Slides to accompany CS777 discussion

Analysing the Predictive Accuracy of Software Reliability Models & Recalibrating to Improve Upon Predictions

NOTE: This material is drawn from Chapter 4 of Lyu & supplemented by the tutorial material of Prof Bev Littlewood & Pete Mellor (of CSR, City University, London) – with the kind permission of Pete Mellor


TRUST
NO ONE
The X-Files - Bottom's Evasion 2000

Sample Data Set – SYS1

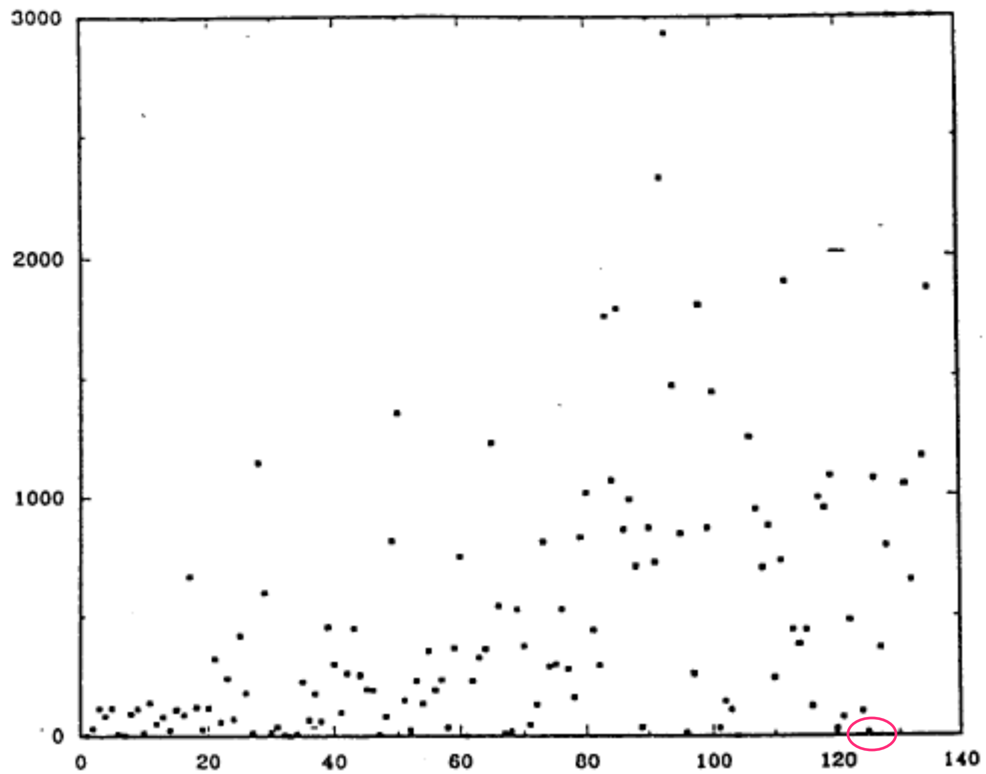
3	30	113	81	115
9	2	91	112	15
138	50	77	24	108
88	670	120	26	114
325	55	242	68	422
180	10	1146	600	15
36	4	0	8	227
65	176	58	457	300
97	263	452	255	197
193	6	79	816	1351
148	21	233	134	357
193	236	31	369	748
0	232	330	365	1222
543	10	16	529	379
44	129	810	290	300
529	281	160	828	1011
445	296	1755	1064	1783
860	983	707	33	868
724	2323	2930	1461	843
12	261	1800	865	1435
30	143	108	0	3110
1247	943	700	875	245
729	1897	447	386	446
122	990	948	1082	22
75	482	5509	100	10
1071	371	790	6150	3321
1045	648	5485	1160	1864
4116				

Failure data set collected by John Musa of Bell Labs during operational testing of a command & control system [Musa 1979]. This is available online from Lyu's data directory of failure data samples (SYS1_DAT).

Note that this data set depicts time-between-failures data (i.e. time to failure & not failure count). The time unit is seconds & the set needs to be read left to right, line by line, down the page.

SYS1 – Plotting Time Between Individual Failures

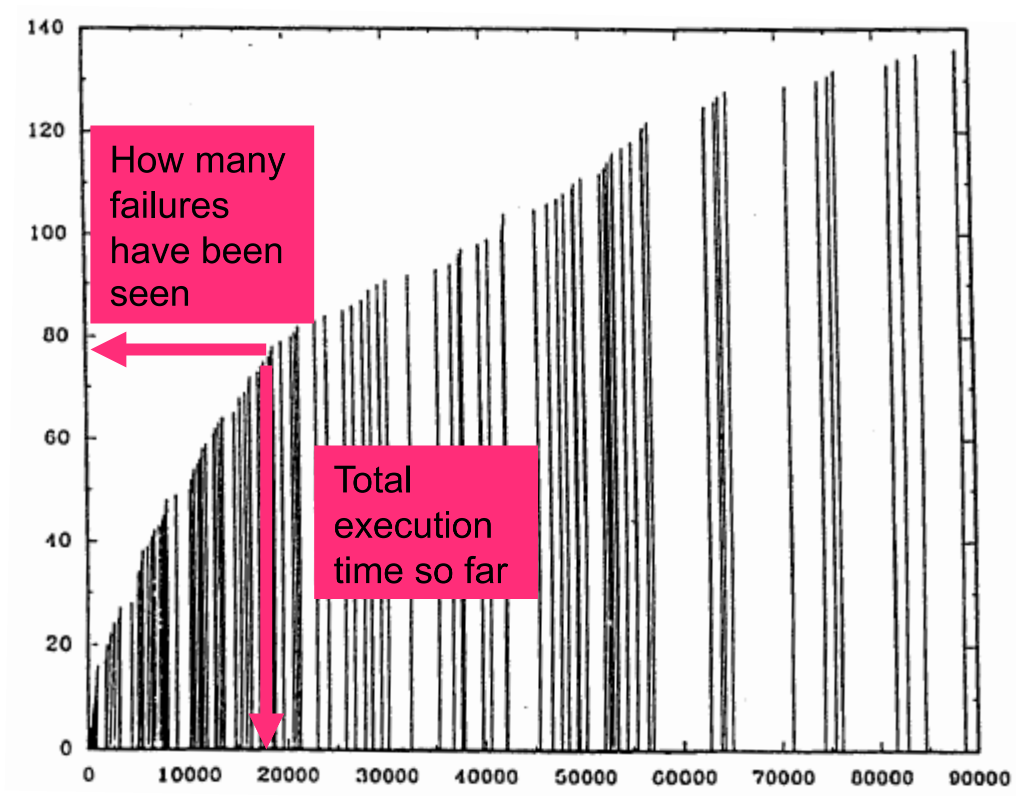
t_i inter-failure time, in seconds



Failure number, i

SYS1 – Plotting Cumulative Failure Over Time

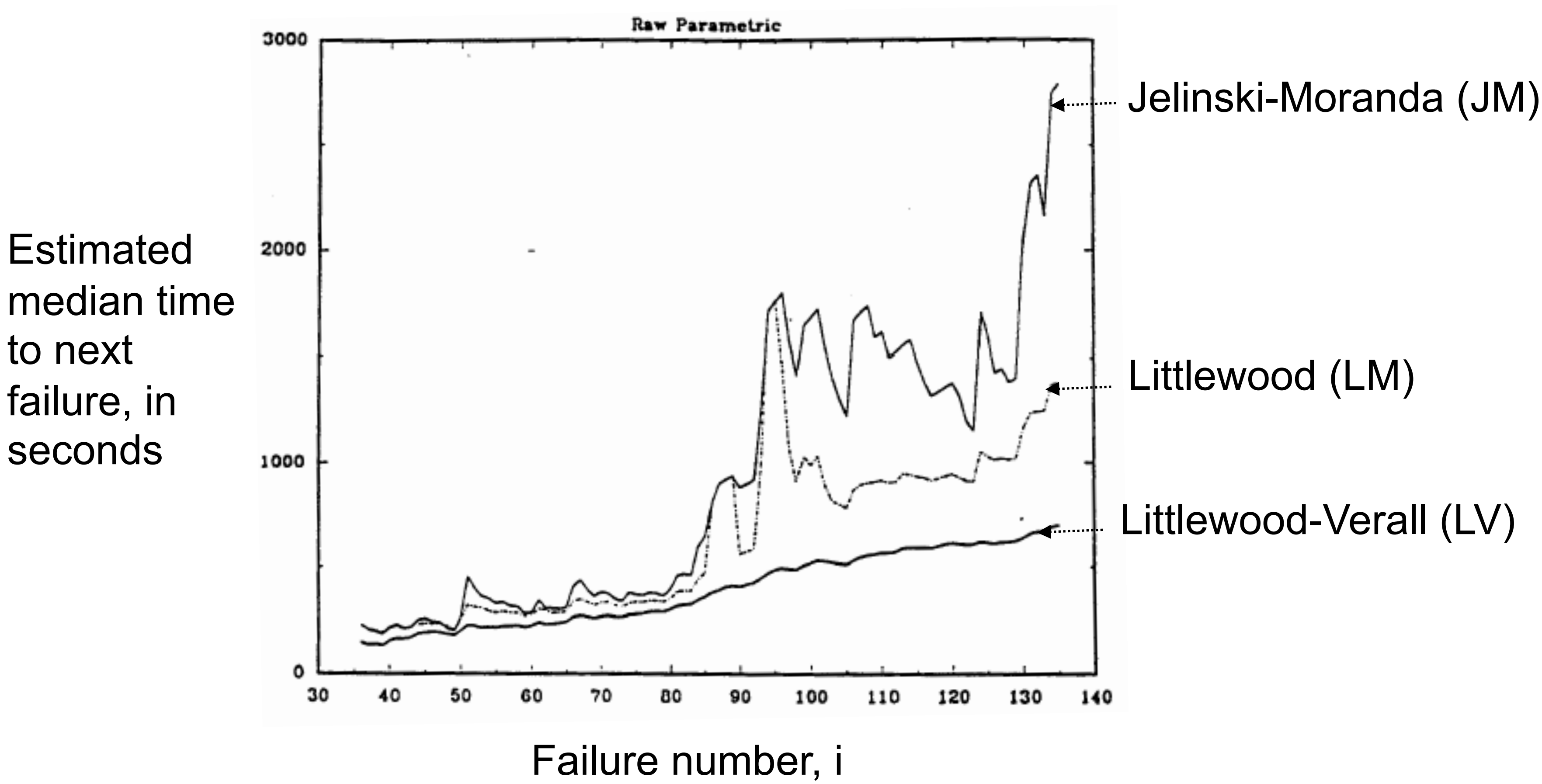
Cumulative number of failures



Total elapsed time, in seconds

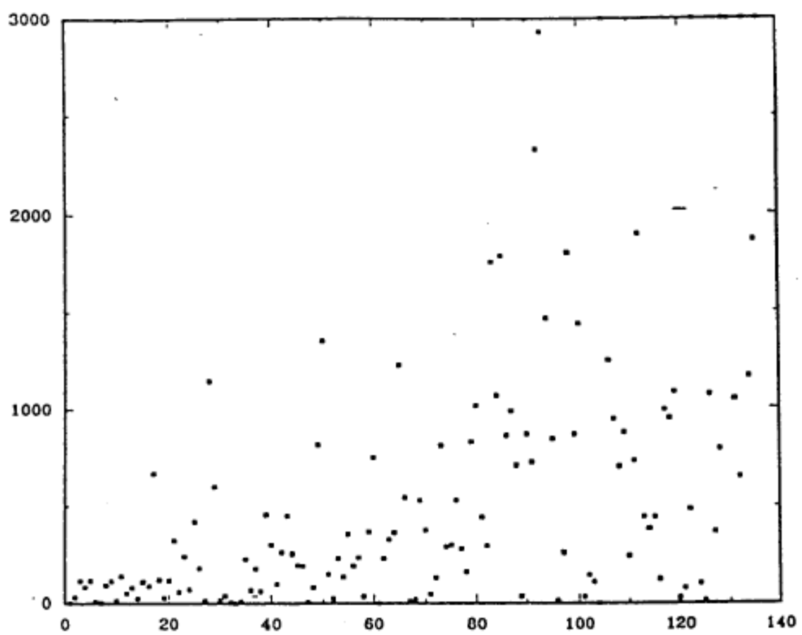
SYS1 - Sequential Estimates of Current Reliability

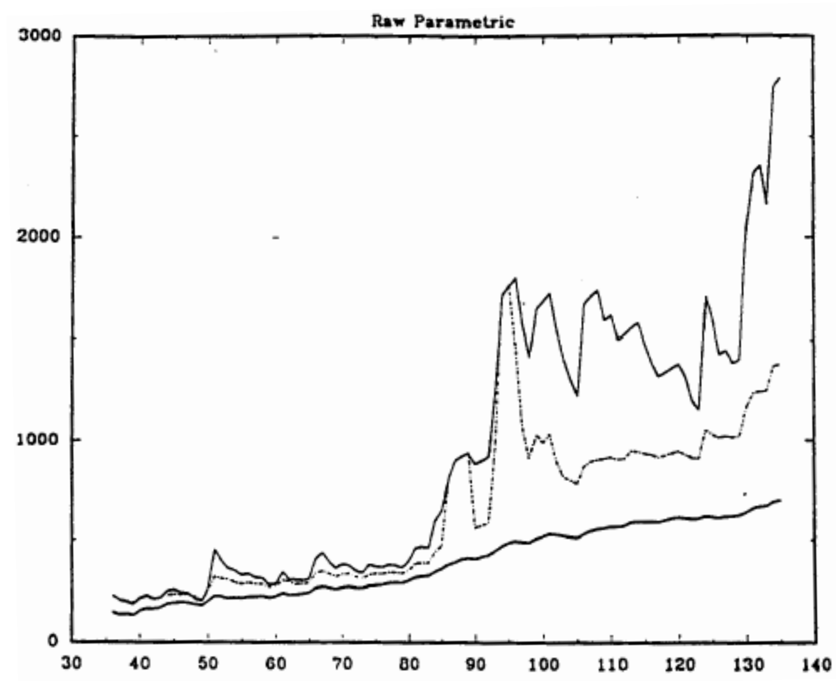
Plotting Median Time to Next Failure
(1 Step-ahead prediction) Using 3 Models



Is the Truth Out There?

Crude Way to Judge Whether Median Plots are Reasonable

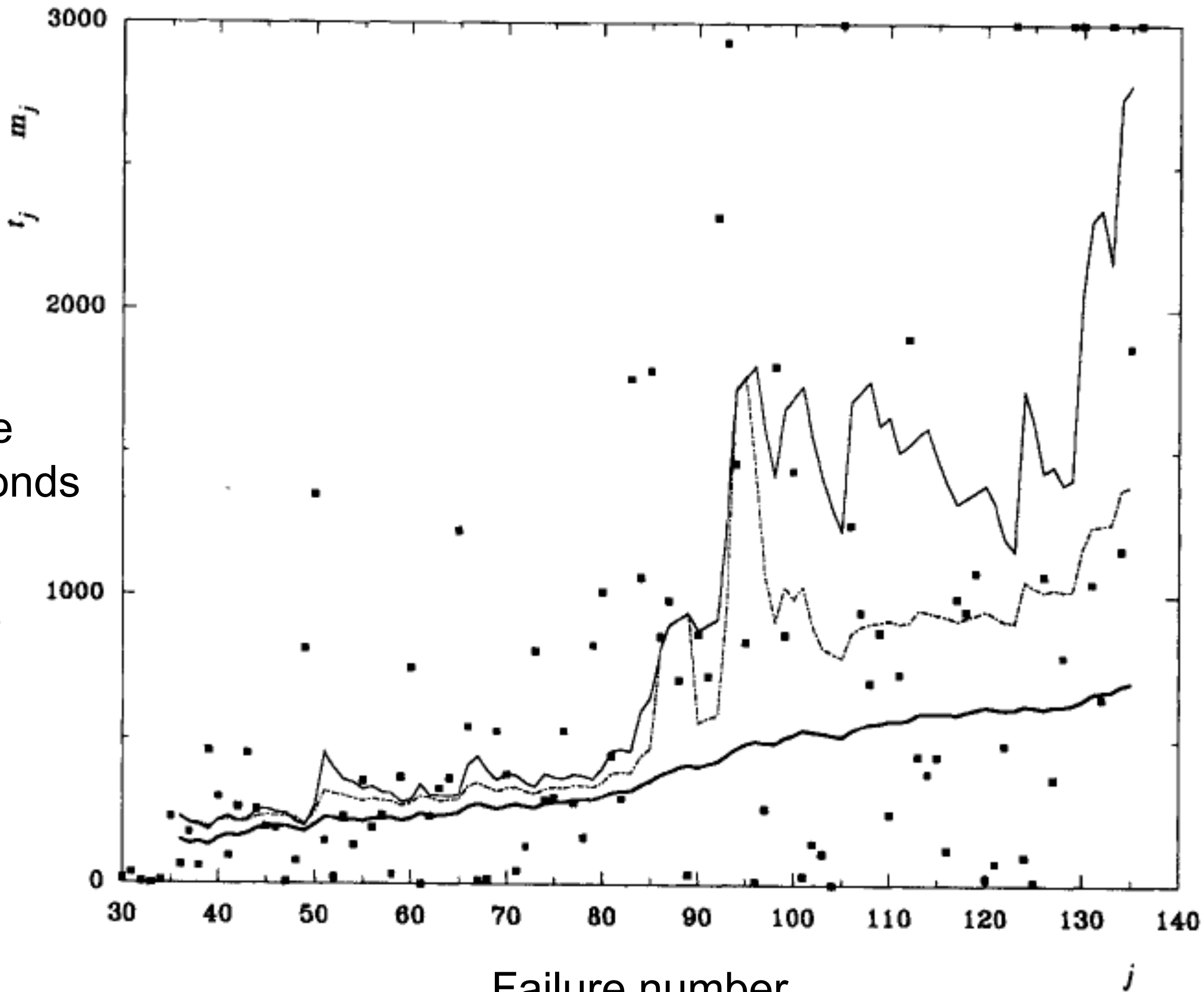




50% Above? 50% Below?

t_i inter-failure time, in seconds

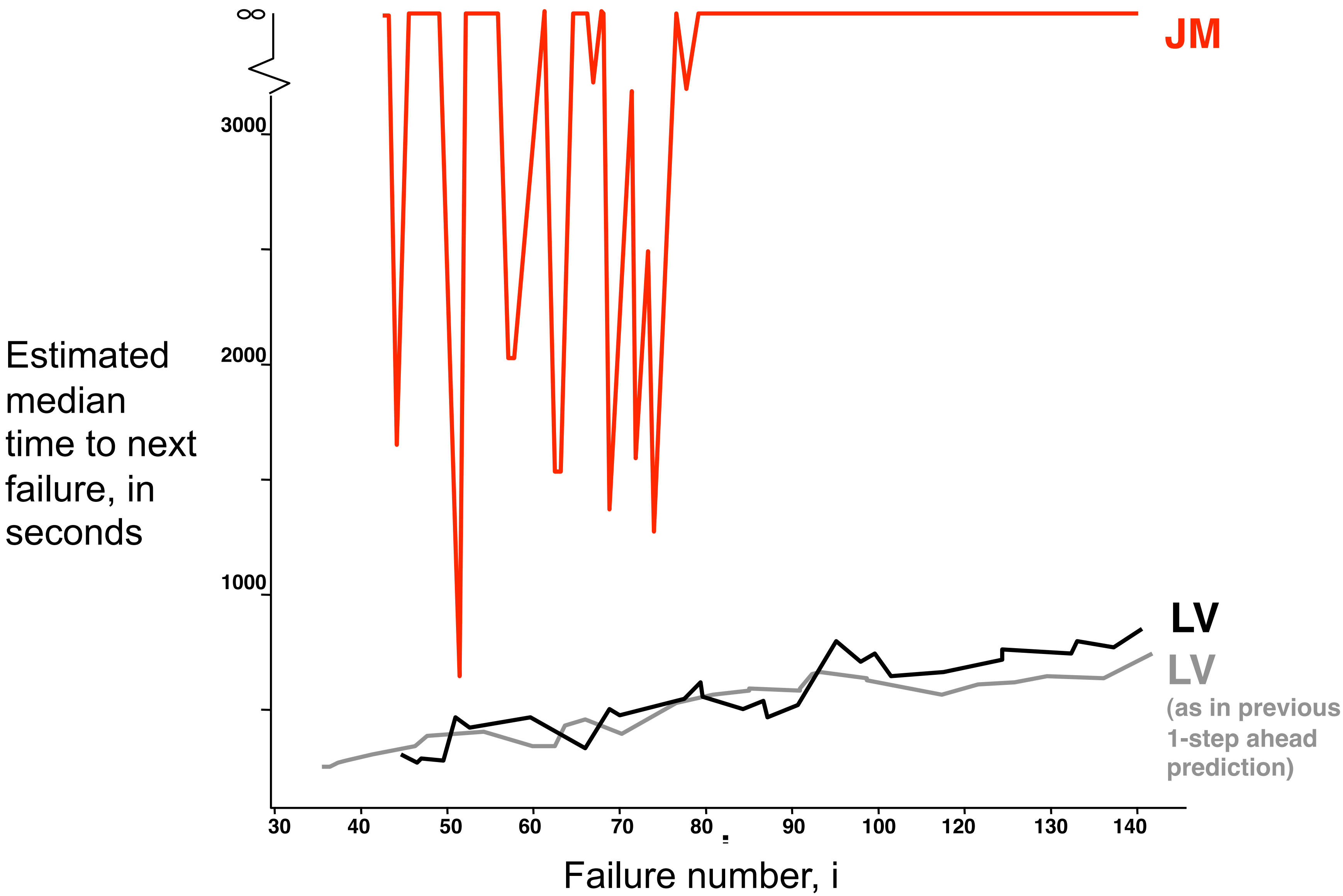
Estimated median time to next failure, in seconds



Failure number

Are 50:50 Odds Good Enough?

SYS1 - 20 Step-ahead Predictions



Yesterday's Weather

Sun	Mon	Tue	Wed	Thu	Fri	Sat
1 OBSERVED N/A Hi 35°F Lo 28°F Precip (in) 0.54in.	2 OBSERVED N/A Hi 28°F Lo 16°F Precip (in) 0.11in.	3 OBSERVED Hi 26°F Lo 12°F Precip (in) 0in.	4 OBSERVED Hi 34°F Lo 11°F Precip (in) 0in.	5 OBSERVED Hi 43°F Lo 22°F Precip (in) 0.01in.	6 OBSERVED Hi 55°F Lo 36°F Precip (in) 0in.	7 OBSERVED Hi 70°F Lo 37°F Precip (in) 0in.
8 OBSERVED Hi 65°F Lo 49°F Precip (in) 0.02in.	9 OBSERVED Hi 50°F Lo 38°F Precip (in) 0.12in.	10 OBSERVED Hi 46°F Lo 37°F Precip (in) 0in.	11 OBSERVED Hi 58°F Lo 40°F Precip (in) 0in.	12 OBSERVED Hi 49°F Lo 31°F Precip (in) 0in.	13 OBSERVED Hi 39°F Lo 27°F Precip (in) 0in.	14 OBSERVED Hi 50°F Lo 33°F Precip (in) 0in.
15 OBSERVED Hi 51°F Lo 43°F Precip (in) 0in.	16 OBSERVED Hi 49°F Lo 40°F Precip (in) 0in.	17 OBSERVED Hi 54°F Lo 35°F Precip (in) 0in.	18 OBSERVED Hi 64°F Lo 40°F Precip (in) 0.01in.	19 OBSERVED Hi 54°F Lo 42°F Precip (in) 0.15in.	20 OBSERVED Hi 45°F Lo 33°F Precip (in) 0.04in.	21 OBSERVED Hi 49°F Lo 31°F Precip (in) 0in.
22 OBSERVED Hi 54°F Lo 35°F Precip (in) 0in.	Today Hi 40°F Lo 24°F Precip 0 % Wind NNW at 17 mph	24 Hi 49°F Lo 31°F Precip 0 % Wind N at 14 mph	25 Hi 51°F Lo 38°F Precip 20 % Wind ENE at 7 mph	26 Hi 49°F Lo 40°F Precip 40 % Wind SSE at 6 mph	27 Hi 59°F Lo 40°F Precip 20 % Wind NNW at 7 mph	28 Hi 52°F Lo 41°F Precip 30 % Wind E at 8 mph

Monday, Mar 23 — More Details/AccuPop™
Bright and sunny, but colder — High: 40 °F Low: 24 °F

Tuesday, Mar 24 — More Details
Sunny, but chilly — High: 46 °F Low: 34 °F

Wednesday, Mar 25 — More Details
Mostly sunny and milder — High: 54 °F Low: 40 °F

Thursday, Mar 26 — More Details
Rain and drizzle possible — High: 54 °F Low: 44 °F

Friday, Mar 27 — More Details
Partly sunny and pleasant — High: 57 °F Low: 46 °F

Saturday, Mar 28 — More Details
Mostly cloudy — High: 56 °F Low: 40 °F

Sunday, Mar 29 — More Details
Rain — High: 55 °F Low: 42 °F

Monday, Mar 30 — More Details
A shower in the morning — High: 53 °F Low: 35 °F

Tuesday, Mar 31 — More Details
Colder; a shower in the a.m. — High: 46 °F Low: 32 °F

Methods to Analyse Predictive Accuracy (1 Step-ahead)

$\mathbf{T_i}$

- At stage i, we have prediction of the distribution of time to next failure, $\hat{F}_i(t)$
- We want this to be close to the unknown true distribution, $F_i(t)$
- Observe what actually happens, t_i
- Repeat for many i in a time period
- Sequence of $(\hat{F}_i(t), t_i)$ tells us about accuracy of predictions

The u-plot

- For each prediction, calculate:

Observation

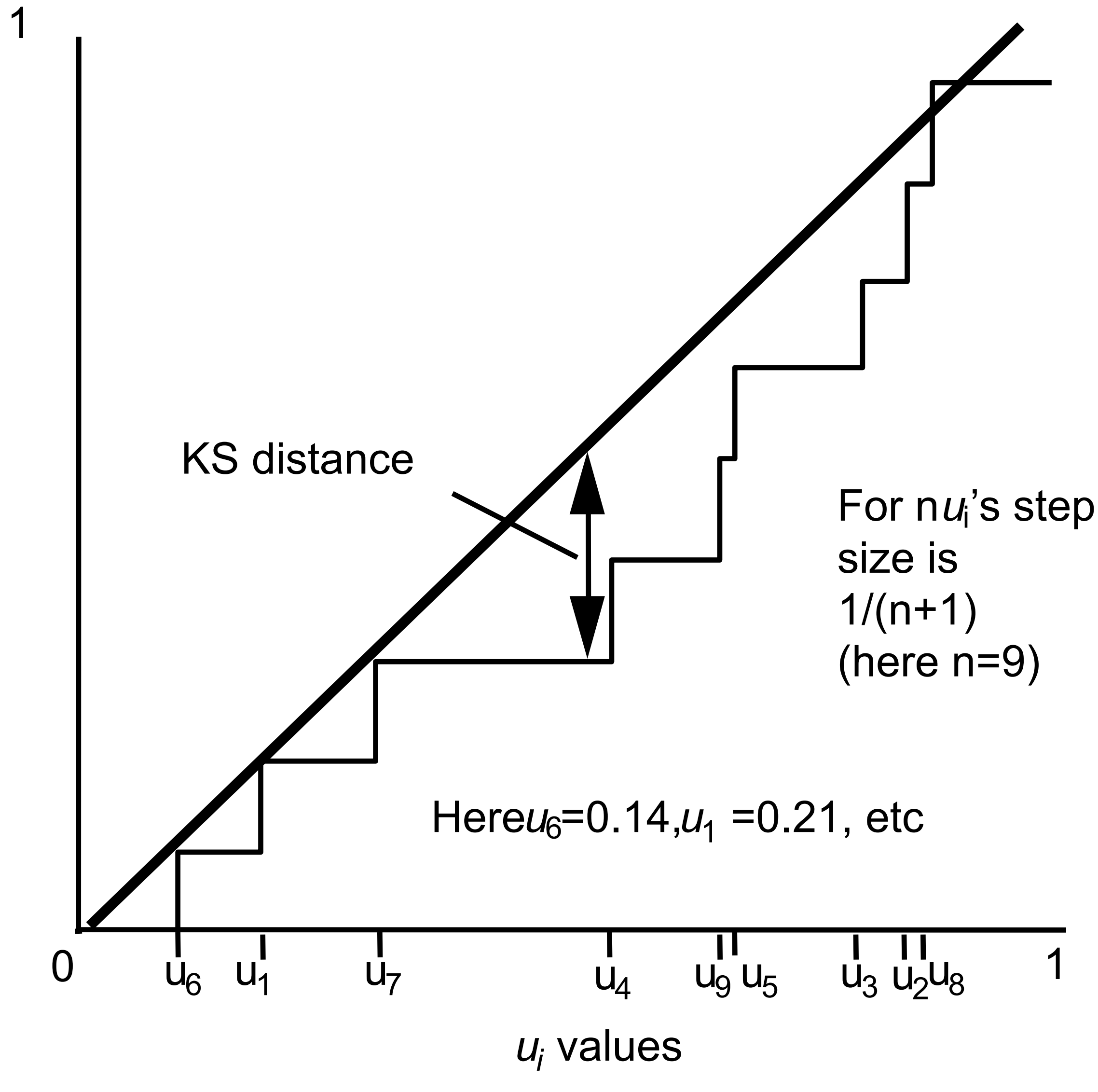
$$u_i = \hat{F}_i(t_i)$$

Prediction

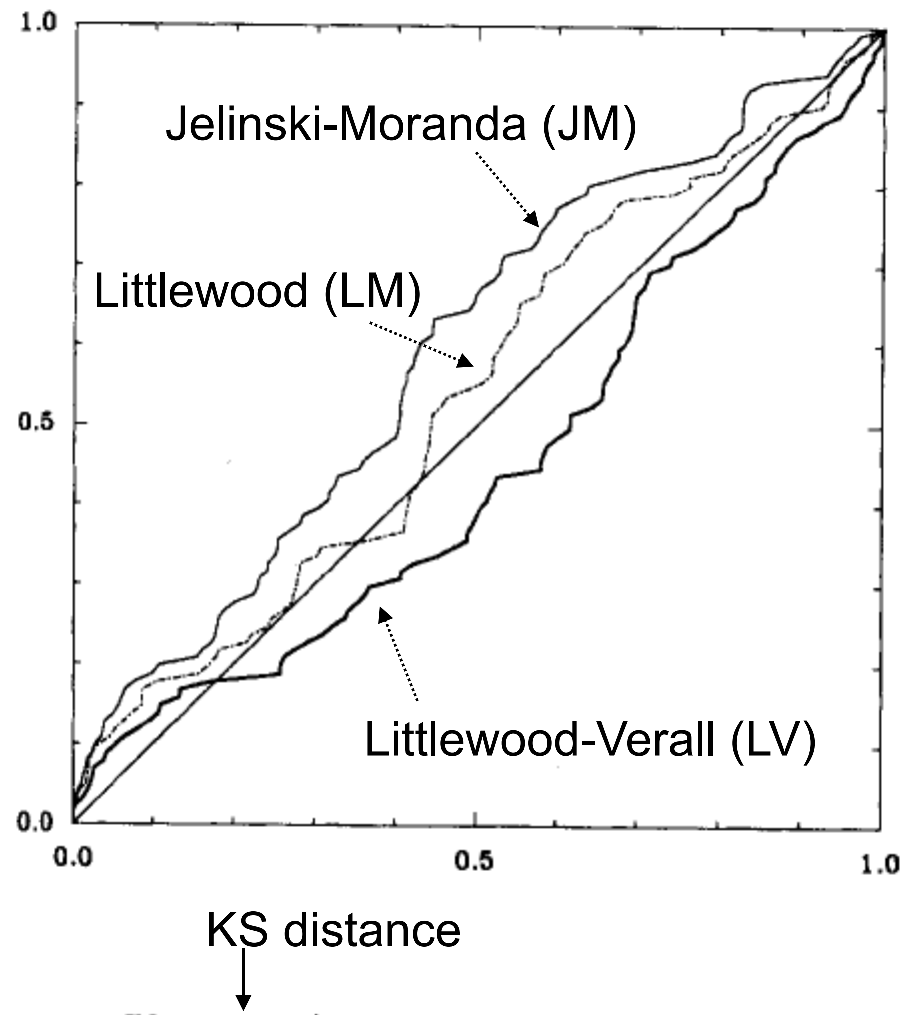
- u-plot is sample distribution function of the u_is

- Tells us about what happens “on average”

How to Draw a u-plot



SYS1 – u-plots for 3 Models



——	JM	0.181 (significant at 1% level)
------	LM	0.103 (insignificant at 20% level)
——	LV	0.148 (2%–5%)

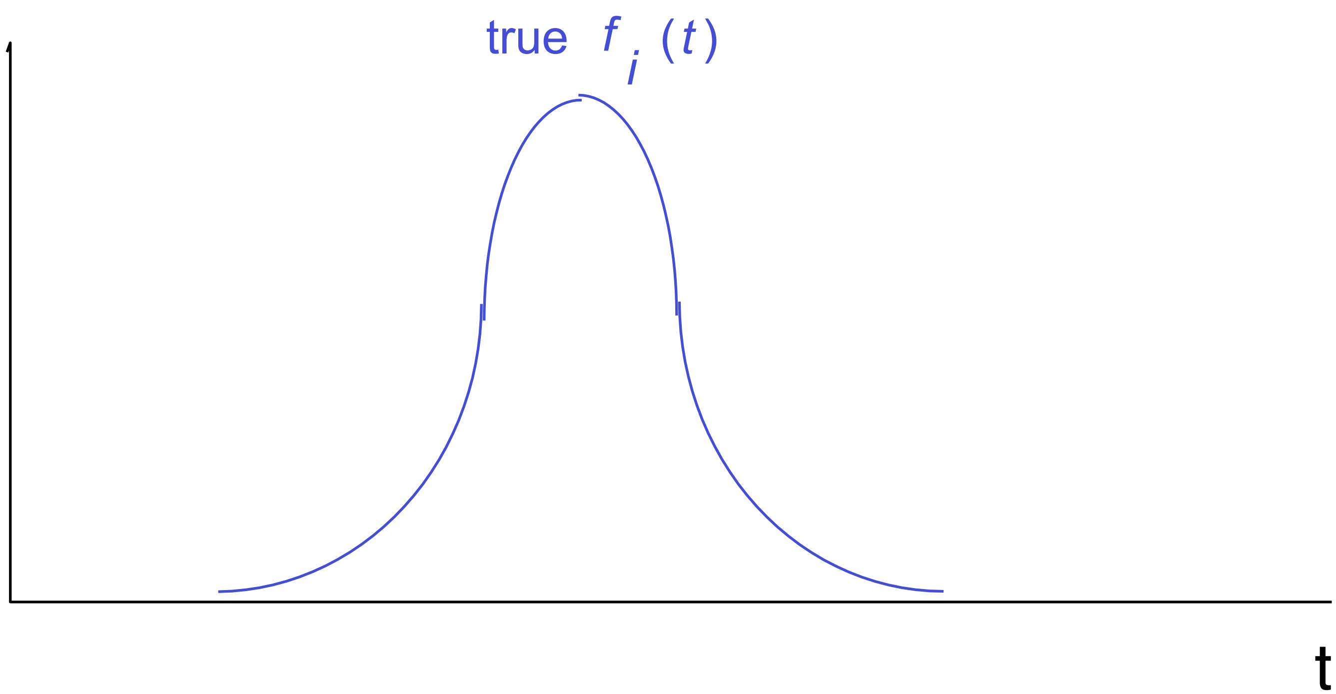
1-step ahead predictions, 100 plots

Below line of unit slope – pessimistic predictions

Above line of unit slope – optimistic predictions

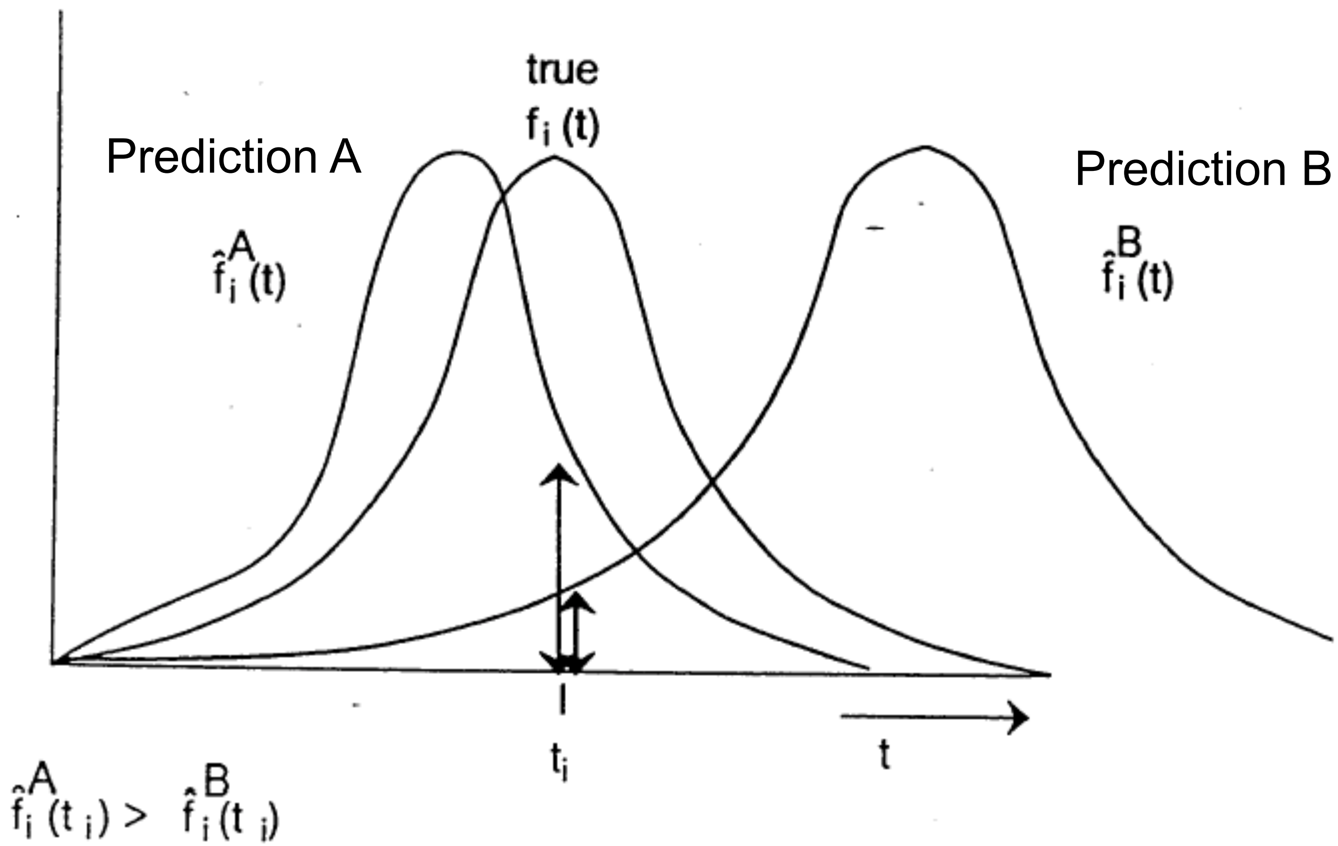
True Probability Density Function

- Probability distribution of T_i over values it may assume
- $f(t) >= 0$ for all t; $\int_{-\infty}^{\infty} f(t)dt = 1$



Hypothetical Predictions from 2 Models versus "Truth"

If prediction system B is consistently less accurate than prediction system A, the actual values will tend to appear less likely according to B than according to A



Prequential Likelihood

- To compare several competing sets of predictions on the same data source
- Select the one which has given the globally most accurate predictions
- Detect *consistent bias* & *inappropriate noise* in a prediction system

Prequential Likelihood Function

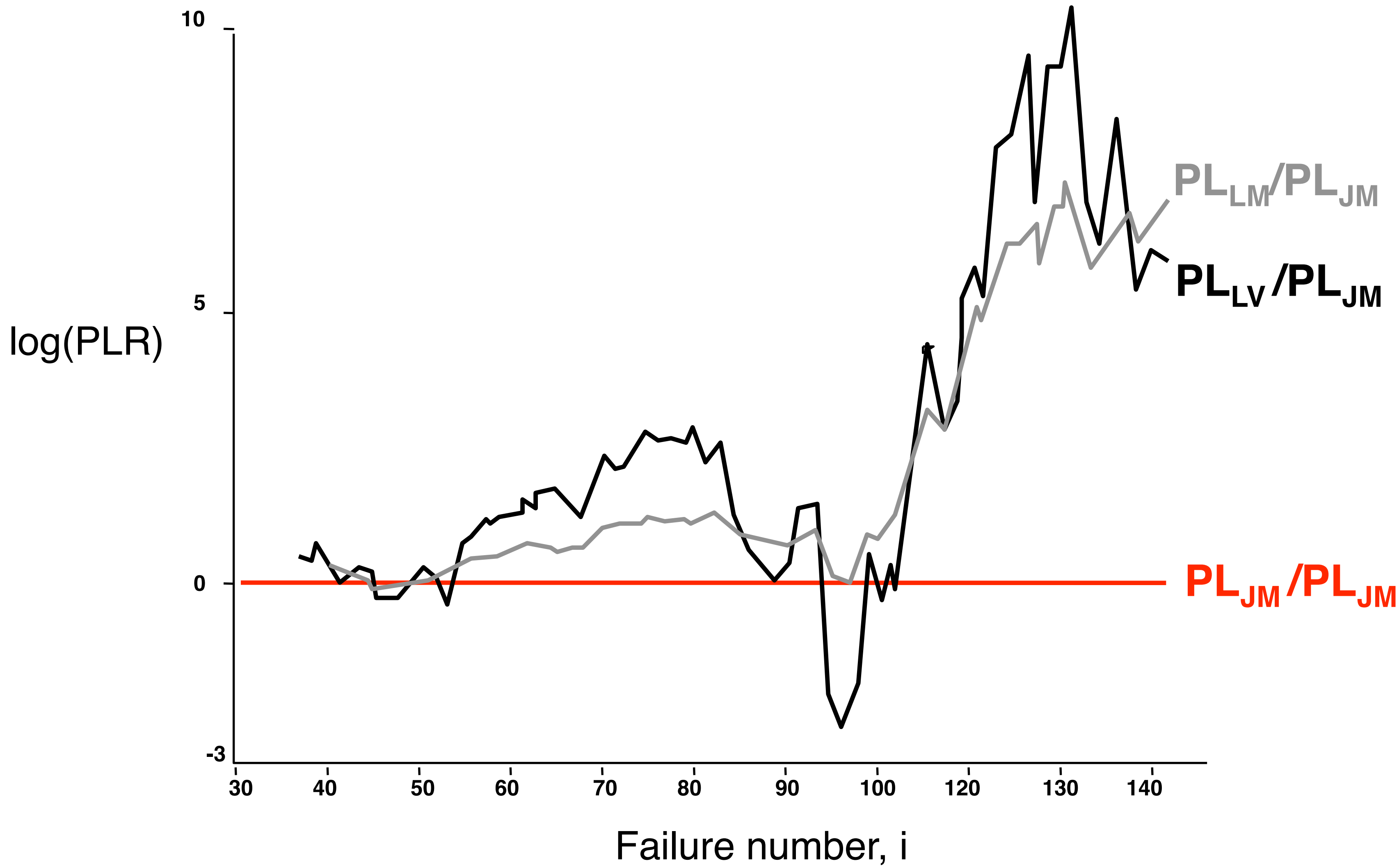
- Problem: estimating true cdf $F_i(t)$ *of* T_i, on the basis of the observed $t_1, t_2, \ldots t_{i-1}$
- Apply prediction system A to a sequence $i=m$ through $i=n$
- After some time from each prediction you observe actual t_i
- The *prequential likelihood function* for these predictions is $PL = \prod_{i=m}^{i=n} \hat{f}_i(t_i)$

Prequential Likelihood Ratio (PLR)

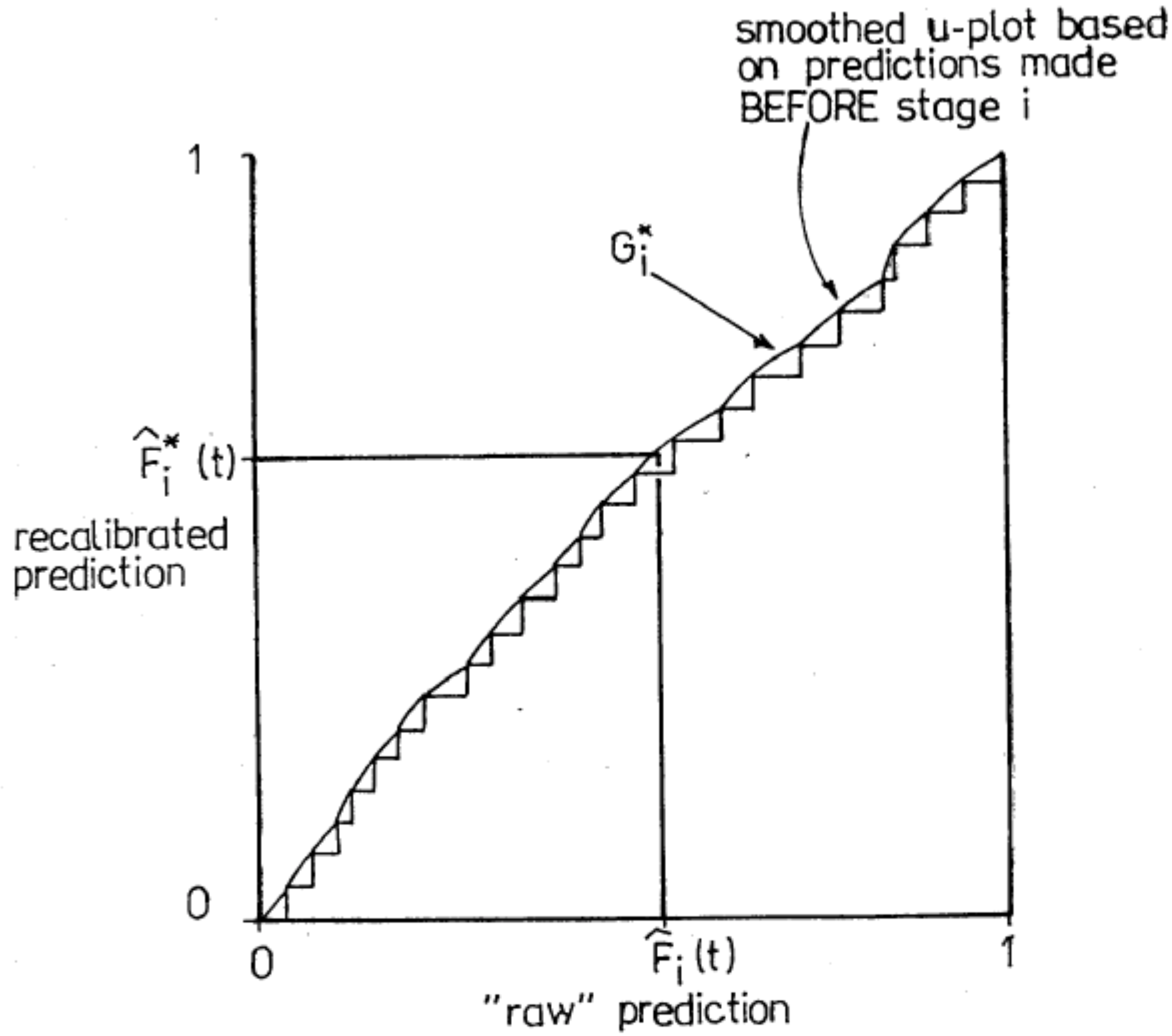
- To compare with a second prediction system, B
- Compute B's predictions, PL function for same sequence
- Consider prequential likelihood ratio: $\dfrac{PL_A}{PL_B}$

- If ratio consistently increases with added predictions, then A is more accurate than B

SYS1 – Log(PLR) with JM as Reference Model



Towards Recalibration



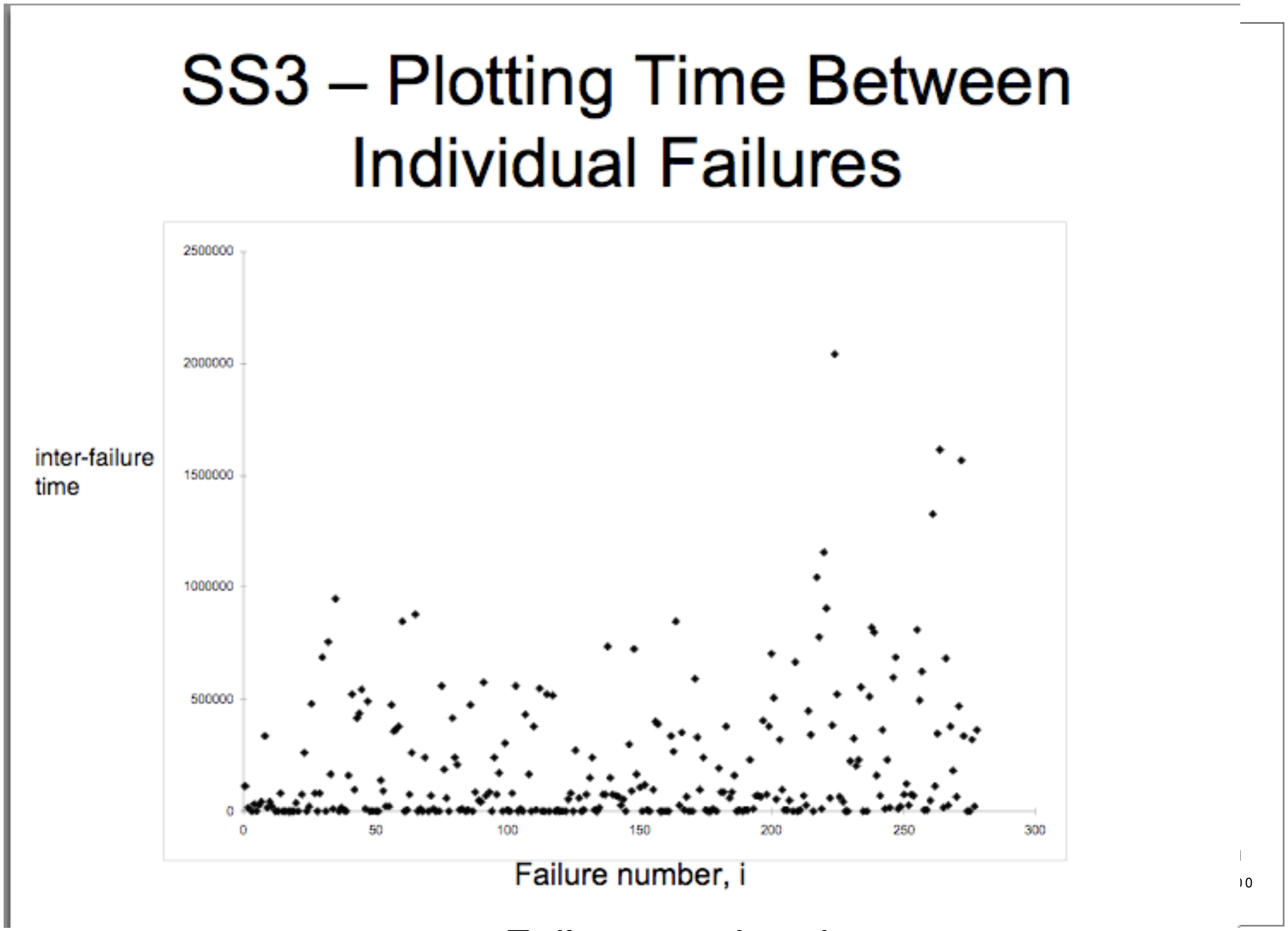
Musa’s SS3 Data Set

107400	17220	180	32880	960	26100	44160	333720	17820
40860	18780	960	960	79860	240	120	1800	480
780	37260	2100	72060	258704	480	21900	478620	80760
1200	80700	688860	2220	758880	166620	8280	951354	1320
14700	3420	2520	162480	520320	96720	418200	434760	543780
8820	488280	480	540	2220	1080	137340	91860	22800
22920	473340	354901	369480	380220	848640	120	3416	74160
262500	879300	360	8160	180	237920	120	70800	12960
300	120	558540	188040	56280	420	414464	240780	206640
4740	10140	300	4140	472080	300	87600	48240	41940
576612	71820	83100	900	240300	73740	169800	1	302280
3360	2340	82260	559920	780	10740	180	430860	166740
600	376140	5100	549540	540	900	521252	420	518640
1020	4140	480	180	600	53760	82440	180	273000
59880	840	7140	76320	148680	237840	4560	1920	16860
77040	74760	738180	147000	76680	70800	66180	27540	55020
120	296796	90180	724560	167100	106200	480	117360	6480
60	97860	398580	391380	180	180	240	540	336900
264480	847080	26460	349320	4080	64680	840	540	589980
332280	94140	240060	2700	900	1080	11580	2160	192720
87840	84360	378120	58500	83880	158640	660	3180	1560
3180	5700	226560	9840	69060	68880	65460	402900	75480
380220	704968	505680	54420	319020	95220	5100	6240	49440
420	667320	120	7200	68940	26820	448620	339420	480
1042680	779580	8040	1158240	907140	58500	383940	2039460	522240
66000	43500	2040	600	226320	327600	201300	226980	553440
1020	960	512760	819240	801660	160380	71640	363990	9090
227970	17190	597900	689400	11520	23850	75870	123030	26010
75240	68130	811050	498360	623280	3330	7290	47160	1328400
109800	343890	1615860	14940	680760	26220	376110	181890	64320
468180	1568580	333720	180	810	322110	21960	363600	

See Lyu’s data directory of failure data samples (SS3_DAT) – execution times in minutes of a telephone switch in early operational use

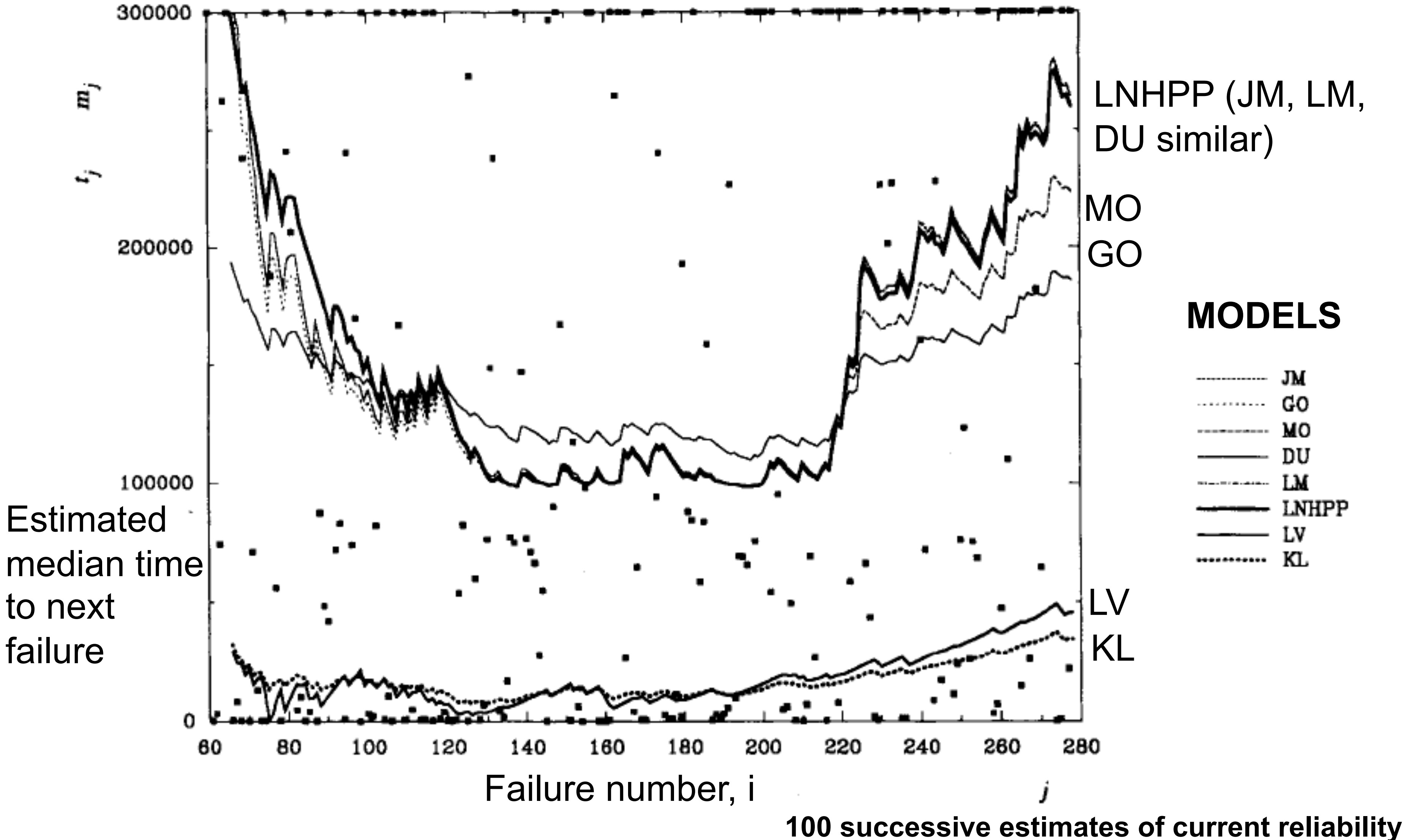
SS3 – Plotting Time Between Individual Failures

inter-failure time

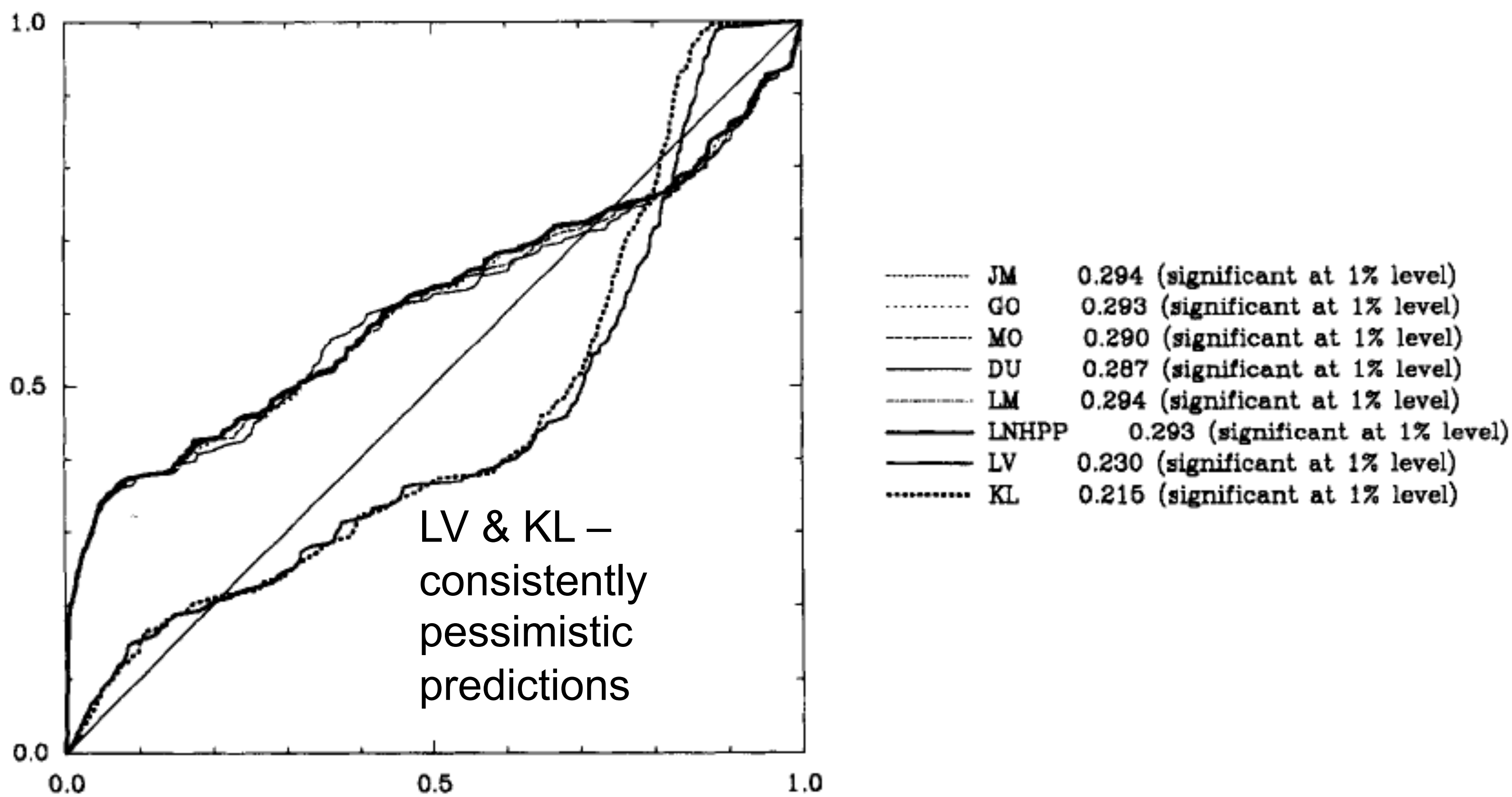


Failure number, i

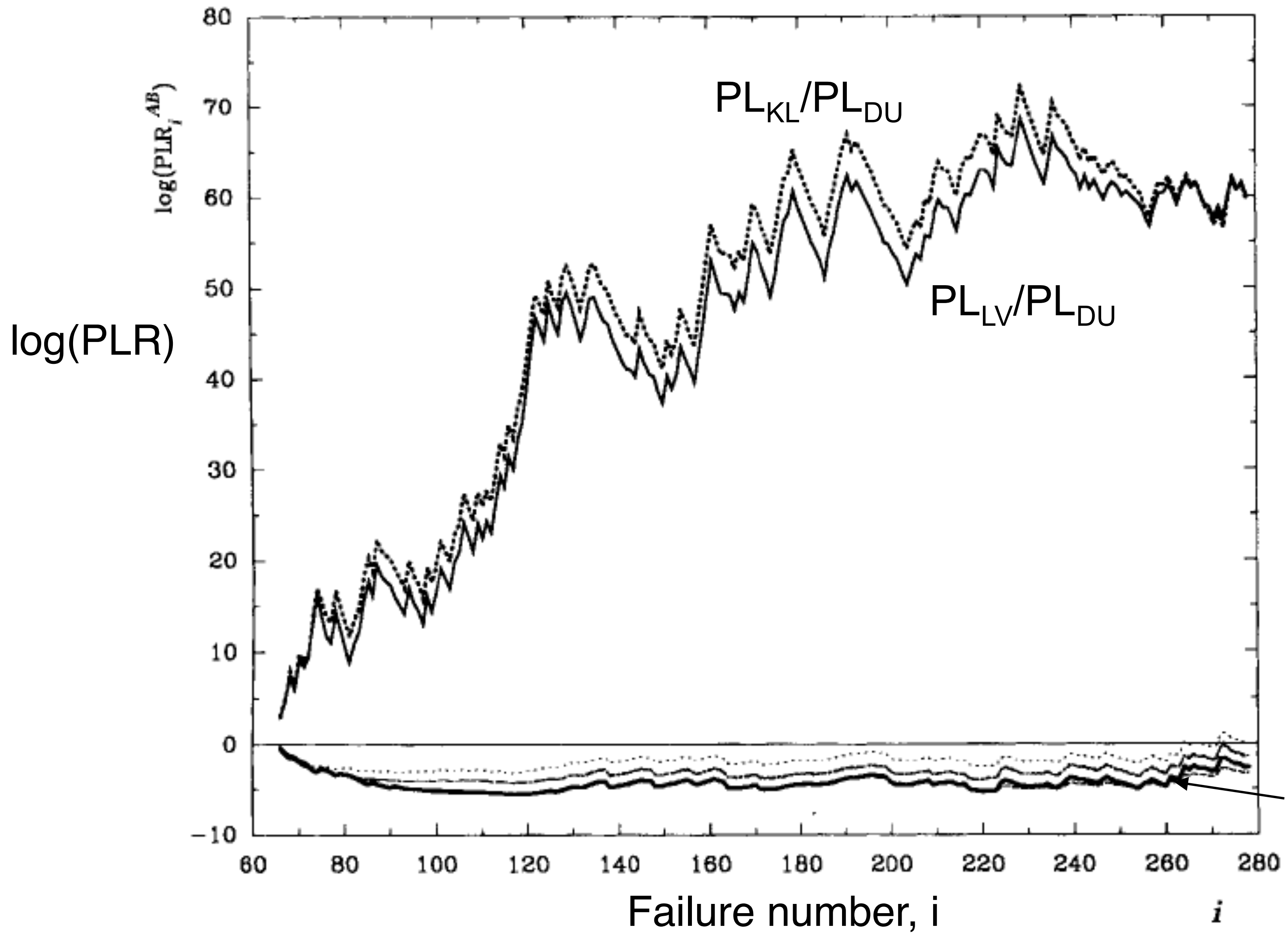
Time to Next Failure – 1 Step-ahead Median Prediction for 8 Models



SS3 - u-plots for the 8 Models



SS3 – Log(PLR) with Duane as Reference Model

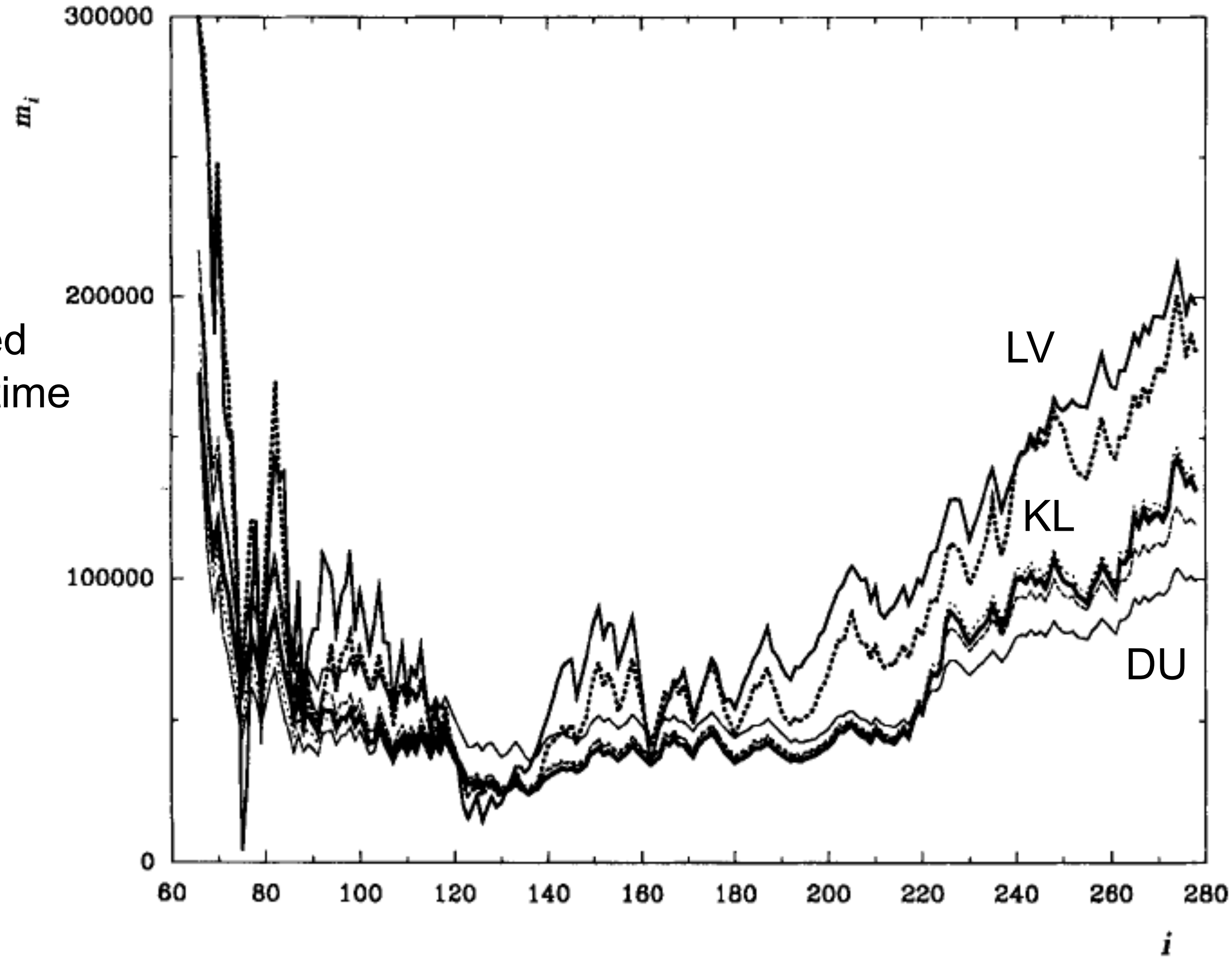


e^{60} - huge & very much better!

All other 5 models in agreement here

SS3 - Recalibrated Models, Medians

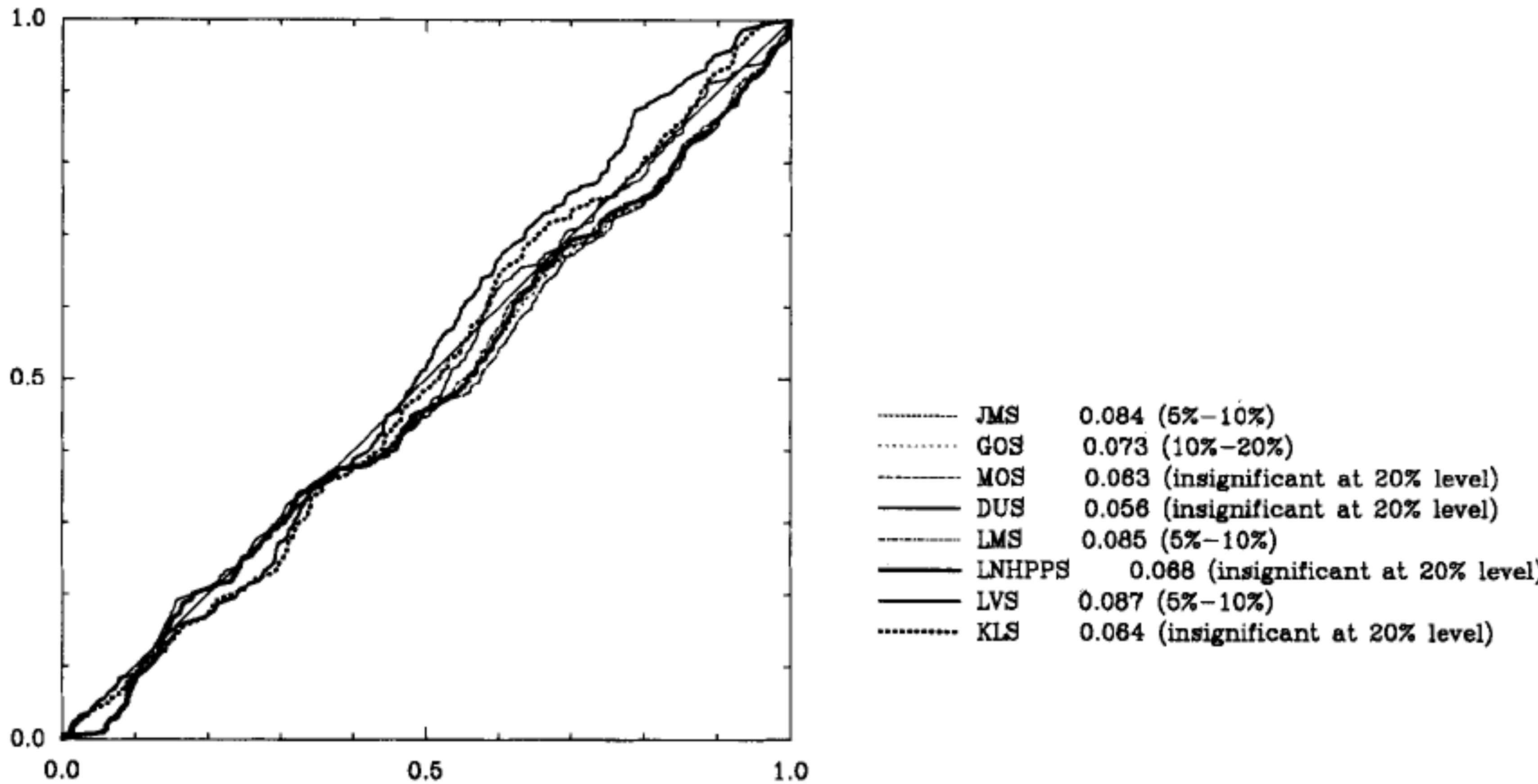
Estimated median time to next failure



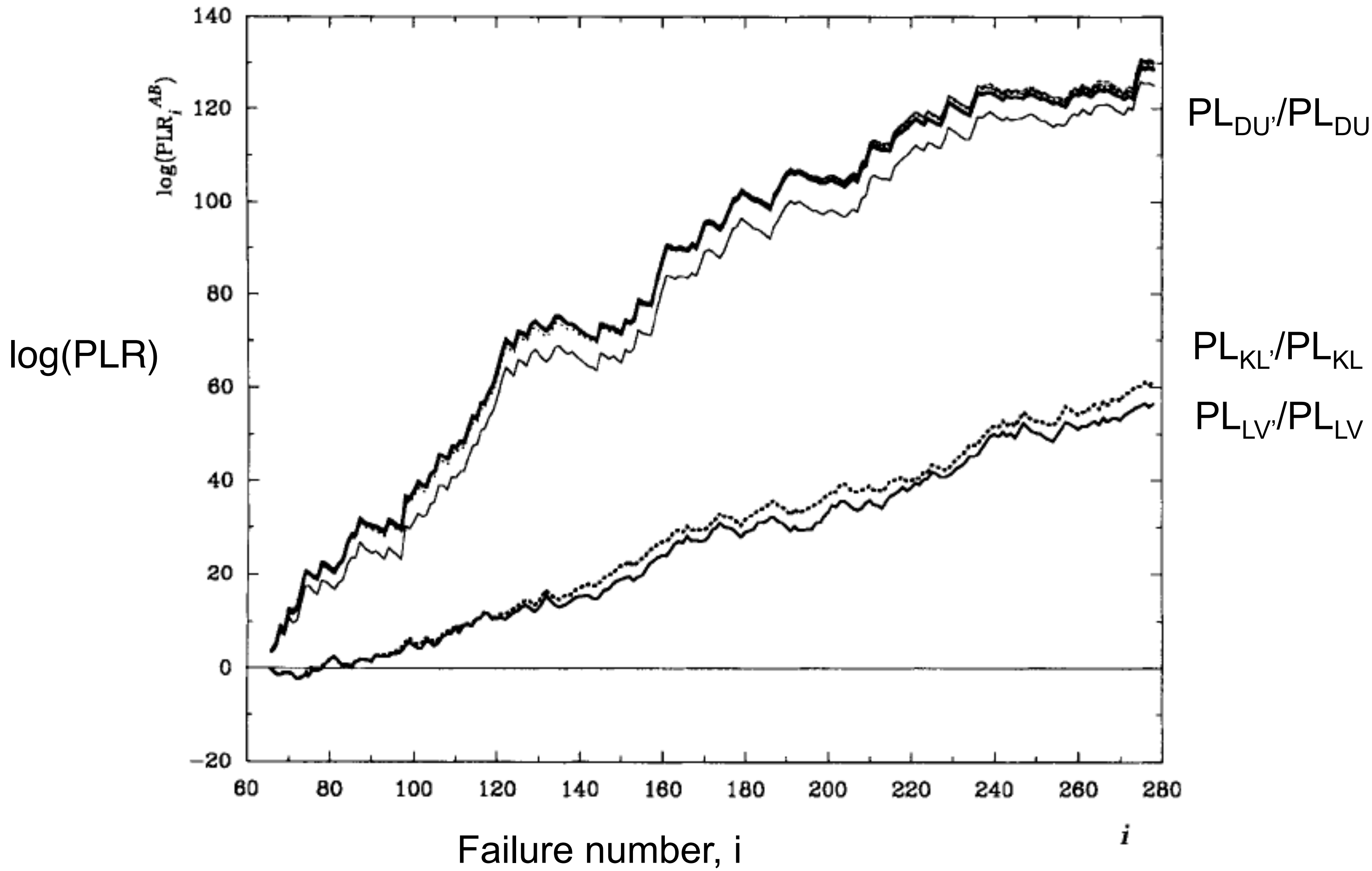
LNHPP (JM, LM, MO, GO similar)

Failure number, i

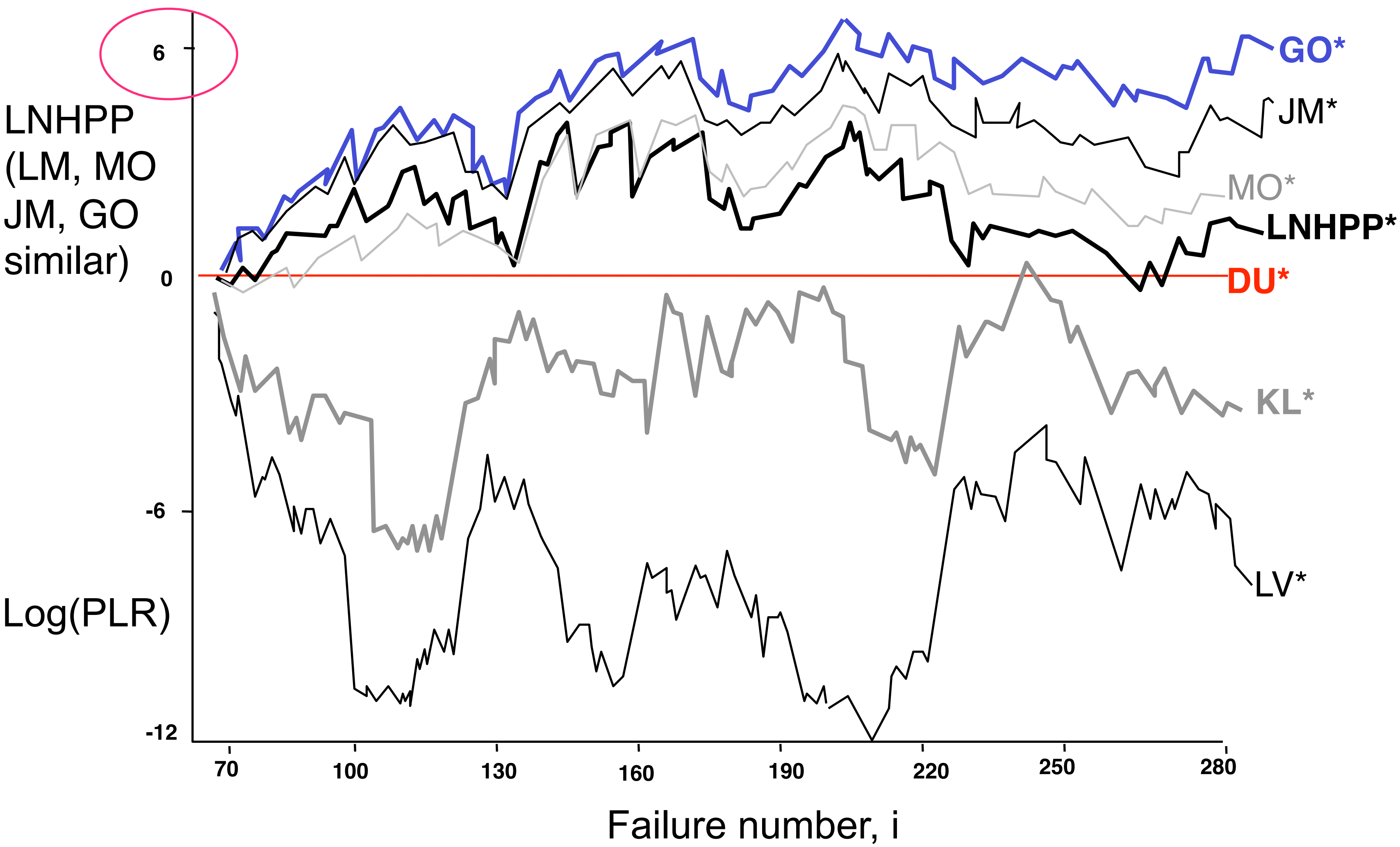
SS3 - Recalibrated Models, u-plot



SS3 – Log(PLR) – Recalibrated Model versus Raw Model



SS3 – Log(PLR) – Recalibrated Models Compared with Duane as Reference



Reading

- Chapter 4 of the Lyu text accompanies the material from this session
- New Ways to Get Accurate Software Reliability Measures, Sarah Brocklehurst and Bev Littlewood (see website)
- Recalibrating Software Reliability Models, Sarah Brocklehurst et al. (see website)
- Read over these to get a feel for the process (not the math!)

Future

- You will be required to install a tool that implements a number of the SRG models and supports the techniques described here – it will perform calculations and draw graphs for you
- I will direct you to a data set to import and explore with this tool
- I will expect an analysis from you of your findings, along the lines of that described in these slides
- We will look at a tool next time… so ensure you are ok with the general idea first, so do the reading!