ORIGINAL PAPER

CrossMark

Multistep quantile forecasts for supply chain and logistics operations: bootstrapping, the GARCH model and quantile regression based approaches

Joanna Bruzda[1] (ORCID)

Abstract

In this paper, we discuss and compare empirically various ways of computing multistep quantile forecasts of demand, with a special emphasis on the use of the quantile regression methodology. Such forecasts constitute a basis for production planning and inventory management in logistics systems optimized according to the cycle service level approach. Different econometric methods and models are considered: direct and iterated computations, linear and nonlinear (GARCH) models, simulation and non-simulation based procedures and parametric as well as semiparametric specifications. These methods are applied to compute multiperiod quantile forecasts of the monthly microeconomic time series from the popular M3 competition database. According to various accuracy measures for quantile predictions, the best procedures are based on simulation techniques using predictive distributions generated by either the quantile regression methodology combined with random draws from the uniform distribution or parametric and nonparametric bootstrap techniques. These methods lead to large reductions in the total costs of logistics systems as compared with non-simulation based procedures. For example, in the case of forecasting 12 months ahead, relatively short time series and a high cycle service level, the quantile regression based simulation approach reduces the average supply chain cost per unit of output by about 70–85%. At the shortest horizons, the GARCH model should be seriously considered among the preferred forecasting solutions for production and inventory planning.

Keywords Production planning · Inventory management · Supply chain forecasting · Quantile forecasts · LINLIN loss · M3 competition

JEL Classification C14 · C22 · C53

✉ Joanna Bruzda
  Joanna.Bruzda@uni.torun.pl

1   Department of Logistics, Nicolaus Copernicus University, Gagarina 11, 87-100 Toruń, Poland

1 Introduction

Demand forecasting in supply chain and logistics management is important for several reasons, including service level targeting, costs minimization and managing the overall risk of supply chains. It constitutes the first step an organization should undertake in its strategic and operational planning (see, for example, Chopra and Meindl 2007, Chap. VII). Because of this, forecasting is among the most important components of logistics systems. The uncertainty concerning future demand makes supply chain management very challenging, while the forecast accuracy to a large extent defines the overall cost of a supply chain (see Syntetos et al. 2016, and compare, for example, Thomassey 2010), even if managers not always consider costs and key financial measures when evaluating their forecasts (see results of the survey studies discussed by McCarthy et al. 2006). From the perspective of risk forecasting and management, the so-called demand risk is considered as one of the most important components of the total risk faced by a supply chain, which consists of risks within a company such as process risks and control risks, risks outside of a company but within the supply chain, encompassing demand risks and supply risks, and risks outside of the supply chain—so-called environmental risks (see Pohl et al. 2010). Thus, the demand risks are considered as a part of the within supply chain risks. On the other hand, the supply chain risks are also divided into operational risks associated with some 'usual' uncertainties and so-called disruption risks of natural or anthropogenic kind such as natural disasters, terrorist attacks or economic crises (see, for example, Tang 2006; Choi and Chiu 2012, section 1.1). The 'usual' uncertainties defining operational risks are those resulting from demand and supply uncertainty or costs variability, and thus the demand risks are basically treated as operational uncertainties.[1] However, in principle, the 'unusual' uncertainties will also influence the variability of demand, the uncertainty of supply as well as the total variability of operational costs. This may call for considering demand risk within a different notion of uncertainty, namely that called Knightian uncertainty in recognition of the work of Knight (1921), in which case the probability model of a random variable of interest to the decision-maker is unknown and not estimable with a reasonable accuracy and which differs from the notion of risk under a known probability distribution. It also calls for a broader usage of semiparametric methods in more objectively dealing with what Knight (1921) names a third type of uncertainty and which is the error associated with estimating a probability distribution.

Quantile forecasts of demand are of interest to both supply chain managers, who want, for example, to find the optimal level of product availability or the supply chain's production or distribution potential, as well as managers in a single firm willing to optimize the firm's inventory level (see, for example, Pearson 2006; Chopra and Meindl 2007, Chap. XII; Ferbar et al. 2009; Tratar 2010, 2015; Wang and Petropoulos 2016;

---

[1] A comprehensive review of definitions and classifications of supply chain risks that were developed in the literature can be found in Sodhi et al. (2012) and Heckmann et al. (2015). There are also different mathematical approaches to mitigate supply chain risks. A non-exhaustive review of them is given in Choi and Chiu (2012, Chapter 1), and covers the use of such diversified concepts as von Neumann–Morgenstern utility functions, profit target probability measures, Value-at-Risk and Conditional Value-at-Risk, and mean-risk (mean–variance, mean-semivariance) analysis (see Choi and Chiu 2012, Table 1.1).

and the dynamic formulation of the newsvendor model in Alwan et al. 2016).[2] In particular, multiperiod quantile forecasts are indispensable under the forecast based periodic review (T, S) policy in the popular cycle service level (CSL) framework (see, for example, Strijbosch and Moors 2005; Syntetos and Boylan 2008; and Syntetos et al. 2010).[3] In such policy, the inventory position is reviewed at equidistant points in time and replenished to a certain level depending on the assumed service level requirements. To explicitly define this inventory policy, we introduce the following notation:

T	Prespecified review (replenishment) period (in months),
S	Order-up-to (target) inventory level (in units),
L	Constant lead time (in months),
y_t	Random demand in period t,
$y_{t,n}$	Cumulated random demand over n consecutive periods starting in t, i.e., $y_{t,n} = \sum_{i=0}^{n-1} y_{t+i}$,
$\hat{y}_t, \hat{y}_{t,n}$	Forecasts of the mean values of y_t and $y_{t,n}$, respectively, i.e., the conditional expectations $E(y_t\|\Omega_{t-1})$ and $E(y_{t,n}\|\Omega_{t-1})$ given the set of information available in period $t-1$, Ω_{t-1},
$\hat{\sigma}_t^2, \hat{\sigma}_{t,n}^2$	Forecasts of the variances of y_t and $y_{t,n}$, respectively, i.e., the conditional variances $Var(y_t\|\Omega_{t-1}) = E(y_t^2\|\Omega_{t-1}) - E^2(y_t\|\Omega_{t-1})$ and $Var(y_{t,n}\|\Omega_{t-1}) = E(y_{t,n}^2\|\Omega_{t-1}) - E^2(y_{t,n}\|\Omega_{t-1})$,
$\Phi(\cdot), \Phi_n(\cdot)$	Standardized conditional cumulative distribution functions of y_t and $y_{t,n}$, respectively, which are assumed to be absolutely continuous and constant over time, and
CSL	Quantile order of interest to the decision-maker, i.e., service level requirement in terms of the probability of no stockout during the replenishment cycle.

Then, the order-up-to level S is defined through the equation:

$$P\{y_{t,T+L} \leq S\} = CSL,$$

i.e., as the quantile of order *CSL* of the cumulated actual demand over the period $T+L$. This is solved as:

$$S = \hat{y}_{t,T+L} + \Phi_{T+L}^{-1}(CSL) \cdot \hat{\sigma}_{t,T+L}. \tag{1}$$

---

[2] It is worth mentioning that quantile forecasts are also widely discussed in the context of energy forecasting. For example, the Global Energy Forecasting Competition (GEFC) is concerned with comparing probabilistic forecasting methods through the computation of forecasts of load or price at 99 quantiles on a rolling basis—see Hong et al. (2016) for a review of the winning solutions in the GEFC 2014, encompassing both popular and new more advanced methods (often making use of the quantile regression methodology). This discussion supplements the stream of research concerning the use of simplified and advanced solutions to the problem of constructing mean value forecasts in energy forecasting (see, for example, De Cosmis et al. 2013; Leopold 2016; Nalcaci et al. 2018).

[3] Compare Dallery and Babai (2005), Syntetos et al. (2010) and Hill et al. (2015) for an introduction to quantile forecasting in a simple reorder point based system and certain approximate solutions to this task. Also note that reorder point systems will in principle require the use of continuous time models as in, for example, Silver et al. (2009).

Under the Gaussianity, stationarity and independence assumptions, (1) can be rewritten as:

$$S = \hat{y} \cdot (L + T) + \Phi^{-1}(CSL) \cdot \hat{\sigma} \cdot \sqrt{L + T}. \tag{2}$$

The scaling or 'square root of $T + L$' method (2) is a popular textbook solution and an industry standard for computing quantile forecasts of demand, and is also used as an approximation in the case of correlated demand.

In this paper, we undertake the task of an empirical evaluation of different procedures of multiperiod quantile forecasting which do not require the stationarity and independence assumptions, encompassing direct and iterated approaches, maximum likelihood (ML) and quantile regression estimation, simulation and non-simulation based methods and models with homoscedastic as well as GARCH errors. The basis for this study constitutes a popular dataset of demand data—the monthly microeconomic time series from the M3 forecast competition database (see Makridakis and Hibon 2000). The main part of this paper is preceded by an introductory discussion concerning alternative decision-theoretic approaches that can be of use in supply chain and logistics forecasting. Then, Sect. 3 presents the analyzed methods together with our evaluation framework, encompassing different measures of accuracy for quantile forecasts and tests of quantile unbiasedness, and Sect. 4 discusses the main findings. Finally, the last section offers brief conclusions.

2 Forecasting for supply chain and logistics operations

It is worth recalling the basic fact that the CSL parameter can be optimized assuming that the costs of under- and overstocking are linear functions of the forecast error (see, for example, Chopra and Meindl 2007, Chap. XII). Thus, it is assumed that the supply chain cost is given as:

$$L(y, \hat{y}) = \begin{cases} c_u(y - \hat{y}) & \text{for } y \geq \hat{y} \\ c_o(\hat{y} - y) & \text{for } y < \hat{y} \end{cases}, \tag{3}$$

where c_u and c_o are the unit costs of under- and overstocking, and y and $\hat{y}$ are the realization and the forecast, respectively. The function (3) is the well known LINLIN (doubly linear) forecast loss (cost) function, which is of the 'prediction error' type, i.e., it is an example of a forecast loss dependent solely on the forecast error. Then, applying to (3) the Bayes rule:

$$E[L(Y, \hat{y})] \rightarrow \min \tag{4}$$

leads to optimal forecasts in the form of quantile predictions (see, for example, Gneiting 2011a, b; Komunjer 2013), i.e., the optimal forecast is as follows:

$$\hat{y} = F^{-1}\left(\frac{c_u}{c_u + c_o}\right) = F^{-1}(\tau), \tag{5}$$

where Y is distributed according to the forecaster's subjective or objective predictive distribution F of demand and, for simplification, we assume throughout the discussion that the distribution F is absolutely continuous and strictly increasing. Thus, under the loss function (3), optimal forecasts in a given logistics system take the form of quantiles of order $CSL = \tau = \frac{c_u}{c_u + c_o}$ of the demand distribution.

An alternative way of optimizing a logistics system is based on the notions of Expected Shortage (ES) and Fill Rate (FR). The Expected Shortage is defined as the mean level of lost sales under a prespecified product availability S, i.e., as:

$$ES(S) = E(Y - S)^+ = \int_S^{+\infty} (y - S)f(y)dy,$$

where f is the probability density function associated with F. On the other hand, the Fill Rate is the fraction of demand immediately available from stock. In its so-called infinite period version, it is defined as follows:

$$FR(S) = \frac{E\min\{S,\ Y\}}{EY} = 1 - \frac{E(Y - S)^+}{EY} = 1 - \frac{ES(S)}{EY}, \tag{6}$$

The Fill Rate (6) is also known as the type II service level, while the CSL is called the type I service level or the vendor service level (see, for example, Lewis 1997; Brandimarte and Zotteri 2007).

Then, the optimization of a logistics system according to the type II service level requires solving:

$$1 - \frac{E(Y - S)^+}{EY} = \tau \tag{7}$$

or, alternatively,

$$ES(S) = (1 - \tau)EY,$$

where $\tau \in (0,\ 1)$ denotes the required service level. In practice, this is often performed under the Gaussianity assumption, $Y \sim N(m_Y, \sigma_Y)$, in which case we have:

$$ES(S) = \sigma_Y\, G\left(\frac{S - m_Y}{\sigma_Y}\right) \tag{8}$$

with

$$G(z) = \int_z^{\infty} (y - z)\varphi(y)dy = \varphi(z) - z[1 - \Phi(z)],$$

where φ and Φ are the probability density function and the distribution function of the standard normal distribution.

On the contrary, the so-called single period version of the Fill Rate is given as (see Chen et al. 2003; Thomas 2005):

$$FR^1(S) = E\frac{\min\{S, \ Y\}}{Y} \tag{9}$$

and, under the assumption that the demand variable is positive, fulfils the condition $FR(S) \leq FR^1(S)$ (see Chen et al. 2003). Thus, a calibration of a logistics system utilizing FR^1 will require solving the equation:

$$1 - E\frac{(Y - S)^+}{Y} = \tau. \tag{10}$$

It is worth noting that, similar to the case of the type I service level, we can specify forecast loss functions whose expected values are uniquely minimized at the statistical functionals of the distributions of demand which solve the Eqs. (7) and (10). In other words, these statistical functionals are elicitable in the sense of Gneiting (2011b). This is explicitly stated in the following two lemmas formulated with the help of the so-called identification function approach of Osband (1985) (for details, see Gneiting 2011b and the references therein).

Lemma 1 *Let F be a class of absolutely continuous distributions with a positive mean value $0 < EY < \infty$, and let $T : F \rightarrow R^+$ be the functional defined according to the infinite period Fill Rate requirement of the form:*

$$1 - \frac{E[Y - T(Y)]^+}{EY} = \tau \tag{11}$$

for a certain service level $\tau \in (0, \ 1)$. Then the loss function $L : R \times R^+ \rightarrow R$:

$$L(y, \ \hat{y}) = \begin{cases} \hat{y}\varphi'(\hat{y}) - \varphi(\hat{y}) - \tau\varphi'(\hat{y})y + a(y) & \text{for } y \geq \hat{y}, \\ (1 - \tau)\varphi'(\hat{y})y - \varphi(y) + a(y) & \text{for } y < \hat{y} \end{cases}, \tag{12}$$

where a is F-integrable and φ is two times differentiable and strictly convex with derivative φ' and such that $\varphi \mathbf{1}_{(-\infty, \, x]}$ is F-integrable for every $x \in R$, is strictly consistent for T.

Proof Let $S = T(Y)$ for a certain distribution in F, i.e.,

$$(1 - \tau)EY = \int\limits_{S}^{\infty} (y - S)\, dF(y), \quad F \in F,$$

and let $f(y)$ be the probability density function of Y. In order to check that $EL(Y, \hat{y})$ is uniquely minimized at S, let us consider:

$$EL(Y, \hat{y}) = \left[\hat{y}\varphi'(\hat{y}) - \varphi(\hat{y})\right] P(Y \geq \hat{y}) - \tau\varphi'(\hat{y}) \int\limits_{\hat{y}}^{\infty} yf(y)dy + (1 - \tau)\varphi'(\hat{y}) \int\limits_{0}^{\hat{y}} yf(y)dy$$

$$-\int_{-\infty}^{\hat{y}} \varphi(y)f(y)dy + Ea(Y) = \left[\hat{y}\varphi'(\hat{y}) - \varphi(\hat{y})\right] - \left[\hat{y}\varphi'(\hat{y}) - \varphi(\hat{y})\right]\int_{-\infty}^{\hat{y}} f(y)dy$$

$$- \tau\varphi'(\hat{y})EY + \varphi'(\hat{y})\int_{-\infty}^{\hat{y}} yf(y)dy - \int_{-\infty}^{\hat{y}} \varphi(y)f(y)dy + Ea(Y).$$

Taking the derivative with respect to $\hat{y}$, we have:

$$\frac{\partial EL(Y,\hat{y})}{\partial \hat{y}} = \hat{y}\varphi''(\hat{y}) - \hat{y}\varphi''(\hat{y})\int_{-\infty}^{\hat{y}} f(y)dy - \tau\varphi''(\hat{y})EY + \varphi''(\hat{y})\int_{-\infty}^{\hat{y}} yf(y)dy.$$

The first-order condition is then as follows:

$$\varphi''(\hat{y})\left[\hat{y} - \hat{y}\int_{-\infty}^{\hat{y}} f(y)dy - \tau EY + \int_{-\infty}^{\hat{y}} yf(y)dy\right] = 0$$

or, equivalently,

$$\varphi''(\hat{y})\left[(1-\tau)EY - \int_{\hat{y}}^{\infty} (y-\hat{y})f(y)dy\right] = 0. \tag{13}$$

The thesis follows since φ is strictly convex and the expression in square brackets in (13) is negative for $\hat{y} < S$ and positive for $\hat{y} > S$. □

Lemma 2 *Let F be a class of absolutely continuous distributions on the positive half-axis such that $E\left(\frac{1}{Y}\right)$ exists, and let $T : \mathrm{F} \to R^+$ be the functional defined according to the single period Fill Rate requirement of the form:*

$$1 - E\left\{\frac{[Y - T(Y)]^+}{Y}\right\} = \tau \tag{14}$$

for a certain service level $\tau \in (0, 1)$. Then the loss function $L : R^+ \times R^+ \to R$:

$$L(y, \hat{y}) = \begin{cases} \frac{\hat{y}\varphi'(\hat{y}) - \varphi(\hat{y})}{y} - \tau\varphi'(\hat{y}) + a(y) \text{ for } y \geq \hat{y} \\ \frac{y\varphi'(\hat{y}) - \varphi(y)}{y} - \tau\varphi'(\hat{y}) + a(y) \text{ for } y < \hat{y} \end{cases}, \tag{15}$$

where a is F-integrable and φ is two times differentiable and strictly convex with derivative φ' and such that $E\left(\frac{\varphi(Y)}{Y}\mathbf{1}_{\{Y \leq x\}}\right)$ exists for every $x \in R^+$, is strictly consistent for T.

Proof Let $S = T(Y)$, i.e.,

$$1 - \tau = \int\limits_{S}^{\infty} \frac{y - S}{y} dF(y), \quad \text{for a certain } F \in F,$$

and let $f(y)$ be the probability density function of Y. As previously, let us consider the expected value of the examined loss function:

$$EL(Y, \hat{y}) = \left[\hat{y}\varphi'(\hat{y}) - \varphi(\hat{y})\right] \int\limits_{\hat{y}}^{\infty} \frac{1}{y} f(y) dy + \int\limits_{-\infty}^{\hat{y}} \left[\varphi'(\hat{y}) - \frac{\varphi(y)}{y}\right] f(y) dy - \tau \varphi'(\hat{y}) + Ea(Y)$$

$$= \left[\hat{y}\varphi'(\hat{y}) - \varphi(\hat{y})\right] E\left(\tfrac{1}{Y}\right) - \left[\hat{y}\varphi'(\hat{y}) - \varphi(\hat{y})\right] \int\limits_{0}^{\hat{y}} \frac{1}{y} f(y) dy + \varphi'(\hat{y}) \int\limits_{0}^{\hat{y}} f(y) dy$$

$$- \int\limits_{-\infty}^{\hat{y}} \frac{\varphi(y)}{y} f(y) dy - \tau \varphi'(\hat{y}) + Ea(Y).$$

Then, differentiating with respect to $\hat{y}$, we arrive at:

$$\frac{\partial EL(Y, \hat{y})}{\partial \hat{y}} = \hat{y}\varphi''(\hat{y}) E\tfrac{1}{Y} - \hat{y}\varphi''(\hat{y}) \int\limits_{0}^{\hat{y}} \frac{1}{y} f(y) dy + \varphi''(\hat{y}) \int\limits_{0}^{\hat{y}} f(y) dy - \tau \varphi''(\hat{y})$$

$$= \hat{y}\varphi''(\hat{y}) \int\limits_{\hat{y}}^{\infty} \frac{1}{y} f(y) dy - \varphi''(\hat{y}) \int\limits_{\hat{y}}^{\infty} f(y) dy + (1 - \tau)\varphi''(\hat{y}).$$

Thus, the first-order condition is the following:

$$\varphi''(\hat{y}) \left[1 - \tau - \int\limits_{\hat{y}}^{\infty} \frac{y - \hat{y}}{y} f(y) dy \right] = 0. \tag{16}$$

As previously, the thesis follows since φ is strictly convex and the expression in square brackets in (16) is negative for $\hat{y} < S$ and positive for $\hat{y} > S$. $\qquad\square$

Thus, for a large class of distributions fulfilling certain moment conditions, the functionals defined according to (7) and (10) are elicitable in the sense of Gneiting (2011b). If a statistical functional of interest to the decision-maker is elicitable, the loss function associated with it can be utilized at both the computation and evaluation steps of forecasting procedures in a way similar to the LINLIN loss and the MAPE (Mean Absolute Percentage Error) accuracy measure (see Weiss and Andersen 1984; Weiss 1996; Engle and Manganelli 2004; and Bruzda 2016, 2018) or the loss functions discussed recently in the context of the estimation and forecasting of financial risk (see

Fissler and Ziegel 2016; Dimitriadis and Bayer 2017; Maume-Deschamps et al. 2017; Patton et al. 2017), even if such loss functions are not of the 'prediction error' type.[4]

Example loss functions (12) and (15) are obtained if one takes $\phi(y) = y^2$ and $a(y) = \tau y^2$ in (12), and $\phi(y) = \frac{1}{2} y^2$ and $a(y) = \frac{1}{2} y$ in (15), which leads, respectively, to:

$$L(y, \hat{y}) = \begin{cases} \tau(y - \hat{y})^2 + (1 - \tau)\hat{y}^2 & \text{for } y \geq \hat{y} \\ -(1 - \tau)(y - \hat{y})^2 + (1 - \tau)\hat{y}^2 & \text{for } y < \hat{y} \end{cases}, \tag{17}$$

and

$$L(y, \hat{y}) = \begin{cases} \frac{(y - \hat{y})^2}{2y} + (1 - \tau)\hat{y} & \text{for } y \geq \hat{y} \\ (1 - \tau)\hat{y} & \text{for } y < \hat{y} \end{cases}. \tag{18}$$

Loss functions such as (17) and (18) can be interpreted, for example, in terms of costs associated with maintaining a firm's production and/or distribution potential.

In the next section we concentrate on supply chain and logistics forecasting under the LINLIN loss and discuss the many ways of computing multiperiod forecasts of this type. We also present our evaluation framework encompassing tests of quantile unbiasedness and different measures of accuracy for quantile forecasts. Similar considerations under loss functions such as (17) and (18) will constitute an interesting topic for further studies.

3 Quantile forecasts of demand

It is worth emphasizing that in the practice of supply chain management there is a need to compute two sorts of multiperiod quantile forecasts of demand. On the one hand, a decision-maker may be interested in quantiles of the predictive distribution of demand in some future period $n+h$, i.e., in quantile forecasts of the variable D_{n+h}, while, on the other, there is a need to forecast quantiles of sums of the variables D_{n+i} in the whole forecast horizon, i.e., quantiles of the variable $\sum_{i=1}^{h} D_{n+i}$. The first sort of multiperiod quantile forecasts can be utilized in early warning systems signaling changes on the market and possible threads (compare De Nicolò and Lucchetta 2017), in medium- or long-term (production and distribution) capacity planning, in production planning if products are not stocked (for example, in the food sector) and will also be used

---

[4] Properties of optimal forecasts with loss functions of the 'prediction error' type, i.e., depending solely on the forecast error, have been established by Patton and Timmermann (2007). In particular, in the simple case of the conditional mean dynamics, such loss functions lead to optimal forecasts of the form 'the conditional mean + a correction term', where the correction term depends solely on the distribution of innovations and the loss function (also see Granger 1969; Christoffersen and Diebold 1997). For an introduction to the decision-theoretic forecasting in a more general case, see Clements (2005, Chap. VI); Granger (1969, 1999), Granger and Machina (2006), Patton and Timmermann (2010) and Gneiting (2011b).

in the case of delays in information flows required for operational decisions. The second type of quantile forecasts is required in inventory and production management systems operating under the CSL approach and, as was already mentioned, is often computed on the basis of the approximation (2), although exact solutions are also considered under prespecified data generating mechanisms of the ARMA type, for example, AR(1), MA(1) and ARMA(1,1) in Ali et al. (2012), and ARIMA(0,1,1) in Babai et al. (2013). Our further discussion is mainly focused on the second type of quantile forecasts, which are also exclusively considered later in the empirical part of this paper.

In the general linear case of (possibly nonstationary) ARMA processes, the exact forecasting solution is directly derived as follows. Let us consider an ARMA(p, q) process of the form:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \cdots + \phi_q \varepsilon_{t-q}, \tag{19}$$

where $\varepsilon_t \sim i.i.d.$. The process (19) can be rewritten in the equivalent form of an infinite moving average:

$$y_t = m + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-1} + \cdots. \tag{20}$$

Then, the forecasts of the sums of the process in the forecast horizon h are given as:

$$E\left(\sum_{i=1}^{h} y_{t+i} | \Omega_t\right) = hm + (\theta_1 + \theta_2 + \cdots + \theta_h)\varepsilon_t + (\theta_2 + \theta_3 + \cdots + \theta_{h+1})\varepsilon_{t-1} + \cdots, \tag{21}$$

where Ω_t is the set of information available in period t, and the forecast error is expressed as:

$$\eta_t(h) = \sum_{i=1}^{h} y_{t+i} - E\left(\sum_{i=1}^{h} y_{t+i} | \Omega_t\right) = (1 + \theta_1 + \ldots + \theta_{h-1})\varepsilon_{t+1}$$
$$+ (1 + \theta_1 + \cdots + \theta_{h-2})\varepsilon_{t+2} + \cdots + \varepsilon_{t+h}. \tag{22}$$

Thus, the variance of the forecast error is the following:

$$Var(\eta_t(h)) = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \left(\sum_{l=0}^{i} \theta_l\right)^2, \tag{23}$$

and quantile forecasts of $\sum_{i=1}^{h} y_{t+i}$ are obtained as follows:

$$Q^\tau\left(\sum_{i=1}^{h} y_{t+i} | \Omega_t\right) = E\left(\sum_{i=1}^{h} y_{t+i} | \Omega_t\right) + \sqrt{Var(\eta_t(h))}\Phi_\varepsilon^{-1}(\tau), \tag{24}$$

where Φ is the standardized conditional distribution function of demand.

In the case of ARMA processes with some sort of conditional heteroscedasticity,[5] the conditional variance of either y_{n+h} or $\sum_{i=1}^{h} y_{n+i}$ can be derived directly from its definition. For example, for y_{n+h} we have:

$$
\begin{aligned}
Var(y_{t+h}|\Omega_t) &= E(y_{t+h}^2|\Omega_t) - [E(y_{t+h}|\Omega_t)]^2 \\
&= E[(\varepsilon_{t+h} + \theta_1\varepsilon_{t+h-1} + \cdots + \theta_{h-1}\varepsilon_{t+1})^2|\Omega_t] \\
&= \sum_{i=0}^{h-1} \theta_i^2 E(\varepsilon_{t+h-i}^2|\Omega_t).
\end{aligned}
\tag{25}
$$

However, when computing multiperiod quantile forecasts under conditional heteroscedasticity, one should resort to simulations since in such a case the conditional distribution becomes a nonlinear combination of the error terms ε_t, rendering the use of, for example, quantiles of the standard normal distribution, z_τ, inappropriate. This, together with the need to account for estimation errors, sets simulation techniques in the centre of our interest when computing multiperiod forecasts of demand for inventory management applications.

The basis for our forecasting procedures in the empirical part of the paper constitute autoregressive models with polynomial trend functions, which is justified taking into account the results of unit root tests on the data used in this study, indicating (trend-)stationarity in the majority of cases. The maximum autoregression order is set to $k = 6$ and the maximum degree of polynomial trend functions is set to $p = 2$. The AICC criterion in its versions for the ML and quantile regression estimations is used in model selections. In the case of the direct approach to forecasting, the models are always chosen separately for each forecast horizon, as are also the quantile regression equations for different quantiles, utilized here either in the direct quantile forecasting or in a quantile regression based simulation approach. Since the analyzed data are trending, the approximation (2) has been excluded from the comparison.

The following forecasting procedures are evaluated:

1–2   The least squares estimation of models of the form:

$$
y_t = \sum_{j=0}^{q} \alpha_j t^j + \beta_1 y_{t-1} + \cdots + \beta_m y_{t-m} + \varepsilon_t
\tag{26}
$$

$(q \le p = 2,\ m \le k = 6)$ with quantile forecasts computed according to the formula (24), where quantiles of the standardized conditional distribution are obtained nonparametrically (1) or are taken from the standard normal distribution (2).[6]

---

[5] A conditional heteroscedasticity in demand is usually dealt with smoothing the mean absolute deviation (MAD) and smoothing the mean squared error (MSE) as in, for example, Strijbosch and Moors (2005) and Syntetos et al. (2010a, b), but GARCH models have also been suggested and discussed in this context (Datta et al. 2007, 2009).

[6] In the nonparametric quantile estimation in this paper, we rely on the Matlab function 'quantile'.

3–4   The least squares estimation of the 'direct' (MIDAS-type) equation:

$$D_t^T = \sum_{i=0}^{T-1} y_{t+i} = \sum_{j=0}^{q} \alpha_j t^j + \beta_1 y_{t-1} + \cdots + \beta_m y_{t-m} + \eta_t \quad (27)$$

   ($q \le p = 2$, $m \le k = 6$) with quantile correction computed nonparametrically (3) or assuming the normal distribution (4).

5.   A direct quantile regression approach, in which the quantile regression estimator (see Koenker and Bassett 1978; Koenker 2005) is applied to estimate the (27) and the quantile forecasts are computed in one step.

6.   A simulation approach utilizing the quantile regression estimation of the (26) for different quantile orders τ (compare the method of forecasting the Quantile Autoregressive—QAR—processes in Koenker 2011), specified here separately for each quantile order; in the first step, 1000 quantile regressions are estimated for quantile orders being equally spaced values between 0 and 1; the estimated equations are directly used to obtain the one-step ahead predictive distribution of y_t; then, random sequences of quantile orders from those used in the first step are generated, the one-step ahead quantile forecasts are substituted for actuals in the quantile regressions chosen according to the random sequences of quantile orders to produce the two-step ahead predictive distribution, and the computations are continued in this way to obtain the h-step ahead predictive distribution; finally, the distributions are cumulated and the forecasts of $\sum_i y_{t+i}$ are computed with the Matlab function 'quantile'.

7–8   The basic bootstrap method as presented in Clements (2005), §4.2.3, based on the percentile method of Efron and the approach suggested in Pascual et al. (2001) (see the formula (4.9) in Clements 2005), performed on the (26) without bias correction; the nonparametric (7) or parametric (8) bootstrap is used, i.e., to generate the bootstrap samples, either resampled residuals are utilized or innovations are generated assuming the normal distribution;[7] the predictive distributions of y_{t+i} are cumulated and the final forecasts are computed; the number of bootstrap samples is set to 1000.

9–10   The model (26) is estimated jointly with the GARCH(1,1) equation for the conditional variance assuming the conditional normal distribution; in the estimation, the Matlab functions 'garchfit' and 'garchget' from the Econometrics Toolbox are used; the predictive distributions of y_{t+i} are generated either by resampling standardized residuals (9) or through random draws from the standard normal distribution (10); the predictive distributions are then cumulated and the final forecasts are obtained; the estimation errors are not accounted for;[8] 1000 replications.

---

[7] Under the Gaussianity assumption, it is potentially possible to obtain an analytic solution accounting for estimation errors, but such a solution would be only approximate and rather tedious to derive.

[8] Attempts to apply bootstrap methods to the AR(m)-GARCH(1,1) models with trend functions estimated on the data from the M3 competition were unsuccessful, leading to much worse results as compared with the simulation approach. Also, our experimentation with simulations based on linear ARCH (LARCH)

1L–2L   The model (26) is estimated on logarithms of the data, and either resampled residuals (1L) or random draws from the normal distribution (2L) are used to generate the predictive distributions for logarithms of y_{t+i}; the logarithmic transformation is then reversed, and the distributions are cumulated; 1000 replications.

6L–10L   As 6–10 but all estimations are performed on logarithms and the final forecasts are computed as in 1L–2L.

Our evaluation of the forecast accuracy is based on the following aggregate measures introduced in Bruzda (2018):

- Relative LINLIN costs:

$$LINLIN^{*\tau} = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}\frac{\{\tau u_{it}^{+} + (1-\tau)u_{it}^{-}\}}{\hat{y}_{it}^{\tau}} = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}LINLIN_{it}^{*\tau},$$

(28)

$$LINLIN^{**\tau} = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}\frac{\{\tau u_{it}^{+} + (1-\tau)u_{it}^{-}\}}{y_{it}} = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}LINLIN_{it}^{**\tau},$$

(29)

- Hit Ratio:

$$Hit^{\tau} = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}I_{\{y_{it}-\hat{y}_{it}^{\tau}>0\}},$$

(30)

- Fill Rate:

$$FR^{\tau} = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}\left\{1 - \frac{(y_{it} - \hat{y}_{it}^{\tau})^{+}}{y_{it}}\right\},$$

(31)

where $\hat{y}_{it}^{\tau}$ and y_{it} are, respectively, the tth quantile forecast for quantile of order τ for the ith time series and the corresponding realization, u_{it}^{+} and u_{it}^{-} are the absolute values of the positive and negative forecast errors, K denotes the number of series, T is the number of (fixed horizon) forecasts for a single time series, and I_A is the indicator function for the set A.

It is worth underlining that the aggregate measures (28)–(31) may be interesting not only as a statistical approach to the evaluation of quantile forecasts, but they can also be interpreted in economic terms. In particular, (28) and (29) may be thought of as proportional to supply chain costs per unit of output and per unit of sales, respectively,

Footnote 8 continued
models estimated as in Francq and Zakoïan (2010), which we considered to be close competitors to the simulations utilizing quantile regression estimation, turned out to provide inferior results compared to the other methods when evaluated on our dataset.

if the unit costs of over- and understocking are equal across different products and over time or, alternatively, to supply chain costs per the (Dollar or Euro) values of output and sales if the unit costs of over- and understocking are proportional to the unit prices of products (see Bruzda 2018).

Furthermore, we also examine average ranks of our forecasting procedures in rankings based on the LINLIN loss:

$$LINLIN_{it}^{\tau} = \tau u_{it}^{+} + (1 - \tau) u_{it}^{-} \tag{32}$$

and perform the Kupiec (1995) test for quantile unbiasedness, which is based on the following test statistic:

$$LR = 2 \left\{ \ln \left[\left(\frac{N}{K \cdot T} \right)^{N} \left(1 - \frac{N}{K \cdot T} \right)^{K \cdot T - N} \right] - \ln \left[(1 - \tau)^{N} \tau^{K \cdot T - N} \right] \right\}, \tag{33}$$

where N denotes the number of hits (violations) and $K \cdot T$ is the total number of forecasts for a given quantile. Under the hypothesis that the Hit Ratio is equal to the nominal value $1 - \tau$, the statistic (33) has asymptotically the $\chi^2(1)$ distribution. The forecast evaluation presented in the next section was executed after replacing all negative demand forecasts with zeros.

4 Empirical results and discussion

The dataset used in this study comprises the 474 monthly microeconomic time series from the M3 forecast competition (see Makridakis and Hibon 2000), which are described as 'shipments'.[9] The time series are of different lengths—the first 277 of them are of length 68–69, while the last 197—of length 126. The ratio-to-moving average method with seasonal factors recomputed on the whole range of observations was applied to seasonally adjust the data if the given time series was treated as seasonal in the exponential smoothing procedures in the M3 competition.[10]

The maximum forecast horizon h was set to 12, and all forecasting models were built on data shortened by 12 observations in order to facilitate a forecast evaluation. Since one forecast for each forecast horizon was calculated, in computing the aggregate accuracy measures we set $T = 1$. The following quantile orders τ were examined: 0.5; 0.75; 0.9; 0.95; 0.975 and 0.99, while the detailed statistics are presented mainly for 0.75 and 0.95.[11] Our presentation of the empirical outcomes is given separately for the two subsets of time series. It is worth underlining that the relative LINLIN costs and the Fill Rate statistics were also computed in trimmed (robustified) versions excluding 5 (10, 20)% of the forecasts (compare Bruzda 2018).

Turning to the analysis of our empirical results, first we note that, for the short time series of length 68–69 and the relative LINLIN costs (28) and (29), the dominant

---

[9] See http://www.forecastingprinciples.com/index.php/data.

[10] See https://forecasters.org/resources/time-series-data/m3-competition/.

[11] The other empirical results are available upon request.

solution turns out to be the procedure 6L, making use of random draws of quantile orders and quantile regression estimation performed on logarithms of the data (see Tables 1 and 2 and, for the cost (29), also compare the first two panels in Fig. 1). Only in the case of one- or two-step ahead quantile forecasts (and up to four-step ahead for $\tau = 0.99$) this procedure is outperformed by some other methods—most of all the bootstrap approach and, particularly for one-step ahead forecasting, the simulations based on GARCH models and the normal distribution (the procedures 10 and 10L). The dominance of the procedure 6L (and also the GARCH models one step ahead) over the other methods becomes even more visible if the robustified (trimmed) versions of the costs (28) and (29) are examined (not reported). The bootstrap procedures with ols estimation on logarithms or, alternatively, levels of the data appear to be the second best solution.

It is interesting to note that the simulation based procedures lead to large reductions in the total costs of logistics systems as compared with non-simulation based methods. In particular, in the case of forecasting 12 months ahead and $\tau = 0.95$, the quantile regression based simulation approach reduces the average supply chain cost per unit of output by about 70% (see Table 1). For $\tau \in \{0.975; 0.99\}$ these reductions are even larger and for $\tau = 0.99$ reach about 85%. This indicates that the choice of a forecasting procedure for production and inventory planning can have a substantial impact on the total costs of a logistics system.

On the other hand, in the case of the longer data, for which the mean relative costs are reported in Tables 3 and 4 (also compare the last two panels in Fig. 1), we observe that the simulations utilizing quantile regression (the procedures 6 and 6L) are able to outperform all the other methods usually exclusively for the lower-order quantiles among those considered and at shorter horizons (no longer than 8–9 months in the case of $\tau \in \{0.5; 0.75\}$, and exclusively for two- and three-step ahead forecasts for $\tau = 0.9$). For longer forecast horizons, the dominant solution becomes the bootstrap method, most often when performed on levels of the data and in its parametric version (the method 8). For example, the method 8 produced the best forecasts on average at horizons in the range 7–12 for $\tau = 0.9$ (not reported). Thus, longer-term quantile forecasting in this dataset can be based on linear processes and the normal distribution, although accounting for estimation errors seems then to be crucial taking into account large discrepancies between the values of forecast accuracy measures for the different approaches (see Tables 3, 4). For example, for $h = 12$ and $\tau = 0.95$, the methods 8 and 9 lead to about 60% reductions in the average supply chain costs per unit of output (see Table 3). In general, these reductions are somewhat smaller than those observed for the first dataset, and, furthermore, they also tend to differ to a large extent among the different procedures.

Besides, it is worth noticing that the GARCH models turn out to be very useful in short horizon forecasting (up to three steps ahead), particularly for large quantiles, more often producing the best outcomes than in the case of the shorter time series in the first dataset. For example, for $h = 1$ and $\tau = 0.95$, the GARCH based procedure 10 reduces the average cost (28) by about 17% (or more) as compared with the non-GARCH based methods (see Table 3). On the contrary, for h lying in the middle of the range examined here, the linear methods 1 and 2 (especially 1) often outperform the

Table 1 The aggregate LINLIN cost (28) for the first dataset

Quantile order τ	0.75						0.95					
Method/forecast horizon h	1	2	3	6	9	12	1	2	3	6	9	12
Levels												
1	0.0710	0.0658	0.0654	0.0691	0.0837	0.1021	0.0196	0.0195	0.0192	0.0271	0.0377	0.0493
2	0.0701	0.0638	0.0623	0.0666	0.0800	0.0969	0.0198	0.0204	0.0202	0.0283	0.0397	0.0509
3	0.0710	Inf	Inf	0.1148	Inf	Inf	0.0196	0.0222	0.0270	0.0579	0.2129	Inf
4	0.0701	0.0839	Inf	0.1069	Inf	Inf	0.0198	0.0224	0.0257	0.0567	0.1839	Inf
5	0.0714	0.0815	0.0827	0.1150	Inf	Inf	0.0441	0.0556	Inf	0.1521	Inf	Inf
6	0.0726	0.0654	0.0629	0.0618	0.0665	0.0709	0.0239	0.0251	0.0244	0.0312	0.0373	0.0406
7	0.0721	0.0633	0.0588	0.0519	0.0507	0.0468	0.0197	0.0186	0.0176	0.0169	0.0165	0.0145
8	0.0717	0.0628	0.0580	0.0518	0.0509	0.0467	0.0194	0.0190	0.0177	0.0173	0.0173	0.0147
9	0.0710	0.0653	0.0641	0.0677	0.0796	0.0937	**0.0193**	0.0198	0.0201	0.0279	0.0389	0.0493
10	0.0699	0.0643	0.0634	0.0676	0.0791	0.0937	0.0203	0.0210	0.0217	0.0302	0.0412	0.0511
Logarithms												
1L	0.0740	0.0661	0.064	0.0650	0.0708	0.0731	0.0239	0.0235	0.0230	0.0314	0.0396	0.0435
2L	0.0728	0.0666	0.0635	0.0639	0.0692	0.0711	0.0224	0.0230	0.0221	0.0284	0.0360	0.0396
6L	**0.0679**	**0.0581**	**0.0528**	**0.0455**	**0.0447**	**0.0419**	0.0194	**0.0184**	**0.0162**	**0.0147**	**0.0159**	**0.0137**
7L	0.0739	0.0642	0.0602	0.0527	0.0510	0.0462	0.0228	0.0216	0.0196	0.0191	0.0186	0.0159
8L	0.0738	0.0639	0.0589	0.0520	0.0495	0.0456	0.0220	0.0212	0.0188	0.0173	0.0171	0.0150
9L	0.0743	0.0665	0.0643	0.0649	0.0704	0.0723	0.0227	0.0243	0.0249	0.0334	0.0409	0.0447
10L	0.0726	0.0666	0.0636	0.0643	0.0695	0.0712	0.0227	0.0242	0.0243	0.0324	0.0388	0.0417

The best results are given in bold. 'Inf' means that some quantile forecasts were negative and are replaced with zeros

Table 2 The aggregate LINLIN cost (29) for the first dataset

Quantile order τ	0.75						0.95					
Method/forecast horizon h	1	2	3	6	9	12	1	2	3	6	9	12
Levels												
1	0.1160	0.0810	0.0716	0.0667	0.0743	0.0799	0.0420	0.0290	0.0250	0.0278	0.0344	0.0408
2	0.1206	0.0823	0.0710	0.0662	0.0732	0.0791	0.0409	0.0293	0.0255	0.0284	0.0355	0.0411
3	0.1160	0.0826	0.0741	0.0852	0.1085	0.1281	0.0420	0.0303	0.0299	0.0476	0.0801	0.1091
4	0.1206	0.0829	0.0739	0.0845	0.1076	0.1280	0.0409	0.0298	0.0283	0.0466	0.0796	0.1082
5	0.1132	0.0852	0.0774	0.0893	0.1074	0.1231	0.0528	0.0484	0.0567	0.0848	0.0969	0.1141
6	0.1138	0.0769	0.0676	0.0598	0.0617	0.0638	0.0433	0.0314	0.0276	0.0292	0.0321	0.0335
7	0.1169	0.0819	0.0694	0.0580	0.0573	0.0550	0.0426	0.0295	0.0248	0.0209	0.0203	0.0184
8	0.1226	0.0839	0.0700	0.0583	0.0578	0.0551	0.0413	0.0292	0.0246	0.0212	0.0208	0.0186
9	0.1146	0.0799	0.0707	0.0663	0.0721	0.0774	0.0408	0.0290	0.0256	0.0283	0.0352	0.0400
10	0.1177	0.0804	0.0708	0.0662	0.0718	0.0775	**0.0397**	0.0289	0.0263	0.0297	0.0362	0.0410
Logarithms												
1L	0.1149	0.0790	0.0699	0.0646	0.0679	0.0696	0.0432	0.0312	0.0274	0.0303	0.0354	0.0377
2L	0.1116	0.0810	0.0712	0.0649	0.0680	0.0692	0.0466	0.0335	0.0283	0.0291	0.0336	0.0355
6L	0.1126	**0.0756**	**0.0632**	**0.0512**	**0.0501**	**0.0481**	0.0417	**0.0280**	**0.0221**	**0.0180**	**0.0185**	**0.0165**
7L	0.1138	0.0784	0.0680	0.0571	0.0556	0.0521	0.0433	0.0307	0.0257	0.0220	0.0211	0.0187
8L	0.1121	0.0794	0.0683	0.0578	0.0555	0.0527	0.0463	0.0327	0.0265	0.0216	0.0208	0.0188
9L	0.1145	0.0788	0.0693	0.0639	0.0671	0.0686	0.0419	0.0314	0.0287	0.0318	0.0364	0.0385
10L	**0.1100**	0.0792	0.0695	0.0642	0.0673	0.0687	0.0448	0.0332	0.0294	0.0322	0.0357	0.0369

The best results are given in bold

Table 3 The aggregate LINLIN cost (28) for the second dataset

Quantile order τ	0.75						0.95					
Method/forecast horizon h	1	2	3	6	9	12	1	2	3	6	9	12
Levels												
1	0.0495	0.0416	0.0451	0.0753	0.0896	0.1024	0.0162	0.0126	0.0118	**0.0152**	0.0239	0.0305
2	0.0510	0.0378	0.0388	0.0594	0.0726	0.0826	0.0157	0.0122	0.0116	0.0165	0.0261	0.0329
3	0.0495	0.0380	0.0414	0.0653	0.0835	0.1032	0.0162	0.0128	0.0120	0.0175	0.0331	0.0505
4	0.0510	0.0371	0.0396	0.0602	0.0811	0.1065	0.0157	0.0122	0.0116	0.0180	0.0338	0.0514
5	Inf	0.0439	0.0415	0.0592	0.0715	0.0925	0.0169	0.0133	0.0148	0.0484	0.0525	Inf
6	0.0522	**0.0369**	**0.0355**	0.0526	0.0562	0.0574	0.0154	0.0104	0.0116	0.0210	0.0277	0.0316
7	0.0500	0.0403	0.0452	0.0585	0.0497	0.0445	0.0158	0.0115	0.0134	0.0200	**0.0156**	**0.0137**
8	0.0499	0.0385	0.0426	0.0555	**0.0491**	0.0448	0.0154	0.0114	0.0133	0.0201	**0.0156**	**0.0137**
9	0.0491	0.0423	0.0460	0.0746	0.0817	0.0890	**0.0128**	0.0102	**0.0112**	0.0222	0.0319	0.0364
10	**0.0472**	0.0408	0.0446	0.0710	0.0791	0.0855	0.0132	**0.0100**	0.0119	0.0254	0.0356	0.0418
Logarithms												
1L	0.0539	0.0441	0.0445	0.0640	0.0618	0.0592	0.0156	0.0118	0.0136	0.0299	0.0350	0.0362
2L	0.0542	0.0434	0.0433	0.0622	0.0608	0.0583	0.0154	0.0113	0.0129	0.0281	0.0332	0.0340
6L	0.0508	**0.0369**	0.0372	**0.0519**	0.0500	0.0481	0.0156	0.0105	0.0125	0.0197	0.0223	0.0231
7L	0.0581	0.0479	0.0523	0.0648	0.0510	0.0442	0.0160	0.0120	0.0174	0.0318	0.0247	0.0198
8L	0.0579	0.0473	0.0506	0.0630	0.0498	**0.0434**	0.0157	0.0114	0.0161	0.0294	0.0231	0.0184
9L	0.0531	0.0433	0.0452	0.0665	0.0650	0.0629	0.0153	0.0128	0.0141	0.0279	0.0350	0.0379
10L	0.0532	0.0428	0.0442	0.0647	0.0644	0.0627	0.0146	0.0119	0.0133	0.0264	0.0333	0.0357

The best results are given in bold. 'Inf' means that some quantile forecasts were negative and are replaced with zeros

Table 4 The aggregate LINLIN cost (29) for the second dataset

| Quantile order τ	0.75						0.95					
Method/forecast horizon h	1	2	3	6	9	12	1	2	3	6	9	12
Levels												
1	0.0608	0.0444	0.0454	0.0613	0.0641	0.0663	0.0310	0.0201	0.0179	**0.0183**	0.0226	0.0265
2	0.0704	0.0453	0.0443	0.0549	0.0594	0.0627	0.0286	0.0187	0.0168	0.0186	0.0238	0.0279
3	0.0608	0.0420	0.0442	0.0585	0.0654	0.0707	0.0310	0.0200	0.0182	0.0201	0.0298	0.0387
4	0.0704	0.0446	0.0457	0.0556	0.0646	0.0710	0.0286	0.0186	0.0170	0.0197	0.0289	0.0385
5	0.0628	0.0468	0.0457	0.0569	0.0597	0.0670	0.0281	0.0181	0.0184	0.0406	0.0440	0.0583
6	0.0600	0.0404	**0.0381**	0.0503	0.0503	0.0495	0.0274	0.0145	0.0149	0.0214	0.0247	0.0266
7	0.0612	0.0422	0.0447	0.0532	0.0464	0.0426	0.0304	0.0179	0.0179	0.0214	**0.0171**	**0.0154**
8	0.0687	0.0429	0.0439	0.0521	**0.0463**	0.0428	0.0285	0.0171	0.0174	0.0214	**0.0171**	**0.0154**
9	0.0581	0.0431	0.0446	0.0614	0.0624	0.0642	0.0230	0.0144	**0.0143**	0.0215	0.0279	0.0308
10	0.0598	0.0431	0.0444	0.0598	0.0613	0.0631	**0.0214**	**0.0137**	0.0144	0.0238	0.0296	0.0338
Logarithms												
1L	0.0595	0.0451	0.0445	0.0578	0.0549	0.0527	0.0237	0.0152	0.0159	0.0278	0.0305	0.0315
2L	0.0604	0.0451	0.0439	0.0566	0.0543	0.0523	0.0239	0.0150	0.0154	0.0266	0.0292	0.0300
6L	**0.0579**	**0.0395**	0.0391	**0.0499**	0.0473	0.0453	0.0248	0.0153	0.0163	0.0204	0.0219	0.0224
7L	0.0629	0.0476	0.0500	0.0587	0.0479	0.0427	0.0258	0.0164	0.0202	0.0308	0.0248	0.0210
8L	0.0629	0.0473	0.0489	0.0575	0.0472	**0.0424**	0.0255	0.0160	0.0192	0.0291	0.0238	0.0198
9L	0.0598	0.0440	0.0446	0.0594	0.0571	0.0556	0.0242	0.0165	0.0168	0.0268	0.0311	0.0335
10L	0.0594	0.0440	0.0439	0.0581	0.0568	0.0556	0.0235	0.0160	0.0163	0.0261	0.0301	0.0319

The best results are given in bold
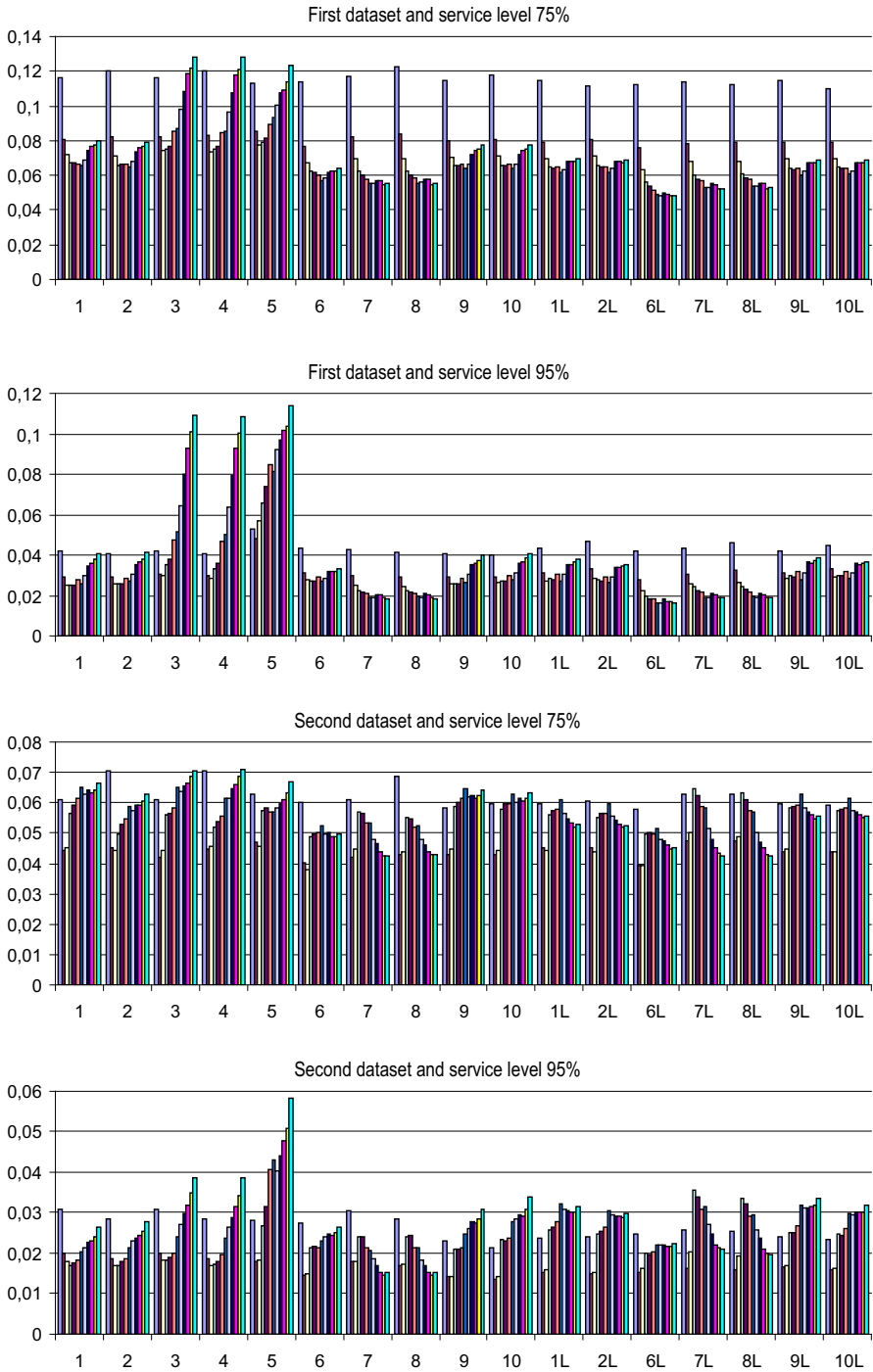
Fig. 1 The aggregate LINLIN cost (29) for forecast horizons in the range 1–12 months. *Note* compare Tables 2 and 4

other approaches. All of this is also confirmed to a large extent when the robustified accuracy measures are analyzed.

It is worth noting that the direct ols or quantile regression based forecasting approaches (the methods 3–5) often provide inferior solutions relative to the iterated ols based methods or the simulations utilizing quantile regression, in particular in the case of short time series, longer forecast horizons and large quantiles. Interestingly, we also observe that generating predictive distributions with a large number of quantile regressions estimated on levels (the method 6) can—and, in fact, often does—provide better one-step ahead quantile forecasts than the direct quantile forecasting (the method 5).

Turning to the analysis of the Hit Ratios, whose values transformed to the Kupiec statistics are collated for $\tau = 0.75$ and 0.95 in Tables 5 and 6, first we see that, in terms of quantile unbiasedness, all the methods perform poorly in multistep forecasting. The only exception to this rule are median forecasts for the first dataset, in which case all procedures show excellent performance for almost all h. As can be seen in Table 5, this generally does not apply to larger quantiles, although it can be noted that, for this dataset, in the case of $\tau = 0.75$, the bootstrap methods (7, 8, 7L, 8L) together with the procedure 6L are able to fulfill the assumed service level requirements at longer horizons, i.e., for h in the range 7–12. For quantiles above 0.75, the performance of the different forecasting procedures applied to the first dataset deteriorates further, with only few methods providing quantile unbiased forecasts at horizons beyond $h = 1$. However, with almost no exceptions, the unbiasedness at horizons longer than $h = 2$ can not be confirmed. Interestingly, the best performance in terms of longer horizon forecasting most often is shown by the parametric bootstrap methods, which produce the smallest and, at the same time, the closest to their nominal values Hit Ratio statistics for large h.

On the other hand, in the case of the second dataset (see the example statistics in Table 6), the performance of the different methods in terms of quantile unbiasedness at higher quantiles improves, although usually the unbiasedness does not hold for forecasts more than three steps ahead, with the exception of the procedure 1, which produces quantile unbiased forecasts up to eight steps ahead for $\tau = 0.99$ and up to six steps ahead for $\tau = 0.975$. At the longest horizons, the smallest (and, at the same time, the closest to their nominal values) Hit Ratios are for the bootstrap methods, especially when performed on levels of the data and in their parametric versions. Interestingly, for $\tau = 0.75$ a good performance at horizons from one to three in terms of the Hit Ratios is achieved under the Gaussianity assumption (i.e., with the procedures 2, 4 and 8).

Figure 2 shows average ranks of the different forecasting procedures for quantiles of orders 0.75 and 0.95. It turns out that, for the first dataset, this part of the forecast evaluation favors the method 6L (usually for forecast horizons starting from $h = 5$–7) and the bootstrap approach, particularly the one based on the normal distribution if shorter forecast horizons are considered. On the other hand, if the second dataset is analyzed, the favored approaches are all based on modeling levels and, for longer forecast horizons, take the form of the bootstrap method (particularly the parametric procedure 8, although for $\tau = 0.95$ the procedure 7 largely dominates the other

Table 5 Tests of quantile unbiasedness for the first dataset

Quantile order τ	0.75						0.95					
Method/forecast horizon h	1	2	3	6	9	12	1	2	3	6	9	12
Levels												
1	**5.40**	**0.01**	20.1	30.3	26.2	20.1	**0.06**	**5.34**	37.6	82.5	102	135
2	**5.40**	**0.10**	13.8	22.5	21.3	17.9	**0.00**	10.4	40.0	92.2	109	135
3	**5.40**	**4.01**	24.9	49.1	70.1	62.1	**0.06**	15.1	64.2	183	260	345
4	**5.40**	**1.13**	21.3	45.7	66.0	58.2	**0.00**	15.1	73.2	187	269	340
5	0.77	8.57	39.2	62.1	80.8	70.1	42.5	89.0	175	366	383	394
6	**0.54**	**0.01**	10.2	20.1	19.0	17.9	**5.34**	18.6	45.0	92.2	102	117
7	6.86	**0.10**	**6.42**	9.4	**5.15**	**1.78**	**0.66**	**3.38**	16.8	30.7	20.4	20.4
8	7.66	**1.04**	**6.42**	10.2	**5.77**	**2.15**	**0.06**	**6.47**	13.5	28.5	20.4	24.4
9	**4.74**	**0.14**	16.8	28.9	27.5	23.7	**0.66**	7.73	35.2	105	113	139
10	**6.11**	**0.14**	17.9	26.2	26.2	21.3	**0.34**	15.1	47.6	105	120	151
Logarithms												
1L	**0.20**	**0.43**	14.7	28.9	24.9	27.5	**1.20**	16.8	40.0	109	128	151
2L	**0.35**	**0.43**	13.8	23.7	23.7	20.1	**0.34**	10.4	30.7	92.2	95.6	120
6L	**4.74**	**0.35**	**4.57**	10.2	**1.44**	**1.78**	**0.10**	**1.82**	10.4	26.4	32.9	20.4
7L	**1.04**	**0.14**	10.2	20.1	8.57	**0.86**	**0.00**	7.73	20.4	40.0	22.3	20.4
8L	**2.10**	**0.01**	8.57	11.0	**4.01**	**0.00**	**0.27**	**5.34**	13.5	24.4	15.1	13.5
9L	**0.20**	**1.13**	19.0	37.6	30.3	27.5	**0.34**	28.5	61.3	120	143	154
10L	**0.03**	**0.63**	16.8	40.7	33.1	26.2	**0.10**	20.4	55.7	102	124	127

The 1% (5%) critical value from the $\chi^2(1)$ distribution is 6.63 (3.84). Statistics for which the quantile unbiasedness is not rejected at the 1% level are given in bold

Table 6 Tests of quantile unbiasedness for the second dataset

| Quantile order τ	0.75						0.95					
Method/Forecast horizon h	1	2	3	6	9	12	1	2	3	6	9	12
Levels												
1	**2.47**	7.97	10.8	64.6	64.6	57.8	**1.82**	**1.82**	**1.82**	13.7	47.8	77.0
2	**1.09**	**0.14**	**0.29**	39.5	43.2	41.3	**1.82**	**1.82**	**0.97**	24.0	42.0	95.2
3	**2.47**	**3.56**	7.97	71.8	64.6	76.8	**1.82**	**3.04**	**0.39**	28.8	84.1	153
4	**1.09**	**0.76**	**0.37**	51.3	49.2	67.0	**1.82**	**1.82**	**0.08**	31.3	84.1	140
5	12.9	7.97	8.86	47.2	43.2	57.8	**4.53**	10.2	11.9	225	268	295
6	**6.31**	**3.00**	**2.47**	47.2	34.1	35.8	**0.46**	**1.64**	**2.45**	60.1	70.1	84.1
7	**6.31**	10.8	11.8	51.3	37.6	27.5	**3.04**	**3.04**	**0.97**	56.9	21.8	24.0
8	**0.29**	**0.37**	**3.00**	45.2	34.1	23.0	**1.82**	**1.82**	**0.97**	56.9	21.8	19.6
9	14.0	17.6	14.0	82.0	67.0	60.0	**0.97**	**0.08**	**0.00**	63.4	95.2	115
10	**2.47**	14.0	14.0	79.4	60.0	60.0	**0.97**	**0.46**	**2.45**	87.7	115	136
Logarithms												
1L	**6.31**	17.6	18.9	90.0	74.3	64.6	**0.00**	**0.97**	**3.42**	107	144	149
2L	**3.00**	15.2	15.2	87.3	69.4	60.0	**0.08**	**0.46**	**4.53**	95.2	131	144
6L	**3.56**	8.86	**4.18**	53.4	27.5	30.7	**0.00**	**0.08**	**0.97**	44.8	56.9	77.0
7L	20.2	27.5	37.6	84.6	41.3	27.5	**0.13**	**0.13**	21.8	111	70.1	66.7
8L	11.8	21.6	29.1	82.0	43.2	24.5	**0.08**	**0.39**	13.7	107	70.1	63.4
9L	8.86	14.0	23.0	92.8	79.4	69.4	**0.46**	**2.45**	**3.42**	107	140	176
10L	7.97	15.2	16.4	92.8	79.4	64.6	**2.45**	**1.64**	**4.53**	91.5	144	158

See *Note* to Table 5
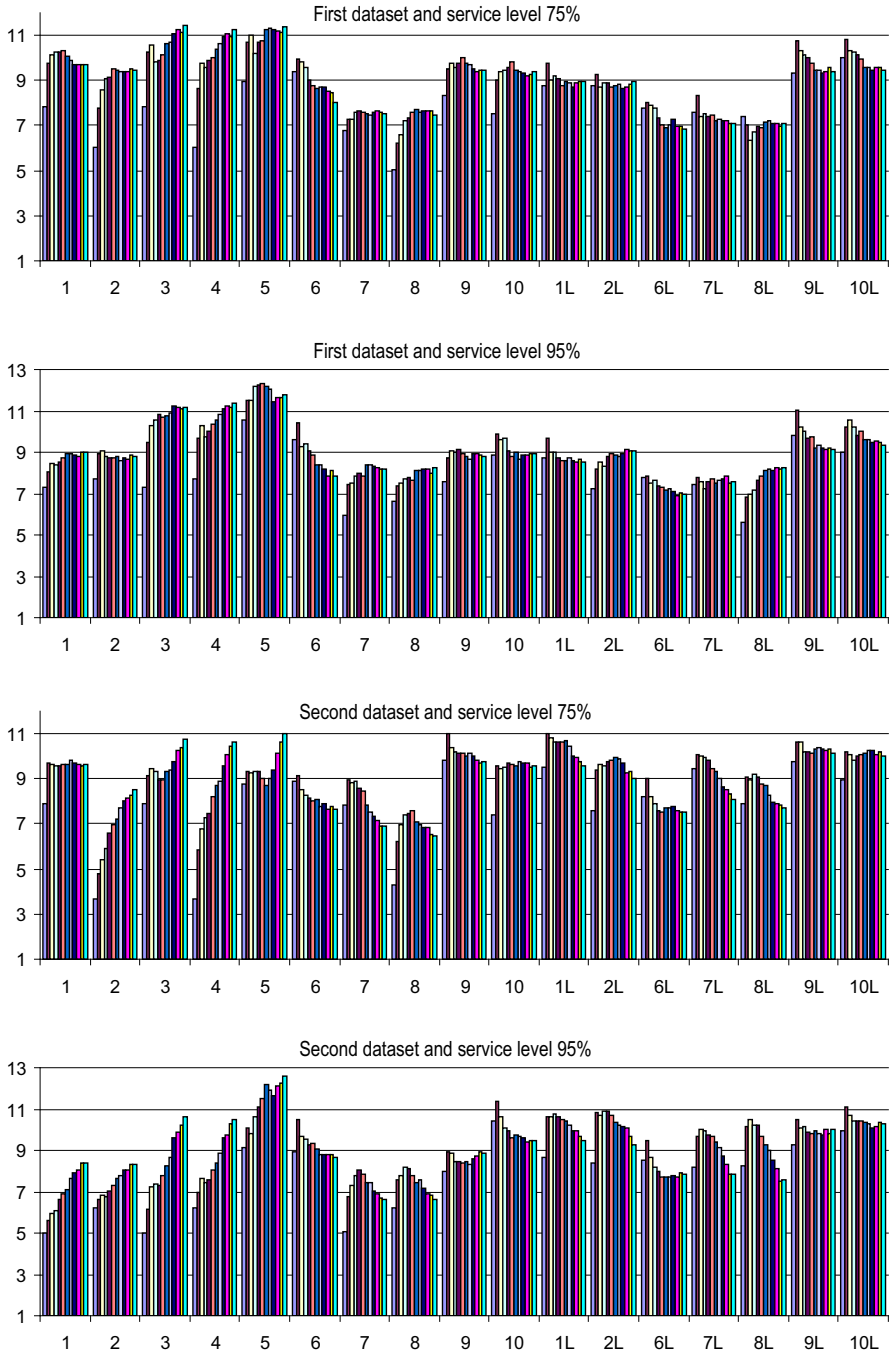
Fig. 2 Average ranks of models for forecast horizons in the range 1–12 months

methods), while in the case of shorter horizon forecasting, the ols based iterated procedures 1 and/or 2 lead to the best outcomes.

Finally, examining the Fill Rates, which are not reported to save space, we observe that, once more, the dominant solution for the time series of length 68–69 are the parametric simulation approach accounting for estimation errors and the simulations utilizing quantile regression, with the latter providing best forecasts at horizons 2–8 or even longer, while the methods 1, 2, 7 and 8 can be distinguished in the case of the second dataset, with the bootstrap approach performing particularly well at the longer horizons among those considered and more often in its parametric form. We also note that, after robustification, for the largest quantiles and particularly in the case of the methods indicated above, the Fill Rate statistics often reach 100%.

The detailed analysis presented above shows that, although certain discrepancies across the accuracy metrics and the datasets are observed, it is safe to say that the most promising approaches to multistep quantile forecasting of demand are those which either explicitly account for estimation errors or make use of predictive distributions generated with the help of quantile regression. In forecasting at the shortest horizons, especially based on larger estimation samples, the AR-GARCH models with trend functions are able to outperform the other approaches examined here. It is also worth underlining that, for our datasets, there are substantial discrepancies among the different methods' performance in terms of the costs associated with forecasts, particularly in the case of longer-term forecasting, and that it is rarely the case that these methods produce quantile unbiased multistep predictions.

5 Conclusions

Multistep forecasting of demand in logistics systems optimized according to the cycle service level approach is often based on exact analytic formulas ignoring estimation errors. Simulation (Monte Carlo or bootstrap) procedures are rarely used to this end. Our empirical results obtained for the popular dataset of 474 monthly microeconomic time series from the M3 forecast competition indicate that simulation based procedures should certainly be taken into account when deciding upon a forecasting approach for supply chain management. In particular, the best forecasts at annual horizons were obtained either with simulation techniques utilizing quantile regression or with procedures able to explicitly account for estimation errors. It is worth underlining that, according to our results, a careful choice of a forecasting procedure for multistep quantile forecasting is crucial since huge discrepancies in the forecast accuracy were observed among the analyzed methods, indicating that such a choice will have a substantial influence on the total costs of a logistics system. From the cost perspective, it may be advisable to invest in fast computational tools, which will deliver better forecasts on a regular (for example, weekly or monthly) basis.

Our empirical study also shows that direct approaches to quantile forecasting (either ols or quantile regression based), which use disaggregated date to forecast aggregates, are often outperformed by some sort of iterated forecasting, especially in the case of high cycle service levels. Finally, we also found that GARCH models, particularly when estimated on longer time series, should be seriously considered among the pre-

ferred solutions for short horizon forecasting of demand in supply chain management applications.

Among future research challenges in the field of supply chain and logistics forecasting one can distinguish a comparison of different approaches to computing other types of optimal forecasts of demand, in particular those fulfilling the type II service level constraints, defined earlier in this paper. It will be worth investigating, for example, if the different functionals of demand can be better forecasted with the help of models belonging to the recently introduced class of generalized autoregressive score models (see Creal et al. 2013; Patton et al. 2017). Furthermore, it is also worth examining if some other characteristics of demand, such as single period Fill Rates for a given CSL used in performance forecasting, which are not elicitable in the sense of Gneiting (2011b) (for proof, see Bruzda 2018), can be made elicitable jointly with other functionals of demand. In a recent paper, Fissler and Ziegel (2016) have shown that Expected Shortfall is elicitable jointly with Value-at-Risk and that this fact has immense implications for the derivation, evaluation and testing of forecasts of this functional (also compare Dimitriadis and Bayer 2017; Patton et al. 2017). This finding opens a new chapter of applied forecasting in finance and sets further research questions concerning possible higher-order elicitability of other functionals of interest to decision makers, including popular demand characteristics. Some steps in this direction have been already taken in Bruzda (2018), but more is needed.

References

Ali MM, Boylan JE, Syntetos AA (2012) Forecast errors and inventory performance under forecast information sharing. Int J Forecast 28:830–841

Alwan LC, Xu M, Yao D-Q, Yue X (2016) The dynamic newsvendor model with correlated demand. Decis Sci 47:11–30

Babai MZ, Ali MM, Boylan JE, Syntetos AA (2013) Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: theory and empirical analysis. Int J Prod Econ 143:463–471

Brandimarte P, Zotteri G (2007) Introduction to distribution logistics. Wiley, Hoboken

Bruzda J (2016) Quantile forecasting in operational planning and inventory management. An initial empirical verification. Dyn Econ Model 16:5–20

Bruzda J (2018) Quantile smoothing in supply chain and logistic forecasting. Working Paper, Nicolaus Copernicus University, Toruń. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3038014

Chen J, Lin DKJ, Thomas DJ (2003) On the single item fill rate for a finite horizon. Oper Res Lett 31:119–123

Choi T-M, Chiu C-H (2012) Risk analysis in stochastic supply chains. A mean-risk approach. Springer, New York

Chopra S, Meindl P (2007) Supply chain management strategy, planning, and operations, 3rd edn. Pearson Education, Upper Saddle River

Christoffersen PF, Diebold FX (1997) Optimal prediction under asymmetric loss. Econ Theory 13:808–817

Clements MP (2005) Evaluating econometric forecasts of economic and financial variables. Palgrave Macmillan, New York

Creal D, Koopman SJ, Lucas A (2013) Generalized autoregressive score models with applications. J Appl Econ 28:777–795

Dallery Y, Babai Z (2005) Inventory management: forecast based approach vs standard approach. In: International conference on industrial engineering and systems management IESM05, May 2005, Marrakech, Morocco, hal-00118642

Datta S, Graham DP, Sagar N, Doody P, Slone R, Hilmola O-P (2009) Forecasting and risk analysis in supply chain management: GARCH proof of concept. In: Wu T, Blackhurst JV (eds) Managing supply chain risk and vulnerability. Tools and methods for supply chain decision makers. Springer, London, pp 187–203

Datta S, Granger CWJ, Barari M, Gibbs T (2007) Management of supply chain: an alternative modelling technique for forecasting. J Oper Res Soc 58:1459–1469

De Cosmis S, De Leone R, Kropat E, Meyer-Nieberg S, Pickl S (2013) Electric load forecasting using support vector machines for robust regression. In: EAIA and MatH '13 Proceedings of the Emerging M&S Applications in Industry & Academia/Modeling and Humanities Symposium, Article No. 9

De Nicolò G, Lucchetta M (2017) Forecasting the tail risks. J Appl Econ 32:159–170

Dimitriadis T, Bayer S (2017) A joint quantile and expected shortfall regression framework. Working Paper, University of Konstanz, arXiv:170402213v1

Engle RF, Manganelli S (2004) CAViaR conditional autoregressive Value at Risk by regression quantiles. J Bus Econ Stat 22:367–381

Ferbar L, Creslovnik D, Mojškerc B, Rajgelj M (2009) Demand forecasting methods in a supply chain: smoothing and denoising. Int J Prod Econ 118:49–54

Fissler T, Ziegel JF (2016) Higher order elicitability and Osband's principle. Ann Stat 44:1680–1707

Francq C, Zakoïan J-M (2010) Inconsistency of the MLE and inference based on weighted LS for LARCH models. J Econ 159:151–165

Gneiting T (2011a) Quantiles as optimal point forecasts. Int J Forecast 27:197–207

Gneiting T (2011b) Making and evaluating point forecasts. J Am Stat Assoc 106:746–762

Granger CWJ (1969) Prediction with a generalized cost of error function. Oper Res Q 20:199–207

Granger CWJ (1999) Outline of forecast theory using generalized cost functions. Span Econ Rev 1:161–173

Granger CWJ, Machina MJ (2006) Forecasting and decision theory. In: Elliott G, Granger CWJ, Timmermann A (eds) Handbook of economic forecasting, vol 1. North Holland, Amsterdam, pp 81–98

Heckmann I, Comes T, Nickel S (2015) A critical review on supply chain risk—definition, measure and modeling. Omega 52:119–132

Hill AV, Zhang W, Burch GF (2015) Forecasting the forecastability quotient for inventory management. Int J Forecast 31:651–663

Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ (2016) Probabilistic energy forecasting: global Energy Forecasting Competition 2014 and beyond. Int J Forecast 32:896–913

Knight FH (1921) Risk, uncertainty and profit. Houghton Mifflin Company, Boston

Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge

Koenker R (2011) Quantile regression, London School of Economics short course, 16–17 May 2011

Koenker R, Bassett G Jr (1978) Regression quantiles. Economics 46:33–50

Komunjer I (2013) Quantile prediction. In: Elliot G, Timmermann A (eds) Handbook of economic forecasting, vol 2. North Holland, Amsterdam, pp 961–994

Kupiec P (1995) Techniques for verifying the accuracy of risk management models. J Deriv 3:73–84

Leopold A (2016) Energy related system dynamic models: a literature review. Cent Eur J Oper Res 24:231–261

Lewis CD (1997) Demand forecasting and inventory control. A computer aided learning approach, Woodhead Publishing, Abington

Makridakis S, Hibon M (2000) The M3-competition: results, conclusions and implications. Int J Forecast 16:451–476

Maume-Deschamps V, Rullière D, Said K (2017) Multivariate extensions of expectiles risk measures. Depend Model 5:20–44

McCarthy TM, Davis DF, Golicic SL, Mentzer JT (2006) The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. J Forecast 25:303–324

Nalcaci G, Özmen A, Weber GW (2018) Long-term load forecasting: models based on MARS, ANN and LR methods. Cent Eur J Oper Res. https://doi.org/10.1007/s10100-018-0531-1

Osband KH (1985) Providing incentives for better cost forecasting. Ph.D. thesis, University of California, Berkeley

Pascual L, Romo J, Ruiz E (2001) Effects of parameter estimation uncertainty on prediction densities: a bootstrap approach. Int J Forecast 17:83–103

Patton AJ, Timmermann A (2007) Properties of optimal forecasts under asymmetric loss and nonlinearity. J Econ 140:884–918

Patton AJ, Timmermann A (2010) Generalized forecast errors, a change of measure, and forecast optimality conditions. In: Bollerslev Y, Russell JR, Watson MW (eds) Volatility and time series econometrics: essays in honor of Robert F Engle. Oxford University Press, Oxford, pp 194–212

Patton AJ, Ziegel JF, Chen R (2017) Dynamic semiparametric models for expected shortfall (and Value-at-Risk). Working Paper, arXiv:170705108v1

Pearson M (2006) The application of prediction capability and adaptive target control to the newspaper business. Eur J Oper Res 168:475–491

Pohl H-C, Köhler H, Thomas D (2010) State of the art in supply chain risk management research: empirical and conceptual findings and a roadmap for the implementation in practice. Logist Res 2:33–44

Silver EA, Naseraldin H, Bischak DP (2009) Determining the reorder point and order-up-to-level in a periodic review system so as to achieve a desired fill rate and a desired average time between replenishments. J Oper Res Soc 60:1244–1253

Sodhi MS, Son B-G, Tang CS (2012) Researchers' perspectives on supply chain risk management. Prod Oper Manag 21:1–13

Strijbosch LWG, Moors JJA (2005) The impact of unknown demand parameters on (R, S)-inventory control performance. Europ J Oper Res 162:805–815

Syntetos AA, Babai Z, Boylan JE, Kolassa S, Nikolopoulos K (2016) Supply chain forecasting: theory, practice, their gap and the future. Eur J Oper Res 252:1–26

Syntetos AA, Babai MZ, Davies J, Stephenson D (2010a) Forecasting and stock control: a study in a wholesaling context. Int J Prod Econ 127:103–111

Syntetos AA, Boylan JE (2008) Demand forecasting adjustments for service-level achievement. IMA J Manag Math 19:175–192

Syntetos AA, Nikolopoulos K, Boylan JE (2010b) Judging the judges through accuracy-implication metrics: the case of inventory forecasting. Int J Forecast 26:134–143

Tang CS (2006) Perspectives in supply chain risk management. Int J Prod Econ 103:451–488

Thomas DJ (2005) Measuring item fill-rate performance in a finite horizon. Manuf Serv Oper Manag 7:74–80

Thomassey S (2010) Sales forecasts in clothing industry: the key success factor of the supply chain management. Int J Prod Econ 128:470–483

Tratar LF (2010) Joint optimisation of demand forecasting and stock control parameters. Int J Prod Econ 127:173–179

Tratar LF (2015) Forecasting method for noisy demand. Int J Prod Econ 161:64–73

Wang X, Petropoulos F (2016) To select or to combine? The inventory performance of model and expert forecasts. Int J Prod Res 54:5271–5282

Weiss AA (1996) Estimating time series models using the relevant cost function. J Appl Econ 11:539–560

Weiss AA, Andersen AP (1984) Estimating time series models using the relevant forecast evaluation criterion. J R Stat Soc Ser A 147:484–487

Other sources

The Forecasting Principles website at http://www.forecastingprinciples.com/index.php/data [accessed 30.06.2018]

The website of the International Institute of Forecasters at https://forecasters.org/resources/time-series-data/m3-competition/ [accessed 30.06.2018]