

## Tourism and Trade Study

---

### Technical Appendix

Deloitte Analytics – Yunfan Li, Aditya Sane, Tom Peters

### Background

The purported importance of travel on commerce and business can be traced for many centuries. In recent history, [1] and [2] have shown a linkage between business travel and exports from an economy. This paper studies the problem of causal linkage between visitors to Canada and changes in foreign trade.

### Materials

This study uses the following data sources

- Quantity and value of exports from Canada to each trade partner country for each month classified into HS06 commodity codes from calendar year 2003 to 2012. This data was sourced from Statistics Canada.
- Quantity and value of imports to Canada from each trade partner country for each month classified into HS06 commodity codes from calendar year 2003 to 2012. This data was sourced from Statistics Canada.
- Number of foreign nationals entering Canada for each calendar quarter. The visitors' intent was recorded and classified into *Business, Pleasure, Visit Friends and Family, and Other*. This data was source from the Canadian Border Services Agency. The number of Canadian citizens entering Canada was not available.
- Gross Domestic Product (GDP) for every country from 2003 to 2012. This data was sourced from the World Bank.
- List of Free Trade Agreements (FTA) between Canada and trading partner countries and the date of effect of the agreement. Data was collated between 2003 and 2012.
- Great circle distance between centroid of Canada and centroid of the trading partner country.
- List of trading partner countries that are land-locked.
- List of trading partner countries that were former colonies or are present colonies of the United Kingdom.
- List of trading partner countries where the primary language spoke in English.

### Methods

[1] suggests using a *Gravitational* model to define the link between trade volume, GDP, and distance. Equation 1 shows the gravitational model from we begin our analysis.

Equation 1 Gravitational Model

$$trade_{i,j,t} \propto \frac{GDP_{i,t}GDP_{j,t}}{d_{i,j}}$$

Where  $(i,j)$  are the index countries,  $t$  is time, and  $d_{i,j}$  is the distance between the countries.

Taking a logarithm of Equation 1 and transforming we get Equation 2.

## Equation 2 Log Transformed Gravitational Model

$$\log(\text{trade}_{i,j,t}) = \log(\text{GDP}_{i,t}) + \log(\text{GDP}_{j,t}) - \log(d_{i,j}) + k_{i,j}$$

Where  $k_{i,j}$  is a constant.

As we are using cross-country time series data, there are many factors or conditions that would also have an influence on Trade variation independent of GDP flow. As such, we propose to generalize Equation 2 into a multiple regression model as specified in Equation 3.

## Equation 3 Regression Model

$$\log(\text{trade}_t) = \beta_1 \log(\text{GDP}_t) + \beta_2 \log(\text{GDP}_{t-1}) + \beta_3 \log(\text{GDP}_{CA,t}) + \beta_4 \log(\text{GDP}_{CA,t-1}) + \beta_5 \log(d) + \beta_6 \text{FTA}_t + \beta_7 \text{ENG} + \beta_8 \text{LL} + \beta_9 \text{UKC} + \beta_{10} \log(\text{trips}_t) + \beta_{11} \log(\text{trips}_{t-1})$$

All  $\beta$  values are regression coefficients.

Other variables in Equation 3 are

- $\text{GDP}_i$ : GDP of the trade partner country in a specified year  $t$
- $\text{GDP}_{CA,t}$ : GDP of Canada in a specified year  $t$
- $\text{FTA}_i$ : Indicator of free trade agreement between Canada and trading partner in specified year  $t$
- $\text{ENG}$ : Indicator variable designating if trade partner country's primary language is English
- $\text{LL}$ : Indicator variable designating if the trade partner country is land locked
- $\text{UKC}$ : Indicator variable designating if the trade partner was a former colony or present colony of the United Kingdom
- $\text{trips}_i$ : Number of visitors from the trade partner country to Canada in year  $t$

We will consider each year for every trade partner country to be an individual observation for the estimating the regression coefficients.

[2] uses *Granger causality* to postulate a "causal" link between visits and trade. A full Granger analysis uses multiple lags of the time to determine causality [3]. In Equation 3, we are employing a one year lag on GDP and number of visits. Hence, one year causality can be determined through analysis of coefficients  $\beta_{10}$  and  $\beta_{11}$ .

We build four models for the variable trade

1. Value of exports from Canada
2. Value of imports to Canada
3. Variety of exports from Canada (using HS06 code)
4. Variety of imports to Canada (using HS06 code)

## Data Description

Table 1 summarizes time invariant data at a high level.

Table 1 Data Summary

Descriptor	Value
Number of countries	225
Number of years	12
Number of valid observations for regression	1025
Countries with English as primary language	54
Landlocked countries	36
Countries that are current or former colonies of the United Kingdom	62

Figure 1 shows the number of countries that have signed free trade agreements with Canada over time.

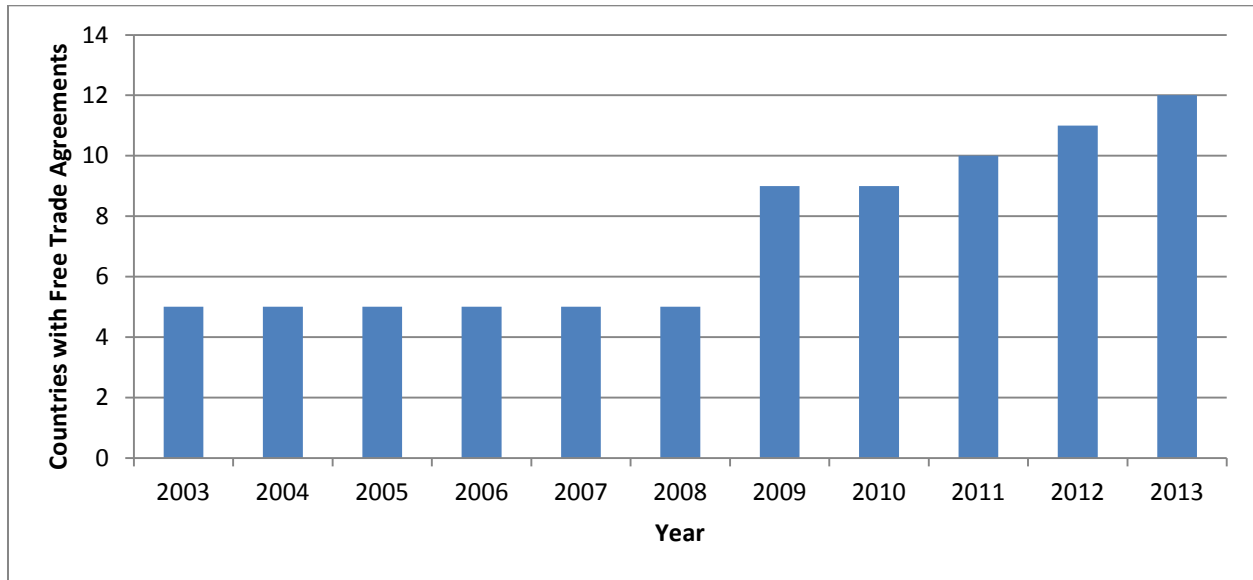


Figure 1 Free Trade Agreements

The GDP for Canada over time is shown in Figure 2.

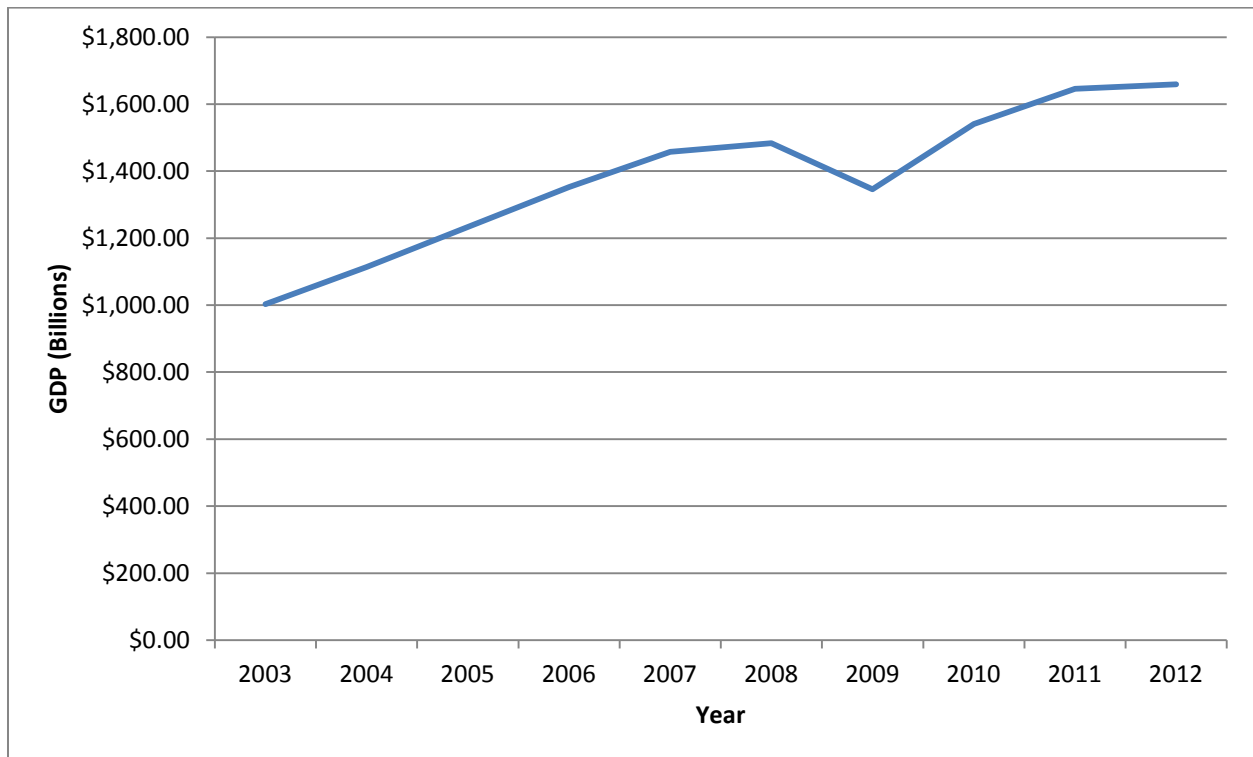


Figure 2 GDP of Canada

A histogram of the logarithm of GDP values of trading partner countries (for all years) is shown in Figure 3.

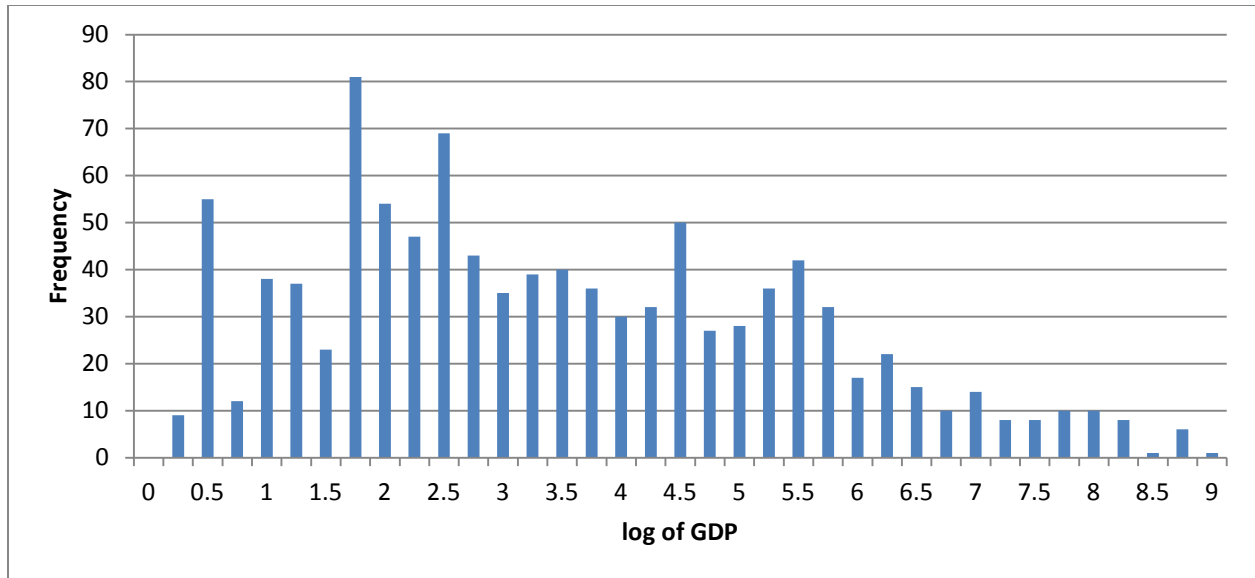


Figure 3 GDP Spread of Trading Partner Countries

The total number of visitors to Canada is shown in Figure 4.

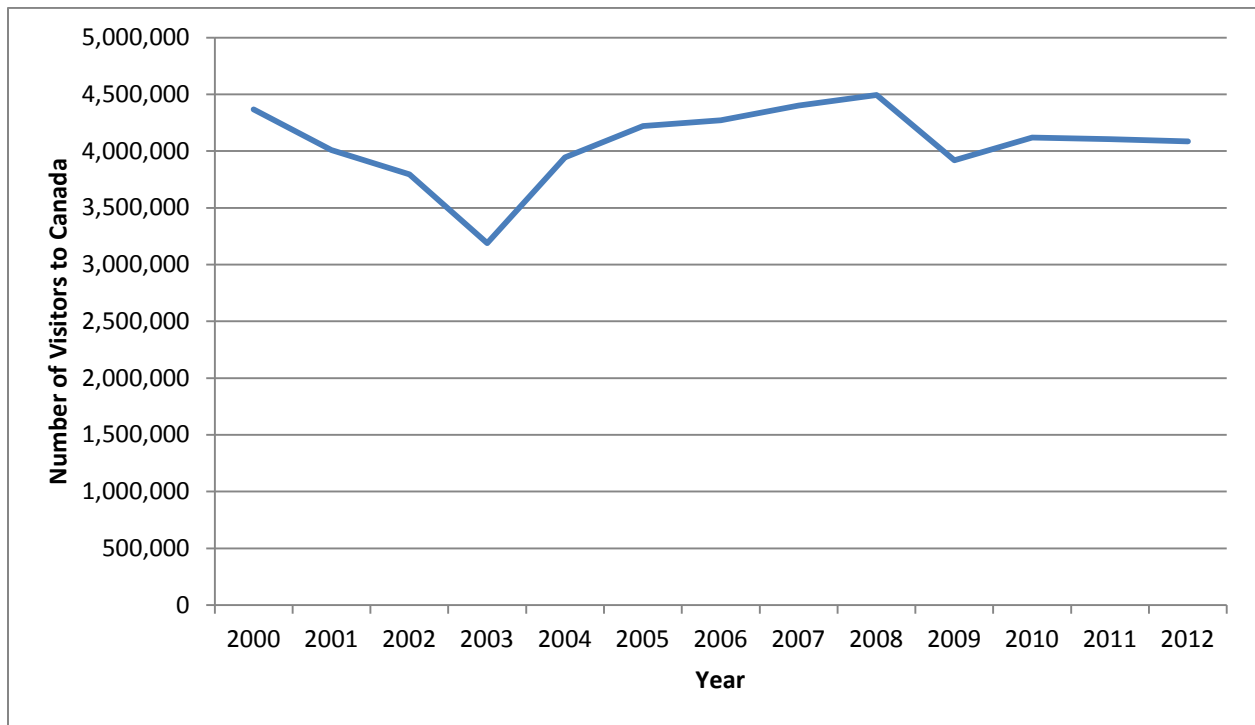


Figure 4 Number of Visitors to Canada

## Regression Modeling

Equation 3 was fit using IBM SPSS Statistics Version 21. The significance of variables in the model was analyzed through p-values. Variables with weak significance at 95% confidence ( $p\text{-value} > 0.05$ ) were removed from the model and regression re-calculated.



## Value of Exports from Canada

Assuming all regression coefficients in Equation 3 are non-zero, the model we calculated is shown in Table 2.

Table 2 Value of Exports Full Model

Model	R	R Square	Adj. R Square	Std. Error	Durbin-Watson
<b>Full Export Value</b>	0.905	0.819	0.817	0.94891	0.536

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		$\beta$	Std. Error	$\beta$			Tolerance	VIF
<b>Full Export Value</b>	(Constant)	7.800	1.969		3.962	.000		
	logGDP	1.422	.351	1.276	4.054	.000	.002	555.593
	logprevGDP	-.703	.352	-.629	-1.997	.046	.002	556.234
	logGDPCA	1.055	.538	.057	1.961	.050	.213	4.684
	logprevGDPCA	.537	.442	.035	1.216	.224	.215	4.645
	logDist	-.532	.082	-.100	-6.466	.000	.748	1.337
	FTA	.434	.147	.041	2.959	.003	.915	1.093
	EngSpeaker	-.141	.084	-.027	-1.677	.094	.675	1.482
	LandLocked	-.478	.078	-.092	-6.154	.000	.804	1.243
	GBRColonized	.324	.086	.058	3.752	.000	.754	1.327
	logTrips	.066	.017	.102	3.935	.000	.265	3.773
	logprevTrips	.090	.017	.138	5.218	.000	.255	3.923

Collinearity tests suggest that variables  $GDP_t$  and  $GDP_{t-1}$  are collinear.  $GDP_{CA,t}$  and  $GDP_{CA,t-1}$  are possibly collinear as well. Removing non-significant regression coefficients ( $p$ -value  $> 0.05$ ), a reduced regression model was derived which is shown in Table 3. Variables removed in Table 2 have shaded rows.

Table 3 Value of Exports Reduced Model

Model	R	R Square	Adj. R Square	Std. Error	Durbin-Watson
<b>Reduced Export Value</b>	0.903	0.816	0.814	0.95686	0.532

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		$\beta$	Std. Error	$\beta$			Tolerance	VIF
<b>Reduced Export Value</b>	(Constant)	7.655	1.946		3.934	.000		
	logGDP	.726	.025	.651	29.324	.000	.367	2.722
	logGDPCA	1.573	.251	.085	6.259	.000	.993	1.007
	logDist	-.500	.079	-.094	-6.369	.000	.834	1.199
	FTA	.412	.147	.039	2.807	.005	.926	1.079

LandLocked	-.510	.078	-.098	-6.568	.000	.816	1.225
logTrips	.069	.017	.105	4.065	.000	.270	3.704
logprevTrips	.085	.017	.131	4.975	.000	.263	3.801

The reduced model has almost the same fit characteristics as the full model with all variables being significant and stable.

## Value of Imports to Canada

Fully specifying all coefficients in Equation 3, we define a model shown in Table 4.

Table 4 Value of Imports Full Model

Model	R	R Square	Adj. R Square	Std. Error	Durbin-Watson
<b>Full Import Value</b>	0.836	0.699	0.695	0.33843	0.594

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		$\beta$	Std. Error	$\beta$			Tolerance	VIF
<b>Full Import Value</b>	(Constant)	4.020	.702		5.726	.000		
	logGDP	.454	.125	1.475	3.629	.000	.002	555.593
	logprevGDP	-.349	.126	-1.131	-2.782	.006	.002	556.234
	logGDPCA	.425	.192	.083	2.213	.027	.213	4.684
	logprevGDPCA	-.061	.158	-.014	-.389	.697	.215	4.645
	logDist	-.384	.029	-.261	-13.074	.000	.748	1.337
	FTA	.019	.052	.007	.363	.717	.915	1.093
	EngSpeaker	-.169	.030	-.118	-5.638	.000	.675	1.482
	LandLocked	-.243	.028	-.169	-8.784	.000	.804	1.243
	GBRColonized	.301	.031	.194	9.787	.000	.754	1.327
	logTrips	.027	.006	.150	4.466	.000	.265	3.773
	logprevTrips	.035	.006	.197	5.754	.000	.255	3.923

Collinearity tests suggest that variables  $GDP_t$  and  $GDP_{t-1}$  are collinear.  $GDP_{CA,t}$ ,  $GDP_{CA,t-1}$ ,  $trips_t$  and  $trips_{t-1}$  are possibly collinear as well. We removed variables that are non-significant (p-value > 0.05) through an iterative model building process. The final reduced model derived is shown in Table 5. Variables removed in Table 4 have shaded rows.

Table 5 Value of Imports Reduced Model

Model	R	R Square	Adj. R Square	Std. Error	Durbin-Watson
<b>Reduced Import Value</b>	0.816	0.666	0.664	1.87082	0.477

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		$\beta$	Std.	$\beta$			Tolerance	VIF

		Error					
<b>Reduced Import Value</b>	(Constant)	14.392	3.803		3.784	.000	
	logGDP	1.117	.045	.689	24.559	.000	.416 2.401
	logGDPCA	1.231	.491	.046	2.506	.012	.994 1.006
	logDist	-1.100	.152	-.142	-7.217	.000	.848 1.180
	FTA	1.176	.287	.077	4.096	.000	.928 1.078
	LandLocked	-.398	.151	-.053	-2.632	.009	.824 1.213
	logprevTrips	.083	.028	.088	3.001	.003	.383 2.610

The reduced model has almost the same fit characteristics as the full model with all variables being significant and stable. It is interesting that the number of visitors in the current year is a non-significant contributor to import values.

## Variety of Exports from Canada

Fully specifying all coefficients in Equation 3, we define a model shown in Table 6.

**Table 6 Variety of Exports Full Model**

Model	R	R Square	Adj. R Square	Std. Error	Durbin-Watson
<b>Full Export Variety</b>	0.836	0.669	0.695	0.33843	0.594

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		$\beta$	Std. Error	$\beta$			Tolerance	VIF
<b>Full Export Variety</b>	(Constant)	4.020	.702		5.726	.000		
	logGDP	.454	.125	1.475	3.629	.000	.002	555.593
	logprevGDP	-.349	.126	-1.131	-2.782	.006	.002	556.234
	logGDPCA	.425	.192	.083	2.213	.027	.213	4.684
	logprevGDPCA	-.061	.158	-.014	-.389	.697	.215	4.645
	logDist	-.384	.029	-.261	-13.074	.000	.748	1.337
	FTA	.019	.052	.007	.363	.717	.915	1.093
	EngSpeaker	-.169	.030	-.118	-5.638	.000	.675	1.482
	LandLocked	-.243	.028	-.169	-8.784	.000	.804	1.243
	GBRColonized	.301	.031	.194	9.787	.000	.754	1.327
	logTrips	.027	.006	.150	4.466	.000	.265	3.773
logprevTrips	.035	.006	.197	5.754	.000	.255	3.923	

Collinearity tests suggest that variables  $GDP_t$  and  $GDP_{t-1}$  are collinear.  $GDP_{CA,t}$ ,  $GDP_{CA,t-1}$ ,  $trips_t$  and  $trips_{t-1}$  are possibly collinear as well. We removed variables that are non-significant (p-value > 0.05) through an iterative model building process. The final reduced model derived is shown in Table 7. Variables removed in Table 6 have shaded rows.

**Table 7 Variety of Exports Reduced Model**

Model	R	R Square	Adj. R Square	Std. Error	Durbin-Watson
-------	---	----------	---------------	------------	---------------

Reduced Export Variety		0.816	0.666	0.664	0.35566	0.551		
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		$\beta$	Std. Error	$\beta$			Tolerance	VIF
Reduced Export Variety	(Constant)	4.164	.723		5.757	.000		
	logGDP	.113	.009	.368	12.304	.000	.368	2.721
	logGDPCA	.334	.093	.065	3.580	.000	.998	1.002
	logDist	-.372	.029	-.253	-12.913	.000	.855	1.170
	LandLocked	-.274	.029	-.190	-9.491	.000	.817	1.224
	logTrips	.027	.006	.151	4.336	.000	.270	3.699
	logprevTrips	.031	.006	.172	4.888	.000	.264	3.789

### Variety of Imports to Canada

Fully specifying all coefficients in Equation 3, we define a model shown in Table 8.

Table 8 Variety of Imports Full Model

Model	R	R Square	Adj. R Square	Std. Error	Durbin-Watson				
Full Import Variety	0.798	0.637	0.633	0.39066	0.456				
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics		
		$\beta$	Std. Error	$\beta$			Tolerance	VIF	
Full Import Variety	(Constant)	3.488	.811		4.304	.000			
	logGDP	-.078	.144	-.240	-.537	.591	.002	555.593	
	logprevGDP	.213	.145	.656	1.468	.142	.002	556.234	
	logGDPCA	.056	.221	.010	.255	.799	.213	4.684	
	logprevGDPCA	.119	.182	.027	.655	.513	.215	4.645	
	logDist	-.207	.034	-.134	-6.103	.000	.748	1.337	
	FTA	.068	.060	.022	1.134	.257	.915	1.093	
	EngSpeaker	-.088	.035	-.059	-2.549	.011	.675	1.482	
	LandLocked	.013	.032	.009	.409	.683	.804	1.243	
	GBRColonized	.111	.036	.068	3.112	.002	.754	1.327	
	logTrips	.038	.007	.200	5.447	.000	.265	3.773	
	logprevTrips	.039	.007	.206	5.482	.000	.255	3.923	

Collinearity tests suggest that variables  $GDP_t$  and  $GDP_{t-1}$  are collinear.  $GDP_{CA,t}$ ,  $GDP_{CA,t-1}$ ,  $trips_t$ , and  $trips_{t-1}$  are possibly collinear as well. We removed variables that are non-significant (p-value > 0.05) through an



iterative model building process. The final reduced model derived is shown in Table 9. Variables removed in Table 8 have shaded rows.

**Table 9 Variety of Exports Reduced Model**

Model	R	R Square	Adj. R Square	Std. Error	Durbin-Watson
<b>Reduced Import Variety</b>	0.795	0.632	0.630	0.39226	0.454

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		$\beta$	Std. Error	$\beta$			Tolerance	VIF
<b>Reduced Import Variety</b>	(Constant)	3.344	.797		4.196	.000		
	logGDP	.139	.010	.429	13.690	.000	.368	2.716
	logGDPCA	.214	.103	.040	2.086	.037	.998	1.002
	logDist	-.222	.031	-.144	-7.089	.000	.880	1.137
	logTrips	.037	.007	.196	5.390	.000	.273	3.663
	logprevTrips	.038	.007	.201	5.463	.000	.266	3.761

## Discussion

A summary of the coefficients for all models is shown in Table 10.

**Table 10 Summary of Model Coefficients**

Variable	Value of Exports	Value of Imports	Variety of Exports	Variety of Imports
<i>Constant</i>	7.655	14.392	4.164	3.344
<i>GDP<sub>t</sub></i>	0.726	1.117	0.113	0.139
<i>GDP<sub>t-1</sub></i>				
<i>GDP<sub>CA,t</sub></i>	1.573	1.231	0.334	0.214
<i>GDP<sub>CA,t-1</sub></i>				
<i>d<sub>t</sub></i>	-0.510	-1.100	-0.372	-0.222
<i>FTA<sub>t</sub></i>	0.412	1.176		
<i>ENG</i>				
<i>LL</i>	-0.510	-0.398	-0.274	
<i>UKC</i>				
<i>trips<sub>t</sub></i>	0.069		0.027	0.037
<i>trips<sub>t-1</sub></i>	0.085	0.083	0.031	0.038

Cells corresponding to unused variables for each model are shaded.

## Effect of GDP

We observe a positive increase in trade, measured through both value and variety, with increased GDP of Canada and trading partner country. This observation can be explained with increased economic activity of both trading countries. Previous year's GDP is collinear with the current year's GDP and was removed from the regression models.

For all the models, GDP of trading partner country and GDP of Canada have the maximum effect on the predicted trade. The Durbin-Watson coefficient [4] for all the models is strictly less than one. This,



combine with the collinearity of lag-1 GDP, suggests that GDP growth create inertia in trade growth as export markets take time to react to bumps in demand.

## Effect of distance

We observe a negative correlation between distance of trading partner country from Canada and trade for all models. We suggest that this is explained through a combination of two effects. First, the United States of America is the nearest country to Canada and its largest trading partner. Second, Canadian export trade the more costly it is to ship goods and clearly, shipping costs would increase with distance.

Of interest is the standardized values of the coefficient correspond to distance. The magnitude of the coefficient for both import measures (value and variety) are approximately 40% higher than exports. We suggest that this reflects the type goods imported vs. exported. We hypothesize that Canada imports a variety of manufactured or finished products and far more heterogeneous with multiple supplier and there for distance will have a greater impact on destination price to the consumer. Our exports are dominated by raw materials – energy and resources and agriculture - or more homogenous, and will be dominated by world prices, long term contracts and our dominant role in supply chain. Hence distance will have less of an effect.

## Effect of Free Trade Agreements

We observe that presence of free trade agreements between Canada and trading partner countries has a positive effect on trade volumes; but is non-significant for trade variety. Further, the model suggests import volumes are enhanced to a higher degree than export volumes, for some of the same reason mention above a regards the distance impact..

## Effect of Language

The initial analysis looked at if language presents a barrier to trade. The fully specified models suggest that trading partner countries where the primary language is not English increases trade. Even though the variable was significant in two of the four models, we have chosen to exclude this.

The People's Republic of China is Canada's second largest trading partner by volume and does not have English as the primary language. We speculate that the emerging growth in Chinese demand for raw materials and Canadian exports thereof, that this effect has nothing to do with language and have therefore chosen to remove this variable from the analysis as we believe it misleading.

## Effect of Land Locked Trading Partner Countries

Land locked trading partner countries have an inherent cost disadvantage due to additional shipping required. Additionally, intermediate countries may choose to levy tariffs or fee that further increases costs to trade. We observe a negative effect on trade when Canada trades with land locked countries. However, the model for import varieties is unaffected by a trading country being land locked. This also indicates that only volume is impacted by costs and that costs impact all variety of goods, so there is nothing specific about the distribution of goods being exported because they are land-locked.

## Effect of Visitors

We observed a positive impact on trade, as measured in volume and variety, attributable to increased number of visitors. We believe there are both a concurrent affect and a delayed (lagged) affect on trade. This reflects the nature of trade deals and transaction costs of getting things done. Further, only import volumes are unaffected by visitors in the present year.

The observations are consistent with findings in [1] and [2]. The effect magnitude though differs.

## Inertia in Trade

The Durbin-Watson coefficient of the models is less than one. This is indicative of autocorrelation in trade. Canadian exports are focused on materials and heavy industrial sectors. Contracts in these sectors tend to be executed over multiple years. As a consequence, there may be *inertia* in the trade values.

To test this hypothesis, we built a regression model that included lag-1 trade as a variable. The resulting model had a Durbin-Watson coefficient of 2.4. However, the variance inflation factor (VIF) for the lag-1 trade variable was 5.5. Further, the t-statistic on the  $trips_t$  variable was -1.27. This would suggest that the inertia in trade was being absorbed into the  $trips_t$  and  $trips_{t-1}$  variables in the reduced models.

## Caveats

We recognize the analysis is subject to a number of limitations, some of which are enumerated below

- The analysis is based on data provided for Canada as sourced.
- All values were normalized in Canadian dollars using historical Bank of Canada daily mid-market rates.
- We conjecture there is an ambiguity in the matter of visitor classification at the border. We have consolidated all visitors into a single pool instead of separating into classes.
- We attempt to invoke Granger causality [3]. However, the analysis is still grounded on a correlational base. Actual causality may not be inferred without appropriate controls.
- We have not independently verified the data values to be consistent with other sources.
- The number of Canadian citizens crossing the borders was not available. The effect of Canadian citizens traveling is un-measured in this study.
- The classification of imports and exports using HS06 codes is non-uniform. That is, for some categories of goods and services the classification is at a fine grain. For other categories, the classification is very broad. We have not *normalized* the grain of classification.
- All statistics were calculated with a 95% confidence level.
- We recognize that a number of variables have time correlation inherently (e.g. GDP). We have not temporally decorrelated the values. An effect of this is the relatively low values of Durbin-Watson statistic on all regression models.
- The analysis is presented on a cross-sectional snapshot of data.
- The regression models are not cross-validated with independent hold-out data. We do not recommend usage of the regression coefficients in predictive capacity.
- We assume homoscedasticity of the data. The regression residual analysis indicates approximate normal distribution of residuals.

## References

- [1] J. P. Poole, "Business Travel as an Input to International Trade," University of California, Santa Cruz, CA, 2010.
- [2] K. Keum, "International tourism and trade flows: A causality analysis using panel data," *Tourism Economics*, vol. 17, no. 5, pp. 949-962, 2011.
- [3] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, vol. 37, no. 3, pp. 424-438, 1969.
- [4] J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression," *Biometrika*, vol. 58, no. 1, pp. 1-19, 1971.