Holiday Effects in Indian Manufacturing Series

Anirban Sanyal ¹
Pratik Mitra ²
Tucker S. McElroy
Anindya Roy

¹,² Reserve Bank of India

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: August 10, 2017

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.
Holiday Effects in Indian Manufacturing Series

Anirban Sanyal¹ Pratik Mitra² Tucker S. McElroy³ and Anindya Roy⁴

Abstract

Moving holidays are known to deeply impact economic data, and can often be modeled through fixed regressors computed from a window of time pertaining to the festival in question. However, determining the width of this time window – corresponding to heightened economic activity in time periods antecedent or subsequent to the festival dates – can be difficult in cases where the holiday’s impact is new. We study the impact of Navaratri and Diwali on Indian manufacturing data, and determine the most appropriate window sizes by systematically comparing p-values of various holiday Wald statistics. This method is demonstrated on one of the manufacturing series with a large holiday effect, where the Navaratri and Diwali windows determined empirically to begin 12 and 10 days respectively before the festival dates, yielding an improved model of the data.

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

1 Introduction

Moving holidays can have a substantial impact on the seasonal patterns of economic time series, particularly in the retail sector (Findley, Wills, and Monsell (2005)) as well as for certain macroeconomic variables. The program X-13ARIMA-SEATS (U.S. Census Bureau, 2015), used around the world to seasonally adjust time series data, has been effectively used to model country-specific holiday effects in Europe, Asia, Latin America, and other regions of the globe (see Soukup and Findley (2000), Monsell (2007), Roberts, Holan, and Monsell

¹Research Officer, Department of Statistics and Information management, Reserve Bank of India, Mumbai, asanyal@rbi.org.in
²Director, Monetary policy Department, Reserve Bank of India, Mumbai, pmitra@rbi.org.in
³Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov
⁴Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, anindya.roy@census.gov
This paper is a study of the effects of the holidays Navaratri and Diwali on Indian economic indicators data. The overall impact is quite narrow, but for two of the series there is some benefit to using these regressors.

Background on moving holiday effects can be found in Bell and Hillmer (1983). A moving holiday is one in which its date of occurrence varies from year to year, so that it is not aliased with the regular seasonal effect. Nevertheless, the failure to account for salient holiday effects in the modeling and adjustment of series can result in deficient seasonal adjustments – see Findley, Wills, and Monsell (2005). However, moving holiday effects are typically only pertinent for certain types of variables; nondurable goods – such as candy retail sales – can exhibit the impact of a holiday, whereas durable goods – such as steel production – would not be expected to be influenced greatly. In Indian context also, consumer durable sales are found to go up during Diwali and Navaratri due to festive impulse. The unique cultural facets of each country or province will dictate whether or not holiday effects are relevant.

A given festival impacts a particular month in proportion to the number of festival days – including anticipatory activity and the aftermath – included in that month. Therefore the length of the holiday activity (called the holiday window size) is a key determinant of the holiday’s impact on the monthly series. For instance, if the celebration date of a given festival happens to occur on the first week of a month, anticipatory festival activity may impact the previous month. Some previous studies (e.g., Monsell (2007)) examine the impact of window size on the modeling and seasonal adjustment of time series, although no systematic treatment of window size estimation has been developed or studied. Current practice is to fit several window specifications (these being chosen by the analyst according to cultural knowledge) and choose between them via some statistical criterion. This paper provides a new approach to window size selection, which is tested upon Indian manufacturing data.

India is a highly populated country with diverse religions and culture, each with their own unique holiday structures. This makes for an interesting study, since the heterogeneous cultural structure is in marked contrast to fairly homogeneous nation-states, such as the USA and the countries of Europe. People of a particular cultural identity tend to be geographically concentrated within specific regions of India, so that regional data could be expected to bear out holiday patterns more clearly; nation-wide aggregates will tend to average the contributions of diverse regions, and to a degree smoothen the holiday impacts. This paper focuses on two important Hindu festivals: Navaratri and Diwali.

Navaratri is a Hindu festival consisting of nine nights and ten days of celebration. The celebration begins eight days before the festival date, on Ashvin Shukla Prathama, and ends the day after on Ashvin Shukla Navami (Wikipedia, 2015). Diwali is another major festival, with activity commencing on Dhanteras two days before the Diwali date, and ending on Bhai Dooj, two days after the Diwali date; see Wikipedia (2015). The holidays are based on the lunar calendar and can vary substantially from year to year. As with many other aspects of Indian society, the holidays exhibit tremendous diversity in terms of how (and when) they are observed and celebrated regionally, as well as the effect they have on regional economies. The national effect is an average of the regional effects, and may be less pronounced than the holiday effects in individual regions.

Complicating this picture is the fact that the date of celebration of Navaratri varies across India; it is primarily celebrated in the northern and western portions of India, and this holiday roughly corresponds to a celebration of regional major festivals (similar to Navaratri) taking place in the southern and eastern parts of India. However, the exact date and duration of festivals in various portions of the country may not exactly match with Navaratri festival. For this reason, the optimal Navaratri window size is not immediately clear, and must be empirically determined. Similar phenomenon can be observed in the case of Diwali, as the exact date of celebration varies across India.
India celebrates different festivals across different regions, cultures and religions over a year. The Hindu festivities surrounding Navaratri and Diwali often overlap with major Muslim (who, after Hindus, constitute the second largest religious group in India’s population) festivals, such as Eid-ul-Fitar and Id-ul-Zuha. Moreover, the Navaratri festival is preceded by Ganesh Chaturthi, another Hindu festival which may have significant regional holiday effects – particularly in large western and southern states such as Maharashtra, Karnataka and Telengana. Thus, one could argue that non-conventional moving holiday regressors are needed to fully capture the effect of this extended celebratory period, where potentially the effect ramps up and peaks immediately prior to the Navaratri days and falls off after the Diwali.

Standard techniques and software (X-13ARIMA-SEATS) exist to model customized cultural activities. However, these methods presume that the window size has been identified. In this paper, we compare p-values of Wald statistics for various holiday window size specifications, and make a selection on the basis of diverging significance. In Section 2 we discuss the thirteen Indian manufacturing series of the Reserve Bank of India (RBI) used in our analysis, and in Section 3 the methodology is presented. Results are summarized in Section 4, and Section 5 concludes.

2 Data

In India, different official agencies publish monthly data on economic indicators. These indicators are frequently used for assessing the macroeconomic condition of the country. Analysis of seasonal behavior helps in differentiating between the seasonal changes and long-run changes of economic time series, which is useful for understanding the underlying economic phenomenon. From a policy maker’s perspective, month on month (m-o-m) or quarter on quarter (q-o-q) movement of an economic series are often perceived to be more informative than year on year (y-o-y) movement as m-o-m or q-o-q growth rates provides movements in these series over last month or quarter rather than a year back.

If seasonality exists in the series, the seasonally adjusted annualised rate estimated from the m-o-m or q-o-q growth rates provides important information on the momentum of the series in recent times. Hence, analysis of seasonality in these series is required for better policy decisions, and in particular the removal of seasonality facilitates clearer decision-making. However, most of the data required for these indicators are collected at a regional level, and hence the regional impact affects the data generating process. In view of the above, thirteen important Indian economic series have been considered for analysing the holiday impact on seasonal patterns. These indicators broadly cover price data, production data and other economic indicators. The monthly data of these indicators cover the period from April 1994 to September 2014.

The wholesale price index (WPI) and consumer price index (CPI)\(^5\) are derived using Laspeyres’ formula\(^6\) in terms of relative price. On the other hand, the production data, i.e., the index of industrial production (IIP) for consumer durable and non-durable commodities, are basically quantity indices derived using the Laspeyres Index\(^7\) with base time 2004-2005. The other production data, i.e., Cement Production and Steel Production, are collected on

---

\(^5\)CPI Industrial Worker
\(^6\)WPI Compilation Manual - Office of Economic Adviser & Methodology for compilation of Consumer Price Index Numbers for Industrial Workers (BASE : 2001=100 )
a monthly basis by the Department of Industrial Policy and Promotion, using a fixed sample frame. Vehicle Production and sales data are official figures released by Society of Indian Automobile Manufacturers. The other economic indicators – namely Passenger kilometre flown and Freight tonne kilometres flown – are published by the Directorate General of Civil Aviation (Government of India), after compiling the figures from different carriers. Item level description of each of these indicators are provided in Table 1.

<table>
<thead>
<tr>
<th>Label</th>
<th>Series Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dur</td>
<td>IIP Consumer Durable</td>
<td>Combined indices of 43 durable item groups’ production volume index. The items are classified as per National Industrial Classification (2004) at 5 digit level.</td>
</tr>
<tr>
<td>Nondur</td>
<td>IIP Consumer Non-durable</td>
<td>89 commodity baskets based on NIC 2004 classification are covered under consumer non-durable.</td>
</tr>
<tr>
<td>Pass</td>
<td>Passenger km flown in domestic scheduled operations</td>
<td>This element represents total kilometers of domestic flight travel undertaken by passengers. This indicator is used as a services sector indicator for assessing transport activity.</td>
</tr>
<tr>
<td>Freight</td>
<td>Freight tonne km flown in domestic scheduled operations</td>
<td>This indicator represents the freight carried by flights during domestic travel. Momentum of this indicator is used as a services indicator for assessing trade and transport growth.</td>
</tr>
<tr>
<td>Veh.sales</td>
<td>Commercial Motor Vehicle Sales</td>
<td>Society of Indian Automobile Manufacturers (SIAM) collect the data of commercial motor vehicle sales from different companies. Momentum of the indicator is used as proxy of consumer demand and transport indicator.</td>
</tr>
<tr>
<td>Veh.prod</td>
<td>Commercial Motor Vehicle Production</td>
<td>This indicator is used to assess the transport sector growth. SIAM is the custodian of the data.</td>
</tr>
<tr>
<td>Cem.prod</td>
<td>Cement Production</td>
<td>Cement production data is collected by Department of Industrial Policy and Promotion (DIPP) at monthly frequency. It comprises of 2.4% of overall industrial production. Cement production has been used as benchmark indicator for estimating gross value added in construction activity by Central Statistical Office (CSO).</td>
</tr>
<tr>
<td>Steel.prod</td>
<td>Steel Production</td>
<td>Steel production data is collected by Department of Industrial Policy and Promotion (DIPP) at monthly frequency. It comprises of 6.7% of overall industrial production. Steel and Cement production have been used as benchmark indicator for estimating gross value added in construction activity by Central Statistical Office (CSO).</td>
</tr>
<tr>
<td>wpi</td>
<td>WPI for all commodities</td>
<td>This index represents the wholesale price indicator of commodities. The overall index constitutes both primary article, Fuel &amp; Power and manufactured items. The inflation figure derived from WPI all commodities provides a notion about the overall rate of price increase. This index closely resembles the producers’ price index in an Indian context.</td>
</tr>
<tr>
<td>wpi.prime</td>
<td>WPI Primary Article</td>
<td>WPI Primary Article mainly includes items which are available readily. It is comprised of Food articles, non-food articles and minerals. The manufactured items and Fuel &amp; Power are covered separately within WPI.</td>
</tr>
<tr>
<td>wpi.prod</td>
<td>WPI Manufactured Products</td>
<td>WPI manufactured products includes food items, beverages &amp; tobacco products, textile, wood products, paper products, chemical products, industrial machinery etc. The contribution of manufactured products in overall WPI is around 65%</td>
</tr>
<tr>
<td>wpi.food</td>
<td>WPI Food Items</td>
<td>This series falls under WPI Primary Articles and it carried weight of 14.3% in overall WPI. Food items are highly volatile in nature and are subjected to high seasonal variation.</td>
</tr>
<tr>
<td>cpi</td>
<td>CPI Industrial Worker</td>
<td>CPI Industrial workers represent the general consumer price level of commodities consumed by industrial workers. This price index is compiled by Labour Bureau at a monthly frequency.</td>
</tr>
</tbody>
</table>

Table 1: Time Series Data used in the article.
3 Methodology

The techniques of analysis for custom holidays are fairly standard (see Roberts et al. (2010)). One can employ univariate RegARIMA models with regressors for outliers, trading day, and moving holiday effects, with the stochastic portions of the data given by SARIMA models. X-13ARIMA-SEATS is used to fit each of the series, allowing for Additive Outliers (AO) and Level Shifts (LS), after modifying the data by a log transformation, if warranted. Typical (flow) trading day regressors can be considered, and may be important for some series. GenHol, a program of the U.S. Census Bureau, can be used to generate the holiday regressors for Navaratri and Diwali; the dates for the holidays were computed, covering the period from 1600 AD through 2400 AD. Based on cultural information regarding these holidays, initial windows were identified for the two holidays. For Navaratri, the festival’s activity commences eight days before the festival date, and ends the day after; for Diwali, activity begins two days before the festival date, and ends two days after.

There is some degree of subjectivity in the selection of these windows, and a degree of mis-specification can be expected to detect effects that are truly present (Findley, Wills, and Monsell (2005)). The Genhol regressors $x_t$ are defined as follows. Let a window size be given, consisting of a number $w$ of consecutive days. For any month of index $t$, we determine the number $n_t$ of those festival days contained in the given month, and

$$x_t = \frac{n_t}{w}.$$  

So if we fix the festival end date and calculate the start date accordingly, the value of $x_t$ for month $t$ including the end date either stays constant at value one (if the start date is also falls in the same month $t$) or decreases with increasing $w$ (if the start date pertains to month $t-1$). A one-day window ($w=1$) corresponds to having a simple dummy for the holiday effect – namely, a value of one if the holiday takes place in a given month, and zero otherwise. The longer window effects allow the modeler to explore the impact of a festival on more than one month. For instance, when Navaratri occurs close to the beginning of October, some economic impact can be expected in both the September and October values of the series.

It is also possible to create customized holiday regressors, using the GenHol program of the U.S. Census Bureau, and in this way explore an economic effect that ramps up gradually over several days/months preceding the holidays. However, we stress that the current methodology (Monsell, 2007) does not directly estimate the window size – or the time location of the ramp effect – from the data; rather, the analyst must posit different suppositions about the window length, and decide between contenders via some criterion, such as the Akaike Information Criterion (AIC). What we propose here is to compare window specifications via examining the p-values of those regressors’ Wald statistics.

With a fixed end date, as we move the start date farther into the past, we expect that our initial specification (with $w=1$) is mis-specified, but as $w$ increases we approach more closely the correct specification, which we denote by $w^\star$. Then for $w > w^\star$, we are again mis-specified, and more severely so as $w$ increases. Because mis-specifications are flagged as low p-values in the Wald statistic, we can expect the p-values to increase with $w$ for $w \leq w^\star$, to peak around $w^\star$, and then to decrease for $w > w^\star$. Heuristically, if we were to plot the p-values as a function of $w$, we expect that a local maximum corresponds to $w^\star$.

Such a procedure could also be used to determine the optimal end date. First, we might fix the end date to be the day after the festival, and determine the optimal start date in the manner described. Once this is done, we might begin to increase the end date as well,
increasing \( w \) into the future, and look for a local maximum in the p-values. This technique could be done for both Navaratri and Diwali, using some benchmark specification for the other holiday. Comparing p-values from fitted non-nested models is not ideal, but seems more attractive than the alternative, which is to do a likelihood ratio test. Because the models are non-nested, the distribution will not be \( \chi^2 \), and the theory is unclear; AIC yields an approximate log likelihood ratio test, but without any asymptotic theory.

4 Results

4.1 Models for Manufacturing Series

For the selected economic indicators, we utilized the initial window sizes based on the known festival dates of Navaratri and Diwali. In each case, we utilized the automdl feature of X-13ARIMA-SEATS to identify initial models, but found that some of the series were difficult to model with this method, indicated by poor Ljung-Box statistics. We then obtained more nuanced ARIMA specifications through extensive analysis, presented as our final results; the final diagnostics were adequate in each case. It may also be noted that residual autocorrelation can cause the standard errors of regression coefficients to be inflated (or deflated), leading to spurious conclusions. Even more important is the impact of AO and LS on the standard errors, as we discovered case by case.

For each series, we included the holiday regressors, and attempted to obtain a decent model for the data – in some cases omitting trading day if it was not warranted. Then we could safely examine the t-statistics and Wald statistics for the holiday effects. Secondly, we reran the models with the holiday effects omitted, if they were deemed irrelevant, to see if there was any preference in terms of AIC or model diagnostics. For this second step, we kept the ARIMA and trading day specification the same, but allowed the outliers to change in order to obtain the best possible fit.

We summarize the results of our analysis in Table 2. A variety of SARIMA models were identified, some of which involved no seasonal differencing – this was true of Dur, wpi, wpi.prime, wpi.prod, and wpi.food. Of these, excepting wpi.prime, all had a significant holiday effect (though the impact on wpi.food was marginal). None of the series requiring seasonal differencing exhibited a holiday effect. For the most part, both Navaratri and Diwali tended to be significant or insignificant in tandem, with the exception of wpi.food and Freight. We also refitted the same models without the holiday regressor, looking for deficiencies in the model diagnostics for those series where the holidays were warranted. We discuss the results for each series in some detail below.

4.2 Discussion of Specific Series

For Dur the seasonal AR parameter was .91 in the initial model, which is strong enough to generate seasonal dynamics; nevertheless, according to the Visual Significance criterion of X-13ARIMA-SEATS, this series was not a candidate for seasonal adjustment. The sample-size corrected AIC (or AICC) was 1066.4967. Leaving the holiday effects out (and allowing the outliers to change), the AICC dropped to 1056.9074 because the likelihood actually increased about three points. This seems to be explained by the identification of a new AO in the non-holiday model; the t-statistic for the Jan 2005 AO was \(-3.77\) in the holiday model, but
Table 2: Modeling results for thirteen manufacturing series of RBI. T-statistics are given for Navaratri and Diwali, and the p-value of their Wald statistic is in the final column.

-3.90 in the non-holiday model, with a cutoff of -3.84. The SARIMA model parameters changed little.

Nodur, Freight, Veh.sales, Veh.prod, Cem.prod, Steel.prod, and cpi share a similar story. The holiday effect is not significant, and its removal gives a slight improvement in terms of AICC. These series require seasonal adjustment, and the X-11 seasonal adjustments are adequate either with or without the holiday regressors. For wpi.prime, there was very little seasonality present (the seasonal AR parameter was .31), and omitting the holiday regressors improves AICC slightly, from 967.7236 to 966.8173.

For wpi, the seasonality is somewhat weak – the seasonal AR parameter is .47. Omission of the holiday regressors does not corrupt the model, as the AICC drops from 449.2430 to 439.7505. This is apparently caused by the selection of a LS outlier: for June 2008 the t-statistic is 3.91 in the holiday model (non-significant), but 4.50 in the model without holiday effects, and is flagged as significant. The SARIMA model parameters are similar in both models.

For wpi.prod the seasonality appears to be weak (AR parameter .72), it is deemed sufficiently palpable by the Visual Significance criterion to warrant adjustment. The AICC increases from 247.7615 to 250.7117 in omitting the holiday effects (and the same LS was retained in both models). SARIMA coefficients were altered little. Here, at least, a case could be made for using the holiday regressors.

Finally, for wpi.food there is moderate seasonality, with a seasonal AR parameter of .97 and Visually Significant peaks in the spectrum. This series has trading day, and a single LS; these effects are retained (though TD is weakened to Wald p-value of .10) after deletion of the holiday effect. The AICC moves slightly down from 1000.3483 to 1000.3210, indicating that the models are essentially equivalent. The SARIMA parameters change little.

In summary, wpi.prod and wpi.food are the only two series for which seasonal adjustment is a necessity and a holiday effect (either Navaratri or Diwali) is significant. However, for the latter series the effect is marginal, and both models (with and without the holiday regressors) are acceptable. In the former series, there is diagnostic evidence as to the merits of including holiday regressors. Next, we examine the question of whether this has any impact on seasonal adjustment.
For wpi.prod, the adjustments with and without holiday effects are adequate, according to autocorrelation plots and the Visual Significance criterion. The same is true of wpi.food. Therefore, in these series there is little evidence that Navaratri and Diwali have a measurable impact on the publication of seasonal adjustments. This may seem counter-intuitive, but a simple explanation can be offered. The dominant frequency of the holiday regressors is $\pi/6$, corresponding to phenomenon that repeats once a year in an alternating fashion – the lag twelve autocorrelation of these regressors is negative. (A similar feature is found in Easter regressors.) Omission of such holiday effects could lead to mis-identification of the spectral behavior of the series at frequency $\pi/6$, and subsequently to seasonal adjustment filters that perform inadequately.

The issue comes down to whether the seasonal adjustment filter smooths the data sufficiently. Longer filters correspond to a more stable seasonality, and sometimes fail to adequately seasonally adjust a series that has a highly evolutive seasonal behavior; on the other hand, shorter filters may over-adjust some series, removing not only the seasonality proper but also some other features that pertain to cyclical or irregular dynamics. (See McElroy (2012) for a discussion of these matters.) Ultimately, the final impact of the filter can be assessed by studying the spectral density of the seasonally adjusted component and the irregular. In the case of wpi.prod and wpi.food, these spectral plots exhibited no peaks at $\pi/6$ (or other seasonal frequencies), and on this basis we claim there is no problem of under-adjustment – although, arguably, there is over-adjustment present, this does not detract from the overall criterion of adequacy. Examination of the sample autocorrelation plots for the seasonally adjusted component also confirms this summary.

### 4.3 IIP Consumer Durables

We further examined the Dur series, attempting to find an optimal window size via examination of Wald statistic p-values. The step-wise analysis of the window length indicates that the start date occurs 12 days before Navaratri’s festival celebration date. The end date was fixed to occur one day afterwards, so as not to be confounded with Diwali effects. For Diwali, analysis of the p-values indicates that its impact remains mostly confined from 10 days before until 9 days after the celebration.

From an economic perspective the identified window lengths are intuitive, as the estimated start date for Navaratri corresponds to the cultural behavior noted in the introduction. On the other hand, the start date for Diwali can be interpreted as the mixed effect of regional festivals, celebrated across different regions of India. The end date of Diwali is found to occur 9 days later, which typically corresponds to the period during which the corporate offers on consumer durables are likely to persist.

Using these window sizes, the seasonal adjustment results of the benchmark specification were recomputed. The order of the identified SARIMA model was unchanged, and the residual diagnostics were actually improved over those obtained from automatic model selection. The new seasonal factors differ slightly from those of the benchmark specification.

### 5 Conclusion

This article is an empirical exploration of moving holiday effects in Indian manufacturing series. Our analysis has focused on 13 series published by RBI, and we assessed the impact of Navaratri and Diwali on these series. Although the actual window size could be much
larger, we found no statistical evidence to warrant such a feature, as our specified models
had adequate diagnostics (and were fairly parsimonious, given the sample size) in each
case. Two of the thirteen series had a significant holiday impact, and also required seasonal
adjustment: wpi.prod and wpi.food. Interestingly, the seasonal adjustment for both these
series was adequate even when the holiday effects were intentionally omitted from the model.

We also have proposed a new method for comparing holiday window specifications, via
looking for local maxima of Wald statistic p-values as a function of window size. This
method, applied to one of the series, indicated a somewhat broader window size for both
Navaratri and Diwali, then those of our default specifications. However, the diagnostics and
seasonal adjustment adequacy were not markedly changed by this modification.

Replication of these results, or extensions to other series, is easily accomplished with
X-13ARIMA-SEATS, used in conjunction with the GenHol software. One only needs the
holiday days, and some conjectures about the window of festival activity, in order to produce
custom holiday regressors via GenHol. As a second step, these regressors are inputted into
a standard run of the X-13ARIMA-SEATS software, and model quality is diagnosed using
standard techniques.

Acknowledgements We are grateful to Osbert Pang for comments and construction of
the regressors in GenHol.

References

of the American Statistical Association 78, 526–534.

using RegARIMA models with X-12-ARIMA. Proceedings of the American Statistical
Association.

adjustment. Taiwan Economic Forecast and Policy 43, 33–70.

series. Proceedings of ICES III.

and holiday regressors with country specific regressors. Journal of Official Statistics,


