State space ARIMA for supply chain forecasting

Ivan Svetunkov[a,*], John E. Boylan[a]

[a] *Centre for Marketing Analytics and Forecasting*
Lancaster University Management School, Lancaster, LA1 4YX, UK

Abstract

ARIMA is seldom used in supply chains in practice. There are several reasons, not the least of which is the small sample size of available data, which restricts the usage of the model. Keeping in mind this restriction, we discuss in this paper a state space ARIMA model with a single source of error and show how it can be efficiently used in the supply chain context, especially in cases when only two seasonal cycles of data are available. We propose a new order selection algorithm for the model and compare its performance with the conventional ARIMA on real data. We show that the proposed model performs well in terms of both accuracy and computational time in comparison with other ARIMA implementations, which makes it efficient in the supply chain context.

Keywords: Forecasting, state space models, ARIMA, supply chain forecasting, order selection, model selection

1. Introduction

ARIMA has always been considered as a statistically sophisticated and complicated model. Although several forecasting competitions showed that simpler methods perform at least as well as statistically sophisticated methods and sometimes outperform ARIMA (Makridakis et al., 1982; Makridakis and Hibon, 1997, 2000; Athanasopoulos et al., 2011), its popularity among researchers has not declined over the years. ARIMA is considered to be a standard model in the statistical literature and is widely used for analytical

*Correspondance: Ivan Svetunkov, Department of Management Science, Lancaster University Management School, Lancaster, Lancashire, LA1 4YX, UK.

Email address: i.svetunkov@lancaster.ac.uk (Ivan Svetunkov)

derivations in the supply chain literature (an extensive review of supply chain forecasting is given in Syntetos et al., 2016). Examples of the application of ARIMA in a supply chain context include Kim et al. (2003), Wang et al. (2010), Hosoda et al. (2008), Disney et al. (2006), Doganis et al. (2008), Svetunkov and Petropoulos (2018), van Gils et al. (2017) and Dellino et al. (2018).

Nevertheless ARIMA is not as widely used in practice as simpler methods, such as exponential smoothing and simple moving averages (Winklhofer et al., 1996; Weller and Crone, 2012). The reason is the complexity of the model. On the one hand it is not always simple to identify the appropriate orders of ARIMA and estimate the model. On the other hand, it is much harder to explain the model to supply chain managers than, for example, exponential smoothing. Furthermore, it is very common for companies working in business to have small samples of data, because managers think that the older data is not useful and not relevant to recent history. As a result companies very often have at most 3 years of data. This makes seasonal ARIMA models hard to build, because of estimation problems. Indeed, in order to estimate the simplest conventional seasonal ARIMA, a forecaster needs at least 3 years of data, where the first year is sacrificed for initialisation of the model and the last two are needed for model fitting. Having less than 3 years means that the model will overfit the second season and inevitably will produce poor forecasts. Furthermore, in order to include ARIMA in appropriate forecasting evaluation against simpler forecasting methods, the sample needs to be split into training and test sets. This further decreases the number of observations available for estimation purposes, making conventional seasonal ARIMA inapplicable.

Having limited data in the training set also means that parametric statistical tests may be inaccurate because of their low power on small samples. This additional complication means that unit root tests and tests for seasonality may be unreliable, which in turn leads to problems in the identification of the correct order of ARIMA.

Finally, a typical forecasting task for the supply chain involves producing forecasts for a large dataset with thousands of Stock Keeping Units (SKUs). This means that the forecasting should be done automatically and fast, which is not always the case for ARIMA models, because each time series has its own structure, and the order of ARIMA needs to be selected individually. Order selection for ARIMA is in general slow, because it either implies analysis of Auto Correlation Functions (ACF and PACF), or applying several ARIMA models of different orders to data and selecting the optimal one (using some criterion).

All of this explains the lack of popularity of ARIMA models in applied supply chain forecasting. At the same time interest in ARIMA models has been recently rising, and overcoming the aforementioned limitations could allow using the flexibility of ARIMA for supply chain forecasting. However, this means that supply chain ARIMA should at least satisfy the following requirements:

1. Order selection and model estimation should work with seasonal data on small samples with at least two years of data;
2. Order selection should be done without statistical tests;
3. The order selection algorithm should be fast.

We propose using ARIMA in state space form with a Single Source of Error (originally proposed in Snyder, 1985), which allows meeting all the three requirements. First of all, state space models can be initialised in period zero, which saves some observations and may increase the number of degrees of freedom. Secondly, a state space model allows estimating ARIMA using likelihood and applying model selection based on information criteria for all the possible models without a need for hypotheses testing. The only issue that needs to be addressed is the order selection algorithm, which should be smart, choosing only those orders that are relevant to the data.

In this paper we discuss state space ARIMA and the methodology of order selection and estimation of the model that satisfies all three requirements. The proposed implementation of ARIMA can be efficiently applied to a wide variety of data, and, as we show in the paper, performs well in terms of forecasting accuracy, given the computational time restriction observed in practical supply chains.

2. State space ARIMA

ARIMA in state space form has been known for at least 40 years. Harvey and Phillips (1979) discuss a state space model with multiple sources of errors (MSOE) underlying a general regression with ARMA errors. Pearlman (1980) uses their finding and proposes a modification of the state space model with a single source of error (SSOE). He points out that this model can be used when the AR order is greater than or equal to MA order, but he does not investigate the model further. Snyder (1985) analyses the SSOE state space model and its connection with ARIMA in more detail. He discusses several basic ARIMA models, showing how the model can be formulated using measurement and transition equations. Snyder et al. (2001) discuss ARIMA in state space form and demonstrate how the prediction intervals can be constructed for this model. Finally, a more detailed explanation of the connection between ARIMA and SSOE state space models is given in (Hyndman et al., 2008, pp. 173 - 174). We use their derivations as the basis for our model.

The general form of state space model with SSOE is (Hyndman et al., 2008):

$$\begin{array}{l} y_t = \mathbf{w}'\mathbf{v}_{t-1} + \epsilon_t \\ \mathbf{v}_t = \mathbf{F}\mathbf{v}_{t-1} + \mathbf{g}\epsilon_t \end{array}, \tag{1}$$

where $\mathbf{v}_t$ is the vector of states, ϵ_t is the error term (usually assumed to be distributed normally with zero mean and variance σ^2), $\mathbf{F}$ is the transition matrix, $\mathbf{w}$ is the measurement vector, $\mathbf{w}'$ is the transposed $\mathbf{w}$ and $\mathbf{g}$ is the persistence vector. Hyndman et al. (2008) give general formulae, connecting ARIMA models with their state space counterparts. They derive the state space model for non-seasonal ARIMA without the constant term. Their derivations with minor modifications can be used in order to present the following more general SARIMA$(p,d,q)(P,D,Q)_m$ model (where m is seasonal frequency) in state space form:

$$\phi_p(B)\delta_d(B)\Phi_P(B^m)\Delta_D(B^m)y_t = \theta_q(B)\Theta_Q(B^m)\epsilon_t + \beta, \tag{2}$$

where $\phi_p(B)$ is the non-seasonal AR, $\delta_d(B)$ is the non-seasonal difference, $\theta_q(B)$ is the non-seasonal MA, $\Phi_P(B^m)$ is the seasonal AR, $\Delta_D(B^m)$ is the seasonal differences and $\Theta_Q(B^m)$ is the seasonal MA polynomials, β is the constant term, which in the case of non-zero order of differences acts as drift and B is the backshift operator. We need to note that all the MA polynomials are used in our formulation with a plus sign, while the AR polynomials use the minus sign. So, for example, we formulate ARIMA(1,1,1) as:

$$(1-\phi_1 B)(1-B)y_t = (1+\theta_1 B)\epsilon_t + \beta, \tag{3}$$

where ϕ_1 is AR(1) parameter and θ_1 is MA(1) parameter. By working models in this way we do not cause the confusion with signs of the coefficients.

In order to write ARIMA in state space form, the polynomials in the model (2) need to be expanded:

$$\left(1 - \sum_{j=1}^{K} \varphi_j B^j \right) y_t = \left(1 + \sum_{j=1}^{K} \eta_j B^j \right) \epsilon_t + \beta, \tag{4}$$

where φ_j and η_j are the values of the coefficients for AR and MA polynomials respectively and $K = \max(p+d+P+D, q+Q)$. The max term means that, for example, if $p+d+P+D > q+Q$, then all the η_j for $j > q+Q$ will be equal to zero. A similar property holds for the opposite situation. Regrouping the elements in (4) leads to:

$$y_t = \sum_{j=1}^{K} \varphi_j B^j y_{t-j} + \sum_{j=1}^{K} \eta_j B^j \epsilon_{t-j} + \beta + \epsilon_t. \tag{5}$$

After that the logic of derivation becomes exactly the same as in (Hyndman et al., 2008, pp. 173 - 174) with an exception for the first component of the state space model and an additional component for β. The state space ARIMA model proposed in this paper can be formulated in the following way:

$$\begin{aligned} & y_t = v_{1,t-1} + \epsilon_t \\ & v_{j,t} = \varphi_j v_{1,t-1} + v_{j+1,t-1} + v_{K+1,t-1} + (\varphi_j + \eta_j)\epsilon_t, && \text{for } j = 1 \\ & v_{j,t} = \varphi_j v_{1,t-1} + v_{j+1,t-1} + (\varphi_j + \eta_j)\epsilon_t, && \text{for } 1 < j \leq K \\ & v_{K+1,t} = v_{K+1,t-1}, \end{aligned} \tag{6}$$

where $v_{j,t}$ is the j-th component and $v_{K+1,0} = \beta$. Note that the first and the $K+1$ components are calculated differently than in (Hyndman et al., 2008, pp. 173 - 174), because of the constant term β. If the constant is not needed for a time series, then $v_{K+1,0}$ can be set to zero, and the ARIMA model in state space form (6) becomes equivalent to the one in (Hyndman et al., 2008, p. 174). The model (6) can be written in the compact form (1), with:

$$\mathbf{v}_t = \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ \vdots \\ v_{K,t} \\ v_{K+1,t} \end{pmatrix}, \mathbf{F} = \begin{pmatrix} \varphi_1 & 1 & 0 & \dots & 0 & 1 \\ \varphi_2 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \varphi_K & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \mathbf{w} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \mathbf{g} = \begin{pmatrix} \varphi_1 + \eta_1 \\ \varphi_2 + \eta_2 \\ \vdots \\ \varphi_K + \eta_K \\ 0 \end{pmatrix}. \tag{7}$$

So ARIMA in state space form has $K+1$ components if the constant term is not equal to zero. In cases with seasonal models the matrices in (7) can become large, especially if the seasonality lag m is large and the seasonal orders are high.

One of the advantages of the state space ARIMA model is that the initialisation of the model (6) can be done on observation $t = 0$, which allows preserving observations for estimation purposes. The values of v_0 can be estimated in different ways, the most popular of which are optimisation and backcasting. We propose using the backcasting technique in order to preserve degrees of freedom and minimise the required computations, because then we do not need to estimate all the $K+1$ initial values of the state vector; we only need to optimise the constant β which corresponds to the component $v_{K+1,0}$ and the parameters of the ARMA. Still before constructing the model some preset values for the initial state vector are needed. In order to speed up the convergence to the true value of the initial vector v_0, we use the following heuristics derived from the model (6) (see Appendix A):

$$\begin{array}{ll} v_{1,t-1} = y_t, & \text{for } t = \{1, \dots, K\} \\ v_{2,t-1} = v_{1,t} - \varphi_1 y_t - v_{K+1,0}, & \text{for } t = \{1, \dots, K-1\} \\ v_{j,t-1} = v_{j-1,t} - \varphi_{j-1} y_t, & \text{for } 2 < j \leq K \text{ and } t = \{1, \dots, K-j+1\} \end{array} . \tag{8}$$

In this way, we define $\frac{K(K+1)}{2}$ elements of the first K state vectors. After that the model (6) is applied to the data starting from the $t = 1$ until the last observation T in the sample. Then the reverse state space model is applied:

$$\begin{aligned} y_t &= \mathbf{w}'\mathbf{v}_{t+1} + \epsilon_t \\ \mathbf{v}_t &= \mathbf{F}\mathbf{v}_{t+1} + \mathbf{g}\epsilon_t \end{aligned} \tag{9}$$

until the observation $t = 0$. Then a new initial value of the state vector is obtained and used in the construction of the model using (1). The procedure is repeated several times, refining the initial values of the state vector. In the implementation that we discuss in Section 4, three iterations are sufficient for the initial states to converge.

Having the state space model (6) also solves the problem with application of ARIMA to small samples. While in order to construct the conventional seasonal model it is necessary to have at least three seasonal cycles of data, the model (6) can be constructed even if only two seasonal cycles are available. This is because the initialisation is done on the observation $t = 0$. For obvious reasons the estimates of the parameters on such a small sample can be unreliable and the forecasts may be less accurate than they would be on large samples, but at least some estimates and some forecasts can be produced in this case.

The other important advantage of the model (6), is that all the possible orders of the model can be compared directly with each other using information criteria. Note that ARIMA models in the conventional form can be compared with each other only for pre-specified differences, because taking differences decreases the sample size, automatically leading to incomparable values of information criteria. So there is no need to conduct preliminary unit root tests in order to determine if the time series is stationary or not with the state space ARIMA. There is also no need to test whether the series is seasonal or not, because this can be done by comparing seasonal and non-seasonal ARIMA models in the state space form using an information criterion.

However, taking into account that there are several possible orders in seasonal ARIMA for each of the components of the model, the search of the optimal order can become a cumbersome task. For example, if the maximum orders of the model correspond to SARIMA(3,2,3)(2,1,2)$_m$ for a fixed value of m, then there are 864 potential models. Checking whether the constant β is needed or not, doubles the number of models, giving a pool of 1728 SARIMA models. In order to find a good model that would produce adequate forecasts in a reasonable time, we need to use a smart algorithm for order selection.

3. Order selection in state space ARIMA

In order to select the most appropriate ARIMA, we propose using an information criterion. For example, the Akaike Information Criterion (Akaike, 1974) can be written as:

$$\text{AIC} = 2k - 2\ell, \tag{10}$$

where k is the number of estimated parameters and ℓ is the value of the log-likelihood function extracted from the model.

We propose using the following stepwise order selection algorithm to allow the selection of a good model for the data:

1. All the possible differences are checked with non-zero constant. This includes seasonal and non-seasonal counterparts. In cases of non-zero difference, the constant acts as a drift, allowing the capture of possible trends in time series and model multiplicative seasonality.

2. The residuals of the best model on the step (1) are extracted. All possible types of seasonal and non-seasonal MA are checked. The order is selected via a modified information criterion, where the number of parameters is set to be equal to the sum of all the parameters estimated on the current and the previous steps:

$$\text{AIC}_2 = 2(k_1 + k_2) - 2\ell_2, \tag{11}$$

where the index in the subscript stands for the step in the algorithm, so that the number of parameters in AIC_2 is equal to sum of all the estimated parameters on step 2 and before. ℓ_2 is the value of the log-likelihood function for the model on step 2.

3. The residuals of the best model on the step (2) are extracted. All possible types of seasonal and non-seasonal AR orders are checked. The information criterion on this step uses the sum of all the estimated parameters on steps (1), (2) and (3), substituting k_1+k_2 from (11) with $k_1 + k_2 + k_3$ and ℓ_2 with ℓ_3, the value of log-likelihood function from the model on step 3.
4. The model of the selected orders is re-estimated on the original data in order to remove a potential bias in estimates of parameters.
5. The model (4) is compared with the same model without the constant. The model with the lowest information criterion is then selected for the forecasting purpose.

Note that if some other criterion is preferred, then the formula (11) should be substituted by the desired formula, preserving the number of estimated parameters and using the value of the log-likelihood function extracted for each specific step.

This algorithm allows for a substantial reduction in the pool of models. For example, in the case of SARIMA(3,2,3)(2,1,2)$_m$ only 31 models need to be checked instead of 1728. This does not guarantee that the selected model will have the lowest AIC among all the 1728 potential SARIMA models, but it gives a reasonable model, as will be demonstrated later in the paper, which should suffice for forecasting purposes.

In order to further decrease the pool of the potential models, higher orders of AR or MA can be checked before the lower orders. In this case when the higher order leads to the higher information criterion, then there is no need to check lower orders, meaning that they can be skipped altogether. For example, if the true model is AR(1) and we first compare AR(0) and AR(3), then the latter should have a lower information criterion, as the AR(3) model includes the correct order as a first element $\varphi_1 y_{t-1}$. AR(2) in turn should be better than AR(3) in terms of information criterion, because it does not contain the redundant term $\varphi_3 y_{t-3}$, and finally AR(1) is expected to have the lowest information criterion as it does not contain any redundant terms. If for some data we find that AR(2) has greater information criterion than AR(3), then the check of AR(1) can be skipped. This shortcut allows saving computational time further by decreasing the pool of models.

4. Evaluation of state space ARIMA performance

In order to see how the state space ARIMA performs, we test it in a real time series experiment.

The state space ARIMA with the described order selection algorithm is implemented in `auto.ssarima()` function in `smooth` package version 2.1.1 for R (Svetunkov, 2017). This model is denoted as "SSARIMA" in the experiment. The maximum order of the model was restricted to SARIMA$(3,2,3)(2,1,2)_m$. This restriction is motivated by the following. The differences of the non-seasonal part should not exceed 2 because this might cause over-differencing with the corresponding loss of information (Box and Jenkins, 1976, p.175). Similarly, there is no point in going beyond the first difference of the seasonal part of the model. Given that we deal with short data, we restrict the maximum seasonal orders of AR and MA to 2, which corresponds to two years of data. Finally, the restriction of AR and MA to the maximum order of 3 should be sufficient for such short data (this is similar to Hyndman and Khandakar, 2008, who also investigated automatic model selection).

We have also applied state space ARIMA with optimised initials using `auto.ssarima()` function (denoted "SSARIMA Opt") in order to see the influence of different initialisation techniques on forecasting accuracy.

In addition we used `auto.ssarima()` with backcasting and the switched off mechanism of skipping orders (controlled by `workFast=FALSE` parameter), discussed in the last paragraph of Section 3 in order to see if it improves the performance of the model or not (this is denoted as "SSARIMA NSO").

Furthermore, we have used the extensive search for state space ARIMA ("SSARIMA Ext"), applying models with all the possible orders and selecting the one with the lowest AIC. This took the most computational time, but allowed us to evaluate the proposed algorithm of order selection.

Finally we have also used a benchmark in the experiment – conventional ARIMA implemented in `auto.arima()` function from `forecast` v8.4 package for R (Hyndman and Khandakar, 2008), denoted as "ARIMA". This implementation allows selecting between seasonal and non-seasonal models using information criteria, but the model itself is formulated in the conventional way.

In order to assess the accuracy of the proposed model, we use the data of an American retail company. This is typical supply chain data, containing 4267 series with 36 monthly observations each. All the time series in the dataset can be categorised as shown in Table 1. The classification was done ex post, by applying the `auto.arima()` function to each of the time series and the whole 36 observations. We used the rule, according to which the time series is considered as seasonal, if seasonal AR, I or MA has non-zero order. If the resulting ARIMA model contained either a non-zero order of non-seasonal difference or a drift component, then the series was flagged as non-stationary. The categorisation in Table 1 is provided for information,

	Non-seasonal	Seasonal
Stationary	25.4%	23.5%
Non-stationary	17.4%	33.7%

Table 1: Categories of time series in the supply chain dataset.

showing the variety of different processes in the dataset, and it was not used for order selection or parameter evaluation.

In order to assess the accuracy of models, we withheld the last 9 observations, which leaves 27 observations in the training set. This is a small sample from a conventional ARIMA perspective, but typical for supply chains and sufficient for simpler forecasting models. We do fixed origin evaluation, producing one to nine steps ahead forecasts, and then calculating the following error measures:

1. MPE – Mean Percentage Error, which assesses bias of forecasts:

$$\mathrm{MPE} = \frac{1}{h}\sum_{j=1}^{h} \frac{e_{t+j}}{y_{t+j}},$$

   where e_{t+j} is the j-steps ahead forecast error, and $h = 9$ is the forecasting horizon, for this and all the other error measures.

2. MAPE – Mean Absolute Percentage Error, which assesses accuracy of forecasts and is commonly used in practice:

$$\text{MAPE} = \frac{1}{h} \sum_{j=1}^{h} \frac{|e_{t+j}|}{y_{t+j}}.$$

   This is considered by many forecasters as a biased error measure as it encourages under-forecasting (Makridakis, 1993).

3. MASE – Mean Absolute Scaled Error, measure proposed by Hyndman and Koehler (2006), which is less biased than MAPE:

$$\text{MASE} = \frac{\frac{1}{h} \sum_{j=1}^{h} |e_{t+j}|}{\frac{1}{t-1} \sum_{i=2}^{t} |y_i - y_{i-1}|};$$

4. sMAE – scaled Mean Absolute Error by Petropoulos and Kourentzes (2015), which is similar to MASE, but has easier interpretation, close to the one of MAPE:

$$\text{sMAE} = \frac{\frac{1}{h} \sum_{j=1}^{h} |e_{t+j}|}{\bar{y}},$$

   where $\bar{y} = \frac{1}{t} \sum_{i=1}^{t} y_i$.

5. ARMAE – Average Relative Mean Absolute Error from Davydenko and Fildes (2013) which was shown to be the least biased error measure among (2) – (5):

$$\text{ARMAE} = \frac{\frac{1}{h} \sum_{j=1}^{h} |e_{1,t+j}|}{\frac{1}{h} \sum_{j=1}^{h} |e_{2,t+j}|},$$

   where $e_{1,t+j}$ is the j-steps ahead forecast error of the model under consideration and $e_{2,t+j}$ is the j-steps ahead forecast error of the Naïve method. Note that when ARMAE is aggregated over all the series, the geometric mean is used instead of arithmetic.

The error measures have been calculated for each separate time series. After that the mean and the median values of each error measure across all the series have been calculated. The results are presented in Tables 2 and 3. The best values in the tables are shown in boldface; the second best values are in italic.

Model	MPE	MAPE	MASE	sMAE	ARMAE
ARIMA	-18.2	49.4	*119.6*	41.5	91.0
SSARIMA	-15.4	*48.4*	**119.2**	*41.3*	90.0
SSARIMA NSO	*-14.8*	**48.1**	**119.2**	*41.3*	*89.9*
SSARIMA Opt	-15.1	48.8	120.2	**41.2**	**89.9**
SSARIMA Ext	**-11.4**	50.9	126.5	44.0	95.2

Table 2: Mean error measures (percentages).

As can be seen from Table 2, SSARIMA with backcasting and the proposed order selection method performs better or at least not worse than more complicated SSARIMA models, including the one with the extensive search: the differences in performance of SSARIMA with the other versions of the model are very small.

Similar conclusions can be drawn from the analysis of Table 3, where SSARIMA performed slightly better than the other models in terms of MASE and ARMAE. Note that SSARIMA performed better than the conventional ARIMA across all measures. This can be explained by the ability of the former to better identify seasonality on small samples.

Model	MPE	MAPE	MASE	sMAE	ARMAE
ARIMA	-4.1	33.5	103.9	34.8	97.4
SSARIMA	*-1.9*	*32.5*	**100.0**	*34.4*	**92.9**
SSARIMA NSO	-2.1	32.7	*100.6*	34.4	*93.2*
SSARIMA Opt	-1.0	**32.4**	102.8	**34.0**	96.7
SSARIMA Ext	**-0.5**	35.0	105.0	35.8	95.8

Table 3: Median error measures (percentages).

It is worth noting that the extensive search of the optimal order does not improve upon the accuracy of SSARIMA model – although the order selected by SSARIMA is not optimal in the sense of AIC, it performs better in terms of forecasting accuracy. Furthermore, the optimisation procedure does not bring significant improvements and the SSARIMA NSO performs slightly worse than SSARIMA in many cases. In addition, the SSARIMA with the optimised initials outperforms SSARIMA with backcasting in some cases, but it does not demonstrate substantial improvement.

In order to further investigate the performance of SSARIMA versus ARIMA, we summarise the ARMAE error measures for the four categories from Table 1. The results for the other error measures look similar, so we have decided to focus on ARMAE, as it is the least biased error measures of the ones in our pool (Davydenko and Fildes, 2013). These values are presented in Table 4.

Series type	ARIMA	SSARIMA
Non Seasonal, Stationary	**0.798**	0.824
Non Seasonal, Non Stationary	**1.050**	1.155
Seasonal, Stationary	0.872	**0.817**
Seasonal, Non Stationary	0.961	**0.906**

Table 4: ARMAE of ARIMA and SSARIMA for different categories of the data.

It can be noted from Table 4, that while ARIMA performs better than SSARIMA on non-seasonal time series, SSARIMA is much better on the seasonal data, thus showing the improvement in the overall forecasting accuracy. In fact, it seems that SSARIMA overfits the non-seasonal data, selecting wrongly the seasonal orders, while ARIMA underfits the seasonal data, not selecting the necessary orders. Given the prevalence of seasonal data in the dataset (57.2% according to Table 1), the summary value of the ARMAE for SSARIMA is lower than that of ARIMA.

Finally, Table 5 summarises the computational time for each of the models for the whole dataset (calculated in serial on Intel Core i7 of 5th generation).

Model	Time in minutes
ARIMA	39.82
SSARIMA	47.86
SSARIMA NSO	309.67
SSARIMA Opt	289.68
SSARIMA Ext	17,989.72

Table 5: Time of computation for each model in minutes for all 4267 series.

Although SSARIMA could not outperform ARIMA in terms of time, the difference in their performance is not large. At the same time SSARIMA with the proposed algorithm produced forecasts faster than any other implementation, and in a practical time. As expected, SSARIMA with backcasting and the new order selection performed much faster than other SSARIMA algorithms. Note that the extensive search took almost 18 thousand minutes of computational time, which is equivalent to 300 hours or 12.5 days. Nevertheless, it performed worse than the faster algorithms. So, although SSARIMA does not necessarily beat other models in forecasting accuracy, it is much more efficient and faster than its competitors. Taking into account the accuracy of the state space ARIMA and its speed of calculation, it can be concluded that the model in the proposed form can efficiently be used in a supply chain context, especially for seasonal data.

5. Conclusions

ARIMA is seldom used in a supply chain context because of the limitations of the data and general complexity of the model. We have discussed the state space form of ARIMA with a single source of error and showed that it overcomes some of the limitations of the conventional ARIMA. We have shown that the state space ARIMA simplifies some of the steps in forecasting and can be used even on data with a short history.

All of the above allows using seasonal ARIMA on small samples, containing at least 2 seasonal cycles, something that ARIMA in the conventional form cannot do. In addition the state space form permits comparing different models directly using information criteria, because they can be initialised in the zero period, making sample sizes for models with different orders of differences the same.

We have also proposed an algorithm of order selection for state space ARIMA, which substantially decreases the pool of models under consideration. This algorithm does not employ hypothesis testing, an important feature in cases of small samples, which are very common in a supply chain context. We tested the state space ARIMA with the proposed order selection algorithm on supply chain data and showed that it outperforms the implementation of the conventional ARIMA from `forecast` package for R in terms of accuracy and that it works fast. It seems to perform especially well on seasonal data. Furthre research is ongoing to find improvements in the algorithm for non-seasonal data.

Furthermore, the practicality of our proposed approach is evidenced by its introduction into commercial software by the *Demand Works* company. The new software module, called ARIMA, is based on the SSARIMA model discussed in this paper. However, it has been subject to several modifications and adjustments which cannot be disclosed because of confidentiality reasons. Nevertheless, we can report that the implemented SSARIMA module demonstrates further improvements in the accuracy and significant reduction in computational times in comparison with the implementations discussed in this paper. Furthermore, *Demand Works* software is used by over 400 corporations, demonstrating that the approach discussed in this paper has reasonable commercial applicability. In summary, we can conclude that ARIMA in state space form is a practical and efficient option for supply-chain forecasting and, indeed, for any context, where the historical time series is limited with few complete seasonal cycles.

The focus of this paper was on state-space ARIMA applied to supply chain data. However, this is not the only possible area of application, and we think that developing and exploring the efficient algorithms for ARIMA application in other business contexts is an interesting direction for future research. This means that as a future work, the state-space ARIMA should be tested on other datasets and compared with other popular forecasting methods. Finally, another interesting direction for future work would be to compare the performance of the proposed approach with the other approaches in terms of inventory measures, such as service level and costs of stocking, similar to the analysis by Syntetos and Boylan (2006).

Acknowledgement

We would like to thank *Demand Works* and especially Bill Tonetti and Eric Townson for their support and cooperation in testing of the models and their help in implementation of the ARIMA module in Smoothie.

Appendix A. Derivation of initial values of state vector

We assume that $\epsilon_j = 0$ for each $j = 1, \dots, K$, which gives us K estimates of the first component of the vector based on the measurement equation in (6):

$$\begin{aligned} v_{1,0} &= y_1 \\ v_{1,1} &= y_2 \\ &\vdots \\ v_{1,K-1} &= y_K \end{aligned} . \tag{A.1}$$

This means that the state space model (6) simplifies to:

$$\begin{array}{ll} y_t = v_{1,t-1} & \\ v_{1,t} = \varphi_1 v_{1,t-1} + v_{2,t-1} + v_{K+1,t-1}, & \text{for } j = 1 \\ v_{j,t} = \varphi_j v_{1,t-1} + v_{j+1,t-1}, & \text{for } 1 < j \leq K \\ v_{K+1,t} = v_{K+1,t-1} & \end{array} \tag{A.2}$$

Every $j+1$ component for $1 < j \leq K$ in (A.2) can be expressed the following way:

$$v_{j+1,t-1} = v_{j,t} - \varphi_j v_{1,t-1}, \tag{A.3}$$

meaning that it can be expressed using the values of the previous component and the very first one. The second component is expressed as:

$$v_{2,t-1} = v_{1,t} - \varphi_1 v_{1,t-1} - v_{K+1,t-1}. \tag{A.4}$$

Substituting values from (A.1) into (A.3) and (A.4) leads to the following system:

$$\begin{array}{ll} v_{1,t-1} = y_t, & \text{for } t = \{1, \dots, K\} \\ v_{2,t-1} = v_{1,t} - \varphi_1 y_t - v_{K+1,t-1}, & \text{for } t = \{1, \dots, K-1\} \\ v_{j,t-1} = v_{j-1,t} - \varphi_{j-1} y_t, & \text{for } 2 < j \leq K \text{ and } t = \{1, \dots, K-j+1\} \end{array} . \tag{A.5}$$

So the procedure of the initialisation of the state vector of state space ARIMA is iterative, the components are defined one after another, starting from the first and finishing with the K-th. The value of $K+1$ component in this case is defined by the optimiser.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Athanasopoulos, G., Hyndman, R. J., Song, H., Wu, D. C., 2011. The tourism forecasting competition. International Journal of Forecasting 27 (3), 822–844.

Box, G., Jenkins, G., 1976. Time series analysis: forecasting and control. Holden-day, Oakland, California.

Davydenko, A., Fildes, R., 2013. Measuring Forecasting Accuracy: The Case Of Judgmental Adjustments To SKU-Level Demand Forecasts. International Journal of Forecasting 29 (3), 510–522.

Dellino, G., Laudadio, T., Mari, R., Mastronardi, N., Meloni, C., 2018. A reliable decision support system for fresh food supply chain management. International Journal of Production Research 56 (4), 1458–1485.

Disney, S., Farasyn, I., Lambrecht, M., Towill, D., de Velde, W. V., 2006. Taming the bullwhip effect whilst watching customer service in a single supply chain echelon. European Journal of Operational Research 173 (1), 151 – 172.

Doganis, P., Aggelogiannaki, E., Sarimveis, H., 2008. A combined model predictive control and time series forecasting framework for production-inventory systems. International Journal of Production Research 46 (24), 6841–6853.

Harvey, A. C., Phillips, G. D., 1979. Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances. Biometrika 66 (1), 49–58.

Hosoda, T., Naim, M. M., Disney, S. M., Potter, A., 2008. Is there a benefit to sharing market sales information? linking theory and practice. Computers & Industrial Engineering 54 (2), 315 – 326.

Hyndman, R. J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 27 (3), 1–22.

Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22 (4), 679–688.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with Exponential Smoothing. Springer Berlin Heidelberg.

Kim, J. S., Shin, K. Y., Ahn, S. E., 2003. A multiple replenishment contract with ARIMA demand processes. Journal of the Operational Research Society 54 (11), 1189–1197.

Makridakis, S., 1993. Accuracy concerns measures : theoretical and practical concerns. International Journal of Forecasting 9, 527–529.

Makridakis, S., Andersen, A. P., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R. L., 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. Journal of Forecasting 1 (2), 111–153.

Makridakis, S., Hibon, M., 1997. ARMA models and the Box-Jenkins methodology. Journal of Forecasting 16, 147–163.

Makridakis, S., Hibon, M., 2000. The M3-Competition: results, conclusions and implications. International Journal of Forecasting 16, 451–476.

Pearlman, J. G., 1980. An algorithm for the exact likelihood of a mixed autoregressive moving average process. Biometrika 67 (1), 232–233.

Petropoulos, F., Kourentzes, N., 2015. Forecast combinations for intermittent demand. Journal of the Operational Research Society 66 (6), 914–924.

Snyder, R. D., 1985. Recursive Estimation of Dynamic Linear Models. Journal of the Royal Statistical Society, Series B (Methodological) 47 (2), 272–276.

Snyder, R. D., Ord, J. K., Koehler, A. B., 2001. Prediction Intervals for ARIMA Models. Journal of Business & Economic Statistics 19 (2), 217–225.

Svetunkov, I., 2017. smooth: Forecasting Using Smoothing Functions. R package version 2.1.1.
URL `https://github.com/config-i1/smooth`

Svetunkov, I., Petropoulos, F., 2018. Old dog, new tricks: a modelling view of simple moving averages. International Journal of Production Research 56 (18), 6034–6047.

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., Nikolopoulos, K., 2016. Supply Chain Forecasting: Theory, Practice, their Gap and the Future. European Journal of Operational Research 252, 1–26.

Syntetos, A. A., Boylan, J. E., 2006. On the stock control performance of intermittent demand estimators. International Journal of Production Economics 103 (1), 36–47.

van Gils, T., Ramaekers, K., Caris, A., Cools, M., 2017. The use of time series forecasting in zone order picking systems to predict order pickers' workload. International Journal of Production Research 55 (21), 6380–6393.

Wang, S. J., Huang, C. T., Wang, W. L., Chen, Y. H., 2010. Incorporating ARIMA forecasting and service-level based replenishment in RFID-enabled supply chain. International Journal of Production Research 48 (9), 2655–2677.

Weller, M., Crone, S., 2012. Supply Chain Forecasting: Best Practices and Benchmarking Study. Tech. rep., Lancaster Centre for Forecasting.

Winklhofer, H., Diamantopoulos, A., Witt, S. F., 1996. Forecasting practice: A review of the empirical literature and an agenda for future research. International Journal of Forecasting 12 (2), 193–221.