

# Exponential Smoothing Forecasts: Taming the Bullwhip Effect when Demand is Seasonal.

Maximiliano Udenio<sup>a</sup>, Eleni Vatamidou<sup>b</sup>, and Jan C. Fransoo<sup>c</sup>

<sup>a</sup>Faculty of Economics and Business, KU Leuven, Leuven, Belgium.

<sup>b</sup>The Faculty of Business and Economics, University of Lausanne, Quartier UNIL-Chamberonne Bâtiment Extranef, 1015 Lausanne, Switzerland.

<sup>c</sup>Tilburg University, Tilburg, The Netherlands.

## ABSTRACT

In this paper, we study the influence of seasonal demands and forecasts on the performance of an Automatic Pipeline, Variable Inventory, Order-Based, Production Control System (APVIOBPCS) using linear control theory. In particular, we consider a system with independent adjustments for the inventory and pipeline feedback loops that uses a seasonal forecast based on a no-trend, additive-seasonality exponential-smoothing model, and compare its performance to an equivalent system using (non-seasonal) simple exponential smoothing. To quantify the performance of these systems, we first conduct extensive numerical experimentation to calculate the Bullwhip Effect for orders and inventory under a number of different behavioral parametrizations and seasonality conditions. Then, we run simulation experiments under different stochastic seasonal demands to characterize the robustness of each of the systems to behavioral biases and misspecification of parameters. Finally, we perform an analysis of the response of the system under real-life demand streams taken from the recent M5 forecasting competition dataset. We find that the system with seasonal forecasting significantly outperforms the system with simple exponential smoothing under certain demand assumptions. With optimal parameter settings, the forecast error of the seasonal model can be up to 40% lower. However, we also find that the forecast superiority does not necessarily translate to the performance of the system measured through the bullwhip metrics. In addition, the seasonal forecasting model is very sensitive to the demand frequency and smoothing parameters, while the simple exponential smoothing model is very robust. This implies that the real life benefits of implementing a seasonal forecasting model are not obvious and depend on the particular situation; under a large number of settings (e.g., low sea-

sonality, high-smoothing), the good performance of simple exponential smoothing certainly justifies its popularity in the industry and research worlds alike.

#### **KEYWORDS**

Control Theory; Bullwhip Effect; Inventory Control; Forecasting; Seasonal Demands.

Word count: 7958.

## **1. Introduction**

In this paper, we use linear control theory to study the influence of seasonal demands and seasonal forecasts on the performance of an APVIOBPCS (Automatic Pipeline, Variable Inventory, Order-Based Production Control System) design. In particular, we extend the behavioral APVIOBPCS design from Udenio et al. (2017) to allow for the explicit forecasting of seasonal demands using a no-trend, additive-seasonality exponential-smoothing model. Throughout the paper, we compare the performance of an APVIOBPCS design that relies on such a seasonal forecast to the performance of an equivalent APVIOBPCS design using a simple exponential smoothing forecasting model. We do so through a mix of numerical analysis of frequency response plots, extensive simulation experiments, and the analysis of the systems' performance using real-life demand streams taken from the latest M competition, M5 (Makridakis et al., 2020b).

Our motivation is simple: It is a relevant topic that has been, thus far, been paid little attention to. Demand seasonality—which is common in practice—is known to dampen the appearance of the bullwhip effect; Cachon et al. (2007) show that highly seasonal industries tend to smooth demand volatility, Chen and Lee (2012) provide the theoretical argumentation as to why this is the case, and Cantor and Katok (2012) replicate this result in a laboratory setting. Nonetheless, the scarcity of bullwhip effect studies that consider decision-making under such conditions is noteworthy—specially considering the vast literature dedicated to analyze multiple aspects of the bullwhip (see Wang and Disney, 2016, for a comprehensive review of bullwhip effect research). In the control theory literature in particular, in spite of the popularity of the bullwhip

effect as a topic (see, a.o. Dejonckheere et al., 2003; Disney and Towill, 2003; Disney, 2008; Hoberg et al., 2007; Hoberg and Thonemann, 2014; Udenio et al., 2017), there are—to the best of our knowledge—no studies that quantify the potential benefits of using seasonal forecasting methodologies under seasonal demands.

In fact, with limited exceptions (see, for example, Li et al., 2014, for an application of the dampened trend forecasting model; used for trending demands), simple exponential smoothing (SES) is typically used as the default forecasting strategy in the field. There is good reason for this. SES is optimal for a wide range of demand structures, such as stationary, ARIMA(0,1,1), and random walks (Muth, 1960; Harrison, 1967). SES was also long held to be the best method of one-period-ahead forecasting (Makridakis et al., 1982) but more recent studies (see, e.g., Petropoulos et al., 2019; Makridakis et al., 2020a) have shown other much more complex methods outperforming it. Nevertheless, these same studies have shown that SES often remains a strong competitor. Moreover, SES is also known to be a robust forecasting methodology that performs well in sub-optimal situations. For example, it is very robust under misspecifications due to incorrect belief in the stationarity of the generating process (Bossons, 1966) and it is even very accurate under a Poisson demand process, particularly with step changes in the mean (Gross and Craig, 1974).

This said, it stands to reason that seasonal forecasting methods are surely to be preferred when demand is seasonal. These methods, however, can be more cumbersome to adopt in practice; manual implementations require a deeper knowledge of the underlying structure of demand and, while modern automatic implementations work well (Hyndman et al., 2002; Babai et al., 2021), they still require a knowledgeable user to make the most of them. Moreover, being a method tailored to a particular demand structure can come at the cost of its robustness; indeed, seasonal models tend to fare modestly in studies that benchmark multiple forecasting models using multiple datasets (e.g., Taylor, 2003; Petropoulos et al., 2019; Li et al., 2022). We explore the tradeoffs behind implementing seasonal forecasting in a APVIOBPCS inventory/production design by quantifying its potential performance advantage relative to its sensitivity to demand and model parameters.

Even though we are interested in the performance of different forecasting models,

what matters in an inventory/production setting is not the performance of the forecast but the performance of the *system*. Thus, we measure the performance of each of the forecasting models through a series of well-known performance metrics related to the amplification of orders and inventory. For notational convenience, we adopt the shorthand notation from Hyndman et al. (2002), whereupon exponential smoothing models are denoted using a pair of letters according to the type of Trend and Seasonality they assume (Additive, Multiplicative, or None). Thus, the models relevant to this work are (N,N), i.e., simple exponential smoothing, and (N,A), i.e., exponential smoothing with additive seasonality. Our main objective is to compare the performance of the (N,A) and (N,N) models in terms of their: (1) theoretical performance (bullwhip of orders and inventory, maximum amplification), (2) empirical performance (bullwhip of orders and inventory) with stochastic demand streams of different seasonality strengths, (3) empirical performance (bullwhip of orders and inventory) with real-life demand streams, and (4) robustness to parameter (smoothing and seasonality) specifications. In this sense, our study is conceptually aligned with Petropoulos et al. (2019) and Spiliotis et al. (2021), who compare the empirical performance of different forecasting methods through a similar lens.

From a theory perspective, we contribute to the literature by using control-theoretic methods to derive closed-form transfer functions for the APVIOBPCS design with (N,A) forecasts. This allows us to characterize the performance of such a system under arbitrary demands as well as quantify its performance relative to (N,N) under a wide array of parametrizations.

We confirm that, if the seasonality of demand is known and deterministic, adopting a seasonal forecast can result in significant performance benefits (Hyndman et al., 2002). This is intuitive. However, these potential benefits come at the cost of an increased sensitivity to parameter selection; the superior performance of the system using (N,A) forecasts can vanish if parameters are not well chosen. This implies that the real-life benefits of adopting the (N,A) model depend on the ability of decision-makers to correctly set model parameters or implement automated solutions, such as the *fable* package in *R* or Facebook’s Prophet open-source models. Furthermore, we show that the performance of the system is also dependent on the structure of the demand stream.

In fact, the potential performance advantage of the (N,A) model against the (N,N) model increases with the seasonality strength.

Seasonality strength, however, is also associated with an increase in the sensitivity of the (N,A) model and a decrease in the absolute value of the bullwhip; factors which may negate (some of) the benefits of adopting the (N,A) model in a real-life setting. Therefore, our analysis suggests that, from a managerial perspective, the choice of forecasting methodology needs to be made while taking into account features of the demand such as the strength of its seasonality and whether or not the seasonal cycles are regular and predictable. Under a large number of settings (e.g., low seasonality, high-smoothing), the simplicity and robustness of the (N,N) model certainly justify its popularity, even under non-optimal conditions. Finally, using M5 competition data we show that the order and inventory performance of the (N,A) model is non-linear and sensitive, a small change in the smoothing parameter can significantly change its response. In contrast, the (N,N) model exhibits an almost linear response in terms of the order and inventory performance, a small change in the smoothing parameter introducing predictable, small changes in its response.

The rest of the paper is laid out as follows. First, we extend the APVIOBPCS design to allow for seasonal forecasts, derive its transfer functions, and present a set of performance metrics in Section 2. We detail the experimental set-up in Section 3.1, discuss theoretical results in Section 3.2, and present extensive computational experiments in Section 3.3. Finally, we illustrate the practical application of our model through the implementation of a real-life example based on the M5 competition data in Section 3.5. We present our conclusions in Section 4.

## **2. Model Description and Performance Metrics**

The APVIOBPCS family of models traces its lineage directly to Simon (1952); the first study to use control theoretic tools to analyze the behavior of a production and control system. In the decades since, various extensions and methods have been published. Of note is the foundational research of Towill (1982), which introduced the first model of its kind—the Inventory Order-Based Production Control System (IOBPCS)—and the

associated control-theoretic methods for its analysis.

Among the issues studied in subsequent extensions are: the inclusion of the pipeline in the ordering policy (e.g., John et al., 1994); the optimization of decision parameters in terms of order and inventory variability (e.g., Dejonckheere et al., 2003); service level considerations (e.g., Disney et al., 2006); the theoretical equivalence of discrete and continuous time analysis (e.g., Warburton and Disney, 2007); the influence of inventory policies (e.g., Hoberg et al., 2007); and the effect of behavioral policies (e.g., Udenio et al., 2017). We refer the interested reader to the literature review study of Lin et al. (2016) for further details on the history, evolution, and future of the field.

Of importance to this study, however, is the relative scarcity of studies comparing the behavior of the system under different forecasting models. With limited exceptions (e.g., Dejonckheere et al., 2003; Li et al., 2014), the vast majority of analytical studies assume forecasts are calculated via simple exponential smoothing, i.e., (N,N) forecast models. Largely due to the complexity of the resulting closed-form expressions, more complex forecasting methodologies are typically studied using continuous-time models (see, Zhan et al., 2018, for an application to VMI-managed seasonal apparel sales) or through simulation (e.g., Wright and Yuan, 2008, find that the bullwhip can be mitigated by using Holt’s and Brown’s forecasting methods). To the best of our knowledge, the influence of seasonal forecasts on the behavior of the system has not yet been studied through analytic, discrete-time, control-theoretic methods. In this paper, we implement seasonal forecasting by extending the discrete-time, periodic-review, single-echelon, general APVIOBPCS design of Udenio et al. (2017) to explicitly allow for a no-trend, additive-seasonality forecast model, i.e., (N,A).

### 2.1. *APVIOBPCS Model*

We now present the general APVIOBPCS model. The structural parameters of the system are the inventory coverage ( $C \in \mathbb{R}^+$ ), the delivery lead time ( $L \in \mathbb{N}$ ), and the forecast smoothing parameter  $\alpha \in [0, 2]^1$ . The system maintains a target inventory ( $\hat{i}$ ) equal to the expected demand over  $C$  periods and a target pipeline ( $\hat{p}$ ) equal to

---

<sup>1</sup>Note that this is less restrictive than the commonly used values of  $0 \leq \alpha \leq 1$ . Even though the extended range has little practical value, it can be shown that the system is mathematically stable when  $0 \leq \alpha \leq 2$  (Hyndman et al., 2008; Udenio et al., 2017).

the expected lead time demand. The lead time is assumed deterministic and defined as the time elapsed between the placement and receipt of a replenishment order under non-stockout conditions. Replenishment orders ( $o$ ) depend on the gap between actual and desired values of the inventory ( $i$ ) and pipeline ( $p$ ). The behavioral parameters of the system are the inventory ( $\gamma_I \in \mathbb{R}$ ) and pipeline ( $\gamma_P \in \mathbb{R}$ ) adjustment factors. The behavioral parameters specify the fraction of the gap between target and actual values that are taken into account to calculate orders:  $\gamma_I$  is the fraction of the inventory gap to be closed and  $\gamma_P$  is the fraction of the pipeline gap to be closed. For instance, a system with  $\gamma_I = 1$  and  $\gamma_P = 0$  completely closes the inventory gap with every order, while it ignores the pipeline entirely. (See Udenio et al. (2017) for detailed insights on the influence of these parameters on the behavior of the model.)

Demand forecasts are calculated via the (N,A) model. Let  $f_{t,t+m}$  be the demand forecast calculated at period  $t$  for period  $t+m$ ;  $\delta$ , the seasonal smoothing parameter; and  $\rho$  the number of periods in each season, which we call the model seasonality. Then, forecasts are maintained according to the following system of equations:

$$\begin{aligned}
 s_t &= \alpha (d_t - j_{t-\rho}) + (1 - \alpha) s_{t-1}, \\
 j_t &= \delta (d_t - s_t) + (1 - \delta) j_{t-\rho}, \\
 f_{t,t+m} &= s_t + j_{t+m-\rho},
 \end{aligned} \tag{1}$$

where  $s_t$  and  $j_t$  are auxiliary variables that keep track of the level and seasonal components of the final forecast. Note that the seasonal terms of the (N,A) forecast ( $j_t$ ) are updated once every seasonal cycle, but the structure of the seasonal pattern ( $\rho$ ) is assumed invariant (Hyndman and Athanasopoulos, 2018).

Formally, the sequence of events and the equations in the model are as follows: at the beginning of each period ( $t$ ) a replenishment order ( $o_t$ ) based on the previous period's demand forecast ( $f_{t-1,t+L}$ ) is placed with the supplier. Following this, the orders that were placed  $L$  periods prior are received. Next, the demand for the period ( $d_t$ ) is observed and served. Excess demand is back-ordered. Then, the demand forecast ( $f_{t,t+L+1}$ ) is updated. The forecast represents the expected demand and is used to

compute the target levels of both inventory and pipeline,

$$\hat{i}_t = \sum_{k=1}^C f_{t,t+k},$$

$$\hat{p}_t = \sum_{k=1}^L f_{t,t+k}.$$

The orders that will be placed in the following period ( $o_{t+1}$ ) are generated according to an anchor and adjustment-type procedure,

$$o_{t+1} = \gamma_I(\hat{i}_t - i_t) + \gamma_P(\hat{p}_t - p_t) + f_{t,t+L+1}.$$

The balance equations for inventory ( $i$ ) and pipeline ( $p$ ) are:  $i_t = i_{t-1} + o_{t-L} - d_t$ , and  $p_t = p_{t-1} + o_t - o_{t-L}$ . Orders and inventories can be negative (i.e., backlogs and returns are allowed) to maintain the linearity of the model. Note that when  $\delta = 0$ , the (N,A) model is equivalent to a simple exponential smoothing forecast, (N,N) model, with smoothing parameter  $\alpha$  and  $f_{t,t+m} = f_t, \forall m \in \mathbb{N}$ . In such cases, this model reduces to the (N,N) model described in Udenio et al. (2017).

As is common in the study of the family of APVIOBPCS models, we use a mathematical transformation (the Z-transform) to simplify the operations on the discrete-time signal (Warburton and Disney, 2007). Using the linearity and time-delay properties of the Z-transform, we can write all system parameters in the frequency domain. Thus, the discrete time transfer function of Equation (1) is given by:

$$F_{m,\rho}(z)/D(z) = \frac{\alpha z(z^\rho - 1) - (\alpha - 1)\delta z^m(z - 1)}{z^{1+\rho} + z^\rho(\alpha - 1) + z(\delta - \alpha\delta - 1) + (\alpha - 1)(\delta - 1)},$$

and the transfer functions for orders and inventories can be derived through algebraic manipulation of:

$$O(z) = \frac{1}{z} \left( \gamma_I(\hat{I}(z) - I(z)) + \gamma_P(\hat{P}(z) - P(z)) + F_{L+1,\rho}(z) \right), \quad (2)$$

$$I(z) = \frac{z}{z-1} (O(z)z^{-L} - D(z)), \quad (3)$$



$$\hat{I}(z) = \sum_{k=1}^C F_{k,\rho}(z), \quad (4)$$

$$\hat{P}(z) = \sum_{k=1}^L F_{k,\rho}(z). \quad (5)$$

Note that Equations (4) and (5) do not have a closed form solution for general values of  $\rho$ ,  $C$ , and  $L$ . Therefore, for our presentation purposes, without loss of generality, we fix these parameters to  $\rho = 50$ ,  $C = 3$ , and  $L = 5$ . The transfer functions for orders and inventory for the (N,A) model are then defined by  $G_O(z) = A_O(z)/B_O(z)$  and  $G_I(z) = A_I(z)/B_I(z)$ , where,

$$\begin{aligned} A_O(z) = & (z-1)z^4(\gamma_I z((z^{50}-1)(z-1+\alpha) - (z-1)(\alpha-1)\delta) \\ & + (z-1)(z(z^{50}-1)\alpha - z^6(z-1)(\alpha-1)\delta)) \\ & + z(\gamma_I(3(z^{50}-1)\alpha - (z^3-1)(\alpha-1)\delta) + \gamma_P(5(z^{50}-1)\alpha - (z^5-1)(\alpha-1)\delta)), \end{aligned} \quad (6)$$

$$\begin{aligned} A_I(z) = & -z^2((z^{50}-1)(z-1+\alpha) - (z-1)(\alpha-1)\delta) + \gamma_I z(z-1)((\gamma_I - \gamma_P) \\ & + z^5(z-1+\gamma_P))((z^{50}-1)(z-1+\alpha) - (z-1)(\alpha-1)\delta) \\ & + ((z^{50}-1)\alpha - (z-1)z^6(\alpha-1)\delta)(z-1)((\gamma_I - \gamma_P) + z^5(z-1+\gamma_P)) \\ & + (\gamma_I(3(z^{50}-1)\alpha - (z^3-1)(\alpha-1)\delta)z(z-1)((\gamma_I - \gamma_P) + z^5(z-1+\gamma_P)) \\ & + (\gamma_P(5(z^{50}-1)\alpha - (z-1)(\alpha-1)\delta)z(z-1)((\gamma_I - \gamma_P) + z^5(z-1+\gamma_P))), \end{aligned} \quad (7)$$

$$\begin{aligned} B_O(z) = & B_I(z) \\ = & z(z-1)((\gamma_I - \gamma_P) + z^5(z-1+\gamma_P))((z^{50}-1)(z-1+\alpha) - (z-1)(\alpha-1)\delta). \end{aligned} \quad (8)$$

As noted before, when  $\delta = 0$ , the above expressions simplify to the same values of  $G_O(z)$  and  $G_I(z)$  as in the (N,N) model developed in Udenio et al. (2017) and thus, the models are equivalent as long as the initial seasonal factors are zero. Moreover, given that the changes in the model are present in the forecasting feed-forward loop, the stability conditions defined in the above-mentioned article for the (N,N) model also hold for the (N,A) model.

## 2.2. Demand Models

To model the seasonal demand stream used in our numerical experiments, we use the form:

$$d_t = \epsilon(t) + A \sin(\omega t). \quad (9)$$

The periodicity of the seasonal component in the demand stream is denoted by  $\omega \in [0, 2\pi]$ . For a demand that is observed every period, the seasonality—defined as the number of periods within a season—is equal to  $P = 2\pi/\omega$  periods. The amplitude of the seasonal component is denoted by  $A$ . This sinusoidal model can generate deterministic streams of varying frequency by setting  $\epsilon(t) = 0$  as well as random realizations of stochastic-seasonal demands by defining  $\epsilon(t)$  as a Gaussian i.i.d. stream.

We use the deterministic demand model in Section 3.2 and the stochastic demand model in Sections 3.3 and 3.4. Then, to evaluate the performance of the system under realistic demands, we introduce real-life demand streams (taken from the M5 forecasting competition) in Section 3.5. Additionally, given the restrictiveness of the demand generating process determined by Equation (9), we also evaluate the performance of the system against an alternative stochastic demand stream (specifically, based on a seasonal ARIMA(0,1,1) process) in Appendix B.

## 2.3. Performance Metrics

We quantify the behavior of the system using a combination of numerical and simulation-based empirical performance metrics for the amplification of orders and inventory. Furthermore, given our interest in understanding the influence of demand seasonality in the performance of the system, we also define a theoretical metric of “seasonality strength”, i.e., how strong the seasonal component of a demand stream is.

### 2.3.1. Numerical Model Performance

We define three measures to quantify the amplification of variability: the amplification ratio, the worst-case amplification, and the bullwhip ratio. These three measures quantify different aspects of the system's response to different demand streams. The amplification ratio measures the amplitude of the systems' output (i.e., orders/inventory) as a response to a sinusoidal input (i.e., demand) of a fixed frequency and unit amplitude. The maximum amplification measures the maximum amplification ratio of the system over all possible input frequencies. Finally, the bullwhip ratio represents the ratio of order/inventory variability to demand variability, under the assumption that demand follows a unit normal distribution.

Let  $|G(e^{i\omega})|$  be the modulus of the transfer function evaluated at the frequency  $\omega$ . Formally, we define  $A_{O,\omega}$  as the amplification ratio of orders for a sinusoidal demand of frequency  $\omega$  and  $A_{I,\omega}$  as the amplification ratio of inventory for a sinusoidal demand of frequency  $\omega$ , i.e.:

$$A_{O,\omega} = |G_O(e^{i\omega})| \quad \text{and} \quad A_{I,\omega} = |G_I(e^{i\omega})|.$$

The worst-case amplification is defined as the maximum value of amplification over the entire frequency range, i.e.,

$$A_O^\infty = \sup_{\forall \omega \in [0, \pi)} |G_O(e^{i\omega})| \quad \text{and} \quad A_I^\infty = \sup_{\forall \omega \in [0, \pi)} |G_I(e^{i\omega})|.$$

Disney and Towill (2003) show that if the input of a system consists of a Gaussian stream with zero mean and unity variance, the ratio of input variance to output variance (typically used to quantify the bullwhip effect in the time domain) is proportional to the “noise bandwidth” of the system, i.e., the square of the area below its frequency response plot. Thus, when the input to our system is a stationary i.i.d. normal demand stream, we can calculate metrics for the bullwhip of orders ( $BW_O$ )

and inventory ( $BW_I$ ):

$$BW_O = \frac{1}{\pi} \int_0^\pi |G_O(e^{i\omega})|^2 d\omega \quad \text{and} \quad BW_I = \frac{1}{\pi} \int_0^\pi |G_I(e^{i\omega})|^2 d\omega.$$

### 2.3.2. Empirical Model Performance

We calculate the empirical model performance by measuring the ratio between input and output variance directly,  $\sigma_{out}^2/\sigma_{in}^2$ . Therefore, the empirical bullwhip of orders ( $\widetilde{BW}_O$ ) can be measured directly in the time domain through  $\widetilde{BW}_O = \sigma_O^2/\sigma_D^2$ . Analogously, the empirical bullwhip of inventory is calculated through:  $\widetilde{BW}_I = \sigma_I^2/\sigma_D^2$ .

### 2.3.3. Demand Seasonality Strength

Wang et al. (2006) propose to quantify the strength of the seasonality of a time series using the ratio of the variance of its seasonal and de-seasonalized components. Formally, the seasonality strength,  $F_S$ , is defined as  $F_S = 1 - (\text{Var}[Y'_t]/\text{Var}[Y_t])$ , where  $Y_t$  is a seasonal, de-trended, time series and  $Y'_t$  is the de-seasonalized, de-trended component of the time series—note that this metric is equivalent to the *seasonality ratio*, independently introduced by Cachon et al. (2007). It can be shown that the variance of a time series generated by a deterministic sinusoid is  $\text{Var}[A \sin(\omega t)] = A^2/2$ . Thus, the seasonality strength of the demand series defined by  $d(t) = \epsilon(t) + A \sin(\omega t)$  with an i.i.d. error term,  $\epsilon(t)$  is defined by:

$$F_s = \frac{A^2}{2 \text{Var}[\epsilon(t)] + A^2}. \quad (10)$$

When the error term is Gaussian with  $\mu = 0$  and  $\sigma = 1$ , the seasonality strength is entirely defined by the amplitude of the sinusoid, such that  $F_S = A^2/(A^2 + 2)$ .

## 3. Results

In this section, we present a number of numerical performance comparisons between an APVIOBPCS design using an (N,A) method of seasonal forecasting (as described in Section 2) and an equivalent design using the widely studied (N,N) forecast method-

ology (used, for example in Hoberg et al., 2007; Udenio et al., 2017).

Our objectives are to test:

- (O.1) The theoretical performance of both systems.
- (O.2) The performance of the systems under stochastic-seasonal demands.
- (O.3) The robustness of the systems w.r.t parameter setting.
- (O.4) The robustness of the systems under different behavioral policies.
- (O.5) The real-life performance of the systems.

We use different demand streams and parameter settings to test the different objectives. Our experimental design is as follows.

### 3.1. *Experimental Set-up*

For each of the systems—(N,N) and (N,A)—we fix the structural parameters and the smoothing factor with the same values as the experiments in Udenio et al. (2017):  $C = 3$ ,  $L = 5$ ,  $\alpha = 0.3$ , and sweep over the seasonal smoothing parameter by letting  $\delta \in \{0.1, 0.3, 0.5, 0.9\}$ . The (N,N) model is set with the same parameters, except for  $\delta = 0$ . For the baseline experiments, we assume no behavioral bias, i.e., we set the inventory and pipeline adjustments such that  $\gamma_I = \gamma_P = 0.5$ . Moreover, we set the seasonality of the baseline (N,A) model at  $\rho = 50$  and compute its performance when demand is misspecified by letting the actual demand seasonality,  $P$ , vary across a range of values. For convenience, we refer to the demand seasonality in terms of periods  $P$ , noting that these can be translated to the frequency in radians through  $\omega = 2\pi/P$ .

We present the results as follows. We test (O.1) by using deterministic sinusoidal demand streams to construct the frequency response (FR) plots of the (N,A) and (N,N) models and compute the numerical metrics, i.e., amplification, worst case amplification, and bullwhip (Section 3.2). We then test (O.2)–(O.4) using stochastic seasonal demand streams as inputs for numerical experimentation and sensitivity analyses (Sections 3.3–3.4). Finally, we test (O.5) by utilizing real-life demand streams, consisting of data from the M5 forecasting competition, as inputs to our models (Section 3.5). In the interest of concision, we include two additional robustness tests as appendices. Appendix A details the frequency response plots and theoretical performance metrics

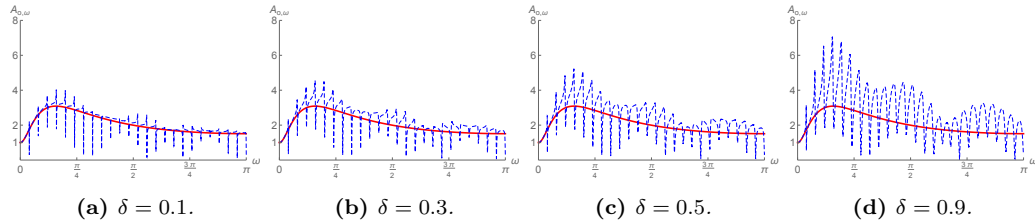
of a system with a different seasonality; namely,  $\rho = 7$ . Appendix B replicates the computational experiments using a different stochastic demand generating process; namely seasonal ARIMA (0,1,1) streams.

### 3.2. Numerical Performance

We compute the frequency response and the associated performance metrics of the system (Section 2.3.1) directly from the transfer functions for orders and inventories, Equations (6) through (8).

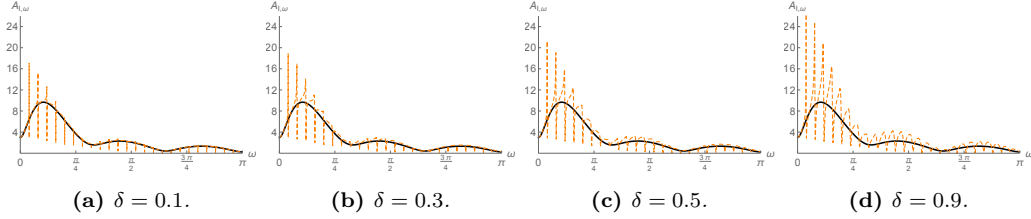
#### 3.2.1. Response of the System Without Behavioral Biases

Figure 1 shows the frequency response of orders of the system without behavioral biases (i.e.,  $\gamma_I = \gamma_P = 0.5$ ) with  $\delta \in \{0, 0.1, 0.3, 0.5, 0.9\}$ . For ease of comparison, we overlay the FR of the (N,N) model ( $\delta = 0$ , plotted in solid red) on top of each of the 4 different parameterizations of the (N,A) model (plotted in dashed blue in each subfigure). Figure 2 shows the equivalent plots for the inventory response, with the (N,N) model overlaid in solid black and the (N,A) models plotted in dashed orange.



**Figure 1.** Order FR of (N,N) (red, solid) and (N,A) models with varying  $\delta$  (blue, dash).  
**Figure 1 Alt Text:** This figure shows the frequency response of the transfer function of orders of the (N,N) and (N,A) models under different values of the seasonal smoothing parameter,  $\delta$ . The response of the seasonal system follows the response of the (N,N) model but exhibiting significant oscillations. Troughs are observed at the frequency corresponding to the seasonality parameter and its harmonics. The amplitude of oscillations increases in  $\delta$ .

The (N,A) model seems to track the response of the (N,N) model but exhibits significant oscillations at specific frequencies. The number and frequency of the peaks (and troughs) is independent of  $\delta$  whereas their amplitude increases in  $\delta$ . Inspection of the figures suggests that for every parametrization, the (N,A) model has a trough at the frequency of the model seasonality (in this case,  $\omega = 2\pi/50$ ) and its harmonics. This suggests that, regardless of the complexity of the frequency response, the (N,A)



**Figure 2.** Inventory FR of (N,N) (black, solid) and (N,A) models with varying  $\delta$  (orange, dash).

**Figure 2 Alt Text:** This figure shows the frequency response of the transfer function of inventory of the (N,N) and (N,A) models under different values of the seasonal smoothing parameter,  $\delta$ . The response of the seasonal system follows the response of the (N,N) model but exhibiting significant oscillations. Troughs are observed at the frequency corresponding to the seasonality parameter and its harmonics. The amplitude of oscillations increases in  $\delta$ .

model will perform well when the demand seasonality is equal to the model seasonality, i.e., when  $\rho = P$ . The oscillatory nature of the frequency response, however, implies that the performance of the (N,A) model is sensitive to a mismatch between these parameters, i.e., when  $\rho \neq P$ . In contrast, the order response of the (N,N) model is relatively smooth across all frequencies, while the inventory response amplifies low frequency demands but attenuates high frequency ones. Finally, given that the bullwhip metric is proportional to the square of the area below the frequency plot, we expect the bullwhip to increase in the seasonal parameter, with low values of  $\delta$  exhibiting a similar bullwhip performance as the (N,N) model.

Table 1 confirms these observations. It shows the values of the amplification of orders and inventory ( $A_{O,\omega}$ ,  $A_{I,\omega}$ ) at seasonality  $P = 50$ , the worst-case amplification ( $A_O^\infty$ ,  $A_I^\infty$ ), and the bullwhip metric ( $BW_O$ ,  $BW_I$ ) for the same set of model parameters.

**Table 1** Theoretical performance metrics of (N,A) and (N,N) models with  $\alpha = 0.3$ .

Model:	(N,N)	(N,A) ( $\delta = 0.1$ )	(N,A) ( $\delta = 0.3$ )	(N,A) ( $\delta = 0.5$ )	(N,A) ( $\delta = 0.9$ )
$A_O^\infty$	3.09	4.01	4.55	5.50	7.03
$A_I^\infty$	9.67	16.92	18.94	21.23	27.58
$A_{O,2\pi/50}$	1.76	0.28	0.35	0.42	0.58
$A_{I,2\pi/50}$	6.43	2.94	2.94	2.94	2.94
$BW_O$	4.59	4.93	5.73	6.69	9.36
$BW_I$	15.47	16.31	18.21	20.52	26.92

The (N,A) model clearly outperforms the (N,N) model when  $\rho = P$ . The (N,N) model, however, has a lower worst-case amplification and bullwhip metric. These results are consistent with the intuition that a seasonal forecast outperforms simple

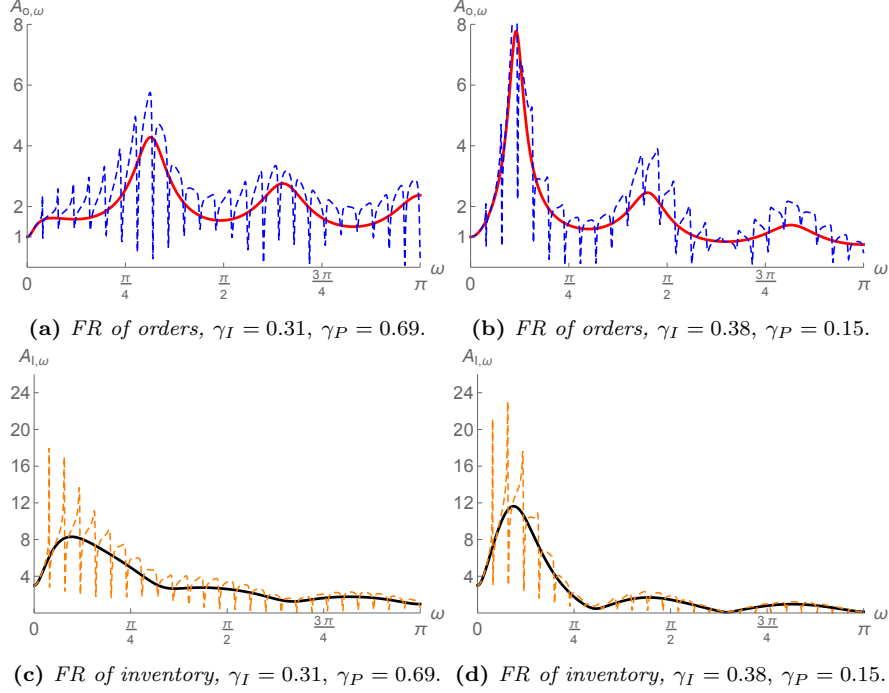
exponential smoothing when correctly parameterized. However, they are also aligned with recent findings highlighting the tradeoffs of more complex forecasting methodologies, where much complexity is required to achieve (sometimes small) increases in performance (see, e.g., Petropoulos et al., 2019; Spiliotis et al., 2021); the sensitivity of the (N,A) model makes it such that mispecifications of the seasonal parameter are heavily penalized—highlighting the importance of the actual implementation (i.e., software packages and parametrization strategies) in practice. Moreover, the observation that  $BW$  metrics increase in the seasonality parameter  $\delta$  is consistent with the effect of the smoothing parameter  $\alpha$  on  $BW$  metrics (see Udenio et al., 2017, for a detailed discussion regarding the effect of the smoothing and structural parameters on the system’s behavior).

### 3.2.2. Response of a Behaviorally-biased System

Figure 3 displays the frequency response of the (N,A) and (N,N) models under different behavioral policies. The two plots at the top (Figures 3a and 3b) correspond to the order response and the two at the bottom (Figures 3c and 3d), to inventory response. With respect to the behavioral policies, the two plots at the left (Figures 3a and 3c) correspond to a system that overestimates the pipeline, whereas the two at the right (Figures 3b and 3d), to a system that underestimates it. We set the behavioral parameters such that the bullwhip of orders for both behaviorally biased systems is the same as in the unbiased system, thus,  $BW_O = 4.59$  in all cases. Namely, we set  $\gamma_I = 0.38$ ,  $\gamma_P = 0.15$  for underestimation of the pipeline; and  $\gamma_I = 0.31$ ,  $\gamma_P = 0.69$  for its overestimation. In the interest of space, we only plot the (N,A) model with  $\delta = 0.5$  noting that the insights obtained in the previous section regarding the effect of the seasonal parameter hold, i.e., the sensitivity of the (N,A) model increases in  $\delta$ .

Similar to the unbiased case, the response of the (N,A) model tracks the response of the (N,N) model albeit with increased sensitivity. Moreover, consistent with the findings from Udenio et al. (2017), underestimating the pipeline increases the amplification of the orders under low demand frequencies, while overestimating the pipeline flattens it. In contrast, the frequency response of inventories appears robust to behavioral changes.





**Figure 3.** *FR of (N,N) (solid) and (N,A) models (dashed) under behavioral biases with  $\delta = 0.5$ . **Figure 3 Alt Text:** This figure shows the frequency response of the transfer function of orders of the (N,N) and (N,A) models under different behavioral biases—under- and over-estimation of the pipeline. The response of the seasonal system follows the response of the (N,N) model but exhibiting significant oscillations. Troughs are observed at the frequency corresponding to the seasonality parameter and its harmonics.*

From the analysis of the systems' frequency response, we conclude that the (N,A) model outperforms the (N,N) model at the model seasonality frequency and its harmonics, at the cost of increased sensitivity to other demand frequencies. Given that real life demands seldom resemble a pure sinusoid, we next test the response of the system to stochastic demand streams with different seasonalities both in terms of seasonality frequency and seasonality strength.

### 3.3. Computational Experiments

We run a series of computational experiments to better understand the response of the system to demands that better approximate what could be observed in real life. We use Equation (9) with a Gaussian i.i.d. error term of mean 0 and unity variance to model a stochastic, seasonal demand. To test the robustness of the (N,A) model to small mispecifications of the demand seasonality, we compute the performance of the system when the model seasonality is optimal, i.e.,  $\{\rho = P = 50\}$  as well as under two

parameter misestimations, specifically by letting  $\{\rho = 50 \neq P = 51\}$  and  $\{\rho = 50 \neq P = 49\}$ . Note that these values are equivalent to streams with weekly observations and a periodicity 49 and 51 weeks, respectively. For completeness, we also compute the performance of the (N,A) model when the underlying demand is not seasonal, i.e.,  $\{\rho = 50 \neq P = 0\}$ . For all base experiments, we set the amplitude of the seasonality as  $A = 1$ , which is equivalent to a seasonality strength  $F_s = 1/3$ . We initialize the first smoothed values in all experiments by using zero values, i.e.,  $s_0 = j_0 = 0$ . We allow for an ‘initialization period’ of 1000 periods for experiments using stochastic demand streams (Sections 3.3 and 3.2.2) and 704 periods for experiments using the M5 data (Section 3.5). Prior work has shown that in the vast majority of cases, the forecasting accuracy of smoothing models is not affected by the initial values, provided long enough time-series (see Makridakis and Hibon, 1991). (We confirm that this is the case for each of the demand streams we use in this study.)

We coded our experiments directly from the transfer functions, using Mathematica 12 with System Modeler 5.2, and run all experiments in a MacBook Pro with a quad-core Intel i7 processor with 2.3 Ghz and 16Gb of memory. We ran each experiment for 100000 time periods and computed all empirical performance metrics directly from the time-series.<sup>2</sup> Each of the stochastic demand streams were generated with a random number seed. The same demand stream was used for every parameterization to ensure that results are structural and not due to random artifacts.

### 3.3.1. *Response to Stochastic Seasonal Demands*

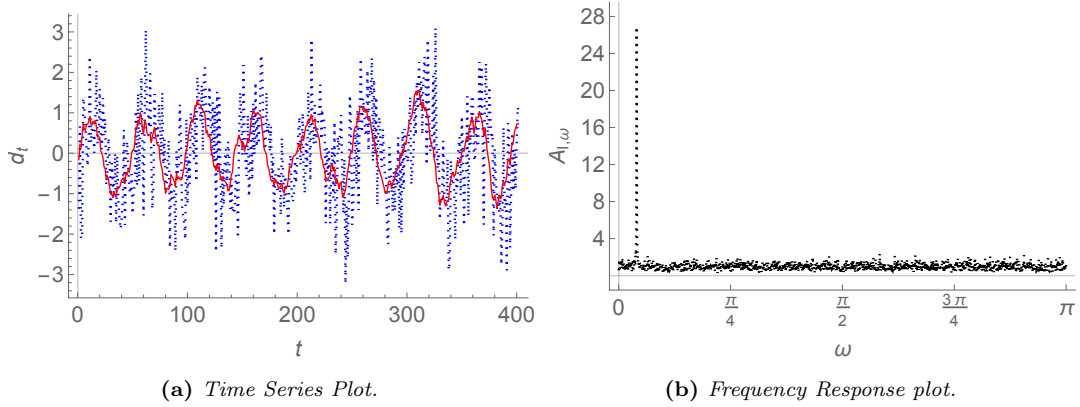
From Section 3.2, we know that the (N,N) model responds with relative robustness to all frequency components whereas the (N,A) model attenuates the amplification at the model seasonality frequency (i.e., when  $\rho = P$ ) but is sensitive in its response to other frequencies.

In this section, we further quantify these observations by studying the behavior of the system under stochastic seasonal demands. Such demands—as defined by Equation (9)—result from adding a deterministic cyclical component to a stochastic component; therefore they exhibit a peak at the dominant cycle frequency, with all other

---

<sup>2</sup>Specifically, for each experiment we ran 20 iterations of 6000 periods each, discarding the first 1000.

frequencies contributing randomly. This is illustrated in Figure 4. The left plot shows a 400 period sample of a seasonal stochastic demand stream with a dominant cycle of 50 periods (dotted blue line) and its 10 period moving average (solid red line); the right plot shows the frequency response of this demand stream. The Time Series plot (Figure 4a) shows a clear seasonality pattern, which is reflected on the FR plot (Figure 4b) as a substantial peak at the corresponding frequency ( $\omega = 2\pi/50$ ).



**Figure 4.** Time (left) and FR (right) plots of a stochastic, seasonal demand with  $P = 50$ . **Figure 4 Alt Text:** This figure shows a stochastic time-series generated using Equation (9) with seasonal periodicity,  $P = 50$ , and its corresponding frequency response plot. The left-side plot shows the original time-series as a dashed line and a 10 period moving average as a solid line. The right-side plot shows that the frequency response has a clear peak at the seasonality frequency.

Table 2 shows the results of  $BW_O$  and  $BW_I$  for the computational experiments performed on the base system (i.e., without behavioral biases) under demands with different seasonality. For each of the demand streams, the best performing model is indicated in bold.

**Table 2**  $BW_O$  and  $BW_I$  baseline results (no behavioral bias).  $F_s = 1/3$ ,  $\rho = 50$ .

Demand Periodicity	(N,N) ( $\delta = 0$ )	(N,A) ( $\delta = 0.1$ )	(N,A) ( $\delta = 0.3$ )	(N,A) ( $\delta = 0.5$ )	(N,A) ( $\delta = 0.9$ )
Order Variability ( $BW_O$ )					
$P = 0$	<b>4.60</b>	4.95	5.75	6.71	9.40
$P = 51$	<b>4.08</b>	4.48	5.50	4.97	6.47
$P = 50$	4.10	<b>3.63</b>	4.33	5.01	6.79
$P = 49$	<b>4.13</b>	4.28	4.71	5.27	6.98
Inventory Variability ( $BW_I$ )					
$P = 0$	<b>15.68</b>	16.48	18.41	20.72	27.25
$P = 51$	<b>23.95</b>	32.08	79.92	100.30	31.38
$P = 50$	24.31	<b>13.19</b>	15.96	17.43	21.20
$P = 49$	24.67	22.35	20.43	<b>20.14</b>	22.93

As expected, the (N,A) model outperforms the (N,N) model when the demand seasonality corresponds exactly to the seasonality assumed in the forecast. The performance, however, also depends on the seasonality parameter—the bullwhip metric increases in  $\delta$ . When the seasonality of the model is misspecified (i.e.,  $\rho \neq P$ ) by one period, we observe that the performance of the (N,A) deteriorates significantly. Note, however, that the system appears to be more robust to misspecification in terms of the bullwhip for orders than to the bullwhip for inventory. In fact, the (N,N) model outperforms the (N,A) in terms of  $BW_O$ , and outperforms, or is comparable, in terms of  $BW_I$ . With regards to the latter, the misspecification of  $P$  can introduce significant performance deterioration (up to 6 times worse in this case).

### 3.3.2. Influence of Behavioral Policies

To understand the influence of behavioral policies on the performance, we repeat the numerical experiments for cases in which a behavioral bias is present, i.e., the pipeline is either under- or over-estimated. We use the same behavioral parameter specification as in Section 3.2.2. Namely,  $\gamma_I = 0.38$ ,  $\gamma_P = 0.15$  for the under-estimation of the pipeline and  $\gamma_I = 0.31$ ,  $\gamma_P = 0.69$  for the over-estimation of the pipeline. Tables 3 and 4 show, respectively, the results for each of the models with under- and over-estimation of the pipeline. Bold indicates the best performing model per stream.

**Table 3**  $BW_O$  and  $BW_I$  behavioral results: Over-estimation of the pipeline.  $F_s = 1/3$ ,  $\rho = 50$ .

Demand Periodicity	(N,N) ( $\delta = 0$ )	(N,A) ( $\delta = 0.1$ )	(N,A) ( $\delta = 0.3$ )	(N,A) ( $\delta = 0.5$ )	(N,A) ( $\delta = 0.9$ )
Order Variability ( $BW_O$ )					
$P = 0$	<b>4.59</b>	4.89	5.61	6.47	8.89
$P = 51$	<b>3.86</b>	4.19	5.06	4.92	6.27
$P = 50$	3.86	<b>3.53</b>	4.21	4.81	6.40
$P = 49$	<b>3.87</b>	4.03	4.44	4.97	6.52
Inventory Variability ( $BW_I$ )					
$P = 0$	<b>10.63</b>	11.13	12.35	13.82	18.00
$P = 51$	<b>22.45</b>	32.94	98.41	129.00	27.28
$P = 50$	22.59	<b>10.31</b>	10.41	11.03	13.32
$P = 49$	22.68	19.02	15.64	<b>14.35</b>	15.16

As expected, the (N,A) model outperforms the (N,N) model in both metrics when  $P = 50$ . Similar to the baseline experiments, however, the (N,N) model shows greater

**Table 4**  $BW_O$  and  $BW_I$  behavioral results: Under-estimation of the pipeline.  $F_s = 1/3, \rho = 50$ .

Demand Periodicity	(N,N) ( $\delta = 0$ )	(N,A) ( $\delta = 0.1$ )	(N,A) ( $\delta = 0.3$ )	(N,A) ( $\delta = 0.5$ )	(N,A) ( $\delta = 0.9$ )
Order Variability ( $BW_O$ )					
$P = 0$	<b>4.53</b>	4.82	5.53	6.40	8.82
$P = 51$	<b>4.07</b>	4.45	5.51	4.84	6.02
$P = 50$	4.10	<b>3.57</b>	4.16	4.77	6.39
$P = 49$	<b>4.14</b>	4.24	4.58	5.07	6.59
Inventory Variability ( $BW_I$ )					
$P = 0$	<b>35.78</b>	37.51	41.77	46.93	61.47
$P = 51$	<b>36.93</b>	44.45	84.99	101.00	52.84
$P = 50$	37.40	<b>27.19</b>	32.42	35.93	45.02
$P = 49$	37.92	<b>36.73</b>	36.80	38.59	46.85

robustness when demand periodicity is misspecified. This results, again, in (N,N) outperforming (N,A) in terms of  $BW_O$ , and in a relatively flat response in terms of  $BW_I$ . In contrast, the  $BW_I$  of the (N,A) model is extremely sensitive to the misspecification.

### 3.4. Robustness Checks

In this section, we test the sensitivity of the system to changes in (i) the forecasting parameters,  $\alpha$  and  $\delta$ ; (ii) the strength of the seasonality,  $F_S$ ; and (iii) the periodicity of the model seasonality,  $P$ . For space considerations, all robustness checks are presented for the baseline (unbiased) system, since the insights derived from behaviorally biased parameter settings are similar.

#### 3.4.1. Forecasting Smoothing Parameters

Comparing the performance of the (N,A) and (N,N) models with a fixed forecasting level-smoothing parameter ( $\alpha = 0.3$ ), sweeping over the forecasting seasonality-smoothing parameter,  $\delta \in \{0.1, 0.3, 0.5, 0.9\}$  allows us to understand the behavior of the (N,A) model vis-à-vis an equivalent (N,N) model, and gives an indication of the sensitivity of the different parametrizations with respect to a misspecification of the demand seasonality parameter,  $P$ . We must note, however, that this comparison does not say anything about the *best* performance of each of the models. To this end, we now present a series of results to quantify the performance of the (N,A) and (N,N) models under an optimal parameterization for a stochastic demand of seasonality  $P = 50$ .

We used a grid search to find the smoothing parameters  $(\alpha^*, \delta^*)$  that minimize the mean squared error of the forecast in each of the models to a precision of two decimal points. To conduct the search, we bounded the parameters such that  $0 \leq \alpha \leq 1$  and  $0 \leq \delta \leq 1$  and sampled equally-spaced values at a resolution of 0.01. Moreover, for the (N,A) model, we assumed perfect knowledge of the seasonal frequency; thus, we set  $P = 50$ . In all cases, we initialized the forecasts by setting the first smoothed values to zero and running a 1000-period long initialization, which we discarded prior to calculating the performance metrics. Table 5 shows the optimal parameter settings, the mean squared error of each forecast, their standard deviation, as well as the resulting bullwhip metrics.

**Table 5** *Summary of optimal forecasting parameter settings.*

		(N,N)	(N,A)
Optimal level smoothing	$(\alpha^*)$	0.26	0.01
Optimal seasonality smoothing	$(\delta^*)$	-	0.02
Root Mean Squared Error of the forecast	(RMSE)	0.89	0.83
Standard deviation of forecast errors	$(\sigma^{RMSE})$	0.67	0.63
Bullwhip of Orders	$(BW_O)$	3.57	0.67
Bullwhip of Inventory	$(BW_I)$	24.66	5.52

As expected, the (N,A) model outperforms the (N,N) model by achieving lower forecast error and standard deviation than the (N,N) model in the order of 10%. From an inventory-control perspective, this indicates that the safety stock required to achieve any given service level is similar for both models. However, the difference in performance, as measured by the bullwhip of orders and inventories is significant; the (N,A) model exhibits around 3 times lower bullwhip.

### 3.4.2. Seasonality Strength

To understand the sensitivity of the system's behavior under different seasonality strengths,  $F_S$ , we re-ran the baseline experiments (see Table 2) using stochastic demand streams of high seasonality, i.e.,  $F_S = 2/3$  and low seasonality, i.e.,  $F_S = 1/6$ . To be able to compare the results, we set the parameters that define the demand stream,  $A$  (amplitude of the sinusoid) and  $\sigma(\epsilon(t))$  (standard deviation of the Gaussian noise) such that, in addition to achieving the aforementioned values of seasonality strength, the total demand variability is equal to that of the initial set of experiments; namely,

$\text{Var}[D] = 1.5$ .

For the high seasonality demand stream, we set the standard deviation of the Gaussian noise  $\sigma(\epsilon(t)) = \sqrt{2}/2$  and the amplitude of the sinusoid,  $A = \sqrt{2}$ , resulting in  $F_S = 2/3$  and  $\text{Var}[D] = 1.5$ . For the low seasonality strength, we set  $\sigma(\epsilon(t)) = \sqrt{1.25}$  and  $A = \sqrt{2}/2$ , resulting in  $F_S = 1/6$  and  $\text{Var}[D] = 1.5$ . The results are shown in Tables 6 (low seasonality) and 7 (high seasonality). For each of the demand streams, the best performing model is indicated in bold.

These results suggest that, the higher the strength of the seasonality the more valuable the usage of the (N,A) model becomes. Indeed, when seasonality strength is low, the (N,A) model outperforms the (N,N) in terms of  $BW_O$  by less than 2%, whereas the difference is over 20% when seasonality is high. Moreover, the previous insights regarding the robustness of the (N,N) model appear to hold under varying seasonality conditions: it performs well under a large range of demand periodicities. In terms of the  $BW_I$  metric, all the prior insights hold, with the addition that the higher the seasonality, the higher the sensitivity of the (N,A) model is.

**Table 6**  $BW_O$  and  $BW_I$  results for low seasonality strength ( $F_s = 1/6$ ).

Demand Periodicity	(N,N) ( $\delta = 0$ )	(N,A) ( $\delta = 0.1$ )	(N,A) ( $\delta = 0.3$ )	(N,A) ( $\delta = 0.5$ )	(N,A) ( $\delta = 0.9$ )
Order Variability ( $BW_O$ )					
$P = 51$	<b>4.34</b>	4.71	5.62	5.84	7.93
$P = 50$	4.35	<b>4.28</b>	5.03	5.85	8.09
$P = 49$	<b>4.36</b>	4.61	5.22	5.99	8.18
Inventory Variability ( $BW_I$ )					
$P = 51$	<b>19.84</b>	24.31	49.23	60.67	29.33
$P = 50$	20.03	<b>14.79</b>	17.16	19.05	24.20
$P = 49$	20.21	<b>19.42</b>	19.42	20.43	25.09

However, our results also show that the bullwhip for orders decreases when the strength of the seasonality increases. Therefore, even though the (N,A) model becomes more attractive—in terms of its relative performance against the (N,N) model—when seasonality increases, the absolute impact of the bullwhip effect itself decreases. This is consistent with prior empirical (e.g., Cachon et al., 2007; Chen and Lee, 2012) and simulation-based (e.g., Bayraktar et al., 2008; Costantino et al., 2016) observations. Note, however, that this comes at the cost of an increased bullwhip of inventories.

**Table 7**  $BW_O$  and  $BW_I$  results for high seasonality strength ( $F_s = 2/3$ ).

Demand	(N,N)	(N,A)	(N,A)	(N,A)	(N,A)
Periodicity	( $\delta = 0$ )	( $\delta = 0.1$ )	( $\delta = 0.3$ )	( $\delta = 0.5$ )	( $\delta = 0.9$ )
Order Variability ( $BW_O$ )					
$P = 51$	<b>3.60</b>	3.82	4.47	3.55	4.06
$P = 50$	3.63	<b>2.98</b>	3.31	3.59	4.39
$P = 49$	3.65	<b>3.63</b>	3.68	3.84	4.57
Inventory Variability ( $BW_I$ )					
$P = 51$	<b>32.86</b>	37.87	81.82	99.65	27.01
$P = 50$	33.19	19.12	18.02	<b>16.88</b>	16.88
$P = 49$	33.56	28.17	22.42	19.54	<b>18.56</b>

### 3.5. Implications for real-life applications

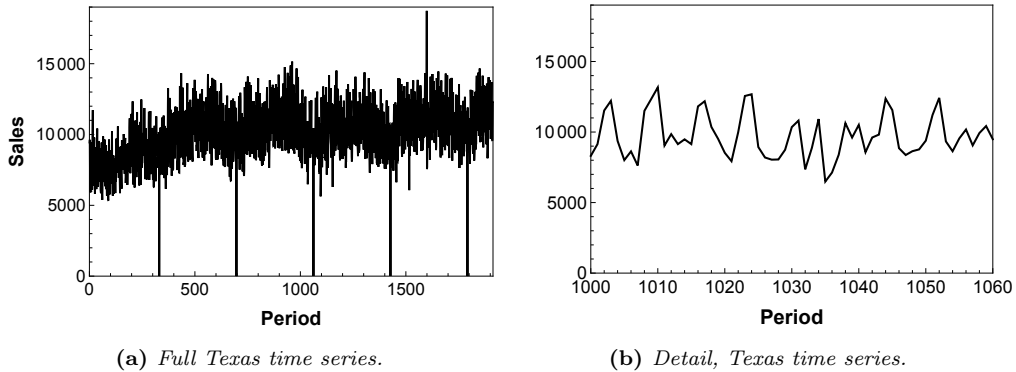
Thus far, we have analyzed the performance of the system from a theoretical perspective (through frequency response analysis) and from a numerical perspective, using stochastic-seasonal demands. The abstract nature of the theoretical analysis and the artificiality of the numerical experimentation leave open the question whether the insights uncovered are relevant in a real-life environment. In other words, are the behaviors uncovered in Sections 3.2.1 through 4 still observed when the demand stream is not artificially generated?

As discussed in Section 3.1, we use data series taken from the M5 forecasting competition, based on 5 years of daily sales data. In summary, the source of the M5 data are daily sales of the American retailer Walmart. In particular, 30490 hierarchically organized data-series of 3049 products were made available, presented at 12 different aggregation levels. Each series contains 1913 daily observations, extracted from January 2011 to April of 2016. Given that the individual series show erratic, lumpy, intermittent, and smooth characteristics, we decided to use two time-series at a relatively high level of aggregation; aggregation by State (aggregation level 2 out of 12)—specifically, the reported time-series from Texas and California. Sales at this level contain no zero-sales days and they exhibit a pronounced, constant, seasonality pattern without a significant trend—making the time series a prime candidate for (N,A) forecasting.

Figure 5 shows the full data of the Texas time-series as well as a 60-day excerpt. The full data clearly shows a yearly seasonal pattern, with high-sales coincident with



holiday periods, whereas the excerpt shows strong weekly seasonality and a more nuanced, but present monthly pattern. There also appears to be a positive trend in the first year of data. For space considerations, and given the similarity of both time series, we only report the results of the analysis conducted with the Texas series—the results using the California series are remarkably consistent.

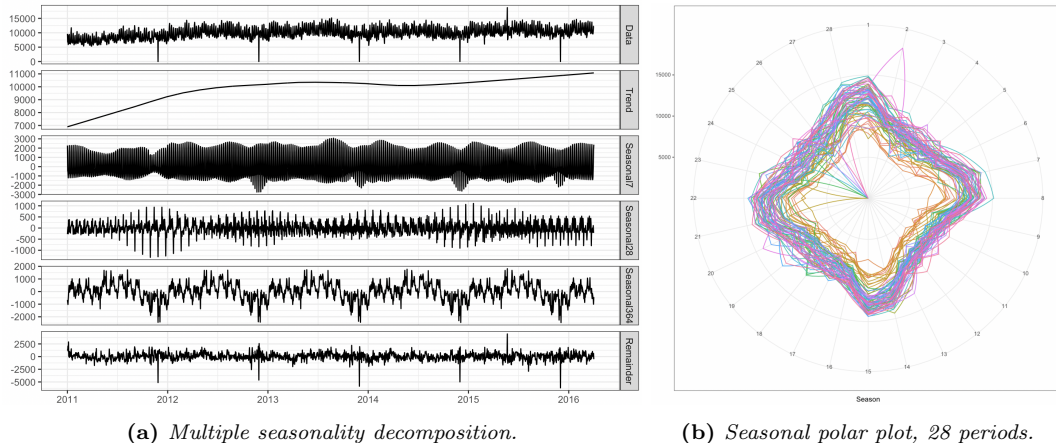


**Figure 5.** Walmart aggregate sales at the state (Texas) level, data from the M5 competition. *Figure 5 Alt Text:* This figure shows the “Texas” time series corresponding to the state-level aggregation of the M5 data. The left-side of the figure shows the complete 1913 periods and the right-side shows a detail of 60 periods. Annual and weekly seasonal patterns are clearly observed.

Given that one of the take-aways from the numerical analysis is the sensitivity of the (N,A) method to the seasonality in the data, we use a number of established methodologies to detect, in an objective manner, the dominant periodicity in the data. In particular, we analyze the frequency periodogram of the demand stream as well as its autocorrelation series to determine the dominant seasonality. These analyses confirm the qualitative assessment of the time-series plot; the time series has strong seasonality with periods of 7, 28, and 364 days. (Note that even though the series exhibit multiple periodicities, the fact that they are multiples of each other simplifies the analysis.) Using these periodicities, we take the approach suggested by Gardner Jr and McKenzie (1988) to determine the most appropriate exponential smoothing model for the time series and find that, despite the initial trend, an (N,A) model with a periodicity of 28 is recommended.

Figure 6a and 6b illustrate the seasonality present in the Texas time series. Both plots confirm the strong weekly, monthly, and yearly seasonality. Moreover, we see that the weekly seasonality is such that weekends exhibit a clear peak in sales, with

those at the beginning of the month having marginally higher sales.

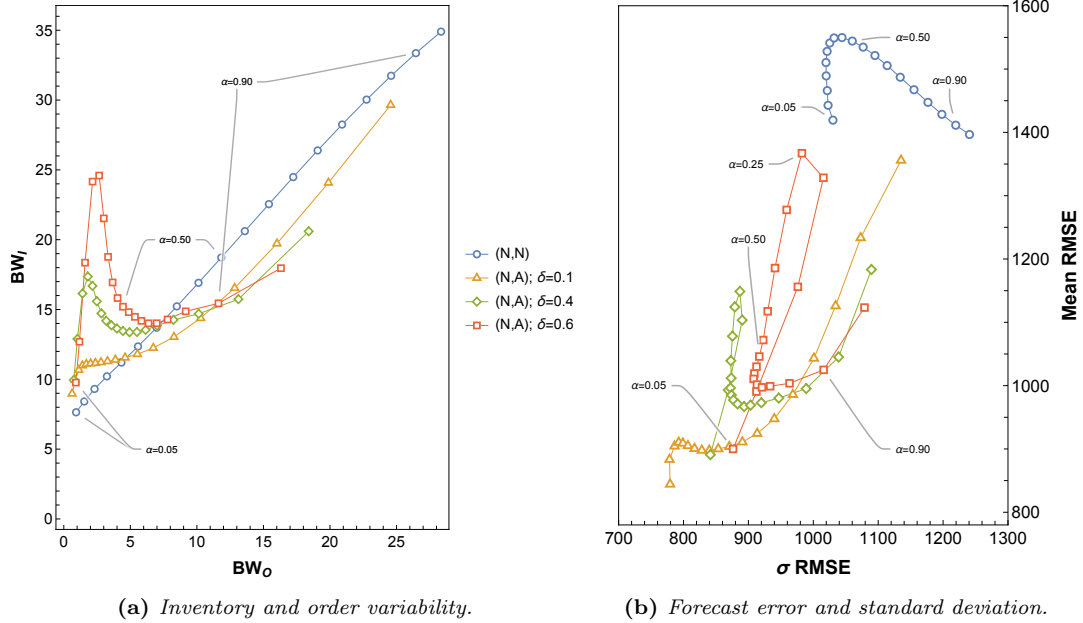


**Figure 6.** Seasonal components of the Texas M5 aggregate series.

**Figure 6 Alt Text:** This figure shows the multiple seasonality decomposition of the “Texas” aggregate M5 series on the left-side, and the corresponding polar seasonal plot on the right-side. The seasonal decomposition shows clear weekly (7-day), monthly (28-day), and yearly (364-day) seasonalities. The polar plot is constructed using a 28-seasonality period and shows a clear pattern of high weekend-sales.

Figure 7 summarizes the order, inventory, and forecast performance of the (N,N) and (N,A) models. We use a correctly specified model, with  $\rho = P = 28$  and plot the results for  $\delta = 0.1, 0.4, 0.6$ . We use the first 704 periods to initialize the forecasts and use the remaining 1029 periods to calculate all performance metrics. Looking at forecasting performance as presented in Figure 7b, it’s clear that the (N,A) model is capable of superior performance; in fact, the best (N,A) parametrization outperforms the best (N,N) parametrization by the order of 40% for RMSE and 25% for  $\sigma$  RMSE. In spite of this, however, we can also clearly see that the (N,A) model is very sensitive to the smoothing parameter  $\alpha$ , and that the performance is non-linear and erratic. This sensitivity increases in  $\delta$ .

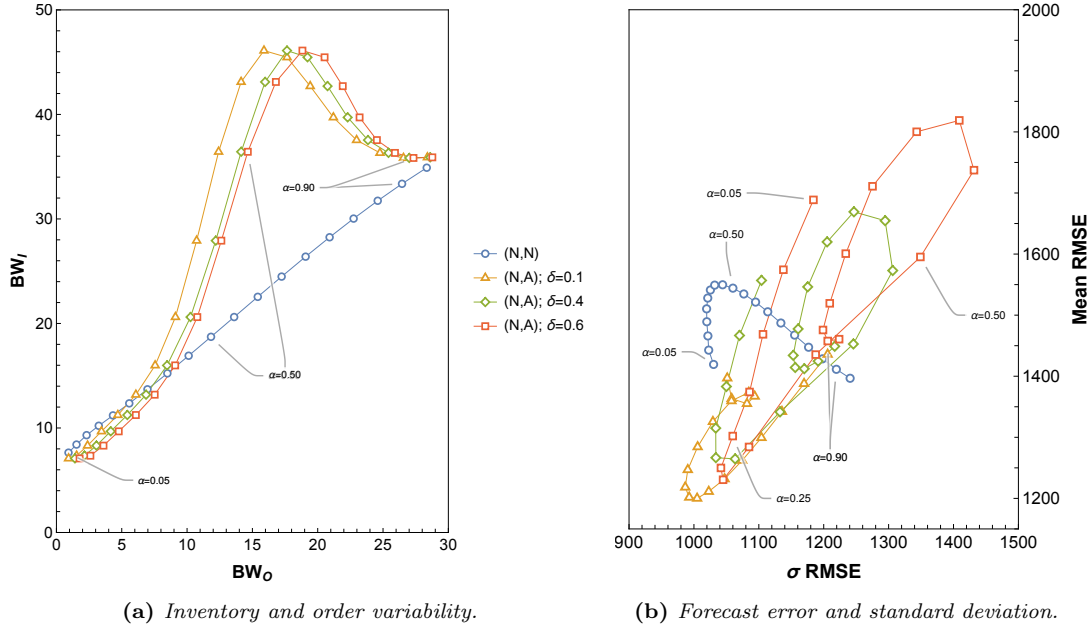
The inventory and order variability plot (Figure 7a) tells a similar story regarding the sensitivity of the models; the response of the (N,N) model appears linear in  $\alpha$ , whereas the response of the (N,A) model is highly non-linear. We also see that the superiority of (N,A) is not as obvious. In terms of order variability, the (N,A) model generally outperforms the (N,N) model for a given value of the smoothing constant  $\alpha$ , however, when we look at the inventory variability, the opposite is true. In fact, in terms of  $BW_I$ , the (N,N) system consistently outperforms the (N,A) system for  $\alpha$



**Figure 7.** Performance metrics when seasonality is correctly specified;  $\rho = P = 30$ .  
**Figure 7 Alt Text:** The left-side figure plots the bullwhip of orders against the bullwhip of inventories for the (N,N) model and (N,A) models of varying smoothing parameters  $\alpha$  and  $\delta$  when the seasonality is correctly specified, i.e.,  $\rho = P$ . We see that the (N,N) model shows a quasi-linear relationship in terms of the smoothing parameter  $\alpha$  whereas the (N,A) model exhibits a strongly non-linear performance. The right-side plot shows the standard deviation of the forecast error against the forecast error for the same models and parameters as the left-side plot.

around 0.3 and lower. These results complement our previous findings and highlight the importance of using operational metrics to choose parameter settings; throughout this paper we have seen that, under different demand settings, the model/parameter combination that minimizes forecast errors will not necessarily minimize the bullwhip of orders, nor the bullwhip of inventories.

To conclude our analysis, we present the results of using a misspecified (N,A) seasonal model. Figure 8 shows the performance metrics of the (N,N) model and an (N,A) model with  $P = 30$ . (The choice of the ‘wrong’  $P$  here is arbitrary and follows the logic that  $P = 30$  would be a reasonable (wrong) approximation of a monthly season.) From the perspective of forecast errors, the misspecified (N,A) model: (i) is sensitive to parameter settings, and (ii) has lost the performance advantage that the correctly specified (N,A) model had. The order and inventory behavior of the model, however, shows that the inventory and order variability at low values of  $\alpha$  (i.e.,  $\alpha \leq 0.3$ ) is much more robust, whereas its performance at higher values of  $\alpha$  is much more sensitive. An



**Figure 8.** Performance metrics when seasonality is misspecified;  $\rho = 28$  and  $P = 30$ .  
**Figure 8 Alt Text:** The left-side figure plots the bullwhip of orders against the bullwhip of inventories for the (N,N) model and (N,A) models of varying smoothing parameters  $\alpha$  and  $\delta$  when the seasonality is not correctly specified, i.e.,  $\rho \neq P$ . We see that the (N,N) model shows a quasi-linear relationship in terms of the smoothing parameter  $\alpha$  whereas the (N,A) model exhibits a strongly non-linear performance. The right-side plot shows the standard deviation of the forecast error against the forecast error for the same models and parameters as the left-side plot.

analysis of the evolution of orders and inventories suggests that the “wrong” seasonality causes a counter-cyclical evolution of the system, which in turn stabilizes the inventory dynamics.

#### 4. Conclusions

In this paper, we developed an extension of the APVIOBPCS model using a forecasting methodology that takes into account no-trend, additive-seasonality in the demand, the (N,A) model. Compared to the simple exponential smoothing (N,N) model, all algebraic manipulations of the frequency-domain representation of the (N,A) model are much more tedious, and even intractable for generic demand seasonality parameters. We have, however, derived the transfer functions for orders and inventory for an array of parametrizations of the model and used these to construct frequency response plots, perform numerical experiments using stochastic seasonal demands, and conduct an

empirical investigation of the behavior of the production/inventory system under real-life demands. With these, we characterized different aspects of the performance of the (N,A) and (N,N) models and analyzed the sensitivity of the results to different parameters.

Our most important findings are as follows:

- (1) When demand is a deterministic cycle (i.e., a sinusoid) of known frequency, the (N,A) model outperforms the (N,N) model by a significant margin.
- (2) When demand is seasonal, stochastic, and of a known frequency, the (N,A) model outperforms the (N,N) model. How much better the (N,A) model performs is proportional to the strength of the demand seasonality; more precisely, the larger the seasonality, the larger the advantage of using the (N,A) model.
- (3) The (N,N) model is more robust to misspecifications of the demand seasonality; when demand frequency is not known with certainty, the (N,N) model outperforms the (N,A) model in the majority of set-ups. This sensitivity depends on the expected demand frequency and the strength of its seasonality. Additional experimentation, provided in the appendices, suggests that these insights hold for different model seasonalities and less restrictive demand assumptions.
- (4) All the insights related to the influence of behavioral biases on the model are consistent for both (N,A) and (N,N) models.
- (5) Our theoretical and numerical findings are consistent with observations made using the system with real-life demand streams. In particular, we observe that parameter choices that minimize the forecast error do not necessarily minimize order or inventory variability, and vice versa. This reinforces calls from prior researchers in the field for decision-makers to consider performance metrics relevant to their business when optimizing their inventory models.

These findings confirm, and extend, the empirical results of Petropoulos et al. (2019) who show that the real-world performance (i.e., inventory performance) of the (N,A) model is sub-par. The overall message is that the (N,A) model excels when parameters are properly specified, and thus managers should consider it if their firms have the capability to properly analyze their demands and maintain their models accordingly.

Otherwise, it's hard to justify the added complexity of the forecasting mechanism. In the end, as Spiliotis et al. (2021) state in a recent analysis of forecasting methods used in the M5 competition, “the value added of more sophisticated methods needs to be carefully examined based on their impact on actual inventory management”.

From a theoretical perspective, our study further contributes to the control-theoretic literature by quantifying the robustness of these two forecasting methods and by providing a first analysis on the mechanisms behind their performance. Our numerical results can be directly used to estimate the performance of two “typical” parameterizations of the (N,A) model (i.e., weekly and yearly model seasonality) under any arbitrary demand stream; and the explicit formulation of the transfer functions, Equations (6) through (8), can be used to obtain the frequency responses of any other specific parameterization.

Based on our findings, we see several promising avenues for future research. First, our theoretical analysis could be extended to include more complex forecasting models, for example (A,A) models with a trend component. We expect, however, that closed-form solutions of the transfer functions to be intractable and thus such work would have to be simulation-based. Second, it's important to note that a limitation of this study is the stylized way in which seasonality, and misspecifications of seasonality, are modelled. It's possible that more complex (and potentially more serious) forms of seasonal misspecification could be encountered in real-life. Therefore, our approach could be used as a basis to evaluate the inventory performance of the different forecasting methods under demand streams with erratic periodic demands, which would in turn generate complex misspecification patterns—for example demands with repeating seasons of unequal length—that typically require much more complex forecasting models, e.g., models with multiple and/or time-varying seasonality parameters. On the one hand, the sensitivity of the (N,A) to the seasonality parameter suggests that such methods best be avoided under complex demand patterns, but on the other hand, its complex non-linear response could be exploited within an inventory system to be able to consistently achieve good performance in terms of inventory availability/variability in spite of sub-optimal results from the point of view of “pure” forecast performance.

## Data Accessibility Statement

The data generated to conduct the experiments of this study are freely accessible online at <http://www.mudenio.net/data> or by request to the authors. The empirical data from the M5 forecasting competitions are openly available in Kaggle at <https://www.kaggle.com/c/m5-forecasting-accuracy/data>.

## Acknowledgements

The authors thank the anonymous associate editor and four reviewers for their guidance and constructive suggestions during the review process. Our paper is much improved thanks to them.

## References

- Babai, M Zied, John E Boylan, Bahman Rostami-Tabar. 2021. Demand forecasting in supply chains: a review of aggregation and hierarchical approaches. *International Journal of Production Research* 1–25.
- Bayraktar, Erkan, SC Lenny Koh, A Gunasekaran, Kazim Sari, Ekrem Tatoglu. 2008. The role of forecasting on bullwhip effect for e-scm applications. *International Journal of Production Economics* **113**(1) 193–204.
- Bossoms, John. 1966. The effects of parameter misspecification and non-stationarity on the applicability of adaptive forecasts. *Management Science* **12**(9) 659–669.
- Cachon, Gérard P, Taylor Randall, Glen M Schmidt. 2007. In search of the bullwhip effect. *Manufacturing & Service Operations Management* **9**(4) 457–479.
- Cantor, David E, Elena Katok. 2012. Production smoothing in a serial supply chain: A laboratory investigation. *Transportation research part e: logistics and transportation review* **48**(4) 781–794.
- Chen, Li, Hau L Lee. 2012. Bullwhip effect measurement and its implications. *Operations Research* **60**(4) 771–784.
- Costantino, Francesco, Giulio Di Gravio, Ahmed Shaban, Massimo Tronci. 2016.

- Smoothing inventory decision rules in seasonal supply chains. *Expert Systems with Applications* **44** 304–319.
- Dejonckheere, J., S.M. Disney, M.R. Lambrecht, D.R. Towill. 2003. Measuring and avoiding the bullwhip effect: A control theoretic approach. *European Journal of Operational Research* **147**(3) 567–590.
- Disney, S.M. 2008. Supply chain aperiodicity, bullwhip and stability analysis with jury’s inners. *IMA Journal of Management Mathematics* **19**(2) 101–116.
- Disney, S.M., I. Farasyn, M.R. Lambrecht. 2006. Taming the bullwhip effect whilst watching customer service in a single supply chain echelon. *European Journal of Operational Research* **173**(1) 151–172.
- Disney, S.M., D.R. Towill. 2003. On the bullwhip and inventory variance produced by an ordering policy. *Omega* **31**(3) 157–167.
- Gardner Jr, Everette S, Ed McKenzie. 1988. Model identification in exponential smoothing. *Journal of the Operational Research Society* **39**(9) 863–867.
- Gross, D, R Craig. 1974. A comparison of maximum likelihood, exponential smoothing and Bayes forecasting procedures in inventory modelling. *International Journal of Production Research* **12**(5) 607–622.
- Harrison, P J. 1967. Exponential smoothing and short-term sales forecasting. *Management Science* **13**(11) 821.
- Hoberg, K., J.R. Bradley, U.W. Thonemann. 2007. Analyzing the effect of the inventory policy on order and inventory variability with linear control theory. *European Journal of Operational Research* **176**(3) 1620–1642.
- Hoberg, K., U.W. Thonemann. 2014. Modeling and analyzing information delays in supply chains using transfer functions. *International Journal of Production Economics* **156** 132–145.
- Hyndman, Rob, Anne B Koehler, J Keith Ord, Ralph D Snyder. 2008. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media.
- Hyndman, Rob J, George Athanasopoulos. 2018. *Forecasting: principles and practice*. OTexts.
- Hyndman, Rob J, Anne B Koehler, Ralph D Snyder, Simone Grose. 2002. A state



- space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting* **18**(3) 439–454.
- John, S., M.M. Naim, D.R. Towill. 1994. Dynamic analysis of a wip compensated decision support system. *International Journal of Management Systems Design* **14**(4) 283–297.
- Li, Q., S.M. Disney, G. Gaalman. 2014. Avoiding the bullwhip effect using Damped Trend forecasting and the Order-Up-To replenishment policy. *International Journal of Production Economics* **149** 3–16.
- Li, Xixi, Fotios Petropoulos, Yanfei Kang. 2022. Improving forecasting by subsampling seasonal time series. *International Journal of Production Research* 1–17.
- Lin, J, M M Naim, L Purvis, J Gosling. 2016. The extension and exploitation of the inventory and order based production control system archetype from 1982 to 2015. *Intern. Journal of Production Economics* **194** 1–0.
- Makridakis, S, E Spiliotis, V Assimakopoulos. 2020a. The m5 accuracy competition: Results, findings and conclusions. *Int J Forecast* .
- Makridakis, S, E Spiliotis, V Assimakopoulos. 2020b. The m5 competition: Background, organization, implementation and data. Tech. rep., Working paper.
- Makridakis, Spyros, Allan Andersen, Robert Carbone, Robert Fildes, Michele Hibon, Rudolf Lewandowski, Joseph Newton, Emanuel Parzen, Robert Winkler. 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of forecasting* **1**(2) 111–153.
- Makridakis, Spyros, Michèle Hibon. 1991. Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy. *International Journal of Forecasting* **7**(3) 317–330.
- Muth, J F. 1960. Optimal properties of exponentially weighted forecasts. *Journal of the american statistical association* .
- Petropoulos, Fotios, Xun Wang, Stephen M Disney. 2019. The inventory performance of forecasting methods: Evidence from the m3 competition data. *International Journal of Forecasting* **35**(1) 251–265.
- Simon, Herbert A. 1952. On the application of servomechanism theory in the study of production control. *Econometrica: Journal of the Econometric Society* 247–268.

- Spiliotis, Evangelos, Spyros Makridakis, Anastasios Kaltsounis, Vassilios Assimakopoulos. 2021. Product sales probabilistic forecasting: An empirical evaluation using the m5 competition data. *International Journal of Production Economics* **240** 108237.
- Taylor, James W. 2003. Exponential smoothing with a damped multiplicative trend. *International journal of Forecasting* **19**(4) 715–725.
- Towill, Denis R. 1982. Dynamic analysis of an inventory and order based production control system. *The international journal of production research* **20**(6) 671–687.
- Udenio, Maximiliano, Eleni Vatamidou, Jan C. Fransoo, Nico Dellaert. 2017. Behavioral causes of the bullwhip effect: an analysis using linear control theory. *IIE Transactions* **49**(10) 1–21. URL <http://dx.doi.org/10.1080/24725854.2017.1325026>.
- Wang, Xiaozhe, Kate Smith, Rob Hyndman. 2006. Characteristic-based clustering for time series data. *Data mining and knowledge Discovery* **13**(3) 335–364.
- Wang, Xun, Stephen M Disney. 2016. The bullwhip effect: Progress, trends and directions. *European Journal of Operational Research* **250**(3) 691–701.
- Warburton, Roger DH, Stephen M Disney. 2007. Order and inventory variance amplification: The equivalence of discrete and continuous time analyses. *International Journal of Production Economics* **110**(1-2) 128–137.
- Wright, David, Xin Yuan. 2008. Mitigating the bullwhip effect by ordering policies and forecasting methods. *International Journal of Production Economics* **113**(2) 587–597.
- Zhan, Xueli, Qin Zhang, Wenfeng Xie. 2018. Dynamic performance and stability research of vmi-apiobpcs in apparel industry based on control theory. *International Journal of Enterprise Information Systems (IJEIS)* **14**(2) 56–76.

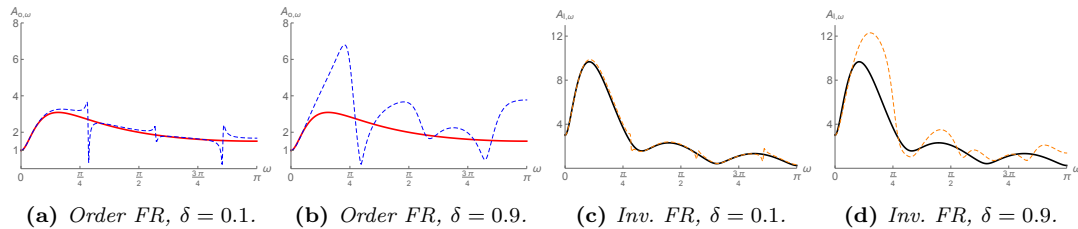
## **Appendix A: The Influence of Model Seasonality**

All theoretical and numerical results from our study are derived based on a baseline seasonality of  $\rho = 50$ . The objective of this Appendix is to investigate the robustness of response of the model to this seasonality parameter,  $\rho$ . In particular, we are interested

to see whether the results discussed are structural, or an artifact of the parameter choices. To this end, Figure 9 shows the frequency response plots for orders and inventories of the (N,A) model with  $\rho = 7$  for  $\delta = 0.1$  and  $\delta = 0.9$ . This represents an (N,A) model that observes demand daily, with a model seasonality of one week.

It is apparent that the sensitivity of the frequency response is directly related to the parameter  $\rho$ . Same as we observed in Figures 1 and 2, the number of troughs and peaks in the responses is proportional to the periodicity of the forecast, with the frequency response for orders and inventories being attenuated at the frequency corresponding to the model periodicity (in this case,  $\omega = 2\pi/7$ ) and its harmonics. This is consistent with prior analyses of frequency response of the moving average forecast, which is sinusoidal in nature, with the number of cycles proportional to the number of periods it averages on (see Dejonckheere et al., 2003).

Similar to the case of  $\rho = 50$ , the (N,A) model is more sensitive than the (N,N) model to the specific frequency and this sensitivity increases in  $\delta$ . This suggests that the insights obtained for the experiments carried out with  $\rho = 50$  hold for other values of  $\rho$ . However, given the lower number of harmonics, one can expect a smoother response to mispecifications in the (N,A) model for lower values of  $\rho$ ; making it more robust to small mispecifications of the demand seasonality parameter.



**Figure 9.** Order and inventory FR of (N,N) (solid) and (N,A) (dashed) models with  $\rho = 7$ . **Figure 9 Alt Text:** This figure shows the frequency response of the transfer function of orders of the (N,N) and (N,A) models under different values of the seasonal smoothing parameter,  $\delta$  with seasonal periodicity,  $\rho = 7$ . The response of the seasonal system follows the response of the (N,N) model but exhibiting significant oscillations. Troughs are observed at the frequency corresponding to the seasonality parameter and its harmonics. The amplitude of oscillations increases in  $\delta$ .

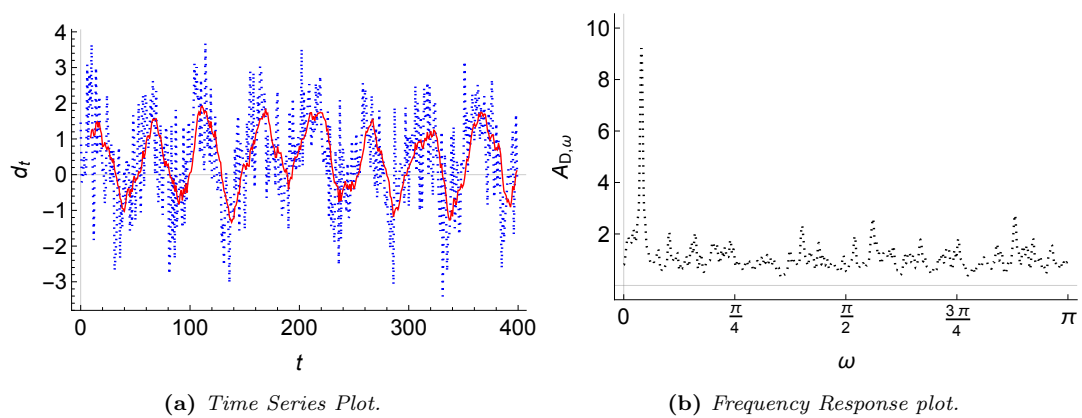
## Appendix B: ARIMA Demands

The objective of this Appendix is to investigate the robustness of the response of the forecasting models to a different stochastic demand generating process. To this end, we replicate the analysis of Section 3.2.1 using the same experimental set-up, with the difference that the input demand is obtained by simulating seasonal ARIMA(0,1,1) processes. As in Section 3.2.1, we ran 20 iterations of 6000 periods for each experiment, discarding the first 1000 to ensure steady-state. The demand process used is illustrated in Figure 10, showing a 400 period sample of a seasonal ARIMA(0,1,1) demand stream with seasonality  $P = 50$ . The Time Series plot (Figure 10a) highlights the seasonal pattern, which is confirmed in on the FR plot (Figure 10b).

Table 8 shows the results of  $BW_O$  and  $BW_I$  for the computational experiments performed on the base system using seasonal ARIMA(0,1,1) demands with different seasonality and a constant seasonality strength,  $F_S = 1/3$ . For each of the demand streams, the best performing model is indicated in bold. These results are consistent with those obtained using demands generated by Equation (9) with an i.i.d. gaussian error. The (N,A) model outperforms the (N,N) model when the demand seasonality corresponds exactly to the seasonality assumed in the forecast model. Results for other experiments are omitted for space, but are similarly consistent. This suggests that the insights obtained in our study are not conditional on the demand generating process and can thus be extended to other demand structures for which the (N,A) model is a reasonable choice.

**Table 8**  $BW_O$  and  $BW_I$  results for Seasonal ARIMA(0,1,1) demands.  $F_s = 1/3$ ,  $\rho = 50$ .

Demand Periodicity	(N,N) ( $\delta = 0$ )	(N,A) ( $\delta = 0.1$ )	(N,A) ( $\delta = 0.3$ )	(N,A) ( $\delta = 0.5$ )	(N,A) ( $\delta = 0.9$ )
Order Variability ( $BW_O$ )					
$P = 51$	<b>1.98</b>	2.14	2.55	2.63	3.10
$P = 50$	2.02	<b>1.58</b>	1.73	1.87	2.21
$P = 49$	<b>2.01</b>	2.04	2.13	2.25	2.62
Inventory Variability ( $BW_I$ )					
$P = 51$	<b>13.53</b>	15.68	27.28	33.33	15.99
$P = 50$	14.01	<b>9.97</b>	10.62	10.98	11.83
$P = 49$	14.27	13.53	12.80	<b>12.54</b>	12.96



**Figure 10.** Time (left) and FR (right) plots of a seasonal ARIMA(0,1,1) stream.  $P = 50$ ,  $F_S = 1/3$ .

**Figure 10 Alt Text:** This figure shows a stochastic time-series generated using a seasonal ARIMA(0,1,1) process with a periodicity,  $P = 50$ , and its corresponding frequency response plot. The left-side plot shows the original time-series as a dashed line and a 10 period moving average as a solid line. The right-side plot shows that the frequency response has a clear peak at the seasonality frequency.