Teaching Time Series and Regression Analysis in Classes of Business Analytics Using Data from the Ports of Los Angeles and Long Beach

Ardavan Asef-Vaziri

David Nazarian College of Business and Economics
California State University, Northridge. 18111 Nordhoff Street, Northridge, CA 91330-824
Email: aa2035@csun.edu

ABSTRACT

The combined ports of Los Angeles and Long Beach in California are among the world's top ten busiest container ports. The data on the volume of activities in these ports provide an excellent dataset to teach time series and regression analysis. We use 26 years of data on the activities of these ports to teach forecasting models, including moving averages, exponential smoothing, trend-adjusted exponential smoothing, and regression analysis. We also use 312 monthly data for teaching seasonality-enhanced regression, multivariate seasonality regression using dummy variables, and trend and seasonality-adjusted exponential smoothing. Excel functions and formulas are fully embedded in the models we develop. We have learned when theoretical concepts are taught through real-life applications, they positively impact students' mental presence and intellectual engagement inside the classrooms. This manuscript can be used as teaching material or a case study in a business analytics foundation or a supply chain analytics course.

Keywords: freight transportation; ports of Los Angeles and Long Beach, predictive analytics, time series analysis, moving average, trend and seasonality adjusted exponential smoothing, seasonality enhanced regression.

1. INTRODUCTION

Competitive firms need forecasting to develop integrated resources and processes; nourish multi-dimensional and structurally integrated capabilities; understand the revolving business eco-system; create value; and reshape the business organization towards achieving the plans of the enterprises. Marketing, finance, and operations are the three key building blocks of manufacturing, service, and distribution systems. Planning, organizing, budgeting, executing, and controlling are the primary responsibilities of the three key managers. Operations Managers need forecasting for capacity planning, inventory management, and scheduling. Financial Managers need forecasting for investment analysis, revenue and cost analysis, and cash flow planning. Marketing Managers need forecasting for pricing, sales force planning, and promotions. Good forecasting facilitates matching customer value propositions with product attributes, and product attributes with process competencies in the four-dimensional space of cost, quality, time, and variety. While marketing, finance, and operation managers may be interested in forecasting different variables, they have a common interest in the volume of activities and investment plans. They are all interested in long-term and short-term forecasts for strategic, tactical, and operational decisions.

Table 1 shows the world's container port throughput (in twenty-foot equivalent units- or TEUs) in ten countries and ten ports. The combined ports of Los Angeles and Long Beach (LA/LB) are ranked 10th in the world. They comprise the largest port complex in the Western Hemisphere.

Table 1. Container port and country rankings

Container Throughput (Port Ranking)			
(Million TEU)			
Rank	Port	Country	MTEUs
1	Shanghai	China	43.5
2	Singapore	Singapore	36.6
3	Ningbo-Zhoushan	China	28.7
4	Shenzhen	China	26.6
5	Guangzhou Harbor	China	23.2
6	Busan	South Korea	21.6
7	Qingdao	China	22.0
8	Hong Kong, S.A.R	China	18.0
9	Tianjin	China	18.4
10	SPB (LA/LB)	USA	17.3

(a) Top 10 ports: 33%

Container Throughput (Country Ranking)			
(Million TEU)			
Rank	Country	MTEUs	% to World
1	China	245.1	31.2%
2	United States	55.0	7.0%
3	Singapore	36.9	4.7%
4	Korea	28.4	3.6%
5	Malaysia	26.7	3.4%
6	Japan	21.4	2.7%
7	United Arab Emirates	19.3	2.5%
8	Germany	18.0	2.3%
9	Hong Kong SAR, China	18.0	2.3%
10	Spain	17.4	2.2%

(b) Top 10 countries: 62%

Source: American Association of Port Authorities, 2020.

Approximately 1/3 of US seaborne containers move through the LA/LB ports. According to the International Trade Outlook, the value of two-way trade in Southern California customs exceeded 10% of total US international trade in goods. Around 75% of this value passes through to LA/LB ports.

The inbound and outbound volumes of the loaded and empty containers in LA/LB ports provide an attractive data set to teach the basics of time series and analytics. This manuscript is a complete teaching material for time series and regression analysis.

We have learned when theoretical concepts are taught through real-life applications, they positively impact students' mental presence and intellectual engagement inside the classrooms. Our study is especially fit for California's business schools. Only the California State University (CSU) and the University of California (US) systems home close to 800,000 students. The Institute for Advanced Analytics has ranked four California cities in the top 57 cities nationwide, representing the number of data scientists/business analysts. Another objective of this paper is to constitute a bridge between port administrations looking for good quantitative research and econometricians eager to apply their skills to the complex world of modern ports.

Excel functions and formulas are fully embedded in the models we develop. We include additional mathematical manipulations, Excel formulas, and visualization capabilities in four appendices. Our spreadsheet models can serve as templates for other real-life applications.

This manuscript can be used as teaching material or as a case study to enhance teaching materials. We have used it as teaching material in an undergraduate course in business analytics foundations and as a case study in a supply chain analytics graduate course. While we use the total volume of loaded and empty inbound and outbound containers, all data are included for four combinations of inbound, outbound, loaded, and empty volumes. Finally, we hope the work can answer many questions port administrations usually have when trying to understand the research they have commissioned to third parties, often at a very high cost.

We will have a short literature review in Section 2. In Section 3, we estimate yearly port throughput levels using moving averages and exponential smoothing. Measures of forecast accuracy and variability are discussed in Section 4. The level and trend for yearly data are discussed in Section 5 using linear regression and trend-adjusted exponential smoothing. Section 6 estimates monthly data's level, trend, and seasonality using seasonality-enhanced regression analysis, multivariate seasonality regression using seasonal dummy variables, and trend and exponential seasonality smoothing. Conclusions follow in Section 7. In Appendix A, we implement Excel's functional and visualization capabilities by examining a general any-period moving average and its dynamic tables and graphs. In Appendix B, we review the basic mathematics of Exponential Smoothing. Appendix C explains the foundations of the computation of Regression metrics in Excel and provides insight for piecewise regression analysis. All our Excel worksheets are in Appendix D.

2. LITERATURE REVIEW

Time series analysis and regression form 1-2 chapters in almost all operations management books. In this study, we have benefited from Chase, Aquilano, and Jacobs (2000), Stevenson (2014), Cachon and Terwiesch (2020), and especially Chopra (2019) and Iravani (2021).

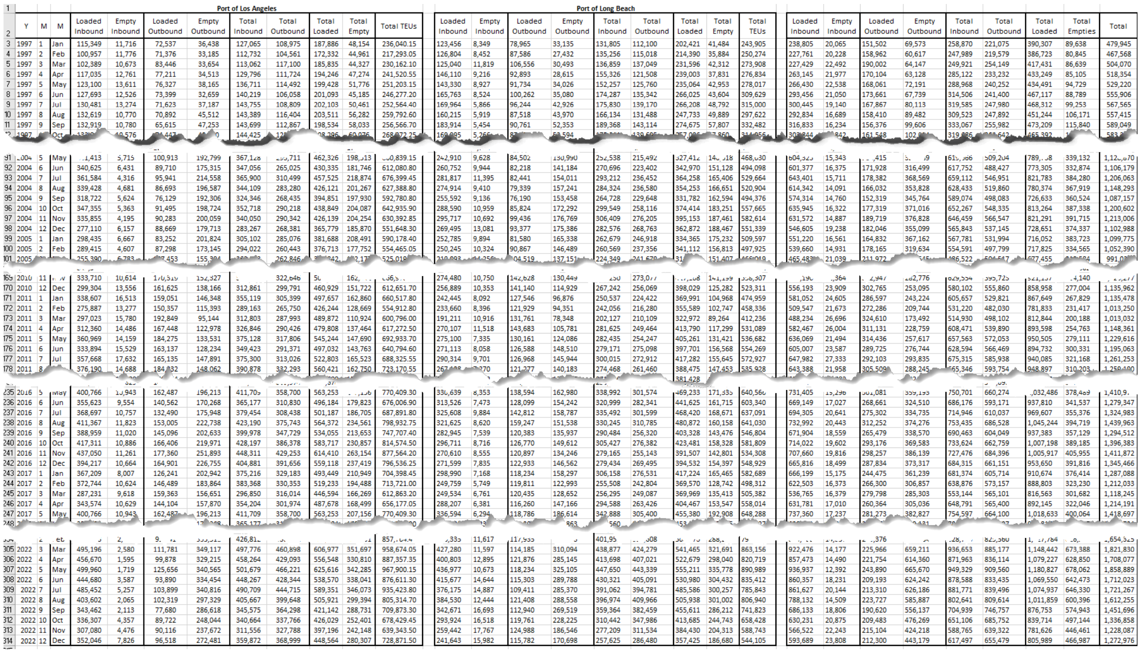
3. HISTORICAL DATA IN LA/LB PORTS AND FORECASTING CHARACTERISTICS

Time series analyzes past data to identify systematic and random components.; to extend systematic components into the future and provide measures of variability. We use 26 years of data on the total inbound and outbound volume of loaded and empty containers in LA/LB ports to experience moving averages, simple exponential smoothing, trend-adjusted exponential smoothing, and regression analysis. We also use 312 monthly data for seasonality-enhanced regression, multivariate seasonality regression using dummy variables, and trend and seasonality-adjusted exponential smoothing. Excel functions and formulas are fully embedded in these computations.

3.1 Historical Data at LA/LB Ports

Table 2 presents parts of 26 years of monthly data for LA and LB, including loaded inbound, loaded outbound, empty inbound, and empty outbound - 312 records with 2496 fields.

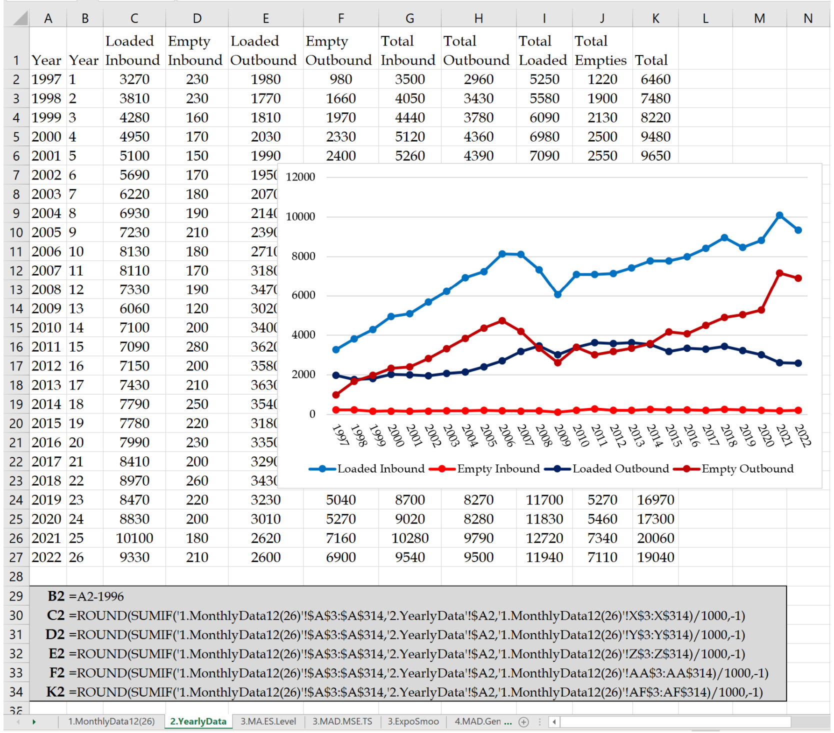
Table 2. 26-Years Monthly TEUs handling in LA/LB Ports



1				Port of Los Angeles									Port of Long Beach																	
2	Y	M	M	Loaded Inbound	Empty Inbound	Loaded Outbound	Empty Outbound	Total Inbound	Total Outbound	Total Loaded	Total Empty	Total TEUs	Loaded Inbound	Empty Inbound	Loaded Outbound	Empty Outbound	Total Inbound	Total Outbound	Total Loaded	Total Empty	Total TEUs	Loaded Inbound	Empty Inbound	Loaded Outbound	Empty Outbound	Total Inbound	Total Outbound	Total Loaded	Total Empties	Total
3	1997	1	Jan	115,349	11,716	72,537	36,438	127,065	108,975	187,886	48,154	236,040.15	123,456	8,349	78,965	33,135	131,805	112,100	202,421	41,484	243,905	238,805	20,065	151,502	69,573	258,870	221,075	390,307	89,638	479,945
4	1997	2	Feb	100,957	11,776	71,376	33,185	112,732	104,561	172,332	44,961	217,293.05	126,804	8,452	87,586	27,432	135,256	115,018	214,390	35,884	250,274	227,761	20,228	158,962	60,617	247,989	219,579	386,723	80,845	467,568
5	1997	3	Mar	102,389	10,673	83,446	33,654	113,062	117,100	185,835	44,327	230,162.10	125,040	11,819	106,556	30,493	136,859	137,049	231,596	42,312	273,908	227,429	22,492	190,002	64,147	249,921	254,149	417,431	86,639	504,070
6	1997	4	Apr	117,035	12,761	77,211	34,513	129,796	111,724	194,246	47,274	241,520.55	146,110	9,216	92,893	28,615	155,326	121,508	239,003	37,831	276,834	263,145	21,977	170,104	63,128	285,122	233,232	433,249	85,105	518,354
7	1997	5	May	123,100	13,611	76,327	38,165	136,711	114,492	199,428	51,776	251,203.15	143,330	8,927	91,734	34,026	152,257	125,760	235,064	42,953	278,017	266,430	22,538	168,061	72,191	288,968	240,252	434,491	94,729	529,220
8	1997	6	Jun	127,693	12,526	73,399	32,659	140,219	106,058	201,093	45,185	246,277.20	165,763	8,524	100,262	35,080	174,287	135,342	266,025	43,604	309,629	293,456	21,050	173,661	67,739	314,506	241,400	467,117	88,789	555,906
9	1997	7	Jul	130,481	13,274	71,623	37,187	143,755	108,809	202,103	50,461	252,564.40	169,964	5,866	96,244	42,926	175,830	139,170	266,208	48,792	315,000	300,445	19,140	167,867	80,113	319,585	247,980	468,312	99,253	567,565
10	1997	8	Aug	132,619	10,770	70,892	45,512	143,389	116,404	203,511	56,282	259,792.60	160,215	5,919	87,518	43,970	166,134	131,488	247,733	49,889	297,622	292,834	16,689	158,410	89,482	309,523	247,892	451,244	106,171	557,415
11	1997	9	Sep	132,919	10,780	65,615	47,253	143,699	112,867	198,534	58,033	256,566.70	183,914	5,454	90,761	52,353	189,368	143,114	274,675	57,807	332,482	316,833	16,234	156,376	99,606	333,067	255,982	473,209	115,840	589,049
91	2004	5	May	[illegible]	5,715	100,913	192,799	367,128	[illegible]	462,326	[illegible]	[illegible]	242,910	9,628	84,502	130,990	252,538	215,492	327,412	[illegible]	468,030	604,323	15,343	[illegible]	[illegible]	[illegible]	509,204	[illegible]	339,132	[illegible]
92	2004	6	Jun	340,625	6,431	89,710	175,315	347,056	265,025	430,335	181,746	612,080.80	260,752	9,944	82,218	141,184	270,696	223,402	342,970	151,128	494,098	601,377	16,375	171,928	316,499	617,752	488,427	773,305	332,874	1,106,179
93	2004	7	Jul	361,584	4,316	95,941	214,558	365,900	310,499	457,525	218,874	676,399.45	281,817	11,395	82,441	154,011	293,212	236,452	364,258	165,406	529,664	643,401	15,711	178,382	368,569	659,112	546,951	821,783	384,280	1,206,063
94	2004	8	Aug	339,428	4,681	86,693	196,587	344,109	283,280	426,121	201,267	627,388.80	274,914	9,410	79,339	157,241	284,324	236,580	354,253	166,651	520,904	614,342	14,091	166,032	353,828	628,433	519,860	780,374	367,919	1,148,293
95	2004	9	Sep	318,722	5,624	76,129	192,306	324,346	268,435	394,851	197,930	592,780.80	255,592	9,136	76,190	153,458	264,728	229,648	331,782	162,594	494,376	574,314	14,760	152,319	345,764	589,074	498,083	726,633	360,524	1,087,157
96	2004	10	Oct	347,355	5,363	91,495	198,724	352,718	290,218	438,849	204,087	642,935.90	288,590	10,959	85,824	172,292	299,549	258,116	374,414	183,251	557,665	635,945	16,322	177,319	371,016	652,267	548,335	813,264	387,338	1,200,602
97	2004	11	Nov	335,855	4,195	90,283	200,059	340,050	290,342	426,139	204,254	630,392.85	295,717	10,692	99,436	176,769	306,409	276,205	395,153	187,461	582,614	631,572	14,887	189,719	376,828	646,459	566,547	821,291	391,715	1,213,006
98	2004	12	Dec	277,110	6,157	88,669	179,713	283,267	268,381	365,779	185,870	551,648.30	269,495	13,081	93,377	175,386	282,576	268,763	362,872	188,467	551,339	546,605	19,238	182,046	355,099	565,843	537,145	728,651	374,337	1,102,988
99	2005	1	Jan	298,435	6,667	83,252	201,824	305,102	285,076	381,688	208,491	590,178.40	252,785	9,894	81,580	165,338	262,679	246,918	334,365	175,232	509,597	551,220	16,561	164,832	367,162	567,781	531,994	716,052	383,723	1,099,775
100	2005	2	Feb	289,415	4,607	87,298	173,145	294,022	260,443	376,713	177,752	554,465.05	250,245	10,324	90,867	146,489	260,569	237,356	341,112	156,813	497,925	539,660	14,931	178,165	319,634	554,591	497,799	717,825	334,565	1,052,390
169	2010	11	[illegible]	333,710	10,614	170,319	152,327	[illegible]	322,646	[illegible]	[illegible]	[illegible]	274,480	10,750	142,628	130,449	[illegible]	273,077	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	282,776	629,554	395,725	[illegible]	[illegible]	[illegible]
170	2010	12	Dec	299,304	13,556	161,625	138,166	312,861	299,791	460,929	151,722	612,651.70	256,889	10,353	141,140	114,929	267,242	256,069	398,029	125,282	523,311	556,193	23,909	302,765	253,095	580,102	555,860	858,958	277,004	1,135,962
171	2011	1	Jan	338,607	16,513	159,051	146,348	355,119	305,399	497,657	162,860	660,517.80	242,445	8,092	127,546	96,876	250,537	224,422	369,991	104,968	474,959	581,052	24,605	286,597	243,224	605,657	529,821	867,649	267,829	1,135,478
172	2011	2	Feb	275,887	13,277	150,357	115,393	289,163	265,750	426,244	128,669	554,912.80	233,660	8,396	121,929	94,351	242,056	216,280	355,589	102,747	458,336	509,547	21,673	272,286	209,744	531,220	482,030	781,833	231,417	1,013,250
173	2011	3	Mar	297,023	15,780	192,849	95,144	312,803	287,993	489,872	110,924	600,796.00	191,211	10,916	131,761	78,348	202,127	210,109	322,972	89,264	412,236	488,234	26,696	324,610	173,492	514,930	498,102	812,844	200,188	1,013,032
174	2011	4	Apr	312,360	14,486	167,448	122,978	326,846	290,426	479,808	137,464	617,272.50	270,107	11,518	143,683	105,781	281,625	249,464	413,790	117,299	531,089	582,467	26,004	311,131	228,759	608,471	539,890	893,598	254,763	1,148,361
175	2011	5	May	360,969	14,159	184,275	133,531	375,128	317,806	545,244	147,690	692,933.70	275,100	7,335	130,161	124,086	282,435	254,247	405,261	131,421	536,682	636,069	21,494	314,436	257,617	657,563	572,053	950,505	279,111	1,229,616
176	2011	6	Jun	333,894	15,529	163,137	128,234	349,423	291,371	497,032	143,763	640,794.60	271,113	8,058	126,588	148,510	279,171	275,098	397,701	156,568	554,269	605,007	23,587	289,725	276,744	628,594	566,469	894,732	300,331	1,195,063
177	2011	7	Jul	357,668	17,632	165,135	147,891	375,300	313,026	522,803	165,523	688,325.55	290,314	9,701	126,968	145,944	300,015	272,912	417,282	155,645	572,927	647,982	27,333	292,103	293,835	675,315	585,938	940,085	321,168	1,261,253
178	2011	8	[illegible]	376,190	14,688	[illegible]	148,062	390,878	332,293	560,421	162,750	723,170.55	[illegible]	[illegible]	121,277	140,183	274,468	261,460	388,475	147,453	535,928	643,388	21,958	305,509	288,245	[illegible]	593,754	948,897	310,203	[illegible]
235	2016	5	May	400,766	[illegible]	162,487	196,213	[illegible]	358,700	563,253	[illegible]	770,409.30	[illegible]	8,353	138,594	162,980	338,992	301,574	469,233	[illegible]	[illegible]	731,405	[illegible]	[illegible]	[illegible]	750,701	660,274	[illegible]	[illegible]	[illegible]
236	2016	6	Jun	355,623	9,554	140,562	170,268	365,177	310,830	496,184	179,823	676,006.90	313,526	7,473	128,099	154,242	320,999	282,341	441,625	161,715	603,340	669,149	17,027	268,661	324,510	686,176	593,171	937,810	341,537	1,279,347
237	2016	7	Jul	368,697	10,757	132,490	175,948	379,454	308,438	501,187	186,705	687,891.80	325,608	9,884	142,812	158,787	335,492	301,599	468,420	168,671	637,091	694,305	20,641	275,302	334,735	714,946	610,037	969,607	355,376	1,324,983
238	2016	8	Aug	411,367	11,823	153,005	222,738	423,190	375,743	564,372	234,561	798,932.75	321,625	8,620	159,247	151,538	330,245	310,785	480,872	160,158	641,030	732,992	20,443	312,252	374,276	753,435	686,528	1,045,244	394,719	1,439,963
239	2016	9	Sep	388,959	11,020	145,096	202,633	399,978	347,729	534,055	213,653	747,707.40	282,945	7,539	120,383	135,937	290,484	256,320	403,328	143,476	546,804	671,904	18,559	265,479	338,570	690,463	604,049	937,383	357,129	1,294,512
240	2016	10	Oct	417,311	10,886	166,406	219,971	428,197	386,378	583,717	230,857	814,574.50	296,711	8,716	126,770	149,612	305,427	276,382	423,481	158,328	581,809	714,022	19,602	293,176	369,583	733,624	662,759	1,007,198	389,185	1,396,383
241	2016	11	Nov	437,050	11,261	177,360	251,893	448,311	429,253	614,410	263,154	877,564.20	270,610	8,555	120,897	134,246	279,165	255,143	391,507	142,801	534,308	707,660	19,816	298,257	386,139	727,476	684,396	1,005,917	405,955	1,411,872
242	2016	12	Dec	394,217	10,664	164,901	226,755	404,881	391,656	559,118	237,419	796,536.25	271,599	7,835	122,933	146,562	279,434	269,495	394,532	154,397	548,929	665,816	18,499	287,834	373,317	684,315	661,151	953,650	391,816	1,345,466
243	2017	1	Jan	367,209	8,007	126,241	202,942	375,216	329,183	493,449	210,949	704,398.45	298,990	7,168	118,234	158,297	306,158	276,531	417,224	165,465	582,689	666,199	15,175	244,475	361,239	681,374	605,714	910,674	376,414	1,287,088
244	2017	2	Feb	372,744	10,624	146,489	183,864	383,368	330,353	519,233	194,488	713,721.00	249,759	5,749	119,811	122,993	255,508	242,804	369,570	128,742	498,312	622,503	16,373	266,300	306,857	638,876	573,157	888,803	323,230	1,212,033
245	2017	3	Mar	287,231	9,618	159,363	156,651	296,850	316,014	446,594	166,269	612,863.20	249,534	6,761	120,435	128,652	256,295	249,087	369,969	135,413	505,382	536,765	16,379	279,798	285,303	553,144	565,101	816,563	301,682	1,118,245
246	2017	4	Apr	343,574	10,629	144,104	157,870	354,204	301,974	487,678	168,499	656,177.05	288,207	6,381	116,260	147,166	294,588	263,426	404,467	153,547	558,014	631,781	17,010	260,364	305,036	648,791	565,400	892,145	322,046	1,214,191
247	2017	5	May	400,766	10,943	162,487	196,213	411,709	358,700	563,253	207,156	770,409.30	336,594	6,294	118,786	186,614	342,888	305,400	455,380	192,908	648,288	737,360	17,237	281,273	382,827	754,597	664,100	1,018,633	400,064	1,418,697
305	2022	3	Mar	495,196	2,580	111,781	349,117	497,776	460,898	606,977	351,697	958,674.05	427,280	11,597	114,185	310,094	438,877	424,279	541,465	321,691	863,156	922,476	14,177	225,966	659,211	936,653	885,177	1,148,442	673,388	1,821,830
306	2022	4	Apr	456,670	1,595	99,878	329,215	458,264	429,093	556,548	330,810	887,357.35	400,803	12,895	121,876	285,145	413,698	407,021	522,679	298,040	820,719	857,473	14,490	221,754	614,360	871,963	836,114	1,079,227	628,850	1,708,077
307	2022	5	May	499,960	1,719	125,656	340,565	501,679	466,221	625,616	342,285	967,900.15	436,977	10,673	118,234	325,105	447,650	443,339	555,211	335,778	890,989	936,937	12,392	243,890	665,670	949,329	909,560	1,180,827	678,062	1,858,889
308	2022	6	Jun	444,680	3,587	93,890	334,454	448,267	428,344	538,570	338,041	876,611.30	415,677	14,644	115,303	289,788	430,321	405,091	530,980	304,432	835,412	860,357	18,231	209,193	624,242	878,588	833,435	1,069,550	642,473	1,712,023
309	2022	7	Jul	485,452	5,257	103,899	340,816	490,709	444,715	589,351	346,073	935,423.80	376,175	14,887	109,411	285,370	391,062	394,781	485,586	300,257	785,843	861,627	20,144	213,310	626,186	881,771	839,496	1,074,937	646,330	1,721,267
310	2022	8	Aug	403,602	2,065	102,319	297,329	405,667	399,648	505,921	299,394	805,314.70	384,530	12,444	121,408	288,558	396,974	409,966	505,938	301,002	806,940	788,132	14,509	223,727	585,887	802,641	809,614	1,011,859	600,396	1,612,255
311	2022	9	Sep	343,462	2,113	77,680	286,618	345,575	364,298	421,142	288,731	709,873.30	342,671	16,693	112,940	269,519	359,364	382,459	455,611	286,212	741,823	686,133	18,806	190,620	556,137	704,939	746,757	876,753	574,943	1,451,696
312	2022	10	Oct	336,307	4,357	89,722	248,044	340,664	337,766	426,029	252,401	678,429.45	293,924	16,518	119,761	228,225	310,442	347,986	413,685	244,743	658,428	630,231	20,875	209,483	476,269	651,106	685,752	839,714	497,144	1,336,858
313	2022	11	Nov	307,080	4,476	90,116	237,672	311,556	327,788	397,196	242,148	639,343.50	259,442	17,767	124,988	186,546	277,209	311,534	384,430	204,313	588,743	566,522	22,243	215,104	424,218	588,765	639,322	781,626	446,461	1,228,087
314	2022	12	Dec	352,046	7,826	96,518	272,481	359,872	368,999	448,564	280,307	728,871.50	241,643	15,982	115,782	170,698	257,625	286,480	357,425	186,680	544,105	593,689	23,808	212,300	443,179	617,497	655,479	805,989	466,987	1,272,976

Using the Excel SUMIF function, the monthly data are summarized and pictorially presented in Table 3. The data are in 1000 TEUs; the last digit was rounded to zero. Details of all Excel formulas with a gray background are shown in the following tables.

Table 3. 26-Year Container Handling at LA/LB Ports



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Year	Year	Loaded Inbound	Empty Inbound	Loaded Outbound	Empty Outbound	Total Inbound	Total Outbound	Total Loaded	Total Empties	Total			
2	1997	1	3270	230	1980	980	3500	2960	5250	1220	6460			
3	1998	2	3810	230	1770	1660	4050	3430	5580	1900	7480			
4	1999	3	4280	160	1810	1970	4440	3780	6090	2130	8220			
5	2000	4	4950	170	2030	2330	5120	4360	6980	2500	9480			
6	2001	5	5100	150	1990	2400	5260	4390	7090	2550	9650			
7	2002	6	5690	170	1950									
8	2003	7	6220	180	2070									
9	2004	8	6930	190	2140									
10	2005	9	7230	210	2390									
11	2006	10	8130	180	2710									
12	2007	11	8110	170	3180									
13	2008	12	7330	190	3470									
14	2009	13	6060	120	3020									
15	2010	14	7100	200	3400									
16	2011	15	7090	280	3620									
17	2012	16	7150	200	3580									
18	2013	17	7430	210	3630									
19	2014	18	7790	250	3540									
20	2015	19	7780	220	3180									
21	2016	20	7990	230	3350									
22	2017	21	8410	200	3290									
23	2018	22	8970	260	3430									
24	2019	23	8470	220	3230	5040	8700	8270	11700	5270	16970			
25	2020	24	8830	200	3010	5270	9020	8280	11830	5460	17300			
26	2021	25	10100	180	2620	7160	10280	9790	12720	7340	20060			
27	2022	26	9330	210	2600	6900	9540	9500	11940	7110	19040			
28														

```
29      B2 =A2-1996
30      C2 =ROUND(SUMIF('1.MonthlyData12(26)'!$A$3:$A$314,'2.YearlyData'!$A2,'1.MonthlyData12(26)'!X$3:X$314)/1000,-1)
31      D2 =ROUND(SUMIF('1.MonthlyData12(26)'!$A$3:$A$314,'2.YearlyData'!$A2,'1.MonthlyData12(26)'!Y$3:Y$314)/1000,-1)
32      E2 =ROUND(SUMIF('1.MonthlyData12(26)'!$A$3:$A$314,'2.YearlyData'!$A2,'1.MonthlyData12(26)'!Z$3:Z$314)/1000,-1)
33      F2 =ROUND(SUMIF('1.MonthlyData12(26)'!$A$3:$A$314,'2.YearlyData'!$A2,'1.MonthlyData12(26)'!AA$3:AA$314)/1000,-1)
34      K2 =ROUND(SUMIF('1.MonthlyData12(26)'!$A$3:$A$314,'2.YearlyData'!$A2,'1.MonthlyData12(26)'!AF$3:AF$314)/1000,-1)
```

3.2 Characteristics of Forecasting Techniques

All forecasting techniques have three main characteristics in common.

(I) Forecasts always deviate from actual observations. Since the world is not deterministic – at least to us – all forecasts are almost always inaccurate Forecasts provide the average value for the variable of interest – sales or demand. Demand is a random variable usually following Poisson distribution estimated by Normal distribution. Thus, besides the average demand, we need a measure of variability- standard deviation, variance, or coefficient of variation. If the average forecast for the next period is F, and the standard deviation of F is S, the coefficient of variation CV= S/F provides a measure of variability; the lower the coefficient of variation, the more confident we are with the forecast.

(II) Forecasts of aggregate values are more accurate than individual item forecasts. Aggregate forecasts reduce variability. The forecast for all container ports in the world is more accurate than the forecast for US container ports, the forecast for US container ports is more accurate than the forecast for California's ports, and the forecast for California's ports is more accurate than the forecast for the port of Oakland in Northern California. Aggregate forecasts reduce the relative variability with respect to the average forecast. One can intuitively understand that the forecast for the summation of two products is more accurate than the forecast for each product because the high demand for one product may compensate for the low demand for the other. From a mathematical point of view, the variance of the sum of two variables is equal to the sum of the variances of the two variables. Therefore, the standard deviation of the summation of the two variables (the numerator of CV) is less than the sum of the two standard deviations. If the standard deviations of the following year's volume of activities in each of LA and LB ports are equal and are shown by σ, then the variance for the volume of activities in the combined port is = $\sigma^2+\sigma^2= 2\sigma^2$. Therefore, the next year's activities volume standard deviation for the combined LA/LB ports is not 2σ but SQRT(2)σ.

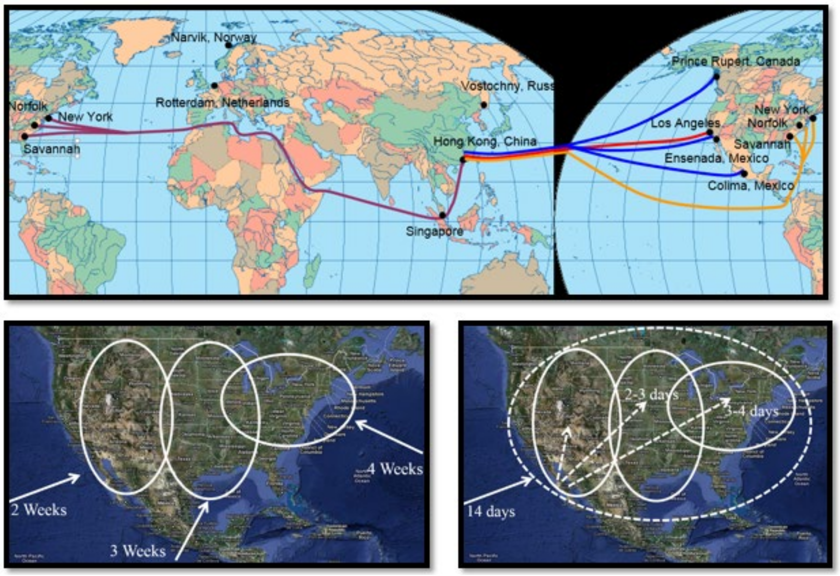
(III) Long-term forecasts are less accurate than short-term forecasts. Forecast accuracy diminishes as we look further into the future. As we get closer to the demand time, we get better information and make better predictions. The forecast for next year's LA/LB activities is more accurate than the forecast for ten years in the future.

3.3. Impact of Characteristics of Forecasting Techniques on LA/LB Ports Throughput.

What are the competing edges of LA/LB ports? Deepwater facilities for post-Panama ships containing close to 20,000 containers? State-of-the-art on-dock facilities to transfer containers between ship and train? Intermodal transfer between sea, rail, and road? Consolidation and distribution facilities for trans-loading from 20- and 40-foot containers to 56-foot containers allowed on California roads? According to Leachman (2010), the characteristics of forecasting techniques are one of the key reasons behind the attractiveness of LA/LB ports.

As pictorially shown in Figure 1, shipping containers from the far east to the East Coast may take 4 weeks. This shipment takes 2 weeks to the west coast and 2-4 weeks from the far-east to the mid-US. For shipments from the far-east to the east coast, one needs to forecast the demand of the east-coast four weeks in advance. But the demand forecast two weeks in advance is enough for shipping to the west coast. According to forecasting characteristics (III), the west-coast demand forecast for east-cost will be more accurate since it is less into the future than east-Asia.

Figure 1. Forecasting-Based Competing Edges of LA/LB Ports



Furthermore, according to forecasting characteristic (II), forecasting the US aggregate demand is more accurate than forecasting demand for any smaller region in the US. Therefore, instead of forecasting for the three regions 14, 21, and 28 days ahead, one may forecast the total US aggregate demand 14 days ahead. It will take 1-3 days to drayage the containers to the three regions. Instead of estimating the demand of the east coast alone, which is less accurate than the demand for the whole US, and instead of forecasting it four weeks ahead, one can forecast for 14+3 days ahead with more accuracy.

4. CURRENT LEVEL AND FORECAST FOR THE NEXT PERIOD

In this section, we estimate the level of demand using moving averages and exponential smoothing. By using these two techniques, we can forecast for the next period. We also provide estimates for the standard deviation of demand. The forecast for all other future periods remains the same as the next period (a straight line) until new data is added. In Section 5, we include trends, and in Section 6, we include seasonality in the levels estimated in this section. All the formulas in all tables are summarized in a set of cells with a gray or white background.

4.1. Basics of Moving Average

Given the annual volume of container handling at the LA/LB ports, a progressive (or naïve) analyst may assume last year's demand as the demand forecast for this year. That is $F_{27} = A_{26}$. A conservative and perhaps irrational analyst may consider the average of all years as the demand forecast for next year. That is $F_5 = (A_4+A_3+A_2+A_1)/4$, $F_6 = (A_5+A_4+A_3+A_2+A_1)/5$, $F_{27} = \text{AVERAGE}(A_{26}+A_{25}+\ldots\ldots+A_2+A_1)$.

Ordinary people, however, may stay between these two extremes and estimate the demand for the next year based on the observations in the past n-periods. An n-period moving average forecast for year 26 is defined as $MA_{26} = \text{AVERAGE}(A_{26}, A_{25,} \ldots\ldots, A_{26-n})$. The forecast for year 27 is then defined as the n-period moving average in year 26; $F_{27} = MA_{26}$. The 4-period moving average forecast in year 27 equals the 4-period moving average in year 26; $F_{27} = (A_{26}+A_{25}+ A_{24}+A_{23})/4$. In general, $F_{t+1} = (A_t+A_{t-1}+ \ldots\ldots + A_{t-n})/n$. Note that the n-period moving average first appears in period n, while the n-period moving average forecast first appears in period n+1. In Appendix A, we develop a general dynamic formula applicable to every n-period moving average, along with dynamic tables and graphs.

4.2. Exponential Smoothing

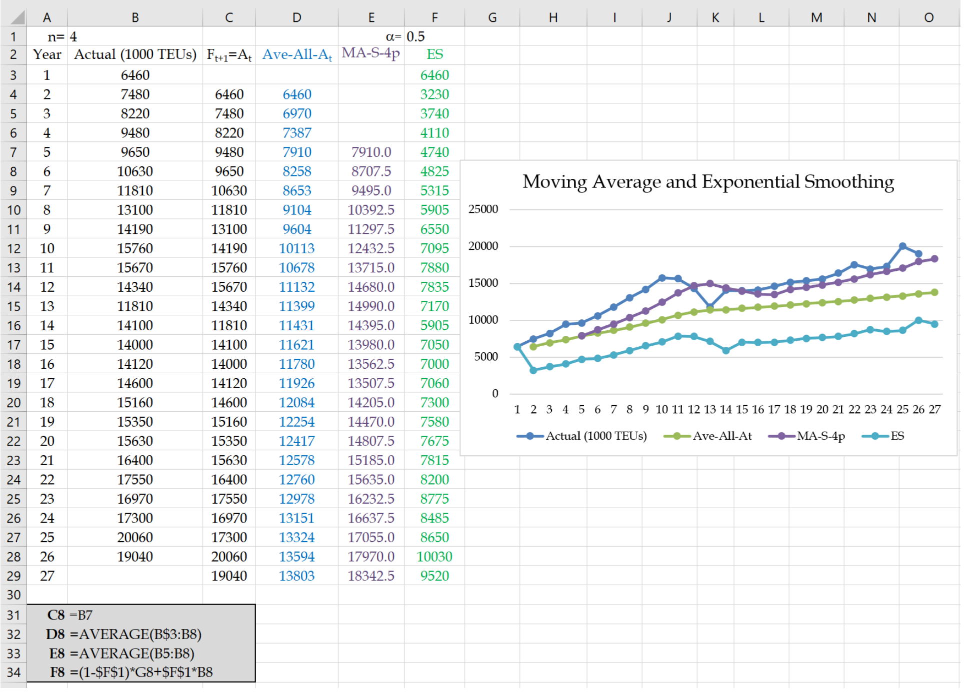
In exponential smoothing, the forecast for the next period equals the forecast for this period plus a fraction of the gap between the actual and forecast values in this period. $F_{t+1} = F_t + \alpha(A_t - F_t)$, where $0 \leq \alpha \leq 1$. A minor manipulation can modify it to $F_{t+1} = (1-\alpha) F_t + \alpha A_t$. That is, the forecast for the next period is the weighted average of the forecast and actual for this period. Exponential smoothing smooths the gap between the actual demand and its forecast.

To start, we need to have a forecast for period 1. There are at least three ways to compute F_1. (*i*) $F_1=A_1$, (*ii*) F_1= average of all existing actual values, (*iii*) F_1= intercept of the linear regression line (discussed later) passing the existing actual values. We assume $F_1=A_1$.

For $\alpha=0.5$, the formula is transformed into $F_{t+1} = 0.5Ft + 0.5A_t = (Ft + A_t)/2$. The forecast for the next period is equal to the average of the actual and the forecast for that period. For $\alpha=1$, the formula is transformed into $F_{t+1} = A_t$. To forecast the next period, we set it equal to the actual for this period. For $\alpha=0$, the formula is transformed into $F_{t+1} = F_t$. To forecast, the next period is the same as the forecast for this period.

We usually start with $\alpha=0.5$ and use an optimization tool, such as Excel's standard SOLVER add-ins or Data Table, to find the optimal α minimizing one of the metrics discussed in the next section. In Appendix B, we show that exponential smoothing is the weighted average of all pieces of data where the weights get smaller and smallest on the older data. Exponential smoothing forecasts using $\alpha=0.5$ are in column F of Table 4. This table also shows the graph for alternative forecasting techniques that can be prepared using Excel's scatter graph or line chart. Key formulas are shown in the gray box.

Table 4. Alternative Moving Average and Exponential Smoothing Forecasts.



	A	B	C	D	E	F
1	n=	4			α=	0.5
2	Year	Actual (1000 TEUs)	$F_{t+1}=A_t$	Ave-All-A_t	MA-S-4p	ES
3	1	6460				6460
4	2	7480	6460	6460		3230
5	3	8220	7480	6970		3740
6	4	9480	8220	7387		4110
7	5	9650	9480	7910	7910.0	4740
8	6	10630	9650	8258	8707.5	4825
9	7	11810	10630	8653	9495.0	5315
10	8	13100	11810	9104	10392.5	5905
11	9	14190	13100	9604	11297.5	6550
12	10	15760	14190	10113	12432.5	7095
13	11	15670	15760	10678	13715.0	7880
14	12	14340	15670	11132	14680.0	7835
15	13	11810	14340	11399	14990.0	7170
16	14	14100	11810	11431	14395.0	5905
17	15	14000	14100	11621	13980.0	7050
18	16	14120	14000	11780	13562.5	7000
19	17	14600	14120	11926	13507.5	7060
20	18	15160	14600	12084	14205.0	7300
21	19	15350	15160	12254	14470.0	7580
22	20	15630	15350	12417	14807.5	7675
23	21	16400	15630	12578	15185.0	7815
24	22	17550	16400	12760	15635.0	8200
25	23	16970	17550	12978	16232.5	8775
26	24	17300	16970	13151	16637.5	8485
27	25	20060	17300	13324	17055.0	8650
28	26	19040	20060	13594	17970.0	10030
29	27		19040	13803	18342.5	9520
30						
31	**C8**	=B7				
32	**D8**	=AVERAGE(B$3:B8)				
33	**E8**	=AVERAGE(B5:B8)				
34	**F8**	=(1-$F$1)*G8+$F$1*B8				

4.4. Age of data in Moving Average and Exponential Smoothing

A 4-period moving average forecast can be computed only after period 4, and then it is set as the forecast for period 5; F_5=MA_4. The newest piece of data in F5 belongs to period 4 and is 1 period old. The oldest data belongs to period 1 and is 4 periods old. Therefore, in a 4-period moving average, the age of data is (1+4)/2 = 2.5 periods. In an n-period moving average, the age of data is (n+1)/2 periods.

The age of data in Exponential Smoothing is 1/α (it is proved in Appendix B). Given 2.5 as the age of data in a 4-period moving average, the data in an exponential smoothing with 1/α =2.5, i.e., α =0.4, has the same age. An exponential smoothing forecast with α = 0.6667 is equivalent to a 2-period moving average forecast, and an exponential smoothing forecast with α = 0.1 is equivalent to a moving average forecast with about 19 periods. The smaller the α (i.e., the larger the number of periods in the moving average), the higher the tendency to smooth out the recent fluctuations. Larger values for α (i.e., the smaller the number of periods in the moving average) result in higher responsiveness to recent fluctuations. A value of α = 1 represents a trend in the past few years and states that the best forecast for the following year's volume of activities is the actual of the current year.

5. Measuring Forecast Accuracy and Variability

In this section, we provide foundations to answer two questions. How to measure the suitability of a forecasting technique for a specific dataset? How can one compare the quality of several forecasting techniques for a specific dataset?

5.1. A Basic Forecast Accuracy and Variability Measure

Given the actual data and forecast (A_t and F_t) and error ($E_t = A_t - F_t$), we define the sum of forecast error SFE = SUM(E_t) and average error BIAS = AVERAGE(E_t). Since the error values are positive or negative, they cross each other out in a forecasting method. Therefore, SFE and BIAS are expected to be small and close to zero. A forecasting approach may be considered of high quality on the foundations of SFE and BIAS. At the same time, there may be significant gaps between actual and forecast values in both positive and negative directions. WE can resolve this problem by considering the absolute value of the gaps. Mean Absolute Deviation (MAD) is defined as MAD = AVERAGE(ABS(E_t)).

MAD serves two essential purposes. First, it compares two or more forecasting techniques and identifies the best based on the lowest MAD value. Second, 1.25MAD provides an estimate of the standard deviation of the demand forecast. We may use any forecasting method to compute F_{t+1} as our estimate for the average demand in the next period. We also provide 1.25 times the most recent MAD as the standard deviation of the forecast for the next period. In other words, $A_{t+1} \sim N(F_{t+1}, 1.25MAD_t)$; demand for the next period follows a normal distribution with an average of F_{t+1} and a standard deviation of $1.25MAD_t$.

Tracking signal (TS) is defined as SFE divided by MAD. It is a positive or negative number divided by a positive number. If the parameters are identified correctly, the summation of all errors has an expected value of zero. TS should be close to zero while jumping up and down on the positive and negative sides due to randomness in the actual data. We can also define the upper control; limit (UCL) and lower control limit (LCL) for TS=SFE/MAD. In some textbooks, it is stated that TS moves between LCL=-4 and UCL=+4. In Appendix B, we will mathematically prove that the limits of ± 4 are incorrect.

TS serves two essential purposes. First, we expect it to stay within UCL and LCL that we define over time. Second, we do not expect a pattern over time. For example, we do not expect to see an always positive or consistently negative TS. In the first case, our forecasting technique underestimates the demand (we have the summation of At-Ft in the numerator); in the second case, it overestimates the demand. We also do not expect to see a cyclic pattern since there may be seasonality in the data that we have not incorporated into our forecasting.

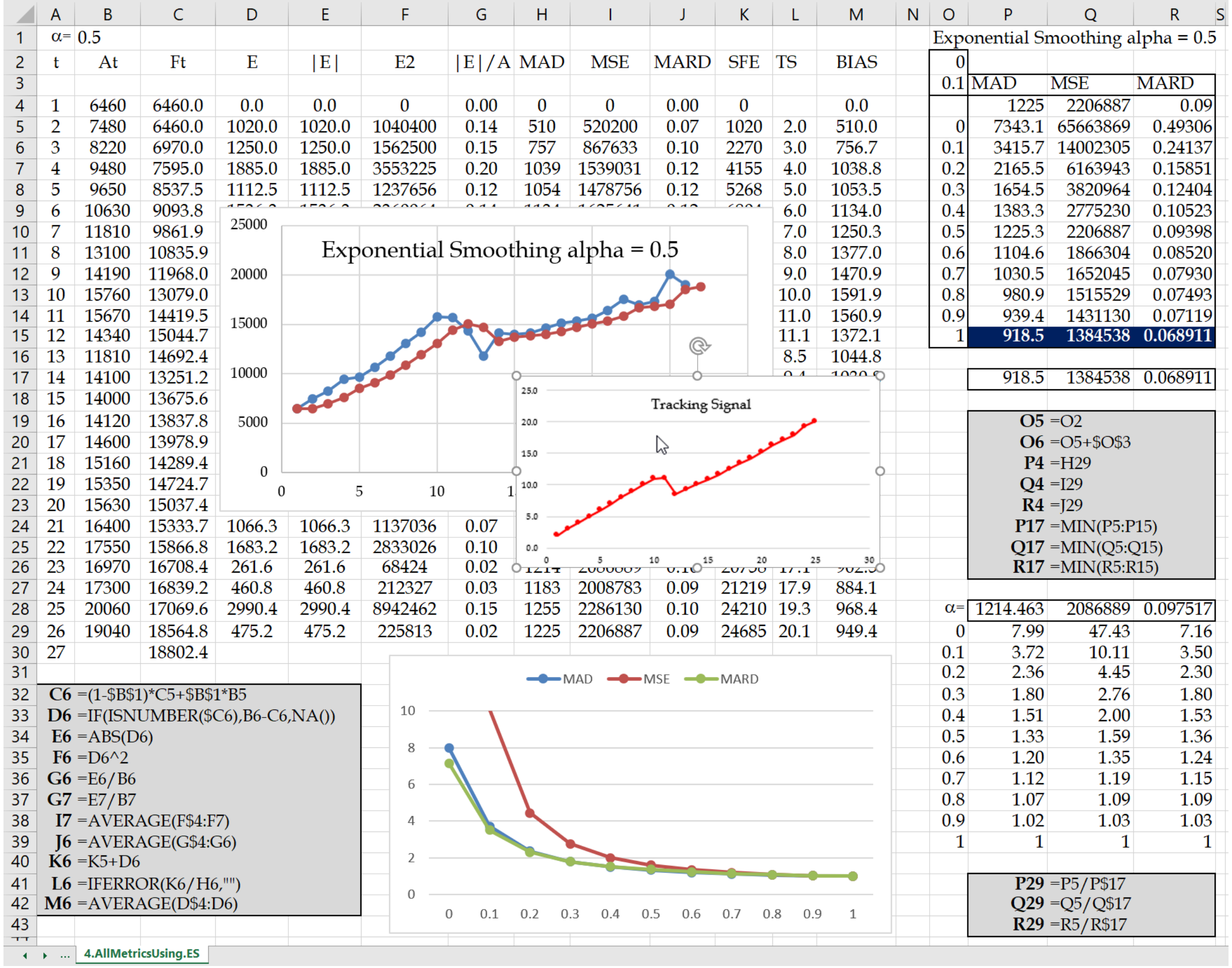
Sometimes we may assign a higher weight to positive gaps than to a negative gap. In the latter case, we are over stock, while in the first case, we have lost sales. Usually, the cost of overstock is less than the cost of lost sales. In these cases, we may assign a coefficient greater than 1 to positive Et=At-Ft values. We may also benefit from the insight into a problem recognized as the newsvendor problem - to find a good tradeoff coefficient of underestimating and overestimating demand.

4.2. Alternative Forecast Accuracy and Variability Measures

An alternative approach to removing negative signs is to square the errors and replace MAD with Mean Squared Error (MSE) = AVERAGE(E_t^2). MSE prevents large gaps between forecast and actual values since the errors are squared. MAD computation was more straightforward when implemented long before calculators and sliding rulers. However, working with an absolute value in mathematical expressions is difficult. It is not difficult to deal with squared values in mathematical expressions. The square root of MSE provides another estimate for the standard deviation of the forecast. That is $A_{t+1} \sim N(F_{t+1}, SQRT(MSE_t))$.

There is also a third method that we refer to it as Mean Absolute Relative Deviation (MARD). Instead of averaging |Et| values, we average $|E_t|/A_t$ values. For example, a $|E_t|$ of 10 states that there were 10 units of deviations between A_t and F_t. If A_t is 200, then 10 relative to 200 is a .05 (or 5% gap). In MARD, the relative absolute gaps (relative to the demand) are computed instead of the absolute gaps. There are still other methods. For example, we may minimize the maximum absolute deviation between actual and forecast. Table 5 shows the computations of error (E), the sum of forecast error (SFE), average error (BIAS), absolute error, MAD, TS, MSE, and MARD for exponential smoothing with α=0.5.

Table 5. All Metrics for Forecast Accuracy and Reliability



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	α=	0.5											
2	t	At	Ft	E	|E|	E2	|E|/A	MAD	MSE	MARD	SFE	TS	BIAS
3													
4	1	6460	6460.0	0.0	0.0	0	0.00	0	0	0.00	0		0.0
5	2	7480	6460.0	1020.0	1020.0	1040400	0.14	510	520200	0.07	1020	2.0	510.0
6	3	8220	6970.0	1250.0	1250.0	1562500	0.15	757	867633	0.10	2270	3.0	756.7
7	4	9480	7595.0	1885.0	1885.0	3553225	0.20	1039	1539031	0.12	4155	4.0	1038.8
8	5	9650	8537.5	1112.5	1112.5	1237656	0.12	1054	1478756	0.12	5268	5.0	1053.5
9	6	10630	9093.8									6.0	1134.0
10	7	11810	9861.9									7.0	1250.3
11	8	13100	10835.9									8.0	1377.0
12	9	14190	11968.0									9.0	1470.9
13	10	15760	13079.0									10.0	1591.9
14	11	15670	14419.5									11.0	1560.9
15	12	14340	15044.7									11.1	1372.1
16	13	11810	14692.4									8.5	1044.8
17	14	14100	13251.2										
18	15	14000	13675.6										
19	16	14120	13837.8										
20	17	14600	13978.9										
21	18	15160	14289.4										
22	19	15350	14724.7										
23	20	15630	15037.4										
24	21	16400	15333.7	1066.3	1066.3	1137036	0.07						
25	22	17550	15866.8	1683.2	1683.2	2833026	0.10						
26	23	16970	16708.4	261.6	261.6	68424	0.02						
27	24	17300	16839.2	460.8	460.8	212327	0.03	1183	2008783	0.09	21219	17.9	884.1
28	25	20060	17069.6	2990.4	2990.4	8942462	0.15	1255	2286130	0.10	24210	19.3	968.4
29	26	19040	18564.8	475.2	475.2	225813	0.02	1225	2206887	0.09	24685	20.1	949.4
30	27		18802.4										

32 **C6** =(1-$B$1)*C5+$B$1*B5
33 **D6** =IF(ISNUMBER($C6),B6-C6,NA())
34 **E6** =ABS(D6)
35 **F6** =D6^2
36 **G6** =E6/B6
37 **G7** =E7/B7
38 **I7** =AVERAGE(F$4:F7)
39 **J6** =AVERAGE(G$4:G6)
40 **K6** =K5+D6
41 **L6** =IFERROR(K6/H6,"")
42 **M6** =AVERAGE(D$4:D6)

	O	P	Q	R
1	Exponential Smoothing alpha = 0.5			
2	0			
3	0.1	MAD	MSE	MARD
4		1225	2206887	0.09
5	0	7343.1	65663869	0.49306
6	0.1	3415.7	14002305	0.24137
7	0.2	2165.5	6163943	0.15851
8	0.3	1654.5	3820964	0.12404
9	0.4	1383.3	2775230	0.10523
10	0.5	1225.3	2206887	0.09398
11	0.6	1104.6	1866304	0.08520
12	0.7	1030.5	1652045	0.07930
13	0.8	980.9	1515529	0.07493
14	0.9	939.4	1431130	0.07119
15	1	918.5	1384538	0.068911
16				
17		918.5	1384538	0.068911

O5 =O2
O6 =O5+$O$3
P4 =H29
Q4 =I29
R4 =J29
P17 =MIN(P5:P15)
Q17 =MIN(Q5:Q15)
R17 =MIN(R5:R15)

	O	P	Q	R
28	α=	1214.463	2086889	0.097517
29	0	7.99	47.43	7.16
30	0.1	3.72	10.11	3.50
31	0.2	2.36	4.45	2.30
32	0.3	1.80	2.76	1.80
33	0.4	1.51	2.00	1.53
34	0.5	1.33	1.59	1.36
35	0.6	1.20	1.35	1.24
36	0.7	1.12	1.19	1.15
37	0.8	1.07	1.09	1.09
38	0.9	1.02	1.03	1.03
39	1	1	1	1

P29 =P5/P$17
Q29 =Q5/Q$17
R29 =R5/R$17

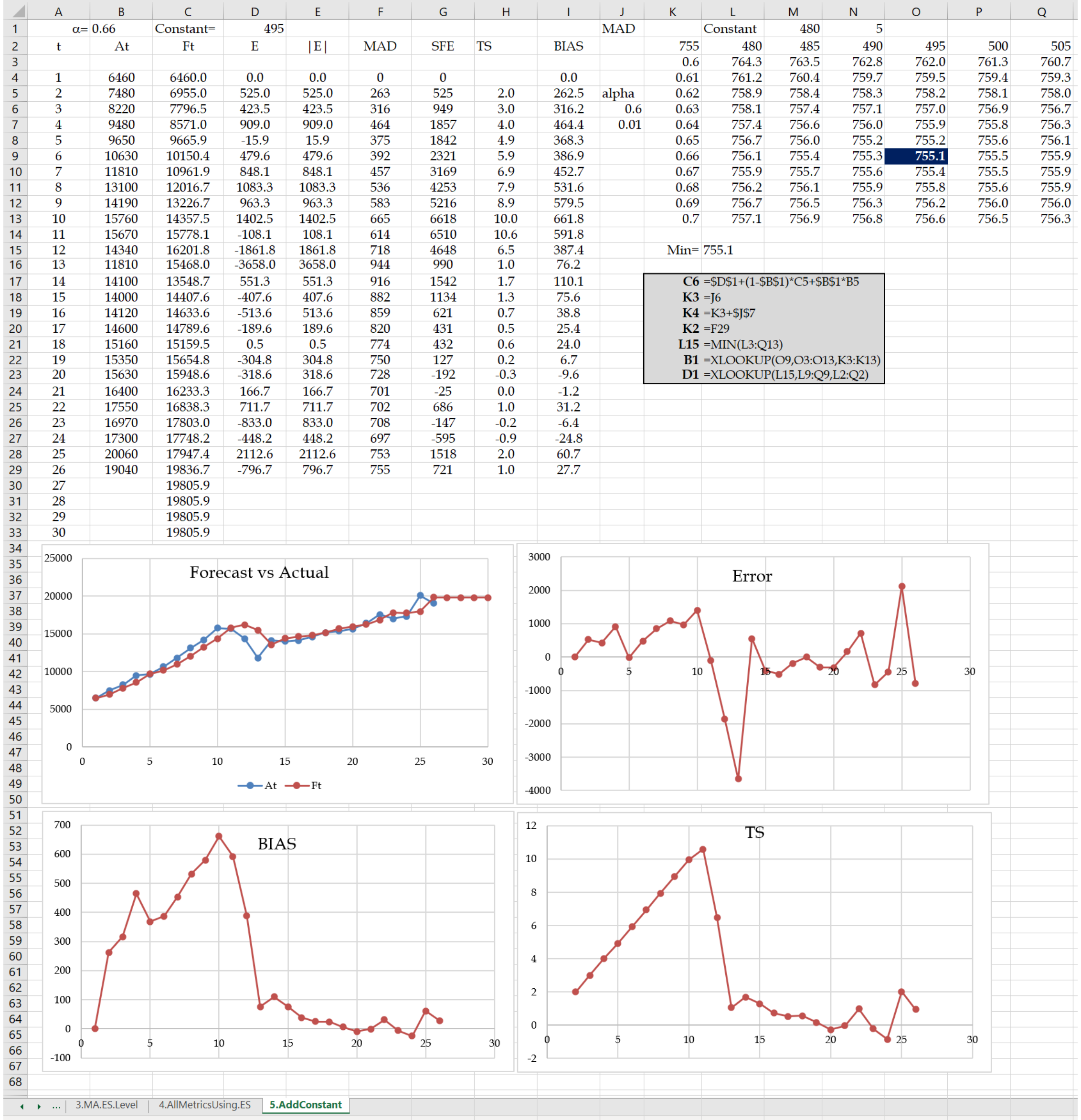
4.4. Optimal α Value

The optimal α value can be computed in at least two ways. (*i*) SOLVER and (*ii*) Data Table. For SOLVER, the objective function is set to one of the three measures of MAD, MSE, or MARD (in cells H29, I29, and J29) to be minimized, and α cell B1 is the changing cell to minimize the objective function value. For the Data Table, we set cells P4, Q4, and R4 equal to cells H29, I29, and J29, respectively. The set of α values are typed one column to the left of MAD and start from one cell below MAD. Using a formula, we can find the value of α in the Data Table to as many as the decimal point that may be desired in SOLVER. This is done by typing the starting α value of 0 and the increment in two arbitrary cells (such as cells O2 and O3 in this example). We then set O5=O2 and O6=O2+$O$3 and copy down from 0 to 1. After setting O4 to R15, → Data → What-if Analysis → Data Table. Since alternative α values are typed in a column (not in a row), inside the column input cell, we point to B1, where the α value is placed. We then find the α value corresponding to the minimal MAD (or MSE or MARD) value. Suppose the α value for the minimal MAD is 0.7. To estimate α with more decimal points we can set cell O2 to 0.65 and O3 to 0.001 and find the minimal α in the range of 0.65 to 0.74. We can continue this procedure to as many decimal points as we wish; to find answers as precisely as SOLVER with Data Table.

Optimal α computations using both solver and Data Table for all three metrics) and normalization (divide each by the minimal value in that column) of these metrics as (changes are included in Table 5). The reader is encouraged to look into all the formulas in gray cells. We have also used conditional formatting to highlight the minimal values. The Tracking Signal curve for α =0.5 is also shown in Table 5.

The reason for an upward tracking signal is the positive overall trend of actual data. That is why the moving average recommends n=1, and exponential smoothing recommends α=1. When the tracking signal shows a continual or increasing positive trend, we may add a constant to the forecast value. In Table 6, we implemented a two-dimensional Data Table to find the optimal value for α = 0.66 plus a constant of 495 to be added to the forecast to minimize MAD. The computations for exponential smoothing and the essential formulas are shown in Table 6.

Table 6. Forecasting Measures under Optimal α and a Constant for Exponential Smoothing

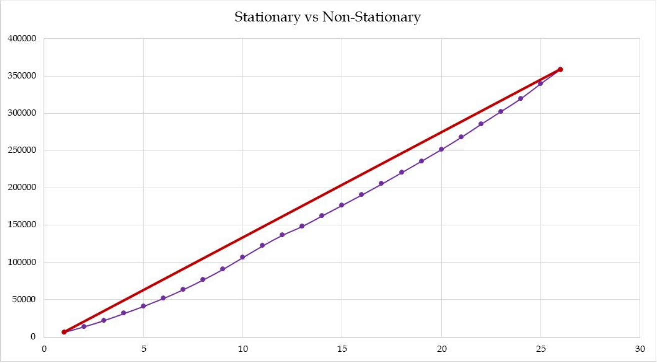


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	α=	0.66	Constant=	495						MAD		Constant	480	5			
2	t	At	Ft	E	|E|	MAD	SFE	TS	BIAS		755	480	485	490	495	500	505
3											0.6	764.3	763.5	762.8	762.0	761.3	760.7
4	1	6460	6460.0	0.0	0.0	0	0		0.0		0.61	761.2	760.4	759.7	759.5	759.4	759.3
5	2	7480	6955.0	525.0	525.0	263	525	2.0	262.5	alpha	0.62	758.9	758.4	758.3	758.2	758.1	758.0
6	3	8220	7796.5	423.5	423.5	316	949	3.0	316.2	0.6	0.63	758.1	757.4	757.1	757.0	756.9	756.7
7	4	9480	8571.0	909.0	909.0	464	1857	4.0	464.4	0.01	0.64	757.4	756.6	756.0	755.9	755.8	756.3
8	5	9650	9665.9	-15.9	15.9	375	1842	4.9	368.3		0.65	756.7	756.0	755.2	755.2	755.6	756.1
9	6	10630	10150.4	479.6	479.6	392	2321	5.9	386.9		0.66	756.1	755.4	755.3	755.1	755.5	755.9
10	7	11810	10961.9	848.1	848.1	457	3169	6.9	452.7		0.67	755.9	755.7	755.6	755.4	755.5	755.9
11	8	13100	12016.7	1083.3	1083.3	536	4253	7.9	531.6		0.68	756.2	756.1	755.9	755.8	755.6	755.9
12	9	14190	13226.7	963.3	963.3	583	5216	8.9	579.5		0.69	756.7	756.5	756.3	756.2	756.0	756.0
13	10	15760	14357.5	1402.5	1402.5	665	6618	10.0	661.8		0.7	757.1	756.9	756.8	756.6	756.5	756.3
14	11	15670	15778.1	-108.1	108.1	614	6510	10.6	591.8								
15	12	14340	16201.8	-1861.8	1861.8	718	4648	6.5	387.4		Min=	755.1					
16	13	11810	15468.0	-3658.0	3658.0	944	990	1.0	76.2								
17	14	14100	13548.7	551.3	551.3	916	1542	1.7	110.1		**C6**	=$D$1+(1-$B$1)*C5+$B$1*B5					
18	15	14000	14407.6	-407.6	407.6	882	1134	1.3	75.6		**K3**	=J6					
19	16	14120	14633.6	-513.6	513.6	859	621	0.7	38.8		**K4**	=K3+$J$7					
20	17	14600	14789.6	-189.6	189.6	820	431	0.5	25.4		**K2**	=F29					
21	18	15160	15159.5	0.5	0.5	774	432	0.6	24.0		**L15**	=MIN(L3:Q13)					
22	19	15350	15654.8	-304.8	304.8	750	127	0.2	6.7		**B1**	=XLOOKUP(O9,O3:O13,K3:K13)					
23	20	15630	15948.6	-318.6	318.6	728	-192	-0.3	-9.6		**D1**	=XLOOKUP(L15,L9:Q9,L2:Q2)					
24	21	16400	16233.3	166.7	166.7	701	-25	0.0	-1.2								
25	22	17550	16838.3	711.7	711.7	702	686	1.0	31.2								
26	23	16970	17803.0	-833.0	833.0	708	-147	-0.2	-6.4								
27	24	17300	17748.2	-448.2	448.2	697	-595	-0.9	-24.8								
28	25	20060	17947.4	2112.6	2112.6	753	1518	2.0	60.7								
29	26	19040	19836.7	-796.7	796.7	755	721	1.0	27.7								
30	27		19805.9														
31	28		19805.9														
32	29		19805.9														
33	30		19805.9														

4.5. Stationary vs. Non-Stationary Data.

In our dataset, the optimal α for all three metrics is equal (this is not the case most of the time) and is equal to 1 (this is not a general observation). Since we have an upward trend almost in all years, an α=1, and therefore $F_{t+1}=A_t$ is the best solution. Moving average and Exponential Smoothing are appropriate for stationary data. We can draw the Cum_t = SUM(A_t) function to check whether data is stationary. The data is stationary if Cum_t is close to a line. Figure 2 shows Cum_t for our data is distant from a line. We will later discuss trend-adjusted exponential smoothing and regression for data with a trend.

Figure 2. Stationary vs. Non-Stationary Data



5. LEVEL AND TREND

This section reviews (*i*) Bi-variate linear regression and (*ii*) Trend adjusted exponential smoothing.

5.1. Bi-variable Linear Regression.

The bi-variable linear regression is generally stated as $y=b_0+b_1x$. Our specific case can be stated as $F_t = b_0 + b_1t$. While we could have continued with the actual years, we set t equal to the current year minus 1996 for simplicity. Nevertheless, no matter how we enumerate the years, while we will have different values for b_0 and b_1, all the analyses and the shape of the regression line remain the same. Alternative linear regression computations are explained below and are summarized in Table 7.

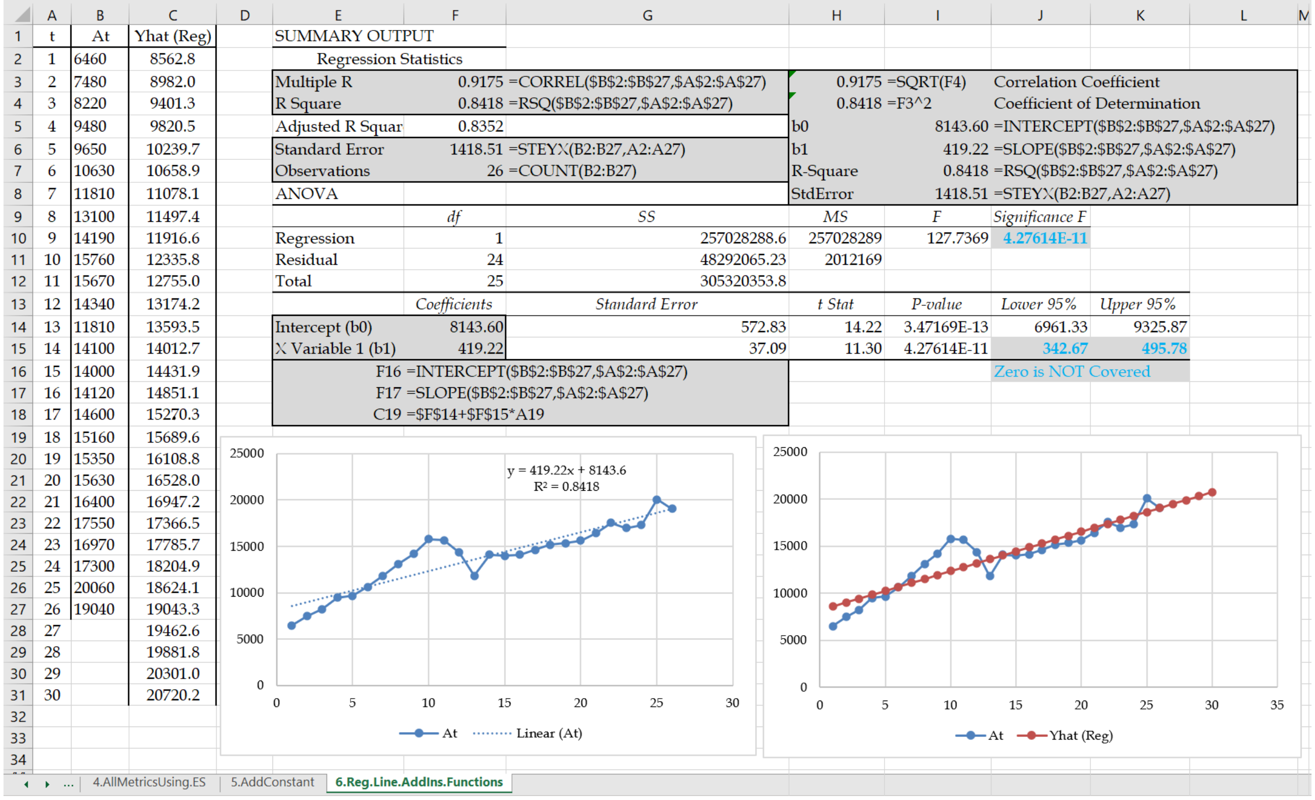
Procedure-1. Add Trend Line. After drawing the data in a scatter graph, we can right-click on the graph and choose to add a trendline. Options of exponential, linear, logarithmic, polynomial, power, and moving average will appear. We chose liner. We also check the display equation and display the R-squared value on the chart. The scatter graph shows the regression equation y = 419.22x + 8143.6 and $R^2 = 0.8418$.

Procedure-2. Data Analysis Add-Ins. Choose Data Tab →Data Analysis → Regression. In the next table, enter the Y variable (A_t), then X variables (t), and select the cell that will be in the east-north of the table (we select cell E1). This approach is not recommended for bi-variable linear regression since we must prepare a new table once a number is changed. Excel functions perform better and are updated as a change is made in the data. As it is shown in the seasonality-enhanced multi-variable regression, Data Analysis Add-Ins is a good choice for bi-variable non-linear and multi-variable linear and non-linear cases,

Procedure 3. Excel Functions. Excel formulas are entered into the Data Analysis Add-Ins output to simplify explanations. There are mainly INTERCEPT, SLOPE, RSQ, STEYX, and CORREL functions that we have added to the table. All the formulas are shown in the gray cells. The larger the R-square ($0 \le R^2 \le 1$), the more reliable the regression line. If the distance between the two blue numbers in the bottom part of the table does not cover zero, there is a relationship between Y and X ($b_1 \neq 0$). If the blues number in the top part of the table is less than 0.05, with more than 95% confidence, not both b_0 and b_1 are zero.

Procedure 4. Using More Fundamental Computations in Excel. In Appendix C, we will provide fundamental insight into the computation of regression metrics through computing SST, SSE, and SSR, as well as a piecewise regression.

Table 7. Alternative Linear Regression Computations



	A	B	C	D	E	F	G	H	I	J	K
1	t	At	Yhat (Reg)		SUMMARY OUTPUT						
2	1	6460	8562.8		Regression Statistics						
3	2	7480	8982.0		Multiple R	0.9175	=CORREL($B$2:$B$27,$A$2:$A$27)	0.9175	=SQRT(F4)	Correlation Coefficient	
4	3	8220	9401.3		R Square	0.8418	=RSQ($B$2:$B$27,$A$2:$A$27)	0.8418	=F3^2	Coefficient of Determination	
5	4	9480	9820.5		Adjusted R Squar	0.8352		b0	8143.60	=INTERCEPT($B$2:$B$27,$A$2:$A$27)	
6	5	9650	10239.7		Standard Error	1418.51	=STEYX(B2:B27,A2:A27)	b1	419.22	=SLOPE($B$2:$B$27,$A$2:$A$27)	
7	6	10630	10658.9		Observations	26	=COUNT(B2:B27)	R-Square	0.8418	=RSQ($B$2:$B$27,$A$2:$A$27)	
8	7	11810	11078.1		ANOVA			StdError	1418.51	=STEYX(B2:B27,A2:A27)	
9	8	13100	11497.4			*df*	*SS*	*MS*	*F*	*Significance F*	
10	9	14190	11916.6		Regression	1	257028288.6	257028289	127.7369	**4.27614E-11**	
11	10	15760	12335.8		Residual	24	48292065.23	2012169			
12	11	15670	12755.0		Total	25	305320353.8				
13	12	14340	13174.2			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
14	13	11810	13593.5		Intercept (b0)	8143.60	572.83	14.22	3.47169E-13	6961.33	9325.87
15	14	14100	14012.7		X Variable 1 (b1)	419.22	37.09	11.30	4.27614E-11	**342.67**	**495.78**
16	15	14000	14431.9		F16 =INTERCEPT($B$2:$B$27,$A$2:$A$27)					Zero is NOT Covered	
17	16	14120	14851.1		F17 =SLOPE($B$2:$B$27,$A$2:$A$27)						
18	17	14600	15270.3		C19 =$F$14+$F$15*A19						
19	18	15160	15689.6								
20	19	15350	16108.8								
21	20	15630	16528.0								
22	21	16400	16947.2								
23	22	17550	17366.5								
24	23	16970	17785.7								
25	24	17300	18204.9								
26	25	20060	18624.1								
27	26	19040	19043.3								
28	27		19462.6								
29	28		19881.8								
30	29		20301.0								
31	30		20720.2								

5.2. Trend Adjusted Exponential Smoothing.

Trend-adjusted exponential smoothing is defined as $F_{t+1} = L_t + T_t$, where Lt and T_t are the level and trend in period t as defined in Chopra (2019) based on Holt (1957).

$$L_{t+1} = \alpha A_t + (1-\alpha)(L_t + T_t)$$

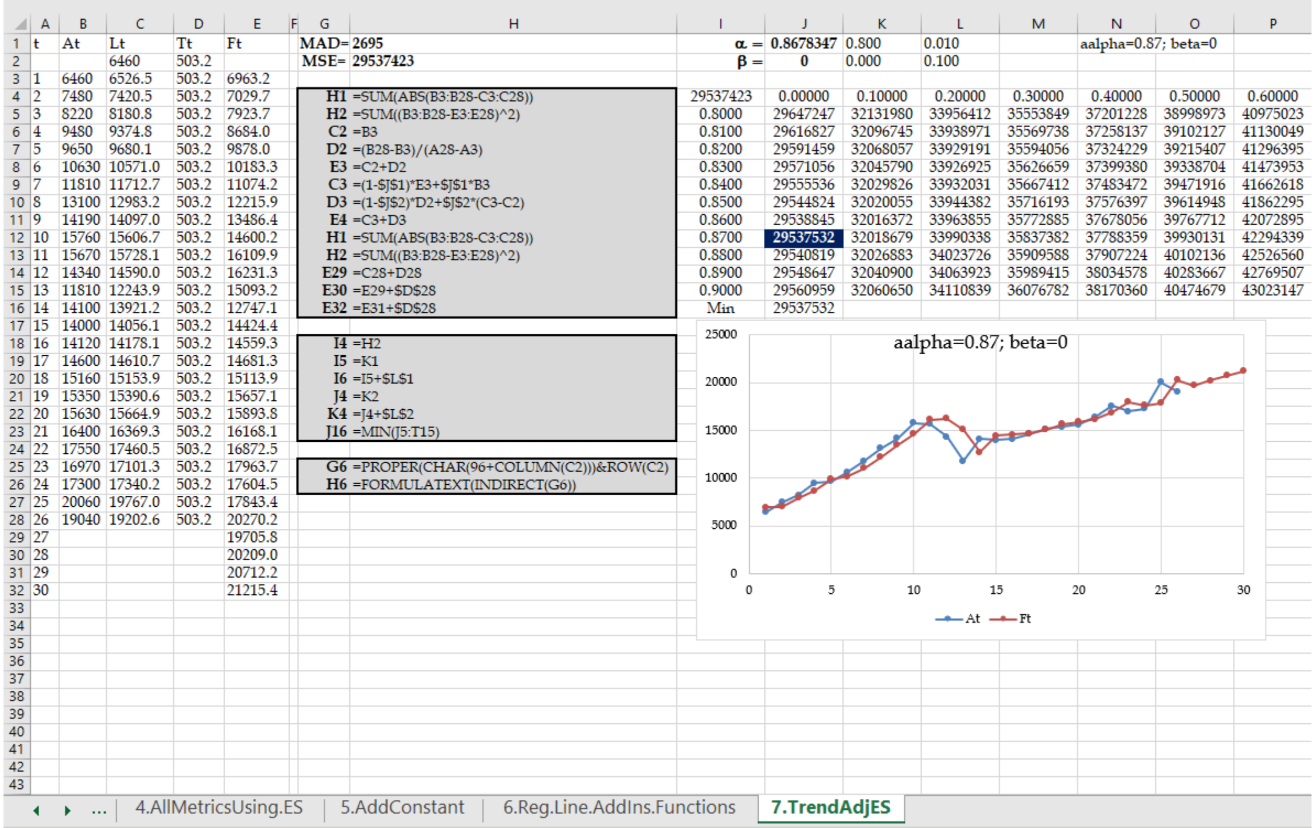
$$T_{t+1} = \beta(L_{t+1} - L_t) + (1-\beta)T_t$$

Trend-adjusted exponential smoothing, or double exponential smoothing, smooths out the level and trend of this period based on the level and trend of the previous period and the actual observation in this period.

Starting L_0 and T_0 can be computed in two ways. We may set L_0 as the demand in the first period and T_0 as the demand of the last period minus the demand of the first period divided by (N-1). In our case, $L_0 = A_1 = 6460$, and $T_0 = (A_{26} - A1)/(26-1) = 503.2$ (proposed in Iravani, 2021). Alternatively, we may set L_0 as the intercept of the regression line and T_0 as its slope. $L_0 = b_0 = 8143.6$, and $T_0 = b_1 = 419.2$ (proposed in Chopra 2019). We follow the first approach. We start from $\alpha = 0.5$ and $\beta = 0.5$ and then use SOLVER or a two-dimensional Data Table to find the optimal values of $\alpha = 0.87$ and $\beta = 0$, as shown in Table 8. Compared to simple exponential smoothing, the MSE and other metrics are lower, and the extension to future periods carries a trend and is not a straight line. Compared to regression, we have a smooth curve going up and down instead of a straight line.

We can also combine linear regression and trend-adjusted exponential smoothing in the form of $F_t = \gamma F_{\text{Trend-Adjusted.ES}} + (1-\gamma) F_{\text{Linear-Regression}}$. The optimal γ value minimizing the MSE of the forecasts from the actual values can then be obtained using SOLVER or Data Table.

Table 8. Trend Adjusted Exponential Smoothing Computations



	A	B	C	D	E
1	t	At	Lt	Tt	Ft
2			6460	503.2	
3	1	6460	6526.5	503.2	6963.2
4	2	7480	7420.5	503.2	7029.7
5	3	8220	8180.8	503.2	7923.7
6	4	9480	9374.8	503.2	8684.0
7	5	9650	9680.1	503.2	9878.0
8	6	10630	10571.0	503.2	10183.3
9	7	11810	11712.7	503.2	11074.2
10	8	13100	12983.2	503.2	12215.9
11	9	14190	14097.0	503.2	13486.4
12	10	15760	15606.7	503.2	14600.2
13	11	15670	15728.1	503.2	16109.9
14	12	14340	14590.0	503.2	16231.3
15	13	11810	12243.9	503.2	15093.2
16	14	14100	13921.2	503.2	12747.1
17	15	14000	14056.1	503.2	14424.4
18	16	14120	14178.1	503.2	14559.3
19	17	14600	14610.7	503.2	14681.3
20	18	15160	15153.9	503.2	15113.9
21	19	15350	15390.6	503.2	15657.1
22	20	15630	15664.9	503.2	15893.8
23	21	16400	16369.3	503.2	16168.1
24	22	17550	17460.5	503.2	16872.5
25	23	16970	17101.3	503.2	17963.7
26	24	17300	17340.2	503.2	17604.5
27	25	20060	19767.0	503.2	17843.4
28	26	19040	19202.6	503.2	20270.2
29	27				19705.8
30	28				20209.0
31	29				20712.2
32	30				21215.4

	G	H
1	MAD=	2695
2	MSE=	29537423
4	H1	=SUM(ABS(B3:B28-C3:C28))
5	H2	=SUM((B3:B28-E3:E28)^2)
6	C2	=B3
7	D2	=(B28-B3)/(A28-A3)
8	E3	=C2+D2
9	C3	=(1-$J$1)*E3+$J$1*B3
10	D3	=(1-$J$2)*D2+$J$2*(C3-C2)
11	E4	=C3+D3
12	H1	=SUM(ABS(B3:B28-C3:C28))
13	H2	=SUM((B3:B28-E3:E28)^2)
14	E29	=C28+D28
15	E30	=E29+$D$28
16	E32	=E31+$D$28
18	I4	=H2
19	I5	=K1
20	I6	=I5+$L$1
21	J4	=K2
22	K4	=J4+$L$2
23	J16	=MIN(J5:T15)
25	G6	=PROPER(CHAR(96+COLUMN(C2)))&ROW(C2)
26	H6	=FORMULATEXT(INDIRECT(G6))

	I	J	K	L	M	N	O	P
1	α =	0.8678347	0.800	0.010		aalpha=0.87; beta=0		
2	β =	0	0.000	0.100				
4	29537423	0.00000	0.10000	0.20000	0.30000	0.40000	0.50000	0.60000
5	0.8000	29647247	32131980	33956412	35553849	37201228	38998973	40975023
6	0.8100	29616827	32096745	33938971	35569738	37258137	39102127	41130049
7	0.8200	29591459	32068057	33929191	35594056	37324229	39215407	41296395
8	0.8300	29571056	32045790	33926925	35626659	37399380	39338704	41473953
9	0.8400	29555536	32029826	33932031	35667412	37483472	39471916	41662618
10	0.8500	29544824	32020055	33944382	35716193	37576397	39614948	41862295
11	0.8600	29538845	32016372	33963855	35772885	37678056	39767712	42072895
12	0.8700	29537532	32018679	33990338	35837382	37788359	39930131	42294339
13	0.8800	29540819	32026883	34023726	35909588	37907224	40102136	42526560
14	0.8900	29548647	32040900	34063923	35989415	38034578	40283667	42769507
15	0.9000	29560959	32060650	34110839	36076782	38170360	40474679	43023147
16	Min	29537532						

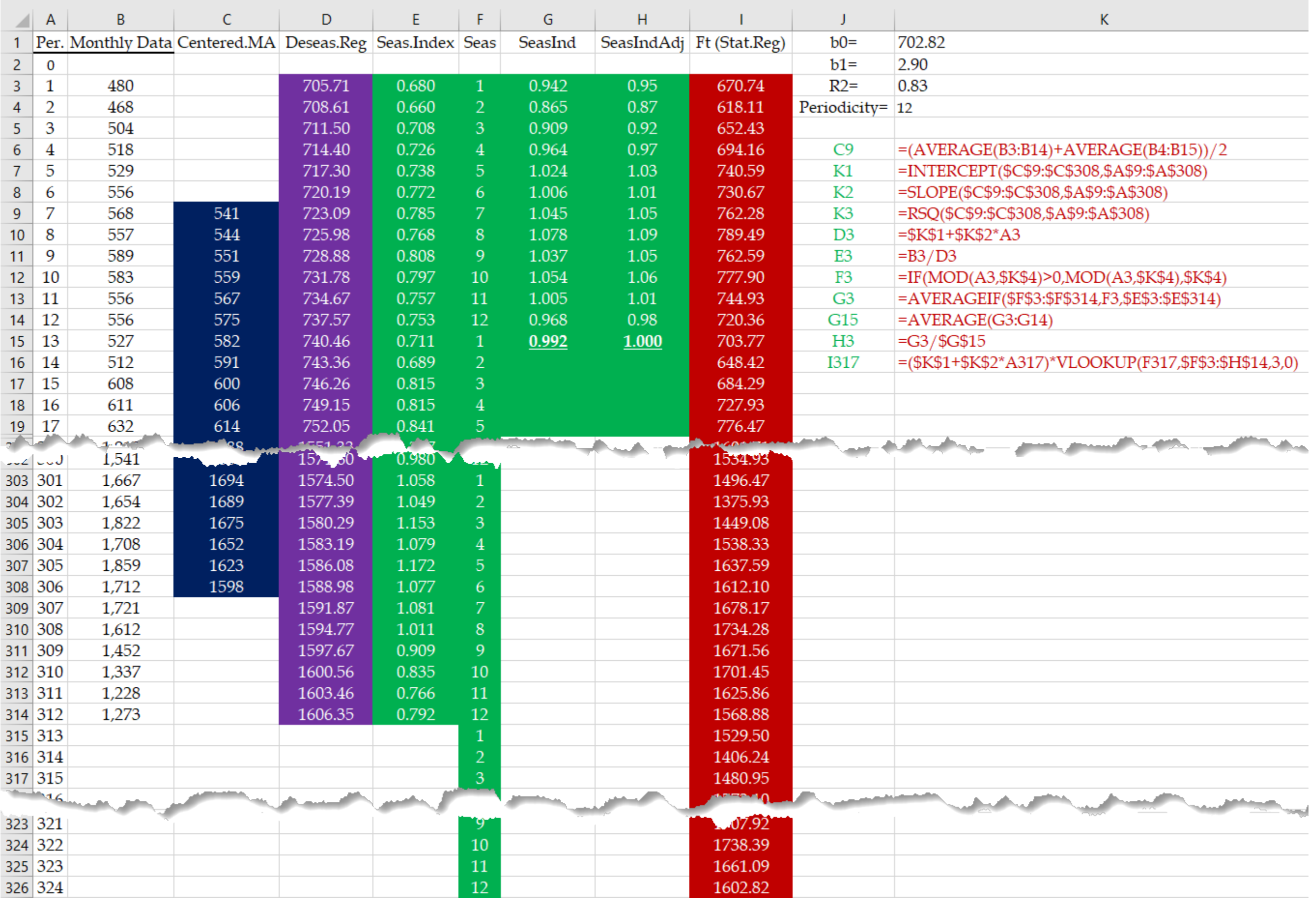
6. LEVEL, TREND, AND SEASONALITY

In this section, we review (*i*) seasonality-enhanced bi-variable linear regression, (*ii*) seasonality-enhanced multi-variable regression using dummy variables, and (*iii*) trend and seasonality-adjusted exponential smoothing.

6.1. Seasonality Enhanced Bi-Variable Linear Regression.

Our approach is recognized as Winter's Model (Chopra 2019). The monthly data shown in Table 2 for 12(26) months (in 1000 TEUs) are copied into Table 9. We consider a periodicity of 12, where periods repeat every 12 months. One may add three months of data and consider the periodicity of four seasons, periodicity of 7 days over a week, or periodicity of 24 hours a day.

Table 9. Computations for Static Seasonality Enhanced Bi-Variate Linear Regression



	A	B	C	D	E	F	G	H	I	J	K
1	Per.	Monthly Data	Centered.MA	Deseas.Reg	Seas.Index	Seas	SeasInd	SeasIndAdj	Ft (Stat.Reg)	b0=	702.82
2	0									b1=	2.90
3	1	480		705.71	0.680	1	0.942	0.95	670.74	R2=	0.83
4	2	468		708.61	0.660	2	0.865	0.87	618.11	Periodicity=	12
5	3	504		711.50	0.708	3	0.909	0.92	652.43		
6	4	518		714.40	0.726	4	0.964	0.97	694.16	C9	=(AVERAGE(B3:B14)+AVERAGE(B4:B15))/2
7	5	529		717.30	0.738	5	1.024	1.03	740.59	K1	=INTERCEPT($C$9:$C$308,$A$9:$A$308)
8	6	556		720.19	0.772	6	1.006	1.01	730.67	K2	=SLOPE($C$9:$C$308,$A$9:$A$308)
9	7	568	541	723.09	0.785	7	1.045	1.05	762.28	K3	=RSQ($C$9:$C$308,$A$9:$A$308)
10	8	557	544	725.98	0.768	8	1.078	1.09	789.49	D3	=$K$1+$K$2*A3
11	9	589	551	728.88	0.808	9	1.037	1.05	762.59	E3	=B3/D3
12	10	583	559	731.78	0.797	10	1.054	1.06	777.90	F3	=IF(MOD(A3,$K$4)>0,MOD(A3,$K$4),$K$4)
13	11	556	567	734.67	0.757	11	1.005	1.01	744.93	G3	=AVERAGEIF($F$3:$F$314,F3,$E$3:$E$314)
14	12	556	575	737.57	0.753	12	0.968	0.98	720.36	G15	=AVERAGE(G3:G14)
15	13	527	582	740.46	0.711	1	**0.992**	**1.000**	703.77	H3	=G3/$G$15
16	14	512	591	743.36	0.689	2			648.42	I317	=($K$1+$K$2*A317)*VLOOKUP(F317,$F$3:$H$14,3,0)
17	15	608	600	746.26	0.815	3			684.29		
18	16	611	606	749.15	0.815	4			727.93		
19	17	632	614	752.05	0.841	5			776.47		
		1,541		[illegible]	0.980				1534.93		
303	301	1,667	1694	1574.50	1.058	1			1496.47		
304	302	1,654	1689	1577.39	1.049	2			1375.93		
305	303	1,822	1675	1580.29	1.153	3			1449.08		
306	304	1,708	1652	1583.19	1.079	4			1538.33		
307	305	1,859	1623	1586.08	1.172	5			1637.59		
308	306	1,712	1598	1588.98	1.077	6			1612.10		
309	307	1,721		1591.87	1.081	7			1678.17		
310	308	1,612		1594.77	1.011	8			1734.28		
311	309	1,452		1597.67	0.909	9			1671.56		
312	310	1,337		1600.56	0.835	10			1701.45		
313	311	1,228		1603.46	0.766	11			1625.86		
314	312	1,273		1606.35	0.792	12			1568.88		
315	313					1			1529.50		
316	314					2			1406.24		
317	315					3			1480.95		
323	321					9			[illegible]		
324	322					10			1738.39		
325	323					11			1661.09		
326	324					12			1602.82		

When we compute the average of 12 months, it is pure of seasonality since high and low seasons cross each other out. This is true for any other periodicity; the average of all seasons does not contain seasonality. Instead of placing the moving average of n seasons in front of the last season (as we did in our moving average computations), we place it at the center of the data incorporated in each moving average; centered moving average.

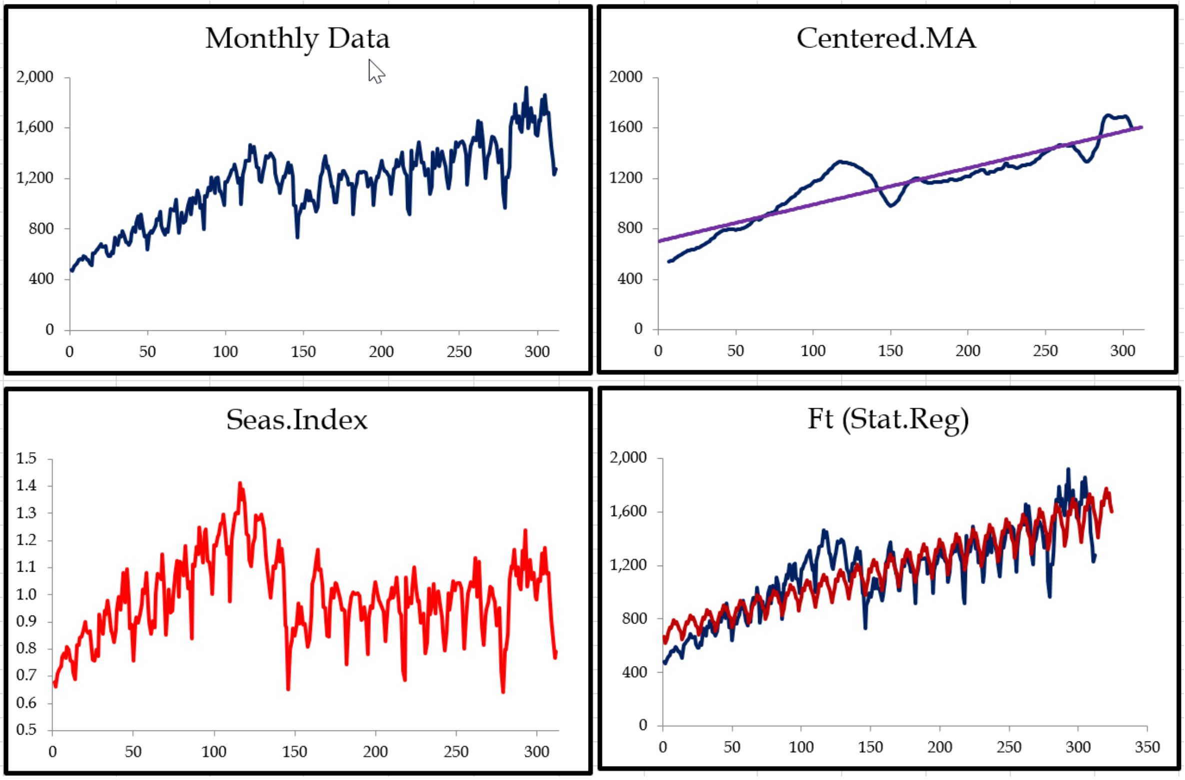
Step 1. Removing Seasonality. If we were considering seasonality over 7 days of weeks since 7 is odd, we could have placed the average in front of period 4, compared the actual period 4 with the centered moving average, and estimated the seasonality of period 4. But there is no middle period for even periodicity. Therefore (and the procedure is the same for all other even periodicities), we first compute the average of the 12 months and assume it is placed at the boundary of months 6 and 7. We also compute the period 2 to period 13 average and assume it is at the boundary of months 7 and 8. Next, we compute the average of these two centered moving averages and place it in front of period 7, representing the unseasonal activity volume at period 7. We then copy this formula down to 6 months to the last months. We will generally have the centered moving average for all periods minus periodicity.

Step 2. Trend in the Deseasonalized Data. We apply linear regression on months 7 to 306 to find the level and trend of the data pure of seasonality. It leads to b_0, b_1, and R^2, as shown in columns K of Table 9. The Excel worksheet also shows the formulas for all other computations (as they follow).

Step 3. Seasonality Indices. We divide the actual data of each month by the value obtained from the regression line applied to the deseasonalized data ($A_t/\underline{Y}_t$). The ratios are estimates of the seasonality index in all 12(26) months. By averaging all seasonality indices of each month, the average seasonality index of January (S_1) to December (S_{12}) is computed. The average of the average seasonality indices for all 12 months must equal 1; therefore, to normalize, we divide the average seasonality index of each month by the average of the averages. These computations are in columns G and H.

Step 4. Trend and Seasonality Adjusted Forecasting. Finally, we put seasonality back on the deseasonalized regression line and forecast the future. $F_t = (b_0+b_1t)*S_t$, where S_t has the same monthly value over all years. All formulas are clearly explained in Table 9. The results of the four steps of this process are schematically represented in Figure 3. The above analysis shows that the monthly seasonality is from a minimum of 0.87 to a maximum of 1.09. In a similar analysis, one may study daily seasonality (periodicity of 30) or hourly seasonality (periodicity of 24).

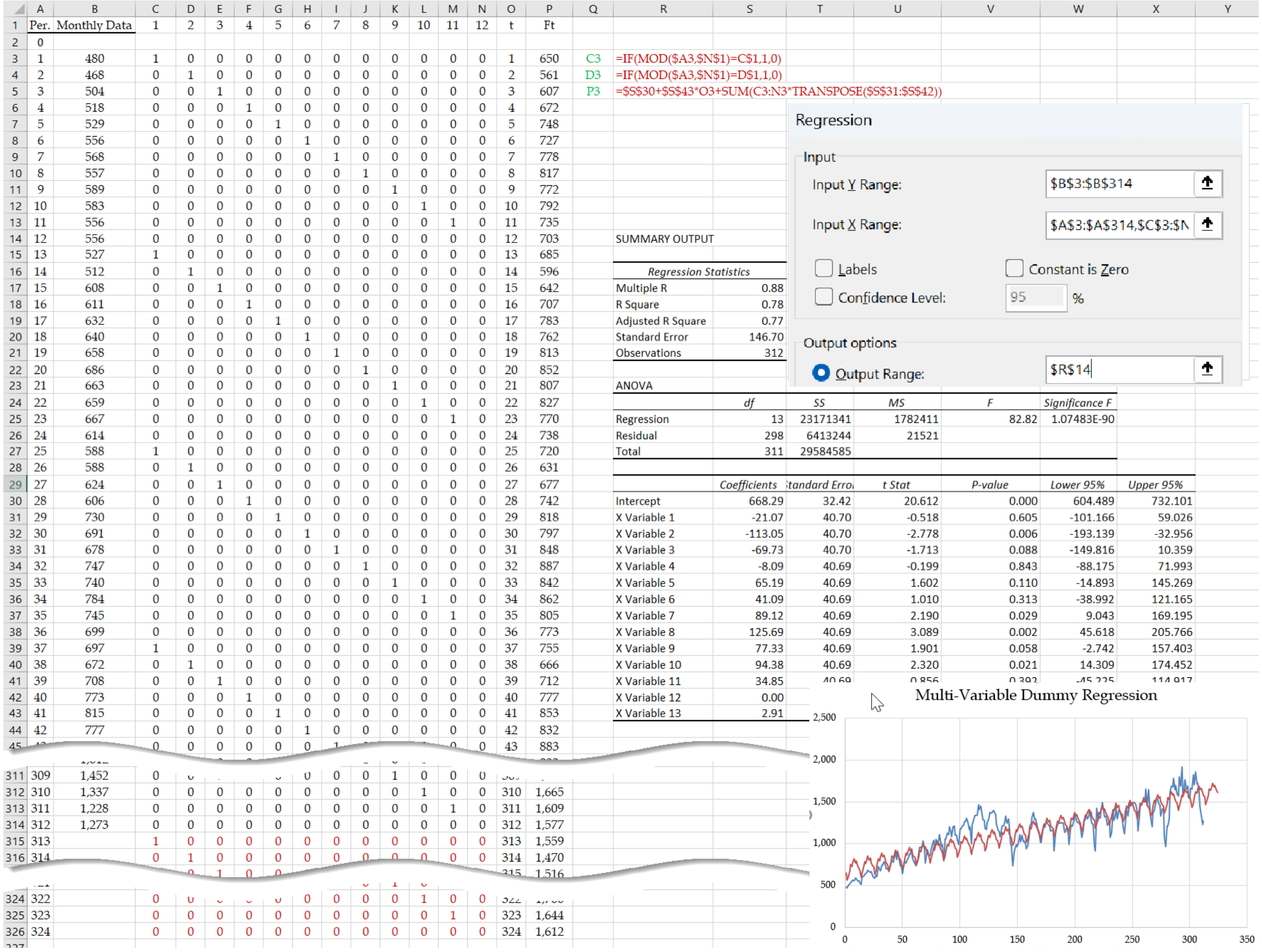
Figure 3. Four Key Steps in Static Seasonality Enhanced Bi-Variable Linear Regression.



6.2. Seasonality Enhanced Multiple Regression Using Dummy Variables.

By implementing a set of binary dummy variables, we use multi-variable regression for another version of static seasonality analysis. For each month, we define a binary variable, which is 1 if we are in that month and 0 otherwise. For periodicity of n periods, we need n-1 dummy binary variables. We compare other periods with a period of choice, where our choice does not affect the analysis outcomes. Since we analyze monthly data over the years, periodicity is 12. We define 11 binary variables for January to November. We will have our Y variable as the volume of activity in the corresponding month, our X variable as the month counter (from 1 to 312), and 11 dummy binary variables. Excel's Data Analysis Add-ins require the independent variables to be in contiguous cells. We, therefore, copy the month variables adjacent to the dummy variables. We can have them before or after the dummy variables. Compared to bi-variable regression, instead of a single column for X variables, we select 12 columns. The output and all the essential formulas are shown in Table 10. The reader may pay attention to the formula to generate 0s and 1 for the dummy variables in each month and, more importantly, to multiply the row of the decision variables by the column of regression coefficients (by using dynamic arrays and transposing one of the two vectors).

Table 10. Seasonality Enhanced Multi-Variable Regression Computations.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Per.	Monthly Data	1	2	3	4	5	6	7	8	9	10	11	12	t	Ft
2	0															
3	1	480	1	0	0	0	0	0	0	0	0	0	0	0	1	650
4	2	468	0	1	0	0	0	0	0	0	0	0	0	0	2	561
5	3	504	0	0	1	0	0	0	0	0	0	0	0	0	3	607
6	4	518	0	0	0	1	0	0	0	0	0	0	0	0	4	672
7	5	529	0	0	0	0	1	0	0	0	0	0	0	0	5	748
8	6	556	0	0	0	0	0	1	0	0	0	0	0	0	6	727
9	7	568	0	0	0	0	0	0	1	0	0	0	0	0	7	778
10	8	557	0	0	0	0	0	0	0	1	0	0	0	0	8	817
11	9	589	0	0	0	0	0	0	0	0	1	0	0	0	9	772
12	10	583	0	0	0	0	0	0	0	0	0	1	0	0	10	792
13	11	556	0	0	0	0	0	0	0	0	0	0	1	0	11	735
14	12	556	0	0	0	0	0	0	0	0	0	0	0	0	12	703
15	13	527	1	0	0	0	0	0	0	0	0	0	0	0	13	685
16	14	512	0	1	0	0	0	0	0	0	0	0	0	0	14	596
17	15	608	0	0	1	0	0	0	0	0	0	0	0	0	15	642
18	16	611	0	0	0	1	0	0	0	0	0	0	0	0	16	707
19	17	632	0	0	0	0	1	0	0	0	0	0	0	0	17	783
20	18	640	0	0	0	0	0	1	0	0	0	0	0	0	18	762
21	19	658	0	0	0	0	0	0	1	0	0	0	0	0	19	813
22	20	686	0	0	0	0	0	0	0	1	0	0	0	0	20	852
23	21	663	0	0	0	0	0	0	0	0	1	0	0	0	21	807
24	22	659	0	0	0	0	0	0	0	0	0	1	0	0	22	827
25	23	667	0	0	0	0	0	0	0	0	0	0	1	0	23	770
26	24	614	0	0	0	0	0	0	0	0	0	0	0	0	24	738
27	25	588	1	0	0	0	0	0	0	0	0	0	0	0	25	720
28	26	588	0	1	0	0	0	0	0	0	0	0	0	0	26	631
29	27	624	0	0	1	0	0	0	0	0	0	0	0	0	27	677
30	28	606	0	0	0	1	0	0	0	0	0	0	0	0	28	742
31	29	730	0	0	0	0	1	0	0	0	0	0	0	0	29	818
32	30	691	0	0	0	0	0	1	0	0	0	0	0	0	30	797
33	31	678	0	0	0	0	0	0	1	0	0	0	0	0	31	848
34	32	747	0	0	0	0	0	0	0	1	0	0	0	0	32	887
35	33	740	0	0	0	0	0	0	0	0	1	0	0	0	33	842
36	34	784	0	0	0	0	0	0	0	0	0	1	0	0	34	862
37	35	745	0	0	0	0	0	0	0	0	0	0	1	0	35	805
38	36	699	0	0	0	0	0	0	0	0	0	0	0	0	36	773
39	37	697	1	0	0	0	0	0	0	0	0	0	0	0	37	755
40	38	672	0	1	0	0	0	0	0	0	0	0	0	0	38	666
41	39	708	0	0	1	0	0	0	0	0	0	0	0	0	39	712
42	40	773	0	0	0	1	0	0	0	0	0	0	0	0	40	777
43	41	815	0	0	0	0	1	0	0	0	0	0	0	0	41	853
44	42	777	0	0	0	0	0	1	0	0	0	0	0	0	42	832
45			0	0	0	0	0	0						0	43	883
311	309	1,452	0					0	0	0	1	0	0	0		
312	310	1,337	0	0	0	0	0	0	0	0	0	1	0	0	310	1,665
313	311	1,228	0	0	0	0	0	0	0	0	0	0	1	0	311	1,609
314	312	1,273	0	0	0	0	0	0	0	0	0	0	0	0	312	1,577
315	313		1	0	0	0	0	0	0	0	0	0	0	0	313	1,559
316	314		0	1	0	0	0	0	0	0	0	0	0	0	314	1,470
324	322		0					0	0	0	0	1	0	0		
325	323		0	0	0	0	0	0	0	0	0	0	1	0	323	1,644
326	324		0	0	0	0	0	0	0	0	0	0	0	0	324	1,612

C3 =IF(MOD($A3,$N$1)=C$1,1,0)
D3 =IF(MOD($A3,$N$1)=D$1,1,0)
P3 =$S$30+$S$43*O3+SUM(C3:N3*TRANSPOSE($S$31:$S$42))

Regression

Input

Input Y Range: $B$3:$B$314

Input X Range: $A$3:$A$314,$C$3:$N

☐ Labels ☐ Constant is Zero

☐ Confidence Level: 95 %

Output options

● Output Range: $R$14

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.88
R Square	0.78
Adjusted R Square	0.77
Standard Error	146.70
Observations	312

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	23171341	1782411	82.82	1.07483E-90
Residual	298	6413244	21521		
Total	311	29584585			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	668.29	32.42	20.612	0.000	604.489	732.101
X Variable 1	-21.07	40.70	-0.518	0.605	-101.166	59.026
X Variable 2	-113.05	40.70	-2.778	0.006	-193.139	-32.956
X Variable 3	-69.73	40.70	-1.713	0.088	-149.816	10.359
X Variable 4	-8.09	40.69	-0.199	0.843	-88.175	71.993
X Variable 5	65.19	40.69	1.602	0.110	-14.893	145.269
X Variable 6	41.09	40.69	1.010	0.313	-38.992	121.165
X Variable 7	89.12	40.69	2.190	0.029	9.043	169.195
X Variable 8	125.69	40.69	3.089	0.002	45.618	205.766
X Variable 9	77.33	40.69	1.901	0.058	-2.742	157.403
X Variable 10	94.38	40.69	2.320	0.021	14.309	174.452
X Variable 11	34.85	40.69	0.856	0.392	-45.225	114.917
X Variable 12	0.00					
X Variable 13	2.91					

6.3. Trend and Seasonality Adjusted Exponential Smoothing.

In the two previous approaches, we used the term static seasonality. Static means we estimate seasonality for all months and keep them as they are; seasonality indexes and all other model parameters remain unchanged. In this third approach, we update seasonality indices (as well as level and trend) as we move forward. It extends the trend-adjusted exponential smoothing (Winter, 1960, Chopra, 2021). The reader may look into the graphs of the three approaches' output to visualize the dynamism inside this third approach.

By applying linear regression on the 12-month centered moving average implemented in seasonality-enhanced bi-variable linear regression, we first estimate the level (L_0 =INTERCEPT) and trend (T_0=SLOPE) in month zero. We also use static seasonality indexes computed in seasonality-enhanced bi-variable linear regression (Chopra 2021). Alternatively, we may set L_0 equal to the average demand in the first 12 months. Given L_N as the average of the last 12 months, we set $T_0=(L_N-L_0)/(12(N-1))$. For seasonality, we may divide the demand of each of the first 12 months by the average of these months and assume them as the seasonality indexes for the first 12 months (Iravani 2021). While the second approach is easier with fewer computations to estimate the starting parameters, since we already have the results as described in the seasonality-enhanced bi-variable linear regression section, we follow the first approach and copy L_0, T_0, S_1, ….S_{12} from Table 9 into Table 11. We first set $\alpha=0.5$, $\beta=0.5$, and $\gamma=0.5$.

Step 1. Compute L_t. Given L_0 = 702.82, T_0= 2.9, and S_1= 0.95; $F_1 =(L_0 + T_0)S_1 = (702.82+ 2.9)* 0.95 = 670.74$. We now move forward to compute L_1, T_1, F_2, and S_{13}, then L_2, T_2, F_3, and S_{14}, and so on. In all exponential smoothing models, we always have one component multiplied by a parameter (α, β, or γ), added to another component multiplied by 1 minus α, β, or γ. The 1 minus part is always easier to compute. We have L_0 = 702.82, T_0= 2.9. Our forecast for level in month 1 is $L_1= L_0 + T_0= 705.71$. This needs to be multiplied by $(1-\alpha)$. That is, $L_1=(1-0.5)* 705.71$. But what is the part that had to be multiplied by α? It is not 480. That is why the computation of the component multiplied by 1 minus α, β, or γ is easier. The actual month 1 data of 480 contains seasonality. We need to remove seasonality. Since S_1= 0.95, month 1 is a low season. We must divide the actual data by S_1 to remove seasonality; 480/0.95 =504.97. Therefore, $L_1 =(1-\alpha)(L_0+T_0)+ \alpha(A_1/S_1) = (1-0.5)(702.82+2.9)+ \alpha(480/0.95) = 605.34$.

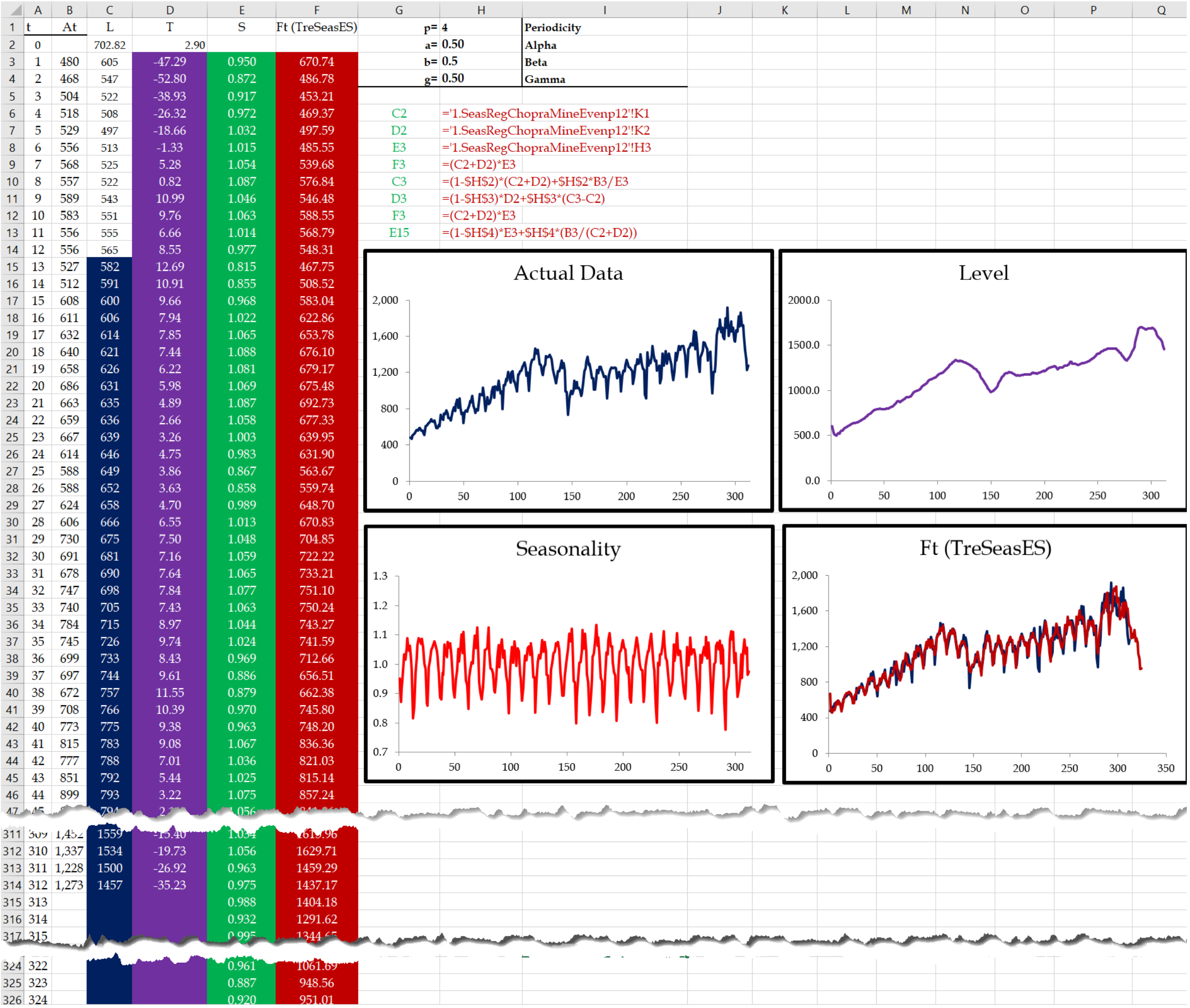
Step 2. Compute T_t. Our forecast for T_1 is T_0. It is multiplied by $(1-\beta)$ to form the first component of T_1. What is the actual T_1? It is the difference between L_0 and L_1 and should be multiplied by β. Therefore $T_1=(1-\beta)T_0+\beta(L_1- L_0) = (1-0.5)* 2.90+0.5(605.34-702.82)=-47.29$.

Step 3. Compute F_{t+1}. The forecast for the next period is simply $F_{t+1}=(L_t+T_t)*S_{t+1}$. For month 2, it is $F_2=(L_1+T_1)*S_2 = (605.34-47.29)0.872 = 486.78$.

Step 4. Compute S_{t+p}. Since periodicity is 12 (p=12), we need to compute S_{1+12}. We first have $(1-\gamma)$ times forecast. Our forecast for period 13 is the same as period 1; S_1=0.96. What is the actual seasonality in period 1? It is the actual data divided by $L_1= L_0 +T_0$. That is $A_1/L_1 = 480/705.71 =0.68$. Therefore, $S_{13}=(1-\gamma)*S_1+\gamma(A_1/L_1) = (1-0.5)(0.96)+0.5(0.68) = 0.82$.

Table 11 shows all the key formulas and curves related to trend and seasonality-adjusted exponential smoothing components.

Table 11. Seasonality Enhanced Multi-Variable Regression Computations.



	A	B	C	D	E	F
1	t	At	L	T	S	Ft (TreSeasES)
2	0		702.82	2.90		
3	1	480	605	-47.29	0.950	670.74
4	2	468	547	-52.80	0.872	486.78
5	3	504	522	-38.93	0.917	453.21
6	4	518	508	-26.32	0.972	469.37
7	5	529	497	-18.66	1.032	497.59
8	6	556	513	-1.33	1.015	485.55
9	7	568	525	5.28	1.054	539.68
10	8	557	522	0.82	1.087	576.84
11	9	589	543	10.99	1.046	546.48
12	10	583	551	9.76	1.063	588.55
13	11	556	555	6.66	1.014	568.79
14	12	556	565	8.55	0.977	548.31
15	13	527	582	12.69	0.815	467.75
16	14	512	591	10.91	0.855	508.52
17	15	608	600	9.66	0.968	583.04
18	16	611	606	7.94	1.022	622.86
19	17	632	614	7.85	1.065	653.78
20	18	640	621	7.44	1.088	676.10
21	19	658	626	6.22	1.081	679.17
22	20	686	631	5.98	1.069	675.48
23	21	663	635	4.89	1.087	692.73
24	22	659	636	2.66	1.058	677.33
25	23	667	639	3.26	1.003	639.95
26	24	614	646	4.75	0.983	631.90
27	25	588	649	3.86	0.867	563.67
28	26	588	652	3.63	0.858	559.74
29	27	624	658	4.70	0.989	648.70
30	28	606	666	6.55	1.013	670.83
31	29	730	675	7.50	1.048	704.85
32	30	691	681	7.16	1.059	722.22
33	31	678	690	7.64	1.065	733.21
34	32	747	698	7.84	1.077	751.10
35	33	740	705	7.43	1.063	750.24
36	34	784	715	8.97	1.044	743.27
37	35	745	726	9.74	1.024	741.59
38	36	699	733	8.43	0.969	712.66
39	37	697	744	9.61	0.886	656.51
40	38	672	757	11.55	0.879	662.38
41	39	708	766	10.39	0.970	745.80
42	40	773	775	9.38	0.963	748.20
43	41	815	783	9.08	1.067	836.36
44	42	777	788	7.01	1.036	821.03
45	43	851	792	5.44	1.025	815.14
46	44	899	793	3.22	1.075	857.24
47	45	[illegible]	794	[illegible]	1.056	[illegible]
311	309	1,452	1559	-15.40	1.034	[illegible]
312	310	1,337	1534	-19.73	1.056	1629.71
313	311	1,228	1500	-26.92	0.963	1459.29
314	312	1,273	1457	-35.23	0.975	1437.17
315	313				0.988	1404.18
316	314				0.932	1291.62
317	315				0.995	[illegible]
324	322				0.961	1061.69
325	323				0.887	948.56
326	324				0.920	951.01

	G	H	I
1	p=	4	Periodicity
2	a=	0.50	Alpha
3	b=	0.5	Beta
4	g=	0.50	Gamma
6	C2	='1.SeasRegChopraMineEvenp12'!K1	
7	D2	='1.SeasRegChopraMineEvenp12'!K2	
8	E3	='1.SeasRegChopraMineEvenp12'!H3	
9	F3	=(C2+D2)*E3	
10	C3	=(1-$H$2)*(C2+D2)+$H$2*B3/E3	
11	D3	=(1-$H$3)*D2+$H$3*(C3-C2)	
12	F3	=(C2+D2)*E3	
13	E15	=(1-$H$4)*E3+$H$4*(B3/(C2+D2))	

7. CONCLUSIONS.

We reviewed and integrated several time series and regression analysis techniques. This manuscript can be used as teaching material or as a case study to enforce the teaching material. While we had our analysis on total loaded and empty for both inbound and outbound throughput, all the data are available to repeat the combination for four combinations of inbound, outbound, loaded, and empty volumes. The same is true for applying these procedures – and the Excel templates – on other data sets.

Appendix A. Computation of Metrics and Drawing the Graphs for an Any-Period Moving Average.

We may develop a general formula applicable to any number of periods in a moving average computation. Consider a 4-period moving average forecast in periods 25 and 26 and examine the differences.

$F_{26} = MA_{25} = (A_{25}+A_{24}+A_{23}+ A_{22})/4 = (A_{25}+A_{24}+ A_{23})/4 + A_{22}/4$

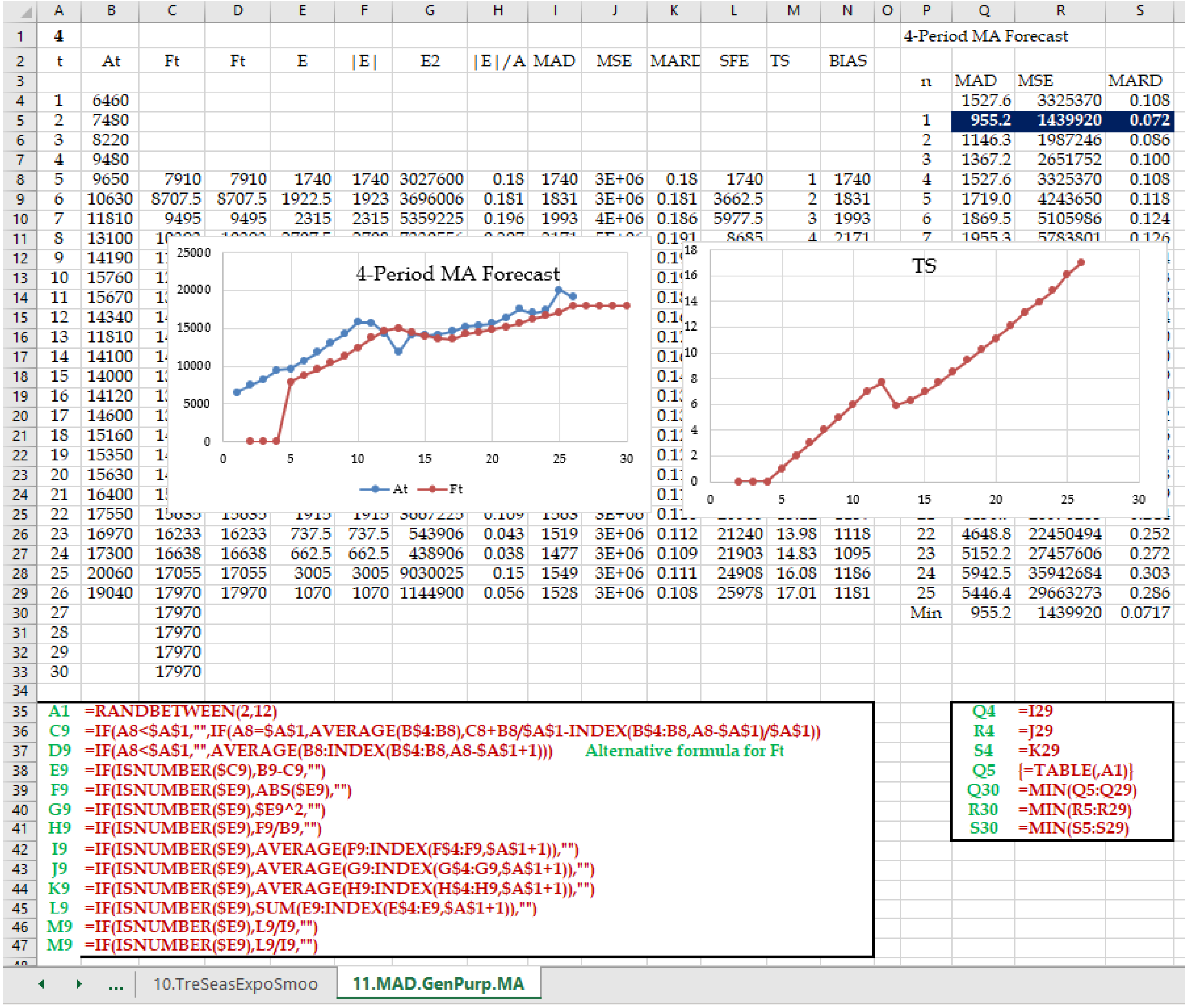
$F_{27} = MA_{26} = (A_{26}+A_{25}+A_{24}+ A_{23})/4 = A_{26}/4 + (A_{25}+ A_{24} +A_{23})/4$.

Therefore, $F_{27} = F_{26} + A_{26}/4 - A_{22}/4$. Generally, $F_{(t+1)} = F_t + (A_t - A_{t-n})/n$. Our forecast for the next period is equal to the forecast for this period (the moving average of the previous period) plus this period's actual data minus the oldest piece of data used on the forecast for the previous period divided by n.

Suppose we enter the number of periods in the moving average is in cell A1 and set it to any number between 2 and 25. Suppose we set it =RANDBETWEEN(2,12); the result is 4 when entered. We now look into the formula in period 6 in row 9 in Table A1. We have the previous forecast and previous actual in row 8, but what is the oldest data in the previous forecast? It is in the row t-n of the actual data. In our example is the data in row 8-4=4 of the Excel sheet. We can use the Excel INDEX function to find the element in a specific row of a vector. IF(A8<$A$1,"",IF(A8=$A$1,AVERAGE(B$4:B8),C8+B8/$A$1-INDEX(B$4:B8,A8-$A$1)/$A$1)) is the forecast formula in cell C9. If the previous year is before year 4, a " " is entered to leave the Excel cell blank. If the previous year is year 4, the average of the actual data for the first four years (from row 4 to 8) is computed and set to the forecast for year 5 (in row 8 of the Excel sheet). For cell C9 which corresponds to year 6>4, we have C8+B8/$A$1-INDEX(B$4:B8,A8-$A$1)/$A$1. Where INDEX(B$4:B8, A8-$A$1) will find the oldest piece of data used in the forecast; INDEX(B$4:B8,5-4) = INDEX(B$4:B8,1) = B4 = 6460. The actual for the previous period is B8 = 9650, and the forecast for the previous period is C8 = 7910. Therefore, the forecast for this period C9= 7910+(9650-6460)/4 =8707.5. The table is adjusted for any number less than 26 that may appear in cell A1.

Since we draw the curves related to some of the columns in Table A1, a " " for the starting years that are less than or equal to the random year that appears in cell A1 will show a Y-value of zero while it is empty and not zero. To resolve this, we replace " " with NA(). To avoid #NA appearing in the table, we use formula-based conditional formatting and switch the font color to white using the IFERROR function for #NA cells. Accordingly, Table A1 and Figure A1 are adjusted automatically no matter what random numbers between 2 and 25 appear in cell A1. Alternatively, we could have the fonts of these columns colored white and switch the font color to black using the ISNUMBER function in conditional formatting.

Table A1. Computation and Evaluation of an Any-Period Moving Average.



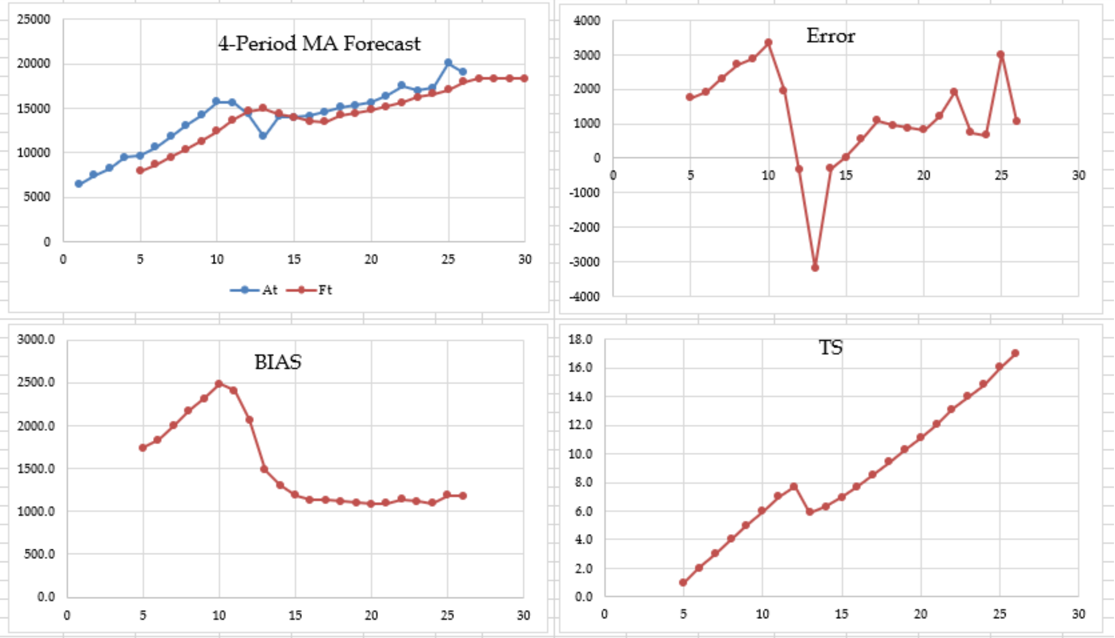
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	4															4-Period MA Forecast			
2	t	At	Ft	Ft	E	|E|	E2	|E|/A	MAD	MSE	MARD	SFE	TS	BIAS					
3																n	MAD	MSE	MARD
4	1	6460															1527.6	3325370	0.108
5	2	7480														1	955.2	1439920	0.072
6	3	8220														2	1146.3	1987246	0.086
7	4	9480														3	1367.2	2651752	0.100
8	5	9650	7910	7910	1740	1740	3027600	0.18	1740	3E+06	0.18	1740	1	1740		4	1527.6	3325370	0.108
9	6	10630	8707.5	8707.5	1922.5	1923	3696006	0.181	1831	3E+06	0.181	3662.5	2	1831		5	1719.0	4243650	0.118
10	7	11810	9495	9495	2315	2315	5359225	0.196	1993	4E+06	0.186	5977.5	3	1993		6	1869.5	5105986	0.124
11	8	13100	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.191	8685	4	2171		7	1955.3	5783801	0.126
12	9	14190																	
13	10	15760																	
14	11	15670																	
15	12	14340																	
16	13	11810																	
17	14	14100																	
18	15	14000																	
19	16	14120																	
20	17	14600																	
21	18	15160																	
22	19	15350																	
23	20	15630																	
24	21	16400																	
25	22	17550	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]
26	23	16970	16233	16233	737.5	737.5	543906	0.043	1519	3E+06	0.112	21240	13.98	1118		22	4648.8	22450494	0.252
27	24	17300	16638	16638	662.5	662.5	438906	0.038	1477	3E+06	0.109	21903	14.83	1095		23	5152.2	27457606	0.272
28	25	20060	17055	17055	3005	3005	9030025	0.15	1549	3E+06	0.111	24908	16.08	1186		24	5942.5	35942684	0.303
29	26	19040	17970	17970	1070	1070	1144900	0.056	1528	3E+06	0.108	25978	17.01	1181		25	5446.4	29663273	0.286
30	27		17970													Min	955.2	1439920	0.0717
31	28		17970																
32	29		17970																
33	30		17970																

```
A1   =RANDBETWEEN(2,12)
C9   =IF(A8<$A$1,"",IF(A8=$A$1,AVERAGE(B$4:B8),C8+B8/$A$1-INDEX(B$4:B8,A8-$A$1)/$A$1))
D9   =IF(A8<$A$1,"",AVERAGE(B8:INDEX(B$4:B8,A8-$A$1+1)))       Alternative formula for Ft
E9   =IF(ISNUMBER($C9),B9-C9,"")
F9   =IF(ISNUMBER($E9),ABS($E9),"")
G9   =IF(ISNUMBER($E9),$E9^2,"")
H9   =IF(ISNUMBER($E9),F9/B9,"")
I9   =IF(ISNUMBER($E9),AVERAGE(F9:INDEX(F$4:F9,$A$1+1)),"")
J9   =IF(ISNUMBER($E9),AVERAGE(G9:INDEX(G$4:G9,$A$1+1)),"")
K9   =IF(ISNUMBER($E9),AVERAGE(H9:INDEX(H$4:H9,$A$1+1)),"")
L9   =IF(ISNUMBER($E9),SUM(E9:INDEX(E$4:E9,$A$1+1)),"")
M9   =IF(ISNUMBER($E9),L9/I9,"")
M9   =IF(ISNUMBER($E9),L9/I9,"")
```

```
Q4    =I29
R4    =J29
S4    =K29
Q5    {=TABLE(,A1)}
Q30   =MIN(Q5:Q29)
R30   =MIN(R5:R29)
S30   =MIN(S5:S29)
```

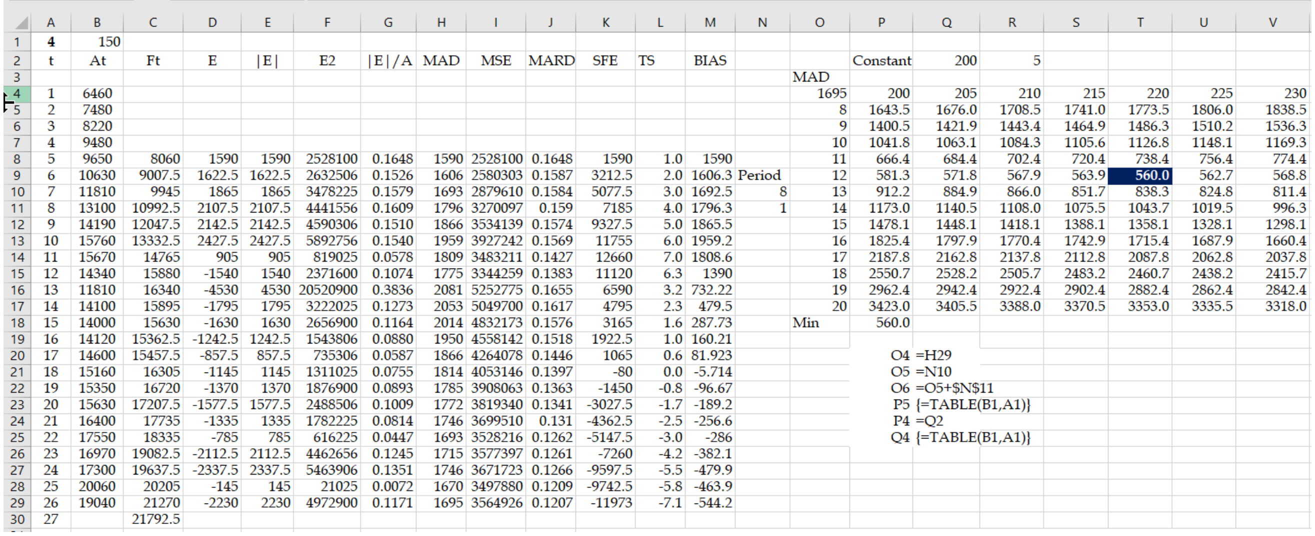
Column D provides an alternative formula for an any-period moving average as follows D9=IF(A6>=$A$1, AVERAGE(B6:INDEX($B$4:$B$30,A6-$A$1+1)),NA()). That is due to the magic inside the AVERAGE(B6:INDEX($B$4:$B$30,A6-$A$1+1) formula. We benefit from this formula in columns E to M to compute the metrics only when the data exist and do not show anything for other years in the graphs, as shown in Figure A1. All the key formulas of Table A1 are re-emphasized by the green and red cells with white backgrounds.

Figure A1. Flexible Graphs for Computation and Evaluation of an Any-Period Moving Average.



Given the new any-period moving average formulas, selecting the number of periods and adding a constant may reduce the gap between the actual data and forecasts. Table A2 shows computations for adding a constant and the number of periods in the moving average using a two-dimensional Data Table for MAD minimization (n= 12, K=220). It can be similarly applied to MSE and MARD minimization.

Table A2. Two-Dimensional Data Table to Add a Constant to an Any-Period Moving Average.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	**4**	150																				
2	t	At	Ft	E	|E|	E2	|E|/A	MAD	MSE	MARD	SFE	TS	BIAS			Constant	200	5				
3															MAD							
4	1	6460													1695	200	205	210	215	220	225	230
5	2	7480													8	1643.5	1676.0	1708.5	1741.0	1773.5	1806.0	1838.5
6	3	8220													9	1400.5	1421.9	1443.4	1464.9	1486.3	1510.2	1536.3
7	4	9480													10	1041.8	1063.1	1084.3	1105.6	1126.8	1148.1	1169.3
8	5	9650	8060	1590	1590	2528100	0.1648	1590	2528100	0.1648	1590	1.0	1590		11	666.4	684.4	702.4	720.4	738.4	756.4	774.4
9	6	10630	9007.5	1622.5	1622.5	2632506	0.1526	1606	2580303	0.1587	3212.5	2.0	1606.3	Period	12	581.3	571.8	567.9	563.9	560.0	562.7	568.8
10	7	11810	9945	1865	1865	3478225	0.1579	1693	2879610	0.1584	5077.5	3.0	1692.5	8	13	912.2	884.9	866.0	851.7	838.3	824.8	811.4
11	8	13100	10992.5	2107.5	2107.5	4441556	0.1609	1796	3270097	0.159	7185	4.0	1796.3	1	14	1173.0	1140.5	1108.0	1075.5	1043.7	1019.5	996.3
12	9	14190	12047.5	2142.5	2142.5	4590306	0.1510	1866	3534139	0.1574	9327.5	5.0	1865.5		15	1478.1	1448.1	1418.1	1388.1	1358.1	1328.1	1298.1
13	10	15760	13332.5	2427.5	2427.5	5892756	0.1540	1959	3927242	0.1569	11755	6.0	1959.2		16	1825.4	1797.9	1770.4	1742.9	1715.4	1687.9	1660.4
14	11	15670	14765	905	905	819025	0.0578	1809	3483211	0.1427	12660	7.0	1808.6		17	2187.8	2162.8	2137.8	2112.8	2087.8	2062.8	2037.8
15	12	14340	15880	-1540	1540	2371600	0.1074	1775	3344259	0.1383	11120	6.3	1390		18	2550.7	2528.2	2505.7	2483.2	2460.7	2438.2	2415.7
16	13	11810	16340	-4530	4530	20520900	0.3836	2081	5252775	0.1655	6590	3.2	732.22		19	2962.4	2942.4	2922.4	2902.4	2882.4	2862.4	2842.4
17	14	14100	15895	-1795	1795	3222025	0.1273	2053	5049700	0.1617	4795	2.3	479.5		20	3423.0	3405.5	3388.0	3370.5	3353.0	3335.5	3318.0
18	15	14000	15630	-1630	1630	2656900	0.1164	2014	4832173	0.1576	3165	1.6	287.73		Min	560.0						
19	16	14120	15362.5	-1242.5	1242.5	1543806	0.0880	1950	4558142	0.1518	1922.5	1.0	160.21									
20	17	14600	15457.5	-857.5	857.5	735306	0.0587	1866	4264078	0.1446	1065	0.6	81.923			O4	=H29					
21	18	15160	16305	-1145	1145	1311025	0.0755	1814	4053146	0.1397	-80	0.0	-5.714			O5	=N10					
22	19	15350	16720	-1370	1370	1876900	0.0893	1785	3908063	0.1363	-1450	-0.8	-96.67			O6	=O5+$N$11					
23	20	15630	17207.5	-1577.5	1577.5	2488506	0.1009	1772	3819340	0.1341	-3027.5	-1.7	-189.2			P5	{=TABLE(B1,A1)}					
24	21	16400	17735	-1335	1335	1782225	0.0814	1746	3699510	0.131	-4362.5	-2.5	-256.6			P4	=Q2					
25	22	17550	18335	-785	785	616225	0.0447	1693	3528216	0.1262	-5147.5	-3.0	-286			Q4	{=TABLE(B1,A1)}					
26	23	16970	19082.5	-2112.5	2112.5	4462656	0.1245	1715	3577397	0.1261	-7260	-4.2	-382.1									
27	24	17300	19637.5	-2337.5	2337.5	5463906	0.1351	1746	3671723	0.1266	-9597.5	-5.5	-479.9									
28	25	20060	20205	-145	145	21025	0.0072	1670	3497880	0.1209	-9742.5	-5.8	-463.9									
29	26	19040	21270	-2230	2230	4972900	0.1171	1695	3564926	0.1207	-11973	-7.1	-544.2									
30	27		21792.5																			

Furthermore, we can add flexibility to our table and graphs by updating them as soon as new data is available. We first put column A in a dynamic form by using =IF(OR(ISNUMBER(B5),ISNUMBER(C5)),A4+1,NA()). If we have actual or forecast value in the corresponding row, the year number will appear otherwise NA(). We then add flexible graphs using the OFFSET function to define dynamic name ranges and then use dynamic name ranges as X and Y of scatter graphs. We define a dynamic range name for the actual data as =OFFSET(click on B4,,,count(B4:B53)). Note that we need to do this in the range name window (not copy and paste from elsewhere) to have =OFFSET('11c.MAD.GenPurp.MA-B'!$B$4,,,COUNT('11c.MAD.GenPurp.MA-B'!$B$4:$B$53)). Fortunately, we can copy this and paste it for other dynamic ranges in other columns. The only change needed is to replace $B$4, and only $B$4, by $A$4, $A$4, $D$4, $H$4, $M$4 in t, Ft, E, BIAS, and TS columns. There is no need to change COUNT('11c.MAD.GenPurp.MA-B'!$B$4:$B$53) since it counts the rows containing data and is the same for all columns. Note that we must also type in t, Ft, E, BIAS, and TS as the names of the corresponding name ranges referred to by OFFSET functions.

We then need to go to each Ft vs. At, E, BIAS, TS, and any other chart we may need and replace the static reference with a dynamic reference. For example, we have ='11c regarding Ft vs. At chart.MAD.GenPurp.MA-B!$A$4:$A$30, ='11c.MAD.GenPurp.MA-B!$B$4:$B$30, and ='11c.MAD.GenPurp.MA-B!$C$4:$C$30 for series X and two Y values. We replace them with ='11c.MAD.GenPurp.MA-B!t, ='11c.MAD.GenPurp.MA-B!At, and ='11c.MAD.GenPurp.MA-B!Ft, respectively. As new actual data is entered into the table, the table and all its graphs are updated dynamically.

Appendix B. Exponential Smoothing Basic Mathematics.

In this Appendix, we show that (*i*) exponential smoothing is a weighted moving average and (*ii*) the age of data is $1/\alpha$.

B.1. Exponential Smoothing a Weighted Moving Average.

The following analytical manipulations show that Exponential Smoothing is a Weighted Moving Average.

$F_1 = A_1$

$F_2 = (1-\alpha)F_1 + \alpha A_1 \Rightarrow F_2 = (1-\alpha)A_1 + \alpha A_1 \Rightarrow F_2 = A_1$

$F_3 = (1-\alpha)F_2 + \alpha A_2 \Rightarrow F_3 = (1-\alpha)A_1 + \alpha A_2$

$F_4 = (1-\alpha)F_3 + \alpha A_3 \Rightarrow F_4 = (1-\alpha)((1-\alpha)A_1 + \alpha A_2) + \alpha A_3 \Rightarrow F_4 = (1-\alpha)^2 A_1 + \alpha (1-\alpha)A_2 + \alpha A_3$

$F_5 = (1-\alpha)F_4 + \alpha A_4 \Rightarrow F_5 = (1-\alpha)^3 A_1 + \alpha ((1-\alpha)^2 A_2 + \alpha((1-\alpha)A_3 + \alpha A_4$

$F_{t+1} = \alpha A_t + \alpha(1 - \alpha)A_{t-1} + \alpha(1 - \alpha)^2 A_{t-2} + \alpha(1 - \alpha)^3 A_{t-3} + \alpha (1 - \alpha)^4 A_{t-4} \dots + \alpha (1 - \alpha)^{t-1} A_1$

The sum of the weights are

$S = \alpha + \alpha(1 - \alpha) + \alpha(1 - \alpha)^2 + \alpha(1 - \alpha)^3 + \alpha (1 - \alpha)^4 \dots + \alpha (1 - \alpha)^{t-1}$

$= \alpha(1 - \alpha) + \alpha(1 - \alpha)^2 + \alpha(1 - \alpha)^3 + \alpha (1 - \alpha)^4 \dots + \alpha (1 - \alpha)^t$

$S-(1-\alpha)S = \alpha - \alpha (1 - \alpha)^t \Rightarrow \alpha S = \alpha(1 - (1 - \alpha)^t) \Rightarrow S = 1 - (1 - \alpha)^t$

When t increases, $(1 - \alpha)^t$ goes to 0, and the sum of the weights S=1.

B.2. Age of Data in Exponential Smoothing.

Through the following analytical manipulations, we show that the age of Data in Exponential Smoothing is $1/\alpha$.

Weights = $\alpha, \alpha(1 - \alpha), \alpha(1 - \alpha)^2, \alpha (1 - \alpha)^3, \alpha(1 - \alpha)^4 ,\ldots\ldots, \alpha(1 - \alpha)^{t-1}$

Ages = 1, 2, 3, 4, ………..t

Weights × Ages = $1\alpha+ 2\alpha(1 - \alpha)+ 3\alpha(1 - \alpha)^2+ 4\alpha(1 - \alpha)^3+ 5\alpha(1 - \alpha)^4 +\ldots\ldots+ t\alpha(1 - \alpha)^{t-1}$

Weights × Ages = $\alpha(1+ 2(1 - \alpha)+ 3(1 - \alpha)^2+ 4(1 - \alpha)^3+ 5(1 - \alpha)^4 +\ldots\ldots+ t(1 - \alpha)^{t-1})$

We have shown $S = \alpha(1+ (1 - \alpha)+(1 - \alpha)^2+(1 - \alpha)^3+(1 - \alpha)^4 +\ldots\ldots+ (1 - \alpha)^{t-1}) =1$

$1+ (1 - \alpha)+(1 - \alpha)^2+(1 - \alpha)^3+(1 - \alpha)^4 +\ldots\ldots+ (1 - \alpha)^{t-1} =1/\alpha$

Derivation with respect to α

$0-1-2(1 - \alpha)^1-3(1 - \alpha)^2-4(1 - \alpha)^3 -\ldots\ldots-(t-1) (1 - \alpha)^{t-2} =-1/\alpha^2$

$\alpha(1+2(1 - \alpha)^1+3(1 - \alpha)^2+4(1 - \alpha)^3 +\ldots\ldots+(t-1) (1 - \alpha)^{t-2})=1/\alpha$

B.3. UCL and LCL in Tracking Signal are larger than $\underline{+}4$

Forecast error $E_t = A_t - F_t$ is a random variable with a mean of 0. MAD estimates the error forecast's standard deviation. StdDev(E_t) = 1.25MAD (for example, Duncan, 2007}.

E_t = Normal (0,1.25MAD)

If x = Normal(μ,σ) ➔ Sum (x) = Normal(μ, SQRT(N)σ)

StdDev [Sum(E_t)] = SQRT(N)StdDev (E_t)

E_t = Normal (0,1.25MAD)

Sum (E_t) = N~(0, SQRT(N)1.25MAD)

3 ≥ ($\sum E_t$ -0)/(SQRT(N)1.25MAD)) ≥ -3.

+ 3SQRT(N)1.25 ≥ ($\sum E_t$ -0)/MAD ≥ - 3SQRT(N)1.25.

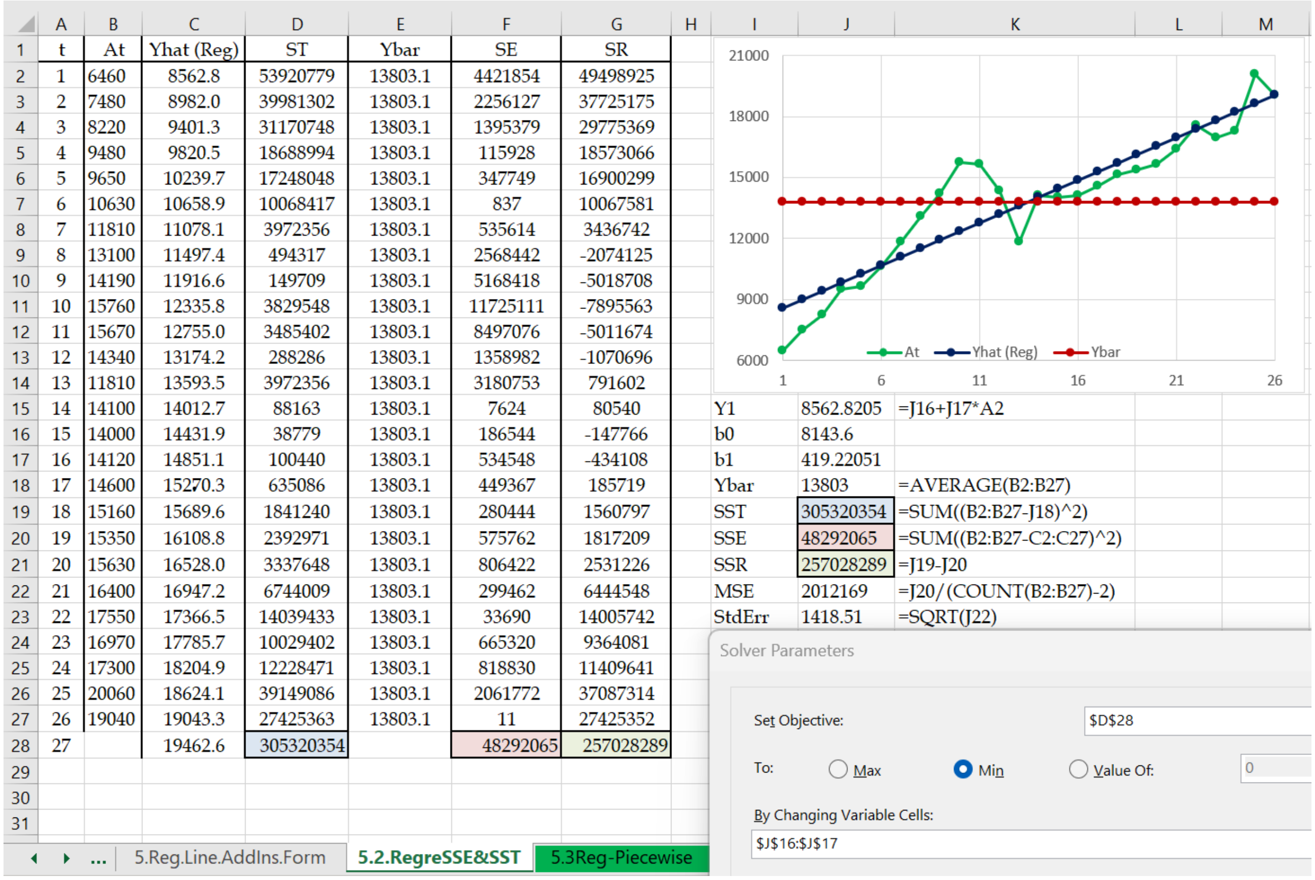
+ 3.75SQRT(N) ≥ ($\sum E_t$ -0)/MAD ≥ - 3.75 SQRT(N)

Therefore, Tracking Signal TS= $\sum E_t$/MAD with samples of size N is normally distributed around 0, and UCL = 3.75 SQRT(N) and LCL =-3.75 SQRT(N).

Appendix C. Foundations of Computation of Regression Metrix in Excel (b_0, B_1, SST, SSE, SSR).

One may design a regression line by minimizing MAD, MSE, MAX(ABS(Error)), or any other measure. Conventionally, regression equations are designed based on MSE minimization (least-square method). We compute MSE or SSE (Sum of Squared Errors) and use SOLVER to find the optimal b_0 and b_1 (which are in cells J16 and J17 in Table A1) to find the optimal values for the SSE (cell D28) objective function. After computing the forecasts in column C using arbitrary but reasonable b_0 and b_1 (in cells J16 and J17), we form column D (the square of the error in each row) and add them to form SSE in cell D26. We then use SOLVER (we can use DataTable too) to find optimal b_0 and b_1 to minimize SSE (or MSE). These optimal values (in cells J16 and J17) are the same as we found using the first three approaches in the regression section. It provides insight into least-squared computations and other regression metrics. Cell D26 can also be computed using dynamic arrays without referencing any values in column D (we can even delete column D). Look at the significant power of dynamic arrays in cell J19 for direct SSE computations.

Table A1. Direct Computation of Regression Coefficients and Key Metrics.



	A	B	C	D	E	F	G
1	t	At	Yhat (Reg)	ST	Ybar	SE	SR
2	1	6460	8562.8	53920779	13803.1	4421854	49498925
3	2	7480	8982.0	39981302	13803.1	2256127	37725175
4	3	8220	9401.3	31170748	13803.1	1395379	29775369
5	4	9480	9820.5	18688994	13803.1	115928	18573066
6	5	9650	10239.7	17248048	13803.1	347749	16900299
7	6	10630	10658.9	10068417	13803.1	837	10067581
8	7	11810	11078.1	3972356	13803.1	535614	3436742
9	8	13100	11497.4	494317	13803.1	2568442	-2074125
10	9	14190	11916.6	149709	13803.1	5168418	-5018708
11	10	15760	12335.8	3829548	13803.1	11725111	-7895563
12	11	15670	12755.0	3485402	13803.1	8497076	-5011674
13	12	14340	13174.2	288286	13803.1	1358982	-1070696
14	13	11810	13593.5	3972356	13803.1	3180753	791602
15	14	14100	14012.7	88163	13803.1	7624	80540
16	15	14000	14431.9	38779	13803.1	186544	-147766
17	16	14120	14851.1	100440	13803.1	534548	-434108
18	17	14600	15270.3	635086	13803.1	449367	185719
19	18	15160	15689.6	1841240	13803.1	280444	1560797
20	19	15350	16108.8	2392971	13803.1	575762	1817209
21	20	15630	16528.0	3337648	13803.1	806422	2531226
22	21	16400	16947.2	6744009	13803.1	299462	6444548
23	22	17550	17366.5	14039433	13803.1	33690	14005742
24	23	16970	17785.7	10029402	13803.1	665320	9364081
25	24	17300	18204.9	12228471	13803.1	818830	11409641
26	25	20060	18624.1	39149086	13803.1	2061772	37087314
27	26	19040	19043.3	27425363	13803.1	11	27425352
28	27		19462.6	305320354		48292065	257028289
29							
30							
31							

	I	J	K
15	Y1	8562.8205	=J16+J17*A2
16	b0	8143.6	
17	b1	419.22051	
18	Ybar	13803	=AVERAGE(B2:B27)
19	SST	305320354	=SUM((B2:B27-J18)^2)
20	SSE	48292065	=SUM((B2:B27-C2:C27)^2)
21	SSR	257028289	=J19-J20
22	MSE	2012169	=J20/(COUNT(B2:B27)-2)
23	StdErr	1418.51	=SQRT(J22)

Solver Parameters

Set Objective: $D$28

To: ○ Max ● Min ○ Value Of: 0

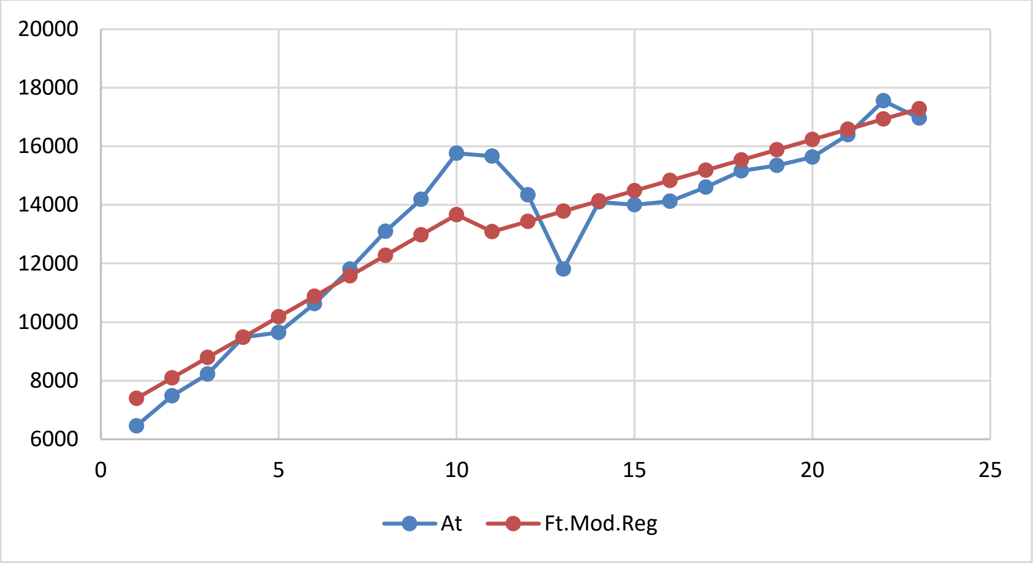
By Changing Variable Cells: $J$16:$J$17

In Regression Analysis, we usually compute three SST, SSE, and SSR metrics. SST is the summation of the squares of the gap between each piece of data with the average. Table C1 shows the gap between the green curve (actual data) and the red curve (average of all data). The total squared error measures how each data element differs from the average. We then have SSE, the squared gap between the green curve (actual data) and the blue curve (regression data). The total squared error measures how each data element differs from the value obtained on the regression line. The difference between these two (SSR) represents how well the regression line could replace the average line representing the data. The reader may compare the computations in cells D28, F28, and G28 with those of J19, J20, and J21 to better understand dynamic arrays (and may delete columns D, E, F, and G).

R-squared is computed as SSR/SST, reaching the same value as computed directly using the RSQ function. The MSE (and Standard Error) computations in regression slightly differ from what we discussed earlier. When you benefit from other statistics extracted from the same data set in the computation of an average, you use degrees of freedom. In the computation of SSE, we have used two parameters b_0 and b_1. Therefore, we lose two degrees of freedom when we average SSE over n years (26 in this example). Therefore, MSE is not SSE/26 but SSE/(26-2).

Given the background provided in this Appendix, we can apply a piecewise regression to find b_{01}, b_{11}, b_{02}, and b_{12} for the first and second piece of the regression line and T as the year to switch from the first regression line to the second. The above five items form the changing cells, and MSE is the objective function to be minimized. The result is shown in Figure A1.

Figure A1. Piecewise Regression on LA/LB ports Annual Data.



Appendix D. All Worksheets Used in This Study. Since there are many computations in different worksheets of this workbook, recalculating all elements on all pages slows down the process. It is recommended to move each worksheet to an individual workbook.

References

Cachon, G. and Terwiesch, C., 2020. Matching Supply and Demand: An Introduction to Operations Management. 4th Edition, McGraw-Hill. ISBN: 978-0-07-808665-5.

Chase, R. B., Aquilano, N. J., and Jacobs, R. F., 2000. Irwin/McGraw-Hill, ISBN 10: 0072395303ISBN 13: 9780072395303.

Chopra, S., 2019. Supply Chain Management, Pearson, ISBN 978-0-13-473188-9.

Duncan, R. M., 2007. CFPIM, Training Manager, E/Step Software Inc. http://www.estepsoftware.com/papers/madrsquare.pdf

Holt, C. C., 1957. *Forecasting seasonals and trends by exponentially weighted averages* (ONR Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA. https://www.sciencedirect.com/science/article/abs/pii/S0169207003001134?via%3Dihub Retrieved 7/12/2023.

Iravani, S.M.R., 2021. Operations Engineering and Management: Concepts, Analytics, and Principles for Improvement. McGraw-Hill. ISBN 978-1-260-46183-1.

Leachman, R.C., 2010. Port and Modal Elasticity Study, Final Report, Southern California Association of Governments.

Port of Los Angeles Statistics. https://www.portoflosangeles.org/ Last Checked 7/12/2023.

Port of Long Beach Statistics. https://polb.com/ Retrieved 7/12/2023.

Stevenson, W. J., 2014. Operations Management. 12th Edition, McGraw-Hill. ISBN: 978-0078024108.

The Institute for Advanced Analytics (North Carolina State University). Survey of Graduate Degree Programs in Analytics. http://analytics.ncsu.edu/?page _id=4184. Retrieved 7/12/2023.

Winters, P. R., 1960. Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324–342. https://pubsonline.informs.org/doi/10.1287/mnsc.6.3.324 Retrieved 7/12/2023.