A New Variant of Croston's Method

Steve Prestwich[1] Roberto Rossi[2]
S. Armagan Tarim[3] Brahim Hnich[4]

[1]University College Cork, Ireland
[2]University of Edinburgh, UK
[3]Hacettepe University, Turkey
[4]Izmir University of Economics, Turkey

s.prestwich@cs.ucc.ie
Insight Centre for Data Analytics
Department of Computer Science
University College Cork
Ireland

Introduction

For forecasting intermittent demand the well-known Croston method, or one of its variants, is often applied.

But when *demand obsolescence* occurs (all demands are 0 after a given time period) most variants continue to forecast the same nonzero demand forever.

This motivated two recent variants that are designed to handle obsolescence. They differ qualitatively in the way their forecasts decay when demand is 0:

- those of the *Teunter-Syntetos-Babai* (TSB) method decay exponentially via smoothing
- those of the *Hyperbolic-Exponential Smoothing* (HES) method decay hyperbolically via Bayesian updating

Introduction

Both are unbiased on stochastic intermittent demand* and are competitive with SBA and SY on various demand patterns.

> **The occurrence of a nonzero demand is a Bernoulli event occurring at each time period with some probability. The magnitude of the demands may have any distribution.*

In experiments TSB handled obsolescence better than HES, while HES had more robust performance under parameter change.

In this talk I describe a new variant called *Linear-Exponential Smoothing* (LES) that asymptotically handles obsolescence better than both.

First some background...

Background

Simple Exponential Smoothing (SES) generates estimates $\hat{y}_t$ of the demand by exponentially weighting previous observations using the formula

$$\hat{y}_t = \alpha y_t + (1 - \alpha)\hat{y}_{t-1}$$

where $\alpha \in (0, 1)$ is a *smoothing parameter*.

SES is known to perform poorly on stochastic intermittent demand.

Background

A well-known method for handling intermittency is *Croston's method* (CR) which applies SES to demand size y and inter-demand interval τ independently, where $\tau = 1$ for non-intermittent demand:

$$\hat{y}_t = \alpha y_t + (1-\alpha)\hat{y}_{t-1}$$
$$\hat{\tau}_t = \beta \tau_t + (1-\beta)\hat{\tau}_{t-1}$$

(α, β might be different.) The CR forecast is

$$f_t = \frac{\hat{y}_t}{\hat{\tau}_t}$$

Both $\hat{y}_t$ and $\hat{\tau}_t$ are updated at each time t for which $y_t \neq 0$.

Background

CR was shown by [Syntetos & Boylan 2005] to be biased on stochastic intermittent demand. They corrected the bias (SBA) by modifying the forecasts to:

$$f_t = \left(1 - \frac{\beta}{2}\right) \frac{\hat{y}_t}{\hat{\tau}_t}$$

SBA works well for intermittent demand but is biased for non-intermittent demand, as its forecasts are those of SES multiplied by $(1 - \beta/2)$. This problem is avoided by [Syntetos 2001] who uses a forecast (SY)

$$f_t = \left(1 - \frac{\beta}{2}\right) \frac{\hat{y}_t}{\hat{\tau}_t - \beta/2}$$

This removes the bias on non-intermittent demand, though it increases the forecast variance.

Background

Another variant is described by [Levén & Segerstedt 2004] who apply SES to the ratio of demand size and inter-demand period when a nonzero demand occurs:

$$f_t = \alpha \left(\frac{y_t}{\tau_t} \right) + (1 - \alpha) f_{t-1}$$

This turns out to be biased on stochastic intermittent demand.

None of these variants handles obsolescence well. When obsolescence occurs they continue to forecast a fixed nonzero demand forever.

Background

The first Croston variant explicitly designed to handle obsolescence is TSB [Teunter, Syntetos, Babai 2011] which updates an estimate of the *demand probability* instead of the inter-demand interval.

Instead of $\hat{\tau}_t$ it uses a smoothed probability estimate $\hat{p}_t$ where p_t is 1 when demand occurs at time t and 0 otherwise. $\hat{p}_t$ is updated at every period while $\hat{y}_t$ is only updated when demand occurs.

The TSB forecast is

$$f_t = \hat{p}_t \hat{y}_t$$

It is unbiased.

Background

Another Croston variant designed to handle obsolescence is HES [Prestwich *et al.* 2014]. Like most Croston variants HES separates demands into y_t and τ_t. Its forecasts are

$$f_t = \begin{cases} \hat{y}_t/\hat{\tau}_t & \text{if } y_t > 0 \\ \hat{y}_t/(\hat{\tau}_t + \beta\tau_t/2) & \text{if } y_t = 0 \end{cases}$$

As usual $\hat{y}$ is updated when demand occurs, but $\hat{\tau}$ is updated at every period.

Between demands τ increases linearly, producing a hyperbolic decay in the forecasts: this can be viewed as Bayesian updating with a Beta prior distribution. HES is unbiased.

This talk

In this talk I describe a new variant called *Linear-Exponential Smoothing* (LES) that:

- is unbiased
- decays linearly to zero *in a finite time*
- asymptotically handles obsolescence better than TSB and HES
- performs well in experiments

(Full description available at `http://arxiv.org/abs/1409.1609`)

The LES method

LES is similar in form to HES but uses forecasts

$$f_t = \begin{cases} \hat{y}_t/\hat{\tau}_t & \text{if } y_t > 0 \\ (\hat{y}_t/\hat{\tau}_t)(1 - \beta\tau_t/2\hat{\tau}_t)^+ & \text{if } y_t = 0 \end{cases}$$

where x^+ denotes $\max(0, x)$.

When obsolescence occurs the forecasts decay linearly to zero at a rate determined by β. When they reach zero they remain there until nonzero demands reoccur. No extra parameters are needed beside α, β.

We can show that LES is unbiased on stochastic intermittent demand under the assumption that $1 - \beta\tau_t/2\hat{\tau}_t \geq 0$. If this assumption does not hold (which may occur if we set β to a high value) then the term will be replaced by 0, causing a positive bias, but in experiments this effect is negligible.

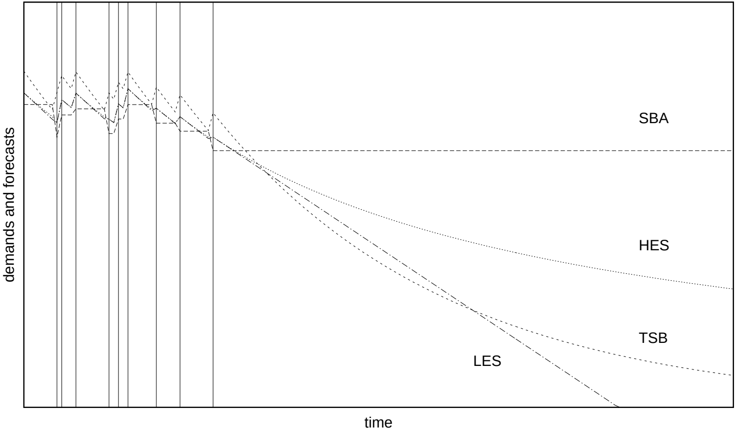
The LES method

Pseudocode for LES:

$\hat{y} \leftarrow 1$, $\tau \leftarrow 1$, $\hat{\tau} \leftarrow 1$
at each time period
  $y \leftarrow$ current demand
  if $y \neq 0$
    $\hat{y} \leftarrow \alpha y + (1 - \alpha)\hat{y}$
    $\hat{\tau} \leftarrow \beta\tau + (1 - \beta)\hat{\tau}$
    $f \leftarrow \hat{y}/\hat{\tau}$
    $\tau \leftarrow 1$
  else
    $f \leftarrow (\hat{y}/\hat{\tau})(1 - \beta\tau/2\hat{\tau})^{+}$
    $\tau \leftarrow \tau + 1$

The LES method

Behaviour of SBA, TSB, HES and LES:



(All forecasters use $\alpha = \beta = 0.1$, except that TSB uses $\beta = 0.02$ because a smaller value is recommended.)

Experiments

We compare LES, HES and TSB using experiments from [Teunter, Syntetos, Babai 2011].

Stationary demand (no obsolescence).

Demand is nonzero with probability p_0 where p_0 is either 0.2 or 0.5. Demand size is logarithmically distributed with two values to simulate low demand and lumpy demand. We test several combinations of smoothing factors and report the best. To compare forecasters we use Mean Error (ME) to measure bias, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Results. TSB and HES have lowest bias (ME) while HES and LES have lowest deviation (MAE and RMSE). LES has low bias (though not the lowest) so $1 - \beta\tau_t/2\hat{\tau}_t$ rarely becomes negative if we use reasonable β.

Decreasing demand

As above, but the probability of a nonzero demand decreases linearly from p_0 in the first period to 0 during the last period. As pointed out by Teunter *et al.*, none of the forecasters use trending to model the decreasing demand so all are positively biased.

Results. Under ME, MAE and RMSE, TSB ranks first, LES second and HES third.

Sudden obsolescence

Demand probability is reduced instantly to 0 after half the time periods.

Results. LES wins under ME and MAE, TSB wins under RMSE.

Summary

Winners:

demand	ME	MAE	RMSE
stationary	TSB+HES	HES+LES	HES+LES
decreasing	TSB	TSB	TSB
sudden	LES	LES	TSB

No clear winner emerges as the rankings depend on many factors: demand pattern, how long we compare forecasters before and after obsolescence occurs, which error measures we use for the comparison...

But all perform well on stationary demand, TSB is the clear winner under decreasing demand, and LES wins more often under sudden obsolescence. LES is highly competitive under all three error measures so it is a reasonable forecaster.

Asymptotic obsolescence error

Can we say more about which best handles obsolescence?

When average case analysis is hard we may resort to worst case analysis. We analyse the asymptotic behaviour of TSB, HES and LES to look for a definitive answer.

Worst-case scenario. Highly intermittent demand, sudden obsolescence, and a large number of forecasts after obsolescence occurs.

This represents a scenario in which an automated inventory control system continues to forecast nonzero demand for an obsolete item for a long time, because it believes demand to be highly intermittent based on previous data.

Asymptotic obsolescence error

We ignore the machine-dependent issue of arithmetic errors causing truncation to 0 as forecasts become small.

All the forecasters are unbiased so we assume they have the same forecast f_0 when obsolescence occurs at time 0.

We compute error measures for the 3 forecasters, using times starting from just after obsolescence occurs at time 0, up to some large $T \to \infty$.

But which error measures should we use? Not as easy as it seems as *most turn out to be inapplicable.*

Asymptotic obsolescence error

A surprising variety of measures have been used in the literature. There is no consensus on which is best so it is generally recommended to use several.

We consider all measures listed in the surveys of [de Gooijer, Hyndman 2005] and [Hyndman, Koehler 2006] and the article [Wallström, Segerstedt 2010].

Scale-dependent measures are based on the mean error $e_t = y_t - \hat{y}_t$ or mean square error e_t^2, and include Mean Error, Mean Square Error, Root Mean Square Error, Mean Absolute Error and Median Absolute Error. As $T \to \infty$ all these tend to zero so they cannot be used for an asymptotic comparison.

Asymptotic obsolescence error

Percentage errors are based on the quantities $p_t = 100e_t/y_t$ and include Mean Absolute Percentage Error, Median Absolute Percentage Error, Root Mean Square Percentage Error, Root Median Square Percentage Error, Symmetric Mean Absolute Percentage Error, and Symmetric Median Absolute Percentage Error. As $y_t = 0$ for all $t > 0$ these are undefined for almost all times.

Asymptotic obsolescence error

Relative error-based measures are based on the quantities $r_t = e_t/e_t^*$ where e_t^* is the error from a baseline forecaster, and include Mean Relative Absolute Error, Median Relative Absolute Error, and Geometric Mean Relative Absolute Error.

The baseline forecaster is usually the *random walk* (or *naive method*) which simply forecasts that the next demand will be identical to the current demand.

For almost all times $e_t^* = 0$ so these measures are undefined. We could use another baseline but we would still have the problem that the mean and median e_t are zero, so these cannot be used for a comparison.

Asymptotic obsolescence error

The *relative measures* are usually the ratio of (i) an error measure, and (ii) the same measure applied to a baseline forecaster. These include Relative Mean Absolute Error, Relative Mean Squared Error, and Relative Root Mean Squared Error (for example the U2 statistic). The baseline is again usually the random walk. Both measures $\rightarrow 0$ as $T \rightarrow \infty$ so these cannot be used.

A different form of relative measure is Percent Better, which computes the percentage of times a forecaster has smaller absolute error $|e_t|$ than a baseline forecaster, again usually random walk. Random walk has asymptotically perfect performance so Percent Better cannot be used.

Another is Percent Best in which no baseline forecaster is used: instead it computes the percentage of times each forecaster being tested has smaller absolute error than the others. We shall use this measure.

Asymptotic obsolescence error

The *scaled errors* include MAD/Mean Ratio [Kolassa, Schütz 2007]. It cannot be used because the denominator (the mean error) tends to zero.

Another is Mean Absolute Scaled Error [Hyndman, Koehler 2006]. It cannot be used because it is proportional to e_t which tends to zero.

Asymptotic obsolescence error

There are also three recent measures designed for intermittent demand.

Cumulative Forecast Error is the sum of all errors over the time periods under consideration. Not taking averages means that errors do not become vanishingly small, so this measure can be used.

The related (but not mentioned) Cumulative Squared Error can also be used. But Number of Shortages and Periods in Stock cannot be used.

Summary. Percent Best (PBt), Cumulative Forecast Error (CFE) and Cumulative Squared Error (CSE) are *the only error measures we know of that can be used for our comparison.*

Asymptotic obsolescence error

Analytical results:

	CFE	CSE	PBt
TSB	f_0/β	f_0^2/β^2	0%
HES	∞	$2f_0^2\hat{\tau}_0/\beta$	0%
LES	$f_0\hat{\tau}_0/\beta$	$2f_0^2\hat{\tau}_0/3\beta$	100%

(Derivations given in the online paper.)

CFE: HES is worst with infinite error. TSB beats LES if they use the same smoothing factor β, but it is recommended to use a smaller β for TSB. We do not know how β and $\hat{\tau}_t$ are related (inverse?) so the TSB & LES results are incomparable.

CSE: TSB is incomparable with HES and LES, but LES is 3 times better than HES.

PBt: LES beats both HES and TSB.

Asymptotic obsolescence error

LES is not beaten by TSB or HES under CFE or CSE, but beats both under PBt, so we rank LES as the best variant for handling obsolescence.

TSB beats HES under CFE, draws with it under PBt and is incomparable under CSE, so we rank it second best.

HES ranks third best.

Conclusion

LES is a variant of Croston’s method with several good features:

- It is unbiased on stochastic intermittent demand.
- It only requires two smoothing parameters to be tuned.
- It performs well in experiments.
- When obsolescence occurs its forecasts decay to 0 in a finite time.
- It asymptotically handles obsolescence better than all other variants.

THE END WE COULD WELCOME FEEDBACK ON HOW USEFUL THESE FEATURES ARE IN PRACTICE.