Stochastic Loss Reserving Using Bayesian MCMC Models

Glenn Meyers – FCAS, MAAA, CERA, Ph.D.

Casualty Loss Reserve Seminar
September 17, 2013

Background

- Risk based capital proposals, e.g. EU Solvency II and USA SMI rely on stochastic models.
  - VaR@99.5% and TVaR@99%
- There are many stochastic loss reserve models that claim to predict the distribution of ultimate losses.

How good are these models?

- This presentation describes tests of the predictions of currently popular stochastic loss reserve models on real data from 50 insurers in each of four lines of insurances.
- It proposes two new models that improve the predictions.

The CAS Loss Reserve Database Created by Meyers and Shi
With Permission of American NAIC

- Schedule P (Data from Parts 1-4) for several US Insurers
  - Private Passenger Auto
  - Commercial Auto
  - Workers' Compensation
  - General Liability
  - Product Liability
  - Medical Malpractice (Claims Made)
- Available on CAS Website

  http://www.casact.org/research/index.cfm?fa=loss_reserves_data

Notation

- w = Accident Year $w = 1,\ldots,10$
- d = Development Year $d = 1,\ldots,10$
- $C_{w,d}$ = Cumulative (either incurred or paid) loss
- $I_{w,d}$ = Incremental paid loss $= C_{w,d} - C_{w-1,d}$

Illustrative Insurer – Incurred Losses

		Cumulative Incurred Losses										
Premium	AY/Lag	1	2	3	4	5	6	7	8	9	10	Source
5812	1988	1722	3830	3603	3835	3873	3895	3918	3918	3917	3917	1997
4908	1989	1581	2192	2528	2533	2528	2530	2534	2541	2538	2532	1998
5454	1990	1834	3009	3488	4000	4105	4087	4112	4170	4271	4279	1999
5165	1991	2305	3473	3713	4018	4295	4334	4343	4340	4342	4341	2000
5214	1992	1832	2625	3086	3493	3521	3563	3542	3541	3541	3587	2001
5230	1993	2289	3160	3154	3204	3190	3206	3351	3289	3267	3268	2002
4992	1994	2881	4254	4841	5176	5551	5689	5683	5688	5684	5684	2003
5466	1995	2489	2956	3382	3755	4148	4123	4126	4127	4128	4128	2004
5226	1996	2541	3307	3789	3973	4031	4157	4143	4142	4144	4144	2005
4962	1997	2203	2934	3608	3977	4040	4121	4147	4155	4183	4181	2006

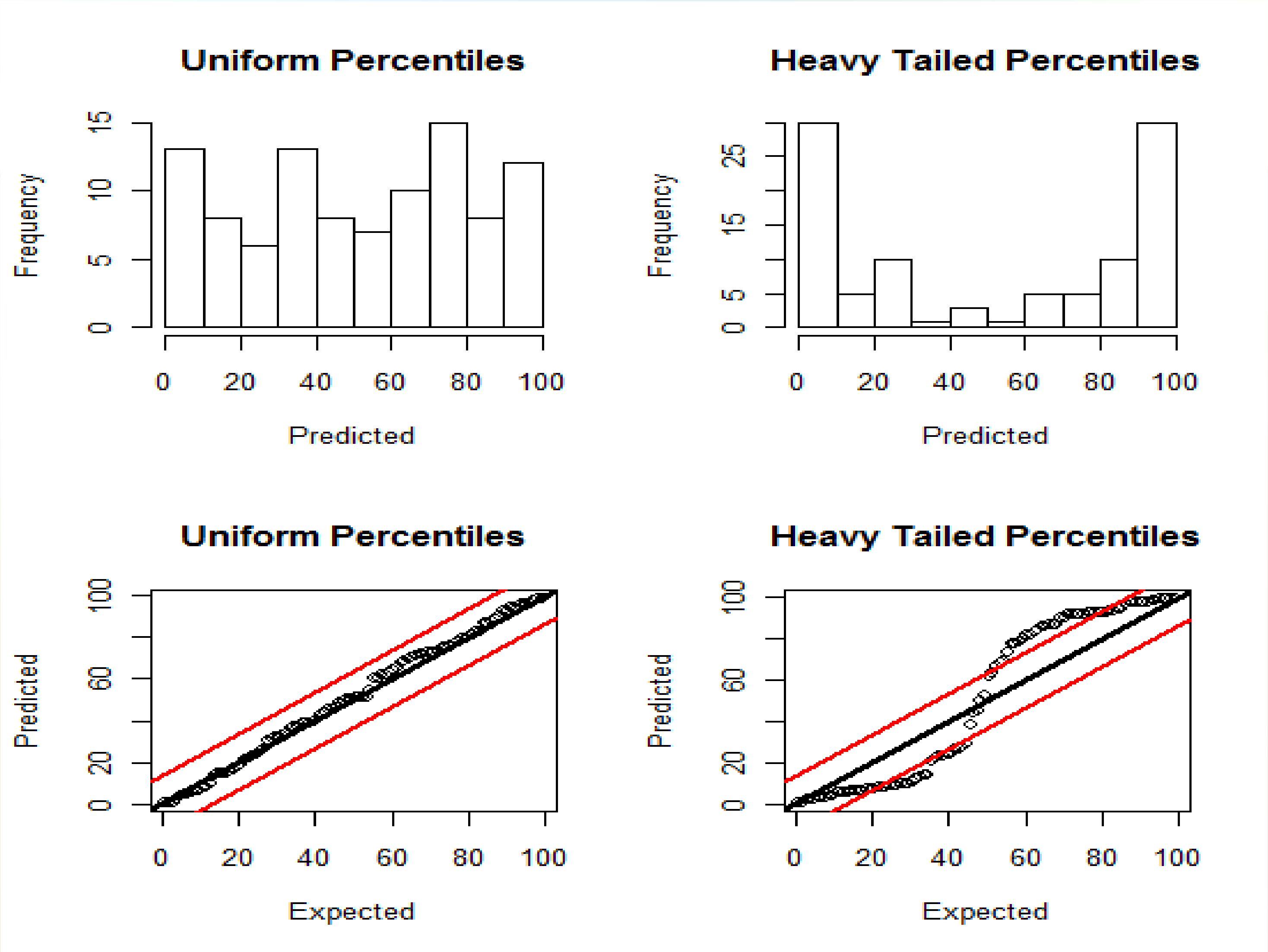
Illustrative Insurer – Paid Losses

		Cumulative Paid Losses										
Premium	AY/Lag	1	2	3	4	5	6	7	8	9	10	Source
5812	1988	952	1529	2813	3647	3724	3832	3899	3907	3911	3912	1997
4908	1989	849	1564	2202	2432	2468	2487	2513	2526	2531	2527	1998
5454	1990	983	2211	2830	3832	4039	4065	4102	4155	4268	4274	1999
5165	1991	1657	2685	3169	3600	3900	4320	4332	4338	4341	4341	2000
5214	1992	932	1940	2626	3332	3368	3491	3531	3540	3540	3583	2001
5230	1993	1162	2402	2799	2996	3034	3042	3230	3238	3241	3268	2002
4992	1994	1478	2980	3945	4714	5462	5680	5682	5683	5684	5684	2003
5466	1995	1240	2080	2607	3080	3678	4116	4117	4125	4128	4128	2004
5226	1996	1326	2412	3367	3843	3965	4127	4133	4141	4142	4144	2005
4962	1997	1413	2683	3173	3674	3805	4005	4020	4095	4132	4139	2006

Criteria for a "Good" Stochastic Loss Reserve Model

- Using the upper triangle "training" data, predict the distribution of the outcomes in the lower triangle
  - Can be observations from individual (AY, Lag) cells or sums of observations in different (AY,Lag) cells.
- Using the predictive distributions, find the percentiles of the outcome data.
- The percentiles should be uniformly distributed.
  - Histograms
  - Test with PP Plots/KS tests
    - Plot Expected vs Predicted Percentiles
    - KS 95% critical values = $\frac{136}{\sqrt{n}}$ =19.2 for n = 50 and 9.6 for n = 200

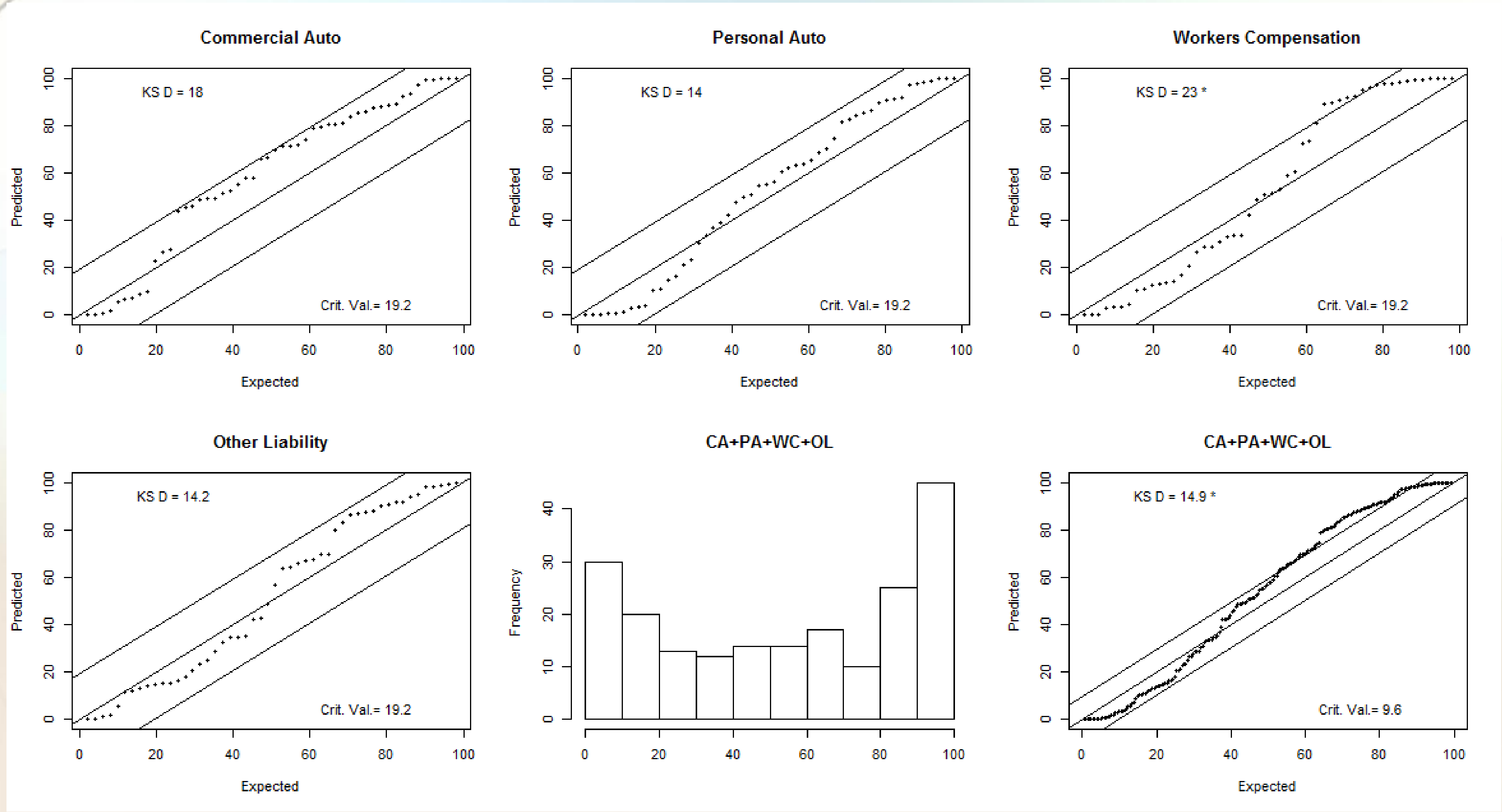
Illustrative Tests of Uniformity



Data Used in Study

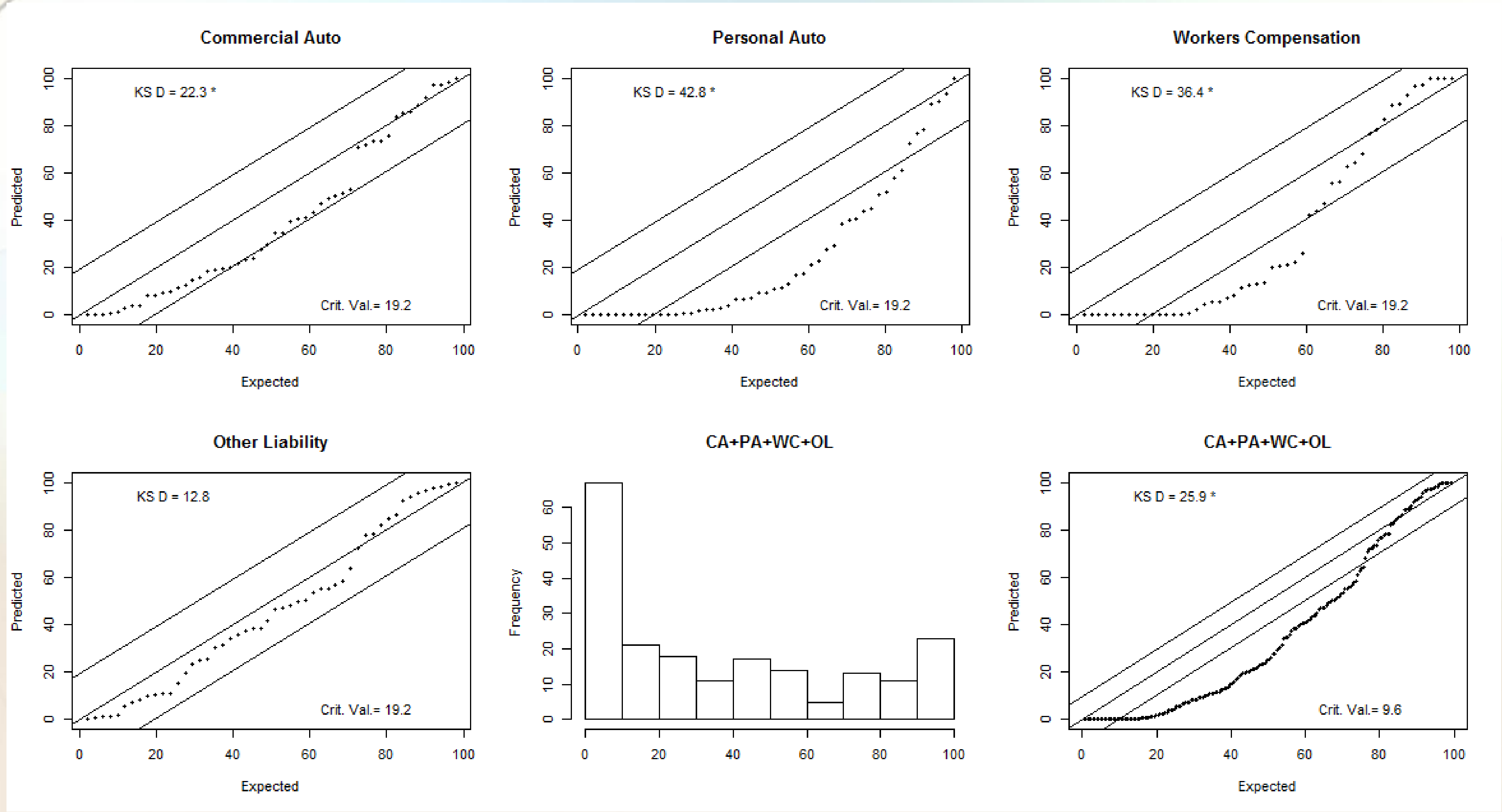
- Insurers listed in Meyers – Summer 2012 e-Forum
- 50 Insurers from four lines of business
  - Commercial Auto
  - Personal Auto
  - Workers’ Compensation
  - Other Liability
- Both paid and incurred losses

Test of Mack Model on Incurred Data



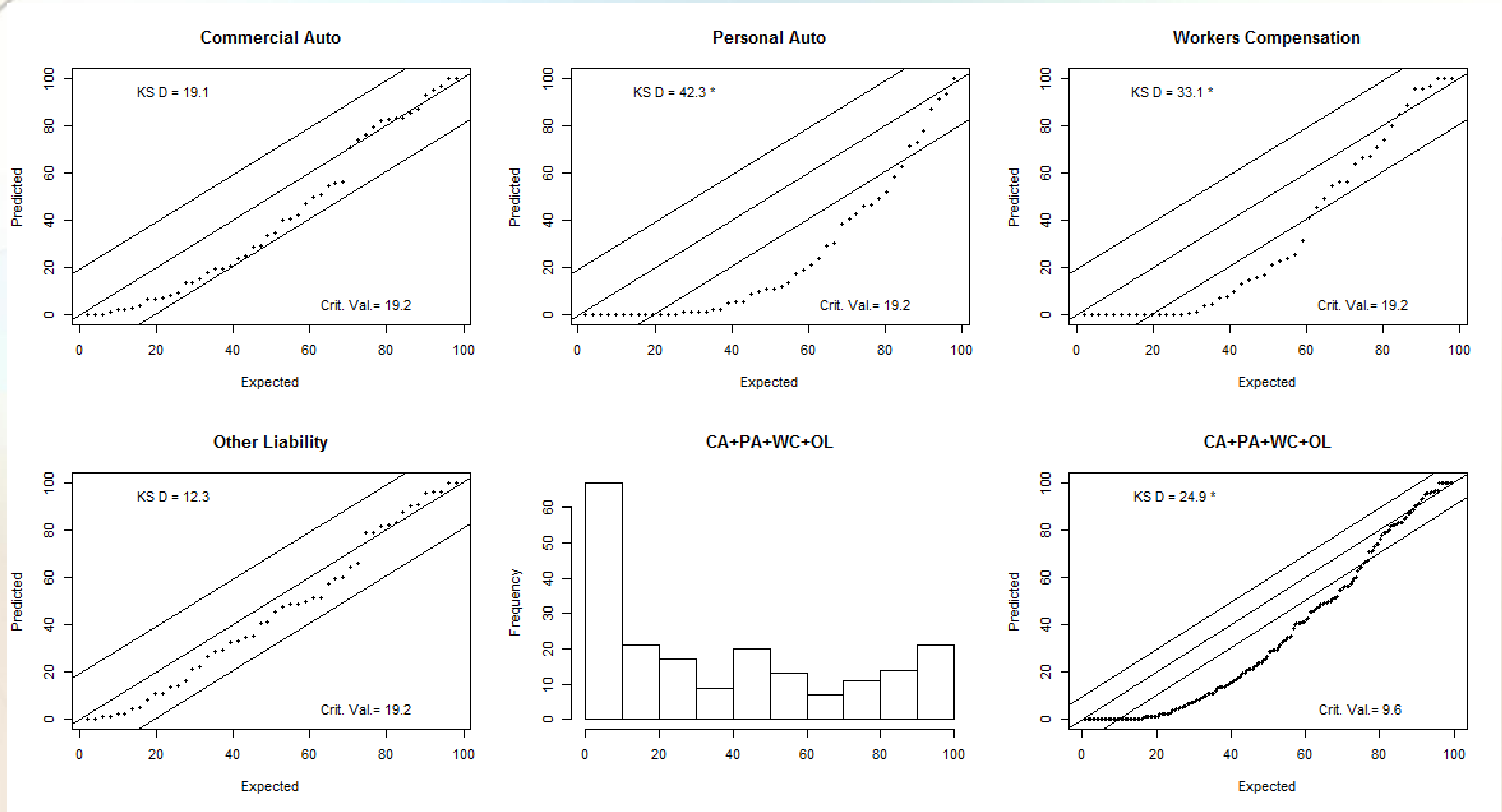
Conclusion – The Mack model predicts tails that are too light.

Test of Mack Model on Paid Data



Conclusion – The Mack model is biased upward.

Test of Bootstrap ODP on Paid Data



Conclusion – The Bootstrap ODP model is biased upward.

Possible Responses to the model failures

- The “Black Swans” got us again!
  - We do the best we can in building our models, but the real world keeps throwing curve balls at us.
  - Every few years, the world gives us a unique “black swan” event.
- Build a better model.
  - Use a model, or data, that sees the “black swans.”
  - Proposed models are Bayesian
    - Computations done by Bayesian Markov-Chain Monte Carlo (MCMC) simulations.

The Problem With Bayesian Analyses

Particularly Applicable to Loss Reserving

- Let θ be an *n*-parameter vector (e.g. development factors).
- Let *X* be a set of observations (e.g. a loss development triangle).

$$f(\theta \mid X) = \frac{f(X \mid \theta) \cdot \pi(\theta)}{\int_{\vartheta_1} \cdots \int_{\vartheta_n} f(X \mid \vartheta) \cdot \pi(\vartheta) \cdot d\vartheta}$$

- $f(X \mid \theta)$ is the likelihood of *X* given θ.
- $\pi(\theta)$ is the prior distribution of θ.
- $f(\theta \mid X)$ is the posterior distribution of θ.
- Calculating the *n*-dimensional integral is intractable.

A New World Order

- This impasse came to an end ~1990 when a simulation-based approach to estimating posterior probabilities was introduced.
    - (Circa the fall of the Soviet empire and Francis Fukuyama's "end of history")

Sampling-Based Approaches to Calculating Marginal Densities

ALAN E. GELFAND AND ADRIAN F. M. SMITH*

© 1990 American Statistical Association
Journal of the American Statistical Association
June 1990, Vol. 85, No. 410, Theory and Methods

Markov Chains

- Let Ω be a finite state with random events $X_1, X_2, \ldots, X_t, \ldots$
- A Markov chain P satisfies $\Pr\{X_t = y | X_{t-1}=x, \ldots, X_1=x_1\} = \Pr\{x_t = y | x\} \equiv P(x,y)$
  - The probability of an event in the chain depends only on the immediate previous event.
  - P is called a transition matrix

The Markov Convergence Theorem

- There is a branch of probability theory, called Ergodic Theory, that gives conditions for which there exists a unique stationary distribution π such that $P^t(x,y) \to \pi(y)$ as $t \to \infty$.
- Counterexamples that do not satisfies these conditions.
  - Periodic paths
  - Absorption states – Once a chain enters one of these states, it does not leave that group of states.
- Jackman[1] (Section 5.1.1) demonstrates that the Markov chain defined by the Metropolis Hastings algorithm satisfies the conditions of the Markov Convergence Theorem. Moreover the stationary distribution, π, is the posterior distribution.

1. Simon Jackman, *Bayesian Analysis for the Social Scientists*, Wiley - 2009

The Metropolis Hastings Algorithm
A Very Important Markov Chain

1. Time t=1: select a random initial position θ_1 in parameter space.
2. Select a **proposal distribution** $p(\theta|\theta_{t-1})$ that we will use to select proposed random steps away from our current position in parameter space.
3. Starting at time t=2: repeat the following until you get convergence:

a) At step t, generate a proposal $\theta^* \sim p(\theta|\theta_{t-1})$

b) Generate $U \sim$ uniform(0,1)

c) Calculate $$R = \frac{f(\Theta^* \mid X)}{f(\Theta_{t-1} \mid X)} \cdot \frac{p(\Theta_{t-1} \mid \Theta^*)}{p(\Theta^* \mid \Theta_{t-1})}$$

d) If $U < R$ then $\theta_t = \theta^*$. Else, $\theta_t = \theta_{t-1}$.

Dodging the Intractable Integral

$$R = \frac{f(\Theta^* \mid X)}{f(\Theta_{t-1} \mid X)} \cdot \frac{p(\Theta_{t-1} \mid \Theta^*)}{f(\Theta^* \mid \Theta_{t-1})}$$

$$R = \frac{\dfrac{f\left(X \mid \Theta^*\right) \cdot \pi\left(\Theta^*\right)}{\cancel{\int_{\vartheta_1} \cdots \int_{\vartheta_n} f(X \mid \vartheta) \cdot \pi(\vartheta) \cdot d\vartheta}}}{\dfrac{f\left(X \mid \Theta_{t-1}\right) \cdot \pi\left(\Theta_{t-1}\right)}{\cancel{\int_{\vartheta_1} \cdots \int_{\vartheta_n} f(X \mid \vartheta) \cdot \pi(\vartheta) \cdot d\vartheta}}} \cdot \frac{p(\Theta_{t-1} \mid \Theta^*)}{p(\Theta^* \mid \Theta_{t-1})}$$

The Metropolis Hastings Algorithm Restated

1. Time t=1: select a random initial position θ_1 in parameter space.
2. Select a **proposal distribution** $p(\theta|\theta_{t-1})$ that we will use to select proposed random steps away from our current position in parameter space.
3. Starting at time t=2: repeat the following until you get convergence:
   a) At step t, generate a proposed $\theta^* \sim p(\theta|\theta_{t-1})$
   b) Generate $U \sim$ uniform(0,1)
   c) Calculate $$R = \frac{f(X|\theta^*)\pi(\theta^*)}{f(X|\theta_{t-1})\pi(\theta_{t-1})} \cdot \frac{p(\theta_{t-1}|\theta^*)}{p(\theta^*|\theta_{t-1})}$$
   d) If $U < R$ then $\theta_t = \theta^*$. Else, $\theta_t = \theta_{t-1}$.

Metropolis Hastings in Practice

- "Tune" the proposal distribution, $p(\theta|\theta_{t-1})$, to minimize autocorrelation between θ_t and θ_{t-1}.
- Convergence - Determine the interval t to $t+m$ that contains a representative sample of the posterior distribution.
- There are several software packages for Bayesian MCMC that work with the R programming language.
  - WINBUGS
  - OpenBUGS
  - JAGS
  - Stan (New)
- There was a CLRS workshop last Sunday that covered the nuts and bolts of Bayesian MCMC for loss reserving and other analyses.

Bayesian MCMC Models

- Use R and JAGS (Just Another Gibbs Sampler) packages
- Get a sample of 10,000 parameter sets from the posterior distribution of the model
- Use the parameter sets to get 10,000 simulated outcomes
- Calculate summary statistics of the simulated outcomes
  - Mean
  - Standard deviation
  - Percentile of the actual outcome

The Correlated Chain Ladder (CCL) Model

- $logelr$ ~ uniform(-5,0)
- α_w ~ normal(log(Premium$_w$)+$logelr$,$\sqrt{10}$) – a wide distribution
- $\beta_1 = 0$, β_d ~ uniform(-5,5), for d=2,…,10 – a wide distribution
- $\mu_{1,d} = \alpha_1 + \beta_d$
- a_i ~ uniform(0,1)
- $\sigma_d = \sum_{i=d}^{10} a_i$ Forces σ_d to decrease as d increases
- $C_{1,d}$ ~ lognormal($\mu_{1,d}$, σ_d)
- ρ ~ uniform(-1,1)
- $\mu_{w,d} = \alpha_w + \beta_d + \rho \cdot (\log(C_{w-1,d}) - \mu_{w-1,d})$ for w = 2,…,10
- $C_{w,d}$ ~ lognormal($\mu_{w,d}$, σ_d)

The First 5 of 10,000 Samples on Illustrative Insurer

Done in JAGS

MCMC Sample Number

	1	2	3	4	5
α_1	7.5798	7.6012	7.6129	7.6185	7.5977
α_2	7.1464	7.1700	7.1717	7.1845	7.1849
α_3	7.6415	7.6433	7.6653	7.6778	7.6400
α_4	7.6502	7.6931	7.7027	7.7307	7.7268
α_5	7.4946	7.4769	7.5035	7.5450	7.5675
α_6	7.4059	7.4748	7.4473	7.4062	7.5084
α_7	7.8438	7.8985	8.0043	7.9140	7.8849
α_8	7.7122	7.5296	7.5820	7.5887	7.6104
α_9	7.8490	7.7618	7.6659	7.6620	7.5529
α_{10}	7.5480	7.6489	7.5134	7.3747	7.5481
β_1	0	0	0	0	0
β_2	0.4159	0.4538	0.4848	0.4370	0.4671
β_3	0.4175	0.5576	0.5025	0.5347	0.5758
β_4	0.6403	0.6190	0.6283	0.6448	0.6188
β_5	0.6861	0.6710	0.6530	0.6315	0.6441
β_6	0.6934	0.6854	0.6628	0.6377	0.6261
β_7	0.7095	0.6715	0.6628	0.6337	0.6926
β_8	0.6889	0.7025	0.6629	0.6364	0.6622
β_9	0.6894	0.6665	0.6590	0.6626	0.6574
β_{10}	0.6863	0.6538	0.6649	0.6508	0.6826
σ_1	0.1912	0.1683	0.2205	0.3440	0.1743
σ_2	0.1628	0.1543	0.1301	0.2260	0.1285
σ_3	0.1115	0.0778	0.0981	0.0942	0.0684
σ_4	0.0427	0.0455	0.0420	0.0518	0.0568
σ_5	0.0393	0.0357	0.0216	0.0252	0.0346
σ_6	0.0358	0.0195	0.0173	0.0240	0.0337
σ_7	0.0328	0.0180	0.0142	0.0225	0.0299
σ_8	0.0242	0.0178	0.0137	0.0207	0.0295
σ_9	0.0155	0.0128	0.0108	0.0087	0.0187
σ_{10}	0.0091	0.0102	0.0068	0.0064	0.0137
ρ	0.3443	0.2373	0.0854	0.2218	0.1554

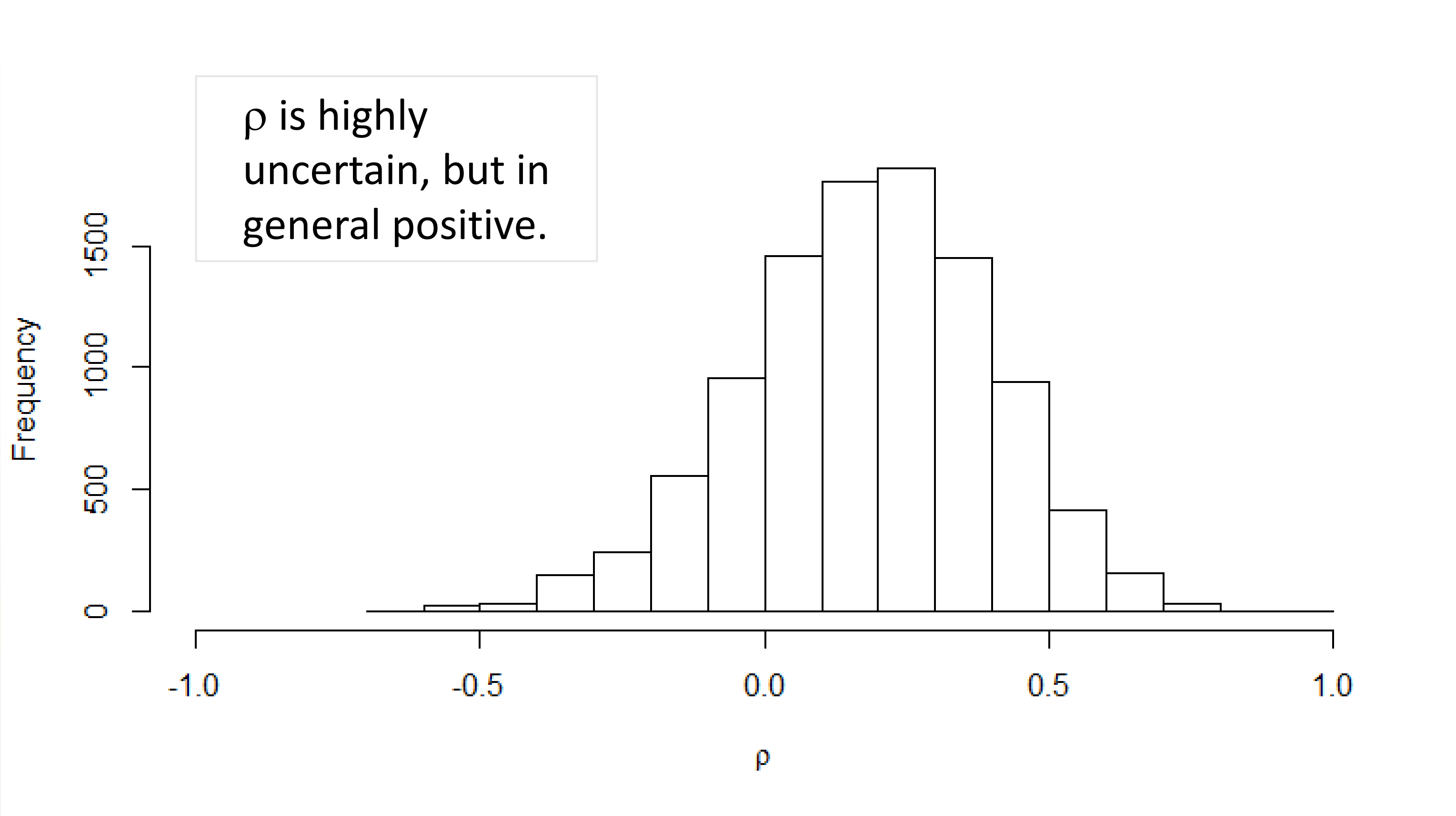
Done in R

$\mu_{1,10}$	8.2661	8.2550	8.2778	8.2694	8.2803
$C_{1,10}$	3917	3917	3917	3917	3917
$\tilde{\mu}_{2,10}$	7.8351	7.8281	7.8362	7.8362	7.8664
$\tilde{C}_{2,10}$	2541	2528	2529	2523	2630
$\tilde{\mu}_{3,10}$	8.3296	8.2989	8.3301	8.3280	8.3239
$\tilde{C}_{3,10}$	4151	4041	4110	4171	4066
$\tilde{\mu}_{4,10}$	8.3370	8.3482	8.3668	8.3833	8.4073
$\tilde{C}_{4,10}$	4194	4243	4311	4384	4485
$\tilde{\mu}_{5,10}$	8.1824	8.1319	8.1686	8.1963	8.2503
$\tilde{C}_{5,10}$	3574	3365	3486	3628	3953
$\tilde{\mu}_{6,10}$	8.0919	8.1261	8.1111	8.0570	8.1960
$\tilde{C}_{6,10}$	3231	3424	3343	3158	3648
$\tilde{\mu}_{7,10}$	8.5262	8.5553	8.6695	8.5650	8.5685
$\tilde{C}_{7,10}$	4975	5124	5765	5236	5203
$\tilde{\mu}_{8,10}$	8.3936	8.1802	8.2460	8.2391	8.2912
$\tilde{C}_{8,10}$	4396	3602	3847	3813	4094
$\tilde{\mu}_{9,10}$	8.5334	8.4178	8.3316	8.3144	8.2395
$\tilde{C}_{9,10}$	5098	4451	4195	4099	3781
$\tilde{\mu}_{10,10}$	8.2354	8.2987	8.1792	8.0265	8.2305
$\tilde{C}_{10,10}$	3721	4056	3601	3050	3707

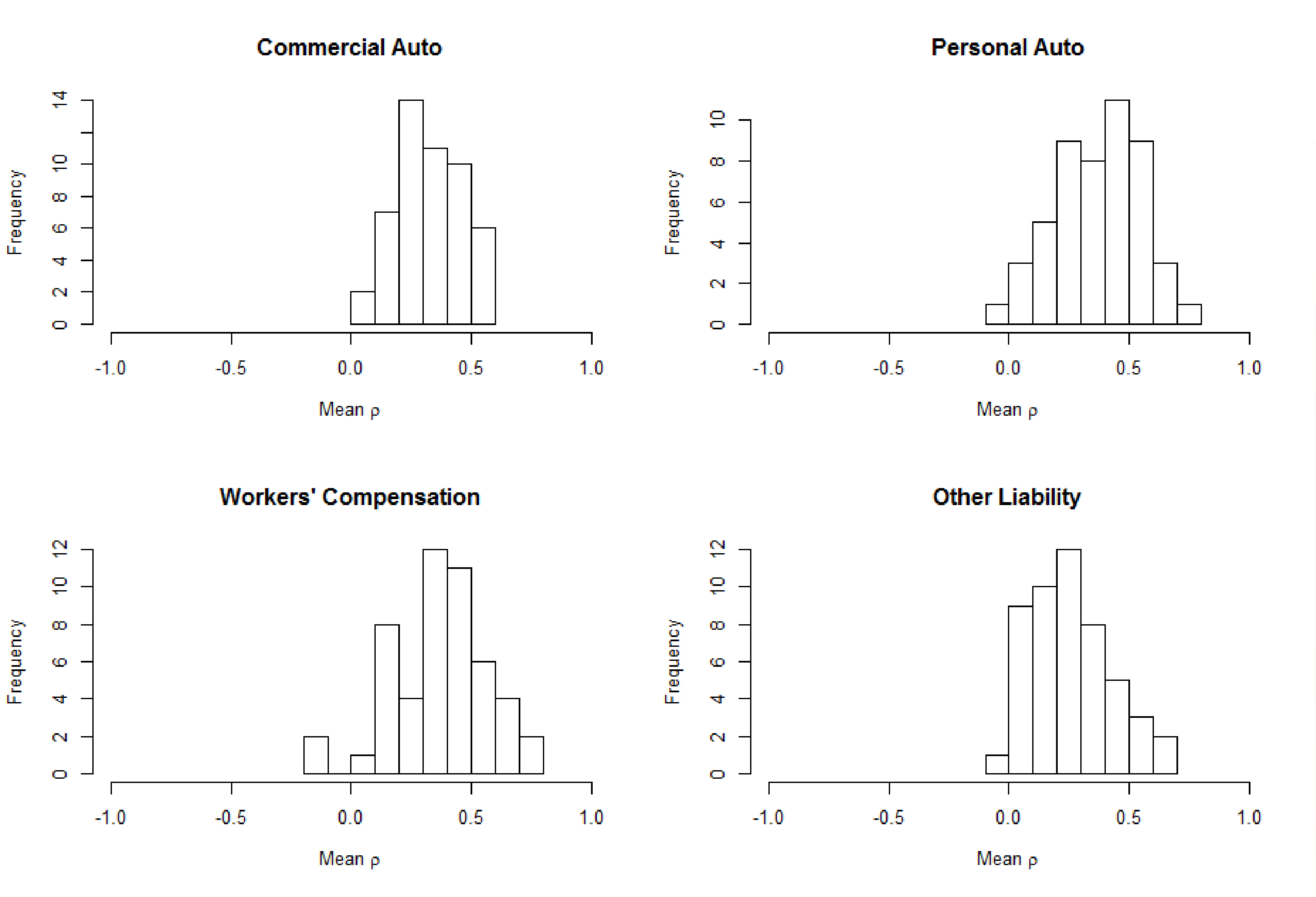
The Correlated Chain Ladder Model Predicts Distributions with Thicker Tails

- Mack uses point estimations of parameters.
- CCL uses Bayesian estimation to get a posterior distribution of parameters.
- Chain ladder applies factors to last ***fixed*** observation.
- CCL uses ***uncertain*** “level” parameters for each accident year.
- Mack assumes independence between accident years.
- CCL allows for correlation between accident years,
  - $Corr[\log(C_{w-1,d}),\log(C_{w,d})] = \rho$

Posterior Distribution of ρ for Illustrative Insurer



Generally Positive Posterior Means of ρ



Predicting the Distribution of Outcomes

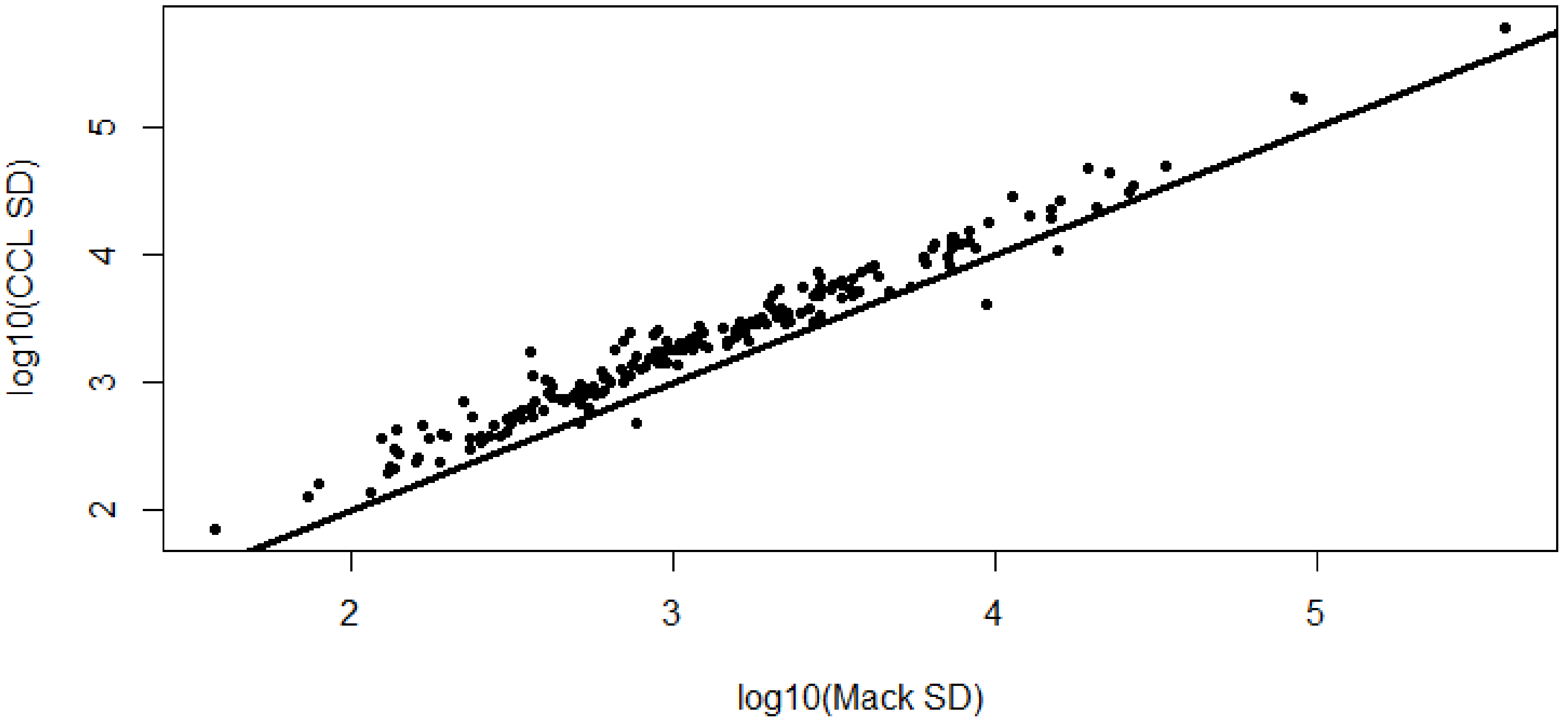
- Use JAGS software to produce a sample of 10,000 $\{\alpha_w\}$, $\{\beta_d\}$,$\{\sigma_d\}$ and $\{\rho\}$ from the posterior distribution.
- For each member of the sample
  - $\mu_{1,10} = \alpha_1 + \beta_{10}$
  - For w = 2 to 10
    - $C_{w,10}$ = random lognormal $(\alpha_w + \beta_{10} + \rho\cdot(\log(C_{w-1,10}) - \mu_{w-1})),\sigma_{10})$
  - Calculate $\sum_{w=1}^{10} C_{w,10}$
- Calculate summary statistics, e.g. $E\left[\sum_{w=1}^{10} C_{w,10}\right]$ and $Var\left[\sum_{w=1}^{10} C_{w,10}\right]$
- Calculate the percentile of the actual outcome by counting how many of the simulated outcomes are below the actual outcome.

Results for the Illustrative Incurred Data

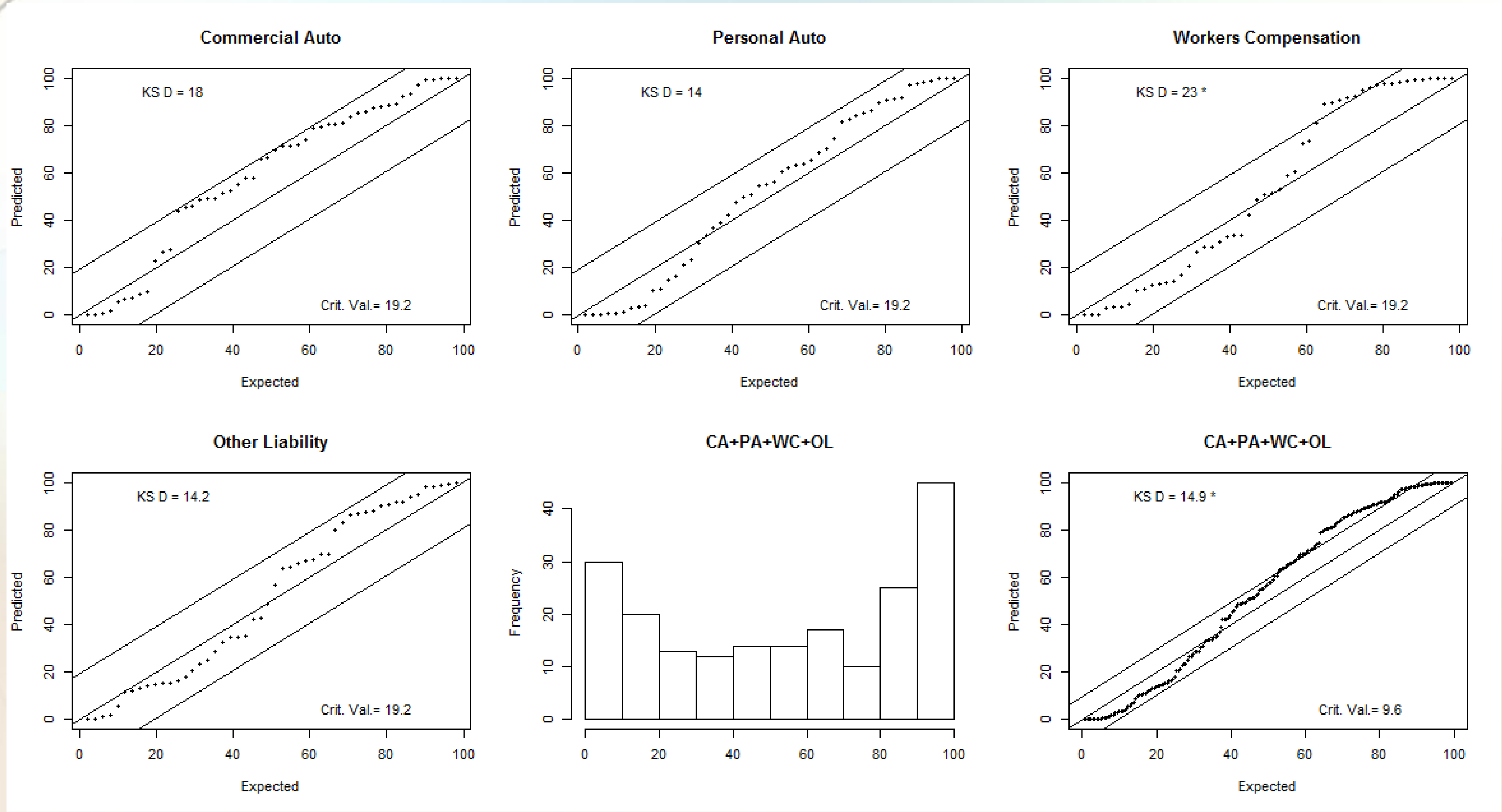
	CCL				Mack				Outcome
w	C_{w,10}	SD	CV		C_{w,10}	SD	CV		C_{w,10}
1	3,917	0	0.0000		3,917	0	0.0000		3,917
2	2,546	62	0.0244		2,538	0	0.0000		2,532
3	4,111	119	0.0289		4,167	3	0.0007		4,279
4	4,316	136	0.0315		4,367	37	0.0085		4,341
5	3,552	126	0.0355		3,597	34	0.0095		3,587
6	3,321	150	0.0452		3,236	40	0.0124		3,268
7	5,285	295	0.0558		5,358	146	0.0272		5,684
8	3,805	335	0.0880		3,765	225	0.0598		4,128
9	4,180	615	0.1471		4,013	412	0.1027		4,144
10	4,141	1,371	0.3311		3,955	878	0.2220		4,181
Total	39,174	1,869	0.0477		38,914	1,057	0.0272		40,061
Percentile		73.40				86.03			

Note the increase in the standard error of CCL over Mack.

Compare SDs for All 200 Triangles

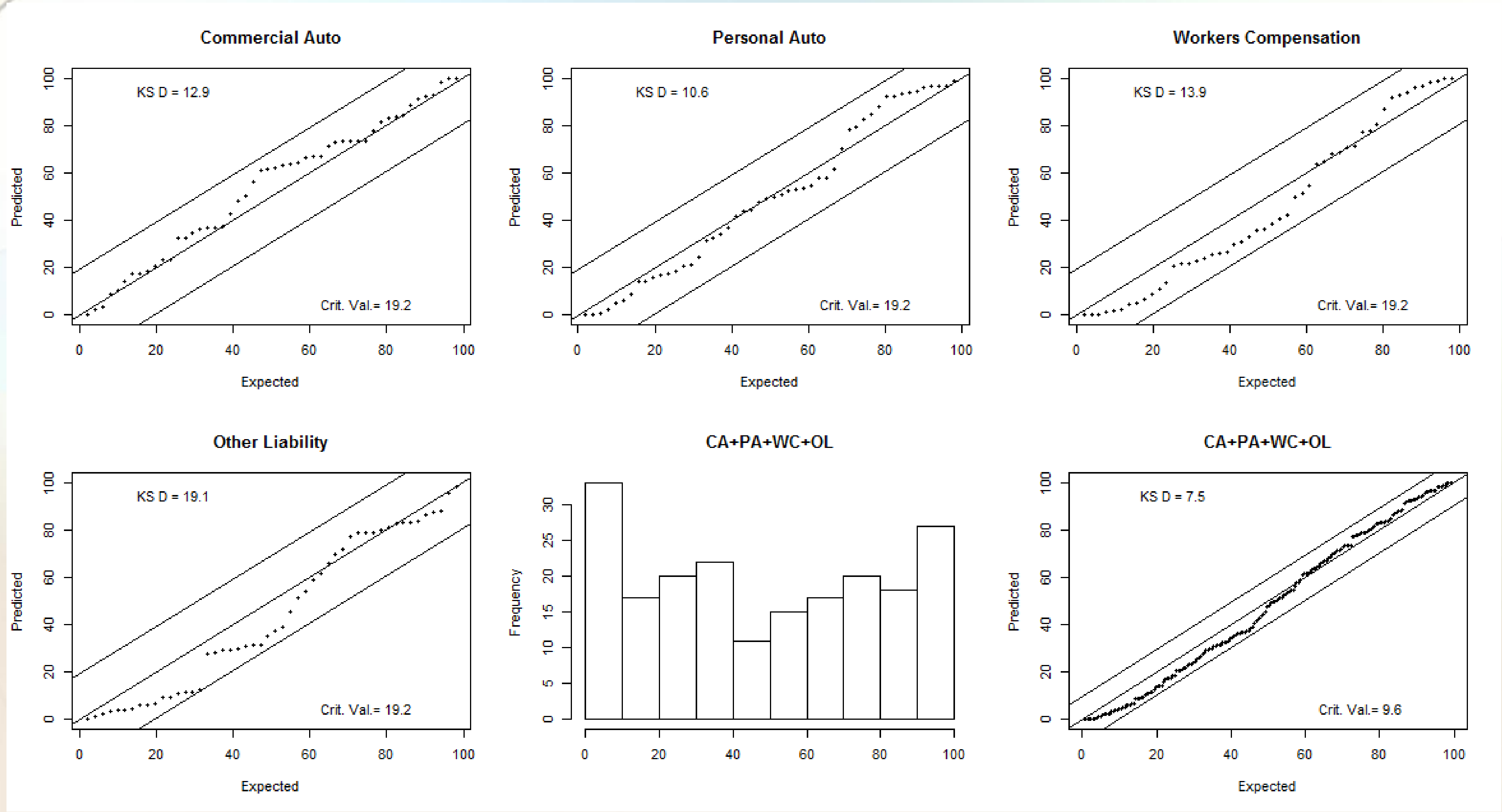


Test of Mack Model on Incurred Data



Conclusion – The Mack model predicts tails that are too light.

Test of CCL on Incurred Data



Conclusion – CCL model percentiles lie within KS statistical bounds.

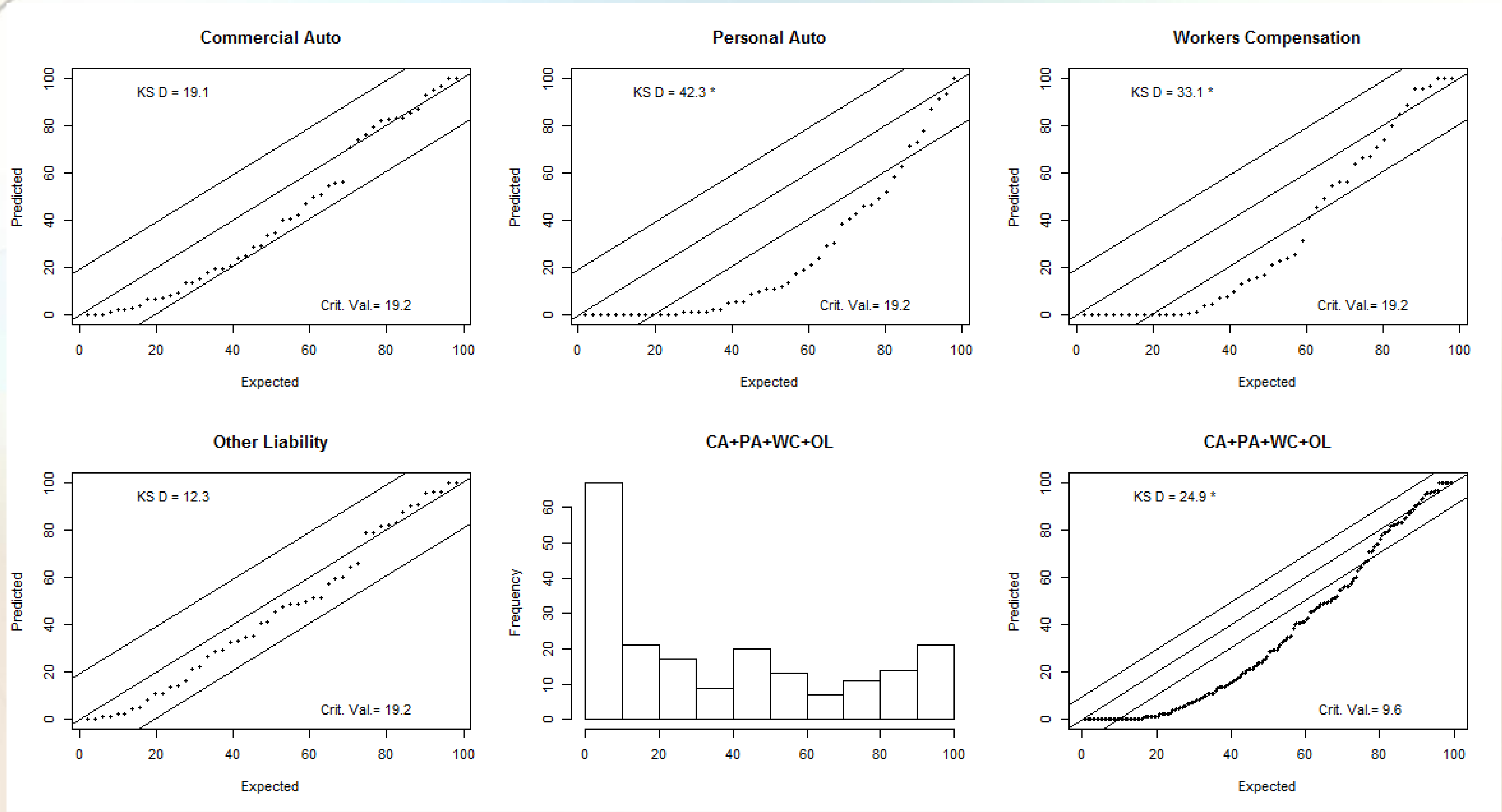
Improvement with Incurred Data

- Accomplished by "pumping up" the variance of Mack model.

What About Paid Data?

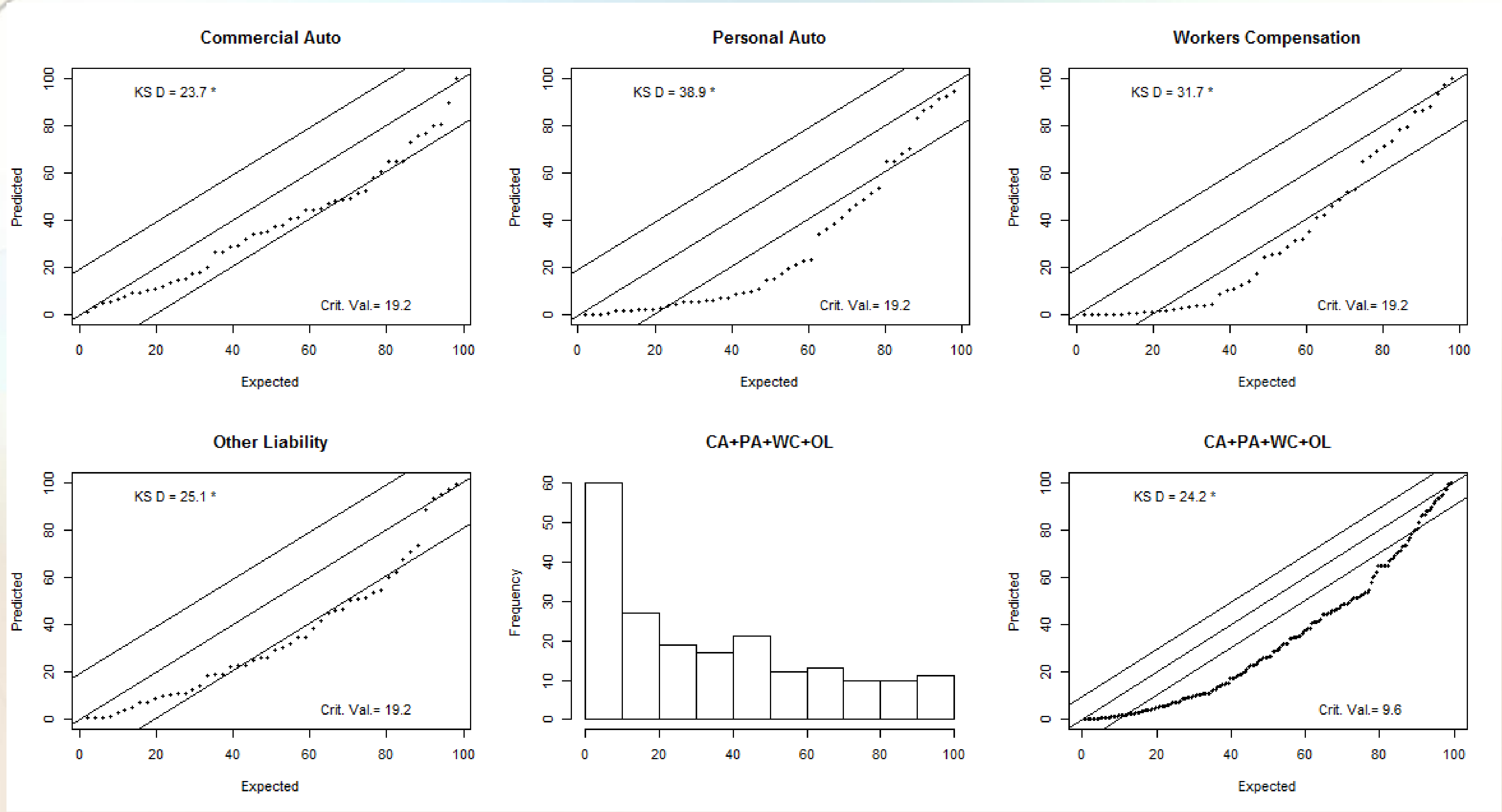
- Start by looking at CCL model on cumulative paid data.

Test of Bootstrap ODP on Paid Data



Conclusion – The Bootstrap ODP model is biased upward.

Test of CCL on Paid Data



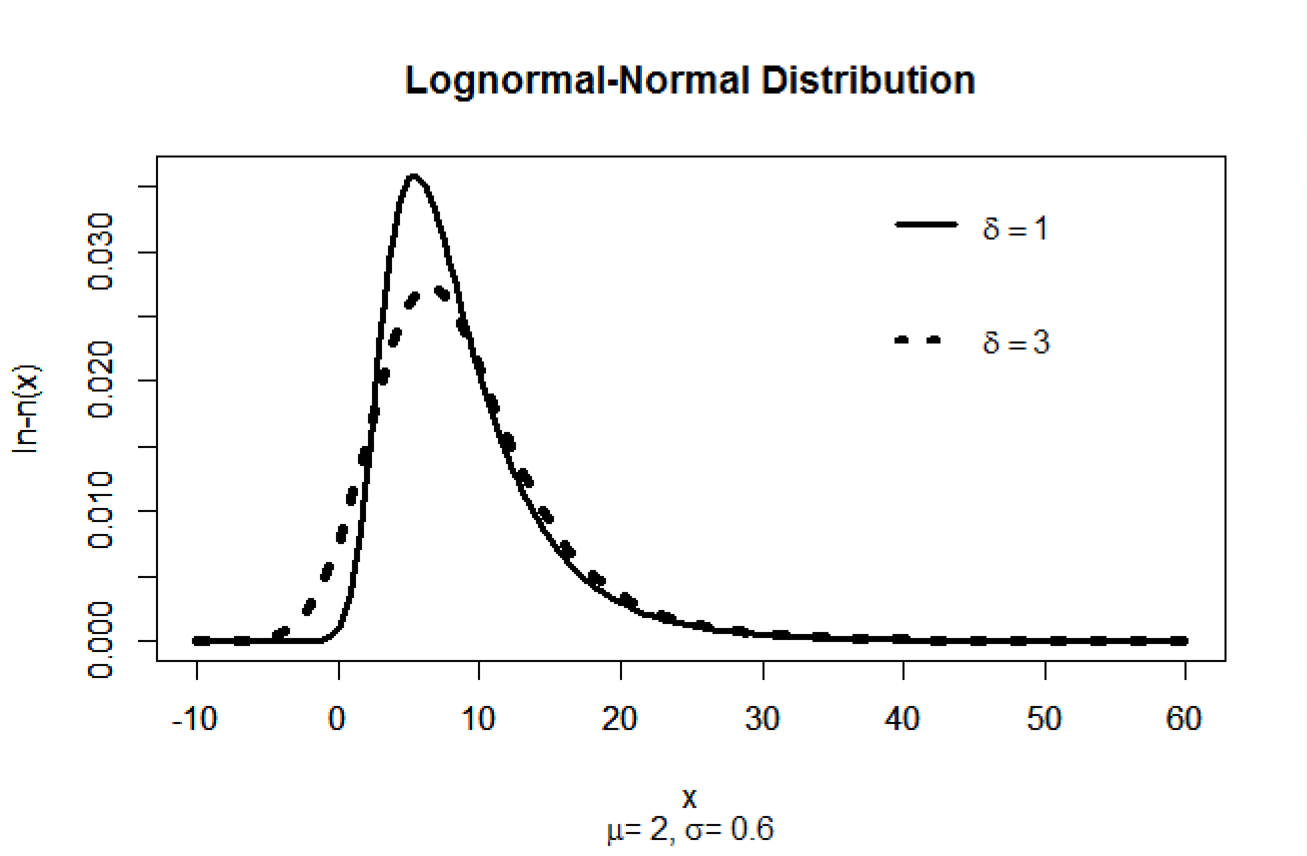
Conclusion – Roughly the same performance a bootstrapping and Mack

How Do We Correct the Bias?

- Look at models with payment year trend.
  - Ben Zehnwirth has been championing these for years.
- Payment year trend does not make sense with cumulative data!
  - Settled claims are unaffected by trend.
- Recurring problem with incremental data – Negatives!
  - We need a skewed distribution that has support over the entire real line.

The Lognormal-Normal (ln-n) Mixture

$X \sim$ Normal(Z,δ), $Z \sim$ Lognormal(μ,σ)



The Correlated Incremental Trend (CIT) Model

- $\mu_{w,d} = \alpha_w + \beta_d + \tau \cdot (w + d - 1)$
- $Z_{w,d} \sim \text{lognormal}(\mu_{w,d}, \sigma_d)$ subject to $\sigma_1 < \sigma_2 < \ldots < \sigma_{10}$
- $I_{1,d} \sim \text{normal}(Z_{1,d}, \delta)$
- $I_{w,d} \sim \text{normal}(Z_{w,d} + \rho \cdot (I_{w-1,d} - Z_{w-1,d}) \cdot e^{\tau}, \delta)$

- Estimate the distribution of $\sum_{w=1}^{10} C_{w,10}$

- "Sensible" priors
  - Needed to control σ_d
  - Interaction between τ, α_w and β_d.

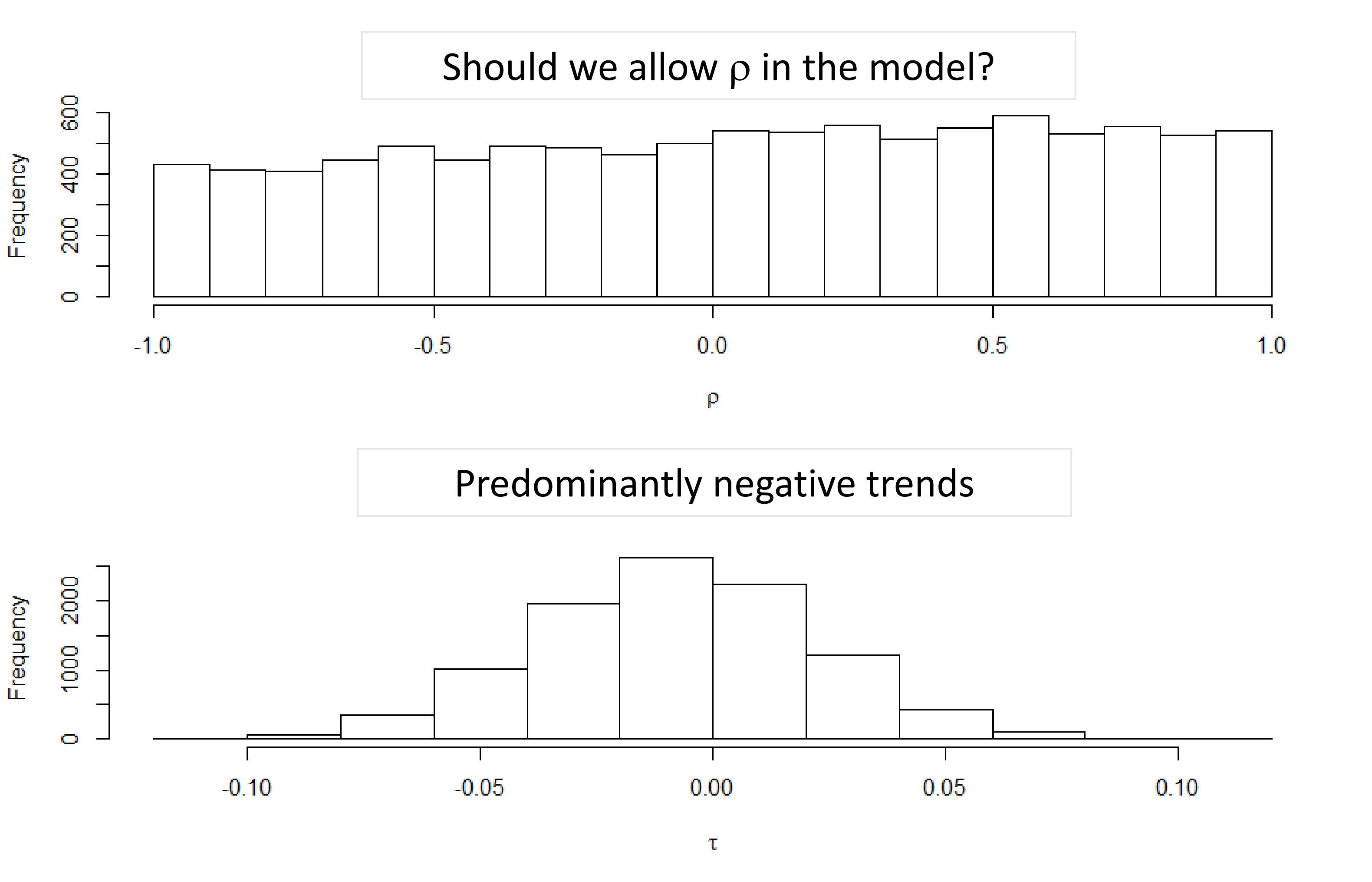
Prior Distribution for CIT Model JAGS Script

```
#
# set up sig2
#
  for (i in 1:length(w)){
    sig2[i]<-sigd2[d[i]]
  }
  sigd2[1]~dunif(.000001,0.5)
  for (j in 2:10){
    sigd2[j]~dunif(sigd2[j-1],sigd2[j-1]+.1)             # control growth of sigma
  }
#
# specify priors
#
  for (i in 1:numlev){
    alpha[i]~dnorm(log(premium[i])+logelr,0.1)           # std dev of alpha = 1/sqrt(.1) = 3.16
    }
  beta[1]<-0
  for (i in 2:4){
    beta[i]~dunif(-5,5)
    }
  for (i in 5:10){
    beta[i]~dunif(-5,beta[i-1])                          # force beta to decrease for d > 4
    }
rho~dunif(-1,1)
logelr~dunif(-5,1)
tau~dnorm(0,1000)                                        # std dev of tau = 1/sqrt(1000) = 0.0316
delta~dunif(0,sum(premium)/10)
```

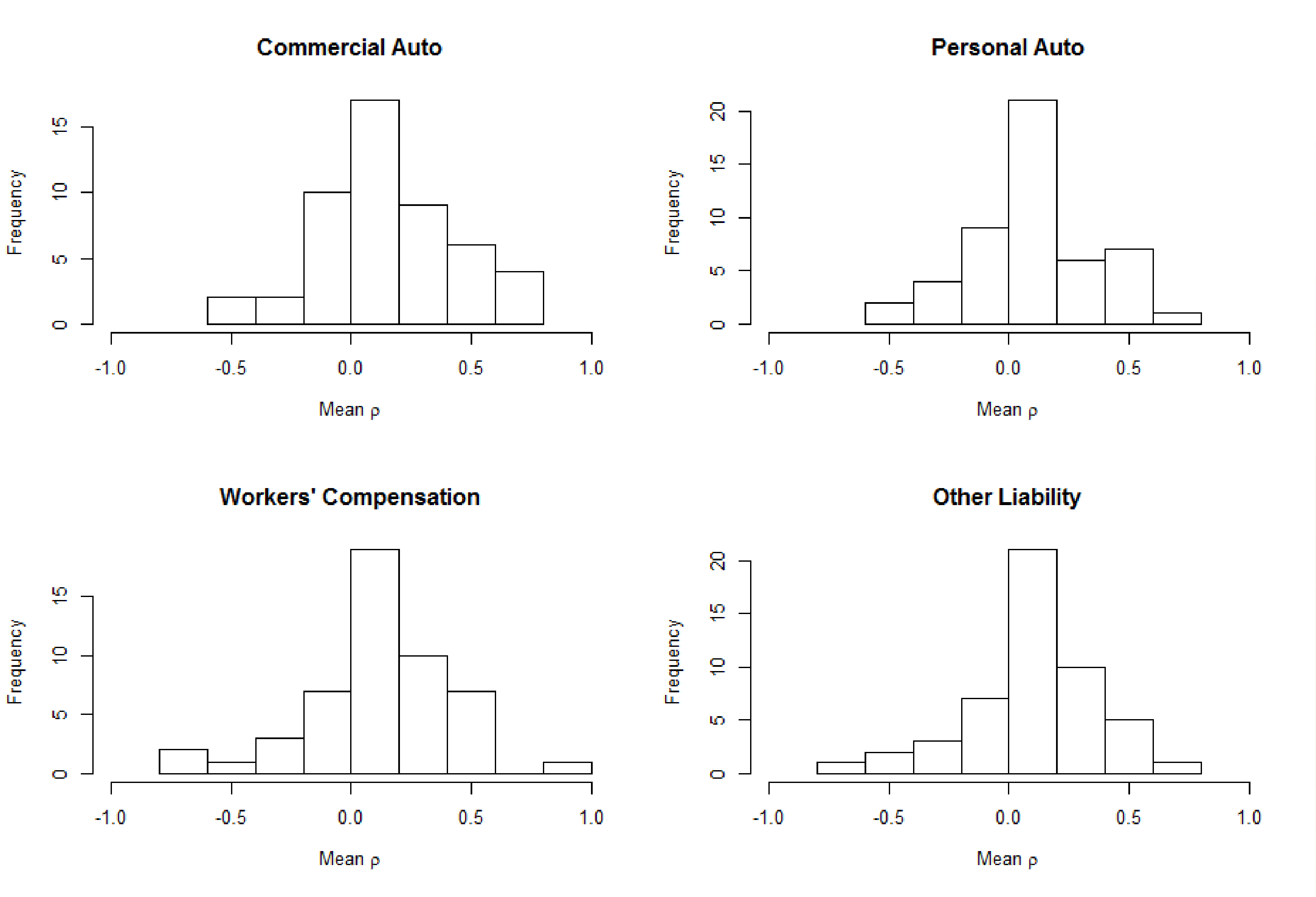
CIT Model for Illustrative Insurer

	CIT				CCL				Outcome
w	$C_{w,10}$	SD	CV		$C_{w,10}$	SD	CV		$C_{w,10}$
1	3912	0	0		3912	0	0.0000		3912
2	2536	5	0.002		2563	110	0.0429		2527
3	4175	11	0.0026		4153	189	0.0455		4274
4	4378	29	0.0066		4320	224	0.0519		4341
5	3539	35	0.0099		3570	207	0.0580		3583
6	3043	105	0.0345		3403	255	0.0749		3268
7	5037	114	0.0226		5207	465	0.0893		5684
8	3501	556	0.1588		3649	467	0.1280		4128
9	3980	710	0.1784		4409	895	0.2030		4144
10	4661	1484	0.3184		5014	2435	0.4856		4139
Total	38763	1803	0.0465		40200	3070	0.0764		40000
Percentile		81.87				51.24			

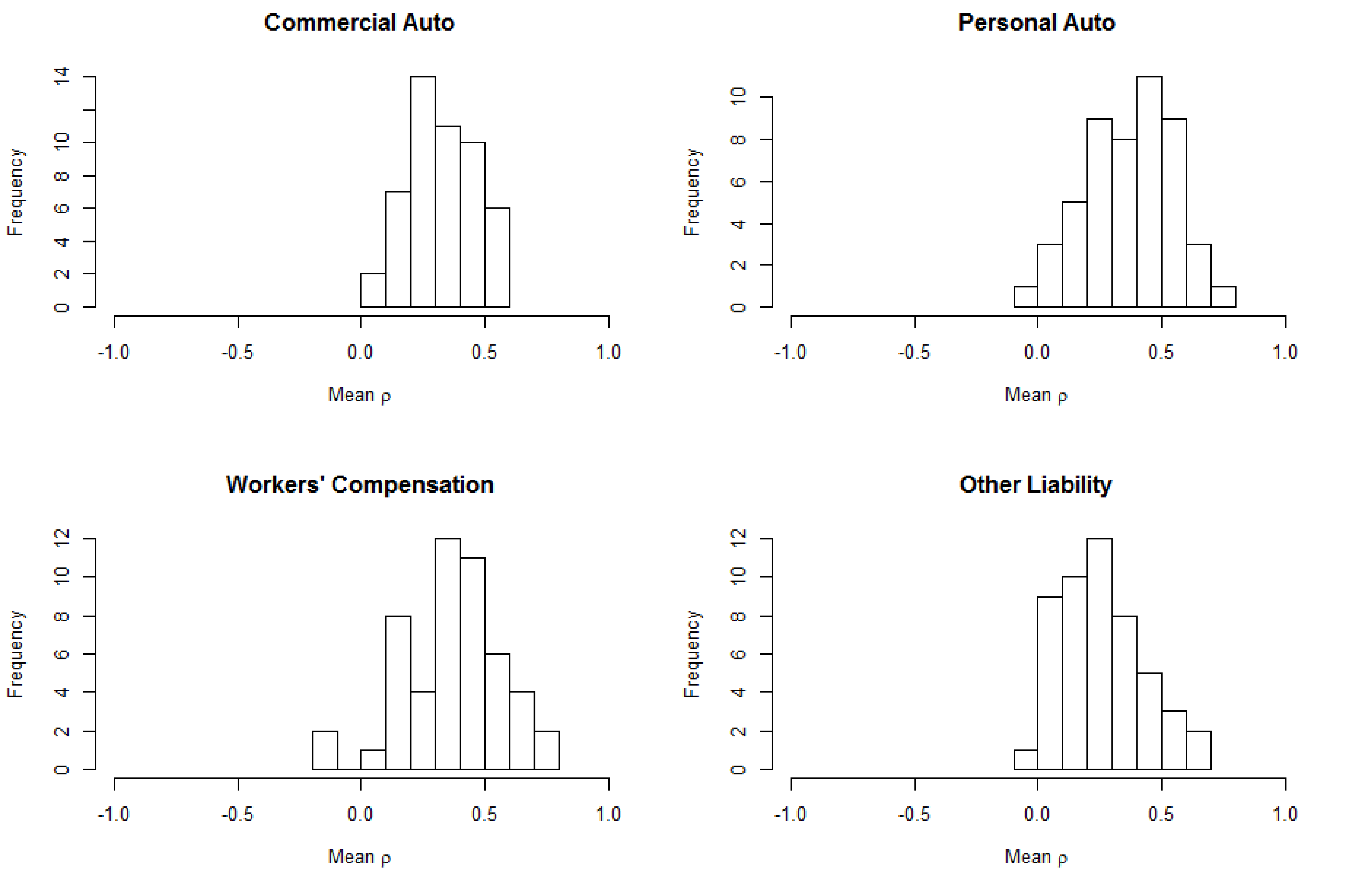
Posterior Distribution of μ and τ for Illustrative Insurer



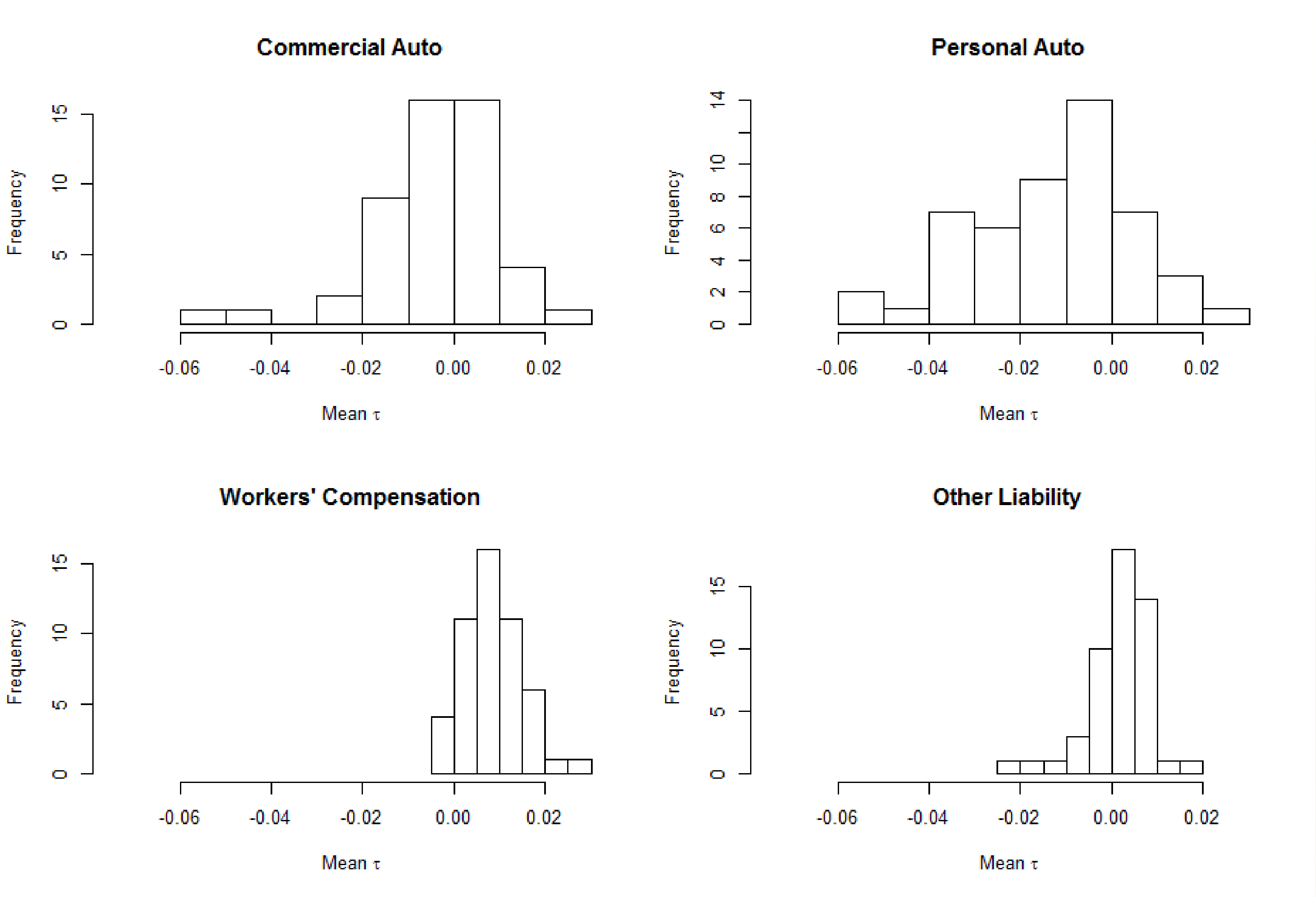
Posterior Mean ρ for All Insurers On Paid Data



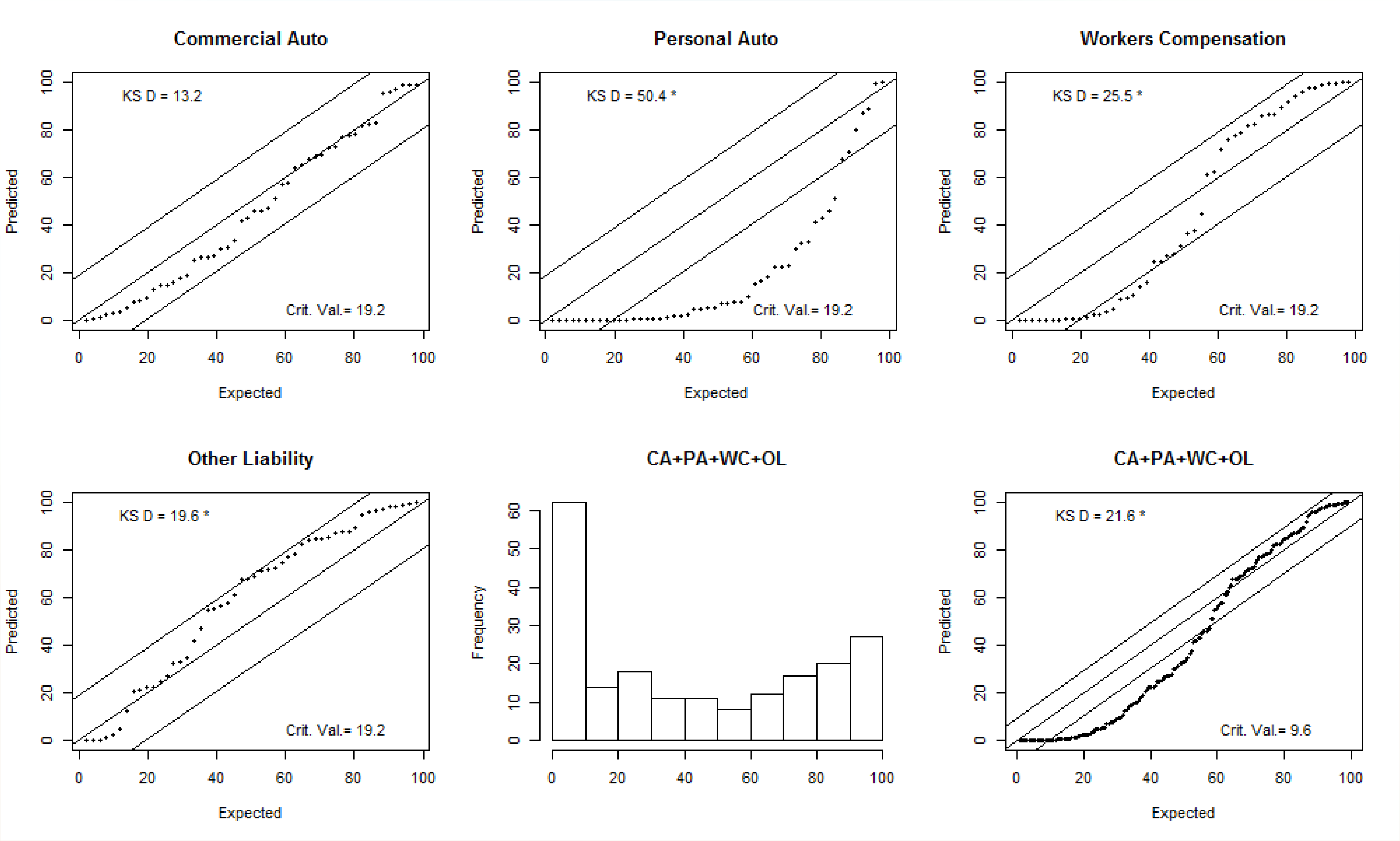
Posterior Mean ρ for All Insurers On Incurred Data



Posterior Mean τ for All Insurers

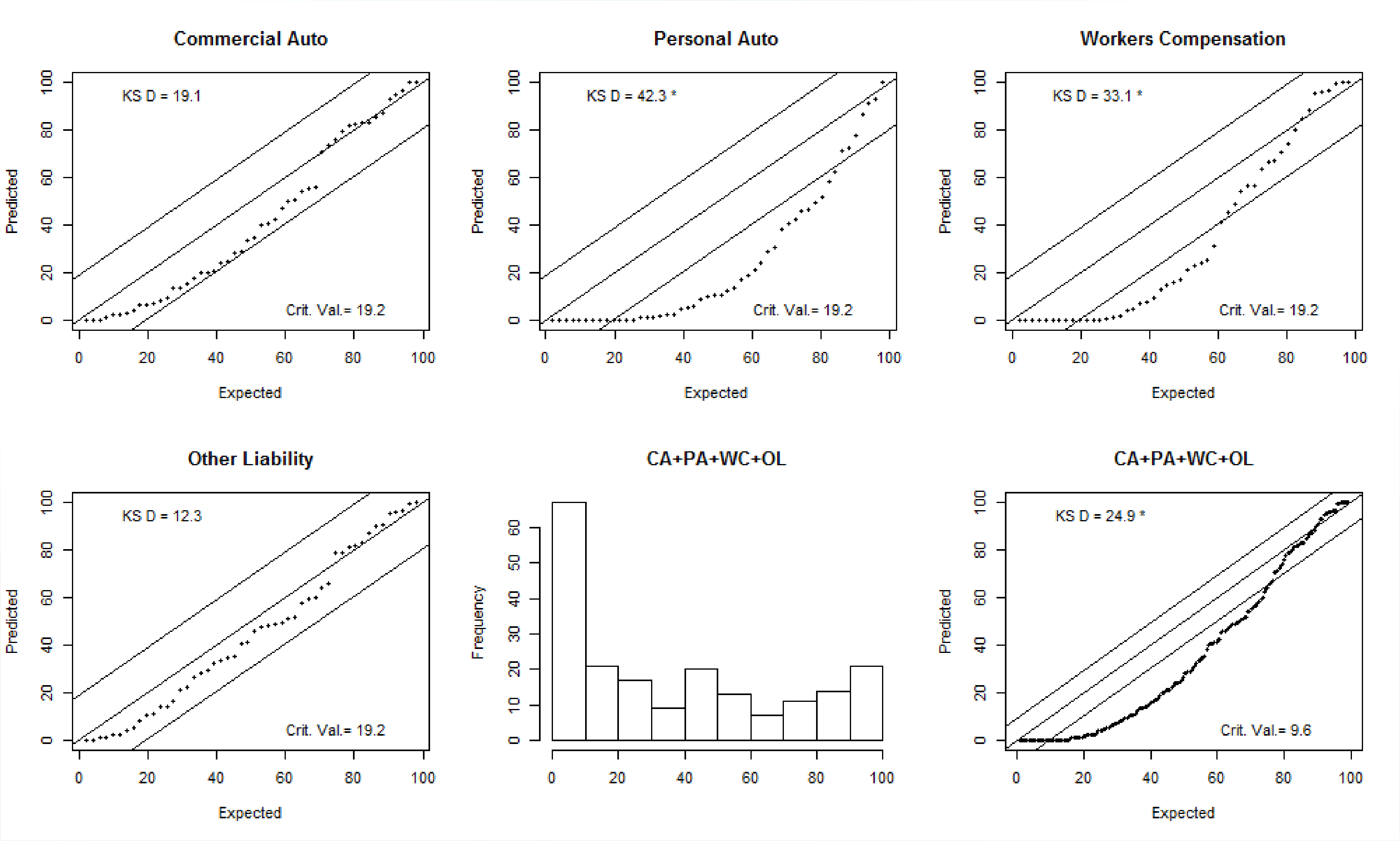


Test of CIT with $\rho = 0$ on Paid Data



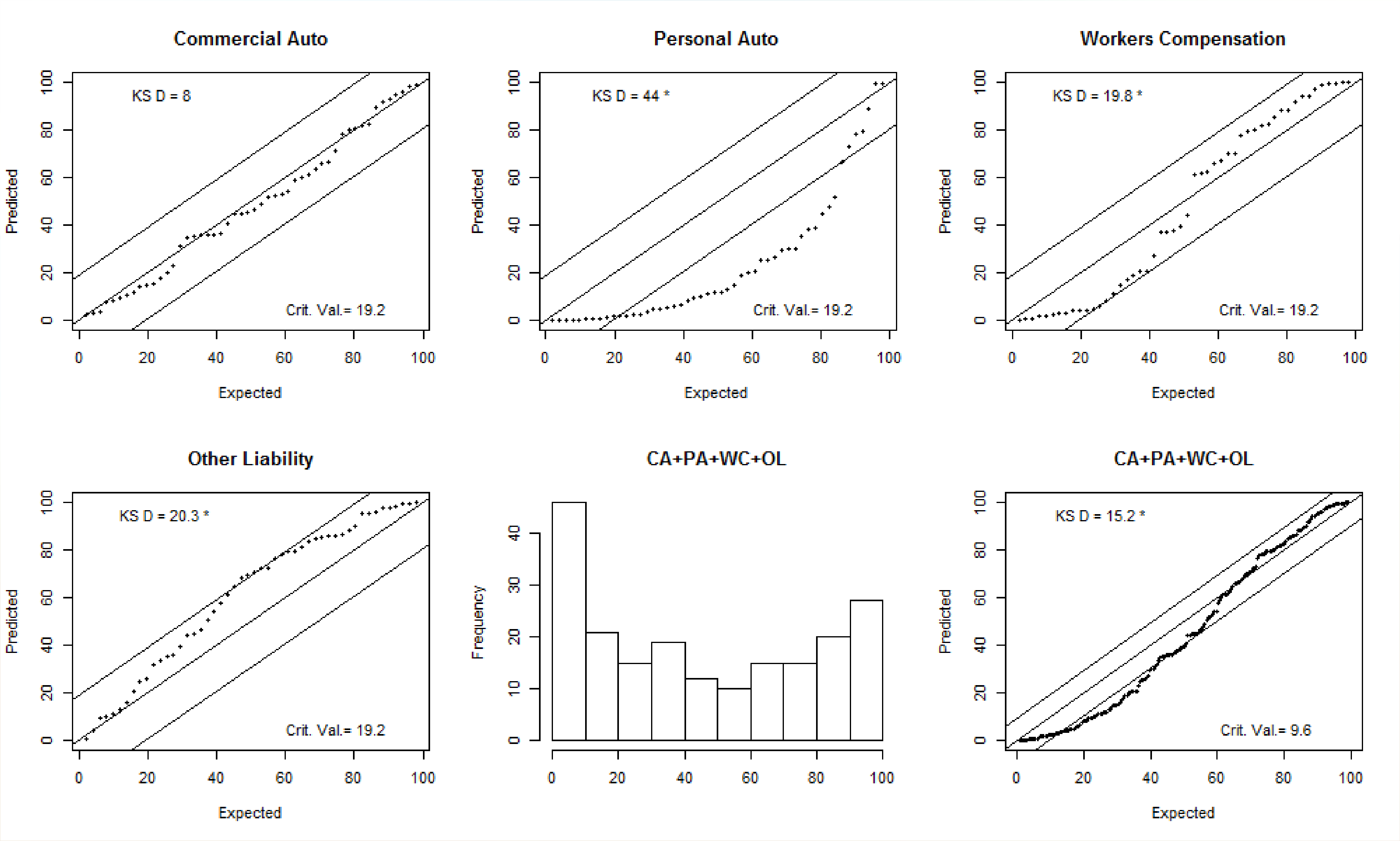
Conclusion – Overall improvement but look at Personal Auto

Test of Bootstrap ODP on Paid Data



Conclusion – The Bootstrap ODP model is biased upward.

Test of CIT on Paid Data



Conclusion – CIT model percentiles are an improvement but do not lie within the KS bounds.

Summary

- Mack underpredicts the variability of outcomes with incurred data.
- Both Mack and Bootstrap ODP are biased high with paid data.
- Bayesian MCMC models
  - Easily modified to produce new models.
  - Easily implemented to produce predictive distributions of outcomes.
- CCL model improves significantly on predictions with incurred data.
  - Important feature – Correlation between accident years
- CIT model improves somewhat on predictions with paid data.
  - Important features – Payment year trend and correlation between accident years
- Shortcoming – Study needs to be repeated on different time periods.