Sanchita Panja
Luis Estrada
Jackson Ver Steeg

Using Multiple Regression to Calculate Global Costs for Importing & Exporting

Outline

- 1. Introduction of Problem
- 2. Data set & Basic Regression Model
- 3. Specific Applications

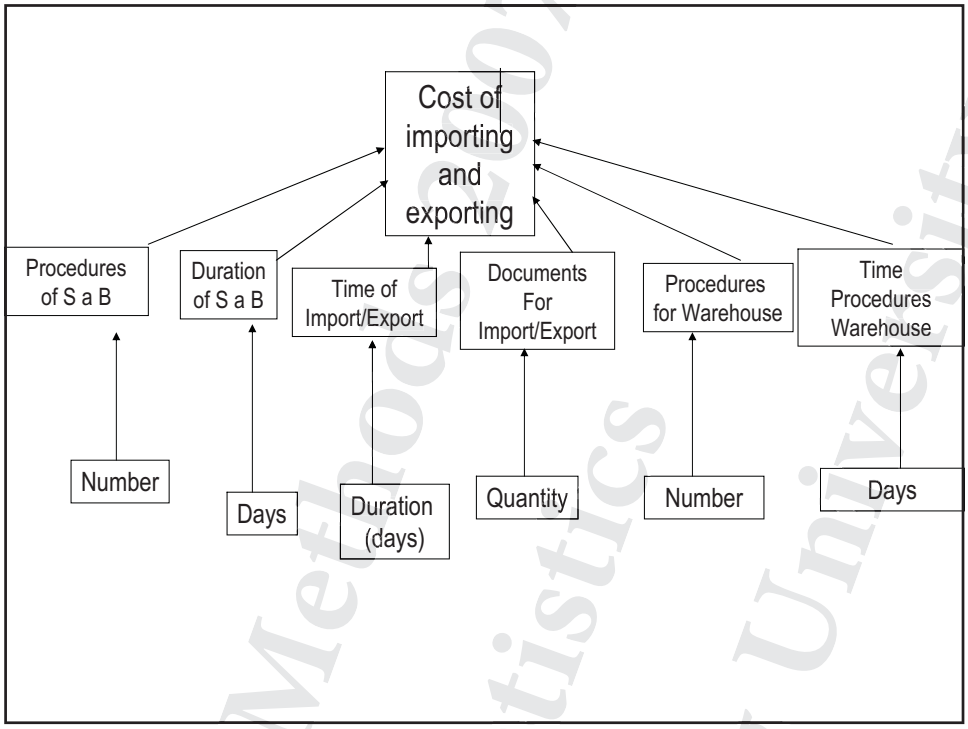
Introduction

- World Bank global economic data (total: 42)
  - Starting a business (4)
  - Warehouse licensing (3)
  - Employing Workers (6)
  - Registering Property (3)
  - Getting Credit (4)
  - Protecting Investors (4)
  - Paying Taxes (6)
  - Trading Across Borders (6)
  - Enforcing Contracts (3)
  - Closing a Business (3)

Variables can be chosen from among most likely factors: an infinite variety!

Regression 1: Cost of Importing & Exporting

- Starting a business:
  - Cost (% of GNI) (money required to start business)
  - Procedures (#)
  - Duration (days)
- Trading:
  - Documents required for import & export process
  - Average time for export & import process
- Employment:
  - Rigidity of Employment (an index reflecting how easy it is for companies to ask workers for overtime or weekend shifts)
- Warehouse Procedures:
  - Procedures (#)
  - Average Duration (days)
  - Associated Costs (% of per capita income)



Model and Required Conditions

- We allow for k independent variables to potentially be related to the dependent variable

Coefficients

Random error variable

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

Dependent variable

Independent variables

Estimating the Coefficients and Assessing the Model

- Data were collected from 177 different economies around the world, and ran for the following suggested model:

- Cost of Imp Exp = β 0 + β 1Procedures of S a B
  + β 2Duration of S a B + β 3Time for Imp/Exp
  + β 4Docum. For Imp/Exp
  + β 5Procedures for Warehouse + β 6Time for Warehouse + ε

	Import & Ex	Starting a Business				Warehouse
Region or Economy	Cost of Import/export	Procedures (number)	Duration (days)	Time of Import/Export	Documents For Import/Export	Procedures
Afghanista	0	4	9	0	0	13
Albania	0	10	36	0	0	24
Algeria	0	14	24	0	0	22
Angola	0	12	119	0	0	14
Antigua an	0	7	21	0	0	14

Regression 1

Regression Statistics	
Multiple R	0.824658607
R Square	0.680061818
Adjusted R Square	0.66281964
Standard Error	900.9560477
Observations	177

ANOVA

	df	SS	MS	F	Significance F
Regression	9	288141624	32015736	39.44176	6.25189E-37
Residual	167	135557541	811721.8		
Total	176	423699165			

	Coefficients	tandard Erro	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1247.939036	279.45735	4.4655795	1.465E-05	696.2145122	1799.6636	696.21451	1799.6636
Cost (% GNI per ca	-0.45084091	0.7022093	-0.642032	0.5217327	-1.837192318	0.9355105	-1.837192	0.9355105
Procedures (numb	41.07589053	24.285166	1.6913984	0.092626	-6.86960605	89.021387	-6.869606	89.021387
Duration (days)	-2.838923915	1.3093213	-2.168241	0.0315555	-5.423878857	-0.253969	-5.423879	-0.253969
Time of Import/Ex	36.29809487	2.1262475	17.071435	1.85E-38	32.10030628	40.495883	32.100306	40.495883
Documents For Im	-9.368555093	3.4668013	-2.702363	0.0075961	-16.21296009	-2.52415	-16.21296	-2.52415
Rigidity of Employ	2.568968788	4.0653476	0.6319186	0.5283042	-5.457128838	10.595066	-5.457129	10.595066
Procedures	-19.13227123	10.1168	-1.891139	0.0603373	-39.10557547	0.841033	-39.10558	0.841033
Duration	-1.161486685	0.5328005	-2.179965	0.0306598	-2.213379283	-0.109594	-2.213379	-0.109594
Cost (% income pe	0.014264421	0.0147651	0.9660872	0.3353975	-0.014885983	0.0434148	-0.014886	0.0434148

Regression 2

Regression Statistics	
Multiple R	0.823155087
R Square	0.677584296
Adjusted R Square	0.666204919
Standard Error	896.4218556
Observations	177

ANOVA

	df	SS	MS	F	Significance F
Regression	6	287091901	47848650	59.544934	2.85165E-39
Residual	170	136607264	803572.14		
Total	176	423699165			

	Coefficients	tandard Errc	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1277.660635	271.14377	4.7121151	5.076E-06	742.4183215	1812.9029	742.41832	1812.9029
Procedures (number)	43.75970297	23.145899	1.8906028	0.0603787	-1.930688068	89.450094	-1.930688	89.450094
Duration (days)	-3.005431263	1.2846638	-2.339469	0.0204733	-5.541378936	-0.469484	-5.541379	-0.469484
Time of Import/Export	36.17795181	2.0696817	17.479959	8.078E-40	32.09236563	40.263538	32.092366	40.263538
Documents For Import/Export	-8.969654161	3.4102482	-2.630206	0.0093167	-15.70154102	-2.237767	-15.70154	-2.237767
Procedures	-19.78149714	9.9807681	-1.981961	0.0490953	-39.48369972	-0.079295	-39.4837	-0.079295
Duration	-1.015825608	0.5134733	-1.978342	0.0495058	-2.029430503	-0.002221	-2.029431	-0.002221

Model Assessment

- The model is assessed using three tools:
  - The *standard error of estimate*
  - The *coefficient of determination*
  - The *F-test of the analysis of variance*
- The standard error of estimate participates in building the other tools.

Standard Error of Estimate

- From the printout, s_ε = 896.42
- Calculating the mean value of y we have 2641
- It seems s_ε is not particularly small.
- Question:

Can we conclude the model does not fit the data well?

$$s_\varepsilon = \sqrt{\frac{SSE}{n-k-1}} \qquad \hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k \qquad SSE = \sum (y - \hat{y})^2$$

Coefficient of Determination

- From the printout, R^2 = 0.6776
- 67.76% of the variation in cost of exp/imp is explained by the six independent variables. 32.24% remains unexplained.
- When adjusted for degrees of freedom, Adjusted R^2 = 1-[SSE/(n-k-1)] / [SS(Total)/(n-1)] =

= 66.62%

$$R^2 = 1 - \frac{SSE}{\sum (y_i - \bar{y})^2} = \frac{SSR}{\sum (y_i - \bar{y})^2} \qquad Adjusted\ R^2 = 1 - \frac{SSE/(n-k-1)}{\sum (y_i - \bar{y})^2/(n-1)}$$

Validity of the Model

- We pose the question:
  Is there at least one independent variable linearly related to the dependent variable?
- To answer the question we test the hypothesis

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$
H_1: At least one β_i is not equal to zero.

- If at least one β_i is not equal to zero, the model has some validity.

Testing the Validity of the Regression Model

[Variation in y] = SSR + SSE.

Large **F** results from a large SSR. Then, much of the variation in y is explained by the regression model; the model is useful, and thus, the null hypothesis should be rejected. Therefore, the rejection region is…

$$F = \frac{SSR/k}{SSE/n-k-1} = \frac{MSR}{MSE}$$

Rejection region

$F > F_{\alpha,k,n-k-1}$

	df	SS	MS	F	Significance F
Regression	6	287091900.6	47848650	59.54493	2.85165E-39
Residual	170	136607264.3	803572.1		
Total	176	423699165			

$F_{\alpha,k,n-k-1} = F_{0.05,6,176-6-1}=2.10$
$F = 59.54 > 2.10$

Also, the p-value (Significance F) = 2.885E-39
Reject the null hypothesis.

- **Conclusion:** There is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.
- At least one of the β_i is not equal to zero. Thus, at least one independent variable is linearly related to y.
- **This linear regression model is valid**

Interpreting the Coefficients

- **b_0 = 1277.66.** This is the intercept, the value of **y** when all the variables take the value zero. Since the data range of all the independent variables do not cover the value zero, do not interpret the intercept.

- **b_1 = 43.75.** In this model, for each additional procedure for S a B, the cost of Imp/exp increases on average by 43.75Usd (assuming the other variables are held constant).

Interpreting the Coefficients

- b_2 **= -3.00.** In this model, for each additional day spent for S a B, the cost of Imp/Exp increases on average by 3Usd when the other variables are held constant.

- b_3 **= 36.17.** For each additional day spent in registering property, the cost Imp/Exp will increase on average by 36.17Usd when the other variables are held constant.

Interpreting the Coefficients

- b_4 **= -8.96.** For each additional document for Imp/Exp the cost of Imp/Exp decreases on average by 8.96Usd when the other variables are held constant.

Interpreting the Coefficients

- **b_5 = -19.78.** For additional procedure for opening a warehouse, the cost of Imp/Exp decreases on average by 19.78Usd, when the other variables remain constant.

- **b_6 = -1.01.** For each additional day increase, the cost of Imp/Exp decreases on average by 1.01Usd when the other variables are held constant.

Testing the Coefficients

- The hypothesis for each β_i is

$$H_0: \beta_i = 0$$
$$H_1: \beta_i \neq 0$$

Test statistic

$$t = \frac{b_i - \beta_i}{s_{b_i}} \quad \text{d.f.} = n - k - 1$$

- Excel printout

	Coefficient	ndard Er	t Stat	P-value	ower 95%	pper 95%
Intercept	1277.661	271.1438	4.712115	5.08E-06	742.4183	1812.903
Procedures (number) of	43.7597	23.1459	1.890603	0.060379	-1.93069	89.45009
Duration (days) of S a E	-3.00543	1.284664	-2.33947	0.020473	-5.54138	-0.46948
Time of Import/Export	36.17795	2.069682	17.47996	8.08E-40	32.09237	40.26354
Documents For Import/I	-8.96965	3.410248	-2.63021	0.009317	-15.7015	-2.23777
Procedures for Wareho	-19.7815	9.980768	-1.98196	0.049095	-39.4837	-0.07929
Time of each procedure	-1.01583	0.513473	-1.97834	0.049506	-2.02943	-0.00222

Applications

- Predicted Values for New Observations
  - However, since all countries have been surveyed, we have to use it to predict how changes in variables can affect cost of import & export.
- Policy Decisions Regarding Lowering Costs to Consumers
  - Can be used by port managers or individual shipping yard managers

Predictions

- Predict the Cost of Imp/Exp in a country with the following characteristics:
  - Procedures (number) of S a B
  - Duration (days) of S a B
  - Time of Import/Export
  - Documents For Import/Export
  - Procedures for Warehouse
  - Time for Procedures for Warehouse

Cost of Imp/Exp = 1277.66 **+** 43.76**(1) –** 3.00**(1)** + 36.18**(1)** **-** 8.97**(1) –** 19.78**(1) –** 1.01**(1) =**

Predicted Values for New Observations

Region or Economy	Cost of Import/ export	Procedures (number)	Duration (days)	Time of Import/ Export	Documents For Import/ Export	Procedures	Duration
Japan	2036	8	23	21	11	15	177
Korea	1490	10	17	21	26	13	34
Taiwan	1494	8	48	25	12	30	198
Hong Kong, China	1050	5	11	11	41	23	155
Singapore	783	5	5	8	13	11	102
Indonesia	1290	12	105	48	81	19	196
Malaysia	817	9	24	32	42	25	285
Thailand	1401	8	33	31	39	11	156
Philippines	1600	15	58	35	11	21	177
Vietnam	1550	11	50	47	14	13	194

30% time reduction

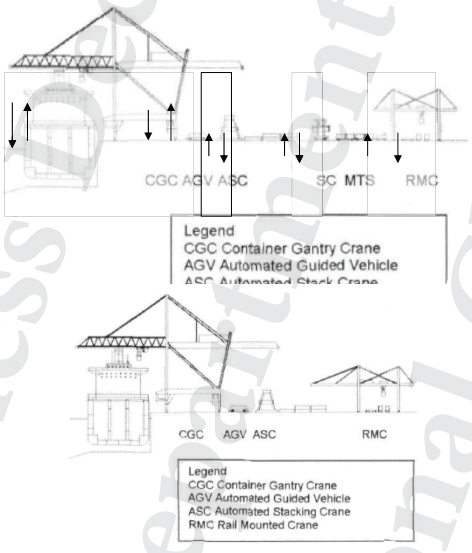


TABLE I Process times of conventional inter-terminal transport of a train with 50 containers

Phase	Activity	Distance (m)	Time/box (min)	Time/train (min)
Train arrival	de-coupling	–	–	2
	departure el. loco	1000	–	3
Shunting	travel shunting loco	1000	–	5
	walking driver	700	–	3
	shunting train set	2000	–	10
	de-coupling	–	–	2
	departure shunt. loco	1000	–	3
Unloading	travel MTS at RMC	1000	–	50
	travel RMC	100	1	50
	transhipment to MTS	10	1	50
Intra-terminal transport	travel MTS to stack	1000	–	50
	transhipment by SC	–	1	50
	transhipment by SC	–	1	50
	travel MTS to RMC	1000	–	50
Loading	travel RMC	100	1	50
	transhipment to train	10	1	50
Shunting	travel shunting loco	1000	–	5
	coupling	–	–	4
	walking driver	700 -	–	3
	shunting train set	2000	–	10
	de-coupling	–	–	2
	departure shunting loco	1000	–	3
Train departure	travel el. loco	3000	–	10
	coupling	–	–	5
	inspection, brake test	–	–	60
	waiting for time slot	–	–	10
Total			12	590

TABLE
Phase
Train an
Shunting
Unloadir
Loading
Shunting
Train departure
travel el. loco 3000 – 10
coupling el. loco – – 5
inspection, brake test – – 60
waiting for time slot – – 10
Total 8 400

Predicted Values for New Observations

- *If time of importing & exporting is reduced by 30% in Indonesia:*

- CostImEx=1277+43.76(12)-3.00(105)+36.18(48)-8.97(81)-19.78(19)-1.01(196) = $1924.07
- to this:
- CostImEx=1277+43.76(12)-3.00(105)+36.18(34)-8.97(81)-19.78(19)-1.01(196) =$1415.76

- there is a 27% decrease in costs
- Prediction Interval….*±20% ???*

Prediction Interval	
	Cost of Import/ex
Predicted value	1415.7649
Prediction Interval	
Lower limit	-395.46034
Upper limit	3226.99014
Interval Estimate of Expected Value	
Lower limit	1029.47006
Upper limit	1802.05974

Policy Decisions Regarding Lowering Costs to Consumers

- Port of Kaohsiung (currently 6th worldwide)
  - Facing loss of shipping
  - Port managers could use information to focus efforts to lower cost to consumers
  - invest in reducing time
  - (example: an automated shunting system which reduces time by 30%)

- Thank you for your attention!

Appendix 1: Stepwise Regression of Asian Economies

- Can we use the model to forecast a smaller group of economies?

Stepwise Regression of Asian Economies

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.931169								
R Square	0.867075								
Adjusted R Squ	0.810107								
Standard Error	907.0334								
Observations	21								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	6	75131868	12521978	15.22041	2.06192E-05				
Residual	14	11517935	822709.7						
Total	20	86649803							
	Coefficients	andard Erro	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%	
Intercept	1688.446	1037.215	1.627865	0.125841	-536.1585898	3913.052	-536.159	3913.052	
Procedures (nu	32.69684	96.56374	0.338604	0.739933	-174.4117763	239.8055	-174.412	239.8055	
Duration (days)	-16.3865	8.059602	-2.03317	0.061448	-33.67262866	0.899626	-33.6726	0.899626	
Time of Import	50.8553	7.889633	6.445839	1.53E-05	33.93372473	67.77688	33.93372	67.77688	
Documents For	-90.9257	62.64975	-1.45133	0.168725	-225.2960879	43.44462	-225.296	43.44462	
Procedures	4.628371	38.02693	0.121713	0.904856	-76.93127425	86.18802	-76.9313	86.18802	
Duration	-0.86315	1.720575	-0.50166	0.6237	-4.553419225	2.827115	-4.55342	2.827115	

Stepwise Regression of Asian Economies

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.930164							
R Square	0.865206							
Adjusted R Squa	0.831507							
Standard Error	854.3971							
Observations	21							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	4	74969893	18742473	25.67482	8.63357E-07			
Residual	16	11679910	729994.4					
Total	20	86649803						
	Coefficients	andard Erro	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	1939.282	767.4486	2.526921	0.022424	312.3638742	3566.201	312.3639	3566.201
Duration (days)	-15.1854	7.12076	-2.13256	0.048801	-30.28076295	-0.09009	-30.2808	-0.09009
Time of Import/	51.4154	6.07728	8.460266	2.66E-07	38.53214649	64.29866	38.53215	64.29866
Documents For	-88.2957	58.71331	-1.50384	0.152103	-212.7623454	36.17095	-212.762	36.17095
Duration	-0.86418	1.498965	-0.57652	0.572287	-4.041842685	2.313484	-4.04184	2.313484

Stepwise Regression of Asian Economies

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.928657881							
R Square	0.86240546							
Adjusted R Sc	0.838124071							
Standard Erro	837.4520099							
Observations	21							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	3	74727263	24909088	35.51714	1.53171E-07			
Residual	17	11922540	701325.9					
Total	20	86649803						
	Coefficients	tandard Erro	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	1825.565072	726.958	2.511239	0.022426	291.817711	3359.312	291.8177	3359.312
Duration (day	-16.38927807	6.672692	-2.45617	0.025096	-30.46742655	-2.31113	-30.4674	-2.31113
Time of Impo	50.89924135	5.891755	8.639063	1.26E-07	38.46872503	63.32976	38.46873	63.32976
Documents Fo	-88.18497308	57.54855	-1.53236	0.143832	-209.6018	33.23185	-209.602	33.23185

Stepwise Regression of Asian Economies

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.918368254							
R Square	0.84340025							
Adjusted R Squa	0.826000278							
Standard Error	868.2465562							
Observations	21							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	2	73080465	36540233	48.47136	5.66395E-08			
Residual	18	13569337	753852.1					
Total	20	86649803						
	Coefficients	tandard Erro	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	922.8988451	441.6506	2.089658	0.051115	-4.974624712	1850.772	-4.97462	1850.772
Duration (days)	-18.95114265	6.697397	-2.82963	0.011108	-33.02185081	-4.88043	-33.0219	-4.88043
Time of Import/	45.02921689	4.641034	9.702411	1.42E-08	35.27876632	54.77967	35.27877	54.77967

Appendix 2: Graphs