Severity GLMs: A Forgotten Distribution

Christopher Monsour

CAS Predictive Modeling Seminar
6 October 2008

Motivation I — Annoyances of Traditional Models

- Models that “go to school” on small claims
  - For example, the gamma GLM model has $V(\mu) = \phi\mu^2$
    - This is just constant coefficient-of-variation (CV)
  - Thus, if an observed claim size is twice the predicted claim size, this is just as big a residual for a $20 claim as for a $20,000 claim
  - Do you really believe constant CV holds over such a large range?
  - Even if you do believe this, do you really want this behavior?

Motivation I — Annoyances of Traditional Models

- Models that can't handle negative values for claim sizes
  - With a gamma model, it isn't merely the choice of link function that constrains
    - Bi-infinite values can't be accommodated by the underlying distribution
  - Why would you care?
    - Maybe salvage and subrogation should be modeled separately
      - Positive and negative flows qualitatively different
      - But even this isn't clear, since the flows are highly correlated to payments, do you really want a separate model?
    - But suppose the dependent variable is profits
    - Or suppose it is a change in evaluation

Motivation II — Central Role of Variance Function

- Even more important than the underlying distribution
  - You never need calculate the underlying distribution
    - Just the variance and deviance
  - Variance function determines how “big” each residual is
- In some ways more important than the link function
  - GLMs “go to school” on the size of residuals on the dependent variable scale, *not* on the linear scale
- Typical variance functions
  - Power law $V(\mu)=\phi\mu^p$, p=0 or p>=1 (also exists for p<0)
  - $V(\mu)=\mu\ (1-\mu)$ [logistic regression]

Motivation II — Central Role of Variance Function

- Suppose we start with the variance function—what underlying distributions are available, and what properties do they have?
  - Sensible question as a variance function either:
    - Does not correspond to a natural exponential family
    - OR uniquely determines a one-parameter natural exponential family of distributions (parameterized by μ)
      - [which may or may not be a subfamily of a two-parameter family]

Morris's Theorem

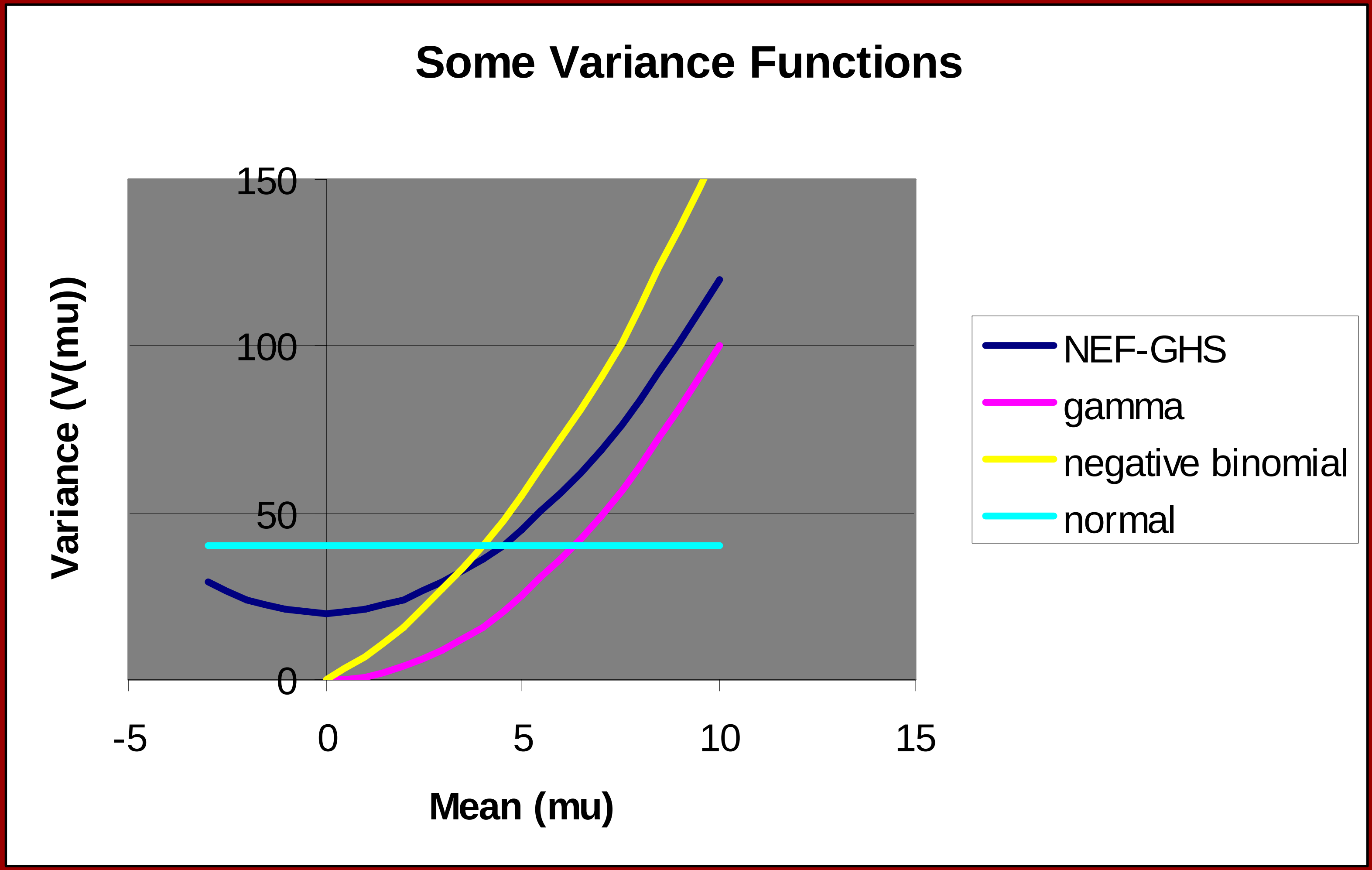
- Classifies natural exponential family distributions with quadratic variance functions ($\phi>0$ in what follows)
  - $V(\mu)=\phi$ [constant]
    - Normal with variance f and mean the parameter m
  - $V(\mu)=\phi(\mu-a)$ [linear]
    - Shifted [Over-/Under-]Dispersed Poisson
    - Distribution supported on a, $a+\phi$, $a+2\phi$, ….
    - m can be any real greater than a
  - $V(\mu)=\phi(\mu-a)^2$ [double root]
    - Shifted Gamma
    - Distribution and m both supported on reals greater than a

Morris's Theorem

- Classifies natural exponential family distributions with quadratic variance functions ($\phi>0$ in what follows)
  - $V(\mu)=(\mu-a)(b-(\mu-a))/n$ [two real roots, leading coeff<0]
    - Shifted rescaled binomial
    - n *must be* a positive integer, or there is no corresponding NEF…thus no ϕ here
    - Distribution supported on a, a + (b-a)/n, a + 2(b-a)/n,…,b
    - μ takes values on the open interval (a,b)
    - a=0, b=1, n=1 should look familiar
  - $V(\mu)=\phi(\mu-a)(b+(\mu-a))$ [two real roots, leading coeff>0]
    - Shifted [over-/under-]dispersed negative binomial
    - Takes values on a, a+ϕ, a+2ϕ, …
    - μ can be any real greater than a

Morris's Theorem, the NEF-GHS

- Classifies natural exponential family distributions with quadratic variance functions (ϕ>0 in what follows)
  - Finally, what about NO real roots?
  - $V(\mu)=\phi(\tau^2+(\mu-a)^2)$
    - Generalized Hyperbolic Secant distribution
    - Support is the real line, and μ takes values on all reals
      - Shares this in common with the normal, alone among quadratic variance NEFs
    - Best called NEF-GHS, to distinguish from other generalizations of the hyperbolic secant distribution that are *not* NEF
      - Although, all these generalization share:
        - Bi-infinite potentially skew, potentially heavy-tailed
        - The non-NEF generalizations can also be made to be light-tailed


Some Variance Functions
150
100
50
0
-5
0
5
10
15
Variance (V(mu))
Mean (mu)
NEF-GHS
gamma
negative binomial
normal

NEF-GHS

- In what follows, let us take a=0.
  - This means the distribution is symmetric when $\mu=0$
  - $V(\mu)=\phi(\tau^2+\mu^2)$
  - Note that we are adding $\phi\tau^2$ to what would be the constant CV variance (and ϕ would be CV^2)
  - Can also get from a negative binomial starting point if the constant added to the variance function is large enough (though in this case a<0)

Skewness and Kurtosis

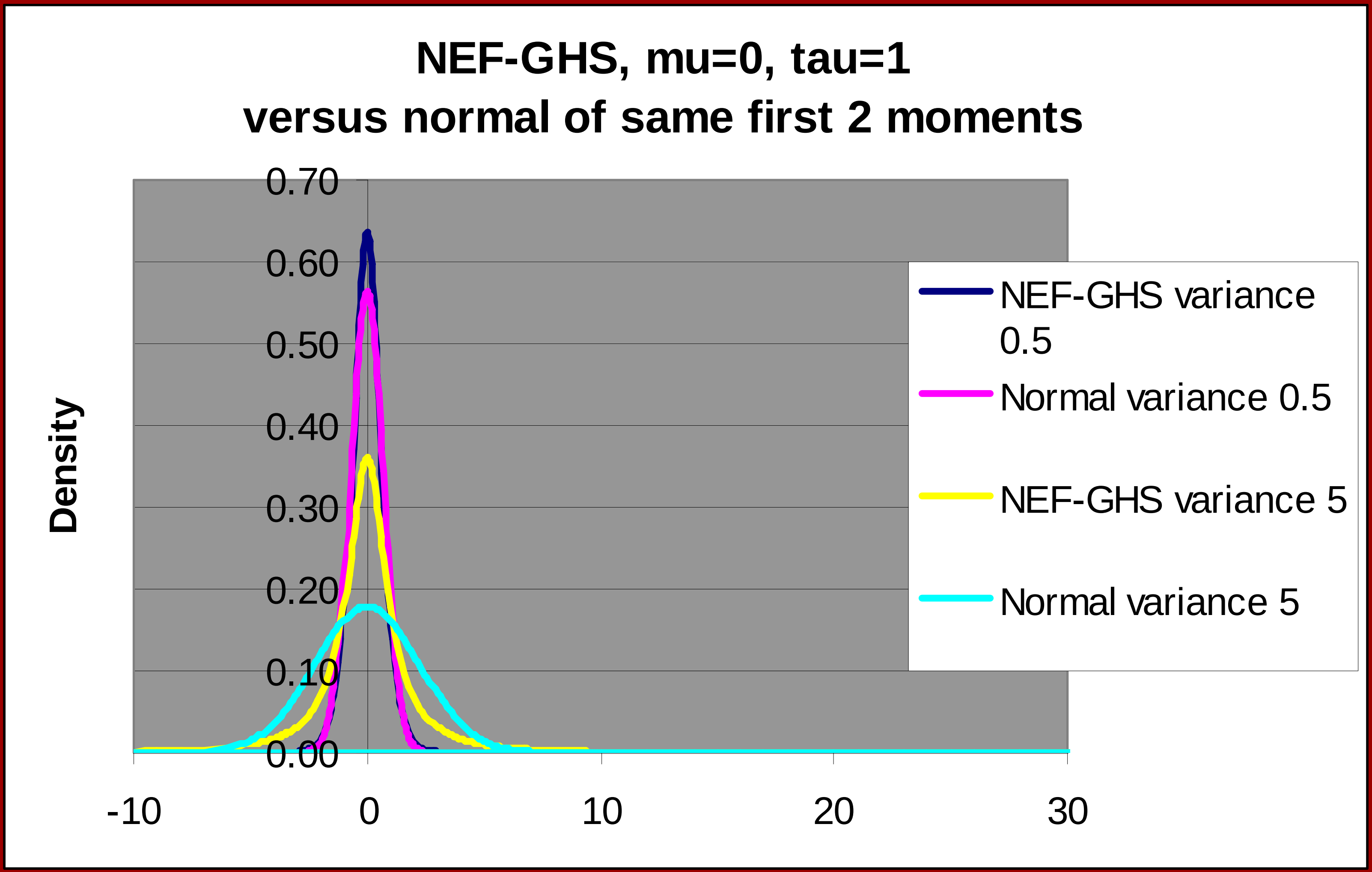
- The skewness is

$$\frac{2\sigma\mu}{\sqrt{\mu^2+\tau^2}}$$

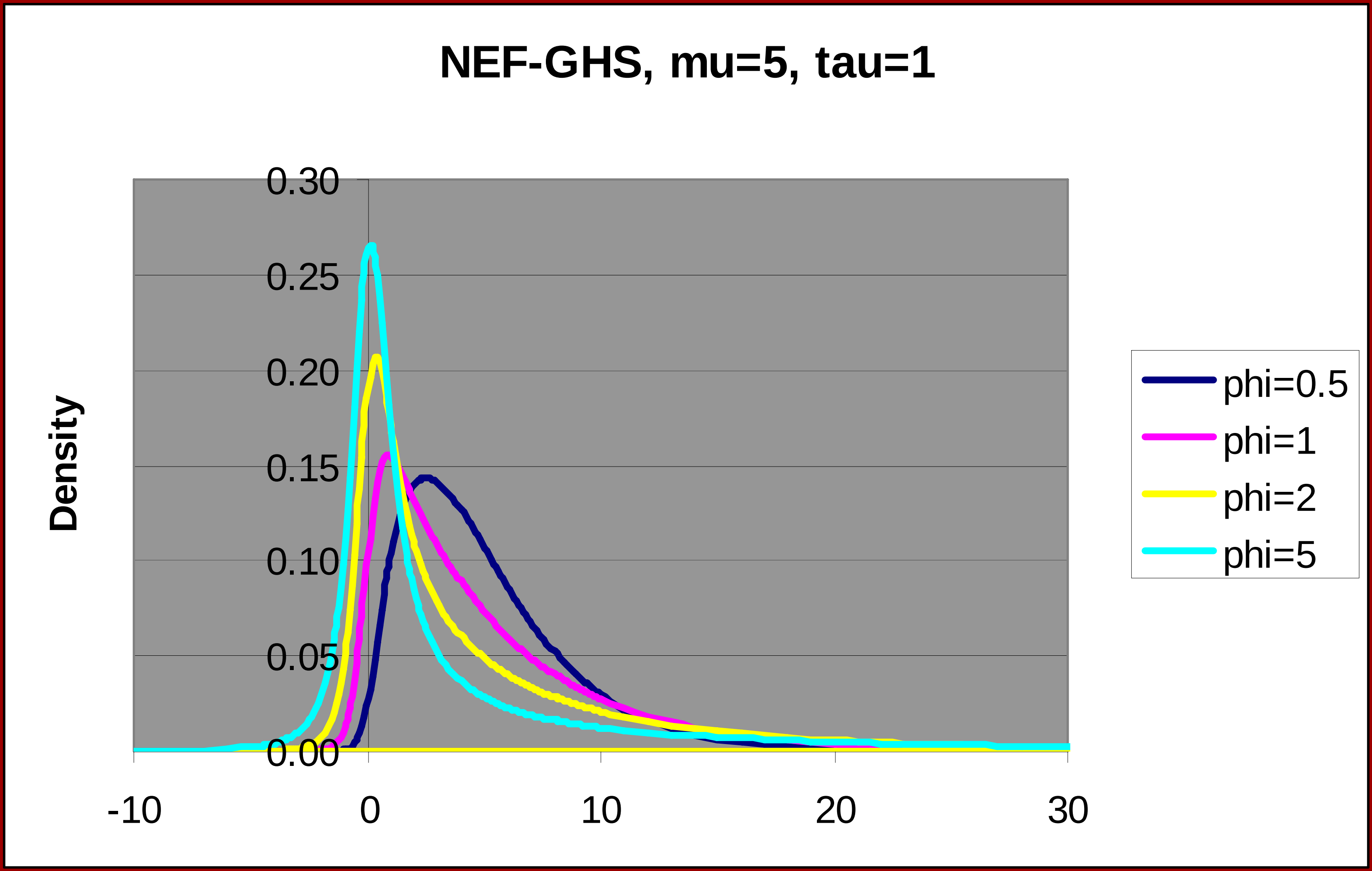
- Where $\sigma^2=\phi$
- The skewness approaches 2σ as μ increases
- The excess kurtosis (with normal = 0) is

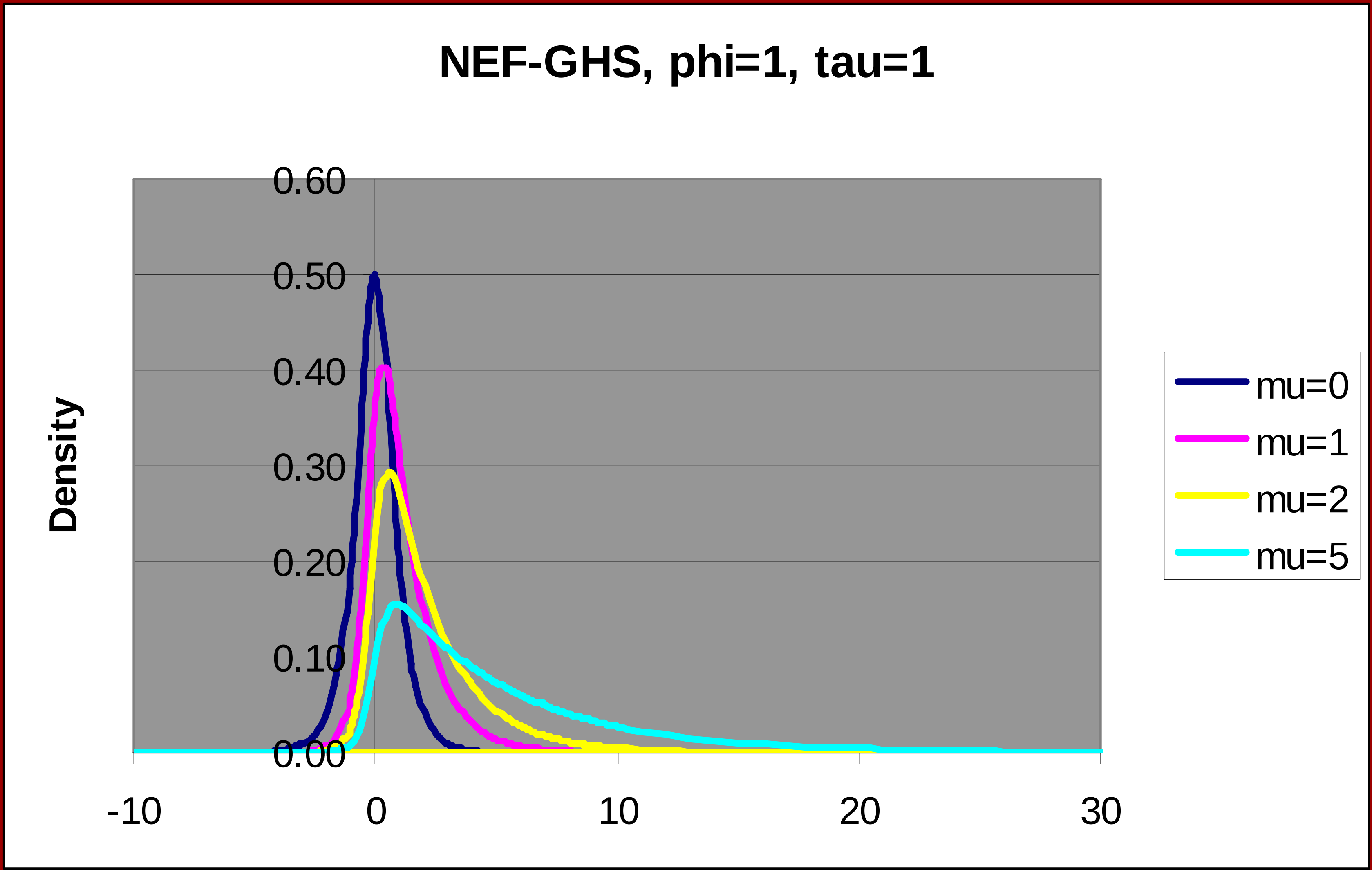
$$2\sigma^2\left(1+\frac{\mu^2}{\mu^2+\tau^2}\right)$$

- Note that this does *not* approach 0 for $\mu=0$



NEF-GHS, mu=0, tau=1
versus normal of same first 2 moments
Density
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-10
0
10
20
30
NEF-GHS variance 0.5
Normal variance 0.5
NEF-GHS variance 5
Normal variance 5


NEF-GHS, mu=5, tau=1
Density
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-10
0
10
20
30
phi=0.5
phi=1
phi=2
phi=5


NEF-GHS, phi=1, tau=1
Density
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-10
0
10
20
30
mu=0
mu=1
mu=2
mu=5

P(X<0) and Other Numeric Properties (for τ=1, a=0)

ϕ	μ	P(X<0)	Skewness	Excess Kurtosis
1	0	0.500	0.000	2.000
1	1	0.220	1.414	3.000
1	2	0.119	1.789	3.600
1	5	0.047	1.961	3.923
0.5	5	0.007	1.387	1.962
1	5	0.047	1.961	3.923
2	5	0.131	2.774	7.846
5	5	0.265	4.385	19.615

Note that allowing a to be different from zero allows flexible control of P(X<0)

Deviance

- Recall that the unit deviance has the form

$$\boxed{2\int_y^\mu \frac{y-t}{V(t)}dt}$$

- For the NEF-GHS, this reduces to

$$\frac{2y}{\tau}(\arctan(\frac{y}{\tau})-\arctan(\frac{\mu}{\tau}))+\ln(1+(\frac{\mu}{\tau})^2)-\ln(1+(\frac{y}{\tau})^2)$$

Simulation

- Set up:
  - Three continuous predictors x_1,x_2,x_3 drawn from a multivariate normal, positively correlated with each other, and capped at boundaries
    - Specifically, before capping they are mean 0, variance 2, and have pairwise correlations of 0.5. x_1 is then capped into the interval [-1,2], and the others into [0,1].
  - "Underlying" dependent variable
    - Mean is $1000e^{x1+x2+x3}$
    - is gamma-distributed with CV 2
  - Used log link except for simulation with negative values
  - Compare normal, gamma, and NEF-GHS models with various values of τ

Recall our Motivations

- Note that the "underlying" parameter values β_1, β_2, β_3 will always be 1.0
  - Don't necessarily expect to estimate those values exactly after data have been distorted—robustness is relative to gamma model
- Use simulations to observe behavior with respect to:
  - Measurement error
  - Appropriate cost function
  - Limits
  - Handling negative values

Measurement Error

- Added normal errors to the “underlying” dependent variable
  - The noise had a standard deviation of 500
- Dropped observation if resulting value not positive
- Since the gamma was chosen with $\phi=4$, expected that $\tau=250$ would handle this (then $\phi\tau^2=500^2$)
- Didn’t work that way

Parameter Estimates

τ	β_1	β_2	β_3
infinity (normal)	0.952	0.919	0.935
5000	0.913	0.897	0.916
2500	0.874	0.876	0.888
1000	0.811	0.820	0.822
500	0.766	0.765	0.756
250	0.742	0.729	0.713
0 (gamma)	0.731	0.711	0.692

Appropriate Cost Function

- Instead of adding a measurement error, 20% of the data kept as is, the other 80% replaced by much smaller random values (gamma with mean 100 and CV 2) that were independent of the covariates
- If you really cared about how the *larger* losses depended on the covariates, NEF-GHS did outperform gamma
  - And it did *not* require an extreme value of τ – this model very like a gamma model for the data in the range you care about

Parameter Estimates

τ	β_1	β_2	β_3
infinity (normal)	0.915	0.967	0.848
500	0.890	0.965	0.856
250	0.855	0.940	0.834
0 (gamma)	0.816	0.898	0.784

Limits

- Data simulated as in the “measurement error” section
- Half of observations chosen as “limited” and capped at $1,000 if their values is greater (and identified as x_4=1)
- This additional covariate is known and available to the model
- Note on the next slide that even the normal model is biased in predicting the overall mean
  - This is because the link function is not the identity
  - A Poisson assumption would yield unbiased predictions, given the log link

Predicted and Actual Means by X_4

τ	X_4=0 (unlimited)	X_4=1 ($1,000 cap)	Overall mean
infinity (normal)	$3,768	$300	$2,034
5000	$3,755	$313	$2,034
2500	$3,700	$336	$2,018
1000	$3,431	$458	$1,945
500	$3,057	$688	$1,873
250	$2,825	$872	$1,849
0 (gamma)	$2,739	$955	$1,847
Actual means	$3,822	$638	$2,230

Negative Values

- Set up as for measurement errors, but negative observations not dropped
- To keep multiplicative model on positive values, used link:
  - Log($x\beta$) if $x\beta > 1$
  - $x\beta-1$ if $x\beta <= 1$
- Can't even compare to gamma, so just compared to normal, with the same link function
- No point showing a table of parameter estimates
  - They all came out close to 1.000 in all case
  - Increased the size of the measurement errors ten-fold
    - Only effect was NEF-GHS with a very small value of τ (less than the size of the measurement error) did less well
    - The funky link function worked very well!

Possible Extensions

- Could "adjust" other GLMs also
- Variance functions tangent to the x-axis perhaps merit special consideration for this
- All NEFs with cubic variance functions have also been classified
- Thus, can "adjust" the inverse Gaussian
  - Can't accommodate bi-infinite dependent variable, though
    - Since a cubic function must cross the x-axis
  - But can add a linear term and make it not be tangent
    - See Letac & Mora

Conclusions

- Using the NEF-GHS distribution in models represents a compromise between gamma and normal assumptions
- Unlike the Poisson, which is a different type of compromise, NEF-GHS looks like the normal for relatively small predicted values and like the gamma for relatively large ones
- NEF-GHS is thus more robust against certain phenomenon than are gamma models
- NEF-GHS may have use in reserve variability methodology because, unlike many distributions used in such methodologies, it
    - Accommodates continuous dependent variables
    - Accommodates negative dependent variable values (and increments in triangles can be negative), and yet is still heavy-tailed and skew

References

- Bent Jørgensen, *The Theory of Dispersion Models*, 1997
- Gérard Letac, Marianne Mora, “Natural Real Exponential Families with Cubic Variance Functions”, *Ann. Stat.* 18 (1990)
- Carl N. Morris, “Natural Exponential Families with Quadratic Variance Function”, *Ann. Stat.* 10 (1982) and *Ann. Stat.* 11 (1983)