HARVARD | BUSINESS | SCHOOL

Quantile Forecasts of Product Life Cycles Using Exponential Smoothing

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

Working Paper 19-038

Quantile Forecasts of Product Life Cycles Using Exponential Smoothing

Xiaojia Guo
University College London

Kenneth C. Lichtendahl Jr.
University of Virginia

Yael Grushka-Cockayne
Harvard Business School

Working Paper 19-038

Quantile Forecasts of Product Life Cycles Using Exponential Smoothing

Xiaojia Guo
UCL School of Management, University College London, London, UK, x.guo.11@ucl.ac.uk
Kenneth C. Lichtendahl Jr.
Darden School of Business, University of Virginia, Charlottesville, VA 22903, lichtendahlc@darden.virginia.edu
Yael Grushka-Cockayne
Darden School of Business, University of Virginia, Charlottesville, VA 22903, grushkay@darden.virginia.edu

We introduce an exponential smoothing model that a manager can use to forecast the demand of a new product or service. The model has five features that make it suitable for accurately forecasting product life cycles at scale. First, the trend in our model follows the density of a new distribution called the tilted-Gompertz distribution. This model can capture the wide range of skewed diffusions commonly found in practice—diffusions of innovations described as having "extra-Bass" skew. Second, its parameters can be updated via exponential smoothing; therefore, the model can react to local changes in the environment. This model is the first exponential smoothing model to incorporate a life-cycle trend. Third, the model relies on multiplicative errors, instead of the additive errors primarily used in existing models. Multiplicative errors ensure that all quantile forecasts are strictly positive. Fourth, the model includes prior distributions on its parameters. These prior distributions become regularization terms in the model and allow the manager to make accurate forecasts from the beginning of a life cycle, which is notoriously difficult. The model's skewed shape, time-varying, regularized parameters, and multiplicative errors can make its quantile forecasts more accurate than leading diffusion models, such as the Bass, gamma/shifted-Gompertz, and trapezoid models. Fifth, the model's estimation procedure is based on an efficient optimization routine, which can be used to forecast product life cycles at scale. In two empirical studies, one of search interest in social networks and the other of new computer sales, we demonstrate that our model outperforms leading diffusion models in out-of-sample forecasting. Our model's point and other quantile forecasts are more accurate. Accurate quantile forecasts at different horizons are critical to many operational decisions, such as capacity and inventory management.

Key words: new product development; demand forecasting; product adoption; innovation diffusion.
Date: June 22, 2018

1. Introduction

Many important business decisions rely on a manager's forecast of a product or service's life cycle. Generating such forecasts is especially hard when the product or service is an innovation. For instance, the decision about when an innovative service, such as Twitter, should add server capacity to handle network demand will depend, in part, on predictions of growth in new users.

Other decisions inside a firm may occur more frequently and on a large-scale, such as managing the inventory levels of thousands or millions of new products, such as next-generation computer hardware products. To support these decisions, managers need a life-cycle model that produces accurate forecasts (Hu et al. 2017, Baardman et al. 2017). Many such decisions rely on the quantile of a demand forecast's distribution (Petruzzi and Dada 1999). In these cases, managers need accurate quantile forecasts.

Since their introduction in the 1950s and 1960s, exponential smoothing models have become one of the most widely used forecasting techniques in business (Holt 1957/2004, Brown 1959, Winters 1960, Brown 1963, Gardner and McKenzie 1985). Their popularity stems largely from how accurate they are in practice (Hyndman 2015). Exponential smoothing models are also efficient to estimate using open-source statistical computing software. In addition, they benefit from the simple and intuitive way in which their parameters are updated and their forecasts are generated. In this paper, we introduce an exponential smoothing model to forecast product (or service) life cycles.

In exponential smoothing, a time series is decomposed into three components: error, trend, and seasonality. For non-seasonal time series, the main component is the trend. In the well-known class of exponential smoothing models (Hyndman et al. 2008), there are five possible trend types: (a) no trend, (b) additive trend, (c) additive damped trend, (d) multiplicative trend, and (e) multiplicative damped trend. See Figure 1(a)-(e) for a schematic of these five trend types.

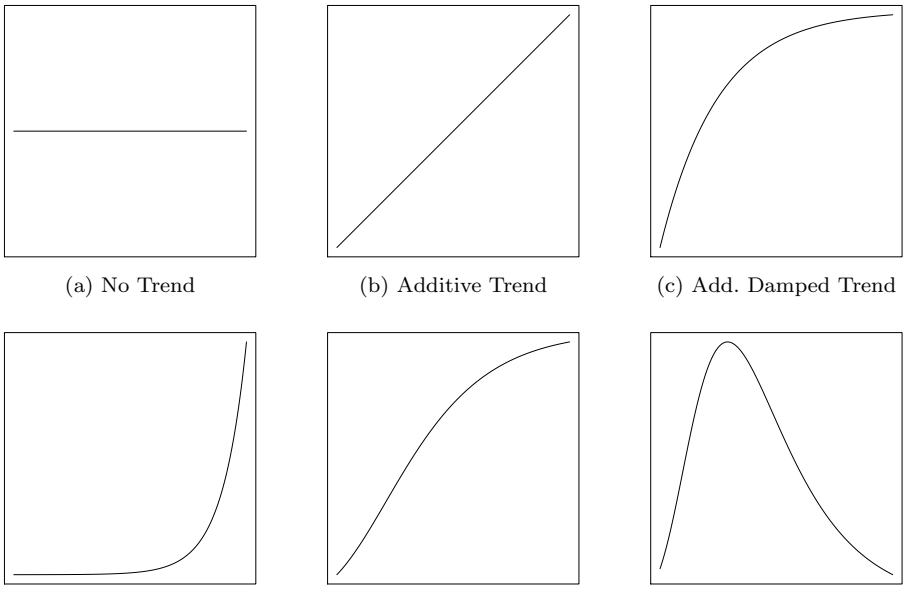


Figure 1 Five Existing Trends and New Life-Cycle Trend in Exponential Smoothing Models.

The five trends in Figure 1(a)-(e) are either increasing, decreasing, or constant. We introduce an important new trend to this list—the life-cycle trend in Figure 1(f). This is a key contribution because exponential smoothing techniques have never been applied to life-cycle models before, and their advantages described above are especially useful for life-cycle modelling in practice.

Our life-cycle trend is most closely related to the two multiplicative trends in this list. Pegels' (1969) multiplicative trend applies to a series that is growing exponentially. Taylor's (2003) multiplicative damped trend applies to a series where the exponential growth dampens and eventually plateaus. Our life-cycle trend captures the notion that a time series initially grows exponentially, slows in its growth until it peaks, and then begins to decline. If it grows up slower than it declines, the trend is left-skewed, and if it grows up faster than it declines, the trend is right-skewed.

Our model's trend follows the probability density function of a new distribution, the tilted-Gompertz distribution. This distribution results from reflecting, truncating, and tilting a Gumbel distribution. When the model is applied to the adoption of a new product, the tilted-Gompertz distribution describes the uncertain time it takes a consumer to adopt the new product. Its density can fit a wide range of left-skewed, nearly symmetric, and right-skewed shapes.

From the marketing literature, we know that the ability to fit skewed diffusions is important because many new-product diffusions exhibit long right or left tails (Dixon 1980, Van den Bulte and Lilien 1997, Van den Bulte and Joshi 2007). Bemmaor and Lee (2002) say these diffusions show "extra-Bass" skew because the popular Bass (1969) model is symmetric around its mode. As extensions, several asymmetric models have been proposed to describe skewed diffusions. Notable among them are the nonuniform influence model (Easingwood et al. 1983), the Gompertz model (Mahajan et al. 1986), the gamma/shifted-Gompertz model (Bemmaor 1994), and the asymmetric influence model (Van den Bulte and Joshi 2007).

From these studies, two models emerge as natural benchmarks: the Bass and gamma/shifted-Gompertz models. Although the Bass model is symmetric around its mode, we consider it here because it is intuitively appealing, tractable, and often accurate in practice. Of the asymmetric models proposed, the gamma/shifted-Gompertz model stands out because it is a tractable extension of the Bass model.

From the operations literature, another asymmetric benchmark emerges. The trapezoid model, recently proposed and applied by Hu et al. (2017), is quite appealing. It is both easy to estimate and easy to explain. Its density follows the shape of a triangle, but with a flat peak corresponding to the life cycle's maturity stage. Hu et al. (2017) note that the trapezoid model applies well to

left-skewed life cycles of new products like the high-technology products described in Goldman (1982).

We compare these three benchmark models to the tilted-Gompertz diffusion model—a diffusion model derived from our exponential smoothing model with time-invariant parameters. Specifically, we compare the ranges of skewness that the Bass, gamma/shifted-Gompertz, trapezoid, tilted-Gompertz models can achieve. One finding here is that of these four models, only the tilted-Gompertz and trapezoid models can fit the full range of skewed diffusions. In examining the skewness of these models, we identify a bias in the leading measure of skewness in the diffusion literature (Easingwood et al. 1983, Mahajan et al. 1990). Consequently, we introduce a new measure of skewness. It is designed specifically to address this bias and to measure "extra-Bass" skew more precisely.

In two empirical studies, we compare the out-of-sample forecasts of the Bass, gamma/shifted-Gompertz, trapezoid, and tilted-Gompertz models. We estimate each of these models using a regularization approach that allows us to incorporate prior information into the model. With prior information as part of the model, we can forecast a life cycle from its very beginning, which is a notoriously difficult problem (Venkatesan et al. 2004). Importantly, our approach for incorporating prior information is both computationally more efficient and more accurate than traditional Bayesian methods such as the extended Kalman filter with continuous state and discrete observations (EKF-CD) introduced by Xie et al. (1997). Consequently, we think our regularization approach, which is new to the literature on product life cycles, may be attractive for use at scale.

Many firms, such as Google and Facebook, forecast demand at long-term horizons, at scale, and on a frequent rolling basis. Moreover, many time series in these settings go through life cycles. Tassone and Rohani (2017) describe Google's forecasting challenges as follows, "Our typical use case was to produce a time series forecast at the daily level for a 12-24 month forecast horizon based on a daily history two or more years long... We wanted to forecast a variety of quantities: overall search query volume and particular types of queries; revenue; views and minutes of video watched on Google-owned YouTube... our forecasting task was easily on the order of tens of thousands of forecasts. And then we wanted to forecast these quantities every week, or in some cases more often." The data scientists at Google were also interested in prediction intervals as a way to estimate tail events.

The data in our empirical studies are time series of Google Trends search interest in 122 social networks and sales of new Dell computers. In these studies, we find that the tilted-Gompertz model, with either time-invariant or time-varying parameters, performs favorably in out-of-sample

forecasting when compared to the benchmark models. The tilted-Gompertz model has, in most instances, more accurate point forecasts and prediction intervals. The tilted-Gompertz model with time-invariant parameters performs well because it can fit a wider range of skewed diffusions. With time-varying parameters, the model performs even better because it can both fit a wider range of skewed diffusions and react to local changes in the environment.

2. Exponential Smoothing Model with a Life-Cycle Trend

To model a time series $y_1, \ldots, y_n$ using exponential smoothing, the level states $\ell_1, \ldots, \ell_n$ and the growth states $b_1, \ldots, b_n$ are used to describe the time series' trend in a recursive fashion. For example, Taylor's (2003) exponential smoothing model with additive errors and a multiplicative damped trend has the following formulation, called a state-space formulation:

$$\begin{aligned} \text{Measurement Equation:} \quad & y_t = \ell_{t-1} b_{t-1}^{\phi} + \varepsilon_t \\ \text{Transition Equation for the Level:} \quad & \ell_t = \ell_{t-1} b_{t-1}^{\phi} + \alpha \varepsilon_t \\ \text{Transition Equation for the Growth:} \quad & b_t = b_{t-1}^{\phi} + \beta \varepsilon_t / \ell_{t-1}, \end{aligned} \tag{1}$$

where $0 < \ell_0 < \infty$ and $0 < b_0 < \infty$ are the initial level and growth, respectively, $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are the smoothing parameters, and $0 < \phi < 1$ is a damping parameter. The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed according to a normal distribution with mean zero and variance σ^2. With $\phi = 1$ in (1), the model becomes Pegel's (1969) multiplicative trend model.

The exponential smoothing model we introduce below is an important extension of Taylor's (2003) model. To construct our model, we multiply each occurrence of b_{t-1}^{ϕ} in Taylor's model by a parameter τ. We also replace Taylor's additive error terms with multiplicative errors. With multiplicative errors, the transitions for the growth states are guaranteed to be positive and thus well-defined when taking b_{t-1} to the power ϕ. Also, the endpoints of our model's prediction intervals are guaranteed to be positive, unlike the intervals from Taylor's model. In most contexts, a manager will expect to have prediction intervals with positive endpoints for the demand of a new product or service.

2.1. State-Space Formulation

We define an exponential smoothing model with a life-cycle trend using the state-space formulation:

$$\begin{aligned} \text{Measurement Equation:} \quad & y_t = \ell_{t-1} b_{t-1}^{\phi} \tau (1 + \varepsilon_t) \\ \text{Transition Equation for the Level:} \quad & \ell_t = \ell_{t-1} b_{t-1}^{\phi} \tau (1 + \varepsilon_t)^{\alpha} \\ \text{Transition Equation for the Growth:} \quad & b_t = b_{t-1}^{\phi} \tau (1 + \varepsilon_t)^{\beta}, \end{aligned} \tag{2}$$

where $0 < \ell_0 < \infty$ and $0 < b_0 < \infty$ are the initial level and growth states, respectively, $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are smoothing parameters, $0 < \phi < \infty$ (and $\phi \neq 1$) is a damping parameter, and $0 < \tau < 1$

is a turn-down parameter. The random variables $1+\varepsilon_1, \ldots, 1+\varepsilon_n$ are independent and identically distributed according to a lognormal distribution where each $\varepsilon_t^* = \log(1+\varepsilon_t)$ is distributed normally with mean zero and variance σ^2.

Note that the multiplicative errors in our model are modified versions of those in typical exponential smoothing models (Hyndman et al. 2008, Chapter 15.2.2). In our transition equations, we use $(1+\varepsilon_t)^\alpha$ and $(1+\varepsilon_t)^\beta$ rather than $1+\alpha\varepsilon_t$ and $1+\beta\varepsilon_t$, respectively. The reason we use modified multiplicative errors in our model is so that we can express the median and mean of any h-step-ahead prediction distribution in closed form. For any exponential smoothing model with a multiplicative trend and typical multiplicative errors, it is intractable to express the median and mean of an h-steps-ahead prediction distribution (Hyndman et al. 2008).

To generate forecasts from our model, we apply a logarithmic transformation to each side of each equation in (2). This transformed model becomes a convenient exponential smoothing model with an additive trend:

$$\begin{aligned} \text{Measurement Equation:} \quad & y_t^* = \ell_{t-1}^* + \phi b_{t-1}^* + \log(\tau) + \varepsilon_t^* \\ \text{Transition Equation for the Level:} \quad & \ell_t^* = \ell_{t-1}^* + \phi b_{t-1}^* + \log(\tau) + \alpha\varepsilon_t^* \\ \text{Transition Equation for the Growth:} \quad & b_t^* = \phi b_{t-1}^* + \log(\tau) + \beta\varepsilon_t^*, \end{aligned} \tag{3}$$

where $y_t^* = \log(y_t)$, $\ell_t^* = \log(\ell_t)$, $b_t^* = \log(b_t)$, and $\varepsilon_t^* = \log(1+\varepsilon_t)$ is distributed normally with mean zero and variance σ^2.

2.2. Forecasting Method and Prediction Distribution

Next we show how to generate forecasts from our exponential smoothing model with a life-cycle trend. We denote an h-steps-ahead point forecast for y_{t+h}^* by $\hat{y}_{t+h}^*$. Our forecasting method below is comprised of three equations: a point forecasting equation and two equations for updating the transformed states. Given the transformed observations $y_1^*, \ldots, y_t^*$, we update beliefs about the transformed states recursively from periods 1 to t. Then, given the transformed states as of time t, we use the point forecasting equation to make forecasts h-steps-ahead.

For the transformed model in (3), the forecasting method is given by the following:

$$\begin{aligned} \text{Point Forecasting Equation:} \quad & \hat{y}_{t+h}^* = \ell_t^* + \sum_{i=1}^{h} \phi^i b_t^* + \sum_{i=1}^{h} (1+\phi+\cdots+\phi^{i-1}) \log(\tau) \\ \text{Level Updating Equation:} \quad & \ell_t^* = \alpha y_t^* + (1-\alpha)(\ell_{t-1}^* + \phi b_{t-1}^* + \log(\tau)) \\ \text{Growth Updating Equation:} \quad & b_t^* = \beta^*(\ell_t^* - \ell_{t-1}^*) + (1-\beta^*)(\phi b_{t-1}^* + \log(\tau)), \end{aligned} \tag{4}$$

where $\beta^* = \beta/\alpha$ if $\alpha > 0$; otherwise, $\beta^* = 0$. To interpret β^* as a weight, we restrict $\beta \leq \alpha$ unless $\alpha = 0$. See the Supplement for a derivation of this forecasting method and the prediction distribution below.

With the forecasting method above, we can see the intuitive way in which the transformed model updates beliefs about its states. The current level is a weighted average of the most recent observation of the time series and the previous period's one-step-ahead forecast (which is based, in part, on the previous period's level). The current growth is a weighted average of the most recent growth estimate (based on the ratio of the two most recent levels) and a function of the previous period's growth. Like other exponential smoothing models, if we increase either of the smoothing parameters α or β, beliefs about the current trend become more reactive to recent (or local) changes in the environment.

For the original time series, the h-steps-ahead prediction distribution, made at time t, is described by the equation

$$y_{t+h} = \ell_t b_t^{\phi+\phi^2+\cdots+\phi^h} \tau^{\sum_{i=1}^{h}(1+\phi+\cdots+\phi^{i-1})} e^{\eta_h}$$

where $\eta_h = \sum_{i=1}^{h-1}(\alpha+\beta(\phi+\cdots+\phi^i))\varepsilon^*_{t+h-i} + \varepsilon^*_{t+h}$ is normally distributed with mean zero and variance $Var[\eta_h] = \big(\sum_{i=1}^{h-1}(\alpha+\beta(\phi+\cdots+\phi^i))^2+1\big)\sigma^2$. Consequently, y_{t+h} is lognormally distributed with the following statistics:

$$\begin{aligned}
\text{Median:} \quad & \hat{y}_{t+h} = \ell_t b_t^{\phi+\phi^2+\cdots+\phi^h} \tau^{\sum_{i=1}^{h}(1+\phi+\cdots+\phi^{i-1})} \\
\text{Mean:} \quad & \hat{y}_{t+h} e^{Var[\eta_h]/2} \\
\text{Variance:} \quad & [e^{Var[\eta_h]} - 1]\hat{y}^2_{t+h} e^{Var[\eta_h]} \\
p\text{-Quantile:} \quad & Q_p = \hat{y}_{t+h} e^{Var[\eta_h]^{1/2}\Phi^{-1}(p)},
\end{aligned}$$

where Φ is the cdf of a standard normal random variable. The $0 < u < 1$ central prediction interval is given by $[Q_{(1-u)/2}, Q_{1-(1-u)/2}]$.

Throughout the rest of the paper, we take the median of y_{t+h} to be the point forecast of y_{t+h}:

$$\text{Point Forecasting Equation (Original Model):} \quad \hat{y}_{t+h} = \ell_t b_t^{\phi+\phi^2+\cdots+\phi^h} \tau^{\sum_{i=1}^{h}(1+\phi+\cdots+\phi^{i-1})}. \tag{5}$$

Note that $\hat{y}_{t+h} = e^{\hat{y}^*_{t+h}}$, which provides a connection between the transformed model's point forecast and the original model's point forecast. There are two reasons why we choose the median to be the model's point forecast. First, the median, in this case, is the zero-error point forecast. That is, it follows from the state-space model in (2) with the errors $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$ all set to zero. Zeroing out future errors is a common way to produce a point forecast from an exponential smoothing model with a multiplicative trend (Hyndman et al. 2008). Second, the median, in any case, is the optimal point forecast under the mean absolute error (Gneiting 2011).

Next we provide the continuous-time version of the discrete-time point forecasting equation in (5) and show that it is proportional to the density of the tilted-Gompertz ($TiGo$) distribution. We denote the resulting point forecasting function by $\hat{y}_t(h)$, a function that can be evaluated for any real number $h \geq 0$. This function follows from re-expressing the summations in the point forecasting equation in (5) so that we can evaluate $\hat{y}_{t+h}$ for any real number $h > 0$.

From the point forecasting equation for $\hat{y}_{t+h}$ in (5), we have

$$\text{Point Forecasting Function:} \quad \hat{y}_t(h) = m_t c_t e^{-\lambda\delta h} e^{-\rho_t e^{-\lambda h}}, \tag{6}$$

where $\lambda = -\log(\phi)$, $\delta = \frac{\log(\tau)}{\log(\phi)(1-\phi)}$, $\rho_t = \frac{\phi}{1-\phi}\big(\log(b_t) - \frac{\log(\tau)}{1-\phi}\big)$, $m_t = \ell_t e^{\rho_t}/c_t$, and $c_t = \lambda\rho_t^{\delta}/(\gamma(\delta,\rho_t) - I_{(-\infty,0)}(\lambda)\Gamma(\delta))$. In the constant c_t, $\gamma(\delta,s) = \int_0^s z^{\delta-1}e^{-z}\,dz$ is the lower incomplete gamma function, $\Gamma(\delta) = \int_0^\infty z^{\delta-1}e^{-z}\,dz$ is the gamma function, and the indicator $I_A(z) = 1$ if $z \in A$ and equals zero otherwise. See the Supplement for the details of this derivation.

The point forecasting function in (6) represents our first derivation of the tilted-Gompertz distribution, which has a probability density function equal to $c_t e^{-\lambda\delta h} e^{-\rho_t e^{-\lambda h}}$. In the next section, we introduce the tilted-Gompertz diffusion model.

3. Tilted-Gompertz Diffusion Model

In this section, we introduce the tilted-Gompertz diffusion model. We derive the model from two different perspectives: from a macro-level view and a micro-level view. At the core of either view is a function F of time. In the macro-level view, $F(t)$ is the expected proportion of eventual adopters that have adopted by time t out of a population of m eventual adopters. While in the micro-level view, $F(t)$ is a cumulative distribution function (cdf) that describes the probability that an individual has adopted by time t. In either view, $N(t)$ is the number of adopters that have adopted by time t, and $E[N(t)] = mF(t)$ is the expected number of adopters that have adopted by time t. In the macro-level view, this expectation is assumed to hold, but in the micro-level view, this expectation is derived from basic assumptions.

3.1. Macro-Level View of the Diffusion Process

In the macro-level view, a diffusion process is typically described by a differential equation that governs the population's rate of adoption. The rate of adoption is the rate of change over time in the expected number of adopters that have adopted by time t. The rate of adoption is $mf(t)$, where $f(t)$ is the first derivative of $F(t)$. Then the population's acceleration of adoption is $mf'(t)$, where $f'(t)$ is the second derivative of $F(t)$. Like the rate of adoption, this second-order rate of change is useful for interpreting the dynamics of a diffusion model.

We introduce the tilted-Gompertz diffusion model by way of its acceleration of adoption:

$$mf'(t) = m(\rho e^{-\lambda t} - \delta)\lambda f(t), \tag{7}$$

for $0 \le t < \infty$ where $-\infty < \lambda < \infty$, $\lambda \neq 0$, and $0 < \rho < \infty$. The general solution to this differential equation results in the model's rate of adoption. Under the condition that $\int_0^\infty f(t)\,dt = 1$, the solution is the model's rate of adoption, which is proportional to the pdf of the tilted-Gompertz ($TiGo$) distribution:

$$f_{TiGo}(t) = ce^{-\lambda\delta t}e^{-\rho e^{-\lambda t}},$$

where $c = \lambda\rho^\delta/(\gamma(\delta,\rho) - I_{(-\infty,0)}(\lambda)\Gamma(\delta))$. The model's rate of adoption $mf_{TiGo}(t)$ is equivalent to the expression in (6) with its t set to zero and t used in place of its h. As we will see in the next section, the sign of λ dictates the process's type of skewness. The process is left-skewed when λ is negative, and the process is right-skewed when λ is positive.

Next we compare the tilted-Gompertz model's acceleration of adoption to that of the Bass diffusion model. The Bass model's acceleration of adoption is

$$mf'(t) = m(q(1 - 2F(t)) - p)f(t),$$

where q is the coefficient of imitation and p is the coefficient of innovation. Using these two models' acceleration equations, we can compare the their modes, or the times at which peak sales occur. Each model has a mode at the time where the acceleration is equal to zero. For the tilted-Gompertz model, the mode is $-\log(\delta/\rho)/\lambda$, while for the Bass model, the mode is $-\log(p/q)/(p+q)$, which look quite similar. These similarities suggest interpretations of the tilted-Gompertz's parameters along the lines of imitation and innovation. See Table 1 for additional properties of these models.

One important property is the Bass model's rate of adoption $mf(t) = m(p + qF(t))(1 - F(t))$. Given the initial condition $F(0) = 0$, Bass (1969) provides the solution to this rate-of-adoption equation. The solution is proportional to the cdf of the Bass (Ba) distribution

$$F_{Ba}(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}.$$

Later we will use this cdf and its inflection point to describe the Bass model's skewness. In the next section, we show how to derive the cdf of the tilted-Gompertz distribution. Once we have its cdf in hand, we will use it to describe the skewness of a tilted-Gompertz diffusion.

To fit a diffusion model to data on per-period adoptions $N(t_i) - N(t_{i-1})$, many researchers estimate the model with additive errors:

$$\begin{aligned} N(t_i) - N(t_{i-1}) &= E[N(t_i) - N(t_{i-1})] + \varepsilon_i \\ &= mF(t_i) - mF(t_{i-1}) + \varepsilon_i, \end{aligned} \tag{8}$$

where each time interval $[t_{i-1}, t_i]$ is one period and the errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed according to a normal distribution with mean zero and variance σ^2. The model in (8) is fit by choosing the parameters m, p, and q in $mF_{Ba}(t_i) - mF_{Ba}(t_{i-1})$ and σ^2 so as to maximize the model's normal likelihood function. This type of maximum likelihood estimation is called non-linear least-squares in the literature (Srinivasan and Mason 1986, Van den Bulte and Lilien 1997, Bemmaor and Lee 2002).

3.2. Micro-Level View of the Diffusion Process

At the micro-level, each individual in the population adopts a new product at an uncertain time. The individuals' adoption times are assumed to be independent and identically distributed according to a cdf F. Hence, the number of adopters that have adopted by time t follows a binomial distribution with probability $F(t)$ and trials m, and $E[N(t)] = mF(t)$ (Schmittlein and Mahajan 1982, Meade and Islam 2006).

The tilted-Gompertz distribution can be used to describe the distribution of an individual's adoption time. Next we derive this distribution from other well-known distributions in the diffusion literature. In doing so, we notice some important connections with these distributions. To begin the construction, we reflect the Gumbel distribution about zero (i.e., take its mirror image) and truncate its reflection below zero. We also truncate the Gumbel distribution itself below zero. These truncated distributions become the left- and right-skewed Gompertz distributions, respectively. In our final step, we exponentially tilt the Gompertz distribution. Exponential tilting is a technique used in statistics to approximate the density in a maximum likelihood estimation procedure (Efron 1981, Goutis and Casella 1999).

The result of the steps described above is the tilted-Gompertz distribution, which has the cdf

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

for $-\infty < \lambda < \infty$, $\lambda \neq 0$, and $0 < \delta, \rho < \infty$. The distribution has three parameters: λ, δ, and ρ. We call these parameters the scale, tilting, and shape parameters, respectively. The tilted-Gompertz diffusion model can now be described in terms of this cdf. (For more details on the derivation of this function, see the Supplement.)

When the tilting parameter δ equals one, the tilted-Gompertz distribution specializes to the Gompertz distribution. The left-skewed (right-skewed) Gompertz distribution is the tilted-Gompertz distribution with $\lambda < 0$ ($\lambda > 0$) and $\delta = 1$. The left-skewed Gompertz distribution was introduced by Gompertz (1825) and is well known as a life distribution (Johnson et al. 1995, Marshall and Olkin 2007). Johnson et al. (1995, p. 25) note that the truncation of the reflected-Gumbel distribution results in the left-skewed Gompertz distribution.

4. Measuring a Diffusion's Skewness Around its Mode

Next we introduce a new measure of skewness and show how the tilting parameter δ and the shape parameter ρ work in conjunction to control the degree of skewness of the tilted-Gompertz distribution. For modeling diffusions with long takeoffs or with different rates of growth before the peak and decline afterward, such as the consumer durable goods studied in Golder and Tellis (1997), the shapes that the tilted-Gompertz distribution can achieve may be useful.

To apply our new measure of skewness, we assume that F is a cdf for an adoption time T, taking values on the non-negative interval $[0, t_{max})$, and that F has a unique mode at time $t^* > 0$. The mode here is the time $t^* > 0$ at which the pdf f reaches its single peak. Specifically, $f(t) > 0$ for $t \in (0, t_{max})$, either $f(0) > 0$ or $f(0) = 0$, $f'(t) > 0$ for $t \in [0, t^*)$, $f'(t) = 0$ for $t = t^*$, and $f'(t) < 0$ for $t \in (t^*, t_{max})$. Let $t^{**} > t^*$ be the adoption time (possibly infinite) such that $f(t^{**}) = f(0)$. For distributions that satisfy these assumptions, we define our measure of local skewness with respect to the mode t^* as

$$Skew(F) = 1 - 2\frac{F(t^*)}{F(t^{**})}.$$

A value of $Skew(F)$ in the interval $(-1, 0)$ indicates left skewness, and a value in $(0, 1)$ indicates right skewness.

For the tilted-Gompertz distribution, the mode is at $t^* = -\log(\delta/\rho)/\lambda$, and $t^{**} = -\log(-(\delta/\rho)W(-(\delta/\rho)e^{-\delta/\rho})/\lambda$ where W is the Lambert W function. For $\lambda > 0$, we use the principal branch W_0 of the Lambert W function, and for $\lambda < 0$, we use the non-principal branch W_{-1}. (See Corless et al. 1996 for details on the Lambert W function.) The local skewness of the tilted-Gompertz distribution is

$$Skew(F_{TiGo}) = 1 - 2\frac{\gamma(\delta, \rho) - \gamma(\delta, \delta)}{\gamma(\delta, \rho) - \gamma(\delta, -\delta W(-(\delta/\rho)e^{-\delta/\rho}))}, \tag{9}$$

for $\rho < \delta$ if $\lambda < 0$ (left-skewed) and $\rho > \delta$ if $\lambda > 0$ (right-skewed). Note that the tilted-Gompertz distribution's mode is at zero when $\lambda(\rho - \delta) < 0$; and therefore, its local skewness is undefined in this case.

The cdf evaluated at its inflection point t^* 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. Easingwood et al. (1983) and Mahajan et al. (1990) use $F(t^*)$ to describe asymmetric diffusions. Arnold and Groeneveld (1995) use the related $1-2F(t^*)$ to measure skewness with respect to the mode of any distribution with a single-peaked density.

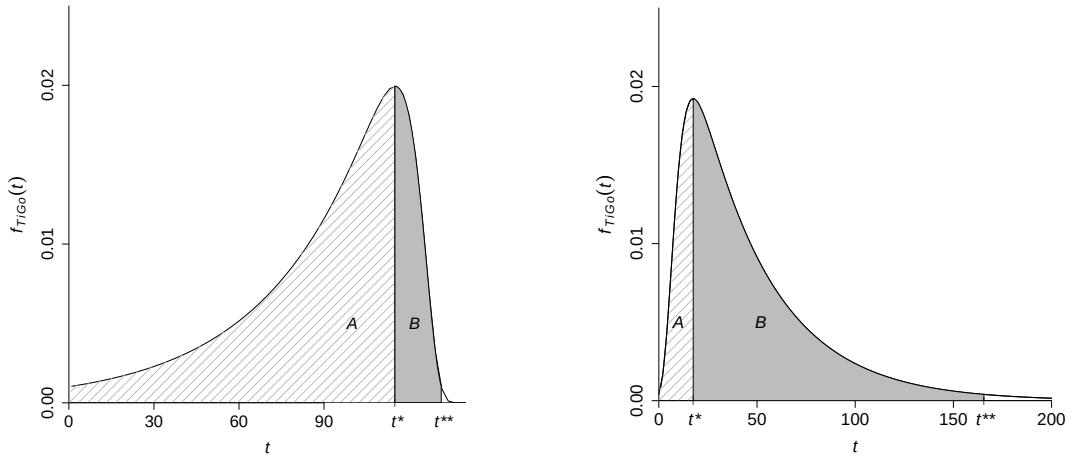
Our new measure is localized version of Arnold and Groeneveld's measure $1-2F(t^*)$, localized to the interval $[0,t^{**}]$ when t^{**} is finite. The main motivation for localizing Arnold and Groeneveld's measure is to measure a diffusion's "extra-Bass" skew, starting from a baseline of zero. The Bass distribution has zero local skewness with respect to its mode: $\mathit{Skew}(F_{Ba})=0$ for $p<q$. This starting point makes sense because the Bass distribution's pdf is formally symmetric around its mode on the interval $[0,2t^*]$. For the Bass distribution, $t^{**}=2t^*$.

And yet, Arnold and Groeneveld's measure of the Bass distribution's skewness is always positive. This indication of right skewness is an artifact of truncation. The Bass distribution is a logistic distribution truncated below zero. This truncation makes sense because an adoption time is naturally non-negative. A manager, however, who hears a summary statistic indicating that a new product's diffusion will be right-skewed—merely because a left tail was truncated—may jump to the conclusion that sales will grow more quickly up to its peak than it will decline thereafter.

To understand how our measure works in more detail, we express it as the difference between two conditional probabilities: $\mathit{Skew}(F)=P(t^*<T\le t^{**}|0<T\le t^{**})-P(0<T\le t^*|0<T\le t^{**})$. Figure 2 illustrates the difference between these probabilities using two areas, labeled A and B. The probability $P(0<T\le t^*|0<T\le t^{**})$ is equal to $A/(A+B)$, and the probability $P(t^*<T\le t^{**}|0<T\le t^{**})$ is equal to $B/(A+B)$. With $A>B$, we have left skewness, as depicted in Figure 2(a). With $A<B$, we have right skewness, as depicted in Figure 2(b).

Our new measure compares the conditional probabilities of adoption in two time intervals: a left interval from zero to the mode and a right interval from the mode to t^{**}. Specifically, $\mathit{Skew}(F)=B/(A+B)-A/(A+B)=1-2A/(A+B)$. In contrast, Arnold and Groeneveld's measure $1-2F(t^*)=1-2A$ compares the conditional probabilities in the same left interval and a right interval from the mode to infinity. Because their right interval is longer, Arnold and Groeneveld's measure is biased toward indicating right skewness. In fact, $\mathit{Skew}(F)<1-2F(t^*)$ always holds.

In Figure 3(a), we see the contours of local skewness values that the tilted-Gompertz distribution can achieve over a large space of settings for δ and ρ. According to the expression in (9), only the parameter values of δ and ρ and the sign of λ affect its local skewness. Here, the local skewness



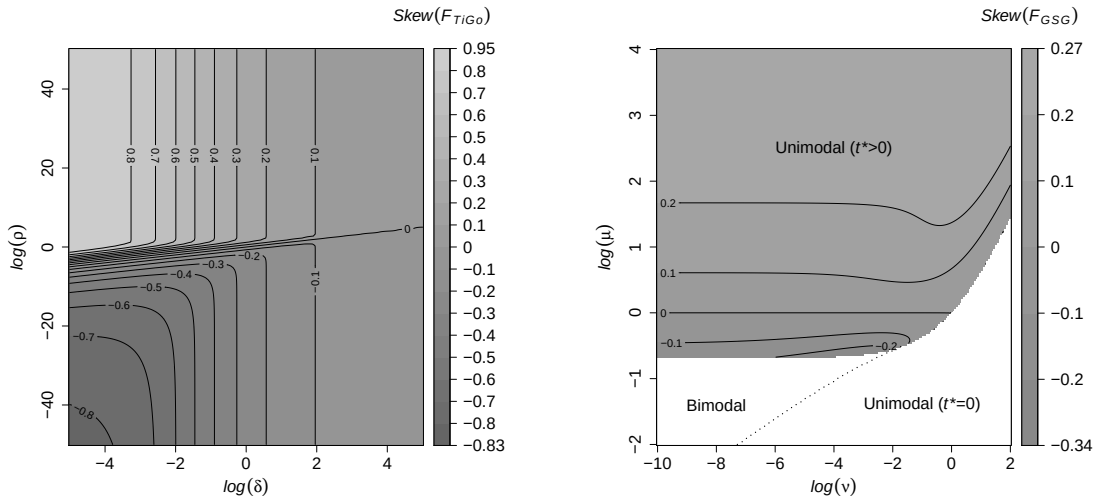
(a) Left skewness: $\lambda = -0.2$, $\delta = e^{-2}$, $\rho = e^{-25}$, and $Skew(F_{TiGo}) = -0.59$

(b) Right skewness: $\lambda = 0.2$, $\delta = e^{-2}$, $\rho = e^{1.5}$, and $Skew(F_{TiGo}) = 0.60$

Figure 2 Two Probability Density Functions from the Tilted-Gompertz Distribution

values range between -0.83 and 0.95. The effect of decreasing δ is to widen the range of skewness values that the tilted-Gompertz distribution can achieve, approaching $(-1, 1)$ as δ gets small.

In Figure 3(b), we see the contours of local skewness values that the gamma/shifted-Gompertz distribution can achieve. These values are always between -0.34 and 0.27. Thus, the gamma/shifted-Gompertz distribution is limited in the amount of "extra-Bass" skew it can capture. We also see in Figure 3(b) the regions for which the gamma/shifted-Gompertz distribution



(a) Tilted-Gompertz Distribution

(b) Gamma/Shifted-Gompertz Distribution

Figure 3 Contours of Local Skewness for the Tilted-Gompertz and Gamma/Shifted-Gompertz Distributions.

is unimodal with its mode at zero, unimodal with a positive mode, or bimodal with one mode at zero. Recall that our measure of local skewness does not apply to distributions with a mode at zero. The condition for a bimodal gamma/shifted-Gompertz model and other properties of this model are given in the Supplement's Table 1.

With regard to skewness, Bemmaor and Lee (2002, p. 211) report that "As shown in [their] Figure 1, the [gamma/shifted-Gompertz] model captures 'extra-Bass' skew in the data for given levels of p and q. When $[\mu]$ is less than 1, there is more right skew than Bass, and when $[\mu]$ is larger than 1, there is more left skew than Bass." But here we find the opposite for the gamma/shifted-Gompertz distribution with μ greater than 1. As μ increases, the gamma/shifted-Gompertz distribution becomes more right-skewed than the Bass distribution. Furthermore, it approaches a limit of around 0.27. Therefore, the gamma/shifted-Gompertz model may not be suitable for strongly right-skewed diffusions with long takeoffs.

A recent proposal that can capture the full range of skewed diffusions is the trapezoid model of Hu et al. (2017). Its rate of adoption is specified by five parameters:

$$mf_{Trap}(t) = \begin{cases} at+b & \text{for} \quad 0 \le t < \tau_1 \\ a\tau_1+b & \text{for} \quad \tau_1 \le t < \tau_2 \\ c(t-\tau_2)+a\tau_1+b & \text{for} \quad \tau_2 \le t \le t_{max} \\ 0 & \text{for} \quad t_{max} < t \end{cases},$$

where $t_{max} = \tau_2 - (a\tau_1 + b)/c$ and $0 < a, b, -c, \tau_1, \tau_2 - \tau_1 < \infty$. The shape of its rate of adoption follows a trapezoid—a triangle, but with a flat peak between τ_1 and τ_2. Its growth curve is given by

$$mF_{Trap}(t) = \begin{cases} at^2/2+bt & \text{for} \quad 0 \le t < \tau_1 \\ a\tau_1^2/2+b\tau_1+(a\tau_1+b)(t-\tau_1) & \text{for} \quad \tau_1 \le t < \tau_2 \\ -a\tau_1^2/2+(a\tau_1+b)\tau_2+c(t^2-\tau_2^2)/2+(a\tau_1+b-c\tau_2)(t-\tau_2) & \text{for} \quad \tau_2 \le t \le t_{max} \\ -a\tau_1^2/2+(a\tau_1+b)\tau_2+c(T^2-\tau_2^2)/2+(a\tau_1+b-c\tau_2)(T-\tau_2) & \text{for} \quad t_{max} < t \end{cases}.$$

Although this model does not have a single peak, we define its local skewness using our local skewness measure, but with t^* set to the trapezoid model's middle peak $(\tau_1 + \tau_2)/2$. For this model, $t^{**} = \tau_2 - a\tau_1/c$ so that $f_{Trap}(t^{**}) = f_{Trap}(0)$. Note that for any applicable growth curve mF, $Skew(F) = Skew(mF)$, since the eventual number of adopters m cancels out in our skewness measure.

5. Empirical Studies

In this section, we present the results of two empirical studies. In the first study, we forecast search interest in 122 social networks. In the second study, we forecast sales of 170 different new Dell computers. We compare the forecasting accuracy of the tilted-Gompertz model to several

benchmark models. We evaluate the accuracy of each model's out-of-sample forecasts using the mean absolute error for point forecasts and the pinball loss function for prediction intervals. Before we present these results, we provide some detail on how we estimated and evaluated the competing models.

The benchmarks we consider are the naïve, Bass, gamma/shifted-Gompertz, and trapezoid models. In these studies, we do not consider any machine learning algorithms as benchmarks. To our knowledge, algorithms, such as random forests, neural networks, and gradient boosted trees, do not work well when forecasting univariate time series. These algorithms are both less accurate and computationally efficient than traditional methods in forecasting univariate time series.

5.1. Model Estimation: Regularized Maximum Likelihood

Because life cycles are notoriously difficult to forecast from pre-peak data (Mahajan et al. 1990, Xie et al. 1997, Bemmaor and Lee 2002, Venkatesan et al. 2004), we incorporate prior information about the model's parameters into our estimation procedure. The typical problem with maximum likelihood estimation is that, before the peak, with only a few data points observed, the model is sensitive to each of the data points. If the last observed data point falls well above the previous few data points, the model's forecast will overshoot the eventual peak. In this situation, if the last observed data point is high due to noise, the model will be overfit to this noise (or error) and will not be well fit to the signal (or trend). Overfit models make poor out-of-sample forecasts. One way to avoid this overfitting problem is to use a Bayesian approach.

Xie et al. (1997) introduce a Bayesian approach to estimating life-cycle models with time-varying parameters. Their approach is an extended Kalman filter with continuous state and discrete observations (EKF-CD). Xie et al. (1997, p. 379) state one of their motivation for their approach as follows: "By incorporating prior estimates of unknown parameters and updating initial estimates as new data become available, time-varying estimation procedures often can provide better early forecasts." This approach, however, can be difficult to implement at scale. Similar to routines such as Markov Chain Monte Carlo (MCMC) simulation, the extended Kalman filter can be slow to run on even hundreds of time series. In our empirical study below, the extended Kalman filter runs slower than our regularized maximum likelihood approach by an order of magnitude. Moreover, the forecasts from the extended Kalman filter in this study are less accurate than those from our approach.

Another popular way to avoid model overfitting is to include a regularization term in a maximum likelihood approach. For instance, in a linear regression setting, the objective is to choose parameters that minimize the negative of the log-likelihood function (i.e., the sum of squared errors)

plus some regularization terms. In a ridge regression (Hoerl and Kennard 1970), the regularization terms are the scaled sum of the squares of the coefficients in a linear regression model. In the lasso (Tibshirani 1996), the regularization terms are the scaled sum of the absolute values of the coefficients in a linear regression model. These regularization terms penalize parameters when they stray too far away from zero. A priori, one may have a strong belief that coefficients in a linear regression model are near zero, so the regularization terms in these objective functions also have a Bayesian interpretation. In the ridge regression, each regularization term is proportional to the logarithm of a normal prior distribution (with mean zero) of a coefficient. In the lasso, each regularization term is proportional to the logarithm of a Laplace prior distribution (with mean zero) of a coefficient.

The approach where one maximizes the log-likelihood function plus the logarithm of the prior distribution is called maximum a posteriori (Geman and Geman 1984). Maximum a posteriori (MAP) involves choosing a model's parameters so as to maximize the posterior distribution of the parameters. This posterior distribution is proportional to the product of the prior distribution of the parameters and the likelihood of the observed data. Consequently, ridge regression and the lasso can be viewed as MAP estimation procedures. Under MAP, a model's parameter estimates represent the mode of the posterior distribution. MAP is often compared to a Bayesian analysis wherein means of the posterior distribution become the model's parameter estimates. Below we describe how we apply MAP to estimate of time-invariant models, including the Bass, gamma/shifted-Gompertz, and tilted-Gompertz models. To estimate our time-varying model, we apply a type of regularized maximum likelihood approach that is closely related to our MAP approach.

An important version of our tilted-Gompertz model is the version with time-invariant parameters. In this version, both smoothing parameters α and β are set to zero. To estimate this version of the model, we choose values for five parameters ϕ, τ, b_0^*, ℓ_0^*, and σ^2 in order to maximize the product of the prior distribution of the parameters and the likelihood of the observed data. Equivalently, we can maximize the logarithm of the posterior distribution.

We start with a prior distribution of the tilted-Gompertz distribution's parameters because this prior distribution is easier to express and confirm from data on previous life cycles. Specifically, we let the parameters $(\lambda, \log(\delta), \log(\rho_0), \log(m_0))$ be jointly normally distributed with density denoted by f_N. We then change variables to find the prior distribution of $(\phi, \tau, b_t^*, \ell_t^*)$. Rather than express the prior distribution of the time-invariant model's parameters in terms of $(\phi, \tau, b_0^*, \ell_0^*)$, we express the prior distribution in terms of $(\phi, \tau, b_t^*, \ell_t^*)$ for convenience. This expression is possible because b_t^* and ℓ_t^* are deterministic functions of $(\phi, \tau, b_0^*, \ell_0^*)$ in the time-invariant model. The reason we express the prior distribution in terms of $(\phi, \tau, b_t^*, \ell_t^*)$ in the time-invariant model is because later

we will use this prior distribution as a regularization term. This term will penalize (or regularize) the parameters at time t in the time-varying model so that they do not stray too far away from our prior beliefs about the time-invariant version of the life cycle. The purpose of this regularization is to combat the overshooting problem.

According to the forecasting method's equations of the time-invariant model, the transformation from $(\lambda, \log(\delta), \log(\rho_0), \log(m_0))$ to $(\phi, \tau, b_t^*, \ell_t^*)$ is given by

$$\begin{aligned}
\lambda &= g_1(\phi,\tau,b_t^*,\ell_t^*) = -\log(\phi) \\
\log(\delta) &= g_2(\phi,\tau,b_t^*,\ell_t^*) = \log\left(\frac{\log\tau}{\log(\phi)(1-\phi)}\right) \\
\log(\rho_0) &= g_3(\phi,\tau,b_t^*,\ell_t^*) = \log\left(\frac{\phi}{1-\phi}\left(\frac{1}{\phi^t}\left(b_t^* - \sum_{i=1}^{t}\phi^{i-1}\log(\tau)\right) - \frac{\log(\tau)}{1-\phi}\right)\right) \\
\log(m_0) &= g_4(\phi,\tau,b_t^*,\ell_t^*) = \ell_t^* - \sum_{i=1}^{t}\phi^i b_t^* - \sum_{i=1}^{t}\frac{1-\phi^i}{1-\phi}\log(\tau) + \rho_0 - \log(\lambda) \\
&\qquad - \delta\log(\rho_0) + \log(\gamma(\delta,\rho_0) - I_{(-\infty,0)}(\lambda)\Gamma(\delta))
\end{aligned} \tag{10}$$

where λ, δ, and ρ_0 in this last equation are $g_1(\phi,\tau,b_t^*,\ell_t^*)$, $\exp(g_2(\phi,\tau,b_t^*,\ell_t^*))$, and $\exp(g_3(\phi,\tau,b_t^*,\ell_t^*))$, respectively. Consequently, the prior distribution of $(\phi,\tau,b_t^*,\ell_t^*)$ is given by

$$\begin{aligned}
&f_N(g_1(\phi,\tau,b_t^*,\ell_t^*), g_2(\phi,\tau,b_t^*,\ell_t^*), g_3(\phi,\tau,b_t^*,\ell_t^*), g_4(\phi,\tau,b_t^*,\ell_t^*))|\det(J)| \\
&\quad = f_N(g_1(\phi,\tau,b_t^*,\ell_t^*), g_2(\phi,\tau,b_t^*,\ell_t^*), g_3(\phi,\tau,b_t^*,\ell_t^*), g_4(\phi,\tau,b_t^*,\ell_t^*))\left|-\frac{1}{\phi}\frac{1}{\tau\log(\tau)}\frac{1}{b_t^* - \frac{\log(\tau)}{1-\phi}}\right|
\end{aligned} \tag{11}$$

where J is the Jacobian of this one-to-one transformation. See the Supplement for a derivation of this prior distribution.

Next we assume that each error's precision σ^{-2} is independent of $(\phi,\tau,b_0^*,\ell_0^*)$ and is distributed according to a gamma distribution with shape $a>1$ and rate $b>0$. The logarithm of the likelihood is $(1/2)\sum_{i=1}^{t}\left(-\log(2\pi)+\log(\sigma^{-2})-\sigma^{-2}(y_i^*-\hat{y}_i^*)^2\right)$ where $\hat{y}_i^*$ is the one-step-ahead forecast at time $i-1$ according to the forecasting method in (4). Thus, the objective we maximize in choosing the parameters ϕ, τ, b_0^*, ℓ_0^*, and σ^2 is given by

$$\begin{aligned}
&\log\left(f_N(g_1(\phi,\tau,b_t^*,\ell_t^*), g_2(\phi,\tau,b_t^*,\ell_t^*), g_3(\phi,\tau,b_t^*,\ell_t^*), g_4(\phi,\tau,b_t^*,\ell_t^*))\right) \\
&+\log\left|-\frac{1}{\phi}\frac{1}{\tau\log(\tau)}\frac{1}{b_t^* - \frac{\log(\tau)}{1-\phi}}\right| + a\log(b) - \log(\Gamma(a)) + (a-1)\log(\sigma^{-2}) - b\sigma^{-2} \\
&+\frac{1}{2}\sum_{i=1}^{t}\left(-\log(2\pi)+\log(\sigma^{-2})-\sigma^{-2}(y_i^*-\hat{y}_i^*)^2\right)
\end{aligned} \tag{12}$$

Here, the MAP estimate of σ^2 is $(2b+\sum_{i=1}^{t}(y_i^*-\hat{y}_i^*)^2)/(2(a-1)+t)$. As we approach a diffuse prior (i.e., $a\to 1$ and $b\to 0$), the MAP estimate of σ^2 goes to the maximum likelihood estimate of σ^2. One can find estimates for the other four parameters via a numerical optimization routine.

For the model with time-varying parameters, we maximize the objective in (12) with the addition of terms for the prior beta distributions of α and β. This new objective (up to an additive constant) is no longer the logarithm of the posterior distribution of the model's parameters. This is because the time-varying parameters b_t^* and ℓ_t^* are functions of many error terms as they evolve from b_0^* and ℓ_0^*. Nonetheless, we can view this objective as a regularized log-likelihood where the time-variant model is penalized for updating too far away from some prior time-invariant model. In other words, the prior distribution of $(\phi, \tau, b_t^*, \ell_t^*, \sigma^2)$ in the time-invariant model becomes a regularization term in the time-varying model.

5.2. Model Evaluation: Pinball Loss

To evaluate forecasting accuracy, we score each model's point forecast and two predictive intervals. The point forecast we score is the median, or 0.50-quantile, denoted by $Q_{0.50}$. We also score four quantiles that describe the central 50% and 90% prediction intervals: (i) the 0.05-quantile ($Q_{0.05}$), (ii) the 0.25-quantile ($Q_{0.25}$), (iii) the 0.75-quantile ($Q_{0.75}$), and (iv) the 0.95-quantile ($Q_{0.95}$). The scoring rule we use is the popular pinball loss function (Hong et al. 2016), which maps directly onto the newsvendor problem (Jose and Winkler 2009, Grushka-Cockayne et al. 2017). The pinball loss of the p-quantile (Q_p) given the realization y is

$$L(Q_p, y) = \begin{cases} p(y - Q_p) & \text{for} \quad Q_p \leq y \\ (1-p)(Q_p - y) & \text{for} \quad Q_p > y \end{cases}.$$

The pinball loss function has a familiar form. For the 0.50-quantile (or median), the pinball loss function is equivalent to one-half times the mean absolute error.

5.3. Study 1: Forecasting Search Interest in Social Networks

For this study, we gathered data from Google Trends (2016) on weekly search interest in 122 social networks from January 2004 to June 2016. The names of these networks came from a list of 211 major, active social networking websites (Wikipedia 2016). The Google trends data represent normalized search volume, a proxy for the use of the network's service. The search volume of each term is normalized by Google so that the volume's peak equals 100. We did not gather data on 89 networks 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 interest in the network, (iii) search interest was high well before the network's launch date, (iv) the network was founded before January 2004 and the first observation is greater than one, or (v) the first observation of the network's longest contiguous stretch of positive search interest (after a previous stretch of positive search interest) 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 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".

For the 122 networks in this study, we disaggregated their weekly data from Google Trends to equal daily amounts each week, and then aggregated these daily data up to the monthly level. Next we found each network's longest stretch of positive search interest after its launch date. The minimum, median, mean, and maximum of these stretches were 28, 108, 104.2, 150 months, respectively. We then deseasonalized the resulting monthly time series using the ratio-to-moving-average approach (Taylor 2003). Finally, we renormalized the resulting deseasonalized time series so that each time series's peak was at 100.

Below we compare the forecasting accuracy of seven models on the 122 time series in this study. The seven models are: (i) the naïve model, (ii) the Bass model estimated by MAP, (iii) the Bass model estimated by EKF-CD, (iv) the gamma/shifted-Gompertz model estimated by MAP, (v) the trapezoid model estimated by MAP, (vi) the tilted-Gompertz model with time-invariant (TI) parameters estimated by MAP, and (vii) the tilted-Gompertz model with time-varying (TV) parameters estimated by our regularized maximum likelihood (RML) approach. For each time series, out-of-sample forecasts were made at 1-step through 24-steps ahead, on a rolling basis, starting from time zero.

To generate the prior distributions for the MAP, EKF-CD, and RML estimation procedures, we use a two-fold process. Each of the two folds has 61 time series randomly assigned to it. We use one fold to inform our prior distributions, called the prior-information fold, and we then make rolling out-of-sample forecasts on the time series in the other fold, called the out-of-sample fold. We go through this process twice and report the average accuracy for both fold's rolling out-of-sample predictions.

To specify the prior distributions in the time-invariant models estimated by MAP and the Bass model estimated by EKF-CD, we take one pass over the prior-information fold. In this single pass over the prior-information fold, we fit time-invariant models using maximum likelihood estimation to the time series in the prior-information fold. To estimate the normal prior distribution's mean, we fit a second-stage time-invariant model to the average of the first-stage models' fitted life cycles. The parameters of this second-stage time-invariant model are jointly robust and become the normal

prior distribution's mean for the out-of-sample fold. From the matrix of estimated parameters (which has one row of parameter estimates for each time series in the prior-information fold), we then calculate the robust covariance and use it as the normal prior distribution's covariance. For the prior gamma distribution of the error term's precision $(1/\sigma^2)$, we use robust estimators: the median and robust variance of the error term's precision from a model fit to each time series in the prior-information fold. These robust estimators become the mean and variance in the prior gamma distribution of the error term's precision in the out-of-sample fold.

To specify the prior distributions in the tilted-Gompertz model with time-varying (TV) parameters, we take a second pass over the prior-information fold. In this second pass, we estimate the parameters in the prior beta distributions for the smoothing parameters (α and β) and the prior gamma distribution of the error term's precision $(1/\sigma^2)$. We use this second pass to estimate the error term's precision (instead of using the first pass) because this precision interacts with the smoothing parameters in the time-varying model. In our second pass, we fit the time-varying model, on a rolling basis, using the prior distributions specified in the first pass. For each rolling training set, we make forecasts 13- through 24-steps ahead and record their average pinball losses for the five quantiles described above. The time-varying model is fit multiple times with different settings for the parameters in the beta and gamma prior distributions. A grid search to minimize the overall average pinball loss determines the final values for these parameters. It is these values that we use when we make rolling forecasts in the out-of-sample fold.

We define the naïve model's forecast at time $t > 0$ for every step ahead h as the most recent observation of the time series y_t. At $t = 0$, the naïve model's forecast is the median y_1 of the time series in the prior-information fold. All data and code from this study and the next are available from the authors.

Tables 1 and 2 present the pinball losses for shorter-term forecasts (1-12 steps-ahead) and longer-term forecasts (13-24 steps-ahead) in Study 1. Each entry in the first five columns is the pinball loss of a model's quantile, averaged over the 12 different steps ahead, over each time series' rolling training sets, and over the 122 time series. Each entry in the last column is the overall average of the five quantiles' pinball losses in that row.

From these tables, we can see that the tilted-Gompertz model with time-varying parameters has the lowest overall average pinball loss at both forecasting horizons. We also notice that among the time-invariant models, the tilted-Gompertz model has the best overall average performance. The next-best model overall is the gamma/shifted-Gompertz model. Consequently, in each table,

	$Q_{0.05}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.95}$	Overall Average
Naïve	—	—	**5.62**	—	—	—
Bass	1.85	5.65	7.98	8.04	5.33	5.77
Bass (EKF-CD)	9.44	18.93	13.81	22.97	10.17	15.07
Gamma/Shifted-Gompertz	1.93	5.08	6.98	7.08	5.01	5.22
Trapezoid	2.57	6.74	8.61	7.85	4.46	6.05
Tilted-Gompertz (TI)	1.72	5.00	6.95	6.67	3.64	4.80
Tilted-Gompertz (TV)	**1.27**	**4.26**	6.17	**6.27**	**3.22**	**4.24**
Difference[a]	-0.66	-0.82	-0.81	-0.81	-1.79	-0.98
Std. Error	0.09	0.13	0.18	0.23	0.28	0.14
Significance[b]	***	***	***	***	***	***

[a] Each difference is the column's Tilted-Gompertz (TV) loss minus Gamma/Shifted-Gompertz loss.
[b] The symbol *** indicates significance at the 99.9% level.

Table 1 Pinball Losses of 1-12 Steps-Ahead Rolling Forecasts of Search Interest in Social Networks.

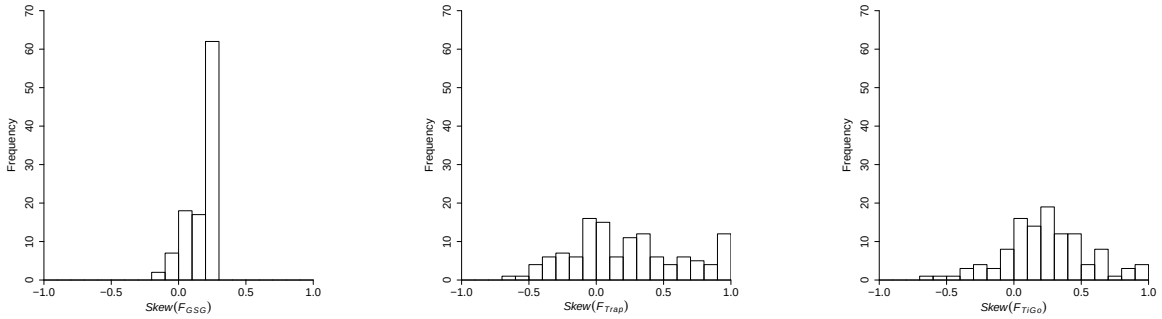
	$Q_{0.05}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.95}$	Avg.
Naïve	—	—	11.48	—	—	—
Bass	2.83	8.47	12.44	13.87	11.72	9.87
Bass (EKF-CD)	9.29	20.38	17.98	23.56	10.31	16.30
Gamma/Shifted-Gompertz	3.01	7.77	11.32	12.80	11.36	9.25
Trapezoid	4.98	10.51	13.36	13.27	9.88	10.40
Tilted-Gompertz (TI)	4.13	9.04	11.85	11.64	7.64	8.86
Tilted-Gompertz (TV)	**2.38**	**7.54**	**11.21**	**11.37**	**6.33**	**7.77**
Difference[a]	-0.62	-0.23	-0.11	-1.42	-5.03	-1.48
Std. Error	0.19	0.27	0.39	0.52	0.68	0.34
Significance[b]	**			**	***	***

[a] Each difference is the column's Tilted-Gompertz (TV) loss minus Gamma/Shifted-Gompertz loss.
[b] The symbols ** and *** indicate significance at the 99%, and 99.9% levels, respectively.

Table 2 Pinball Losses of 13-24 Steps-Ahead Rolling Forecasts of Search Interest in Social Networks.

we report the differences between the tilted-Gompertz (TV) and gamma/shifted-Gompertz models, by quantile and overall average. The tilted-Gompertz (TV) model is statistically significantly better than the gamma/shifted-Gompertz model for most quantiles, at most forecasting horizons. Our time-varying model is also better than the gamma/shifted-Gompertz model overall at both forecasting horizons.

Not surprisingly, the naïve (or no-change) model produces the best point shorter-term forecasts (1-12 steps ahead). This result lends some support to those who advocate for the use of simple forecasting models (Armstrong 2001, Green and Armstrong 2015). At the longer-term horizon (13-24 steps ahead), many of the life-cycle models, however, are more accurate than the naïve model. One might expect a life-cycle model, which can predict growth or decline depending on where in the life cycle the rolling forecast is made, to dominate a no-change forecast further into the future. In the context of life cycles, which often grow and decline in dramatic and systematic ways, a more complicated model may be a more appropriate, even for point forecasting purposes.

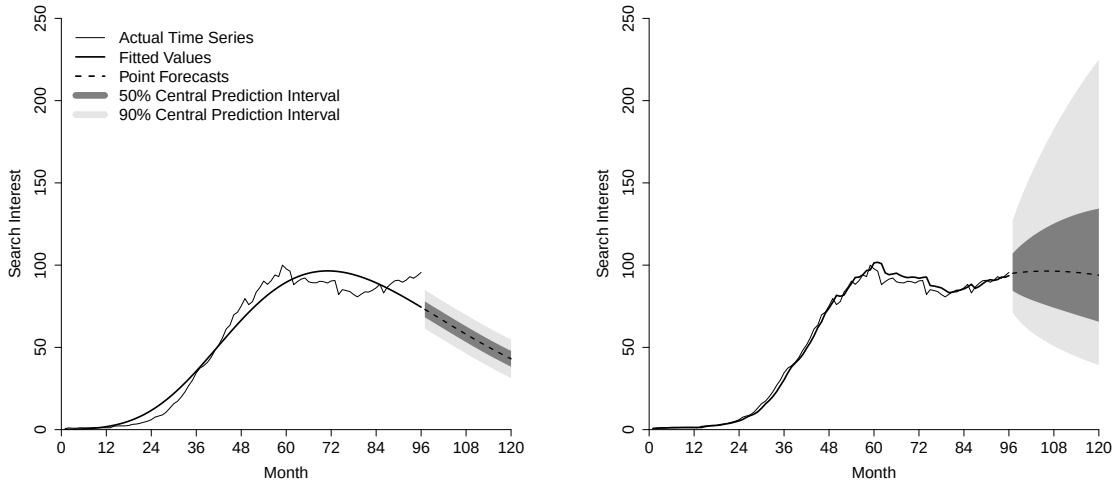


(a) Gamma/Shifted-Gompertz Model (b) Trapezoid Model (c) Tilted-Gompertz (TI) Model

Figure 4 Histograms of Local Skewness Values from Three Models Fitted to 122 Search Interest Life Cycles.

We believe the tilted-Gompertz model outperforms the gamma/shifted-Gompertz model for two reasons. First, the tilted-Gompertz (TI) model can capture "extra-Bass" skew better than the gamma/shifted-Gompertz model can, even though both of these time-invariant models have four parameters. See Figure 4, which contains histograms of the local skewness values calculated for the tilted-Gompertz (TI) and gamma/shifted-Gompertz models fit to each of the 122 time series. From these histograms, we see that the tilted-Gompertz model takes on a wider range of skewness values, which suggests there is more "extra-Bass" skew in these search interest life cycles than the gamma/shifted-Gompertz model can capture. Although the trapezoid model's histogram of local skewness values looks similar to tilted-Gompertz (TI) model's, the trapezoid model, with its five parameters, may not perform as well because its shape cannot capture exponential growth in a life cycle's growth stage, deceleration in its maturity stage, and exponential decline in its decline stage. Its rate of adoption is linear in the growth stage, flat in the maturity stage, and linear in the decline stage.

Second, the best tilted-Gompertz model has time-varying parameters. Its time-varying parameters allow it to cope with local changes in the environment. For example, the model can fit itself to a second peak in a life cycle. Figure 5 demonstrates this effect. For this figure, the tilted-Gompertz and gamma/shifted-Gompertz models are fit to search interest in the social network Tumblr, and forecasts are produced 24-months ahead. Here, the tilted-Gompertz model (with time-varying parameters) is able to forecast a second peak, while the gamma/shifted-Gompertz model (with time-invarying parameters) cannot. Note that we also depict the 50% and 90% prediction intervals for these two models. The tilted-Gompertz (TV) model's prediction intervals are wider because of the model's time-varying parameters and multiplicative errors.



(a) Gamma/Shifted-Gompertz Model

(b) Tilted-Gompertz (TV) Model

Figure 5 Point Forecasts and Prediction Intervals from Gamma/Shifted-Gompertz and Tilted-Gompertz (TV) Models Fit to Search Interest in Tumblr.

We considered multiplicative errors as a third reason why the tilted-Gompertz model outperformed the benchmark models, but we did not find evidence in the study to support this conjecture. When we estimated the Bass and gamma/shifted-Gompertz models with multiplicative errors, these two models performed worse than their additive-error counterparts in terms of overall average pinball loss, except for the Bass model at 13-24 steps-ahead forecasting. Of the four models—the Bass and gamma/shifted-Gompertz models with additive and multiplicative errors—the gamma/shifted-Gompertz model with additive errors was the most accurate in terms of overall average pinball loss. Therefore, we only report in the tables above results for the Bass and gamma/shifted-Gompertz models with additive errors.

5.4. Study 2: Forecasting New Computer Sales

For this study, we used the computer sales data described in Acimovic et al. (2018) and analyzed in Hu et al. (2017). These data include weekly sales of 170 different new Dell computers (fixed and mobile workstations, laptops, and desktops) during the period 2013 to 2016. During this period, 4,037,825 units of these products were sold, worth over one billion dollars in revenue. The data are normalized so that the total sales over any complete life cycle is equal to one. The median length of these life cycles is much shorter than in our previous study: 40 weeks compared to 108 months. The weekly data in this study are also noisier than in the previous study. As such, we forecast the complete life cycle of each new product from its very beginning, rather than make several rolling out-of-sample forecasts. Hu et al. (2017) also make out-of-sample forecasts for complete life cycles.

To forecast the complete life cycles in this study, we again use a two-fold process. Each of the two folds has 85 time series randomly assigned to it. Because we make forecasts of the complete life cycle, we only estimate the time-invariant version of our tilted-Gompertz model. Therefore, we only need to take one pass over a prior-information fold, which we do using the first pass described above in Study 1. The single model fit to the average of the models' fitted life cycles in the prior-information fold is used to forecast each time series in the out-of-sample fold. This type of forecast treats all products in the prior-information fold as if they came from a single cluster. Based on additional data about product features and categories, Hu et al. (2017) use several clusters in their approach. Since we do not have access to this proprietary data, we cannot do better than to use a single cluster.

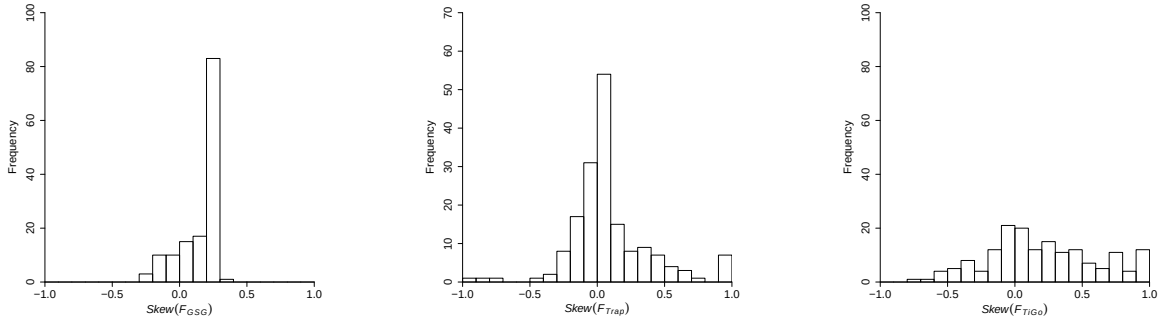
Table 3 presents the pinball losses of the forecasts for the complete life cycles in Study 2. The tilted-Gompertz model has the lowest overall average pinball loss and is best at three of the five quantiles. The difference in the overall average pinball loss between the tilted-Gompertz model and the second best model overall (the gamma/shifted-Gompertz model) is statistically significant at the 99% level.

	$Q_{0.05}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.95}$	Avg.
Bass	0.00162	0.00670	0.01046	0.01172	0.00992	0.00808
Gamma/Shifted-Gompertz	**0.00157**	0.00655	0.01036	0.01173	0.01003	0.00805
Trapezoid	0.00167	0.00672	0.01047	**0.01166**	0.00981	0.00807
Tilted-Gompertz (TI)	0.00180	**0.00644**	**0.01029**	0.01176	**0.00943**	**0.00794**
Difference[a]	0.00023	-0.00011	-0.00007	0.00003	-0.00061	-0.00011
Std. Error	0.00003	0.00003	0.00005	0.00006	0.00011	0.00004
Significance[b]	***				***	**

[a] Each difference is the column's Tilted-Gompertz (TI) loss minus Gamma/Shifted-Gompertz loss.
[b] The symbols ** and *** indicate significance at the 99%, and 99.9% levels, respectively.

Table 3 Pinball Losses of Forecasts for Complete Life Cycles of New Computer Sales.

See Figure 6 for histograms of the local skewness values of the gamma/shifted-Gompertz, trapezoid, and tilted-Gompertz (TI) models fit to the 170 new products. Here again, the tilted-Gompertz model takes on a wider range of skewness values than the gamma/shifted-Gompertz model, which suggests there is more "extra-Bass" skew in these new product life cycles than the gamma/shifted-Gompertz model can capture. Although the trapezoid model's local skewness values are also on a wider range, the model may not perform well in this study, again because of its piecewise linear shape.



(a) Gamma/Shifted-Gompertz Model (b) Trapezoid Model (c) Tilted-Gompertz (TI) Model

Figure 6 Histograms of Local Skewness Values from Three Models Fitted to 170 New Computer Sales Life Cycles.

6. Conclusion

In this paper, we introduce a new life-cycle model based on the principles of exponential smoothing. In two empirical studies, we demonstrate that the model outperforms several benchmarks in out-of-sample forecasting. The trend in our exponential smoothing model follows the density of a new distribution, the tilted-Gompertz distribution. This trend has the ability to fit a wider range of skewed life-cycles than existing models can. When used to model a diffusion process, our life-cycle trend can be interpreted as the distribution for an individual's uncertain adoption time. Because the model has time-varying parameters, it can also adapt to local changes in the marketplace or even to the firm's own marketing mix.

Importantly, the model includes multiplicative errors, instead of the usual assumption of additive errors. These multiplicative errors enable the model to maintain strictly positive prediction intervals. In our empirical study of search volumes, we find that the model's prediction intervals are also more accurate than existing diffusion models with either additive or multiplicative errors. Another reason the model produces accurate quantile forecast is the inclusion of prior information through regularization terms in our model. Including prior information is crucial for accurately forecasting a product life cycle from its beginning. Because our model is based on exponential smoothing, it is also computationally efficient to estimate. Overall, the model appears to be well-suited for practical use in large-scale forecasting environments where key operational decisions depend on quantile forecasts.

The use of our modeling approach in practice will depend, in large part, on access to relevant training sets. In our empirical study, we used half of the time series to form prior distributions in the model. For each new time series in another setting, a manager will need to identify completed

life cycles that are similar to the upcoming life cycle. In a large-scale setting, it may be possible to identify large sets (or clusters) of comparable completed life cycles, as discussed in Hu et al. (2017).

References

Acimovic, Erize F Hu K Thomas DJ, J, JA Van Mieghem. 2018. Product life cycle data set: Raw and cleaned data of weekly orders for personal computers. *Manufacturing & Service Operations Management* Forthcoming.

Armstrong JS. 2001. Standards and practices for forecasting. *Principles of Forecasting*. Springer, 679–732.

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

Baardman L, Perakis G, Levin I, Singhvi D. 2017. Leveraging comparables for new product sales forecasting Working Paper, available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3086237`.

Bass FM. 1969. A new product growth for model consumer durables. *Management Science* **15** 215–227.

Bemmaor AC. 1994. Modeling the diffusion of new durable goods: Word-of-mouth effect versus consumer heterogeneity. Pras B Laurent G, Lilien GL, ed., *Research Traditions in Marketing*. Kluwer, 201–229.

Bemmaor AC, Lee J. 2002. The impact of heterogeneity and ill-conditioning on diffusion model parameter estimates. *Marketing Science* **21** 209–220.

Brown RG. 1959. *Statistical Forecasting for Inventory Control*. McGraw-Hill.

Brown RG. 1963. *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice Hall.

Corless, Gonnet GH Hare DEG Jeffrey DJ, RM, DE Knuth. 1996. On the LambertW function. *Advances in Computational mathematics* **5** 329–359.

Dixon R. 1980. Hybrid corn revisited. *Econometrica* **48** 1451–1461.

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

Efron B. 1981. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics* **9** 139–158.

Gardner Jr. ES, McKenzie E. 1985. Forecasting trends in time series. *Management Science* **31** 1237–1246.

Geman S, Geman D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6) 721–741.

Gneiting T. 2011. Quantiles as optimal point forecasts. *International Journal of Forecasting* **27** 197–207.

Golder PN, Tellis GJ. 1997. Will it ever fly? Modeling the takeoff of really new consumer durables. *Marketing Science* **16** 256–270.

Goldman A. 1982. Short product life cycles: implications for the marketing activities of small high-technology companies. *R&D Management* **12** 81–90.

Gompertz B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* **115** 513–583.

Google Trends. 2016. `https://www.google.com/trends/`. Accessed: June 29-30, 2016.

Goutis C, Casella G. 1999. Explaining the saddlepoint approximation. *The American Statistician* **53** 216–224.

Green KC, Armstrong JS. 2015. Simple versus complex forecasting: The evidence. *Journal of Business Research* **68**(8) 1678–1685.

Grushka-Cockayne Y, Jose VRR, Lichtendahl Jr. KC, Winkler RL. 2017. Quantile evaluation, sensitivity to bracketing, and sharing business payoffs. *Operations Research* **65**(3) 712–728.

Hoerl AE, Kennard RW. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**(1) 55–67.

Holt CC. 1957. Forecasting seasonals and trends by exponentially weighted averages. ONR Memorandum 52, Carnegie Institute of Technology.

Holt CC. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* **20** 5–10.

Hong T, Fan S, Pinson P, Zareipour H, Troccoli A, Hyndman RJ. 2016. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting* **32**(3) 896–913.

Hu K, Erize F, Acimovic J, Thomas DJ, Van Mieghem JA. 2017. Forecasting product life cycle curves: Practical approach and empirical analysis. *Manufacturing & Service Operations Management* Forthcoming.

Hyndman RJ. 2015. R vs Autobox vs ForecastPro vs ... `http://robjhyndman.com/hyndsight/show-me-the-evidence/`.

Hyndman RJ, Ord JK, Koehler AB, Snyder RD. 2008. *Forecasting with Exponential Smoothing*. Springer-Verlag, Berlin, Germany.

Johnson NL, Balakrishnan N, Kotz S. 1995. *Continuous Univariate Distributions, Volume 2*. John Wiley & Sons, New York, NY.

Jose VRR, Winkler RL. 2009. Evaluating quantile assessments. *Operations Research* **57** 1287–1297.

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

Mahajan V, Srinivasan V, Mason CH. 1986. An evaluation of estimation procedures for new product diffusion models. Wind J Mahajan V, ed., *Innovation Diffusion Models of New Product Acceptance*. Ballinger Publishing Co., Cambridge, MA, 203–232.

Marshall AW, Olkin I. 2007. *Life Distributions*. Springer, New York, NY.

Meade N, Islam T. 2006. Modelling and forecasting the diffusion of innovation–a 25-year review. *International Journal of Forecasting* **22** 519–545.

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

Pegels CC. 1969. Exponential forecasting: Some new variations. *Management Science* **12** 311–315.

Petruzzi NC, Dada M. 1999. Pricing and the newsvendor problem: A review with extensions. *Operations Research* **47** 183–194.

Schmittlein DC, Mahajan V. 1982. Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing Science* **1** 57–78.

Srinivasan V, Mason CH. 1986. Nonlinear least squares estimation of new product diffusion models. *Marketing Science* **5** 169–178.

Tassone E, Rohani F. 2017. Our quest for robust time series forecasting at scale. `http://www.unofficialgoogledatascience.com/2017/04/our-quest-for-robust-time-series.html`.

Taylor JW. 2003. Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting* **19** 715–725.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1) 267–288.

Van den Bulte C, Joshi YV. 2007. New product diffusion with influentials and imitators. *Marketing Science* **26** 400–421.

Van den Bulte C, Lilien GL. 1997. Bias and systematic change in the parameter estimates of macro-level diffusion models. *Marketing Science* **16** 338–353.

Venkatesan R, Kumar V, Krishnan TV. 2004. Evolutionary estimation of macro-level diffusion models using genetic algorithms: An alternative to nonlinear least squares. *Marketing Science* **23** 451–464.

Wikipedia. 2016. List of social networking websites. `https://en.wikipedia.org/w/index.php?title=List_of_social_networking_websites&oldid=719695760`. Accessed: June 10, 2016.

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

Winters PR. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science* **6** 324–342.

Xie J, Sirbu M, Song XM, Wang Q. 1997. Kalman filter estimation of new product diffusion models. *Journal of Marketing Research* **34** 378–393.