How to choose a city of residence?

Hyoin SEOL
Rana Shamail NASARULLA
Yashraj VASISHTHA

Ecole Nationale de l'Aviation Civile

January 19, 2024

Presentation Overview

Jack's Dilemma

Who is Jack

A married, corporate worker

Where does he live?

Currently, lives in Toulouse, France

What is his dilemma

Jack's company has offered him to move to a metropolitan city. He has to chose from New York, Vancouver, Brussels and Copenhagen

Objective

Goals

- To explore factors that impact an individual's decision in selecting a city for residency.
- Suggest Jack a place to choose from the list, using statistical data.

Data and Hypothesis

What does data set contain?

Details related to economy, lifestyle, challenges and population migration data of major cities in the world

How many cities are considered?

20

Which year does the data refer to?

2019

Hypothesis

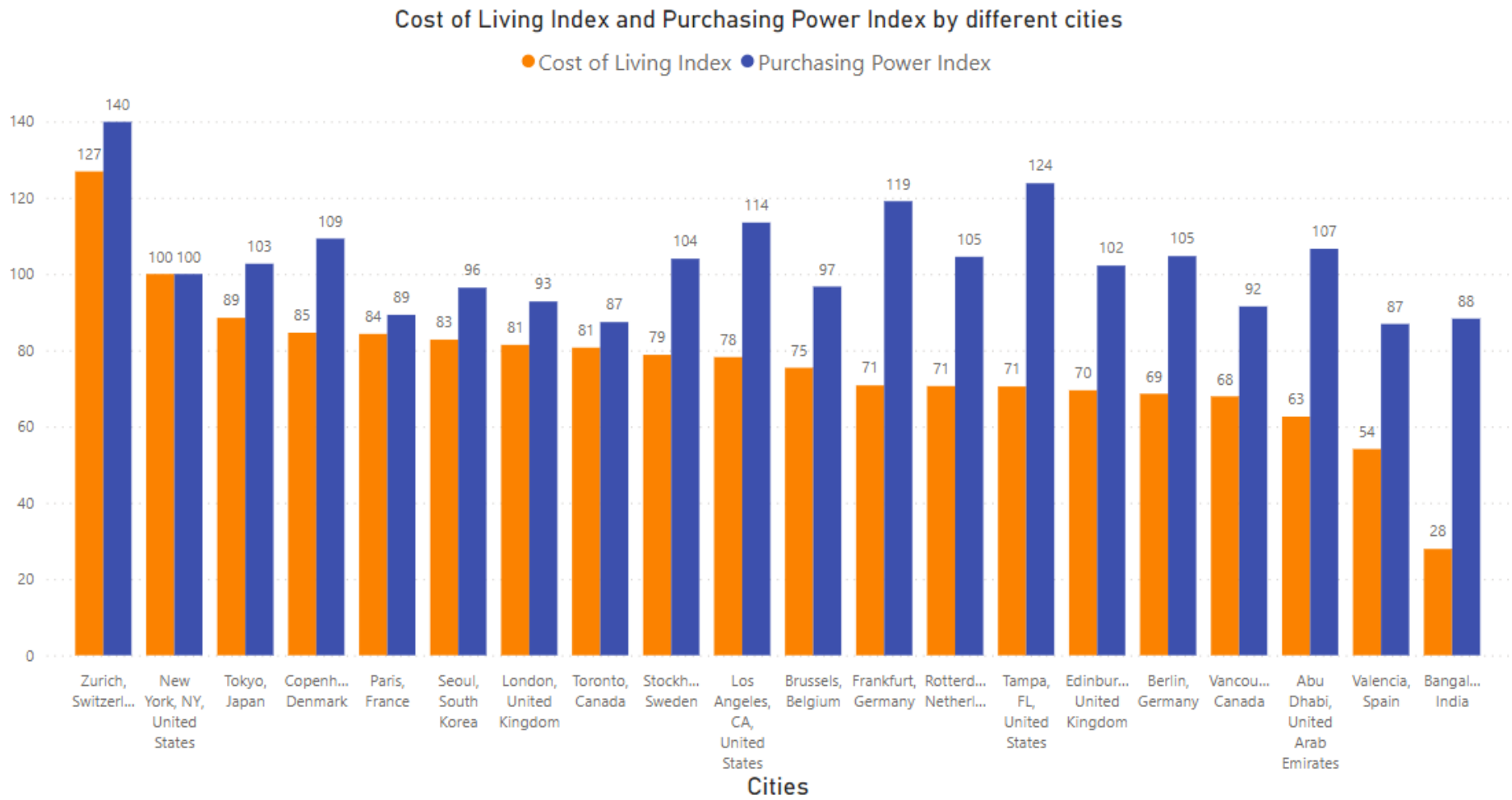
The indices used in variables are correct and represent real world scenario.

Variable Set

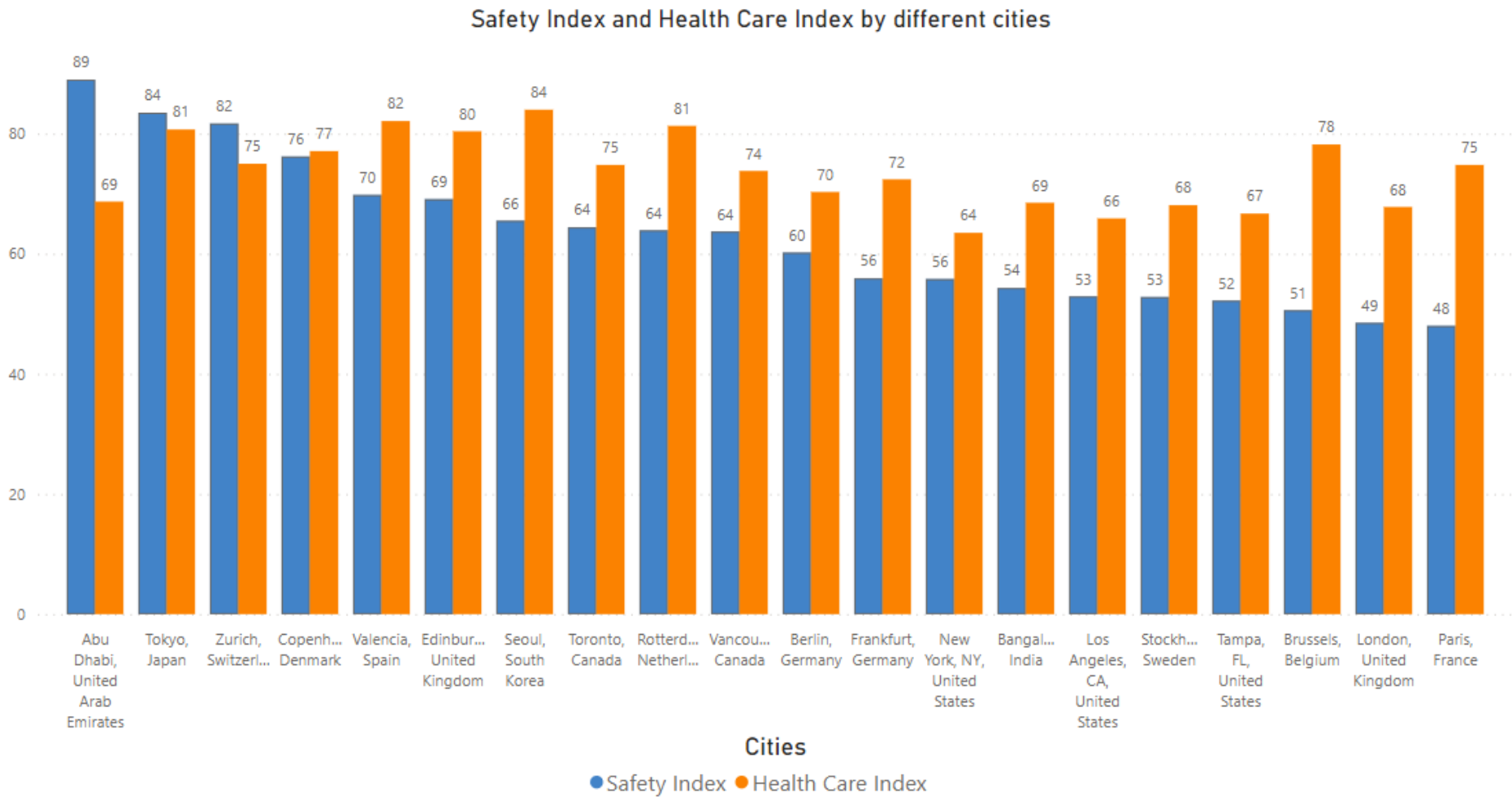
Variable	Explanatory Variables
P	Total population influx
PPI	Purchasing Power Index
SI	Safety Index
H	Health Care Index
L	Cost of Living Index
PR	Property Price to Income Ratio
T	Traffic Commute Time Index
PLI	Pollution Index
CI	Climate Index

Table: Detailed variables explanation

Variable set comparison



Variable set comparison



Linear equation

- $$\log(P) = C + \beta1.\log(PPI) + \beta2.\log(\mathrm{SI}) + \beta3.\log(\mathrm{H}) + \beta4.\log(\mathrm{L}) + \beta5.\log(\mathrm{PR}) + \beta6.\log(\mathrm{T}) + \beta7.\log(PLI) + \beta8.\log(CI)$$

where,

- C is a constant
- $\beta1$, $\beta2$, ..., $\beta8$ are coefficients of respective variables

First Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.58267	10.80260	2.275625	0.0404
LOG(H)	-5.175803	2.542549	-2.035675	0.0627
LOG(S)	1.195448	1.344707	0.889002	0.3902
LOG(PLI)	1.379231	0.431455	3.196695	0.0070
LOG(L)	-0.498625	0.681768	-0.731370	0.4775
LOG(PR)	0.861916	0.403515	2.136018	0.0523
LOG(CI)	-0.218218	1.174233	-0.185839	0.8554

R-squared	0.658446	Mean dependent var	11.23570
Adjusted R-squared	0.500806	S.D. dependent var	1.047677
S.E. of regression	0.740223	Akaike info criterion	2.505485
Sum squared resid	7.123083	Schwarz criterion	2.853992
Log likelihood	-18.05485	Hannan-Quinn criter.	2.573517
F-statistic	4.176888	Durbin-Watson stat	1.335006
Prob(F-statistic)	0.014705		

Figure: Results from First model test

First Model

Result interpretation

- R squared measures how close the points are to the estimated regression line in the scatter plot.
- Closer the value is to 1, better is the model.
- Result highly depends on inverse of healthcare index.

Second regression required.

Second Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.681774	14.58422	-0.663853	0.5205
LOG(PPI)	5.672823	1.862445	3.045901	0.0111
LOG(H)	-1.763459	2.443109	-0.721809	0.4855
LOG(S)	0.471332	1.096454	0.429869	0.6756
LOG(CI)	-1.053550	0.988133	-1.066203	0.3092
LOG(PLI)	1.640805	0.382148	4.293632	0.0013
LOG(PR)	1.453190	0.442566	3.283559	0.0073
LOG(T)	0.804510	0.986631	0.815411	0.4321
LOG(L)	-1.709532	0.672078	-2.543651	0.0273

R-squared	0.817761	Mean dependent var	11.23570
Adjusted R-squared	0.685223	S.D. dependent var	1.047677
S.E. of regression	0.587799	Akaike info criterion	2.077301
Sum squared resid	3.800589	Schwarz criterion	2.525380
Log likelihood	-11.77301	Hannan-Quinn criter.	2.164771
F-statistic	6.170020	Durbin-Watson stat	1.252219
Prob(F-statistic)	0.003594		

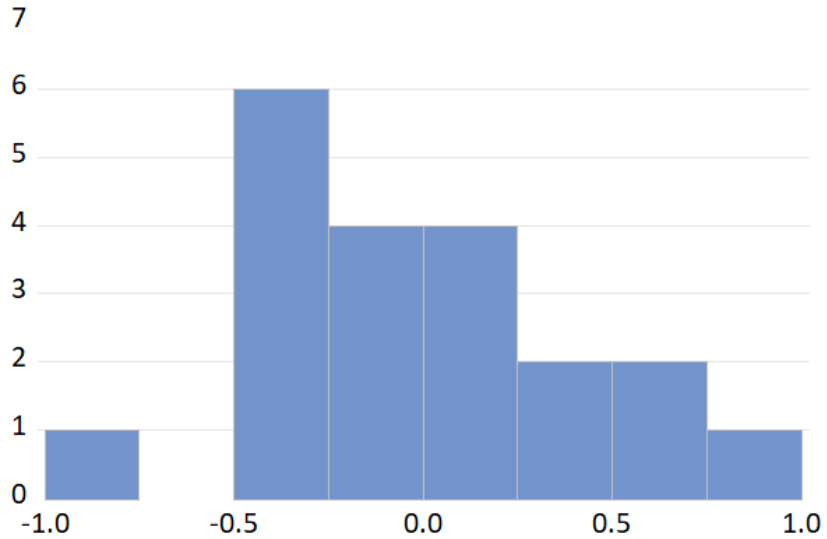
Figure: Results from second model test

Second Model

Result interpretation

- R squared received is 0.818
- Dependency of total population influx is reasonably distributed among variables.

Normality Test



Series: Residuals	
Sample 1 20	
Observations 20	
Mean	-1.48e-15
Median	-0.081181
Maximum	0.785791
Minimum	-0.961692
Std. Dev.	0.447248
Skewness	0.116098
Kurtosis	2.747644
Jarque-Bera	0.097999
Probability	0.952182

Figure: Results from normality test

Heteroskedasticity Test

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	1.429017	Prob. F(8,11)	0.2851
Obs*R-squared	10.19264	Prob. Chi-Square(8)	0.2518
Scaled explained SS	2.694232	Prob. Chi-Square(8)	0.9521

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 01/18/24 Time: 22:50
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.29144	5.885469	1.748618	0.1082
LOG(PPI)	-1.541837	0.751591	-2.051432	0.0648
LOG(H)	-0.861907	0.985918	-0.874218	0.4007
LOG(S)	0.029834	0.442475	0.067425	0.9475
LOG(CI)	0.658747	0.398762	1.651981	0.1268
LOG(PLI)	0.005198	0.154216	0.033703	0.9737
LOG(PR)	-0.098622	0.178598	-0.552204	0.5919
LOG(T)	-0.909818	0.398155	-2.285082	0.0431
LOG(L)	0.276796	0.271218	1.020567	0.3294

R-squared	0.509632	Mean dependent var	0.190029
Adjusted R-squared	0.153001	S.D. dependent var	0.257742
S.E. of regression	0.237207	Akaike info criterion	0.262394
Sum squared resid	0.618938	Schwarz criterion	0.710473
Log likelihood	6.376062	Hannan-Quinn criter.	0.349864
F-statistic	1.429017	Durbin-Watson stat	2.824342
Prob(F-statistic)	0.285130		

Figure: Results from heteroskedasticity test

Residual

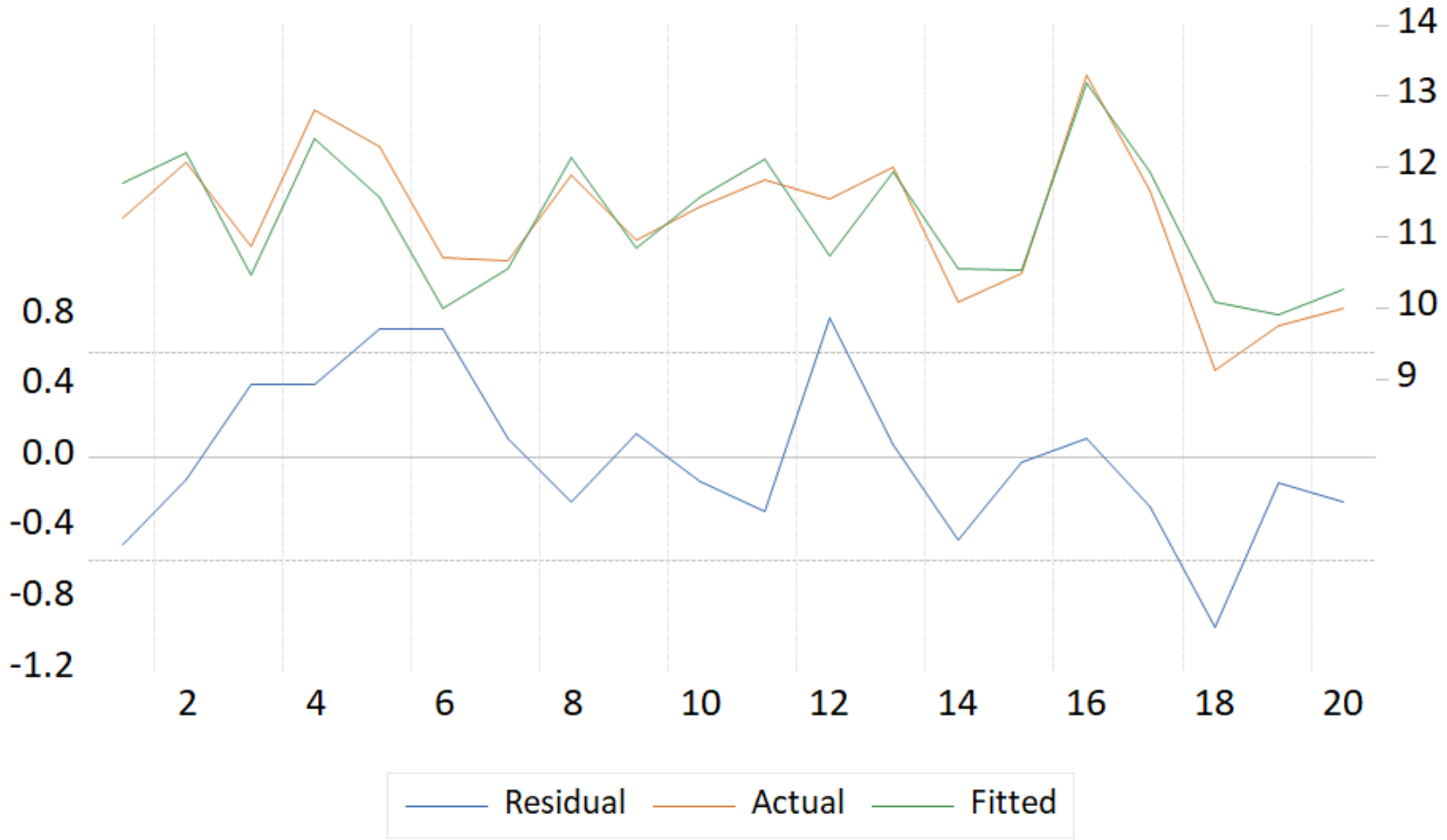


Figure: Residual graph

Conclusion

Result interpretation

- Economic factors play major role in population migration
- Other factors like employment rate, inflation, technological advancements, government policies etc. are also considered while deciding residence city
- Deciding factors can be highly personal and can sometimes go against statistics

Suggestion for Jack

- On statistical foundation, Jack is advised to move to Copenhagen. However, he might have his personal preferences.

Thank you for listening

Questions and comments are welcomed!