# Online Appendices for "Bayesian Ensembles of Exponentially Smoothed Life-Cycle Forecasts"

Xiaojia Guo, Kenneth C. Lichtendahl Jr., Yael Grushka-Cockayne

## A.1. Derivations of Results in the Paper

In this appendix, we provide derivation of our exponential smoothing model's forecasting method, along with the expression and skewness of the tilted-Gompert distribution. We also provide details of the widely applied diffusion models' cumulative distribution, mode, and skewness. Finally, we show how to relate the regret in a newsvendor problem to the pinball loss.

### A.1.1. Derivation of TiGo Exponential Smoothing Model's Forecasting Method

First, we show that the method's level updating equation follows from eliminating $\varepsilon_t$ from the transition equations using the measurement equation:

$$\begin{aligned}\ell_t &= \ell_{t-1} + \phi b_{t-1} + \alpha(\log(y_t) - (\ell_{t-1} + \phi b_{t-1}) \\ &= \alpha \log(y_t) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}).\end{aligned}$$

Second, we show that the method's growth updating equation follows similarly:

$$\begin{aligned}b_t &= \phi b_{t-1} + \log(\tau) + \beta/\alpha(\ell_t - \ell_{t-1} - \phi b_{t-1}) \\ &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1} + \log(\tau).\end{aligned}$$

where $\beta^* = \beta/\alpha$.

Third, we show that the mean and variance of $(\log(y_{t+h})|\boldsymbol{D}_t)$ follow from two induction arguments and substitutions into the measurement equation in (9). We show by induction that $b_{t+h} = \phi^h b_t + (1+\phi+\cdots+\phi^{h-1})\log(\tau) + \beta(\varepsilon_{t+h} + \phi\varepsilon_{t+h-1} + \cdots + \phi^{h-1}\varepsilon_{t+1})$ for any $h \in \{1,\ldots,n-t\}$. For $h=1$, $b_{t+1} = \phi b_t + \log(\tau) + \beta\varepsilon_{t+1}$ holds by the transition equation for the growth in (9). We assume the equation holds for $h$ and show that it holds for $h+1$:

$$\begin{aligned}b_{t+h+1} &= \phi b_{t+h} + \log(\tau) + \beta\varepsilon_{t+h+1} \\ &= \phi[\phi^h b_t + (1+\phi+\cdots+\phi^{h-1})\log(\tau) + \beta(\varepsilon_{t+h} + \phi\varepsilon_{t+h-1} + \cdots + \phi^{h-1}\varepsilon_{t+1})] \\ &\quad + \log(\tau) + \beta\varepsilon_{t+h+1} \\ &= \phi^{h+1} b_t + (1+\phi+\cdots+\phi^h)\log(\tau) + \beta(\varepsilon_{t+h+1} + \phi\varepsilon_{t+h} + \cdots + \phi^h\varepsilon_{t+1}),\end{aligned}$$

where the first equality follows from the transition equation for the growth in (9) and the second equality from our induction hypothesis.

Next we show by induction that $\ell_{t+h} = \ell_t + \sum_{i=1}^{h}\phi^i b_t + \sum_{i=2}^{h}(\phi+\cdots+\phi^{i-1})\log(\tau) + \sum_{i=1}^{h-1}(\alpha + \beta(\phi+\cdots+\phi^i))\varepsilon_{t+h-i} + \alpha\varepsilon_{t+h}$ for any $h \in \{1,\ldots,n-t\}$. For $h=1$, $\ell_{t+1} = \ell_t + \phi b_t + \alpha\varepsilon_{t+1}$ holds by

the transition equation for the level in (9). We assume the equation holds for $h$ and show that it holds for $h+1$:

$$\begin{aligned}
\ell_{t+h+1} &= \ell_{t+h} + \phi b_{t+h} + \alpha\varepsilon_{t+h+1} \\
&= \ell_t + \sum_{i=1}^{h} \phi^i b_t + \sum_{i=2}^{h} (\phi + \cdots + \phi^{i-1}) \log(\tau) + \sum_{i=1}^{h-1} (\alpha + \beta(\phi + \cdots + \phi^i))\varepsilon_{t+h-i} + \alpha\varepsilon_{t+h} \\
&\quad + \phi[\phi^h b_t + (1 + \phi + \cdots + \phi^{h-1}) \log(\tau) + \beta(\varepsilon_{t+h} + \phi\varepsilon_{t+h-1} + \cdots + \phi^{h-1}\varepsilon_{t+1})] + \alpha\varepsilon_{t+h+1} \\
&= \ell_t + \sum_{i=1}^{h+1} \phi^i b_t + \sum_{i=2}^{h+1} (\phi + \cdots + \phi^{i-1}) \log(\tau) + \sum_{i=1}^{h} (\alpha + \beta(\phi + \cdots + \phi^i))\varepsilon_{t+h+1-i} + \alpha\varepsilon_{t+h+1},
\end{aligned}$$

where the first equality follows from the transition equation for the level in (9) and the second equality from our induction hypothesis and our previous induction argument for $b_{t+h}$.

An expression for $\mu_{\boldsymbol{\theta}_t}(h)$, where $\boldsymbol{\theta}_t = (\phi, \tau, \ell_t, b_t)$, follows from substitutions according to the above expressions for $\ell_{t+h-1}$ and $b_{t+h-1}$ into the measurement equation in (9):

$$\begin{aligned}
\log(y_{t+h}|\boldsymbol{\theta}_t) &= \ell_{t+h-1} + \phi b_{t+h-1} + \varepsilon_{t+h} \\
&= \ell_t + \sum_{i=1}^{h-1} \phi^i b_t + \sum_{i=2}^{h-1} (\phi + \cdots + \phi^{i-1}) \log(\tau) + \sum_{i=1}^{h-2} (\alpha + \beta(\phi + \cdots + \phi^i))\varepsilon_{t+h-1-i} + \alpha\varepsilon_{t+h-1} \\
&\quad + \phi[\phi^{h-1} b_t + (1 + \phi + \cdots + \phi^{h-2}) \log(\tau) + \beta(\varepsilon_{t+h-1} + \phi\varepsilon_{t+h-2} + \cdots + \phi^{h-2}\varepsilon_{t+1})] + \varepsilon_{t+h} \\
&= \ell_t + \sum_{i=1}^{h} \phi^i b_t + \sum_{i=2}^{h} (\phi + \cdots + \phi^{i-1}) \log(\tau) + \sum_{i=1}^{h-1} (\alpha + \beta(\phi + \cdots + \phi^i))\varepsilon_{t+h-i} + \varepsilon_{t+h}.
\end{aligned}$$

The mean and variance of $(\log(y_{t+h})|\boldsymbol{D}_t)$ follow directly from this last equation.

Next, we prove that when $\alpha, \beta = 0$, the mean of the forecast follows (4), i.e. $\mu_{\boldsymbol{\theta}_0}(t+h) = \mu_{\boldsymbol{\theta}_t}(h)$. At time $t$, parameters $\ell_t$ and $b_t$ evolve deterministically according to

$$\begin{aligned}
b_t &= \phi^t b_0 + (1 + \phi + \cdots + \phi^{t-1}) \log(\tau), \\
\ell_t &= \ell_0 + \sum_{i=1}^{t} \phi^i b_0 + \sum_{i=2}^{t} (\phi + \cdots + \phi^{i-1}) \log(\tau).
\end{aligned} \tag{A.1}$$

From the previously provided expression of $\mu_{\boldsymbol{\theta}_t}(h)$, we have

$$\begin{aligned}
\mu_{\boldsymbol{\theta}_0}(t+h) = \log(y_{t+h}|\boldsymbol{\theta}_0) &= \ell_0 + \sum_{i=1}^{t+h} \phi^i b_0 + \sum_{i=2}^{t+h} (\phi + \cdots + \phi^{i-1}) \log(\tau), \\
\mu_{\boldsymbol{\theta}_t}(h) = \log(y_{t+h}|\boldsymbol{\theta}_t) &= \ell_t + \sum_{i=1}^{h} \phi^i b_t + \sum_{i=2}^{h} (\phi + \cdots + \phi^{i-1}) \log(\tau).
\end{aligned}$$

Substitute (A.1) into $\mu_{\boldsymbol{\theta}_t}(h)$, we have

$$\begin{aligned}
\mu_{\boldsymbol{\theta}_t}(h) &= \ell_0 + \sum_{i=1}^{t} \phi^i b_0 + \sum_{i=2}^{t} (\phi + \cdots + \phi^{i-1}) \log(\tau) + \sum_{i=1}^{h} \phi^i (\phi^t b_0 + (1 + \phi + \cdots + \phi^{t-1}) \log(\tau)) \\
&\quad + \sum_{i=2}^{h} (\phi + \cdots + \phi^{i-1}) \log(\tau) \\
&= \ell_0 + \sum_{i=1}^{t+h} \phi^i b_0 + \Big( \sum_{i=2}^{t} (\phi + \cdots + \phi^{i-1}) + \sum_{i=1}^{h} (\phi + \cdots + \phi^{i+t-1}) \Big) \log(\tau) \\
&= \ell_0 + \sum_{i=1}^{t+h} \phi^i b_0 + \Big( \sum_{i=2}^{t} (\phi + \cdots + \phi^{i-1}) + \sum_{i=t+1}^{t+h} (\phi + \cdots + \phi^{i-1}) \Big) \log(\tau) \\
&= \ell_0 + \sum_{i=1}^{t+h} \phi^i b_0 + \sum_{i=2}^{t+h} (\phi + \cdots + \phi^{i-1}) \log(\tau) \\
&= \mu_{\boldsymbol{\theta}_0}(t+h).
\end{aligned}$$

### A.1.2. Derivation of Tilted-Gompertz Distribution From Our Exponential Smoothing Model

In this derivation, we use a fact about the geometric series: $\sum_{i=0}^{h} r^i = 1 + \sum_{i=1}^{h} r^i = \frac{1-r^{h+1}}{1-r}$. Hence, we have that

$$\sum_{i=1}^{h} \phi^i = \frac{1-\phi^{h+1}}{1-\phi} - 1 = \frac{\phi}{1-\phi} - \frac{\phi^{h+1}}{1-\phi} = \frac{\phi}{1-\phi} - \frac{\phi}{1-\phi} e^{\log(\phi)h}$$

and

$$\sum_{i=2}^{h} (\phi + \cdots + \phi^{i-1}) = \sum_{i=2}^{h} \sum_{j=1}^{i-1} \phi^j = \sum_{i=2}^{h} \frac{\phi - \phi^i}{1-\phi} = \frac{\phi(h-1)}{1-\phi} - \frac{\phi^2}{(1-\phi)^2} + \frac{\phi}{(1-\phi)^2} e^{\log(\phi)h}.$$

We substitute into $e^{\mu_{\boldsymbol{\theta}_t}(h)}$ according to the above facts:

$$\begin{aligned}
e^{\mu_{\boldsymbol{\theta}_t}(h)} &= \ell_t^* b_t^{* \phi + \phi^2 + \cdots + \phi^h} \tau^{\sum_{i=2}^{h} (\phi + \cdots + \phi^{i-1})} \\
&= \ell_t^* b_t^{* \frac{\phi}{1-\phi} - \frac{\phi}{1-\phi} e^{\log(\phi)h}} \tau^{\frac{\phi(h-1)}{1-\phi} - \frac{\phi^2}{(1-\phi)^2} + \frac{\phi}{(1-\phi)^2} e^{\log(\phi)h}} \\
&= \ell_t^* b_t^{* \frac{\phi}{1-\phi} - \frac{\phi}{1-\phi} e^{\log(\phi)h}} \tau^{\frac{\phi h}{1-\phi} - \frac{\phi}{1-\phi} - \frac{\phi^2}{(1-\phi)^2} + \frac{\phi}{(1-\phi)^2} e^{\log(\phi)h}} \\
&= \ell_t^* b_t^{* \frac{\phi}{1-\phi} - \frac{\phi}{1-\phi} e^{\log(\phi)h}} \tau^{\frac{\phi h}{1-\phi} - \frac{\phi}{(1-\phi)^2} + \frac{\phi}{(1-\phi)^2} e^{\log(\phi)h}} \\
&= \ell_t^* b_t^{* \frac{\phi}{1-\phi} e^{-\frac{\phi b_t}{1-\phi} e^{\log(\phi)h}}} \tau^{-\frac{\phi}{(1-\phi)^2} e^{\frac{\log(\tau)\phi}{1-\phi} h} e^{\frac{\phi \log(\tau)}{(1-\phi)^2} e^{\log(\phi)h}}} \\
&= \ell_t^* b_t^{* \frac{\phi}{1-\phi}} \tau^{-\frac{\phi}{(1-\phi)^2} e^{\frac{\log(\tau)\phi}{1-\phi} h} e^{\left( \frac{\phi \log(\tau)}{(1-\phi)^2} - \frac{\phi b_t)}{1-\phi} \right) e^{\log(\phi)h}}} \\
&= \ell_t^* b_t^{* \frac{\phi}{1-\phi}} \tau^{-\frac{\phi}{(1-\phi)^2} \left( e^{\log(\phi)h} \right)^{\frac{\log(\tau)\phi}{\log(\phi)(1-\phi)}} e^{-\frac{\phi}{1-\phi} \left( b_t - \frac{\log(\tau)}{1-\phi} \right) e^{\log(\phi)h}}} \\
&= m_t c_t e^{-\lambda \delta h} e^{-\rho_t e^{-\lambda h}}.
\end{aligned}$$

where $\ell_t^* = e^{\ell_t}$, $b_t^* = e^{b_t}$, and $c_t = \lambda \rho_t^{\delta} / (\gamma(\delta, \rho_t) - I_{(-\infty,0)}(\lambda)\Gamma(\delta))$. Therefore, the tilted-Gompertz distrbution's pdf is

$$f_{TiGo}(h) = c_t e^{-\lambda \delta h} e^{-\rho_t e^{-\lambda h}}.$$

Given this expression of the pdf, the cdf of the tilted-Gompertz distribution is given by

$$\begin{aligned} F_{TiGo}(h) &= \frac{\lambda}{\gamma(\delta,\rho_t) - I_{(-\infty,0)}(\lambda)\Gamma(\delta)} \int_0^h (\rho_t e^{-\lambda s})^\delta e^{-\rho_t e^{-\lambda s}}\, ds \\ &= \begin{cases} -\frac{1}{\gamma(\delta,\rho_t)-\Gamma(\delta)} \int_{\rho_t}^{\rho_t e^{-\lambda t}} u^{\delta-1} e^{-u}\, du & \text{for} \quad -\infty < \lambda < 0 \\ \frac{1}{\gamma(\delta,\rho_t)} \int_{\rho_t e^{-\lambda t}}^{\rho_t} u^{\delta-1} e^{-u}\, du & \text{for} \quad 0 < \lambda < \infty \end{cases} \\ &= \frac{\gamma(\delta,\rho_t) - \gamma(\delta,\rho_t e^{-\lambda h})}{\gamma(\delta,\rho_t) - I_{(-\infty,0)}(\lambda)\Gamma(\delta)} \end{aligned}$$

where $u = \rho e^{-\lambda s}$ and $du = -\rho\lambda e^{-\lambda s} ds$.

### A.1.3. Derivation of the Tilted-Gompertz Distribution By Exponentially Tilting the Gompertz Distribution

This derivation begins with the Gumbel ($Gu$) distribution which also contains a term with double exponential. The cdf of a Gumbel distribution is $F_{Gu}(x) = e^{-\rho e^{-\lambda x}}$ for $-\infty < x < \infty$ where $0 < \rho, \lambda < \infty$. We denote the random variable $X$ drawn from a Gumbel distribution by $X \sim Gu(\lambda, \rho)$.

For the (possibly) left-skewed Gompertz ($LeGo$) distribution, we define an adoption-time $T$ as a reflected Gumbel random variable $-X$ truncated below zero: $T = (-X | -X \geq 0)$. It has the cdf given by

$$F_{LeGo}(t) = \frac{F_{Gu}(0) - F_{Gu}(-t)}{F_{Gu}(0)} = 1 - e^{-\rho(e^{-\lambda t} - 1)}$$

for $0 < t < \infty$. Its pdf is $f_{LeGo}(t) = \lambda\rho e^{\lambda t} e^{-\rho(e^{\lambda t} - 1)}$.

For the (always) right-skewed Gompertz ($RiGo$) distribution, we define an uncertain adoption-time $T$ as a Gumbel random variable $X$ truncated below zero: $T = (X | X \geq 0)$. It has the cdf given by

$$F_{RiGo}(t) = \frac{F_{Gu}(t) - F_{Gu}(0)}{1 - F_{Gu}(0)} = \frac{e^{-\rho e^{-\lambda t}} - e^{-\rho}}{1 - e^{-\rho}}$$

for $0 < t < \infty$. Its pdf is $f_{RiGo}(t) = \lambda\rho e^{-\lambda t} e^{-\rho e^{-\lambda t}} / (1 - e^{-\rho})$.

We put (possibly) left-skewed and (always) right-skewed Gompertz distributions together and define the Gompertz distribution to be the distribution with the pdf

$$f_{Go}(t) = \frac{\lambda\rho}{I_{(0,\infty)}(\lambda) - e^{-\rho}} e^{-\lambda t} e^{-\rho e^{-\lambda t}}$$

for $-\infty < \lambda < \infty$, $\lambda \neq 0$, $0 < \rho < \infty$, and $0 \leq t < \infty$ where the indicator $I_A(z) = 1$ if $z \in A$ and is zero otherwise.

The Gompertz distribution's moment generating function $M_{Go}$, evaluated at the point $\lambda(1-\delta)$ where $\delta > 0$, is given by

$$\begin{aligned} M_{Go}(\lambda(1-\delta)) &= \int_0^\infty e^{\lambda(1-\delta)t} f_{Go}(t)\, dt = \frac{\rho\lambda}{I_{(0,\infty)}(\lambda) - e^{-\rho}} \int_0^\infty e^{-\lambda\delta t} e^{-\rho e^{-\lambda t}}\, dt \\ &= \begin{cases} \frac{\rho^{1-\delta}}{e^{-\rho}} \int_\rho^\infty u^{\delta-1} e^{-u}\, du & \text{for} \quad -\infty < \lambda < 0 \\ \frac{\rho^{1-\delta}}{1-e^{-\rho}} \int_0^\rho u^{\delta-1} e^{-u}\, du & \text{for} \quad 0 < \lambda < \infty \end{cases} \\ &= \rho^{1-\delta} \frac{\gamma(\delta,\rho) - I_{(-\infty,0)}(\lambda)\Gamma(\delta)}{I_{(0,\infty)}(\lambda) - e^{-\rho}} \end{aligned}$$

where $u = \rho e^{-\lambda t}$, $du = -\rho\lambda e^{-\lambda t}\, dt$, $\gamma(\delta, s) = \int_0^s z^{\delta-1} e^{-z}\, dz$ is the lower incomplete gamma function, and $\Gamma(\delta) = \int_0^\infty z^{\delta-1} e^{-z}\, dz$ is the gamma function. Note that $\gamma(1,\rho) = 1 - e^{-\rho}$.

To tilt the Gompertz distribution, we multiple its pdf by $e^{\lambda(1-\delta)t}$ and divide by the moment generation function $M_{Go}(\lambda(1-\delta))$, which yields the pdf of the tilted-Gompertz ($TiGo$) distribution

$$f_{TiGo}(t) = \frac{e^{\lambda(1-\delta)t} f_{Go}(t)}{M_{Go}(\lambda(1-\delta))} = \frac{\lambda(\rho e^{-\lambda t})^\delta e^{-\rho e^{-\lambda t}}}{\gamma(\delta,\rho) - I_{(-\infty,0)}(\lambda)\Gamma(\delta)}$$

for $0 < \delta < \infty$, $-\infty < \lambda < \infty$, $\lambda \neq 0$, $0 < \rho < \infty$, and $0 < t < \infty$.

### A.1.4. Skewness of the Tilted-Gompertz Distribution

To measure skewness, we use the measure developed by Arnold and Groeneveld (1995). They define the skewness of any distribution with a single-peaked density as $Skew(F) = 1 - 2F(t^*)$, where $F(t^*)$ is the cdf evaluated at its mode $t^*$. This skewness measure ranges from -1 to 1, with negative values indicating left skew and positive values indicating right skew. The greater a positive measure, the more right-skewed we say the life cycle is, and conversely for left-skewed life cycles.

The cdf evaluated at the mode has a long history as a measure of skewness or asymmetry. Pearl and Reed (1925) and Winsor (1932) use $F(t^*)$ to describe the skewness of a general growth curve $mF$. In the marketing literature, Easingwood et al. (1983) and Mahajan et al. (1990) use $F(t^*)$ to describe asymmetric diffusions. Because the mode of the titled-Gompertz distribution is $t^* = -\log(\delta/\rho)/\lambda$, the skewness of the titled-Gompertz diffusion model is given by

$$Skew(F_{TiGo}) = 1 - 2\frac{\gamma(\delta,\rho) - \gamma(\delta,\delta)}{\gamma(\delta,\rho) - I_{(-\infty,0)}(\lambda)\Gamma(\delta)}.$$

When $\lambda > 0$ and $\rho > \delta$ (due to the single peak), the titled-Gompertz model is always right-skewed. In this case, the skewness measure can be written as $Skew(F_{TiGo}) = 2\gamma(\delta,\delta)/\gamma(\delta,\rho) - 1$. This measure decreases from 1 to $2\gamma(\delta,\delta)/\Gamma(\delta) - 1$, as $\rho$ goes from $\delta$ to $\infty$. This follows because $\gamma(\delta,\rho)$ is increasing in $\rho$ and $1/2 < \gamma(\delta,\delta)/\Gamma(\delta) < 1$. These last two inequalities hold because $\gamma(\delta,\delta)/\Gamma(\delta)$ is the cdf of a gamma distribution evaluated at $\delta$. The gamma distribution here has shape parameter $\delta$ and rate parameter 1 with cdf $F_{Ga}(t) = \gamma(\delta,t)/\Gamma(\delta)$. According to Chen and Rubin (1986), $\delta$ is greater than the median of this gamma distribution, so clearly $1/2 < \gamma(\delta,\delta)/\Gamma(\delta) < 1$.

When $\lambda < 0$ and $\rho < \delta$ (due to the single peak), the titled-Gompertz model can be left-skewed or right-skewed, depending on how much of the left tail is truncated. In this case, the skewness measure can be written as $Skew(F_{TiGo}) = 2(\Gamma(\delta) - \gamma(\delta,\delta))(\Gamma(\delta) - \Gamma(\delta)) - 1$. This measure increases from $1 - 2\gamma(\delta,\delta)/\Gamma(\delta)$ to 1, as $\rho$ increases from 0 to $\delta$. Since $1/2 < \gamma(\delta,\delta)/\Gamma(\delta) < 1$, the range of skewness when $\lambda < 0$ is $(-1, 1)$. In Section A.1.5, we provide the expressions of the four diffusion models' cdf, model (or peak time) and skewness.

### A.1.5. Properties of the Diffusion Models

Below we present a table which includes the density function, mode (or peak time), and skewness value for the diffusion models discussed in this paper: the Bass, gamma/shifted-Gompertz, trapezoid, and tilted-Gompertz models.

### A.1.6. Derivation of the Newsvendor's Deviation from Maximum Possible Profit

From the realized profit (17), we have

$$\begin{aligned}
\Pi(y,y) - \Pi(Q_u,y) &= (p-c)y - \big(p\min(Q_u,y) + s\max(Q_u - y, 0) - l\max(y - Q_u, 0) - cQ_u\big) \\
&= p(y - \min(Q_u,y)) - s\max(Q_u - y, 0) + l\max(y - Q_u, 0) + c(Q_u - y) \\
&= p\max(y - Q_u, 0) - s\max(Q_u - y, 0) + l\max(y - Q_u, 0) \\
&\quad + c\max(Q_u - y, 0) - c\max(y - Q_u, 0) \\
&= (p + l - c)\max(y - Q_u, 0) + (c - s)\max(Q_u - y, 0) \\
&= (p + l - s)\big(u\max(y - Q_u, 0) + (1-u)\max(Q_u - y, 0)\big) \\
&= (p + l - s)\begin{cases} u(y - Q_u) & \text{for} \quad Q_u \le y \\ (1-u)(Q_u - y) & \text{for} \quad Q_u > y \end{cases} \\
&= (p + l - s)L(Q_u, y).
\end{aligned}$$

where $u = (p + l - c)/(p + l - s)$.

**Table A.1 Properties of the Bass, Gamma/shifted-Gompertz, Trapezoid and Tilted-Gompertz Diffusion Models**

| | cdf ($F(t)$) | Mode ($t^*$) | Skewness |
|---|---|---|---|
| Bass | $\frac{1-e^{-(p+q)t}}{1+q/pe^{-(p+q)t}}$ | $-\frac{\log(p/q)}{(p+q)}$ if $q>p$ | $1-2\frac{1-p-q}{1+q+q^2/p}$ if $q>p$ |
| | $0<p,q<\infty$ | 0, otherwise | Not applicable otherwise |
| GSG[a] | $\frac{1-e^{-\lambda t}}{(1+1/\nu e^{-\lambda t})^{\mu}}$ | $-\frac{\log(\nu)}{\lambda}$ if $\mu=1$, $\nu<1$ | $2\nu-1$ if $\mu=1$, $\nu<1$ |
| | $0<\lambda,\nu,\mu<\infty$ | $-\frac{\log(x_1)}{\lambda}$ if $0<x_1<1\leq x_2$ | $1-2\frac{1-x_1}{(1+1/2x_1)^{\mu}}$ if $0<x_1<1\leq x_2$ |
| | | 0 and $-\frac{\log(x_1)}{\lambda}$ if $0<x_1<x_2<1$ | Not applicable otherwise |
| | | 0, otherwise | |
| Trap[b] | $\frac{1}{\kappa}(\frac{at^2}{2}+bt)$ if $0\leq t<\tau_1$ | $\frac{\tau_1+\tau_2}{2}$ | $\frac{1}{\kappa}\frac{a\tau_1\tau_2+b\tau_2+b\tau_1}{2}$ |
| | $\frac{1}{\kappa}(C_1+d(t-\tau_1))$ if $\tau_1\leq t<\tau_2$ | | |
| | $\frac{1}{\kappa}(C_2+(d-\frac{c\tau_2}{2}+\frac{ct}{2})(t-\tau_2))$ if $\tau_2\leq t\leq t_m$ | | |
| | $\frac{1}{\kappa}(C_3+(d-c\tau_2)(t_m-\tau_2))$ if $t_m<t$ | | |
| | $0<a,b,-c,\tau_1,\tau_2-\tau_1<\infty$ | | |
| TiGo | $\frac{\gamma(\delta,\rho)-\gamma(\delta,\rho e^{-\lambda t})}{\gamma(\delta,\rho)-I_{(-\infty,0)}(\lambda)\Gamma(\delta)}$ | $-\frac{\log(\delta/\rho)}{\lambda}$ if $\lambda(\rho-\delta)>0$ | $1-2\frac{\gamma(\delta,\rho)-\gamma(\delta,\delta)}{\gamma(\delta,\rho)-I_{(-\infty,0)}(\lambda)\Gamma(\delta)}$ if $\lambda(\rho-\delta)>0$ |
| | $-\infty<\lambda<\infty$, and $0<\delta,\rho<\infty$ | 0, otherwise | Not applicable otherwise |

[a] In the mode of the GSG distribution, $x_1=\frac{-B+\sqrt{B^2-4AC}}{2A}$, $x_2=\frac{-B-\sqrt{B^2-4AC}}{2A}$, where $A=-(\mu-1)^2$, $C=-\nu(\mu+\nu)$, and $B=\mu^2+3\mu\nu-2\nu$.

[b] The support of the Trapezoid model's density is$=0<t<t_m=\tau_2-(a+b\tau_1)/c$. Its normalizing constant $\kappa=-b\tau_1^2/2+(a+b\tau_1)\tau_2+c(t_{max}^2-\tau_2^2)/2+(a+b\tau_1-c\tau_2)(t_{max}-\tau_2)$. The other constants in its cdf are $d=a\tau_1+b$, $C_1=a\tau_1^2/2+b\tau_1$, $C_2=-a\tau_1^2/2+(a\tau_1+b)\tau_2$, and $C_3=-a\tau_1^2/2+(a\tau_1+b)\tau_2$.

## A.2. Supplemental Results and Discussions for the Empirical Studies

In this Appendix, we provide supplemental information for our empirical studies, including: additional study on Dell's computer sales data, sMAPE results for median forecasts, data collection and preparation in Studay 1, time-series cross-validation procedure and results, and the diffusion models' in-sample fit accuracy and their range of skewness.

### A.2.1. Additional Study on Dell's Computer Sales Data

For this study, we use the computer sales data described in Acimovic et al. (2019) and analyzed in Hu et al. (2018). These data include weekly sales of 170 different new Dell computers (fixed and mobile workstations, laptops, and desktops) in the United States from 2013 to 2016. During this period, 4,037,825 units of these products were sold, worth over one billion dollars in revenue. The publicly available data from Acimovic et al. (2019) are normalized so that the total sales over any complete life cycle is equal to 1.

**A.2.1.1. Data Collection and Preparation** Before we apply our approaches to these data, we unnormalize the data to approximately recover their actual sales volumes. Our unnormalization procedure is based on the summary statistics provided in Acimovic et al. (2019, Table 1). For each series, we randomly sample a cumulative sales value according to an approximation of the empirical distribution of cumulative sales. We then multiply each observation in that series by the sampled cumulative sales value.

In Table A.2, we reproduce Acimovic et al's (2019) Table 1. This table provides useful information about the joint distribution of the length of a product's per-period sales series and a product's weekly sales over its lifetime. We use this information to express a distribution for the cumulative sales of each series given its length. For each series, we unnormalize a normalized product life cycle by sampling from its distribution of cumulative sales given length and then multiplying the series by the sampled cumulative sales.

We use Bayes Theorem to inform our distribution of the series' weekly sales quintile $Q$ given its length $N$ is $n$. Here, the quintile $Q$ takes values in $\{1,2,3,4,5\}$, and the length $N$ takes values

**Table A.2 Summary statistics of the life cycle length and weekly sales**

| Weekly Sales | Life cycle length (in weeks) | | | Median of Mean |
|---|---|---|---|---|
| Quintile | 1st Quartile | 2nd Quartile | 3rd Quartile | Weekly Sales |
| 1 | 48 | 53 | 85 | 1021 |
| 2 | 47 | 58 | 75 | 456 |
| 3 | 34 | 49 | 62 | 267 |
| 4 | 23 | 54 | 61 | 166 |
| 5 | 22 | 39 | 58 | 66 |

(inclusively) between 13 and 118, the minimum and maximum lengths in the dataset, respectively. The conditional distribution of $(Q|N=n)$ is given by

$$\Pr(Q=q|N=n) = \frac{\Pr(Q=q)\Pr(N=n|Q=q)}{\Pr(N=n)}, \tag{A.2}$$

where the prior $\Pr(Q=q)=1/5$ for each quintile $q$, $\Pr(N=n|Q=q)$ is estimated from Table A.2, and $\Pr(N=n)=\sum_{q=1}^{5}\Pr(N=n|Q=q)/5$, which follows from the law of total probability. To estimate $\Pr(N=n|Q=q)$, we assume a uniform distribution for lengths within each length quartile in Table A.2. For example, in weekly sales quintile 1, we assume the first length quartile goes from 13 to 48 and the fourth length quartile goes from 86 to 118.

Our unnormalization procedure works as follows. For a series of length $n$, we first sample from the discrete distribution of $(Q|N=n)$ according to (A.2). Given a sampled length quintile $q$, we sample weekly sales $X$ from a normal distribution with mean $\mu_q$ and standard deviation of $0.1\mu_q$, where $\mu_q$ is the quintile $q$'s median of the mean weekly sales from Table A.2. We denote the sampled weekly sales by $x$, and our estimate of a series' cumulative sales $m$ becomes $nx$. We multiply each normalized series values by $nx$.

For example, given a series of length 70, suppose we sample quintile 1 from the discrete distribution of $(Q|N=n)$. Given we sampled quintile 1, we then sample weekly sales $X$ from a normal distribution with mean 1021 and standard deviation 10.21. If the sampled weekly sales is $x=1010$, then each value in the normalized series is multiplied by $70,700$. After this multiplication, we have a reasonable estimate of the unnormalized series.

**A.2.1.2. Accuracy Results** The typical practice at Dell is to make rolling forecasts on a quarterly basis for 34 weeks into the future. Because the median length of the series in this study is short (40 weeks), we make rolling forecasts each week with a horizon of 1-17 weeks ahead. Because we do not have the proprietary data on product features and categories discussed in Hu et al. (2018), we include in our Bayesian ensemble all past completed life cycles for each upcoming/ongoing life cycle. If we did have some proprietary data, we could use them to cluster products and include in our Bayesian ensemble only models fit to the completed life cycles with the upcoming/ongoing life cycle's cluster. Additionally, although we include the trapezoid model proposed in Hu et al. (2018) as a base model in our ensemble approach, we do not consider their data-driven clustering approach as a benchmark. This clustering approach uses the values of the upcoming series which amounts to data leakage, as these values would not be available at the time of prediction.

Our main results from this study are in Table A.3. The leading approach is again the Bayesian ensemble of TiGo-ETS model forecasts over both horizons and in both point and distributional forecasting. Also, the second best approach excluding the Bayesian ensembles is again the QRF

**Table A.3 MAE and MCRPS of the Thirteen Forecasting Approaches in the Supplemental Study.**

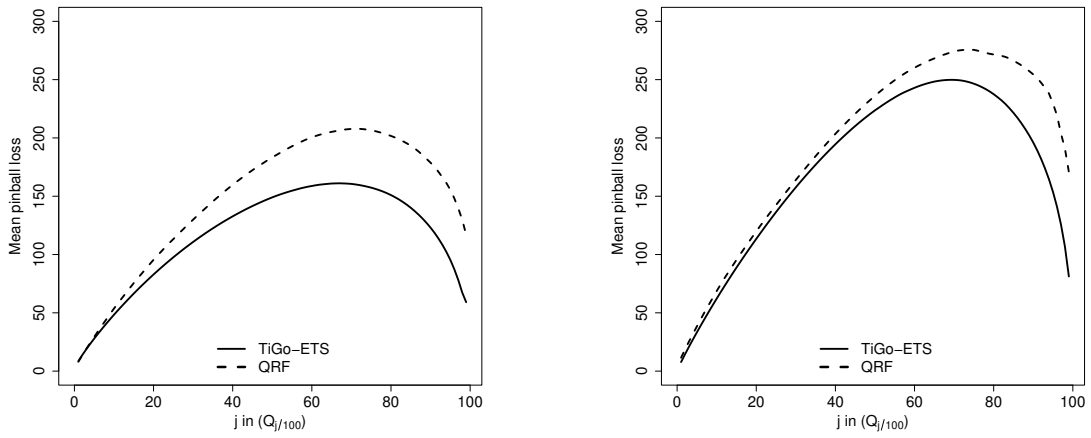
| | | 1-8 steps ahead | | 9-17 steps ahead | |
|---|---|---|---|---|---|
| | | MAE | MCRPS | MAE | MCRPS |
| Bayesian Ensembles of Single Model Type | Bass | 468.2 | 399.4 | 588.3 | 518.4 |
| | GSG | 413.2 | 346.2 | 573.6 | 498.0 |
| | Trap | 406.6 | 337.6 | 558.0 | 488.6 |
| | TiGo | 386.8 | 315.3 | 506.4 | 419.2 |
| | TiGo-ETS | **298.1**[a] | **233.2** | **447.3** | **350.0** |
| Bayesian Ensembles of Multiple Types | First 4 Model Types | 436.1 | 367.1 | 541.1 | 466.6 |
| | All 5 Model Types | 320.1 | 249.7 | 485.7 | 387.4 |
| Bayesian Updating Approaches | Bass-EKF(CD) | 1391.6 | 1312.8 | 1539.3 | 1423.6 |
| | TiGo-KF | 499.3 | 394.2 | 551.7 | 438.6 |
| | Bayesian NP | 469.0 | 375.5 | 654.3 | 500.1 |
| | Byesian FR | 533.8 | 420.8 | 682.7 | 533.5 |
| Machine Learning Approaches | LightGBM | 462.9 | 402.6 | 518.4 | 468.1 |
| | QRF | 367.6 | 298.4 | 473.1 | 395.4 |
| Significance Tests | Difference[b] | -69.5 | -65.2 | -25.8 | -45.4 |
| | Std. Error | 17.1 | 17.7 | 16.1 | 18.4 |
| | Significance[c] | *** | *** | * | ** |

[a] Values in bold indicate the lowest errors.
[b] Each difference is the TiGo-ETS loss minus the QRF loss.
[c] ***, **, and * indicate significance based on t-tests at the 0.1%, 1%, and 5% levels, respectively.

approach. Based on a grid search, the TiGo-ETS model's optimal hyperparameter values over the whole study period for the longer horizon was $\alpha = 0.3$ and $\beta = 0.0003$.

In Figure A.1, we see that the Bayesian ensemble of the TiGo-ETS model forecasts outperforms QRF at nearly every quantile in both shorter and longer horizons. For quantiles above the median, the Bayesian ensemble of the TiGo-ETS model forecasts performs especially well in this study when compared to the QRF approach.

**Figure A.1 Mean Pinball Loss for Each of 99 Quantiles from Two Forecasting Approaches in the Supplemental Study.**




(a) 1-8 weeks ahead

(b) 9-17 weeks ahead

### A.2.2. Measuring Point Forecasting Accuracy with Symmetric Mean Absolute Percentage Error (sMAPE)

In Table A.4, we present the sMAPE results for the two empirical studies in the paper, along with those from the additional study discussed in Appendix A.2.1. The sMAPE is calculated as $2/n\sum_{t=1}^{n}(|y_t-\mu_t|/(|y_t|+|\mu_t|))$, where $\mu_t$ is the prediction and $y_t$ is the realization. Our Bayesian Ensemble of TiGo-ETS models outperforms the other models in most cases. In the study of Dell computer sales, it has a slightly higher sMAPE than QRF for 9-17 weeks ahead forecasts. However, the difference between these two methods is not statistically significant.

### A.2.3. Data Collection and Preparation in Study 1

We gathered data on the 161 social networks used in Study 1 from Wikipedia (2021a and 2021b). We did not gather data on all of the social networks listed on Wikipedia for the following reasons: (i) no data were available from Google Trends, (ii) the network's name was too generic to associate its name with search interest in the network, (iii) search interest was high well before the network's launch date, or (iv) the network was founded before January 2004 and the first observation in the series is greater than one.

To retrieve the data from Google Trends, we manually entered the network's name and, whenever possible, chose the topic "Social network" or a related topic, rather than the generic topic "Search

**Table A.4 Evaluation of the Point Forecasts Using sMAPE**

| | | Study 1 Search Interests | | Study 2 Intel Processor Sales | | Supplemental Study Dell Computer Sales | |
|---|---|---|---|---|---|---|---|
| | | 1-12 steps | 13-24 steps | 1-8 steps | 9-17 steps | 1-8 steps | 9-17 steps |
| Bayesian Ensembles of Single Model Type | Bass | 0.608 | 0.739 | 0.679 | 0.764 | 0.450 | 0.622 |
| | GSG | 0.612 | 0.730 | 0.690 | 0.787 | 0.408 | 0.534 |
| | Trap | 0.637 | 0.772 | 0.705 | 0.792 | 0.440 | 0.564 |
| | TiGo | 0.580 | 0.717 | 0.665 | 0.781 | 0.372 | 0.432 |
| | TiGo-ETS | **0.191**[a] | **0.315** | **0.358** | **0.462** | **0.294** | 0.384 |
| Bayesian Ensembles of Multiple Model Types | First four models | 0.557 | 0.697 | 0.674 | 0.772 | 0.389 | 0.520 |
| | All five models | 0.206 | 0.351 | 0.362 | 0.467 | 0.307 | 0.402 |
| Bayesian Updating Approaches | Bass-EKF(CD) | 0.526 | 0.611 | 0.550 | 0.663 | 0.422 | 0.513 |
| | TiGo-KF | 0.227 | 0.496 | 0.553 | 0.717 | 0.310 | 0.384 |
| | Bayesian NP | 0.444 | 0.503 | 0.686 | 0.778 | 0.405 | 0.497 |
| | Byesian FR | 0.407 | 0.439 | 0.525 | 0.566 | 0.632 | 0.747 |
| Machine Learning Approaches | LightGBM | 0.384 | 0.608 | 0.984 | 0.958 | 0.412 | 0.386 |
| | QRF | 0.207 | 0.342 | 0.441 | 0.563 | 0.330 | **0.350** |
| Significance Tests | Difference[b] | -0.016 | -0.027 | -0.083 | -0.100 | -0.036 | 0.034 |
| | Std. Error | 0.004 | 0.009 | 0.015 | 0.014 | 0.010 | 0.014 |
| | Significance[c] | *** | *** | *** | *** | *** | |

[a] Values in bold indicate the lowest errors.
[b] Each difference is the TiGo-ETS loss minus the QRF loss.
[c] *** indicates significance based on t-tests at the 0.1% level.

term". For networks with generic names, the idea was to separate search interest in the network from search interest in some other item associated with that name. For example, in Google Trends, the term "Delicious" was listed with a topic of "Search term" and separately with a topic of "Social bookmarking website".

Between the start and end dates of some series, we find a few zero series' values. Because our models require all positive series' values and presumably some searching occurs every month for an active network, we impute the zeros in these series. We replace the zeros in these series with the minimum positive value in the series, divided by 2. We note that our results are not sensitive to dividing by as much 5 or 10. Some series have large values at their listed end date or have small values for a long period before their listed end date. For these series, we determine a new end date. The new end date is set to be the first month when search interest less than a threshold of 5 for six consecutive months. Our results are not sensitive to changing this threshold to 1.

### A.2.4. Time Series Cross-Validation and Results

Our cross-validation procedure involves two main folds. In Study 1 presented in the main body of the paper, fold 1 is comprised of rolling forecasts with forecast origin dates running from October 2009 to September 2012, and fold 2 has forecast origin dates running from October 2012 to May 2020. September 2012 is the midpoint between the beginning and ending dates of the data: January 2004 to May 2021. For the three models with tuning parameters, we search for the optimal parameter values for the longer horizon in fold 1, and report in Table A.5 the models' fold 2 performance under these fold 1 optimal parameter settings. We can see from the table that our Bayesian ensemble of TiGo-ETS models performs significantly better than the other two methods in fold 2.

In Study 2, we again split the data at the midpoint, which is week 94 in the range between week 1 and week 187. In the Supplemental Study, fold 1 ends and fold 2 begins by the time half of the life cycles have started (February 1, 2014). We did not use the midpoint between the beginning and ending dates of the data as in Study 1 and 2 because 91 out of 170 time series had ended before this date (January 10, 2015).

**Table A.5 Evaluation of the Three Approaches with Tuning Parameters for Fold 2 in Study 1.**

| | 1-12 months ahead | | 13-24 months ahead | |
|---|---|---|---|---|
| | MAE | MCRPS | MAE | MCRPS |
| TiGo-ETS | **6.73**[a] | **5.36** | **12.24** | **9.66** |
| LightGBM | 16.16 | 7.37 | 33.92 | 15.89 |
| QRF | 7.04 | 5.62 | 13.36 | 11.78 |
| Difference[b] | -0.31 | -0.26 | -1.12 | -2.12 |
| Std. Error | 0.18 | 0.36 | 0.65 | 0.62 |
| Significance[c] | * | * | * | *** |

[a] Values in bold indicate the lowest errors.
[b] Each difference is the TiGo-ETS loss minus the QRF loss.
[c] *** and * indicate significance based on t-tests at the 0.1% and 5% levels, respectively.

**Table A.6 MAE($\times 10^3$) and MCRPS($\times 10^3$) of the Approaches with Tuning Parameters for Fold 2 in Study 2.**

| | 1-12 months ahead | | 13-24 months ahead | |
|---|---|---|---|---|
| | MAE | MCRPS | MAE | MCRPS |
| TiGo-ETS | **16.62**[a] | **12.64** | **22.10** | **16.99** |
| LightGBM | 32.84 | 32.56 | 35.47 | 35.17 |
| QRF | 16.89 | 13.26 | 22.91 | 19.67 |
| Difference[b] | -0.28 | -0.63 | -0.81 | -2.7 |
| Std. Error | 0.51 | 0.49 | 0.58 | 0.71 |
| Significance[c] | | · | · | *** |

[a] Values in bold indicate the lowest errors.
[b] Each difference is the TiGo-ETS loss minus the QRF loss.
[c] *** and · indicate significance based on t-tests at the 0.1% and 10% levels, respectively.

From Table A.6 for Study 2, we see that our TiGo-ETS model outperforms the machine learning methods in fold 2. However, the difference between our new approach and the QRF model is not as significant as observed in the previous study. Based on Supplemental Study's results shown in Table A.7, we can see that our Bayesian ensemble of TiGo-ETS models again significantly outperforms the machine learning models over both horizons and in both point and distributional forecasting.

Overall, Table 2, 3 and A.3 present accuracy results spanning the entire study period, thereby offering a statistically more meaningful interpretation due to the inclusion of more periods. These results also contain periods with few prior life cycles, a scenario frequently encountered in practice. On the other hand, the three tables in this section show results starting from somewhere around the middle of the entire time period covered by the data. During this period, more life cycles have been completed and more data from the current time series is available. The inclusion of these results, however, is valuable as they illustrate the models' performance in real-world scenarios when we need to tune parameters with a validation set. Together, tables in this section and those in Section 4 offer a comprehensive view of the models' performance.

**Table A.7 Evaluation of the Three Approaches with Tuning Parameters for Fold 2 in the Supplemental Study.**

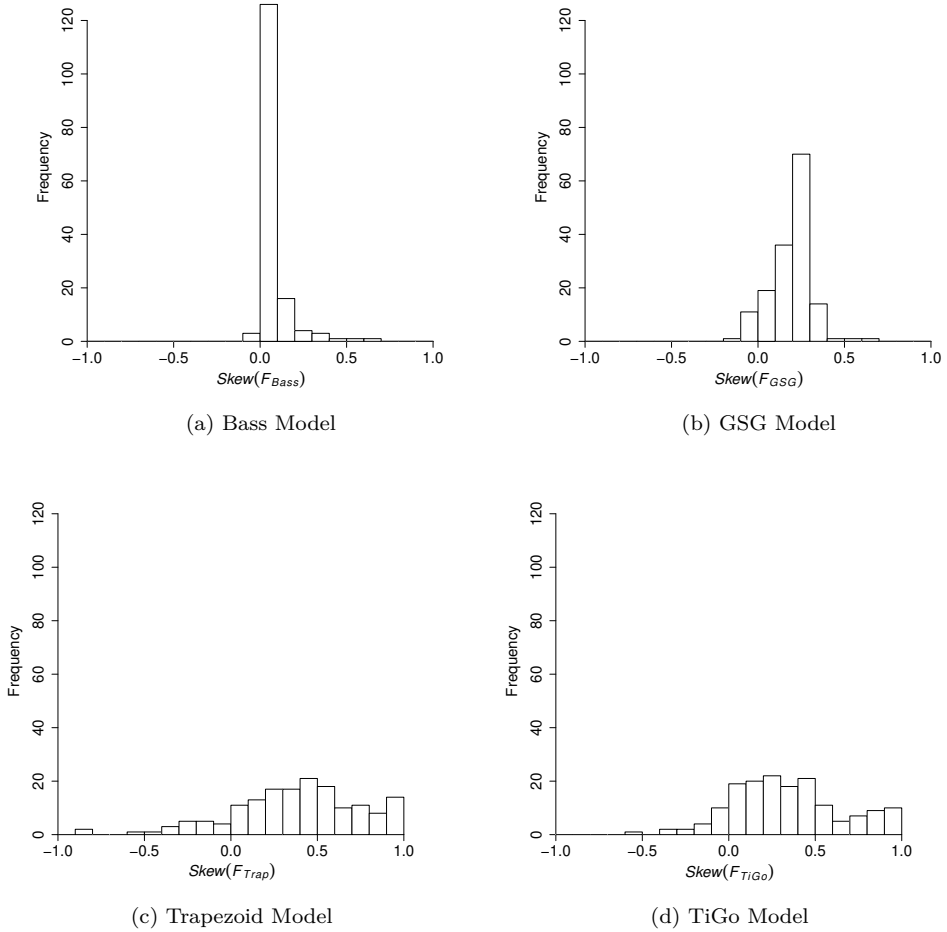
| | 1-12 months ahead | | 13-24 months ahead | |
|---|---|---|---|---|
| | MAE | MCRPS | MAE | MCRPS |
| TiGo-ETS | **313.8**[a] | **246.3** | **445.2** | **350.2** |
| LightGBM | 507.9 | 443.4 | 523.2 | 473.9 |
| QRF | 387.3 | 317.1 | 469.2 | 394.0 |
| Difference[b] | -73.5 | -70.8 | -24.1 | -43.8 |
| Std. Error | 21.6 | 22.1 | 19.3 | 20.4 |
| Significance[c] | *** | *** | . | * |

[a] Values in bold indicate the lowest errors.
[b] Each difference is the TiGo-ETS loss minus the QRF loss.
[c] ***, * and · indicate significance based on t-tests at the 0.1%, 5% and 10% levels, respectively.

### A.2.5. In-Sample Fitting Accuracy and Skewness

Figure A.2 shows histograms displaying skewness values of the Bass, gamma/shifted-Gompertz, trapezoid, and tilted-Gompertz models fit to the 161 social networking websites in Study 1; similar results appear in the other two empirical studies. We measure skewness using Arnold and Groeneveld's method (1995), as detailed in Appendix A.1.4. This skewness measure ranges from -1 to 1, where negative values imply left skew and positive values imply right skew. The histograms clearly show that in comparison to the Bass and GSG models, the tilted-Gompertz model covers a wider range of skewness values. Although the trapezoid model also exhibits a broad range of skewness values, its performance may be hindered by its piecewise linear structure. Combing the ability of capturing skewness and its continuous form, the TiGo model could perform the best in modeling diffusion curves. As shown in Table A.8, it has the lowest mean absolute error (MAE) and root mean squared error (RMSE) for in-sample fit in all three studies.

**Figure A.2 Histograms of Skewness Values from Four Diffusion Models Fit to 161 Social Networking Websites Life Cycles.**




(a) Bass Model

(b) GSG Model

(c) Trapezoid Model

(d) TiGo Model

**Table A.8 Evaluation of In-sample Fitting Performance Using MAE and RMSE**

| | Study 1 Search Interests | | Study 2 Intel Processor Sales | | Supplemental Study Dell Computer Sales | |
|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| Bass | 12.37 | 18.6 | 14083.6 | 20345.7 | 206.52 | 308.47 |
| GSG | 8.79 | 13.65 | 14120.07 | 20849.7 | 195.98 | 296.63 |
| Trap | 9.95 | 14.89 | 13908.52 | 20316.46 | 199.35 | 300.07 |
| TiGo | **7.72**[a] | **12.26** | **12937.14** | **18981.04** | **190.95** | **293.46** |
| Difference[b] | -1.07 | -1.39 | -971.38 | -1335.42 | -5.04 | -3.17 |
| Std. Error | 0.15 | 0.23 | 290.88 | 381.56 | 1.29 | 1.81 |
| Significance[c] | *** | *** | *** | *** | *** | * |

[a] Values in bold indicate the lowest errors.
[b] Each difference is the TiGo-ETS loss minus the second best model in the column.
[c] *** and *indicates significance based on t-tests at the 0.1% and 5% levels, respectively.

## References


Acimovic J, Erize F, Hu K, Thomas DJ, Van Mieghem JA. 2019. Product life cycle data set: Raw and cleaned data of weekly orders for personal computers. *Manufacturing & Service Operations Management* **21**(1) 171–176.

Arnold BC, Groeneveld RA. 1995. Measuring skewness with respect to the mode. *The American Statistician* **49**(1) 34—38.

Chen J, Rubin H. 1986. Bounds for the difference between median and mean of Gamma and Poisson distributions. *Statistics & Probability Letters* **4**(6) 281–283.

Easingwood CJ, Muller E, Mahajan V. 1983. A nonuniform influence innovation diffusion model of new product acceptance. *Marketing Science* **2**(3) 273–295.

Hu K, Acimovic J, Erize F, Thomas DJ, Van Mieghem JA. 2018. Forecasting product life cycle curves: Practical approach and empirical analysis. *Manufacturing & Service Operations Management* **21**(1) 66–85.

Mahajan V, Bass FM, Muller E. 1990. New product diffusion models in marketing: A review and directions for research. *Journal of Marketing* **54**(1) 1–26.

Pearl R, Reed LJ. 1925. Skew-growth curves. *Proceedings of the National Academy of Sciences* **11**(1) 16–22.

Wikipedia. 2021a. List of defunct social networking services. `https://en.wikipedia.org/wiki/List_of_defunct_social_networking_services`. Accessed: June 1, 2021.

Wikipedia. 2021b. List of social networking services. `https://en.wikipedia.org/wiki/List_of_social_networking_services`. Accessed: June 1, 2021.

Winsor CP. 1932. The Gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences* **18**(1) 1–8.