Forecast accuracy measures for count data & intermittent demand

Stephan Kolassa

International Symposium on Forecasting, June 2015

Public

Means, medians and MADs

- Given any distribution…
  – The median minimizes the expected absolute error (Hanley et al., 2001)
  – The mean minimizes the expected squared error

Means, medians and MADs

- Given any distribution…
  - The median minimizes the expected absolute error (Hanley et al., 2001)
  - The mean minimizes the expected squared error

- Translate this into forecasting: given a (correctly specified) predictive distribution…
  - Forecast the median to minimize the expected MAD/MAE
  - Forecast the mean to minimize the expected MSE

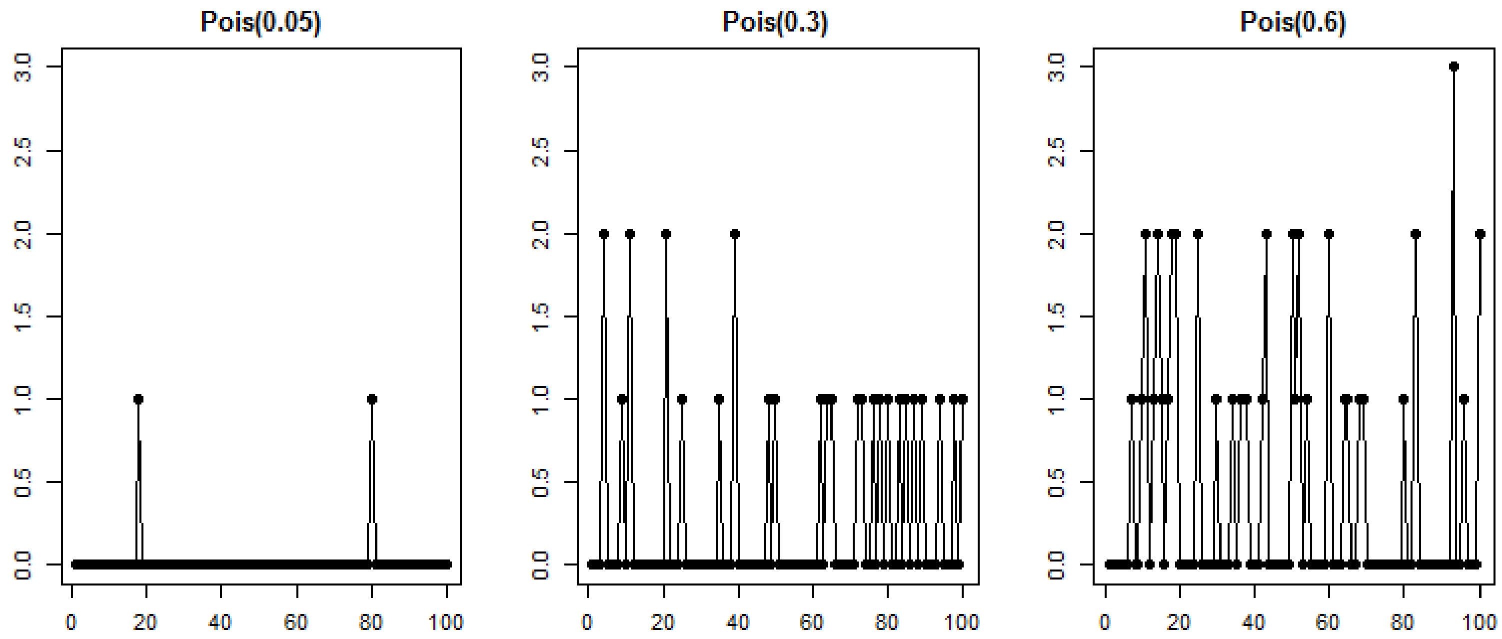
Means, medians and MADs

- Given any distribution…
  - The median minimizes the expected absolute error (Hanley et al., 2001)
  - The mean minimizes the expected squared error

- Translate this into forecasting: given a (correctly specified) predictive distribution…
  - Forecast the median to minimize the expected MAD/MAE
  - Forecast the mean to minimize the expected MSE

- Turn this around:
  - If you optimize your forecast method or parameters to minimize MAD and the future distribution is skewed, *your forecast will be biased* (Morlidge, 2015)!
  - This is particularly relevant for intermittent series (which are usually skewed), but also for non-intermittent low volume count series
  - This also applies to the MASE (Hyndman & Kohler, 2006) and the wMAPE (Kolassa & Schütz, 2007), which are simply scalar multiples of the MAD

An example: forecasting Poisson distributions



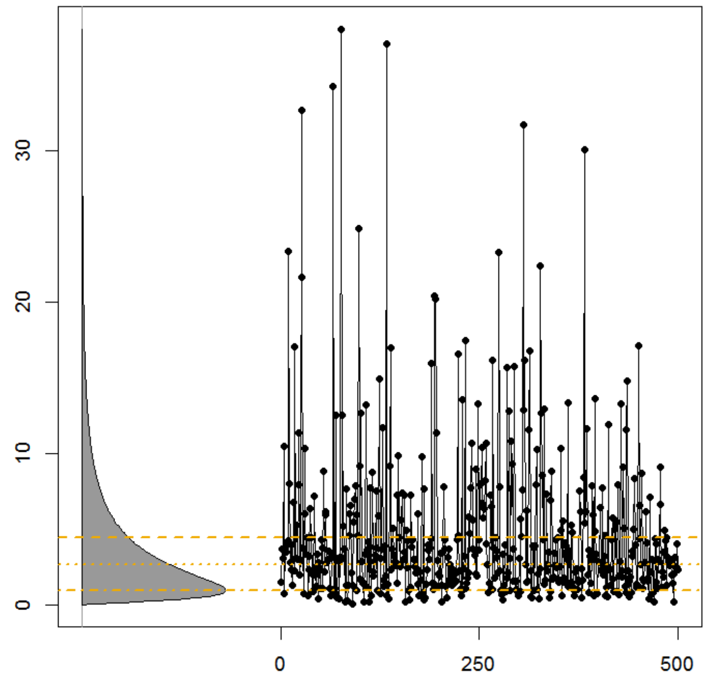
	Pois(0.05)	Pois(0.3)	Pois(0.6)
Forecast to minimize the expected MAD	0	0	0
Forecast to minimize the expected MSE	0.05	0.3	0.6

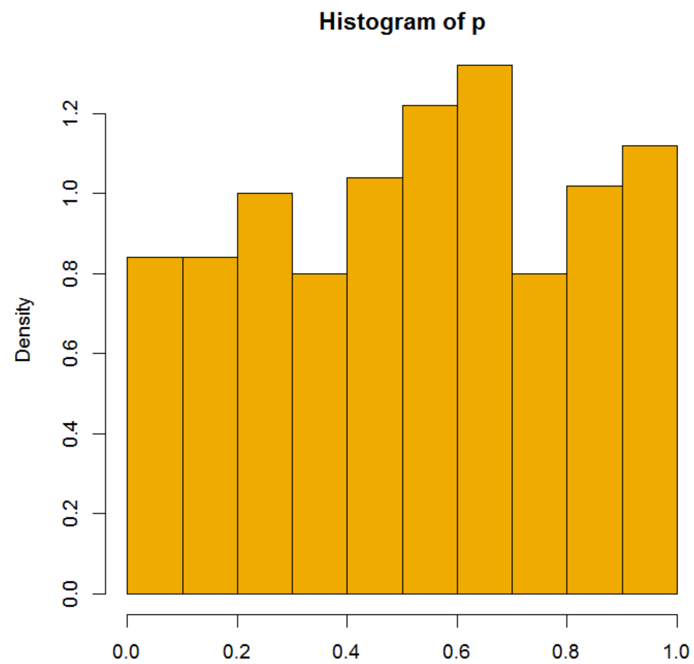
The Probability Integral Transformation (PIT)

- Assume predictive distribution with densities $\hat{f}_t$ and CDFs $\hat{F}_t$
- Transform observations y_t:

$$p_t := \hat{F}_t(y_t) = \int_{-\infty}^{y_t} \hat{f}_t = \hat{P}_t(Y_t \leq y_t)$$

- If the predictive distributions are correct, $\hat{f}_t = f_t$ and $\hat{F}_t = F_t$, then $p_t \sim U(0{,}1)$ – this can be tested (e.g., Ledwina, 1994; Berkowitz, 2001; or others)



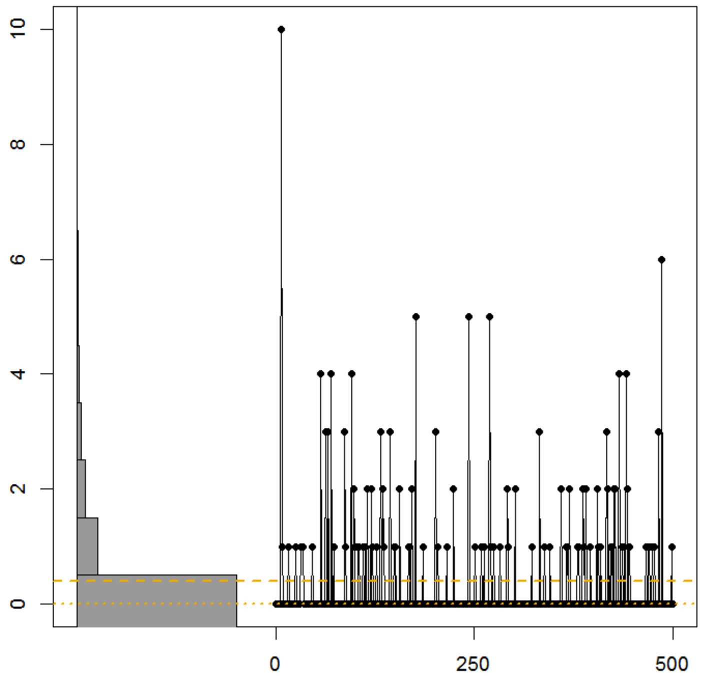


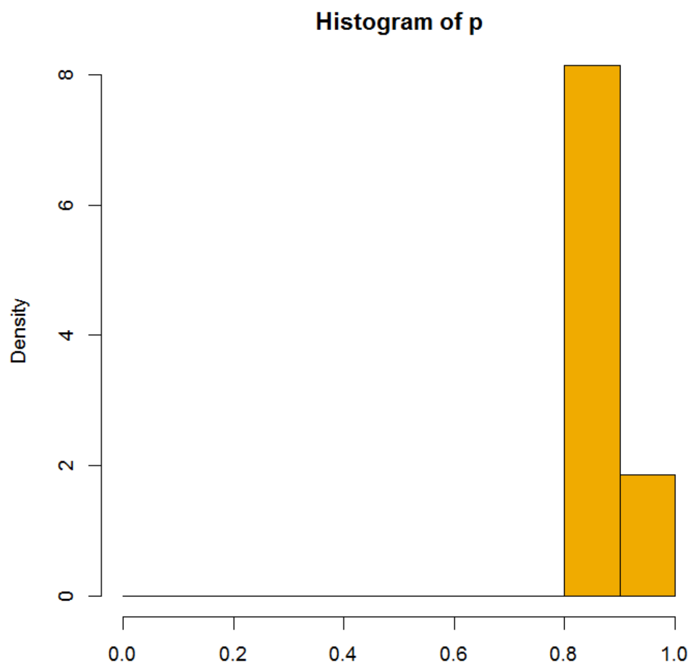
We have a problem with discrete predictive distributions!

- Transform observations y_t:

$$p_t := \hat{F}_t(y_t) = \int_{-\infty}^{y_t} \hat{f}_t = \hat{P}_t(Y_t \leq y_t)$$

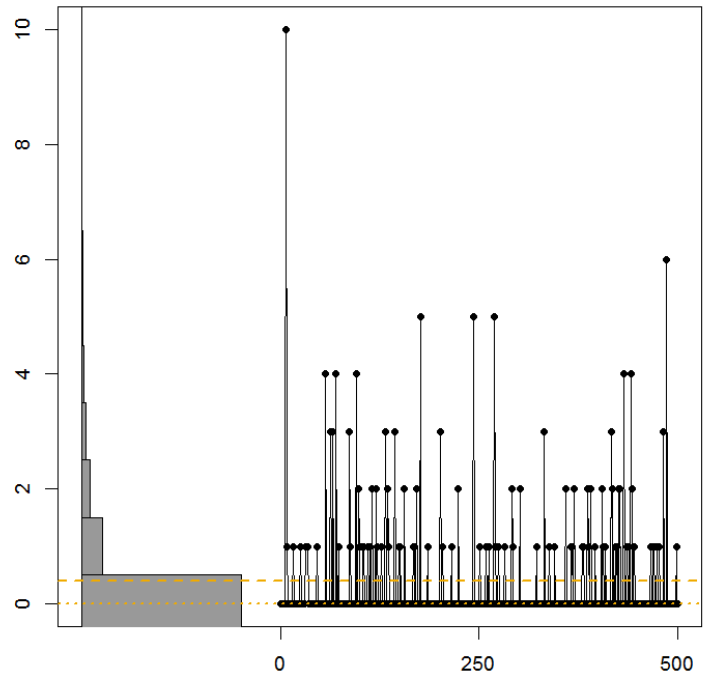
- Even if the predictive distributions are correct, p_t will have a discrete distribution if $\hat{F}_t = F$ is stationary!

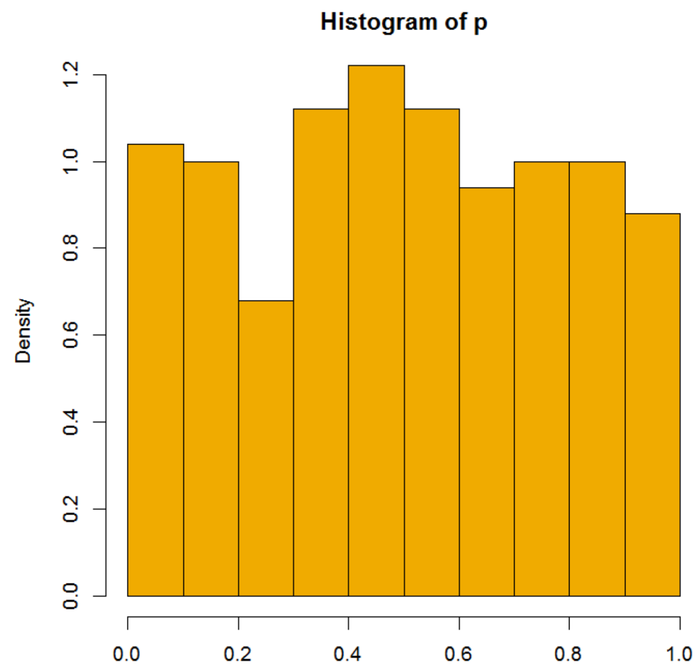




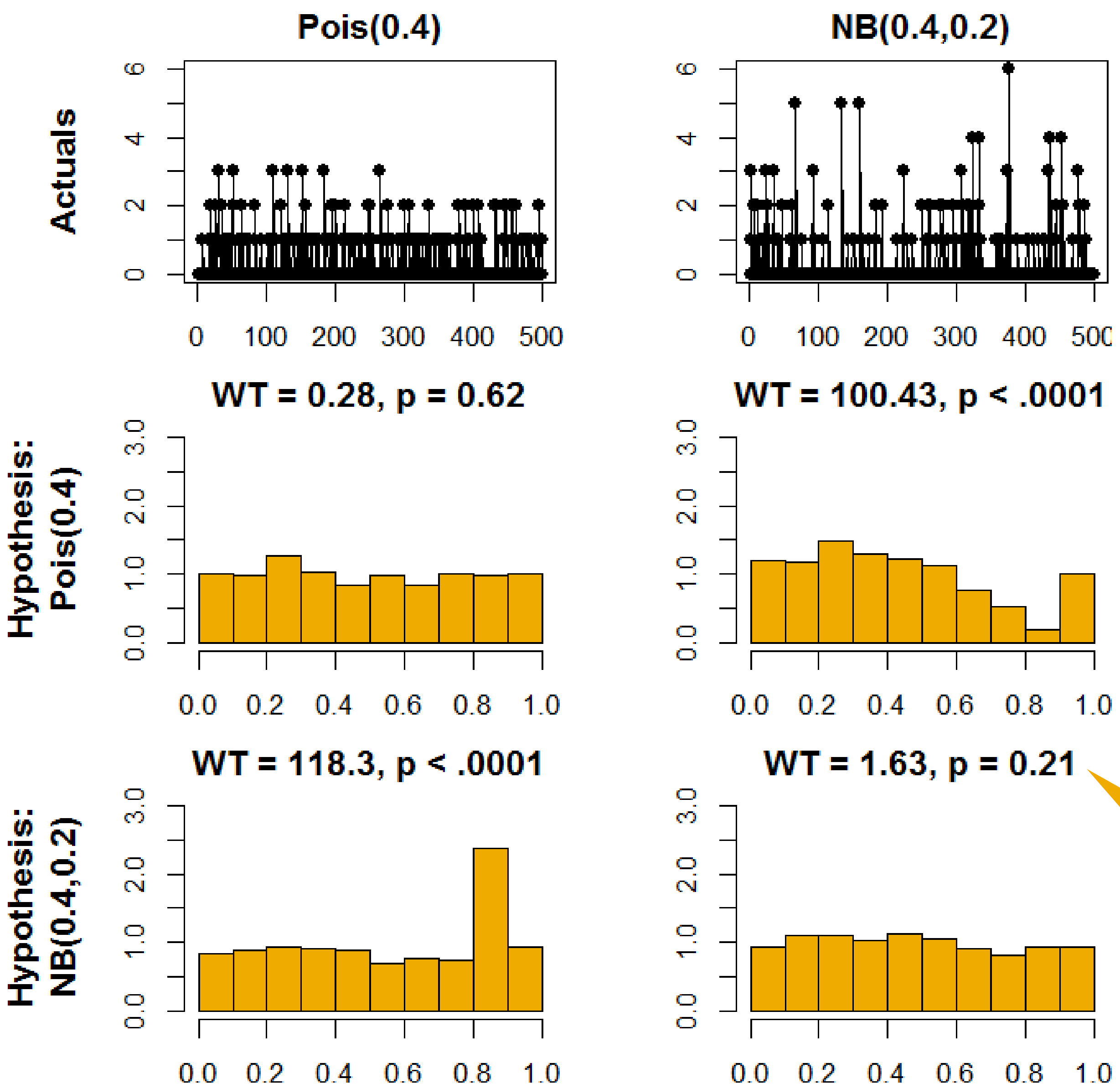
A solution for discrete predictive distributions

- Set $\hat{F}_t(-1) \coloneqq 0$, and draw uniform p_t:
$$p_t \sim U\left(\hat{F}_t(y_t - 1), \hat{F}_t(y_t)\right)$$
- Then $\hat{F}_t = F_t$ again implies that $p_t \sim U(0{,}1)$





Poisson vs. negative binomial



- Actuals and Hypotheses are either Pois(0.4) or NB(0.4,0.2)
  - In both cases, the median and MAD-optimal forecast is 0
  - In both cases, the expectation and MSE-optimal forecast is 0.4
  - Pois and NB differ heavily in the tails:

$$P_{\text{Pois}}(Y > 3) = 0.00078$$
$$P_{\text{NB}}(Y > 3) = 0.026$$
$$\frac{P_{\text{NB}}(Y > 3)}{P_{\text{Pois}}(Y > 3)} = 34.1$$

Data driven smooth tests for uniformity (Biecek & Ledwina, 2012):

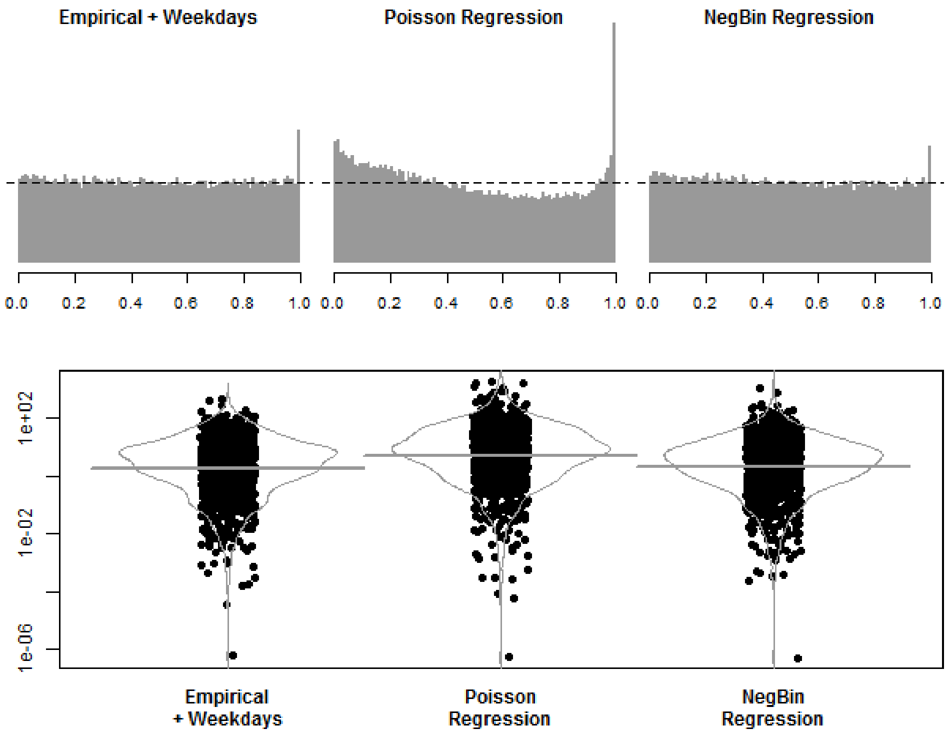
test statistics WT and p values

How to apply this to *multiple* series?

- Two possible summaries:
  - Simply “stack” all p values and test this big vector
  - Test each series’ p values, yielding a test statistic WT for each series – plot, summarize, compare these
- Two datasets with daily sales from European retailers
  - 1000 series each
  - Forecast horizon 100 days for each series
- Try multiple approaches – here, look at three:
  - Empirical + Weekday
  - Poisson Regression
  - NegBin Regression
  - Regressions include day of week, price, trend and Christmas

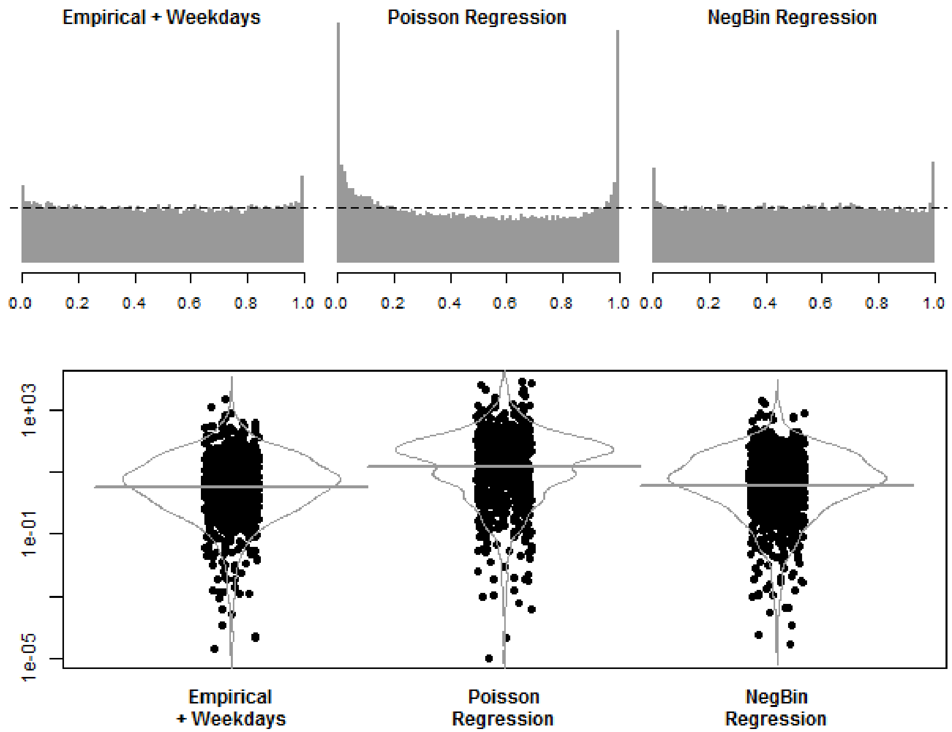
Results

Retailer A



	Empirical + Weekdays	Poisson Regression	NegBin Regression
95%	48.1	117.6	43.3
Median	2.6	6.8	2.7
5%	0.0	0.1	0.0
Mean	10.9	35.9	12.2

Retailer B



	Empirical + Weekdays	Poisson Regression	NegBin Regression
95%	101.5	512.1	110.4
Median	4.6	28.2	5.6
5%	0.0	0.1	0.0
Mean	26.9	142.2	30.8

- Poisson Regression obviously bad
- Empirical + Weekdays comparable to NegBin Regression

Conclusion

- Do not rely on the MAD et al. to find an unbiased forecast
  – If you do need to report MAD/wMAPE/MASE, also report bias

- Better: use a measure that is minimized by the expected value but is still scaled, e.g., a "relative Root Mean Squared Error":

$$\text{rRMSE} := \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}}{\frac{1}{n}\sum_{t=1}^{n} y_t}$$

- Yet better: forecast and assess full predictive densities, as we did here

- Finally: assess the *consequences* of your forecast
  – "Cost of Forecast Error"
  – "Forecast Value Added"
  – These will usually include both interval forecasts/predictive distributions *and* subsequent processes, like logistical optimization for replenishment

References

- Berkowitz, J. (2001). Testing Density Forecasts, With Applications to Risk Management. *Journal of Business and Economic Statistics*, 19, 465-474
- Biecek, P. & Ledwina, T. (2012). ddst: Data driven smooth test. R package version 1.03. http://CRAN.R-project.org/package=ddst
- Kolassa, S. & Schütz, W. (2007). Advantages of the MAD/Mean ratio over the MAPE. *Foresight*, 6, 40-43
- Hanley, J. A.; Joseph, L.; Platt, R. W.; Chung, M. K. & Belisle, P. (2001). Visualizing the Median as the Minimum-Deviation Location. *The American Statistician*, 55, 150-152
- Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688
- Ledwina, T. (1994). Data-Driven Version of Neyman's Smooth Test of Fit. *Journal of the American Statistical Association*, 89, 1000-1005
- Morlidge, S. (2015). Measuring the Quality of Intermittent Demand Forecasts: It's Worse than We've Thought! *Foresight,* 37, 37-42

Thank you!

Contact information:

Stephan Kolassa
Stephan.Kolassa@sap.com
Research Expert
Products & Innovation Suite Engineering Consumer Industries
SAP Switzerland