Teaching time series and regression analysis using ports of Los Angeles and Long Beach data

Ardavan Asef-Vaziri

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

Abstract: The combined ports of Los Angeles and Long Beach (LA/LB) ports are among the world's top ten busiest container ports. Approximately 1/3 of US waterborne containers move through the LA/LB ports. The data on the volume of containerised activities in these ports provide an excellent dataset to teach time series and regression analysis. We use 26 years of data on these ports' activities to teach 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. This manuscript can be used as teaching material, or as a case study in a business analytics foundations or a supply chain management course. A set of useful Excel functions and formulas have been brought together and are fully embedded in the models.

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

Reference to this paper should be made as follows: Asef-Vaziri, A. (2024) 'Teaching time series and regression analysis using ports of Los Angeles and Long Beach data', *Int. J. Information and Operations Management Education*, Vol. 7, No. 4, pp.374–403.

Biographical notes: Ardavan Asef-Vaziri is a Professor of Systems and Operations Management in the David Nazarian College of Business and Economics, California State University, Northridge. He completed his BSc, MSc and PhD.all in Industrial Engineering. His research revolves around applications of optimisation and simulation in facility logistics and process flow analysis. He was among the first wave professors adopting the flipped classroom course delivery format. His publications have appeared in *EJOR*, *IISE Transactions*, *Computers & Operations Research*, *Journal of the Operational Research Society*, *International Journal of Production Economics*, *Decision Science Journal of Innovative Education*, among others.

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 organisation towards achieving the plans of the enterprises. Marketing, finance, and operations are the three key building blocks of manufacturing, service, and distribution systems. Planning, organising, 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, investment plans, operating costs, and revenues. They are all interested in long-term and short-term forecasts for strategic, tactical, and operational decisions.

Approximately 1/3 of US seaborne containers move through the LA/LB ports. The value of two-way trade in Southern California customs exceeds 10% of total US international trade in goods. Around 75% of this value passes through to LA/LB ports. Around 125,000 firms consider the LA/LB ports their export hub, and 175,000 firms consider these ports their import hubs. One out of 10 jobs in Southern California is associated with LA/LB ports.

The inbound and outbound volumes of the loaded and empty containers in the LA/LB ports provide an excellent data set to teach the basics of time series and regression analytics.

Teaching-focused business schools (TFBSs) make up close to 50% of all AACSB (Association to Advance Collegiate Schools of Business) accredited institutions. State-funded teaching-focused business schools (SFTFBSs) are a large subset of (TFBSs). Many of these lower-funded SFTFBSs educate a nontraditional and low-income mixture of first-generation high school or community college graduates. SFTFBS students are often self-supporting and work 20–60 hours per week. With less time dedicated to education, these students require more educational resources and streamlined learning processes than traditional university students.

By fully implementing time series and regression analysis in Excel, we provide a platform where students can learn the basic, intermediate, and advanced Excel functions and formulas. Excel is among the three fundamental skills (communications and time management) employers seek in SFTFBS graduates. We have tried to bring well-known time series and regression techniques under one roof, link them with well-thought-of Excel functions, formulas, and vitalisation tools, and combine them in well-integrated and easy-to-follow Excel sheets. Our spreadsheet models can also serve as templates for other real-life applications students may encounter in their early employment years.

Competitive emphasis on globalisation in today's education and developing case studies in international trade provide suitable teaching material in this direction. Articles of this kind facilitate continuing education and lifelong learning on information and operations management subjects. Manuscripts of this type may also constitute a bridge between port administrations looking for employees with good analytical skills and academic institutions training workforces to apply their skills in modern ports.

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.

We will have a short literature review in Section 2. Historical LA/LB ports container handling data are presented in Section 3. In Section 4, we estimate yearly ports throughput levels using moving averages and exponential smoothing. Measures of forecast accuracy and variability are discussed in Section 5. The level and trend for yearly data are covered in Section 6 using linear regression and trend-adjusted exponential smoothing. Section 7 estimates monthly data's level, trend, and seasonality using seasonality-enhanced regression, multivariate regression using seasonal dummy variables, and trend and seasonality-adjusted exponential smoothing. Conclusions follow in Section 8. In Appendix A, we implement Excel's functional and visualisation capabilities by examining a general any-period moving average and its dynamic tables and graphs. In Appendix B, we review the basic mathematics behind exponential smoothing. Appendix C explains the foundations of the computation of regression metrics in Excel and provides insight for piecewise regression analysis. The Excel workbooks can be obtained by contacting the author.

2 Literature review

Forecasting methods are partitioned into qualitative and quantitative techniques. Qualitative techniques are based on expert opinions and intuitions, such as subjective judgment, surveys, salesforce polling, historical analogies, and the Delphi method.

Time series and regression analysis are among the quantitative forecasting tools. They form one or more chapters in Operations and Supply Chain Management and business analytics foundations books.

For Operations Management and Supply Chain Management books, the reader is referred to Cachon and Terwiesch (2020), Chase et al. (2000), Heizer et al. (2023), Stevenson (2014), Venkataraman and Pino (2018), and especially Chopra (2019) and Iravani (2021). For Business Analytics Foundations books, the reader is referred to Hillier and Hillier (2023), Albright and Winston (2015), Powel and Baker (2017), Ragsdale (2018), Camm et al. (2020), Jaggia et al. (2023), and Krajewski et al. (2016), and Winston (2022). For time series and forecasting specifics, the reader is referred to Holt (1957) and Vandeput (2023).

To limit the length of this manuscript, we do not cover autoregressive models. An autoregressive model is a regression model where the forecasts are based on previous periods. The reader is referred to Chapter 3, Iravani (2021), for a simple introduction to an autoregressive model. The moving average in the autoregressive moving average model (ARMA) differs from the moving average we discuss in this manuscript. In ARMA's moving average model, the forecasts are based on deviations from past forecasts (a taste of exponential smoothing). This means a part of the forecast is based on the past observations (AR part), and another part is based on the deviations from the past observations (MA part). The AR part can be obtained using regression, while the MA part follows a stationary distribution. The integration part is the difference between

ARMA and ARIMA (autoregressive integrated moving average). It identifies the number of time intervals to make the time series stationary. For an in-depth review of ARMA and ARIMA and more advanced techniques, the reader is referred to Box et al. (2015), Brockwell and Davis (2016), and Keating and Wilson (2019).

3 Historical data in la/lb ports and forecasting characteristics

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

3.1 Historical data at LA/LB ports

Figure 1 presents parts of 312 monthly data records, including loaded inbound, loaded outbound, empty inbound, and empty outbound for the Los Angeles and Long Beach (LA/LB) ports (Port of Long Beach Statistics and Port of Los Angeles Statistics). The data is presented in TEUs (20 equivalent units), which are the size of a 20-foot container.

Figure 1 312 months TEUs handling in LA/LB ports

1				Port of Los Angeles								
2	Y	M	M	Loaded Inbound	Empty Inbound	Loaded Outbound	Empty Outbound	Total Inbound	Total Outbound	Total Loaded	Total Empty	Total TEUs
3	1997	1	Jan	115,349	11,716	72,537	36,438	127,065	108,975	187,886	48,154	236,040.15
4	1997	2	Feb	100,957	11,776	71,376	33,185	112,732	104,561	172,332	44,961	217,293.05
5	1997	3	Mar	102,389	10,673	83,446	33,654	113,062	117,100	185,835	44,327	230,162.10
6	1997	4	Apr	117,035	12,761	77,211	34,513	129,796	111,724	194,246	47,274	241,520.55
7	1997	5	May	123,100	13,611	76,327	38,165	136,711	114,492	199,428	51,776	251,203.15
8	1997	6	Jun	127,693	12,526	73,399	32,659	140,219	106,058	201,093	45,185	246,277.20
9	1997	7	Jul	130,481	13,274	71,623	37,187	143,755	108,809	202,103	50,461	252,564.40
28	1999	2	Feb	134,281	7,799	61,074	60,625	142,080	121,698	195,355	68,424	263,778.40
29	1999	3	Mar	133,887	8,039	68,857	64,907	141,925	133,764	202,744	72,945	275,689.25
30	1999	4	Apr	146,258	7,300	63,470	65,718	153,558	129,188	209,729	73,017	282,746.10
31	1999	5	May	175,535	7,439	70,211	82,775	182,975	152,986	245,746	90,215	335,960.70
32	1999	6	Jun	161,751	7,931	67,377	80,109	169,681	147,486	229,127	88,040	317,167.10
33	1999	7	Jul	176,445	8,572	66,948	78,281	185,016	145,229	243,393	86,852	330,245.00
34	1999	8	Aug	180,685	7,596	68,901	90,878	188,281	159,779	249,586	98,474	348,059.95
35	1999	9	Sep	180,796	6,285	66,207	83,648	187,081	149,855	247,003	89,933	336,935.70
36	1999	10	Oct	198,681	9,481	75,968	103,035	208,162	179,003	274,649	112,516	387,165.45
37	1999	11	Nov	174,939	9,723	77,653	85,502	184,662	163,155	252,592	95,225	347,816.25
63	2002	1	Jan	210,394	10,415	83,123	104,305	220,809	187,427	293,517	114,720	408,236.45
64	2002	2	Feb	234,405	9,762	92,401	98,440	244,167	190,841	326,806	108,202	435,007.70
65	2002	3	Mar	218,698	11,256	92,662	113,950	229,954	206,612	311,360	125,206	436,565.70
66	2002	4	Apr	258,026	9,346	91,114	125,812	267,373	216,927	349,141	135,159	484,299.25
67	2002	5	May	274,449	9,935	96,685	150,145	284,385	246,830	371,134	160,080	531,214.20
68	2002	6	Jun	291,966	9,481	95,285	139,729	301,447	235,015	387,251	149,210	536,461.50
69	2002	7	Jul	260,811	11,121	83,740	140,019	271,932	223,758	344,551	151,140	495,690.45
70	2002	8	Aug	327,472	12,217	91,421	174,383	339,689	265,805	418,893	186,601	605,493.80
71	2002	9	Sep	320,578	10,389	89,503	168,772	330,966	258,276	410,081	179,161	589,242.15
72	2002	10	Oct	252,910	12,308	70,208	135,545	265,217	205,753	323,118	147,853	470,970.60
73	2002	11	Nov	291,571	9,761	95,772	148,925	301,332	244,697	387,344	158,686	546,029.50
166	2010	8	Aug	399,151	7,541	147,609	209,537	406,691	357,146	546,759	217,078	763,837.20
167	2010	9	Sep	373,249	6,535	139,800	192,028	379,784	331,829	513,050	198,563	711,613.00
168	2010	10	Oct	349,545	8,695	151,049	173,095	358,241	324,144	500,594	181,790	682,384.50
169	2010	11	Nov	333,710	10,614	170,319	152,327	344,324	322,646	504,029	162,941	666,970.50
170	2010	12	Dec	299,304	13,556	161,625	138,166	312,861	299,791	460,929	151,722	612,651.70
171	2011	1	Jan	338,607	16,513	159,051	146,348	355,119	305,399	497,657	162,860	660,517.80
172	2011	2	Feb	275,887	13,277	150,357	115,393	289,163	265,750	426,244	128,669	554,912.80
173	2011	3	Mar	297,023	15,780	192,849	95,144	312,803	287,993	489,872	110,924	600,796.00
174	2011	4	Apr	312,360	14,486	167,448	122,978	326,846	290,426	479,808	137,464	617,272.50
175	2011	5	May	360,969	14,159	184,275	133,531	375,128	317,806	545,244	147,690	692,933.70
176	2011	6	Jun	333,894	15,529	163,137	128,234	349,423	291,371	497,032	143,763	640,794.60
177	2011	7	Jul	357,668	17,632	165,135	147,891	375,300	313,026	522,803	165,523	688,325.55
302	2021	12	Dec	385,251	2,272	70,872	328,194	387,523	399,066	456,122	330,466	786,588.75
303	2022	1	Jan	427,208	1,267	100,185	336,936	428,474	437,121	527,393	338,202	865,595.35
304	2022	2	Feb	424,073	2,740	95,441	335,511	426,812	430,952	519,514	338,251	857,764.40
305	2022	3	Mar	495,196	2,580	111,781	349,117	497,776	460,898	606,977	351,697	958,674.05
306	2022	4	Apr	456,670	1,595	99,878	329,215	458,264	429,093	556,548	330,810	887,357.35
307	2022	5	May	499,960	1,719	125,656	340,565	501,679	466,221	625,616	342,285	967,900.15
308	2022	6	Jun	444,680	3,587	93,890	334,454	448,267	428,344	538,570	338,041	876,611.30
309	2022	7	Jul	485,452	5,257	103,899	340,816	490,709	444,715	589,351	346,073	935,423.80
310	2022	8	Aug	403,602	2,065	102,319	297,329	405,667	399,648	505,921	299,394	805,314.70
311	2022	9	Sep	343,462	2,113	77,680	286,618	345,575	364,298	421,142	288,731	709,873.30
312	2022	10	Oct	336,307	4,357	89,722	248,044	340,664	337,766	426,029	252,401	678,429.45
313	2022	11	Nov	307,080	4,476	90,116	237,672	311,556	327,788	397,196	242,148	639,343.50
314	2022	12	Dec	352,046	7,826	96,518	227,481	359,872	368,999	448,564	280,307	728,871.50

1	Port of Long Beach								
2	Loaded Inbound	Empty Inbound	Loaded Outbound	Empty Outbound	Total Inbound	Total Outbound	Total Loaded	Total Empty	Total TEUs
3	123,456	8,349	78,965	33,135	131,805	112,100	202,421	41,484	243,905
4	126,804	8,452	87,586	27,432	135,256	115,018	214,390	35,884	250,274
5	125,040	11,819	106,556	30,493	136,859	137,049	231,596	42,312	273,908
6	146,110	9,216	92,893	28,615	155,326	121,508	239,003	37,831	276,834
7	143,330	8,927	91,734	34,026	152,257	125,760	235,064	42,953	278,017
8	165,763	8,524	100,262	35,080	174,287	135,342	266,025	43,604	309,629
9	169,964	5,866	96,244	42,926	175,830	139,170	266,208	48,792	315,000
28	172,482	3,905	80,000	67,743	176,387	147,743	252,482	71,648	324,130
29	171,913	4,659	88,699	82,704	176,572	171,403	260,612	87,363	347,975
30	181,575	3,584	77,040	60,622	185,159	137,662	258,615	64,206	322,821
31	205,640	4,958	87,553	96,317	210,598	183,870	293,193	101,275	394,468
32	194,410	7,196	78,962	93,408	201,606	172,370	273,372	100,604	373,976
33	188,586	7,013	74,155	78,207	195,599	152,362	262,741	85,220	347,961
34	214,709	3,304	80,581	100,711	218,013	181,292	295,290	104,015	399,305
35	219,080	2,890	82,308	98,433	221,970	180,741	301,388	101,323	402,711
36	204,242	6,483	84,349	101,345	210,725	185,694	288,591	107,828	396,419
37	209,444	5,508	91,201	91,150	214,952	182,351	300,645	96,658	397,303
63	180,528	5,896	70,944	87,276	186,424	158,220	251,472	93,172	344,644
64	202,276	4,303	83,333	73,070	206,579	156,403	285,609	77,373	362,982
65	171,874	5,525	87,845	78,362	177,399	166,207	259,719	83,887	343,606
66	227,953	4,373	87,960	89,914	232,326	177,874	315,913	94,287	410,200
67	231,382	4,971	84,062	107,783	236,353	191,845	315,444	112,754	428,198
68	236,112	3,043	75,814	104,009	239,155	179,823	311,926	107,052	418,978
69	249,618	4,254	70,500	127,546	253,872	198,046	320,118	131,800	451,918
70	234,592	3,217	71,086	118,341	237,809	189,427	305,678	121,558	427,236
71	188,874	1,745	53,924	95,796	190,619	149,720	242,798	97,541	340,339
72	163,776	1,137	40,608	95,797	164,913	136,405	204,384	96,934	301,318
73	179,463	2,307	64,946	87,419	181,770	152,365	244,409	89,726	334,135
166	311,240	7,802	126,039	165,922	319,042	291,961	437,279	173,724	611,003
167	288,905	6,744	124,021	155,119	295,649	279,140	412,926	161,863	574,789
168	303,168	8,334	150,581	151,538	311,502	302,119	453,749	159,872	613,621
169	274,480	10,750	142,628	130,449	285,230	273,077	417,108	141,199	558,307
170	256,889	10,353	141,140	114,929	267,242	256,069	398,029	125,282	523,311
171	242,445	8,092	127,546	96,876	250,537	224,422	369,991	104,968	474,959
172	233,660	8,396	121,929	94,351	242,056	216,280	355,589	102,747	458,336
173	191,211	10,916	131,761	78,348	202,127	210,109	322,972	89,264	412,236
174	270,107	11,518	143,683	105,781	281,625	249,464	413,790	117,299	531,089
175	275,100	7,335	130,161	124,086	282,435	254,247	405,261	131,421	536,682
176	271,113	8,058	126,588	148,510	279,171	275,098	397,701	156,568	554,269
177	290,314	9,701	126,968	145,944	300,015	272,912	417,282	155,645	572,927
302	358,687	12,525	113,918	269,184	371,212	383,102	472,605	281,709	754,314
303	389,334	12,492	123,060	276,058	401,826	399,118	512,394	288,550	800,944
304	390,335	11,617	117,935	276,673	401,952	394,608	508,270	288,290	796,560
305	427,280	11,597	114,185	310,094	438,877	424,279	541,465	321,691	863,156
306	400,803	12,895	121,876	285,145	413,698	407,021	522,679	298,040	820,719
307	436,977	10,673	118,234	325,105	447,650	443,339	555,211	335,778	890,989
308	415,677	14,644	115,303	289,788	430,321	405,091	530,980	304,432	835,412
309	376,175	14,887	109,411	285,370	391,062	394,781	485,586	300,257	785,843
310	384,530	12,444	121,408	288,558	396,974	409,966	505,938	301,002	806,940
311	342,671	16,693	112,940	269,519	359,364	382,459	455,611	286,212	741,823
312	293,924	16,518	119,761	228,225	310,442	347,986	413,685	244,743	658,428
313	259,442	17,767	124,988	186,546	277,209	311,534	384,430	204,313	588,743
314	241,643	15,982	115,782	170,698	257,625	286,480	357,425	186,680	544,105

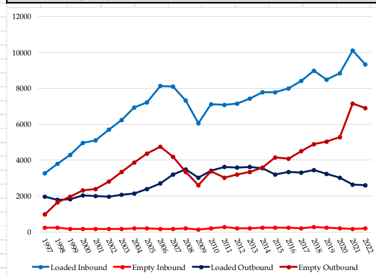
1	LA/LB Ports								
2	Loaded Inbound	Empty Inbound	Loaded Outbound	Empty Outbound	Total Inbound	Total Outbound	Total Loaded	Total Empties	Total
3	238,805	20,065	151,502	69,573	258,870	221,075	390,307	89,638	479,945
4	227,761	20,228	158,962	60,617	247,989	219,579	386,723	80,845	467,568
5	227,429	22,492	190,002	64,147	249,921	254,149	417,431	86,639	504,070
6	263,145	21,977	170,104	63,128	285,122	233,232	433,249	85,105	518,354
7	266,430	22,538	168,061	72,191	288,968	240,252	434,491	94,729	529,220
8	293,456	21,050	173,661	67,739	314,506	241,400	467,117	88,789	555,906
9	300,445	19,140	167,867	80,113	319,585	247,980	468,312	99,253	567,565
28	306,763	11,704	141,074	128,368	318,467	269,442	447,837	140,072	587,909
29	305,800	12,698	157,556	147,611	318,498	305,167	463,356	160,309	623,665
30	327,833	10,884	140,510	126,340	338,717	266,850	468,343	137,224	605,567
31	381,175	12,397	157,764	179,092	393,572	336,856	538,939	191,489	730,428
32	356,161	15,127	146,339	173,517	371,288	319,856	502,500	188,644	691,144
33	365,031	15,585	141,103	156,488	380,616	297,591	506,134	172,073	678,207
34	395,394	10,900	149,482	191,589	406,294	341,071	544,876	202,489	747,365
35	399,876	9,175	148,515	182,081	409,051	330,596	548,391	191,256	739,647
36	402,923	15,964	160,317	204,380	418,887	364,697	563,240	220,344	783,584
37	384,383	15,231	168,854	176,652	399,614	345,506	553,237	191,883	745,120
63	390,922	16,311	154,067	191,581	407,233	345,648	544,989	207,892	752,881
64	436,681	14,065	175,734	171,510	450,746	347,244	612,415	185,575	797,990
65	390,572	16,781	180,507	192,312	407,353	372,819	571,079	209,093	780,172
66	485,979	13,719	179,074	215,726	499,698	394,800	665,053	229,445	894,498
67	505,831	14,906	180,747	257,928	520,737	438,675	686,578	272,834	959,412
68	528,078	12,524	171,099	243,738	540,602	414,837	699,177	256,262	955,439
69	510,429	15,375	154,240	267,565	525,804	421,805	664,669	282,940	947,609
70	562,064	15,434	162,507	292,724	577,498	455,231	724,571	308,158	1,032,729
71	509,452	12,134	143,427	264,568	521,586	407,995	652,879	276,702	929,581
72	416,686	13,445	110,816	231,342	430,131	342,158	527,502	244,787	772,289
73	471,034	12,068	160,718	236,344	483,102	397,062	631,752	248,412	880,164
166	710,391	15,343	273,648	375,459	725,734	649,107	984,039	390,802	1,374,841
167	662,154	13,279	263,821	347,147	675,433	610,968	925,975	360,426	1,286,401
168	652,713	17,029	301,630	324,633	669,742	626,263	954,343	341,662	1,296,005
169	608,190	21,364	312,947	282,776	629,554	595,723	921,137	304,140	1,225,277
170	556,193	23,909	302,765	253,095	580,102	555,860	858,958	277,004	1,135,962
171	581,052	24,605	286,597	243,224	605,657	529,821	867,649	267,829	1,135,478
172	509,547	21,673	272,286	209,744	531,220	482,030	781,833	231,417	1,013,250
173	488,234	26,696	324,610	173,492	514,930	498,102	812,844	200,188	1,013,032
174	582,467	26,004	311,131	228,759	608,471	539,890	893,598	254,763	1,148,361
175	636,069	21,494	314,436	257,617	657,563	572,053	950,505	279,111	1,229,616
176	605,007	23,587	289,725	276,744	628,594	566,469	894,732	300,331	1,195,063
177	647,982	27,333	292,103	293,835	675,315	585,938	940,085	321,168	1,261,253
302	743,938	14,797	184,790	597,378	758,735	782,168	928,728	612,175	1,540,903
303	816,542	13,759	223,245	612,994	830,301	836,239	1,039,787	626,753	1,666,540
304	814,408	14,357	213,376	612,184	828,765	825,560	1,027,784	626,541	1,654,325
305	922,476	14,177	225,966	659,211	936,653	885,177	1,148,442	673,388	1,821,830
306	857,473	14,490	221,754	614,360	871,963	836,114	1,079,227	628,850	1,708,077
307	936,937	12,392	243,890	665,670	949,329	909,560	1,180,827	678,062	1,858,889
308	860,357	18,231	209,193	624,242	878,588	833,435	1,069,550	642,473	1,712,023
309	861,627	20,144	213,310	626,186	881,771	839,496	1,074,937	646,330	1,721,267
310	788,132	14,509	223,727	585,887	802,641	809,614	1,011,859	600,396	1,612,255
311	686,133	18,806	190,620	556,137	704,939	746,757	876,753	574,943	1,451,696
312	630,231	20,875	209,483	476,269	651,106	685,752	839,714	497,144	1,336,858
313	566,522	22,243	215,104	424,218	588,765	639,322	781,626	446,461	1,228,087
314	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 summarised and visualised in Figure 2. The data are in 1,000 TEUs; the last digits were rounded to zero.

Figure 2 26-year TEUs handling at LA/LB ports (see online version for colours)

Year	Year	Loaded Inbound	Empty Inbound	Loaded Outbound	Empty Outbound	Total Inbound	Total Outbound	Total Loaded	Total Empties	Total
1997	1	3270	230	1980	980	3500	2960	5250	1220	6460
1998	2	3810	230	1770	1660	4050	3430	5580	1900	7480
1999	3	4280	160	1810	1970	4440	3780	6090	2130	8220
2000	4	4950	170	2030	2330	5120	4360	6980	2500	9480
2001	5	5100	150	1990	2400	5260	4390	7090	2550	9650
2002	6	5690	170	1950	2820	5860	4770	7630	3000	10630
2003	7	6220	180	2070	3330	6410	5400	8290	3520	11810
2004	8	6930	190	2140	3850	7120	5990	9070	4040	13100
2005	9	7230	210	2390	4370	7440	6760	9620	4570	14190
2006	10	8130	180	2710	4740	8310	7450	10840	4920	15760
2007	11	8110	170	3180	4200	8290	7380	11300	4370	15670
2008	12	7330	190	3470	3350	7510	6820	10800	3540	14340
2009	13	6060	120	3020	2610	6180	5630	9080	2730	11810
2010	14	7100	200	3400	3390	7300	6790	10510	3590	14100
2011	15	7090	280	3620	3010	7370	6630	10710	3290	14000
2012	16	7150	200	3580	3190	7350	6770	10740	3390	14120
2013	17	7430	210	3630	3340	7640	6960	11060	3540	14600
2014	18	7790	250	3540	3590	8040	7120	11320	3840	15160
2015	19	7780	220	3180	4160	8010	7350	10970	4390	15350
2016	20	7990	230	3350	4070	8210	7420	11340	4300	15630
2017	21	8410	200	3290	4500	8610	7790	11700	4700	16400
2018	22	8970	260	3430	4900	9220	8330	12400	5150	17550
2019	23	8470	220	3230	5040	8700	8270	11700	5270	16970
2020	24	8830	200	3010	5270	9020	8280	11830	5460	17300
2021	25	10100	180	2620	7160	10280	9790	12720	7340	20060
2022	26	9330	210	2600	6900	9540	9500	11940	7110	19040

```
B2 =A2-1996
C2 =ROUND(SUMIF(Table1.Data312M!$B$3:$B$314,Table2.Data26Y!$A2,Table1.Data312M!Y$3:Y$314)/1000,-1)
D2 =ROUND(SUMIF(Table1.Data312M!$B$3:$B$314,Table2.Data26Y!$A2,Table1.Data312M!Z$3:Z$314)/1000,-1)
E2 =ROUND(SUMIF(Table1.Data312M!$B$3:$B$314,Table2.Data26Y!$A2,Table1.Data312M!AA$3:AA$314)/1000,-1)
F2 =ROUND(SUMIF(Table1.Data312M!$B$3:$B$314,Table2.Data26Y!$A2,Table1.Data312M!AB$3:AB$314)/1000,-1)
K2 =ROUND(SUMIF(Table1.Data312M!$B$3:$B$314,Table2.Data26Y!$A2,Table1.Data312M!AG$3:AG$314)/1000,-1)
```



3.2 *Characteristics of forecasting techniques*

All forecasting techniques have three key characteristics.

1 *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 a variable of interest such as 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.

2 *Forecasts of aggregate items are more accurate than forecasts of individual items*. Aggregate forecasts reduce variability. The forecast for all container terminals in all ports in the world is more accurate than the forecast for US ports, the forecast for US 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 Hueneme in Channel Island, North-East of Los Angeles. 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 the LA and LB ports are equal and represented by σ, then the variance for the volume of activities in the combined port is $2\sigma^2$. Therefore, the following year's activities volume standard deviation for the combined LA/LB ports is not 2σ. It is SQRT(2)σ ≈ 1.41σ < 2σ.

3 *Long-term forecasts are less accurate than short-term forecasts*. Forecast accuracy diminishes as we look further into the future. We benefit from more accurate information and make better predictions as we get closer to the demand time. Next year's LA/LB activities forecast is more accurate than the ten-year forecast.

3.3 Impact of characteristics of forecasting techniques on LA/LB ports throughput

Figure 3 shows the world's container port throughput in ten countries and ten ports. The combined LA/LB ports are ranked 10th in the world. They comprise the largest port complex in the Western Hemisphere.

Figure 3 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%

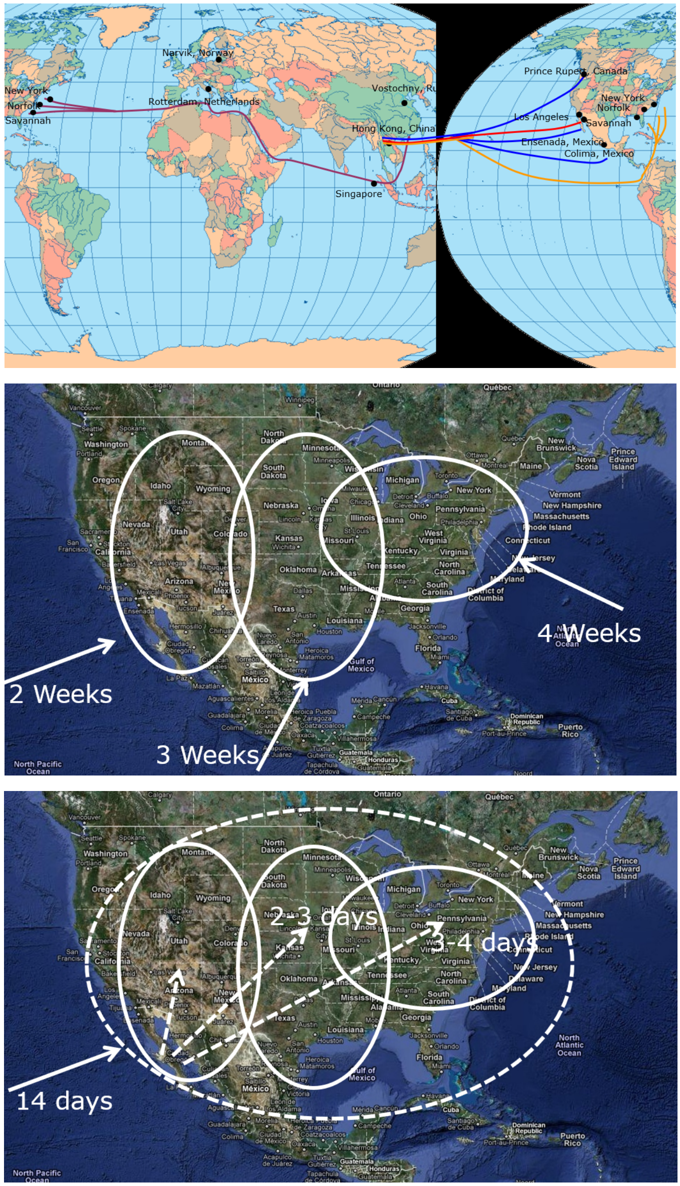
(b) Top 10 countries: 62%

Source: American Association of Port Authorities, 2020.

What are the competing edges of the LA/LB ports? Deepwater facilities with a depth of 60 feet for Post-Panama ships with 18,000 TEUs? Seventy miles of waterfront and about 16,000 acres, about half water, half land area? Fourteen container terminals with 150 ship-to-shore container cranes? State-of-the-art on-dock facilities to transfer containers between ships and trains? 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 the key reason behind the attractiveness of the LA/LB ports.

As pictorially shown in Figure 4, shipping containers from the Far East to the East Coast may take four weeks. This shipment takes two weeks to the west coast. For shipments from the Far East to the East Coast, one needs to forecast the demand for the East Coast four weeks in advance. However, the demand forecast two weeks in advance is enough for shipping to the West Coast. According to forecasting characteristics (III), the forecast for two weeks in the future is more accurate than the forecast for four weeks in the future. In addition, forecasting the east-cost demand when the commodity is in west-cost will be more accurate than in East Asia.

Figure 4 Forecasting-based competing edges of LA/LB ports (see online version for colours)



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 when the product is in the LA/LB ports. It will take 1–3 days to drayage the containers to the final 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. Using these two techniques, we can forecast the average and standard deviation of the next period's activities. The forecast for all future periods is assumed to be the same as the next period as a straight line. The forecasts are updated when the actual data for the next period becomes available. In Section 5, we include trends, and in Section 6, we include seasonality in the levels estimated in this section.

Details of all Excel formulas in all tables and all sections are summarised in a set of cells with grey backgrounds or in red fonts.

4.1 Moving average forecasts

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; $F_{27} = A_{26}$. A conservative analyst may consider the average of all years as the demand forecast for next year; F_{27} = AVERAGE $(A_{26} + A_{25} + \ldots + A_2 + A_1)$.

A more rational analyst, however, may stay between these two extremes and estimate the demand for the next period based on the observations in the past n-periods. An n-period moving average in period 26 is defined as MA_{26} = AVERAGE($A_{26}, A_{25}, \ldots, A_{26-n+1}$). The forecast for period 27 is then defined as the n-period moving average in year 26. The four-period moving average forecast in year 27 equals the four-period moving average in year 26; $F_{27} = MA^4_{26} = (A_{26} + A_{25} + A_{24} + A_{23})/4$. Generally, $F_{t+1} = MA^n_t = (A_t + A_{t-1} + \ldots + A_{t-n+1})/n$. Note that the n-period moving averages do not exist until period n, and n-period moving average forecasts do not exist until period n + 1.

Moving average formulas for one-period, all-period, and four-period moving averages are shown in Figure 16 columns C to E. In Appendix A, we develop a general dynamic formula adaptable to every n-period moving average, along with its dynamic tables and graphs. It provides a playground to practice advanced Excel functions and formulas.

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 $0 \leq \alpha \leq 1$. It has an autoregressive taste. A minor manipulation can restate the formula as $F_{t+1} = (1 - \alpha) F_t + \alpha A_t$. That is, the forecast for the next period is the weighted average of the forecast and actual for this period. It 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 the forecast for the first period: $F_1 = A_1$; F_1 = average of all existing actual values; F_1 = interpret of the linear regression line (discussed later).

We follow the first approach and set $F_1 = A_1$.

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

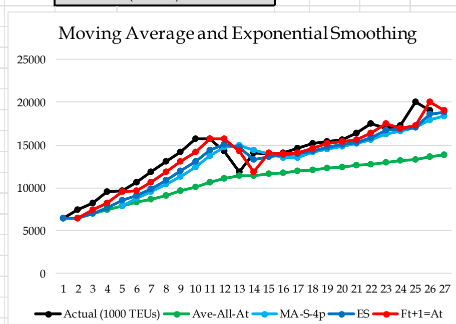
$F_{t\ 1} = F_t$; the forecast for the next period is equal to the forecast for this period and remains the same in all the coming periods.

We usually start with $\alpha = 0.5$ and use an optimisation tool, such as Excel's standard SOLVER add-ins or DataTable, to find the optimal α minimising one of the forecast accuracies measured, 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 continually get smaller on the older data. Exponential smoothing forecasts using $\alpha = 0.5$ are in column F of Figure 5. This figure also shows the graph for alternative forecasting techniques discussed up to this point that can be made using Excel's scatter graph or line chart.

Figure 5 Alternative moving average and exponential smoothing forecasts (see online version for colours)

n=	4			α=	0.5
Year	Actual (1000 TEUs)	$F_{t+1}=A_t$	Ave-All-A_t	MA-S-4p	ES
1	6460				6460
2	7480	6460	6460		6460
3	8220	7480	6970		6970
4	9480	8220	7387		7595
5	9650	9480	7910	7910.0	8537.5
6	10630	9650	8258	8707.5	9093.75
7	11810	10630	8653	9495.0	9861.875
8	13100	11810	9104	10392.5	10835.94
9	14190	13100	9604	11297.5	11967.97
10	15760	14190	10113	12432.5	13078.98
11	15670	15760	10678	13715.0	14419.49
12	14340	15670	11132	14680.0	15044.75
13	11810	14340	11399	14990.0	14692.37
14	14100	11810	11431	14395.0	13251.19
15	14000	14100	11621	13980.0	13675.59
16	14120	14000	11780	13562.5	13837.8
17	14600	14120	11926	13507.5	13978.9
18	15160	14600	12084	14205.0	14289.45
19	15350	15160	12254	14470.0	14724.72
20	15630	15350	12417	14807.5	15037.36
21	16400	15630	12578	15185.0	15333.68
22	17550	16400	12760	15635.0	15866.84
23	16970	17550	12978	16232.5	16708.42
24	17300	16970	13151	16637.5	16839.21
25	20060	17300	13324	17055.0	17069.61
26	19040	20060	13594	17970.0	18564.8
27		19040	13803	18342.5	18802.4

C8	=B7
D8	=AVERAGE(B$3:B8)
E8	=AVERAGE(B5:B8)
F8	=(1-$F$1)*F8+$F$1*B8



4.3 *Age of data in moving average and exponential smoothing*

A four-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 F_5 belongs to period 4 and is 1 period old. The oldest data belongs to period 1 and is 4 periods old. Therefore, in a four-period moving average, the average 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. It is proved in Appendix B that the age of data in exponential smoothing is $1/\alpha$. Given 2.5 as the age of data in a four-period moving average, the data in an exponential smoothing with $1/\alpha = 2.5$, i.e., $\alpha = 0.4$, has the same age. An exponential smoothing forecast with $\alpha = 0.6667$ has an age equivalent to a two-period moving average forecast. An exponential smoothing forecast with $\alpha = 0.1$ has an age equivalent to a 19-period moving average forecast.

A smaller α in exponential smoothing plays the same role as a larger number of periods in moving averages in smoothing out the recent fluctuations. Larger values for α in exponential smoothing, similar to the smaller number of periods in the moving average, result in higher responsiveness to recent fluctuations. An $\alpha = 1$ has the same role as a one-period moving average where the forecast for the next period is equal to the actual in this period.

5 Measuring forecast accuracy and variability

In this section, we answer two questions. How do we 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 if they are added or averaged. 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. Still, there may be significant gaps between actual and forecast values in both positive and negative directions. This problem can be resolved 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. A forecasting method provides F_{t+1} as the estimated average demand in the next period. In addition, 1.25 times the most recent MAD is the standard deviation of the next period forecast. In other words, $A_{t+1} \sim N(F_{t+1}, 1.25MAD_t)$. The demand for the next period follows a normal distribution with an average of F_{t+1} and a standard deviation of $1.25MAD_t$.

The tracking signal is defined as TS = SFE/MAD. It is a positive or negative number divided by a positive number. In an accurate forecasting method, the summation of all errors is expected to be zero. TS can jump up and down on the positive and negative sides due to randomness in the actual data, but in an unbiased forecasting method, it should remain close to zero. We also define the upper control limit (UCL) and lower control limit (LCL). 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 inaccurate.

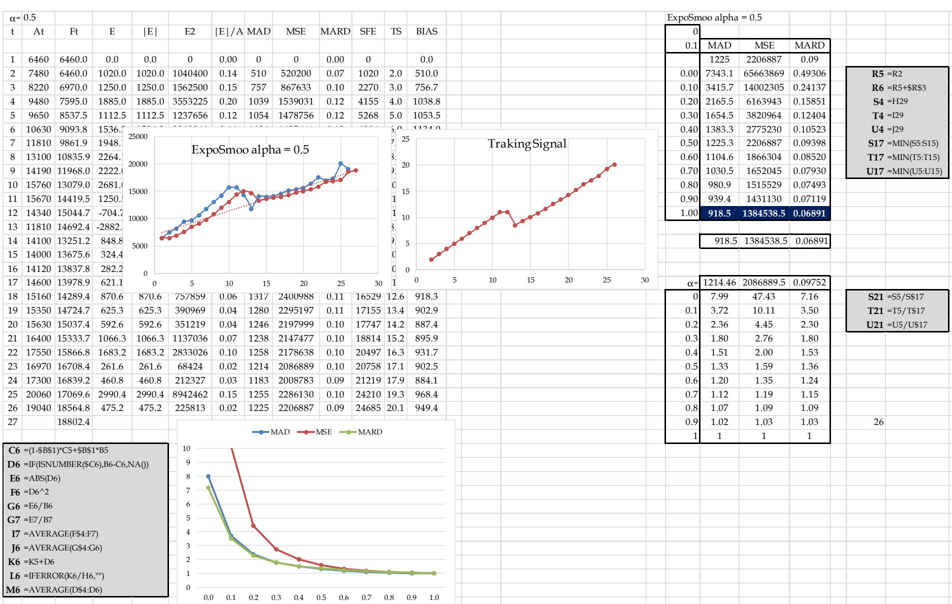
TS serves two essential purposes. First, we expect it to stay within UCL and LCL. Second, we do not expect to see a pattern over time. For example, we do not expect to see an always positive or consistently negative TS. Since we have the summation of $E_t = A_t – F_t$ in the numerator, in the first case,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

We may assign a higher weight to positive gaps than to a negative gap. In the second case, we have overstock; in the first case, we lost sales. Usually, the cost of overstock is less than the cost of lost sales. We may assign a coefficient greater than 1 to positive E_t values. We may also benefit from the newsvendor problem (Arrow et al., 1951; Schweitzer and Cachon, 2000; Iravani, 2021) to find a good tradeoff coefficient of underestimating and overestimating demand.

5.2 *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, for example, computing the derivative of an expression containing an absolute value, is difficult. It becomes easy if the squared values replace absolute values. In addition to 1.25MAD, the square root of MSE provides another estimate for the standard deviation of the forecast: that is $A_{t+1} \sim N(F_{t+1}, SQRT(MSE_t))$.

Figure 6 All metrics for forecast accuracy and reliability (see online version for colours)



There is a third method that we refer to as mean absolute relative deviation (MARD). It is also called mean absolute percentage error (MAPE) when multiplied by 100. Instead of averaging | E_t | values, we average | E_t | / A_t values. For example, | E_t | = 10 states that there were 10 units of deviations between A_t and F_t. If A_t = 200, then 10 relative to 200 is a 0.05 (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 minimise the maximum absolute deviation between actual and forecast.

Figure 6 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 $\alpha = 0.5$. Besides the actual vs. forecast and the tracking signal curves, the third curve shows the MAD, MSE, and MARD ratio to their minimum value as α changes from 0 to 1.

5.3 Optimal α value

The optimal α value can be computed in at least two ways; SOLVER and DataTable.

For SOLVER, the objective function is to set one of the three measures of MAD, MSE, or MARD (in cells H29, I29, and J29) to be minimised, and α, cell B1, is the changing cell to minimise the objective function value. For the DataTable, we set cells P4, Q4, and R4 equal to cells H29, I29, and J29, respectively. The α values start from a cell one column to the left and one row below MAD. Using a formula, we can find the value of α in the DataTable with as many decimal points as the value obtained by SOLVER. This is done by typing the starting α value of 0 and the increment in two arbitrary cells (such as cells R2 and R3 in this example). We then set R5 = R2 and R6 = R2 + \$R\$3 and copy down from 0 to 1. We will then mark R4 to U15, select Data tab, What-if Analysis, and DataTable. Since alternative α values are typed in a column (not in a row), inside the column input cell, point to B1, where the α value is placed (this should be the same cell used as α in the exponential smoothing formula). 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 R2 to 0.65 and R3 to 0.001 and find the minimum (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 the DataTable.

Optimal α computations using both SOLVER and DataTable for all three metrics and their normalisation (divide each by the minimal value in that column) are shown in Figure 6. The reader is encouraged to look into all the formulas in grey cells. We have also used conditional formatting to highlight the minimal values.

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 $\alpha = 1$. When the tracking signal shows a continual or increasing positive trend, we may add a constant to the forecast value. In Figure 7, we implemented a two-dimensional DataTable to find the optimal value for $\alpha = 0.66$ plus a constant of 495 to be added to the forecast to minimise MAD. The computations for exponential smoothing and the essential formulas are shown in Figure 7.

5.4 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 also not the case most of the time). Since we have an upward trend almost in all years, an $\alpha = 1$, 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 Cumt = SUM(A_t) function to check if a data set is stationary. The data is stationary if Cumt is close to a straight line. Figure 8 shows Cumt for our data is distant from a line. We will later discuss trend-adjusted exponential smoothing and regression for data with a trend.

Figure 7 Forecasting measures under optimal α and a constant for exponential smoothing (see online version for colours)

α=	0.66	Constant=	495					
t	At	Ft	E	|E|	MAD	SFE	TS	BIAS
1	6460	6460.0	0.0	0.0	0	0		0.0
2	7480	6955.0	525.0	525.0	263	525	2.0	262.5
3	8220	7796.5	423.5	423.5	316	949	3.0	316.2
4	9480	8571.0	909.0	909.0	464	1857	4.0	464.4
5	9650	9665.9	-15.9	15.9	375	1842	4.9	368.3
6	10630	10150.4	479.6	479.6	392	2321	5.9	386.9
7	11810	10961.9	848.1	848.1	457	3169	6.9	452.7
8	13100	12016.7	1083.3	1083.3	536	4253	7.9	531.6
9	14190	13226.7	963.3	963.3	583	5216	8.9	579.5
10	15760	14357.5	1402.5	1402.5	665	6618	10.0	661.8
11	15670	15778.1	-108.1	108.1	614	6510	10.6	591.8
12	14340	16201.8	-1861.8	1861.8	718	4648	6.5	387.4
13	11810	15468.0	-3658.0	3658.0	944	990	1.0	76.2
14	14100	13548.7	551.3	551.3	916	1542	1.7	110.1
15	14000	14407.6	-407.6	407.6	882	1134	1.3	75.6
16	14120	14633.6	-513.6	513.6	859	621	0.7	38.8
17	14600	14789.6	-189.6	189.6	820	431	0.5	25.4
18	15160	15159.5	0.5	0.5	774	432	0.6	24.0
19	15350	15654.8	-304.8	304.8	750	127	0.2	6.7
20	15630	15948.6	-318.6	318.6	728	-192	-0.3	-9.6
21	16400	16233.3	166.7	166.7	701	-25	0.0	-1.2
22	17550	16838.3	711.7	711.7	702	686	1.0	31.2
23	16970	17803.0	-833.0	833.0	708	-147	-0.2	-6.4
24	17300	17748.2	-448.2	448.2	697	-595	-0.9	-24.8
25	20060	17947.4	2112.6	2112.6	753	1518	2.0	60.7
26	19040	19836.7	-796.7	796.7	755	721	1.0	27.7
27		19805.9						
28		19805.9						
29		19805.9						
30		19805.9						

MAD		Constant	480	5			
	755	480	485	490	495	500	505
	0.6	764.3	763.5	762.8	762.0	761.3	760.7
	0.61	761.2	760.4	759.7	759.5	759.4	759.3
alpha	0.62	758.9	758.4	758.3	758.2	758.1	758.0
0.6	0.63	758.1	757.4	757.1	757.0	756.9	756.7
0.01	0.64	757.4	756.6	756.0	755.9	755.8	756.3
	0.65	756.7	756.0	755.2	755.2	755.6	756.1
	0.66	756.1	755.4	755.3	755.1	755.5	755.9
	0.67	755.9	755.7	755.6	755.4	755.5	755.9
	0.68	756.2	756.1	755.9	755.8	755.6	755.9
	0.69	756.7	756.5	756.3	756.2	756.0	756.0
	0.7	757.1	756.9	756.8	756.6	756.5	756.3
	Min=	755.1					

C6	=$D$1+(1-$B$1)*C5+$B$1*B5
K3	=J6
K4	=K3+$J$7
K2	=F29
L15	=MIN(L3:Q13)
B1	=XLOOKUP(O9,O3:O13,K3:K13)
D1	=XLOOKUP(L15,L9:Q9,L2:Q2)

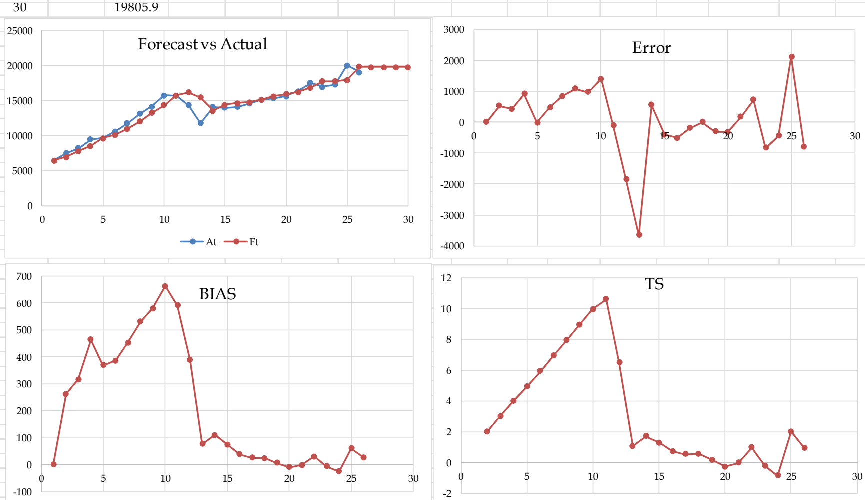
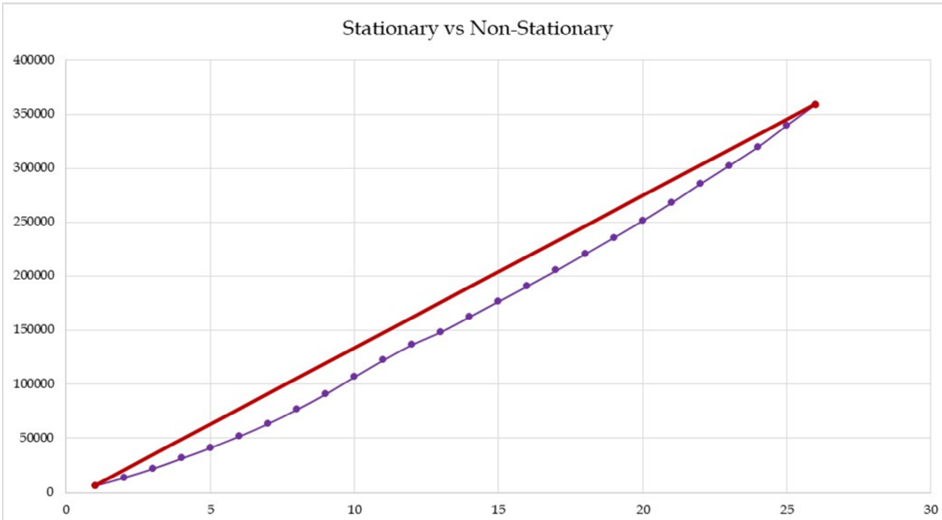


Figure 8 Stationary vs. non-stationary data (see online version for colours)



6 Level and trend

This section reviews:

1 bivariate linear regression

2 trend-adjusted exponential smoothing.

6.1 Bivariable linear regression

The bivariable linear regression is generally stated as $y = b_0 + b_1x$. Our time series 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. 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 tools are explained below and are summarised in Figure 9. Unlike moving average and exponential smoothing, where the forecast for all future periods is equal to the forecast for the next period, regression's forecast for any period t can be computed as $F_t = b_0 + b_1t$.

- *Procedure-1. Add trend line:* after presenting 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 linear. We also check the display equation and display the R-squared value boxes. The scatter graph shows the regression equation y = 419.22x + 8143.6 and $R^2 = 0.8418$. The larger the R-square value ($0 \le R^2 \le 1$), the more reliable the regression line.

- *Procedure-2. Data analysis add-ins:* choose data tab, data analysis, then regression. In the next table, enter the Y variable (A_t), then X variables (t), check the output range, and select a cell to be the east-north of the table (we select cell E1). You may also choose new worksheet to have a table in a new worksheet.

Figure 9 Alternative linear regression computations (see online version for colours)

t	At	Yhat (Reg)
1	6460	8562.8
2	7480	8982.0
3	8220	9401.3
4	9480	9820.5
5	9650	10239.7
6	10630	10658.9
7	11810	11078.1
8	13100	11497.4
9	14190	11916.6
10	15760	12335.8
11	15670	12755.0
12	14340	13174.2
13	11810	13593.5
14	14100	14012.7
15	14000	14431.9
16	14120	14851.1
17	14600	15270.3
18	15160	15689.6
19	15350	16108.8
20	15630	16528.0
21	16400	16947.2
22	17550	17366.5
23	16970	17785.7
24	17300	18204.9
25	20060	18624.1
26	19040	19043.3
27		19462.6
28		19881.8
29		20301.0
30		20720.2

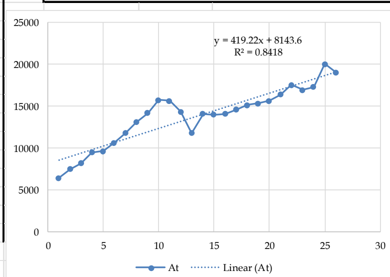
SUMMARY OUTPUT

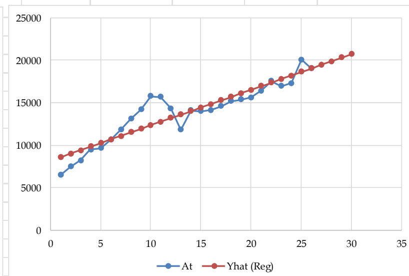
Regression Statistics			
Multiple R	0.9175 =CORREL($B$2:$B$27,$A$2:$A$27)	0.9175 =SQRT(F4)	Correlation Coefficient
R Square	0.8418 =RSQ($B$2:$B$27,$A$2:$A$27)	0.8418 =F3^2	Coefficient of Determination
Adjusted R Squa	0.8352	b0	8143.60 =INTERCEPT($B$2:$B$27,$A$2:$A$27)
Standard Error	1418.51 =STEYX(B2:B27,A2:A27)	b1	419.22 =SLOPE($B$2:$B$27,$A$2:$A$27)
Observations	26 =COUNT(B2:B27)	R-Square	0.8418 =RSQ($B$2:$B$27,$A$2:$A$27)
ANOVA		StdError	1418.51 =STEYX(B2:B27,A2:A27)

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	257028288.6	257028289	127.7369	4.2761E-11
Residual	24	48292065.23	2012169		
Total	25	305320353.8			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept (b0)	8143.60	572.83	14.22	3.47169E-13	6961.33	9325.87
X Variable 1 (b1)	419.22	37.09	11.30	4.27614E-11	342.67	495.78
					Zero is NOT Covered	

F16 =INTERCEPT($B$2:$B$27,$A$2:$A$27)
F17 =SLOPE($B$2:$B$27,$A$2:$A$27)
C19 =$F$14+$F$15*A19





If the range between the two blue numbers in the bottom part of Figure 9 (confidence interval under Lower 95% and Upper 95% for b_1) does not cover zero, there is a relationship between Y and X ($b_1 \neq 0$). Otherwise, we cannot claim a relationship. If the blue number in the top part of the table (under significance F) is less than 0.05, with more than 95% confidence, b_0 and b_1 are not both zero.

This approach is not recommended for bivariable linear regression if we do not need all the information this add-ins provides. That is because:

1 we must reproduce the table if a Y or X value changes

2 it also occupies a portion of the worksheet.

As shown in the seasonality-enhanced multivariate regression, using regression in data analysis add-ins is a good choice for bivariable nonlinear and multivariate linear and nonlinear cases.

- *Procedure-3. Excel functions:* as shown in Figure 9, we can compute most of the data analysis add-ins output using Excel functions such as INTERCEPT, SLOPE, RSQ, STEYX, CORREL, CONFIDENCE.NORM, CONFIDENCE.T, and additional formulas. It is the preferred method since all the metrics are already in the Excel cells (do not need to be read from a graph as in Procedure-1) and are updated continually (as is not the case in Procedure-2).

- *Procedure 4. Using more fundamental computations in Excel:* in Appendix C, we will provide fundamental insight into the computation of regression metrics from computing SST, SSE, and SSR, as well as a piecewise regression.

6.2 Trend-adjusted exponential smoothing

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

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

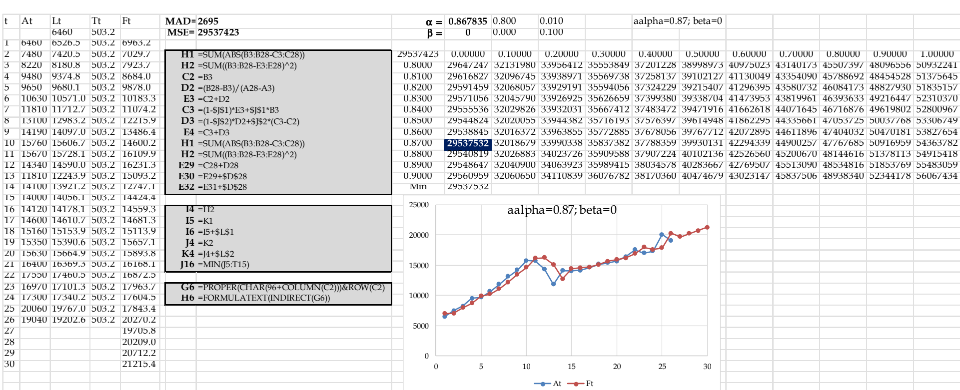
$$T_{t+1} = (1-\beta)\, T_t + \beta\left(L_{t+1} - L_t\right)$$

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, $L_0 = A_1 = 6{,}460$, and $T_0 = (A_{26} - A_1)/(26–1) = 503.2$ [based on Iravani (2021)]. Alternatively, we may set L_0 as the intercept of the regression line and T_0 as its slope. $L_0 = b_0 = 8{,}143.6$, and $T_0 = b_1 = 419.2$ [based on 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 DataTable to find the optimal values of $\alpha = 0.87$ and $\beta = 0$, as shown in Figure 10. 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 $F_t = \gamma F_{Trend\text{-}Adjusted.ES} + (1–\gamma)F_{Linear\text{-}Regression}$. The optimal γ value minimising the MSE of the forecasts from the actual values can then be obtained using SOLVER or DataTable.

Figure 10 Trend-adjusted exponential smoothing computations (see online version for colours)



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

MAD= **2695**
MSE= **29537423**

H1 =SUM(ABS(B3:B28-C3:C28))
H2 =SUM((B3:B28-E3:E28)^2)
C2 =B3
D2 =(B28-B3)/(A28-A3)
E3 =C2+D2
C3 =(1-$J$1)*E3+$J$1*B3
D3 =(1-$J$2)*D2+$J$2*(C3-C2)
E4 =C3+D3
H1 =SUM(ABS(B3:B28-C3:C28))
H2 =SUM((B3:B28-E3:E28)^2)
E29 =C28+D28
E30 =E29+$D$28
E32 =E31+$D$28

I4 =H2
I5 =K1
I6 =I5+$L$1
J4 =K2
K4 =J4+$L$2
J16 =MIN(J5:T15)

G6 =PROPER(CHAR(96+COLUMN(C2)))&ROW(C2)
H6 =FORMULATEXT(INDIRECT(G6))

α =	**0.867835**	0.800	0.010		aalpha=0.87; beta=0						
β =	**0**	0.000	0.100								
29537423	0.00000	0.10000	0.20000	0.30000	0.40000	0.50000	0.60000	0.70000	0.80000	0.90000	1.00000
0.8000	29647247	32131980	33956412	35553849	37201228	38998973	40975023	43140173	45507397	48096556	50932241
0.8100	29616827	32096745	33938971	35569738	37258137	39102127	41130049	43354090	45788692	48454528	51375645
0.8200	29591459	32068057	33929191	35594056	37324229	39215407	41296395	43580732	46084173	48827930	51835157
0.8300	29571056	32045790	33926925	35626659	37399380	39338704	41473953	43819961	46393633	49216447	52310370
0.8400	29555536	32029826	33932031	35667412	37483472	39471916	41662618	44071645	46716876	49619802	52800967
0.8500	29544824	32020055	33944382	35716193	37576397	39614948	41862295	44335661	47053725	50037768	53306749
0.8600	29538845	32016372	33963855	35772885	37678056	39767712	42072895	44611896	47404032	50470181	53827654
0.8700	29537532	32018679	33990338	35837382	37788359	39930131	42294339	44900257	47767685	50916959	54363782
0.8800	29540819	32026883	34023726	35909588	37907224	40102136	42526560	45200670	48144616	51378113	54915418
0.8900	29548647	32040900	34063923	35989415	38034578	40283667	42769507	45513090	48534816	51853769	55483059
0.9000	29560959	32060650	34110839	36076782	38170360	40474679	43023147	45837506	48938340	52344178	56067434
Min	29537532										

7 Level, trend, and seasonality

In this section, we review:

1 seasonality-enhanced bivariable linear regression

2 seasonality-enhanced multivariate regression using dummy variables

3 trend and seasonality-adjusted exponential smoothing.

7.1 Seasonality enhanced bivariable linear regression

The monthly data shown in Figure 2 for 12(26) months (in 1,000 TEUs) are copied into Figure 11. Periodicity is 12 (seasonality repeats every 12 months). One may add three months of data and consider the periodicity of four seasons, provide daily data and the periodicity of a week, or provide hourly data with a periodicity of 24 hours.

Figure 11 Computations for static seasonality enhanced bivariate linear regression (see online version for colours)

Per.	Monthly Data	Centered.MA	Deseas.Reg	Seas.Index	Seas	SeasInd	SeasIndAdj	Ft (Stat.Reg)
0						0.992	1.000	
1	480		705.71	0.680	1	0.942	0.95	670.74
2	468		708.61	0.660	2	0.865	0.87	618.11
3	504		711.50	0.708	3	0.909	0.92	652.43
4	518		714.40	0.726	4	0.964	0.97	694.16
5	529		717.30	0.738	5	1.024	1.03	740.59
6	556		720.19	0.772	6	1.006	1.01	730.67
7	568	541	723.09	0.785	7	1.045	1.05	762.28
8	557	544	725.98	0.768	8	1.078	1.09	789.49
9	589	551	728.88	0.808	9	1.037	1.05	762.59
10	583	559	731.78	0.797	10	1.054	1.06	777.90
11	556	567	734.67	0.757	11	1.005	1.01	744.93
12	556	575	737.57	0.753	12	0.968	0.98	720.36
13	527	582	740.46	0.711	1		0.95	703.77
14	512	591	743.36	0.689	2		0.87	648.42
15	608	600	746.26	0.815	3		0.92	684.29
16	611	606	749.15	0.815	4		0.97	727.93
17	632	614	752.05	0.841	5		1.03	776.47
296	1,762	1681	1560.02	1.130	8		1.09	1696.49
297	1,652	1685	1562.91	1.057	9		1.05	1635.21
298	1,692	1687	1565.81	1.081	10		1.06	1664.51
299	1,557	1685	1568.71	0.993	11		1.01	1590.62
300	1,541	1687	1571.60	0.980	12		0.98	1534.93
301	1,667	1694	1574.50	1.058	1		0.95	1496.47
302	1,654	1689	1577.39	1.049	2		0.87	1375.93
303	1,822	1675	1580.29	1.153	3		0.92	1449.08
304	1,708	1652	1583.19	1.079	4		0.97	1538.33
305	1,859	1623	1586.08	1.172	5		1.03	1637.59
306	1,712	1598	1588.98	1.077	6		1.01	1612.10
307	1,721		1591.87	1.081	7		1.05	1678.17
308	1,612		1594.77	1.011	8		1.09	1734.28
309	1,452		1597.67	0.909	9		1.05	1671.56
310	1,337		1600.56	0.835	10		1.06	1701.45
311	1,228		1603.46	0.766	11		1.01	1625.86
312	1,273		1606.35	0.792	12		0.98	1568.88
313					1		0.95	1529.50
314					2		0.87	1406.24
315					3		0.92	1480.95
316					4		0.97	1572.10
317					5		1.03	1673.47
318					6		1.01	1647.35
319					7		1.05	1714.80
320					8		1.09	1772.07
321					9		1.05	1707.92
322					10		1.06	1738.39
323					11		1.01	1661.09
324					12		0.98	1602.82

b0=	702.82
b1=	2.90
R2=	0.83
Periodicity=	12

C9	=(AVERAGE(B3:B14)+AVERAGE(B4:B15))/2
L1	=INTERCEPT($C$9:$C$308,$A$9:$A$308)
L2	=SLOPE($C$9:$C$308,$A$9:$A$308)
L3	=RSQ($C$9:$C$308,$A$9:$A$308)
D3	=$L$1+$L$2*A3
E3	=B3/D3
F3	=IF(MOD(A3,$L$4)>0,MOD(A3,$L$4),$L$4)
G3	=AVERAGEIF($F$3:$F$314,F3,$E$3:$E$314)
G2	=AVERAGE(G3:G14)
H3	=G3/$G$2
I317	=($L$1+$L$2*A317)*VLOOKUP(F317,$F$3:$H$14,3,0)

Step 1 Removing seasonality: 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. Unlike with the moving average, where we placed the average of n periods in front of the last period, we implement the centred moving average and place the average of the n periods at the centre of the n periods.

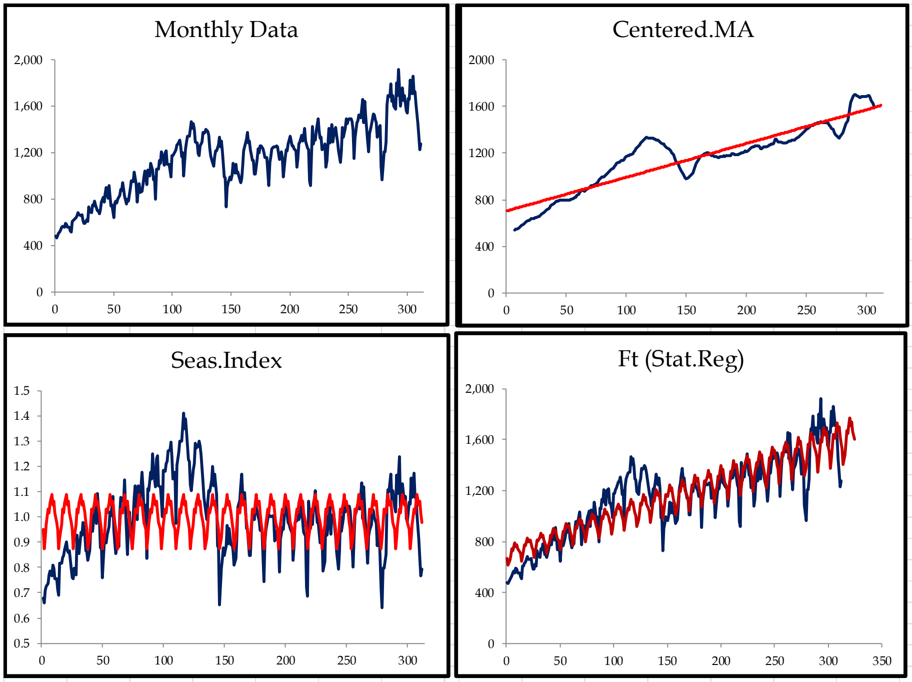
If we were considering seasonality over 7 days of a week, since 7 is odd, we could have placed the average in front of period 4, compared the actual period 4 with the centred moving average, and estimated the seasonality of period 4. But there is no middle period for even periodicity. Therefore, for the 12-period centred moving average, we compute the average of the first 12 periods and assume it is on the boundary of periods 6 and 7. Then, compute the average of periods 2 to 13 and assume it is on the boundary of periods 7 and 8. (In general, for even periodicity of n, we compute the average of periods 1 to n and assume it's on the boundary of periods n/2 and n/2 + 1. Then, compute the average of periods 2 to n + 1 and assume it's on the boundary of periods n/2 + 1 and n/2 + 2.) Next, we compute the average of these two centred moving averages and place it in front of period 7, representing the unseasonal activity volume at period 7. We then copy this formula to 6 months before the last months (month N – n/2). We will generally have the centred moving average for all periods minus periodicity.

Step 2 Trend in the deseasonalised data: we apply linear regression on months 7 (six months after 1) to 306 (six months before 312) 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 and L of Figure 11. The Excel worksheet also shows the formulas for all other computations, as discussed below.

Step 3 Seasonality indices: we divide the actual data of each month by the value obtained from the regression line applied to the deseasonalised data (A_t / 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 normalise, we divide the average seasonality index of each month by the average of the averages. These computations are in columns E–H. These seasonality indices remain fixed for all the past and future months. That is why Chopra (2019) refers to it as a Static method compared to the trend and seasonality-adjusted exponential smoothing, discussed later – as an adaptive method.

Step 4 Seasonality enhanced regression: finally, we put seasonality back on the deseasonalised 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 Figure 11. The results of the four steps of this process are schematically represented in Figure 12. 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) if the data is available.

Figure 12 Four key steps in static seasonality enhanced bivariable linear regression (see online version for colours)



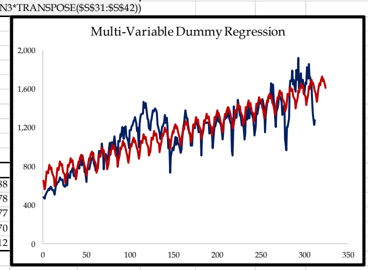
7.2 *Seasonality enhanced multivariate regression using dummy variables*

We use multivariate regression as another Static seasonality analysis approach by implementing a set of binary dummy variables. For each month, we define a binary variable, which is 1 if we are in that month and 0 otherwise. For periodicity of n, we need n – 1 dummy binary variables. Other periods are compared with the period with no binary variable. The period of choice does not affect the outcomes of the analysis. For a periodicity of 12, we define 11 binary variables from January to November. The dependent variable Y is the volume of activity in the corresponding month, and our X variables are 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. Therefore, we copy the month variables adjacent to the dummy variables. They can be in the first column to the left or right of the dummy variables. Compared to bivariable regression, when we select a single column as X or our month variable, we select 12 adjacent columns here. One column is associated with the month, and 11 are associated with 11 dummy variables. The output and all the essential formulas are shown in Figure 13. The reader may pay attention to the formula to generate 0s and 1s 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).

Figure 13 Seasonality enhanced multivariate regression computations (see online version for colours)

Per.	Monthly Data	1	2	3	4	5	6	7	8	9	10	11	12	t	Ft
0															
1	480	1	0	0	0	0	0	0	0	0	0	0	0	1	650
2	468	0	1	0	0	0	0	0	0	0	0	0	0	2	561
3	504	0	0	1	0	0	0	0	0	0	0	0	0	3	607
4	518	0	0	0	1	0	0	0	0	0	0	0	0	4	672
5	529	0	0	0	0	1	0	0	0	0	0	0	0	5	748
6	556	0	0	0	0	0	1	0	0	0	0	0	0	6	727
7	568	0	0	0	0	0	0	1	0	0	0	0	0	7	778
8	557	0	0	0	0	0	0	0	1	0	0	0	0	8	817
9	589	0	0	0	0	0	0	0	0	1	0	0	0	9	772
10	583	0	0	0	0	0	0	0	0	0	1	0	0	10	792
11	556	0	0	0	0	0	0	0	0	0	0	1	0	11	735
12	556	0	0	0	0	0	0	0	0	0	0	0	0	12	703
13	527	1	0	0	0	0	0	0	0	0	0	0	0	13	685
14	512	0	1	0	0	0	0	0	0	0	0	0	0	14	596
15	608	0	0	1	0	0	0	0	0	0	0	0	0	15	642
16	611	0	0	0	1	0	0	0	0	0	0	0	0	16	707
17	632	0	0	0	0	1	0	0	0	0	0	0	0	17	783
18	640	0	0	0	0	0	1	0	0	0	0	0	0	18	762
19	658	0	0	0	0	0	0	1	0	0	0	0	0	19	813
20	686	0	0	0	0	0	0	0	1	0	0	0	0	20	852
21	663	0	0	0	0	0	0	0	0	1	0	0	0	21	807
22	659	0	0	0	0	0	0	0	0	0	1	0	0	22	827
23	667	0	0	0	0	0	0	0	0	0	0	1	0	23	770
24	614	0	0	0	0	0	0	0	0	0	0	0	0	24	738
25	588	1	0	0	0	0	0	0	0	0	0	0	0	25	720
26	588	0	1	0	0	0	0	0	0	0	0	0	0	26	631
27	624	0	0	1	0	0	0	0	0	0	0	0	0	27	677
28	606	0	0	0	1	0	0	0	0	0	0	0	0	28	742
29	730	0	0	0	0	1	0	0	0	0	0	0	0	29	818
30	691	0	0	0	0	0	1	0	0	0	0	0	0	30	797
31	678	0	0	0	0	0	0	1	0	0	0	0	0	31	848
32	747	0	0	0	0	0	0	0	1	0	0	0	0	32	887
33	740	0	0	0	0	0	0	0	0	1	0	0	0	33	842
34	784	0	0	0	0	0	0	0	0	0	1	0	0	34	862
35	745	0	0	0	0	0	0	0	0	0	0	1	0	35	805
36	699	0	0	0	0	0	0	0	0	0	0	0	0	36	773
37	697	1	0	0	0	0	0	0	0	0	0	0	0	37	755
38	672	0	1	0	0	0	0	0	0	0	0	0	0	38	666
39	708	0	0	1	0	0	0	0	0	0	0	0	0	39	712
40	773	0	0	0	1	0	0	0	0	0	0	0	0	40	777
41	815	0	0	0	0	1	0	0	0	0	0	0	0	41	853
313		1	0	0	0	0	0	0	0	0	0	0	0	313	1,559
314		0	1	0	0	0	0	0	0	0	0	0	0	314	1,470
315		0	0	1	0	0	0	0	0	0	0	0	0	315	1,516
323		0	0	0	0	0	0	0	0	0	0	1	0	323	1,644
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))



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	1.0748E-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	0.00	65535	#NUM!	0.000	0.000
X Variable 13	2.91	0.09	31.553	#NUM!	2.730	3.093

7.3 *Trend and seasonality adjusted exponential smoothing*

A crucial difference between regression and trend-adjusted exponential smoothing is that regression has a static trend, while the trend is adaptive in trend-adjusted exponential smoothing. Also, for the two previous seasonality-adjusted approaches discussed in this section, we used the term static seasonality since seasonality indexes and all other coefficients remain unchanged yearly. You can observe unchanged seasonality indices in the two charts at the bottom of Figure 12 and the chart in Figure 13.

In this third approach, we update seasonality indices – along with level and trend – from one period to the next. It extends the trend-adjusted exponential smoothing (Winters, 1960; Chopra, 2019). The reader may look into seasonality in Figure 12 and Figure 13 and compare them with the graphs in Figure 14 to visualise the dynamism of this third approach.

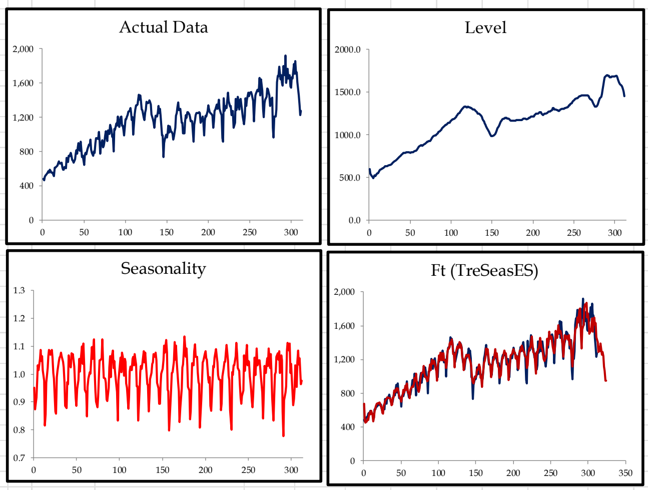
By applying linear regression on the 12-month centred moving average implemented in seasonality-enhanced bivariable linear regression, we first estimate the level (L_0 = INTERCEPT) and trend (T_0 = SLOPE) in month zero. We use static seasonality indexes computed in seasonality-enhanced bivariable linear regression [implemented in Chopra (2019)]. Alternatively, we may set L_0 equal to the average demand in the first 12 months. We also compute L_N as the average of the last 12 months, and 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 L_0 and set them as the seasonality indexes for the first 12 months [implemented in Iravani (2021)]. While the second approach is easier with fewer computations to estimate the starting parameters, since we already have the results for seasonality-enhanced bivariable linear regression, we follow the first approach and copy L_0, T_0, S_1, …, S_{12} from Figure 11 into Figure 14. We first set $\alpha = 0.5$, $\beta = 0.5$, and $\gamma = 0.5$.

Figure 14 Seasonality enhanced multivariate regression computations (see online version for colours)

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

p=	4
α=	0.50
β=	0.5
γ=	0.50

C2	=Table9.SeasRegChopraMineEvenp12!L1
D2	=Table9.SeasRegChopraMineEvenp12!L2
E3	=Table9.SeasRegChopraMineEvenp12!H3
F3	=(C2+D2)*E3
C3	=(1-$I$2)*(C2+D2)+$I$2*B3/E3
D3	=(1-$I$3)*D2+$I$3*(C3-C2)
F3	=(C2+D2)*E3
E15	=(1-$I$4)*E3+$I$4*(B3/(C2+D2))



Step 1 *Compute L_t*. Given $L_0 = 702.82$, $T_0 = 2.9$, and $S_1 = 0.95$; $F_1 = (L_0 + T_0) S_1 = (702.82 + 2.9) \times 0.95 = 670.74$. 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 $L_1 = L_0 + T_0 = 705.72$. This needs to be multiplied by $(1 - \alpha)$. That is, $L_1 = (1 - 0.5)\times 705.71$. 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 $1 - \alpha$, $1 - \beta$, and $1 - \gamma$ is easier. The month 1 actual data of 480 contains seasonality. We need to remove seasonality. Since $S_1 = 0.95$ (month 1 is a low season), we divide the actual data by S_1 to remove seasonality; $480/0.95 = 504.97$. This is the unseasoned value of the actual data in month 1. Accordingly, $L_1 = (1 - \alpha)(L_0 + T_0) + \alpha (A_1/S_1) = (1 - 0.5)\times(702.82 + 2.9) + 0.5\times(480/0.95) = 605.34$.

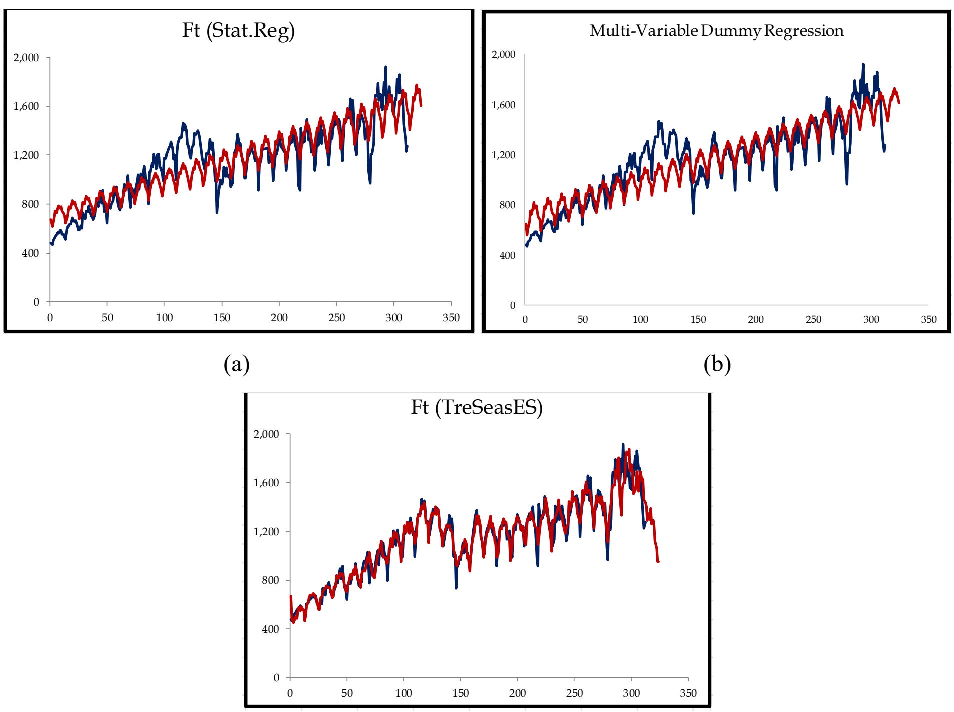
Step 2 *Compute T_t.* Our forecast for T_1 is T_0. It is multiplied by $(1 - \beta)$ to form the first component of T_1. What is the actual T_1? It is the difference between L_0 and L_1 to be multiplied by β. Therefore $T_1 = (1 - \beta)\, T_0 + \beta\, (L_1 - L_0) = (1 - 0.5) \times 2.90 + 0.5(605.34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

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 $F_2 = (L_1 + T_1) \times S_2 = (605.34 - 47.29) \times 0.872 = 486.78$.

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

Figure 14 shows all the key formulas and curves related to trend and seasonality-adjusted exponential smoothing components.

Figure 16 The results of the three seasonality enhanced/adjusted methods, (a) static regression (b) dummy-multivariate regression (c) trend and seasonality adjusted exponential smoothing (see online version for colours)



(a) (b)

(c)

8 Conclusions

We have learned that when theoretical concepts are taught through real-life applications, they positively impact students' mental presence and intellectual engagement inside the classroom.

In this study, we tried to streamline the learning process by applying time series and regression analysis to a significant real-life application.

We reviewed and integrated alternative time series and regression analysis techniques. This manuscript can be used as teaching material or as a case study to enhance 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.

We handpicked a set of intermediate to advanced Excel functions and formulas for step-by-step improvement of Excel skills and side-by-side enrichment of time series and regression knowledge of undergraduate and graduate students at teaching-focused business schools. The approach is tailored to the student population's knowledge, skills, and abilities in teaching-focused business schools. The Excel sheets designed in this manuscript could serve as templates for other real-life applications the students may encounter in their early employment years.

References

Albright, C. and Winston, W. (2015) *Business Analytics: Data Analysis & Decision Making*, 7th ed., Cengage, Boston, MA.

Arrow, K.J., Harris, T. and Marshak, J. (1951) 'Optimal inventory policy', *Econometrica*, Vol. 19, No. 3, pp.250–272.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015) *Time Series Analysis: Forecasting and Control*, 5th ed., Wiley, Hoboken, NJ.

Brockwell, P.J. and Davis, R.A. (2016) *Introduction to Time Series and Forecasting*, 3rd ed., Springer-Verlag, New York.

Cachon, G. and Terwiesch, C. (2020) *Matching Supply and Demand: An Introduction to Operations Management*, 4th ed., McGraw-Hill, New York, NY.

Camm, J.D., Cochran, J.J., Fry, M.J. and Ohlmann, J. (2020) *Business Analytics, Prescriptive, Predictive, Prescriptive*, 5th ed., Cengage, Boston, MA.

Chase, R.B., Aquilano, N.J. and Jacobs, R.F. (2000) *Operations Management for Competitive Advantage*, Irwin/McGraw-Hill, New York, NY.

Chopra, S. (2019) *Supply Chain Management*, Pearson, New York, NY.

Duncan, R.M. (2007) *CFPIM, Training Manager*, E/Step Software Inc [online] http://www.estepsoftware.com/papers/madrsquare.pdf (accessed 7 December 2023).

Heizer, J., Render, B. and Munson, C. (2023) *Operations Management: Sustainability and Supply Chain Management*, 14th ed., Pearson Publishing, New York, NY.

Hillier, F.S. and Hillier, M.S. (2023) *Introduction to Management Science and Business Analytics*, 7th Ed., McGraw Hill, New York, NY.

Holt, C.C. (1957) *Forecasting Seasonals and Trends by Exponentially Weighted Averages*, ONR Memorandum No. 52, = Carnegie Institute of Technology, Pittsburgh USA [online] https://www.sciencedirect.com/science/article/abs/pii/S0169207003001134?via%3Dihub (accessed 12 July 2023).

Iravani, S.M.R. (2021) *Operations Engineering and Management: Concepts, Analytics, and Principles for Improvement*, McGraw-Hill, New York, NY.

Jaggia, S., Kelly, A., Lertwachara, K. and Chen, L. (2023) *Business Analytics, Communicating With Numbers*, 2nd ed., McGraw Hill, New York, NY.

Keating, B. and Wilson, H.J. (2019) *Forecasting and Predictive Analytics*, McGraw Hill, New York, NY.

Krajewski, L.J., Malhotra, M.K. and Ritzman, L.P. (2016) *Operations Management, Processes and Supply Chains*, 11th ed., Pearson, New York, NY.

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

Powell, S.G. and Baker, K.R. (2017) *Business Analytics*, Wiley, Hoboken, NJ.

Ragsdale, C.T. (2018) *Spreadsheet Modeling and Decision Analysis*, 9th ed., Cengage, Boston, MA.

Schweitzer, M.E. and Cachon, G.P. (2000) 'Decision bias in the newsvendor problem with a known demand distribution: experimental evidence', *Management Science*, Vol. 43, No. 3, pp.404–420.

Stevenson, W.J. (2014) *Operations Management*, 12th ed., McGraw-Hill, ISBN: 978-0078024108.

Vandeput, N. (2023) *Demand Forecasting Best Practices*, Manning Publishing, Shelter Island, NY.

Venkataraman, R.R. and Pino, J.K. (2018) *Operations Management, Managing Global Supply Chains*, Sage Publishing, Los Angeles, CA.

Winston, W.L. (2022) *Microsoft Excel Data Analysis and Business Modeling*, Pearson, New York, NY.

Winters, P.R. (1960) 'Forecasting sales by exponentially weighted moving averages', *Management Science*, Vol. 6, No. 3, pp.324–342 [online] https://pubsonline.informs.org/doi/10.1287/mnsc.6.3.324 (accessed 12 July 2023).

Appendix A

Computation of metrics and drawing the graphs for an any-period moving average

Computational considerations. Consider a four-period moving average forecast in periods 25 and 26 and examine the differences.

$$F_{26} = MA_{25} = (A_{25} + A_{24} + A_{23} + A_{22}) / 4 = (A_{25} + A24 + A23)/4 + A_{22}/4$$

$$F_{27} = MA_{26} = (A_{26} + A_{25} + A_{24} + A_{23})/4 = A26/4 + (A_{25} + A_{24} + A_{23})/4.$$

Therefore, $F_{27} = F_{26} + A_{26}/4 - A_{22}/4$.

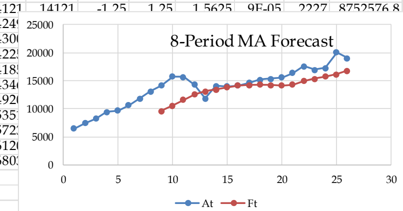
Given this fundamental insight, we develop a general formula applicable to any number of periods in a moving average computation as $F_{(t+1)} = F_t + (A_t - A_{t-n})/n$. This formula directly computes the next period's forecast to be equal to the forecast for this period (the moving average of the previous period) plus this period's actual data (the newest piece of data) minus the oldest piece of data used for this period' forecast. They (the newest and oldest data) are divided by n.

In Figure A1, we enter the number of periods in the moving average in cell A_1 using the RANDBETWEEN(2, 12) function. Let's say it generates 4. Now, consider the formula in row C9, which is the forecast for period 6. We have both the previous forecast and the actual in row 8. The newest piece of data is in A8, but how do we access the oldest piece of data used in the previous forecast? This data is in the row t–n of the actual data, which in our example is row 8–4 = 4 on 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 (referenced by A8) is less than 4 (the number of periods in the moving average), a blank space (" ") is entered in the corresponding forecast cell (C9) to leave it empty. 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 rows 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). It will find the oldest piece of data used in the forecast: INDEX(B$4:B8,5-4) = INDEX(B$4:B8,1) = B4 = 6,460. The actual for the previous period is B8 = 9,650, and the forecast for the previous period is C8 = 7,910. Therefore, the forecast for this period C9 = 7,910 + (9,650–6,460)/4 = 8,707.5. The figure is adjusted for any number less than 26 that may appear in cell A_1.

Figure A1 Computation and evaluation of an any-period moving average (see online version for colours)

8													
t	At	Ft	Ft	E	|E|	E2	|E|/A	MAD	MSE	MARD	SFE	TS	BIAS
1	6460												
2	7480												
3	8220												
4	9480												
5	9650												
6	10630												
7	11810												
8	13100												
9	14190	9603.8	9603.8	4586.3	4586	2.1E+07	0.3232	4586	21033689	0.323	4586.3	1	4586
10	15760	10570	10570	5190	5190	2.7E+07	0.3293	4888	23984895	0.326	9776.3	2	4888
11	15670	11605	11605	4065	4065	1.7E+07	0.2594	4614	21498005	0.304	13841	3	4614
12	14340	12536	12536	1803.8	1804	3253514	0.1258	3911	16936882	0.259	15645	4	3911
13	11810	13144	13144	-1334	1334	1778889	0.1129	3396	13905283	0.23	14311	4.21446	2862
14	14100	13414	13414	686.25	686.3	470939	0.0487	2944	11666226	0.2	14998	5.09397	2500
15	14000	13848	13848	152.5	152.5	23256.3	0.0109	2545	10002945	0.173	15150	5.95201	2164
16	14120	14121	14121	-1.25	1.25	1.5625	9E-05	2227	8752576.8	0.151	15149	6.80126	1894
17	14600	1424								0.137	15500	7.67749	1722
18	15160	1430								0.129	16360	8.59695	1636
19	15350	1422								0.124	17485	9.54279	1590
20	15630	1418								0.121	18930	10.5167	1578
21	16400	1434								0.122	20984	11.5326	1614
22	17550	1492								0.124	23614	12.5778	1687
23	16970	1535								0.122	25233	13.5646	1682
24	17300	1572								0.12	26810	14.5509	1676
25	20060	1612								0.124	30750	15.6418	1809
26	19040	1680								0.124	32988	16.6522	1833
27													
28													
29													
30													



8-Period MA Forecast			
	MAD	MSE	MARD
	1981.0	6167272	0.124
1	955.2	1439920	0.072
2	1146.3	1987246	0.086
3	1367.2	2651752	0.100
4	1527.6	3325370	0.108
5	1719.0	4243650	0.118
6	1869.5	5105986	0.124
7	1955.3	5783801	0.126
8	1981.0	6167272	0.124
9	2002.2	6251843	0.123
10	1935.1	5422733	0.118
11	1883.4	4750157	0.114
12	1988.2	5019616	0.120
13	2312.2	6199787	0.140
14	2473.0	6860163	0.149
15	2678.1	7812002	0.160
16	2925.4	9136685	0.172
17	3187.8	10720558	0.186
18	3450.7	12474581	0.198
19	3762.4	14754591	0.213
20	4123.0	17631550	0.229
21	4456.7	20576203	0.244
22	4648.8	22450494	0.252
23	5152.2	27457606	0.272
24	5942.5	35942684	0.303
25	5446.4	29663273	0.286
Min	955.2	1439920	0.07167

```
A1  =RANDBETWEEN(2,12)
C9  =IF(A8<$A$1,NA(),IF(A8=$A$1,AVERAGE(B$4:B8),C8+B8/$A$1-INDEX(B$4:B8,A8-$A$1)/$A$1))
D9  =IF(A8<$A$1,NA(),AVERAGE(B8:INDEX(B$4:B8,A8-$A$1+1)))  Alternative formula for Ft
E9  =IF(ISNUMBER($C9),B9-C9,NA())
F9  =IF(ISNUMBER($E9),ABS($E9),NA())
G9  =IF(ISNUMBER($E9),$E9^2,NA())
H9  =IF(ISNUMBER($E9),F9/B9,NA())
I9  =IF(ISNUMBER($E9),AVERAGE(F9:INDEX(F$4:F9,$A$1+1)),NA())
J9  =IF(ISNUMBER($E9),AVERAGE(G9:INDEX(G$4:G9,$A$1+1)),NA())
K9  =IF(ISNUMBER($E9),AVERAGE(H9:INDEX(H$4:H9,$A$1+1)),NA())
L9  =IF(ISNUMBER($E9),SUM(E9:INDEX(E$4:E9,$A$1+1)),NA())
M9  =IF(ISNUMBER($E9),L9/I9,NA())
N9  =IF(ISNUMBER($E9),AVERAGE(E9:INDEX(E$4:E9,$A$1+1)),NA())
```

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

- *Combined excel function considerations:* 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$4:B6,A6-$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. All the key formulas of Figure A1 are re-emphasized in the cells with the grey background.

- *Visualization considerations:* since we draw curves for some of the columns in Figure A1, a blank space (“ “) for starting years less than or equal to the random number that appears in cell A_1 will result in a Y-value equal to 0 even though it is empty (not zero). To resolve this, we replace the blank space (“ “) with the NA() function. In that case, the curves will not draw anything when encountering a #NA. However, to avoid #NA appearing in the table, we use formula-based conditional formatting with the IFERROR function and switch the font colour of these cells to white. Accordingly, Figure A1 are adjusted automatically no matter what random numbers between 2 and 25 appear in cell A_1. Alternatively, we could have the fonts of these columns coloured white and switch the font colour to black using the ISNUMBER function in conditional formatting. This way, Figure A1 automatically adjust regardless of the random number between 2 and 25 in cell A_1. Alternatively, we could initially set the font colour of these columns to white and then use conditional formatting with the ISNUMBER function to switch the font colour back to black for valid numbers.

Appendix B

Exponential smoothing basic mathematics

In this appendix, we show that:

1 exponential smoothing is a weighted moving average

2 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.

$$F_1 = A_1$$

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

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

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

$$F5 = (1-\alpha)F_4 + \alpha A_4 \rightarrow F_5 = (1-\alpha)^3 A_1 + \alpha(1-\alpha)^2 A_2 + \alpha(1-\alpha)A_3 + \alpha A_4$$

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

The sum of the weights are

$$S = \alpha + \alpha(1-\alpha) + \alpha(1-\alpha)^2 + \alpha(1-\alpha)^3 + \alpha(1-\alpha)^4 \ldots + \alpha(1-\alpha)^{t-1}$$
$$= \alpha(1-\alpha) + \alpha(1-\alpha)^2 + \alpha(1-\alpha)^3 + \alpha(1-\alpha)^4 \ldots + \alpha(1-\alpha)^t$$
$$S - (1-\alpha)S = \alpha - \alpha(1-\alpha)^t \to \alpha S = \alpha\left(1-(1-\alpha)^t\right) \to S = 1-(1-\alpha)^t$$

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 $1/\alpha$.

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

$$\text{Ages} = 1,\ 2,\ 3,\ 4, \ldots, t$$

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

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

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

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

Derivation with respect to α

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

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

B.3 UCL and LCL in tracking signal are larger than + 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)].

$$E_t = \text{Normal } (0, 1.25\text{MAD})$$
$$\text{If } x = \text{Normal}(\mu, \sigma) \to \text{Sum } (x) = \text{Normal}(\mu, \text{SQRT(N)}\sigma)$$
$$\text{StdDev } \left[\text{Sum}(E_t)\right] = \text{SQRT(N)StdDev}(E_t)$$
$$E_t = \text{Normal } (0, 1.25\text{MAD})$$
$$\text{Sum}(E_t) = N \sim (0,\ \text{SQRT(N)}1.25\text{MAD})$$
$$3 \geq \left(\sum E_t - 0\right) \Big/ \left(\text{SQRT(N)}1.25\text{MAD}\right) \geq -3.$$
$$+3\text{SQRT(N)}1.25 \geq \left(\sum E_t - 0\right) \Big/ \text{MAD} \geq -3\text{SQRT(N)}1.25.$$
$$+\ 3.75\text{SQRT(N)} \geq \left(\sum E_t - 0\right) \Big/ \text{MAD} \geq -\ 3.75\text{SQRT(N)}$$

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 metrics in Excel (b_0, b_1, SST, SSE, SSR)

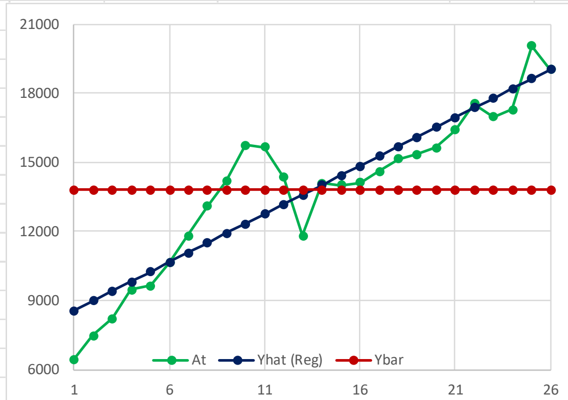
Regression lines can be designed by minimizing various error measures, including MAD, MSE, MAX(| E_t |), or any other measure. Conventionally, regression equations are designed based on MSE minimization (least-square method). However, the conventional approach uses the least-squares method, which uses MSE or SSE (Sum of Squared Errors). We use SOLVER to find the optimal b_0 and b_1 (in cells J16 and J17 in Figure A1) for the minimal value for SSE (cell D28). 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 compute SSE in cell D26. SOLVER is then used (DataTable is another option) to find optimal b_0 and b_1 values minimizing SSE (or MSE). These optimal values in cells J16 and J17 are the same as those found using the three regression tools mentioned in this manuscript. It highlights the significant power of dynamic arrays, as shown by the formula in cell J19 for direct SSE calculations.

Quantitative foundations of regression metrics

We usually compute three SST, SSE, and SSR metrics in regression analysis. SST (total squared deviations from the mean) is the summation of the squares of the gaps between each pair of points on the green (actual data) and the red (average of all data) lines. SSE (total squared deviations from the regression line) represents the summation of the squares of the gaps between each pair of points on the green and the blue curve (regression) lines. SSR = SST-SSE represents how well the regression line could replace the average data line. The reader may compare the computations in cells D28, F28, and G28 with those of J19, J20, and J21 to better understand the efficiency of dynamic arrays (and may delete columns D, E, F, and G).

Figure C1 Direct computation of regression coefficients and key metrics (see online version for colours)

Ybar	SE	SR
13803.1	4421854	49498925
13803.1	2256127	37725175
13803.1	1395379	29775369
13803.1	115928	18573066
13803.1	347749	16900299
13803.1	837	10067581
13803.1	535614	3436742
13803.1	2568442	-2074125
13803.1	5168418	-5018708
13803.1	11725111	-7895563
13803.1	8497076	-5011674
13803.1	1358982	-1070696
13803.1	3180753	791602
13803.1	7624	80540
13803.1	186544	-147766
13803.1	534548	-434108
13803.1	449367	185719
13803.1	280444	1560797
13803.1	575762	1817209
13803.1	806422	2531226
13803.1	299462	6444548
13803.1	33690	14005742
13803.1	665320	9364081
13803.1	818830	11409641
13803.1	2061772	37087314
13803.1	11	27425352
	48292065	257028289



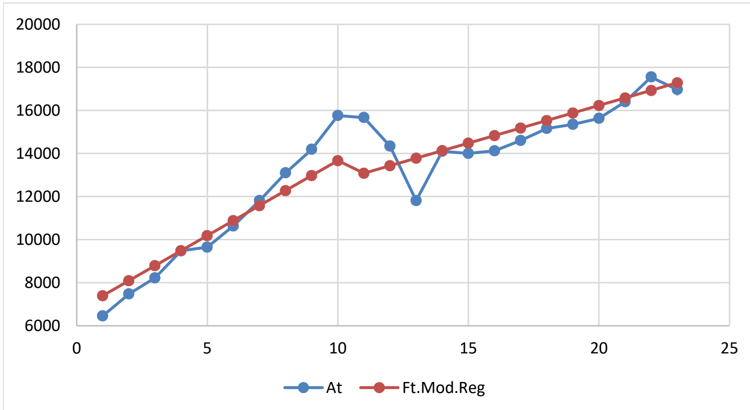
Y1	8562.82	=J16+J17*A2
b0	8143.60	
b1	419.22	
Ybar	13803	=AVERAGE(B2:B27)
SST	305320354	=SUM((B2:B27-J18)^2)
SSE	48292065	=SUM((B2:B27-C2:C27)^2)
SSR	257028289	=J19-J20
R-Squared	0.8418315	=J21/J19
MSE	2012169	=J20/(COUNT(B2:B27)-2)
StdErr	1418.51	=SQRT(J23)

R-squared, calculated as SSR/SST, reflects the proportion of variance explained by the regression model. It provides the same value as the RSQ function. While MSE is traditionally viewed as SSE divided by the number of data points, we must include degrees of freedom in regression analysis. We lose degrees of freedom when we benefit from one statistic extracted from the same data set to compute another statistic. 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).

Piecewise regression

Leveraging the background provided in this Appendix, we can employ piecewise regression to estimate b_{01}, b_{11}, b_{02}, and b_{12} as the parameters for the first and second segments of the regression line and T as the year to switch from the first regression line to the second. These five changing cells, and MSE as the objective function to be minimized, lead to the piecewise regression line visualized in Figure C2.

Figure C2 Piecewise regression on la/lb ports annual data (see online version for colours)



Appendix D

All worksheets used in this study

Due to the extensive calculations across multiple worksheets, recalculating the entire workbook can slow down the computations of the metrics of interest. Splitting the worksheets you need into separate files can significantly improve performance. Please contact the author for the entire workbook used in preparing this manuscript.