E(Var(X|Y)) + Var(E(X|Y))$

Applying these once for each type of uncertainty produces:

- $E(N_t) = x_t'.\mu_t'$
- $Var(N_t) = E(N_t).\{1+c_t.\ E(N_t)\}$

where:

- $c_t = (1+ \rho_x^2 + \varphi_t / x_t').(1+ \rho_c^2).(1+ \rho_e^2) - 1$
- $\varphi_t = \theta + q_t .\rho_h^2 .(1 + \theta)$

In terms of the squared variation coefficient we have:

- $Vco^2(N_t) = 1/E(N_t) + c_t$

Limiting and special cases of general formula

The general formulas for predictive mean and variance are:

- $E(N_t) = x_t'.\mu_t'$
- $Vco^2(N_t) = 1/E(N_t) + c_t$

where:

- $c_t = (1+ \rho_x^2 + \varphi_t / x_t').(1+ \rho_c^2).(1+ \rho_e^2) - 1$
- $\varphi_t = \theta + q_t .\rho_h^2 .(1 + \theta)$

If expected future exposure x_t' tends to infinity, we obtain the limit:

$Vco^2(T_t) = Vco^2(N_t) = (1+ \rho_x^2).(1+ \rho_c^2).(1+ \rho_e^2) - 1$

If variation coefficients are so small that cross-terms can be neglected:

- $Vco^2(N_t) = 1/E(N_t) + \rho_x^2 + \varphi_t / x_t' + \rho_c^2 + \rho_e^2$

Calibration (case $\theta = 0$) (Section 3 of ASTIN paper)

The general formulas for forecast mean and variance are:

- $E(N_t) = x_t'.\mu_t'$
- $Vco^2(N_t) = 1/E(N_t) + c_t$

where (if $\theta = 0$):

- $c_t = (1+ \rho_x^2 + \rho_h^2.q_t /x_t').(1+ \rho_c^2).(1+ \rho_e^2) - 1$

To apply these formulas we need values for all 7 quantities:

- x_t', ρ_x, q_t estimated from business plans and past renewal and lapse rates – judgement more relevant than statistical analysis
- Given sufficient data, μ_t', $\rho_{e,}$ $\rho_{h,}$ ρ_c can be estimated by statistical analysis of past claim numbers and corresponding exposure data
- Data-set comprises data-points (t, x_{tk}, n_{tk}). Subscript k allows for possibility of more than one data-point in single epoch
- To distinguish θ and ρ_h^2 requires several fixed exposure blocks to be observed over several epochs – this not pursued as such data are rarely available.

Distribution assumptions – Process variation

Exposure unit i, epoch t: $E(N_{it}) = \lambda_{it,}$ $Var(N_{it}) = \lambda_{it}.(1+\theta.\lambda_{it})$

$\theta = 0$: Poisson

$\theta < 0$: Binomial

- number of success in r independent trial each with probability p
- $\lambda_{it} = r.p_{it}$ and $\theta = -1/r$,
- $Var(N_{it}) = r.p_{it}.(1-p_{it})$

$\theta > 0$: Negative Binomial:

- number of failures before r^{th} success in independent trials
- $\lambda_{it} = r.(1-p_{it})/p_{it}$ and $\theta = 1/r$
- $Var(N_{it}) = r.(1-p_{it})/p_{it}.\{1+(1-p_{it})/p_{it}\} = r.(1-p_{it})/p_{it}^2$

For both Binomial and Negative Binomial, r is assumed constant across risk units: heterogeneity is modelled by variation in p_{it}

Distribution assumptions – Heterogeneity

- Second moment assumptions: $E(\lambda_{it}) = \lambda_t$ and $Var(\lambda_{it}) = (\rho_h.\lambda_t)^2$
- Distributional assumption: λ_{it} Gamma-distributed across risk-units i
- Combined with Poisson process variation, this implies Negative Binomial distribution for number n_t of claims arising from x_t risk-units selected independently at random:

$$\Pr(n_t \mid x_t, \lambda_t, \rho_h) = \frac{\Gamma(n_t + r_t)}{n_t!\Gamma(r_t)}(1 - p_t)^{n_t}.p_t^{\,r_t}$$

- where the Negative Binomial parameters are:
- $p_t = 1/(1+ \rho_h^2.\lambda_t)$
- $r_t = x_t / \rho_h^2$
- Binomial and Negative Binomial process variation ($\theta \neq 0$) not considered further as mixed distributions become too complex when there is heterogeneity
- Advantage of Gamma distribution for heterogeneity is sum-stability (or infinite divisibility): same distribution family for any number x_t of risk units. Another possibility is Inverse Gauss (see eg Willmot, ASTIN 1986).

Heterogeneity - Alternative distribution assumptions

- Second moment assumptions: single policy: $E(\lambda_i) = \lambda$ and $Var(\lambda_i) = \varphi.\lambda^2$
- For x policies, total Poisson parameter: $E(\Lambda) = x.\lambda$ and $Var(\Lambda) = x.\varphi.\lambda^2$
- If heterogeneity has sum-stable distribution (eg Gamma or Inverse Gauss), then Λ has same distribution family for any x:

$f(\Lambda \mid s,a)$	Mean	Variance	Scale, s	Shape, a
Gamma	s.a	$s^2.a$	$\varphi.\lambda$	x/φ
Inv Guass	s.a	$s^2.a^3$	$x^2.\lambda/\varphi$	φ/x

- Number of claims N from x exposure units is mixture of Poisson(Λ):

$$\Pr(n \mid x, \lambda, \varphi) = \int e^{-\Lambda} \frac{\Lambda^n}{n!} . f(\Lambda \mid s, a).d\Lambda$$

- If f() is Gamma, this gives Neg Bin, if f() is Inverse Gauss gives Poisson-IG
- View $\Pr(n|x,\lambda,\varphi)$ as posterior distribution for λ
- If f() is Gamma, this is Pearson-VI, if f() is IG this is complex and numerical methods needed to evaluate posterior mean and variance of λ

Two methods of estimating trend and contagion parameters (Section 3 of paper describes bits of both)

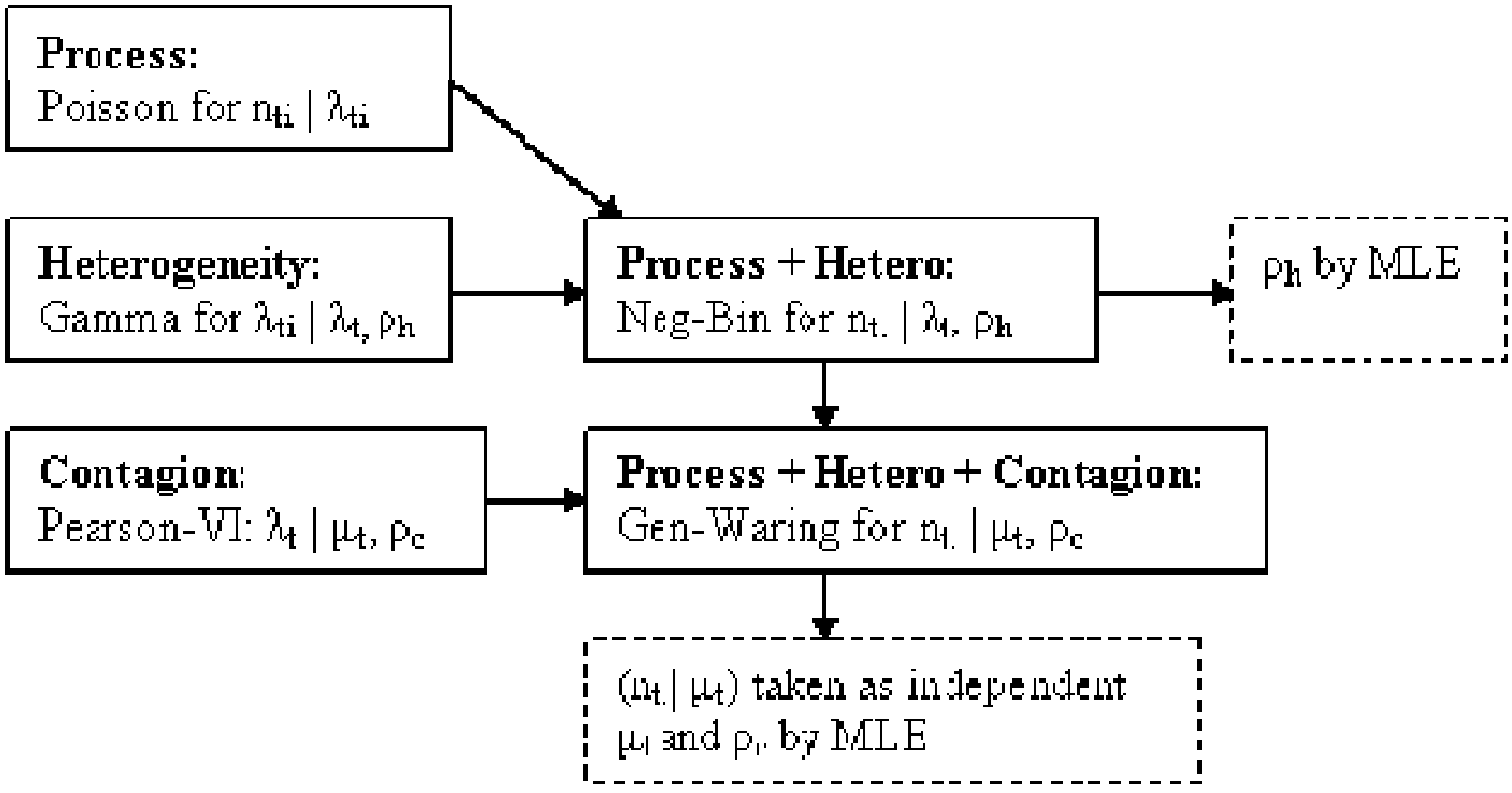
Method A (outlined in Section 3.3.2 of ASTIN paper):

- Assume Pearson-VI distribution for contagion effects.
- Distribution of total number of claims in an epoch (allowing for process variation, heterogeneity and contagion) is then Pearson-VI mixture of Negative Binomials (known as Generalised Waring Distribution).
- Fit this by MLE to estimate trend and contagion parameters.

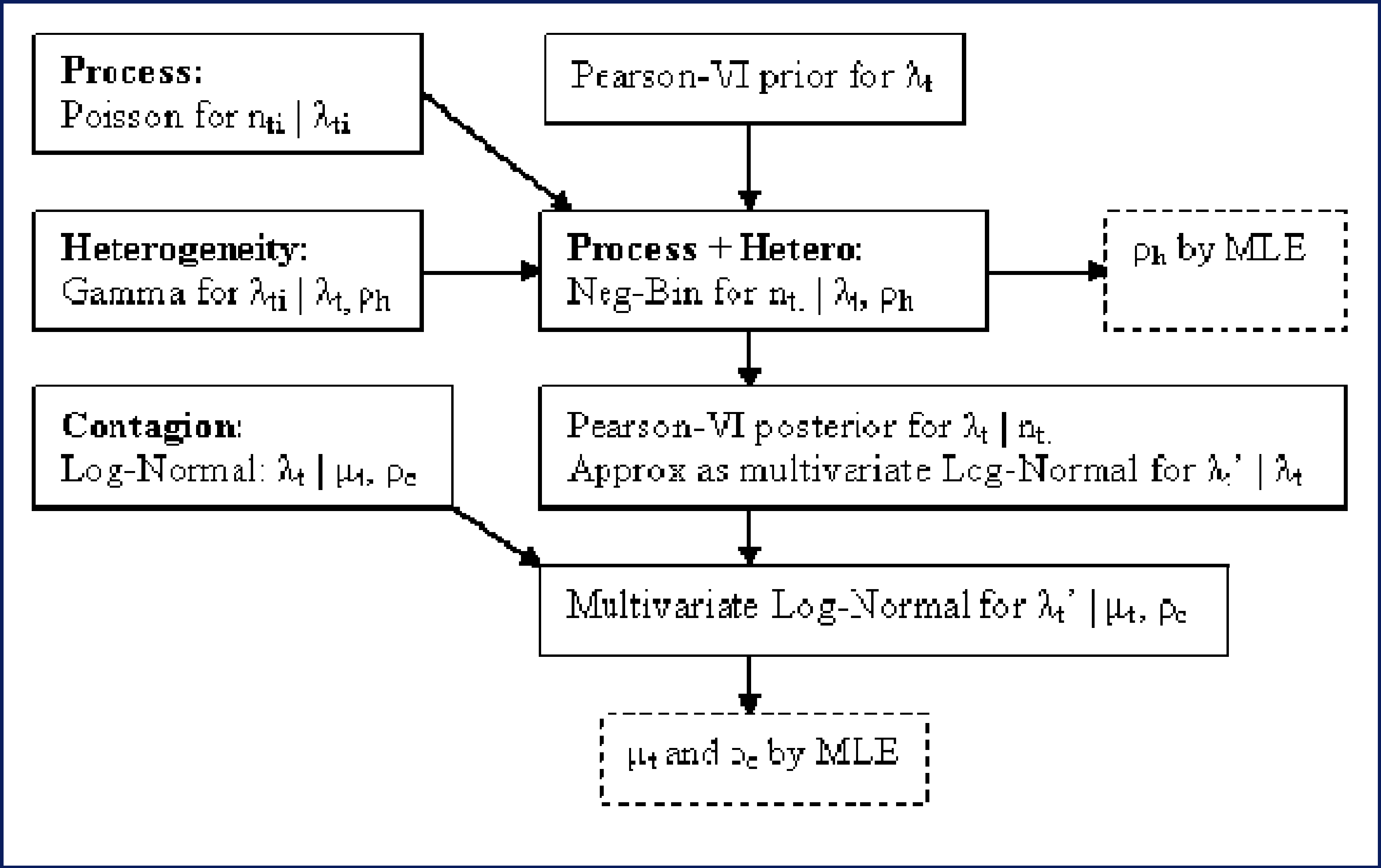
Method B (first part covered in Section 3.2 of ASTIN paper):

- Use Negative Binomial distribution (from Poisson process and Gamma heterogeneity) to produce Bayesian posterior distribution of λ_t for each epoch separately.
- Posterior distribution of λ_t is a Pearson-VI distribution: approximate this using Log-Normal with same mean and variance.
- Analyse these posterior distributions across epochs to estimate trends and contagion parameter ρ_c, on assumption that contagion effects are also Log-Normal.

Estimation of Trend and Contagion parameters: Method A



Estimation of Trend and Contagion parameters: Method B



Comparison of the two methods

Method A (outlined in Section 3.3.2 of ASTIN paper):

- Single stage procedure based on combined model for process, heterogeneity and contagion.
- Disadvantage: assumes total numbers of claims in each epoch are stochastically independent (given contagion and trend parameters): not true if heterogeneous risk units persist across epochs.

Method B (first stage covered in Section 3.2 of ASTIN paper):

- Two stage procedure: first estimate λ_t for each epoch separately, then impose trend and contagion model to smooth across epochs.
- Can take account of stochastic dependence between λ_t estimates (induced by persistence of risk units) using multivariate Log-Normal.
- Generalises to long-tail classes in which number of claims is not fully known for past years: in the first stage, the λ_t are estimated by stochastic triangle projection methods (gives correlated estimates).

Method A – Pearson-VI Contagion

- Second moment assumptions: $E(\lambda_t) = \mu_t$ and $Var(\lambda_t) = (\rho_c.\mu_t)^2$
- Distributional assumption: λ_t has Pearson-VI distribution:

$$f(\lambda_t) = \frac{\Gamma(a_t + b_t).(\lambda_t / s_t)^{a_t - 1}}{s_t.\Gamma(a_t).\Gamma(b_t).(1 + \lambda_t / s_t)^{a_t + b_t}}$$

- Implies: $E(\lambda_t) = s.a/(b-1)$ and $Vco^2(\lambda_t) = (a+b-1)/\{a.(b-1)\}$
- Setting $s = 1/\rho_h^2$ and solving for a and b:
- $a_t = \{1 + \rho_h^2.\mu_t.(1+ \rho_c^2)\} / \rho_c^2$ and $b_t = \{1 + 2.\rho_c^2 + 1 / (\rho_h^2.\mu_t)\} / \rho_c^2$
- By previous assumptions (Poisson process, Gamma heterogeneity) $n_{t.}|\lambda_t$ has a Negative Binomial distribution
- So combination of process, heterogeneity and contagion produces a Pearson-VI mixture of Negative Binomials:

$$\Pr(n_{t.} \mid x_{t.}, \mu_t, \varphi, \rho_c) = \int \frac{\Gamma(n_{t.} + x_{t.}/\varphi).(\varphi.\lambda_t)^{n_{t.}}}{n_{t.}!\Gamma(x_{t.}/\varphi).(1 + \varphi.\lambda_t)^{n_{t.} + x_{t.}/\varphi}}.f(\lambda_t).d\lambda_t$$

$$= \frac{\Gamma(n_{t.} + x_{t.}/\varphi).\Gamma(a_t + b_t).\Gamma(n_{t.} + a_t).\Gamma(x_{t.}/\varphi + b_t)}{n_{t.}!\Gamma(x_{t.}/\varphi).\Gamma(a_t).\Gamma(b_t).\Gamma(n_{t.} + x_{t.}/\varphi + a_t + b_t)}$$

Method A – Pearson-VI mixture of Negative Binomials

$$\Pr(n_{t.} \mid x_{t.}, \varphi, \mu_t, \rho_c) = \\ = \frac{\Gamma(n_{t.} + x_{t.}/\varphi).\Gamma(a_t + b_t).\Gamma(n_{t.} + a_t).\Gamma(x_{t.}/\varphi + b_t)}{n_{t.}!\Gamma(x_{t.}/\varphi).\Gamma(a_t).\Gamma(b_t).\Gamma(n_{t.} + x_{t.}/\varphi + a_t + b_t)}$$

- Meng Shengwang, Wei and Whitmore (ASTIN 1999) call this the “Negative Binomial Pareto Distribution” and use it for combination of process variation and heterogeneity. Klugman, Panjer & Willmot call it “Generalized Waring”.
- Viewed as a distribution for ρ_c and μ_t this is intractable, so instead of attempting complete Bayesian estimation of these quantities, obtain point estimates by MLE (mode of posterior distribution with uninformative prior) and approximate estimation covariance as negative inverse Hessian matrix
- This is sufficient to calibrate the general formulas for first two moments of forecast distribution: need only best estimate and variance of μ_t
- MLE assumes ($n_{t.} \mid x_{t.}, \varphi, \mu_t, \rho_c$) are stochastically independent across epochs t: not true if risk units persist from one epoch to the next.

Method B - Bayesian estimation of λ_t for each epoch t

- The Negative Binomial distribution for $n_t|\lambda_t$ (from Poisson process and Gamma heterogeneity) can be used for Bayesian estimation of λ_t – this eliminates the problem of zero estimates λ_t'
- Writing φ for ρ_h^2 (since we have assumed θ = 0), the Negative Binomial can be expressed:

$$\Pr(n_t \mid x_t, \lambda_t, \varphi) = \frac{\Gamma(n_t + x_t / \varphi).(\varphi.\lambda_t)^{n_t}}{n_t!\Gamma(x_t / \varphi).(1 + \varphi.\lambda_t)^{n_t + x_t / \varphi}}$$

- Given multiple observations (n_{tk}, x_{tk}), the likelihood $f(\lambda_t|n_{tk},x_{tk})$ is the product of this over k. Viewed as a distribution for $\lambda_{t,}$ this is a Pearson-VI distribution.
- From standard results for the Pearson-VI distribution, we have:
- $E(\lambda_t) = (n_{t.} + 1) / (x_{t.} - 2.\varphi) = \lambda_t'$ say [compare MLE $\lambda_t' = n_{t.}/ x_{t.}$]
- $Var(\lambda_t) = \lambda_t' .(1+ \varphi.\lambda_t') / (x_{t.} - 3.\varphi)$ [MLE: $Var(\lambda_t') = \lambda_t.(1+ \varphi.\lambda_t) / x_{t.}$]
- The above assumes φ is known, but in practice φ too must be estimated from the data. The Negative Binomial likelihood (above) gives an intractable posterior distribution for φ: a practical solution is to estimate φ as the mode of this posterior, ie, use MLE to obtain a point estimate of φ.

Method B – Log-Normal Contagion

Moment assumptions for contagion: $E(\lambda_t) = \mu_t$ and $Var(\lambda_t) = (\rho_c.\mu_t)^2$

- Distributional assumption: λ_t independent Log-Normal for each t
- $\lambda_t = \mu_t.u_t$ where u_t is Log-Normal with $E(u_t) = 1$, $Var(u_t) = \rho_c^2$
- $\delta_t = \ln(u_t)$ is Normal with $Var(\delta_t) = \ln(1+ \rho_c^2)$ and $E(\delta_t) = - Var(\delta_t)/2$

Method B – Approximate the Pearson-VI posterior of λ_t as Log-Normal

- Pearson-VI posterior gives $E(\lambda_t) = (n_{t.} + 1) / (x_{t.} - 2.\varphi)$ (denoted λ_t') and $Var(\lambda_t) = \lambda_t'.(1+ \varphi.\lambda_t') / (x_{t.} - 3.\varphi)$
- $\lambda_t = \lambda_t'.w_t$: w_t is Pearson-VI with $E(w_t) = 1$, $Var(w_t) = (1/\lambda_t' + \varphi) / (x_{t.} - 3.\varphi)$
- Approximate w_t as Log-Normal:
- $\zeta_t = \ln(w_t)$ approx Normal with $Var(\zeta_t) = \ln(1+ Var(w_t))$ and $E(\zeta_t) = - Var(\zeta_t)/2$
- Infer $Cov(\lambda_s, \lambda_t)$ (persistence of risk-units from heterogeneous population) from non-Bayesian calculations: $Cov(w_s, w_t) = (1-q_{s+1})\ldots(1-q_t).\varphi / (x_{s.} - 3.\varphi)$
- (No additional stochastic dependence between the λ_t from prior beliefs: this factored-in via trend model)
- Implies $Cov(\zeta_s, \zeta_t) = \ln(1+ Cov(w_s, w_t))$
- Reciprocal of a Log-Normal is also Log-Normal:
- $\lambda_t' = \lambda_t.v_t$: where $v_t = 1/w_t = \exp(- \zeta_t) = \exp(\varepsilon_t)$ say
- $Var(\varepsilon_t) = Var(\zeta_t) = \ln(1+ Var(w_t))$
- $E(\varepsilon_t) = -E(\zeta_t) = Var(\varepsilon_t)/2$
- $Cov(\varepsilon_s, \varepsilon_t) = Cov(\zeta_s, \zeta_t) = \ln(1+ Cov(w_s, w_t))$

Method B – Combined Log-Normal model for Process variation, Heterogeneity and Contagion

For each past epoch we have the estimate $\lambda_t' = (n_{t.} + 1) / (x_{t.} - 2.\varphi)$

Process variation and heterogeneity modelled using: $\lambda_t' = \lambda_t.v_t$

- $\varepsilon_t = \ln(v_t)$ is Multivariate-Normal with
- $Var(\varepsilon_t) = \ln\{1+ (1/\lambda_t' + \varphi) / (x_{t.} - 3.\varphi)\}$, $E(\varepsilon_t) = Var(\varepsilon_t)/2$, and
- $Cov(\varepsilon_{s,} \varepsilon_t) = \ln\{1+ (1-q_{s+1})\ldots(1-q_t).\varphi / (x_{s.} - 3.\varphi)\}$

Contagion modelled using: $\lambda_t = \mu_t.u_t$

- $\delta_t = \ln(u_t)$ is Normal with $Var(\delta_t) = \ln(1+ \rho_c^2)$ and $E(\delta_t) = - Var(\delta_t)/2$

Combining these gives Log-Normal model for process, hetero and contagion:

- $\lambda_t' = \mu_t.u_t.v_t \Rightarrow \ln(\lambda_t') = \ln(\mu_t) + \delta_t + \varepsilon_t \Rightarrow y_t = \ln(\mu_t) + \eta_t$
- where $y_t = \ln(\lambda_t') + [Var(\delta_t) - Var(\varepsilon_t)]/2$, $\eta_t = \delta_t + \varepsilon_t + [Var(\delta_t) - Var(\varepsilon_t)]/2$
- η_t is Multivariate-Normal with $E(\eta_t) = 0$, $Var(\eta_t) = Var(\delta_t) + Var(\varepsilon_t)$,
- $Cov(\eta_s,\eta_t) = Cov(\varepsilon_s,\varepsilon_t) = \ln(1+ Cov(w_{s,}w_t))$

Method B – Combined Log-Normal model for Process variation, Heterogeneity and Contagion

We have: $y_t = \ln(\mu_t) + \eta_t$ where:

- $y_t = \ln(\lambda_t') + [\ln(1+ \rho_c^2) - \ln(1+ Var(w_t)]/2$
- $\lambda_t' = (n_{t.} + 1) / (x_{t.} - 2.\varphi)$
- η_t is Multivariate-Normal with zero mean
- $Var(\eta_t) = \ln(1+ \rho_c^2) + \ln(1+ Var(w_t)$, $Cov(\eta_s,\eta_t) = \ln(1+ Cov(w_{s,}w_t))$
- $Var(w_t) = (1/\lambda_t' + \varphi) / (x_{t.} - 3.\varphi)$, $Cov(w_{s,}w_t) = (1-q_{s+1})\ldots(1-q_t).\varphi / (x_{s.} - 3.\varphi)$

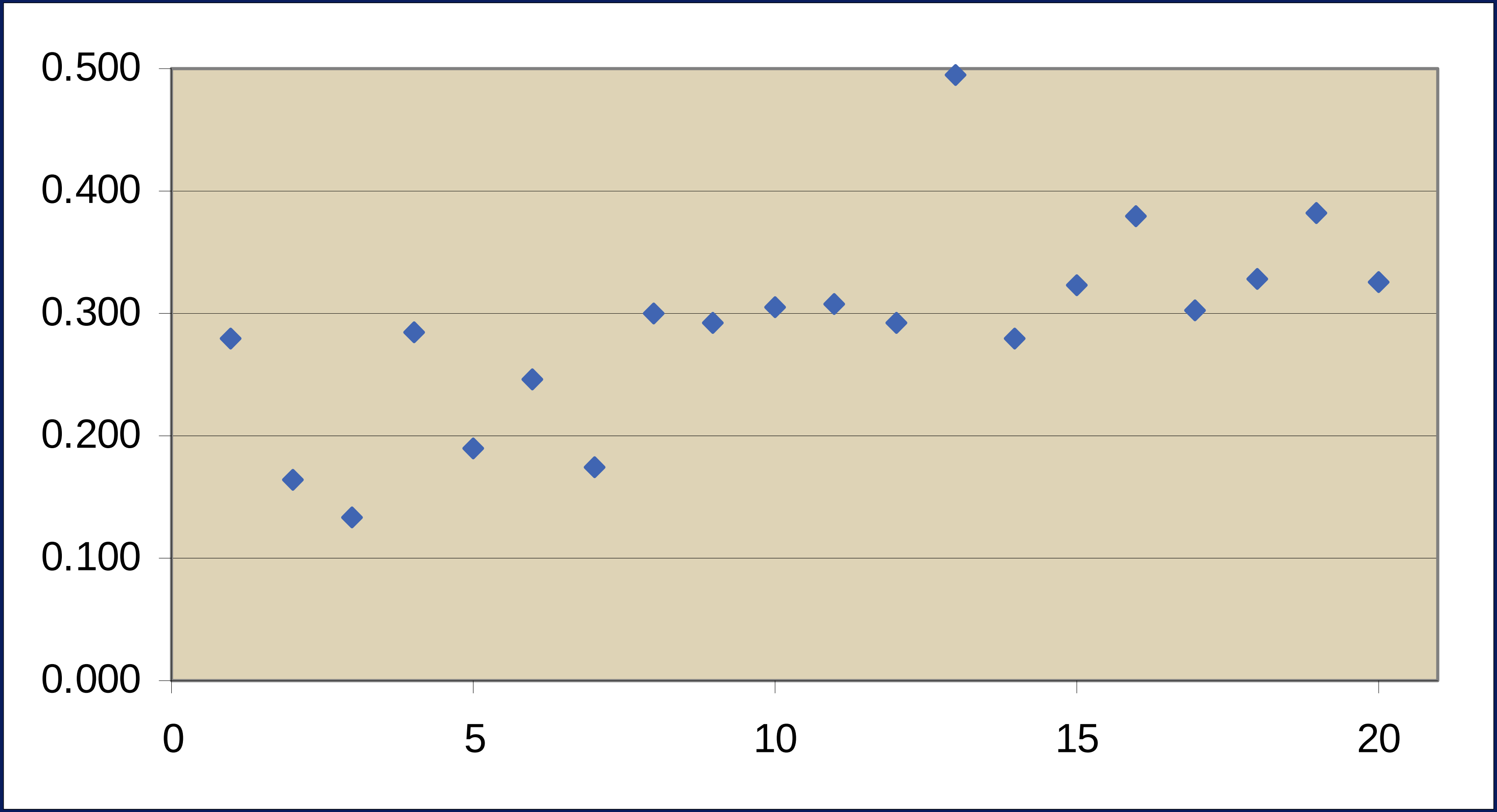
Use linear form for $\ln(\mu_t)$ to model trends/cycles, eg $\ln(\mu_t) = \beta_0 + \beta_1.t + \beta_2.t^2$

Use Normal maximum likelihood to estimate β-parameters and ρ_c:

- Given a value for ρ_c, MLEs of β-parameters are given by:
- $\beta' = (X^T.S^{-1}.X)^{-1}.X^T.S^{-1}.y$ $Var(\beta') = (X^T.S^{-1}.X)^{-1}$
- where X is the design matrix, S is the covariance matrix of η_t
- Calculate the Multivariate Normal log-likelihood from β′, and search for value ρ_c that maximises this

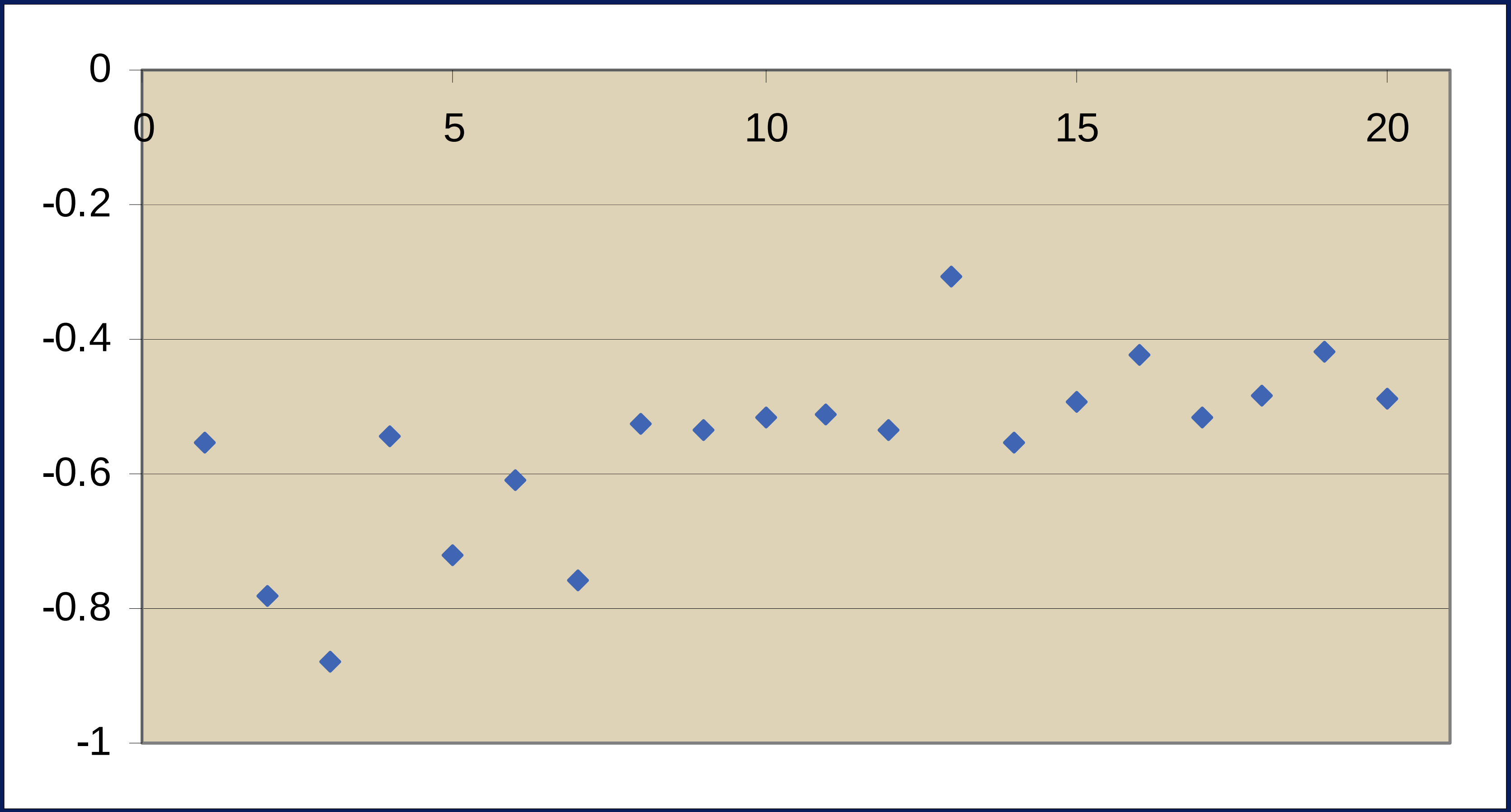
Example data-set – case of one observation in each epoch

Observed frequency against time:



Example data-set – case of one observation in each epoch

Natural log of observed frequency against time:



Example data-set – case of one observation in each epoch

t	x_t	n_t	Observed frequency n_t/x_t	Posterior mean frequency ($\varphi = 0$) $(n_t+1)/x_t$	Posterior mean freq ($\varphi = 0.25$) $(n_t+1)/(x_t - 2.\varphi)$
-20	100	28	0.280	0.290	0.291
-19	97	16	0.165	0.175	0.176
-18	106	14	0.132	0.142	0.142
-17	112	32	0.286	0.295	0.296
-16	111	21	0.189	0.198	0.199
-15	102	25	0.245	0.255	0.256
-14	120	21	0.175	0.183	0.184
-13	117	35	0.299	0.308	0.309
-12	113	33	0.292	0.301	0.302
-11	118	36	0.305	0.314	0.315
-10	124	38	0.306	0.315	0.316
-9	113	33	0.292	0.301	0.302
-8	103	51	0.495	0.505	0.507
-7	93	26	0.280	0.290	0.292
-6	87	28	0.322	0.333	0.335
-5	95	36	0.379	0.389	0.392
-4	89	27	0.303	0.315	0.316
-3	79	26	0.329	0.342	0.344
-2	84	32	0.381	0.393	0.395
-1	77	25	0.325	0.338	0.340

Method B: p-values for hypothesis tests of trends and contagion (assumed to be no heterogeneity, φ = 0)

Model	Parameters estimated by MLE				Residual sum of squares	Maximised log-likelihood
	β_0	β_1	β_2	ρ_c		
A	-1.198	-	-	-	45.552	11.082
B	-0.885	0.0314	-	-	27.955	19.881
C	-1.008	-0.0013	-0.0016	-	26.654	20.532
D	-1.212	-	-	0.203	21.489	15.17
E	-0.886	0.032	-	0.116	19.505	20.714
F	-1.003	0.0007	-0.0015	0.107	19.53	21.14

	Alternative hypothesis					
Null hypothesis	Any (RSS)	B	C	D	E	F
A	0.06%	0.00%	0.01%	0.42%	0.01%	0.02%
B	6.30%		25.40%		19.70%	28.40%
C	6.30%					27.00%
D					0.09%	0.26%
E						35.60%

Parameters estimated by Methods A and B

- Statistical hypothesis tests (previous slide) show Model E fits best (linear trend in log of frequency, plus random contagion effect).
- Parameter estimates from the two methods (on assumption $\varphi = 0$) are reasonably close (standard errors in parentheses):

Parameter	Method A	Method B
β_0 (constant term)	-0.904 (0.101)	-0.886 (0.101)
β_1 (linear trend)	0.0334 (0.0087)	0.0320 (0.0088)
ρ_c (contagion)	0.112	0.116

- Method B preferred as it allows for non-independence across epochs induced by persistence of risk units where there is heterogeneity ($\varphi > 0$)

Methods A and B – comparison of results assuming zero heterogeneity

Variance component	Formula	Method A	Method B
Process variation	E(N)	33.0	32.6
Heterogeneity	$\{E(N).\rho_h\}^2 .q/m$	0	0
Contagion	$\{E(N).\rho_c\}^2$	13.68	14.24
Estimation uncertainty	$\{E(N).\rho_e\}^2$	11.09	9.25
Exposure uncertainty	$\{E(N).\rho_x\}^2$	2.60	2.53
Interactions between the above		0.20	0.18
Total		60.6	58.8

Method B – results based on two alternative heterogeneity assumptions

Variance component	Formula	$\rho_h = 0$	$\rho_h = 0.4,\ q = 1$
Process variation	E(N)	32.6	33.7
Heterogeneity	$\{E(N).\rho_h\}^2 .q/m$	0	2.22
Contagion	$\{E(N).\rho_c\}^2$	14.24	13.44
Estimation uncertainty	$\{E(N).\rho_e\}^2$	9.25	11.65
Exposure uncertainty	$\{E(N).\rho_x\}^2$	2.53	2.71
Interactions between the above		0.18	0.25
Total		58.8	64.0

One further source of uncertainty: past claim numbers not fully known

- Often considerable uncertainty in past claim numbers for long-tail classes
- Instead of data n_t we have claim number development arrays
- Method B generalises to this situation:
  - First stage: Stochastic projection method applied to claim numbers triangles to obtain estimates λ_t' (for all past epochs t) and estimation covariance matrix (arising from process variation and heterogeneity)
  - Covariances between the λ_t' arise from heterogeneity and the fact they have been estimated from a triangle, covariances at this stage do not arise from any structure imposed by contagion and trend assumptions
  - Second stage: use contagion and trend model to smooth estimates λ_t' and to estimate trend and contagion parameters
  - Third stage: project trends and apply general formula for mean-square predictive error of claim number forecasts

Deloitte.