LOOP and the Internet Marketplace

Examining the Importance of Google Search Traffic in Explaining the Existence and Persistance of Arbitrage Opportunities on eBay

Ben Southgate
Carleton College

Abstract

This paper examines 12,887 individual eBay, english-style, completed auctions of the Otterbox Defender iPhone 4/4S cell-phone case over a 90 day period, with sales and purchases occuring in both the USA and Canada, all collected from Terapeak.com. Findings suggest quick reversion of cross-border prices to parity. Specifically the half-life of deviations is estimated to be 10.34 hours. With this evidence in mind, eBay appears ever closer to an Economist's ideal marketplace, with limited price friction, large numbers of informed buyers and sellers, and low barriers to entry.

Author Biography:

Ben Southgate is a recent graduate of Carleton College (class of 2013). He currently works as a research assistant for the Urban Institute in Washington, D.C., doing statistical programming and microsimulation. A California native, his interests include water-polo, learning new programming languages, long road trips through the desert, and posting on bensouthgate.com.

I. Introduction

Since eBay.com's launch in 1995, the company has quickly accumulated users to become the largest online auction-style marketplace. Online, entrepreneurs have brought about a new era of global commerce, pushing international trade closer to an economist's ideal of a frictionless, easily accessible, and transparent marketplace. Merchants can now offer items for sale to the global Online market for a miniscule amount of what it would cost to maintain a traditional brick and mortar shop. Additionally, buyers can easily utilize the (mostly) unrestricted wealth of information contained within the Internet, as many websites offer sufficient resources to find the lowest available price for a good. Currencies can be easily, and sometimes automatically, converted. The massive volume of Internet users provides an integrated marketplace of previously impossible size, further reducing the ability of merchants and firms to charge above marginal cost, and giving consumers a greater range of product choices.

The same attributes of Internet commerce that enable near perfect competition also provide an ideal arena in which to examine economic theory. Studying an economic system that demonstrates many classical economic assumptions (perfect competition, free flow of information, etc.) can give better insight into the validity of theory stemming from those assumptions. Additionally, the explosion of data sources that came with the Internet's growing scope provides ample opportunity for analysis. Companies like eBay and Facebook collect and store massive amounts of data about the habits and interests of users, which may be used to learn more about patterns of consumer behavior or macroeconomic trends. However, since the development of Internet commerce is a relatively recent phenomenon, many possible economic questions have yet to be asked, and many classical theories are left untested in a web-based context.

In this paper, I examine one of these relatively untested theories. Specifically, I examine the Law of One Price in the context of completed eBay english-style (ascending price) online auctions with Canadian and American participants. In doing so, this paper is broken into 5 separate stages. First, I lay out previous analysis conducted concerning the Law of One Price, both on and offline. I discuss in detail the paper seminal to my analysis (Maier, 2010), as well as other foundational literature on adherence to the Law of One Price, and different aspects of eBay auctions. Second, I identify a particular gap in the literature, that of correlating the volume of consumer interest (as measured by Google Searches) in a specific item auctioned on eBay to fluctuations in the mean and variance of its completed sale prices, and transitively on the real exchange rate (calculated specifically for the product). Thirdly, I describe my process of collecting auction data from Terapeak (eBay's data hosting service) and search traffic data from Google, as well as steps taken to clean and prepare the data for analysis. Fourthly, I lay out theoretical and econometric models concerning the relationships between the collected data, with a specific eye towards the effect of consumer interest on price fluctuations both intranationally and internationally. Finally, having laid out the theoretical and statistical structure of my analysis, I then describe the various statistical findings, describing in detail the theoretical conclusions which may be appropriately drawn from the econometric results, providing a self critique of my analysis and discussing potential future research.

The empirical results from this study of the efficiency of eBay as an international marketplace, and the extent to which measured search interest correlates with fluctuations in the price level in Canada and the USA, and the real exchange rate between them, provide a few interesting insights. Firstly, while there does appear to be a border effect, it appears to be negative (at least for the product examined)! Completed auctions which involve sending a product across the USA-Canada border have a statistically lower total price than auctions that are bought and sold entirely within the USA or Canada. The mysterious cost to moving goods across national lines investigated by (Engel and Rogers, 1994) seems to have flipped its sign, at least with this product and timeframe. In response, I offer the theory that in individual online auctions, the anonymity of the users might cause buyers to prefer to deal with merchants in their own country. Additionally, the real exchange rate between the USA and Canada, calculated using the mean daily ending auction price in each country, is estimated to revert quickly to its mean when shocked from parity, suggesting high efficiency of international market dynamics on the website. With regard to investigating the influence of Google Search traffic on price levels and the real exchange rate, there is some evidence that increased search interest in an item will raise its mean ending auction price, but further investigation is needed to determine its relationship with fluctuations in the real exchange rate.

Overall, this paper argues that the internet is highly efficient as a small party market of international exchange. The fact that more and more trade is being pushed online should imply that international trade will get more and more efficient, benefiting everyone through increasing price flexibility and competition. Additionally, there seems to be potential for web search traffic to be used as a metric for predicting fluctuations in international macroeconomic variables, as they have already been shown to be useful in some types of forecasts (Choi and Varian, 2012).

II. Background

2.1 General Foundations: The Law of One Price

If an examination of the adherence to the Law of One Price (LOOP) within online auctions must begin with a review of literature on the general theory of LOOP, then Kenneth Rogoff's 1996 publication, entitled "The Purchasing Power Parity Puzzle", serves as an excellent starting point. In the paper, Rogoff summarizes past economic analysis of the reality of LOOP and its aggregation, Purchasing Power Parity, over the past century. This section will discuss the portions of Rogoff's paper which are relevant to the development of empirical analysis concerning LOOP, while addressing early studies, common themes, and potential suggestions for future investigations.

Rogoff begins by discussing the long history of the theory of international price parity, noting that scholars as far back as the 1500's had been questioning its accuracy and persistence, with LOOP being "first articulated by scholars of the Salamanca school in sixteenth century Spain". Part of the reason why LOOP has such a long history of intellectual attention is its simple structure. It doesn't require much mathematical ability to appreciate its potential power and logical foundation. The idea that for good i, the price of that good in the home country should equal to the price of the same good in other countries when taking into account the real exchange rate makes intuitive sense. However, Rogoff states that while LOOP and PPP may incite "warm fuzzy feelings" in economists due to its theoretical elegance, they are not "a substitute for hard facts". Rogoff notes that empirical analysis has repeatedly arrived at the conclusion that LOOP fails to hold, often spectacularly, in the short run. Additionally, the evidence suggests that deviations from parity seem to die out at a rate of 15% per year, and that only in the long run aggregate does LOOP somewhat converge to parity. The central question Rogoff then advances in his introduction is the following: "How can one reconcile the enormous short-term volatility of real exchange rates with the extremely slow rate at which shocks appear to damp out?". Essentially, what is causing the lack of international price parity in the short run?

In section 4 of his paper, Rogoff expands on the discussion of previous empirical analysis of LOOP, establishing the general trend of finding large, persistent deviations from parity. One of the earlier analyses of LOOP adherence was (Isard, 1977), who examined the comparative movements of disaggregated U.S. and German Industrial and Machinery prices of highly traded goods. Isard examined trade data from 1968-1975, during which Bretton Woods collapsed and exchange rate volatility increased dramatically. He found that fluctuations in the nominal exchange rate have substantial and persistent effects on the real exchange rate. Isard's analysis was early indication of a lack of efficient arbitrage, even among goods seemingly theoretically close to ideal for frictionless trade (that of easily substitutable industrial commodities). Rogoff then notes how Isard's analysis of disaggregated goods was complemented by (Richardson, 1978), which examined industrial commodity trade between the U.S. and Canada, as well as (Giovannini, 1988), which investigated US-Japanese manufacturing commodity trade. Both studies affirmed Isard's conclusions of inefficient arbitrage and volatile fluctuations of the real exchange rate. The consistency and magnitude of deviations from LOOP found in all the studies are a testament to the scope of LOOP's general failure.

After discussing the scope and volatility of real exchange rates between international bodies, Rogoff then points to literature examining the differences between intranational and international trade, motivating the thesis that the real puzzle is the volatility of real exchange rates between, and not within, nations. This distinction is important, as it affirms that the failure of arbitrage as a general concept is not apparent. Rogoff first cites (Engel and Rogers, 1994) which examines trade between the U.S. and Canada, as well as within each nation itself. Disaggregated product data is analyzed, and substantial deviations from international LOOP adherence are found. Engel and Rogers go further in their conclusions to infer the cost of simply having a national border between two trading areas in terms of physical distance, finding that national borders inherently equate to transportation costs of traveling the circumference of the earth three times (75,000 miles), ceteris paribus. In addition to the vast scope of international price level differences relative to differences in intranational price levels, Rogoff notes the equally shocking magnitude of variance in the real exchange rate between nations. (Engel, 1994), for example, examines the variance of price differentials across and within national borders. The paper concludes, amazingly, that prices for substantially different goods within a country have a smaller differential variance than substantially similar goods traded across countries. Somehow, the border effect cancels out the efficiency resulting from a market of similar goods. This finding makes it at a minimum hard to argue that price differentials are a phenomenon constant at any level of geopolitical division. Moreover, it appears that there is some inherent friction to trading between nations, beyond what tariffs, different currencies, and physical distance would initially suggest.

After showing that LOOP consistently fails on the international stage, Rogoff then discusses the extent to which these assertions have been present throughout recent history. Rogoff cites some of his own work (Froot, Kim and Rogoff, 1995) which looks at international grain price differentials over 400 years between England and Holland. The study found that price differential volatility was relatively consistent throughout the period examined, arguing against any assertion that the failure of LOOP internationally is a recent phenomenon caused by new institutional, technological, or cultural factors.

Section 4 of Rogoff's paper concludes by advancing several possible reasons for disparities between international price levels for specific, highly tradable goods. Transportation costs are stated to account for a small but important part of price level discrepancies, as physically moving goods from one place to another is never free. Additionally, nontradable aspects of tradable goods, like the rent of housing the items and other local factor costs can provide additional disparity (costs which are not incurred in an online market). For example, companies might incur specific insurance or local shipping costs in some countries but not others. Finally, there may be elements of price discrimination due to the varying sources and availability of product comparison across borders.

Overall, Rogoff's discussion of the repeated, historical, and substantial failure of international adherence to LOOP offers several key reasons why an analysis of the theory in the context of eBay auctions is the natural next step. First, the fact that nominal exchange rate fluctuations are consistently found to have a large effect on price level disparity between nations might be mitigated by the automatic currency conversion of eBay listings. The automatic conversion of prices removes the barrier to information flow of constantly changing foreign exchange data. This might allow arbitrage channels to be lubricated, as good prices can be quickly and reliably be compared, even in the face of nominal exchange rate fluctuation. Additionally, the fact that LOOP seems to fail much more on the international stage than within a country may suggest that non-visible communication or organizational difficulties between international market actors might be occurring.

By providing a centralized, easily accessible forum for trade across borders eBay auctions may help to remedy these difficulties, providing a marketplace where whatever language is required can be utilized. Additionally, eBay's data hosting company, Terapeak, provides a product which allows merchants to easily see price differences that arise in different countries, promoting more free flowing information. Furthermore, the issue of potential non-tradable costs to tradable goods is largely mitigated by the lack of factor costs to listing items on eBay, leaving only physical shipping costs which are completely visible. Finally, price discrimination, for all intents and purposes, essentially can't happen on eBay, as consumers can easily compare prices for goods listed across country lines in different currencies.

2.2 Physicality to Abstraction: Analysis of Online Markets

While Rogoff's discussion of empirical consideration of LOOP presents several reasons why online markets might be an ideal place to examine the evolution of the theories adherence to reality, it was written before the Internet truly matured as a marketplace and as a source of data. This next section will examine several papers which ask economic questions about online markets, with particular attention to said markets effectiveness in moving towards (more) perfect competition.

While the Internet as a construct may have seemed like a theoretical key to perfection regarding international commerce during its initial adoption by consumers; early research did not entirely agree. In 1999, The Economist published an article entitled "Frictions in Cyberspace" which discussed the new potential of the Internet to remove market frictions. The article gave a mixed initial prognosis, praising the flexibility of prices on the internet while warning against potential price volatility.

> "Studies do show that online retailers tend to be cheaper than conventional rivals, and that they adjust prices more finely and more often. But they also find that price dispersion (the spread between the highest and lowest prices) is often as wide on the Internet as it is in the shopping mall—or even wider." - The Economist, 1999

Early empirics of the Internet's theoretical prospects as a paradigm-shifter seemed to be mixed, but was this evidence of the lack of inherent potential within the Internet, or merely its failure to be sufficiently, and efficiently, utilized by consumers and firms?

A highly cited paper by Michael Baye, John Morgan, and Patrick Scholten (2004), provides both a nice summary of early research on Internet commerce, as well as original empirical analysis on price dispersion in online markets. Most importantly Baye, Morgan, and Scholten utilize a dataset that is a perfect example of why the Internet has not only changed the markets themselves, but also the way in which markets can be analyzed. The paper examines 4 million price observations collected from a price comparison site, data of a size and scope that would have seemed preposterous to Isard when he wrote his 1977 article. That such data can be collected accurately and quickly clearly illustrates how economic analysis has entered a new era since the days of empirical work summarized by Rogoff in his review of LOOP analysis.

After pointing to the evolving literature around price dispersion in online markets, Baye, Morgan, and Scholten turn to their own analysis. The paper set out to examine the scope and persistency of price dispersion among Internet retailers. The authors collected over 4 million prices across ≈1000 goods using a PERL-based web-crawling script which mined prices from a price comparison website, Shopper.com (now part of CNET.com). Additionally, the dataset represented a large number of firms, varying from product to product, allowing the study to examine the effect of seller market size on dispersion.

Before turning to empirical analysis, Baye, Morgan, and Scholten first built a theoretical framework for the concept of a price comparison website, adapting a clearing house model. The model essentially assumed that firms can list prices for products at a clearing house, and consumers can come and choose whether or not to purchase. It also assumed that there are two types of consumers, those who pay the lowest price, and those who don't (due to perceived heterogeneity of the products, or lack of knowledge about the price comparison website/clearinghouse). In all, the theoretical framework assumed some consistant level of price dispersion that should be negatively correlated to the general number of sellers.

Baye, Morgan, and Scholten then noted the strengths and limitations of their data. Strengths were stated to be the size and scope (≈4 million prices, ≈1000 products, over the course of ≈8 months), the high minimum price of products (large price discrepancies due to brand preference are more likely to arise in low-cost products than high-cost products), as well as the "seriousness" of offers listed on the price comparison site, which charges $1000 to merchants for the ability to list, as well as per-listing fees of $.50. The primary limitation discussed was the lack of data on number of goods sold at each price.

The results showed price dispersion to be a clear, pervasive, and consistent effect. Price differentials varied by 10% on average over the time surveyed. The consistency of the dispersion is an argument against its theoretical existence as a dis-equilibrium state, something that real-exchange rate parity implies (even for intra-national trade). Additionally, dispersion seemed to be highly correlated with the number of sellers[1].

Baye, Morgan, and Scholten give a good analysis of online price dispersion within specific retailers, and do so with a vast and comprehensive dataset. However, the theoretical framework, of a pay-to-enter listing service, serving brand loyal buyers for nearly identical goods, presents several limitations that might be approached through analysis of products on eBay. eBay removes the adverse effects of comparing nearly homogenous, but not completely homogenous product prices as the prices for individual transactions are publicly vis-

1 The results in this paper contain somewhat similar findings, with the number of auctions on a given day being negatively correlated to the mean ending auction price on that day (see section 4.1).

ible, and the specificity of goods can be near exact. Additionally, the average size of most eBay "firms" is smaller than the professional retailers which pay large amounts to list prices on comparison sites such as Shopper.com. This may have the theoretical effect of increasing individual attention to sales of products, as merchants have a greater vested interest in selling their particular good or few goods. It is with this comparison and motivation that we next turn to specific analysis of international and domestic price dispersion on eBay itself.

2.3 Motivating Literature: Examining LOOP on eBay

One study of particular interest to the study of LOOP online that cites Baye, Morgan, and Scholten's analysis of online price dispersion, is Philip Maier's 2010 paper, entitled "An Analysis of International Price Differentials on eBay". Maier's paper is currently the only analysis of the validity of LOOP using the massive online marketplace that is eBay.com. This section will discuss Maier's theory, model and results, as well as how this paper intends to build upon his work.

In his introduction, Maier discusses the importance of eBay as marketplace, both within the U.S. and Internationally. In 2005, eBay reported that "180.6 million users from more than 150 countries listed 1.9 billion items", resulting in a "total value of goods sold that year reached nearly 44 billion U.S. dollars." As of August 2012, eBay now accounts for nearly 5% of all U.S. retail sales, making it a formidable market force that continues to grow. Maier also establishes the theoretical importance of examining eBay as a datasource for empirical tests of LOOP, noting that Internet auctions, as opposed to retailers, don't have large fixed non-tradable costs as discussed by (Rogoff, 1996).

After motivating the use of eBay as an important experimental arena, Maier divides his analysis into two parts. Firstly, price level differences are analyzed and their specific sources are examined. Specifically, do physical borders have the same large effect in online transactions as was found in offline trade (Engel and Rogers, 1994), and if so, does the border effect stem from differences in currency? In addition to price level differences, Maier examines price dispersion by correlating the variance of the distribution of the difference in price between various product pairs to different aspects of the transacting countries.

To collect data on auction final bid prices from eBay, Maier used a web crawling script similar to the spider used in (Baye, Morgan and Scholten, 2004) which collected data on 5,648 completed transactions for 16 "Hot" (according to eBay rankings) products distributed over 8 countries. The data concerning each observation included all information available to an average consumer, except individual feedback ratings. Additionally, all of the products examined were simultaneously available in retail markets at the time of collection. The products varied in "genre", as there was a mix of consumer electronics, apparel, and toys. Finally, while the location of the buyer was not listed, Maier tested for arbitrage opportunities by determining whether the price difference for a good between two countries exceeds its shipping cost.

Having collected the data, Maier first conducted an econometric test to determine the existence of price differentials between countries by examining the coefficients on country specific dummy variables, with Germany as the base case. This first model took the form:

$$\hat{p}^i_k = \hat{\alpha}^i_0 + \sum_m \hat{\alpha}^i_{1,m} \text{eBay}^i_{m,k} + \sum_n \hat{\alpha}^i_{2,n} \text{Country}^i_{n,k} + \epsilon^i_k \tag{1}$$

where p^i_k is the logged price of good *i* in auction k, eBay represents the *m* different aspects of the auction, and Country represents the *n* country specific dummy variables.

The results for the price prediction model indicated the importance of different aspects of eBay listings. Displaying a photo increased the price, longer auctions received higher prices, buy-it-now auctions received higher prices (in agreement with (Budish and Takeyama, 2001) which found that fixed "buy it prices" resulted in theoretical risk reduction to consumers which would be compensated for through higher average revenue), more bidders tend to drive up the price, and newer products receive higher prices. With regard to country specific effects, the coefficient indicating the dummy variable for U.K. (relative to a baseline of Germany) was lower than some euro-area countries, directly opposing of euro-area common currency relative price adherence.

The second econometric test modeled the variance of product pair price differentials on auction and country explanatory variables, specifically taking the form:

$$\text{VAR}(\hat{p}^i_{A,B}) = \hat{\beta}^i_0 + \sum_m \hat{\beta}^i_{1,m} \text{eBay}^i_{m,k} + \sum_n \hat{\beta}^i_{2,n} \text{Border}^i_{n,k} + \epsilon^i_k \tag{2}$$

Where $\text{VAR}(\hat{p}^i_{A,B})$ is the variance of the distribution of price differences between all possible auction pairs between country *A* and *B*, and Border represents the different border effect variables, including the existence of a physical border and sharing a common currency. The price dispersion model indicated that sharing a common currency did indeed reduce variation of product pair price differentials. This result is somewhat interesting, as eBay automatically converts currencies on its website, so dealing with differing currencies shouldn't present too much of a hurdle[2].

In summary, Maier's analysis is an important step in examining the scope of adherence to LOOP through use of data from a highly utilized, auction style, international online marketplace. However, there are several limitations to the data and model that I attempt to expand and improve upon. First, the lack of information concerning the buyer's location severely limits identification of cross border trade (Maier had to use logical inference instead). The dataset of individual auctions I use in this paper, collected from Terapeak.com, includes seller and buyer nationality information, allowing my investigation to not only identify the direction of trade, but also to isolate intranational transactions and price dispersion. Second, Maier briefly mentions the differing levels of interest in eBay among different countries, but does not attempt to control for this. In response, I utilize Google Search traffic data collected from Google Trends to proxy both the level of Internet search interest in both an item and the item on eBay in particular, and then model the daily mean price, price variance, and fluctuations in the product real exchange rate using both variables.

2.4 Contribution to the Body of Literature

Economic literature concerning the reality of the Law of One Price is beginning to enter a new phase, as the Internet provides a wealth of data and commercial activity for future analysis. Most of the literature concerning LOOP so far has been related to offline trade of commodities, and the little study that has been conducted concerning online markets (Maier, 2010; Ellison and Ellison, 2009; Brynjolfsson and Smith, 2000) has yet to consider other online variables such as search interest, and how those variables effect international parity. eBay in particular represents an ideal marketplace to examine international good price differentials and their variance, along with consumer search interest data, as it is exceptionally popular and becoming a staple of online small business. This paper sets out to expand the current literature on the Law of One Price by considering new data sources, Terapeak and Google Trends, as well as new theoretically important variables, such as consumer search interest, that have yet to be utilized.

III. Method

3.1 Data Summary

As there is limited aggregate price data for online auction marketplaces, I chose to conduct analysis of individual auctions of a particular, high-volume,

2 This result could be theoretically explained by consumers and merchants pricing in the cost of spot-rate conversion charged by eBay's currency exchange services partner, XE.com.

product and make theoretical aggregate inferences from the results about eBay as an internet goods exchange. Analysis of individual goods in lue of aggregate indicies has been conducted before in examination of online marketplaces (Baye, Morgan and Scholten, 2004; Maier, 2010), and has the added benefit of theoretically faster reversion to parity, potentially negating the downsides of having a short time-frame of 90 days.

To gain sufficient observations, I chose to examine a product that had relatively high sales volume on eBay. Fortunately, as part of their product for eBay merchants, Terapeak offers a search of products by total volume over a specified time period. I examined the most traded, by unit volume, goods over the past 90 days, eventually selecting the Defender case for the iPhone 4/4S made by Otterbox (Figure 1).

I chose the Defender for several theoretical reasons in addition to its sales volume dominance. Firstly, the case is light and easy to ship, suggesting that the logistical costs of trading it online for sale should not have too much of an impact on sellers' responsiveness to price fluctuations. Secondly, the Otterbox Defender case has hundreds of theoretical substitutes[3], suggesting that inference onto a general iPhone case market might be more justified, as prices across the different brands should equilibrate. Finally, the fact that the ratio of N purchases made by Canadians to those made by Americans roughly reflect the relative sizes of each countries' population[4] suggests that consumer interest in the case is relatively aligned across the border. Additionally, correlation between online interest in the Otterbox Defender in the USA and Canada, in terms of Google search indicies, was found to show evidence of significance on a daily basis[5].

The final sample consisted of 12,887 Completed english-style auctions of new, individually listed, Otterbox Defender cases. Of these transactions, 12,136 occurred within the USA and 112 occurred within Canada. Additionally 502 cases were sold by an American to a Canadian, and 137 cases were sold by a Canadian to an American.

Other than the individual auction data collected from Terapeak.com, I also utilized daily search traffic data from Google Trends (http://google.com/trends). Google Trends is just beginning to be recognized for its potential in predicting economic indicators, with autoregressive time series models us-

3 The point can be made that consumers of the Otterbox Defender might prefer a more "rugged" case, which may rule out many other cases on the market. However, even if the market is restricted to rugged cases, the Defender has ample competition such as "Lifeproof", "AQUA TEK S", "iFrogz", and others.

4 Precisely, ≈5% of auctions in the data have a Canadian buyer, while Canada has ≈10% of the population of the USA

5 The Pearson's Product-Moment correlation coefficient between ["Otterbox Defender"(CA)]t and ["Otterbox Defender"(US)]t was 0.6625 with a t-statistic of t = 7.6089.

ing search data beating traditional econometric forecasting models in terms of predictive power (Choi and Varian, 2012). Google Trends presents search traffic in the form of a normalized index, which is calculated using the query share, defined as (Keyword Queries)/(Total Queries) for a specified region. The index is normalized to a scale of 0 - 100, with a value of 100 assigned to the time index with the greatest query share within the specified time interval. For this paper, I collected daily observations from 10/16/12 to 12/30/12 for the keywords "Otterbox Defender" searched in the USA and Canada, as well as the Keywords "Otterbox eBay" in the USA[6].

Figure 1 : Black Otterbox Defender for the iPhone 4/4S

3.2 Data Collection Process

In order to analyze the scope of international price differentials on eBay, I utilized eBay's data hosting service, Terapeak. Terapeak hosts data on all eBay listings and transactions on eBay.com and all country specific eBay sites. The service maintains individual auction level data for the past 90 days. Terapeak provides two services regarding access and analysis of the hosted data. One service is the free developer API, located at https://developer.terapeak.com, which allows software developers interested in developing user-end ebay related software products to make requests for data, using Hypertext Transfer Protocol (HTTP), directly to the server. The other service is a graphical user interface (GUI) search engine directed at eBay merchants, located at http://www.terapeak.com. For 50 USD per month, merchants can track various aggregate statistics about the items they are selling. Specifically, merchants can see the average price of successful sales in different countries, as well as note product description keywords that have relative success in sales.

For the actual collection of individual observations, I chose to use Terapeak's GUI product. The search interface itself, displayed in Figure 2, allowed specification of the following variables: Buyer Country, Seller Country, Item Condition, Auction Type, Listing Duration, Seller Id, Time of Day, and Starting/Ending price ranges[7].

To restrict my sample for analysis, I searched the keywords "otterbox

6 The volume of "Otterbox eBay" among Canadians was too low to allow data collection, see section 3.2.

7 All currency values listed in daily spot-rate converted USD, with the conversion performed by XE.com, a Canadian foreign exchange services company.

iphone 4 defender", and limited results to bid (english style) auctions, of single, new cases. All Auctions were listed on the eBay parent website, http://www.ebay.com. I ran four separate initial searches over the ninety day period from 10/02/12 to 12/30/12, specifying following [buyer, seller] country pairs: [USA, USA], [USA, Canada], [Canada, USA], and [Canada, Canada]. Each search returned all the auctions fitting all of the specified criteria over the time period. Additionally, the average shipping price (calculated as the average shipping costs paid by purchasers over the search results), displayed above the listings themselves, was copied into the spreadsheet as was the [buyer, seller] information.

The resulting spreadsheet, after copying the search results, had 7 columns of attributes for the auctions, as well as the sell date. The first column "description", contained a string consisting of descriptive attributes of the listing written by the seller. The other columns were the starting and ending price in USD, the number of bids, the average shipping cost[8], and the country of the buyer and seller. All formatting in the spreadsheet was cleared and the file was saved as a comma separated values file in preparation for processing by the Python cleaning script.

The other data used in this paper, that of Google Trends data on keywords pertaining to Otterbox Defender Cases, was obtained directly from http://www.google.com/trends. The keywords "Otterbox Defender" were first searched (restricted to the last 90 days)[9], and two separate spreadsheets of daily observations were downloaded using the built-in export feature, one for search trends among Americans and Canadians respectively. Additionally, search traffic for "Otterbox eBay" was also obtained for American Google users only, as there was not enough search traffic to merit index calculation for Canada[10]. All the search traffic variables were combined into a single spreadsheet.

8 This was calculated by Terapeak, and consisted of the average across search results for the specified period. This meant that there was one average shipping cost for the [USA, Canada], [Canada, USA], and [Canada, Canada] searches, but multiple for the [USA, USA] search (one for each 3-day period).

9 Unfortunately, daily observations are limited to 90 days prior to the search, and as the collection of Google data was conducted 14 days after auction data collection, the time periods are slightly askew. This resulted in 14 observations being lost in the aggregate time series analysis, and will be discussed in the Results section.

10 Google Trends doesn't present search traffic data if it falls below a certain volume out of concern for user privacy (Choi and Varian, 2012).

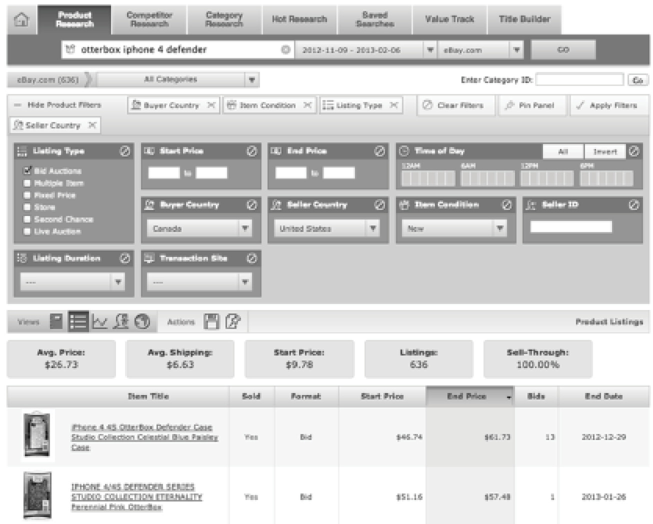

Figure 2: Terapeak graphical search interface, with example output.

3.3 Data Cleaning and Pre-Analysis Processing

Once the relevant individual auctions had been properly collected and dumped into a spreadsheet, pre-analysis processing was conducted to mine the "description" (Labeled as "Item Title" in Figure 2) string variable for important discriminating atributes to be used as controls in the price determination regression, as well as the calculation of daily aggregate variables.

All data cleaning was conducted by an author-coded Python programming language script which found the most common words used in the auction description, calculated the real exchange rate and its variance, and interpolated some missing values (such as average price in canada for days which contained no auctions, see section 5.2 for details.

3.3 Economic Theory and Model Construction

After appropriately cleaning and pre-processing the data, I began the econometric analysis of the validity of LOOP with regard to eBay auctions involving the USA and Canada, as well as the determinates of deviations from LOOP parity. My analysis took four separate stages (Figure 3) to show the existence of deviations from perfect price parity, and then determine whether or not Google Search traffic appeared to play a role in fluctuations of the real exchange rate. Walking through the steps of analysis, I first established the existence of potential arbitrage opportunities, implying deviation from LOOP, through regression of individual auction prices on attributes of the auction and its country of origin. After showing the failure of LOOP on eBay, I investigated the extent to which Google Search traffic correlates with the daily mean total price of auctions, as well as the variance of prices, on each day in the sample. Searches by Canadians and Americans are tested for correlation with the daily mean price (and variance) in both Canada and the USA. Next, the logged real exchange rate (qt) was analyzed to determine if it indeed reverts to a mean of 0, as LOOP suggests. Upon finding mean reversion, I finally investigated the extent to which Google Search traffic appears to relate to changes in the evolution of qt.

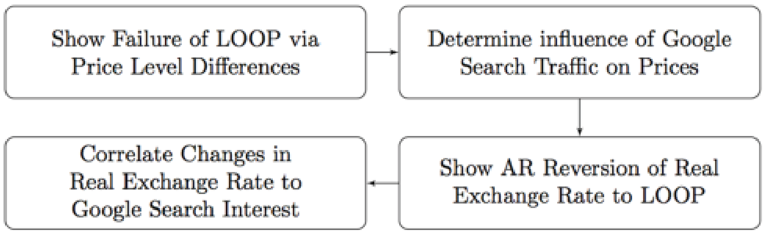

Figure 3: Steps of Analysis

First, using the individual auction level data generated by the Python script, I attempted to model the total price of each auction, using the starting price, number of bids, country of the buyer and seller, and all the descriptive attribute binary variables. Following the specification of logged individual auction prices as linearly dependent on the atributes of the auction and the product itself laid out by (Maier, 2010) (see equation 1), where $\hat{p}$ is the estimated (logged) total price of each auction,

$$\hat{p} = \hat{\alpha}_0 + \hat{\alpha}_1 K + \sum_{i=1}^{x} \hat{\alpha}_{2,i} A_i + \sum_{j=2}^{y} \hat{\alpha}_{3,j} C_j + \sum_{k=2}^{t} \left(\hat{\alpha}_{4,k} D_k + \sum_{j=2}^{y} \hat{\alpha}_{5,j} (C_j * D_k) \right) + \hat{\epsilon} \qquad (3)$$

K indicates if the case had a belt clip. $[A_1, A_1, \ldots, A_x]$ are the x non-color attributes of the auction, specifically including the loggd number of bids, the logged starting price, and the country of the [buyer, seller] (with [USA, USA] omitted as the baseline). $[C_2, C_3, \ldots, C_y]$ are the y different colors the case might take (green was omitted as a baseline). $[D_2, D_3, \ldots, D_t]$ represent dummy variables for each of the *t* days in the sample (with the first day omitted), these were included to control for day specific effects that might be caused by general ebbs and flows in consumer demand for the product. Finally $(C_j * D_k)$ are the interaction terms between each color and each day binary variable, to control for changes in consumer color preference over time. In theory, each of the $\alpha_{2,i}$ coefficients on the binary variables indicating [buyer, seller] should equal 0, as there should be no opportunities for arbitrage and that the direction of trade should be irrelevant (assuming the validity of LOOP).

Switching to analysis of daily aggregates, I attempted to model the mean and variance of daily auction prices in the US and Canada on the number of auctions (N), mean bids per day (B), and Google search traffic data. A theoretical multiplicative model (of identical form in the determination of mean price P_t) of

$$\mathrm{VAR}(p) = e^{\beta_0} * B^{\beta_1} * N^{\beta_2} * \prod_{i=1}^{3} \mathrm{Google}_i^{\beta_{3,i}} \tag{4}$$

was assumed, as this allows for diminishing (or potentially increasing) returns to successive unit increases in N, B, and the Google Trends variables, which all (theoretically) result in increased mean price and reduced price variance. Taking logs, and first differences (in anticipation of variable difference-stationarity), the OLS model is then,

$$\Delta ln(\widehat{\mathrm{VAR}}(p)_t) = \hat{\beta}_1 \Delta ln(B_t) + \hat{\beta}_2 \Delta ln(N_t) + \sum_{i=1}^{3} \hat{\beta}_{3,i} \Delta ln(\mathrm{Google}_{t,i}) + \widehat{\Delta\epsilon_t} \tag{5}$$

Specifically examining the theory behind the model of daily price variance, each of β_n (for $n>0$), should be < 0, as all are different measures of market participation. B and the Google Trends variables are each measures of the participation in the market by buyers. Assuming no collusion by eBay merchants, and as long as there is more than one seller on a given day, more buyer participation should increase the likelihood that all seller prices are observed and compared by buyers. This should prevent the sellers from raising their prices too high, as buyers will simply switch to another seller. Additionally, the number of auctions per day is a proxy for the number of sellers (firms) in the market on that day. As basic microeconomic theory suggests, increasing the number of firms will decrease the ability of firms to raise price above marginal cost. Overall, the increase in competition should pull prices towards their mean value, decreasing the variance of the distribution of individual auction prices on each day.

With regard to modeling the mean auction price (first differenced) on each day, the β coefficients on ΔB and the Google Trends variables should be > 0, while the coefficient on ΔN should be < 0. B and all of the Google Trends variables are proxies for consumer demand on that day, and when demand for the case is higher on a given day, its price should follow. N however, can be seen as a proxy for supply, and with an increase in the number of suppliers, price should be driven down towards whatever price eBay merchants deem marginal cost.

Next, the first difference of the logged real exchange rate ($\Delta q_t \equiv q_t - q_{t-1}$) was modeled using an AR(1) process to test for the presence of a unit root, as well as estimate the first lag coefficient (λ) and use it to calculate the half-life of deviations from LOOP,

$$\widehat{\Delta q_t} = \hat{\lambda} q_{t-1} + \hat{\epsilon}_t \quad \rightarrow \quad \hat{t}_{\frac{1}{2}} = \frac{ln(1/2)}{ln(1+\hat{\lambda})} \tag{6}$$

Allowing for failure of perfect LOOP parity, (λ) is expected to be < 0, indicating the stationarity of Δq_t and the trend towards LOOP parity in the long run. As far as this author knows, eBay auctions have never been used to calculate the real exchange rate, as well as its auto-regressive process, so there is no baseline half-life of deviations from which to derive an expected value for the half-life. In theory, however, the efficiency of online markets, as well as the specificity of the product used to calculate this estimated real exchange rate, should result in relatively fast degradation of LOOP disparities. While many estimates of the half life of real exchange rate deviations (calculated using an aggregate consumption bundle or commodities traded offline) place it in the 3-4 year range (Chen and Engel, 2004; Rogoff, 1996), the fact that qt is here calculated using one good should imply that mean reversion occurs much faster. Since "Canadian" and "American" auctions are actually listed on the same website (and show up in searches by citizens of both countries), any large deviations from price parity should not be allowed to exist for too long, as consumers will quickly take advantage of any difference in price that is immediately apparent on the website.

Finally, having shown Δq_t to evolve in an autocorrelative process, I fit it to a multivariate OLS estimator, including the first difference of the various Google Search traffic variables, in order to estimate their significance in reducing (or perhaps increasing) deviations from LOOP parity. The model takes the linear form

$$\widehat{\Delta q_t} = \sum_{i=1}^{3} \hat{\phi}_i \Delta ln(\text{Google}_{t,i}) + \widehat{\Delta \epsilon_t} \tag{7}$$

Just as the Google Search traffic variables are theoretically important in explaining changes in the mean price of auctions on a given day, they should also influence the volatility of the real exchange rate and its deviation from parity. If more people are paying attention (as indicated by search traffic) to the ending prices of auctions each day, in both countries, prices across the border should be pulled together, or arbitrage initiates. As with the regressions concerning mean price and price variance, the Google Search variables are first differenced (to determine the correlation between successive movements in the search traffic index to successive movements in the exchange rate (Δq_t).

Overall, this multi-staged investigation of eBay as an online international marketplace is intended to show its overall efficiency, while pointing to fluctuations in search traffic to explain fluctuations in the scope of that efficiency. This paper represents the first attempt to clearly isolate the effect that consumer attention plays in promoting price disparity in an otherwise highly flexible and transparent market eBay theoretically represents.

IV. Results

4.1 USA - Canada Price Level Differences on eBay

The first regression analysis conducted (equation 3) concerned the prediction of logged individual auction prices as a function of their product features (color, belt-clip, etc.), auction attributes (starting price, number of bids, etc.), and the country of the seller and purchaser. This was intended to determine the statistical significance of a difference in exchange-rate converted price level between the USA and Canada, motivating further investigation of its cause.

Equation 3 was run twice with different variables relating to the countries involved, using OLS in both cases. The first version (Model 3.1) contained only a dummy variable indicating the country in which the good was purchased (i.e. if it was purchased by an American or Canadian). This was used to determine differences in the price level itself, by showing that either Canadians or Americans pay more for the case on average. The second version (Model 3.2) contained dummy variables indicating the [buyer, seller] pair. This was intended to show possible significance of the direction of trade in determining the total price of the auction. Additionally, both regressions were conducted in a linear form, with the prices of the auctions, the number of bids, and the starting prices all in non-logged form[11]. All versions of the model were tested for heteroskedasticity via Breusch–Pagan test, using the R command ncvTest()[12]. As all four versions were found to exhibit non-constant error variance, violating the classical linear regression assumptions of homoskedasticity, the regression coefficient standard errors were re-calculated using a heteroskedasticity-consistant (HC) covariance matrix created by the R command coeftest()[13]. Theoretically important coefficients, as well as those for common case colors, from all four versions of the model are shown in Table 1 (all other control and interaction variables listed in equation 3 are not included in the table, but were included in the regression).

11 The linear form appeared to have greater explanatory power, and is thus included here.

12 From the car package, see (Fox and Weisberg, 2011).

13 From the lmtest package, see (Zeileis and Hothorn, 2002).

	3.1 (Logged)	3.2 (Logged)	3.1 (Linear)	3.2 (Linear)
(Intercept)	2.87^{***}	2.87^{***}	16.17^{***}	16.21^{***}
	(0.05)	(0.05)	(1.23)	(1.23)
ln(Number of Bids)	0.10^{***}	0.10^{***}		
	(0.00)	(0.00)		
ln(Starting Price)	0.06^{***}	0.06^{***}		
	(0.00)	(0.00)		
Number of Bids			0.46^{***}	0.46^{***}
			(0.01)	(0.01)
Starting Price			0.43^{***}	0.43^{***}
			(0.01)	(0.01)
Canadian Buyer		0.06^{***}		1.83^{***}
		(0.01)		(0.26)
[USA, Canada]	-0.06^{***}		-3.30^{***}	
	(0.02)		(0.65)	
[Canada, USA]	0.07^{***}		2.10^{***}	
	(0.01)		(0.28)	
[Canada, Canada]	-0.02		-0.16	
	(0.02)		(0.61)	
Blue	0.12^{*}	0.12^{*}	1.80	1.79
	(0.07)	(0.07)	(1.58)	(1.58)
Red	0.02	0.01	1.85	1.81
	(0.06)	(0.06)	(1.42)	(1.42)
Yellow	0.05	0.05	0.36	0.34
	(0.05)	(0.05)	(1.24)	(1.24)
Pink	0.07^{***}	0.07^{***}	1.80^{**}	1.79^{**}
	(0.03)	(0.03)	(0.83)	(0.83)
Black	0.21^{***}	0.21^{***}	5.79^{***}	5.79^{***}
	(0.04)	(0.04)	(1.05)	(1.05)
Camo	0.40^{***}	0.40^{***}	10.43^{***}	10.43^{***}
	(0.04)	(0.04)	(1.38)	(1.38)
White	0.25^{***}	0.25^{***}	7.40^{***}	7.39^{***}
	(0.04)	(0.04)	(1.20)	(1.20)
Teal	0.21^{***}	0.21^{***}	4.76^{***}	4.76^{***}
	(0.05)	(0.05)	(1.59)	(1.59)
Plum	0.10	0.10	0.07	0.07
	(0.08)	(0.08)	(2.60)	(2.59)
Orange	-0.09	-0.09	-0.16	-0.17
	(0.06)	(0.06)	(1.59)	(1.59)
Purple	0.19^{***}	0.19^{***}	5.47^{***}	5.43^{***}
	(0.04)	(0.04)	(1.12)	(1.12)
Belt Clip	0.03^{***}	0.03^{***}	0.22	0.23^{*}
	(0.01)	(0.01)	(0.13)	(0.13)
R^2	0.41	0.41	0.56	0.56
Adj. R^2	0.32	0.32	0.50	0.49
Num. obs.	12887	12887	12887	12887

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, HC Corrected S.E. in Parentheses

Table 1: OLS regressions of Individual Auction Prices (*p*) using Equation 3

The first notable result is that of the statistically significant price level difference between the USA and Canada. Noting the coefficient on "Canadian Buyer" in "3.2 (Linear)", it appears that Canadians pay approximately 1.83 USD more in total price for the Otterbox Defender on eBay than Americans, ceteris paribus. This indicates a failure of LOOP online, prompting investigation into potential reasons for disparity. The results regarding the coefficient on the [buyer, seller] dummy variables—indicating the auction was sent across the border ([USA, Canada] and [Canada, USA])—were somewhat perplexing. They indicate that sending a case from the USA to Canada adds 2.10 USD to the price (relative to intranational sales in the USA and Canada), while sending the case from the Canada to the USA subtracts 3.30 USD. Apparently, the direction of trade has a significant effect on the total price of the auction!

An investigation of potential additional costs to trade, however, presents a possible answer. The maximum price of a good allowed to be duty free is in fact lower in Canada than in the USA. Mail items sent into Canada are only free from import duty if they are of retail value less than 20 CAD[14]. If the mail item is worth more than 20 CAD, it is subject to a 5 CAD handling duty (8 CAD if express). For mail goods imported into the USA, the duty free cut-off falls at 100 USD. This means that the Otterbox Defender, with an average price of around 30 USD, falls within the difference. It is possible that sellers are factoring in this handling fee into the shipping costs. From the data sample, [Canada, USA] auctions have a mean shipping cost of 7.09 USD while [USA, Canada] auctions have a mean shipping cost of 2.85 USD, representing a difference similar to the import duty[15]. The approximately 5 USD difference between importing from the USA into Canada and importing from Canada to the USA neatly matches with the handling fee paid by Canadian importers. If we subtract the duty fee from the [Canada, USA] auctions, both directions now appear to have lower total prices than the auctions which only involved one country. It may be that buyers on eBay prefer not to deal with international transactions, and need to be compensated as such. This all boils down to speculation in the end though, and the different effects of trans-border trade on total price deserves further investigation.

Having shown the apparent existence of an eBay price level difference between the USA and Canada, as well as posited a theoretical explanation for the difference in total price related to the direction of trade, I then turned to investigation of the importance of Google Search traffic in explaining the daily aggregates.

4.2 Google Traffic and the Daily μ and σ^2 of Auction Prices

The next stage in the econometric analysis of the adherence of eBay to LOOP involved correlating the daily mean price of auctions, as well as the daily variance of auction prices, to the variables indicating Google Search traffic for particular keywords related to the product and eBay itself.

First, equation 5 was used, with logged mean auction price $\Delta ln(P_{US,t})$ and $\Delta ln(P_{CA,t})$ (both first-differenced) as the dependent variables in models 5.1 (US) and 5.1 (CA) respectively. This was intended to show any correlation between Google Search traffic and the mean price for the day. Additionally, the variance of auction prices on each day was modeled using equation 5 with $\Delta ln(\mathrm{VAR}(p)_{US,t})$ and $\Delta ln(\mathrm{VAR}(p)_{CA,t})$ (again, both were first-differenced to impose stationarity) as the dependent variables in models 5.2 (US) and 5.2 (CA) respectively. Results for all four models are shown in Table 2. As the data examined are time series, all of the variables were run through Augmented Dickey-Fuller tests for

14 Canadian and American import duty information taken from the Canada Border Services Agency (www.cbsa-asfc.gc.ca) and U.S. Customs and Border Protection (help.cbp.gov) respectively.

15 The mean shipping cost for [USA, USA] and [Canada, Canada] auctions were 3.831 and 3.851 USD respectively

stationarity, as most showed the possibility of non-stationarity (as was thought to be likely), the first difference of each variable was calculated and used in the model[16, 17]. Additionally, after running the regressions, all four models were subjected to a Durbin-Watson test for autocorrelation, using the R command dwtest()[18] as well as a Breusch–Pagan test, using the R command ncvTest() discussed in section 4.1. All of the models failed to reject the test null hypotheses of constant error variance and error independence (Breusch-Pagan and Durbin-Watson respectively), giving little indication of heteroskedasticity or error serial correlation, so no recalculation of standard errors was performed.

	5.1 (US)	5.1 (CA)	5.2 (US)	5.2 (CA)
(Intercept)	0.00	0.00	−0.01	−0.02
	(0.00)	(0.01)	(0.03)	(0.08)
$\Delta ln(N_{US,t})$	-0.09^{***}		−0.11	
	(0.02)		(0.14)	
$\Delta ln(B_{US,t})$	0.00		0.27	
	(0.03)		(0.24)	
$\Delta ln(N_{CA,t})$		−0.01		−0.26
		(0.03)		(0.28)
$\Delta ln(B_{CA,t})$		0.03		0.59
		(0.04)		(0.63)
Δ["Otterbox Defender"(US)]$_t$	0.13^{***}	−0.10	0.03	3.38
	(0.02)	(0.14)	(0.31)	(2.58)
Δ["Otterbox Defender"(CA)]$_t$	-0.03^{*}	−0.09	−0.02	0.12
	(0.01)	(0.08)	(0.14)	(0.81)
Δ["Otterbox eBay"(US)]$_t$	0.00	0.16^{*}	0.13	-1.89^{*}
	(0.02)	(0.09)	(0.18)	(1.12)
R^2	0.34	0.07	0.04	0.15
Adj. R^2	0.29	0.00	−0.03	0.09
Num. obs.	75	75	75	75

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2: OLS regressions of $\Delta ln(P_{c,t})$ & $\Delta ln(VAR(p)_{c,t})$ using variables in Equation 5

The first notable result, examining Table 2, is that there is generally poor fit to the models across the board. Other than 5.1 (US), little significance is assigned to any explanatory variables, and the overall explanatory power of the specification appears to be quite low. The fact that the adjusted R2 values for models 5.1 (CA), 5.2 (US), and 5.2 (CA) (predicting daily price variance) fall below, and 1 suggest that minimal inference should be prescribed from them. For the model concerning the mean auction price on each day in the United States, 5.1 (US), the fit is somewhat stronger. Noting the coefficient on the variable indicating the (first differenced) number of auctions on that day

16 The ADF test statistics for $\Delta ln(VAR(p)_{US,t})$, $\Delta ln(VAR(p)_{CA,t})$, $\Delta ln(P_{US,t})$, $\Delta ln(P_{CA,t})$, $\Delta ln(B_{US,t})$, $\Delta ln(B_{CA,t})$, $\Delta ln(N_{US,t})$, and $\Delta ln(N_{CA,t})$ were -5.2958, -5.3445, -5.1168, -6.9154, -4.0278, -5.6108, -4.9528, and -5.648 respectively.

17 A relatively large assumption accompanied all time series analyses, that of difference stationarity of all involved variables. It was assumed in this paper that the relatively short timespan of the analysis would theoretically prevent the time series from having trends or seasonal patterns. This is a potential fault with the analysis of aggregate data performed in this paper, and will be discussed further in Section 5.2.

18 From the lmtest package, see (Zeileis and Hothorn, 2002).

(ΔN), we see that the first difference of the log number of auctions negatively correlates with $\Delta P_{US,t}$. It appears that an increase in the number of auctions purchased by Americans on a given day will bring the difference between the mean price in successive periods towards the mean, reflecting the basic price-quantity tradeoff of consumer demand.

Additionally, in model 5.1 (US), there is a positive coefficient (with a high *p* value) on the variable indicating the volume of searches for "Otterbox Defender" on Google by Americans. This supports the theory expressed in section 3.4, and it appears as if there is a positive correlation between the volume of search interest and the mean price of auctions on a given day. Additionally, the data indicates that an increase in the volume of Google traffic on a given day causes the mean price to grow relative to the previous period. This fits neatly in line with the classical economic theory equating increases in demand to increases in prices. Also, the variable indicating searches for "Otterbox Defender" among Canadians shows possible negative correlation with the American mean price (although at a lower *p*-value). This could suggest that as more Canadians search for the case on Google, the American price level is pulled more towards the Canadian mean price, causing a reduction in volatility in successive periods. However, the correlation is tenuous, and therefore little inference is taken from the coefficient.

While one of the Google Trends variables appears to have relatively strong correlation with the mean price in the USA, why is the fit so poor, and why do the other search terms not correlate? This can perhaps be explained with a thought experiment concerning the routes consumers may take in navigating to eBay auction listings on the internet. Besides directly searching for a specific good, consumers might see an auction listing while they are browsing eBay already, and they might see an advertisement posted on another website or receive a link from a friend about another site among endless other opportunities. Additionally, a search for "Otterbox Defender" might not mean that the individual was looking to purchase the case, but perhaps just doing research for an Economics Integrative Exercise, or simply just looking for reviews. All in all, using Google Trends as a proxy for all online interest is far from ideal, as it sweeps under the rug a host of other methods of arriving at the eBay auction listings (this will be discussed in further detail in Section 5.2).

Understanding that the correlation between all of the Google Search traffic variables and the mean auction price and variance is not exceptionally strong (and non existant with regard to price variance), I still chose to examine the possibility of their correlation with the evolution of the real exchange rate. It is possible that the correlation of search traffic with the real exchange rate (which involves the ratio of both price levels) masks its relationship with the individual price levels themselves, as the individual price levels are in a constant state of stochastic flux.

4.3 Time Series Analysis of the Real Exchange Rate

Having investigated the economic ties between the mean and variance of auction prices and the Google search traffic for specific terms on each day, I know discuss the analysis of the evolution of the logged ratio of prices between the USA and Canada, the logged real exchange rate (q_t).

In determining the correct autoregressive model for q_t, I first demeaned[19] the series and ran it through an Augmented Dicky-Fuller test for stationarity using the adf.Test() R command. While the test statistic (-3.4942) passed the 95% confidence threshold for stationarity, it did so just barely. I decided to take the first difference of q_t, which gave a strong confirmation of stationarity upon an Augmented Dickey-Fuller test (-7.2649)[20]. Using Δq_t as the dependent variable in the model gives the added bonus of easy interpretation of its path, as its value in each successive period represents the growth of volatility of the real exchange rate or, inversely, its reversion to the mean. With a stationary series, I was now able to run the autoregressive model using OLS (Granger and Newbold, 1974), testing the significance of lags up going back up to three periods (shown in Table 3)[21].

	AR(1)	AR(2)	AR(3)	Δq_t and Google TraŒc
(Intercept)	−0.01	−0.01	−0.02	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
q_{t-1}	−0.80***	−0.79***	−0.81***	
	(0.11)	(0.12)	(0.12)	
q_{t-2}		−0.06	−0.08	
		(0.12)	(0.12)	
q_{t-3}			0.00	
			(0.12)	
Δ[ıOtterbox Defenderȷ(US)]$_t$				0.21
				(0.13)
Δ[ıOtterbox Defenderȷ(CA)]$_t$				0.06
				(0.07)
Δ[ıOtterbox eBayȷ(US)]$_t$				−0.17**
				(0.08)
R^2	0.40	0.41	0.44	0.07
Adj. R^2	0.39	0.39	0.41	0.03
Num. obs.	75	74	73	75

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 3: OLS regressions of Δq_t

The first result of interest is the non-significance of the intercept across the board. This makes sense, as the iterative process of first differencing removes the constant. Examining the lag coefficients, we see that only the first lag seems to correlate with Δqt, subtracting ≈.8 of its value from Δq_t each

19 This was done using the mean of qt over the entire sample period.

20 The different time series paths of qt and Δqt are plotted in figure 5 in section 6.

21 While theory points to a simple AR(1) process due to the efficiency of the market, I decided to test two extra lags for added rigor.

period. This equates to quite a fast half-life of price deviations (t1/2 = ln(1/2) / ln(1+ λ)). Plugging in the estimated coefficient (λ) of -.8 into the formula results in a half-life of approximately .431. As the time interval of q_t is days, this equates to a 50% reduction in disparity after only 24*.431=10.34 hours. A visual representation of the evolution of q_t after a positive shock, is given in Figure 4.

The efficiency of q_t with regard to mean reversion is not surprising. Since it is calculated using a single, easily transportable and substitutable, high volume good, it makes sense that prices should be highly flexible and quick in dynamic response to disequalibrium. This result, however, might oppose the findings of (Baye, Morgan and Scholten, 2004), who find that price dispersion on a price comparison website is significant and persistant. This difference could be explained that the transactions examined in this paper were english-style auctions, while (Baye, Morgan and Scholten, 2004) examined more static prices set by retailers. It seems logical that an auction based market place would respond more quickly to price differences, as the customers themselves are setting the prices.

With q_t apparently following its theoretical path of mean reversion, we turn to the multivariate regression including the variables for Google Search traffic. The variables for all three Google Search terms were first differenced to ensure stationarity (as the original series for all three failed to reject the null hypothesis of non-stationarity via individual Dickey-Fuller tests). Augmented Dickey-Fuller test statistics of the first difference of each of the three Google Search variables listed in Table 3 were −6.5882, −5.6693, and −6.9508 respectively. Having been shown to be stationary, the three variables were then included in the multivariate model. The model was checked for heteroskedasticity using the R command discussed in Section 4.1, but the null hypothesis of constant variance was not rejected.

The results from the multivariate specification of Δq_t show little correlation between successive period changes in q_t and the Google Search traffic variables. American search traffic for "Otterbox eBay" shows the greatest possibility of correlation, and has a negative coefficient, but the poor quality of the model makes any interpretation meaningless. It is likely that the specification of the relationship between search traffic and the real exchange rate as a linear equation of differences is far to simplistic. Suggestions for more complex models involving different Google Trends variables of more varying specification are noted in Section 5.2.

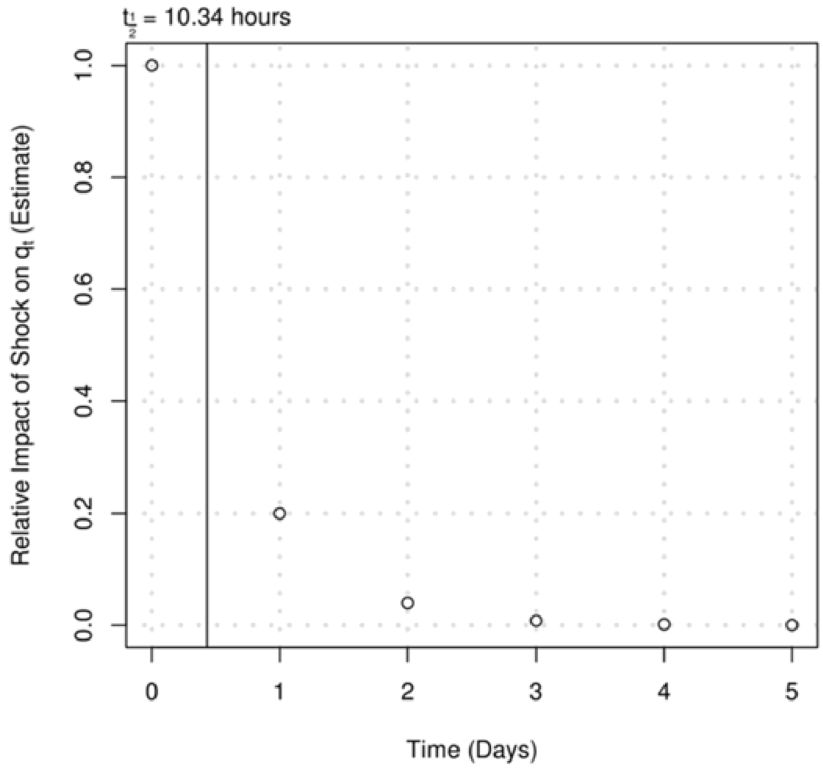

Figure 4: Estimated Decay of a Shock to q_t at $t = 0$

Time series analysis of this, specific and simplified, version of the real exchange rate overall points to the general efficiency of eBay as an international place of exchange, an interesting and important result in and of itself. However, to discern the true relationship between traffic on the internet relating to a product, and the evolution of its international price ratio, further analysis is merited, using more rigorous methods and a wide range of variables.

V. Discussion

5.1 Summary

This Paper intended to provide an answer to a previously unapproached question. Specifically, to what extend does consumer interest in a product (proxied via Google Search traffic data) bring the ratio of a products price in two countries, when accounting for the exchange rate, toward parity? To answer this question, this paper laid out a four step analytical process (shown in Figure 3) to show the extent of LOOP failure, and then determined the extent to which search traffic explained that failure. The results of this process indicate several interesting findings relevant to the theoretical efficiency of eBay as an international market.

Firstly, in examination of a specific product traded frequently between the USA and Canada, a significant price level difference is apparent. This is consistant with (Maier, 2010) finding differing prices levels, even between countries with a common currency. Moreover, taking into account the direction of trade through examination fo the location of the buyer and seller involved in each auction, the mysterious border effect seems to live on, albeit in a negative form. Even in an online marketplace like eBay, where the inclusion of auctions from all countries in the same interface somewhat masks the nationality of merchants and buyers, there is an added effect of simply transporting the good across the border. It is possible that there is some sort of online xenophobia with regard to auctions, pushing buyers to demand lower prices from their neighbors to the north or south. Or, the observed anomaly in this paper may be simply a quality of the specific product and time examined, and will be shown to be a fluke when aggregations are calculated using a wider basket of traded goods. Either way, further investigation is merited, and this author is personally intrigued by the possibility of such study.

However, while there is an apparent price level disparity and cost of transporting across national borders, analysis of the evolution of the real exchange rate shows it to be mean reverting. This suggests that while there may be some unobserved or unmeasured qualities of the auctions causing their relative price in the USA and Canada to deviate from parity, any further deviation to parity via demand or supply shock will quickly dissipate (within a matter of hours). The observed international flexibility of auction prices on eBay puts it miles ahead of the efficiency of brick and mortar stores across the border, with the estimated half-life of good-basket price level deviations in the 3-4 year range (Chen and Engel, 2004; Rogoff, 1996)[22].

Additionally, while the data on Google Search traffic didn't appear to correlate strongly with the evolution of the real exchange rate or daily price variance, there is some evidence of a positive relationship with the mean daily price in individual countries. It is only natural that higher levels of curiosity about a product should translate into greater exposure of products listed by merchants, which should increase competition and lower prices. The correlation that did appear to be present in the analysis conducted for this paper gives a hint to its potential importance and certainly merits further investigation. The power of Google Search traffic data in forecasting different economic indicators has been explored somewhat (Choi and Varian, 2012), and the observed correlation in this study suggests that similar methods (including different types of keywords and a longer time horizon) might be utilized to predict the

22 While some might cry out that the comparison is apples to oranges, it is doubtful that product aggregation would make up the vast time difference.

real exchange rate between countries, or at a minimum the price level within countries.

5.2 Critiques of Method

While the analysis in this paper seems to have arrived at some important conclusions regarding the nature of eBay as an international consumer marketplace, there are several elements of the analytical methods which deserve some thoughtful self-criticism.

Firstly, the search traffic variables used to proxy consumer interest in the Otterbox Defender have a number of downsides. Only two sets of keywords ("Otterbox Defender" and "Otterbox eBay") were utilized, representing very targeted searches of the item. It is possible that more general keywords, such as "iPhone case" or "eBay phone case", might have shown a greater correlation, as many people looking for a product to buy online don't search for the exact brand but instead the general type of item. Additionally, the 90 day period containing the collected Google Search variables began 14 days after the 90 day period containing all the collected auction observations. This meant that the final time series analysis was limited to 76 days, representing a loss of 15% of the original dataset. A dataset of search traffic containing more terms of varying specificity with relation to the examined product might provide a clearer picture of the interaction between web traffic and international auction price fluctuations.

In addition to potential problems stemming from the specificity of the search keywords, the specificity of the product itself might explain the lack of correlation between internet traffic and price fluctuations. If the analysis was instead conducted on a range of phone cases, and more general search keywords included, there might be a clearer connection. While there is an analytical trade off between the level of product specification and aggregate inference, it seems as if the exact product specification used in this paper may have prevented stronger correlation with the search data. As Google itself builds its core search product off of correlative analysis of aggregated user patterns, more aggregated product data could be warranted.

Besides the data themselves, elements of the Python cleaning script could have reduced the accuracy of the data in reflecting reality. Specifically, several values for the Canadian aggregate time series had to be calculated via inference to be included in the models. There were several days when there were only 1 or 0 cases purchased by Canadians. Both of these resulted in the variance taking a value of 0 (and thus unable to be logged, as they would be undefined, and included in the models). To get around this, the variance for 4 missing days was calculated as the average of the variance on the day before and the day after[23]. These imputations distorted the data, and might explain the minimal explanatory power in the regressions concerning Canadian daily aggregate variables. By moving to a larger range of products, this would be much less likely to happen, removing the need for imputation. However, the large amount of time involved in the manual data collection process performed for this paper prevented this author from expanding the selection of examined products[24].

As briefly mentioned in Section 4.2, the assumption that all the time series variables were difference-stationarity may be a large one, and was derived mostly from the short time period examined. It seems unlikely that a general trend in prices or sales could occur with a matter of slightly over 2 months for a specific product, though this is an empirical question that might have been more rigorously addressed.

Finally, with regard to statistical methodology, a threshold autoregression model of the real exchange rate might have added some explanatory power, as most literature points to upper and lower boundaries to which q_t reverts, rather than the mean itself (Yoon, 2010). This is mostly for studies of aggregated goods, however, and the fact that the product studied here was so specific, and the market response to deviations from parity so fast, that the band of inaction would most likely be minimal. Additionally, a vector autoregression might be used to more closely analyse the relationship between the evolution of the real exchange rate and the Google Search variables, though the limited correlation seen here suggests different and possibly more search keywords should be considered, perhaps over a longer time.

5.3 Additional Avenues for Research

Analzying eBay as an internet marketplace with a specefic interest in online consumer interest brings to mind several potential avenues for further research, involving modified or additional variables and more rigorous methods.

A more rigorous approach to determining fluctuations in online consumer interest in an item might employ more computationally intensive methods such as data mining Twitter or Facebook for mentions of the examined products. Additionally, graph theory or cluster analysis might be utilized to determine the most traveled hyperlink channels on eBay itself that lead to the product. There already exists general precedent for analysis of consumer behavior patterns and using some of these techniques (Giudici and Passerone,

23 The number of bids, the average price, and the number of cases all had to be imputed for the one day with no auctions at all.

24 During days when I performed the data collection, the connection speed to Terapeak's servers was excruciatingly slow, with searches taking multiple minutes to complete. Usage of the Terapeak Developer API would greatly reduce the time and labor involved in data collection.

2002; Bollen, Mao and Zeng, 2011), and similar methods might be interestingly applied to the evaluation of the “online real exchange rate”. Creating an overall “interest index” by combining all of these methods (including more Google Trends variables) might give an even better picture of fluctuations in consumer interest.

As discussed in the previous subsection, a larger number of products could result in a better understanding of eBay as an internet marketplace in general. While this has been somewhat explored before (Maier, 2010), data was collected either manually or through purpose written script, crawling through the general user interface. The Terapeak API offers an easier, more targeted way to collect data on large amounts of auctions of different products.

Additionally, by writing an automatic data collection script and collecting auction data every day over a few months or a year, the 90 day limit of Terapeak for stored auction listings could be bypassed. eBay could potentially even create a sort of automatic monitoring system, which calculates the daily real exchange rate via bundles of commonly purchased goods, and gives an almost real time picture of the state of purchasing power parity. The sample could also easily be expanded to include many more countries, as Terapeak maintains data on the nationalities of both the buyer and seller for all countries with access to eBay.

5.4 Online markets and the Future of Macroeconomics

There is no doubt that the rapid development of online commerce has brought a host of new opportunities for consumers and merchants around the world. eBay in particular represents a massive, highly efficient, highly transparent arena in which to conduct trade across state lines and around the globe. As the internet becomes a more important part of the average person’s life, and more trade is pushed online, investigating the economic qualities of internet markets is extremely important. Thus, understanding how fluctuations in consumer interest online translates to changes in macroeconomic variables will be pivotal.

Fortunately, the rate at which the internet becomes a more ubiquitous part of global society is matched by the growth of data generation and storage. An important economic challenge in the coming years will be to adequately sift through these new, massively accumulating metrics of the global economy and use them to make rational and beneficial policy recommendations. Elements of computer science, computational statistics, and economics will need to cohesively work together in tackling and addressing the significan opportunities for social analysis created by the complex institution that is the internet.

VI. Data Appendix

This section contains descriptive statistics concerning the individual auction data (Table 4), as well as the aggregate time series (Table 5). Additionally, time series plots for q_t and Δq_t are included (Figure 5).

Variable	n	Minimum	Mean	Maximum
Total Price	12887	4.340	29.440	104.420
Starting Price	12887	0.010	8.117	98.680
Ending Price	12887	0.970	25.493	100.340
Number of Bids	12887	1.000	12.105	51.000
Average Shipping Cost	12887	0.000	3.947	7.090
[USA, USA] *	12887	0.000	0.942	1.000
[USA, Canada]*	12887	0.000	0.011	1.000
[Canada, USA]*	12887	0.000	0.039	1.000
[Canada, Canada]*	12887	0.000	0.009	1.000
American Buyer*	12887	0.000	0.952	1.000
Canadian Buyer*	12887	0.000	0.048	1.000
Other Color*	12887	0.000	0.150	1.000
Green*	12887	0.000	0.037	1.000
Blue*	12887	0.000	0.101	1.000
Red*	12887	0.000	0.032	1.000
Yellow*	12887	0.000	0.044	1.000
Pink*	12887	0.000	0.281	1.000
Black*	12887	0.000	0.227	1.000
Camo*	12887	0.000	0.147	1.000
Grey*	12887	0.000	0.163	1.000
White*	12887	0.000	0.100	1.000
Teal*	12887	0.000	0.064	1.000
Plum*	12887	0.000	0.056	1.000
Orange*	12887	0.000	0.053	1.000
Zebra*	12887	0.000	0.029	1.000
Thermal*	12887	0.000	0.028	1.000
Ocean*	12887	0.000	0.024	1.000
Purple*	12887	0.000	0.019	1.000
Glacier*	12887	0.000	0.018	1.000
Night*	12887	0.000	0.020	1.000
Gunmetal*	12887	0.000	0.011	1.000
Sky*	12887	0.000	0.013	1.000
Stripes*	12887	0.000	0.011	1.000
Military	12887	0.000	0.010	1.000
Forest*	12887	0.000	0.003	1.000
Sea*	12887	0.000	0.005	1.000
Clip*	12887	0.000	0.575	1.000

Table 4: Individual Auction Data Summary

(* indicates a binary variable, mean implies the ratio of observations fitting that variable)

Variable	n	Min	Median	Mean	Max
$P_{US,t}$	76	26.423	29.051	29.443	33.865
$B_{US,t}$	76	7.539	12.847	12.351	16.013
$\mathrm{VAR}(p)_{US,t}$	76	24.749	51.175	53.384	111.368
$N_{US,t}$	76	54.000	125.000	131.487	241.000
$P^{**}_{CA,t}$	76	22.128	32.225	32.767	46.704
$B^{**}_{CA,t}$	76	1.000	11.917	12.194	27.333
$\mathrm{VAR}(p)^{**}_{CA,t}$	76	0.032	62.067	64.963	179.567
$N^{**}_{CA,t}$	76	1.000	6.000	6.842	20.000
q_t	76	-0.432	-0.105	-0.097	0.270
[ıOtterbox Defenderȷ(CA)]$_t$	76	25.000	43.000	45.645	100.000
[ıOtterbox Defenderȷ(US)]$_t$	76	29.000	44.000	46.513	100.000
[ıOtterbox eBayȷ(US)]$_t$	76	19.000	42.000	45.368	100.000

Table 5: Aggregate Data Summary

(** indicates some imputed values)

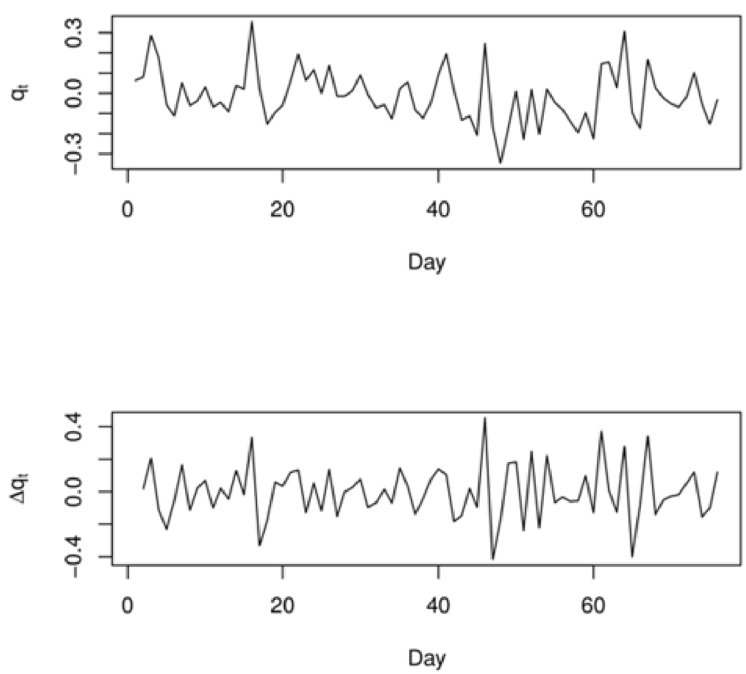

Figure 5: Time Series Plots of q_t and Δq_t

References

Baye, Michael R., John Morgan, and Patrick Scholten. 2004. "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site." The Journal of Industrial Economics, 52(4): pp. 463–496.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter Mood Predicts the Stock Market." Journal of Computational Science, 2(1): 1 – 8.

Brynjolfsson, Erik, and Michael D. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." Management Science, 46(4): pp. 563–585.

Budish, Eric B., and Lisa N. Takeyama. 2001. "Buy Prices in Online Auctions: Irrationality on the Internet? " Economics Letters, 72(3): 325 – 333.

Chen, Shiu-Sheng, and Charles Engel. 2004. "Does "Aggregation Bias" Explain the PPP Puzzle? " National Bureau of Economic Research Working Paper 10304.

Choi, Hyunyoung, and Hal Varian. 2012. "Predicting the Present with Google Trends." Economic Record, 88: 2–9.

Ellison, Glenn, and Sara Fisher Ellison. 2009. "Search, Obfuscation, and Price Elasticities on the Internet." Econometrica, 77(2): pp. 427–452.

Engel, Charles. 1994. "Real Exchange Rates and Relative Prices: An Empirical Investigation." National Bureau of Economic Research Working Paper 4231.

Engel, Charles, and John H. Rogers. 1994. "How Wide is the Border? " National Bureau of Economic Research Working Paper 4829.

Fox, John, and Sanford Weisberg. 2011. An R Companion to Applied Regression. Second ed., Thousand Oaks CA:Sage.

Froot, Kenneth A., Michael Kim, and Kenneth Rogoff. 1995. "The Law of One Price Over 700 Years." National Bureau of Economic Research Working Paper 5132.

Giovannini, Alberto. 1988. "Exchange Rates and Traded Goods Prices." Journal of International Economics, 24(1): 45 – 68.

Giudici, Paolo, and Gianluca Passerone. 2002. "Data Mining of Association Structures to Model Consumer Behaviour." Computational Statistics and Data Analysis, 38(4): 533 – 541.

Granger, Clive W. J., and P. Newbold. 1974. "Spurious Regressions in Econometrics." Journal of Econometrics, 2(2): 111–120.

Isard, Peter. 1977. "How Far Can We Push the "Law of One Price"? " The American Economic Review, 67(5): pp. 942–948.

Maier, Philipp. 2010. "An Analysis Of International Price Differentials On Ebay." Contemporary Economic Policy, 28(3): 307–321.

Richardson, J.David. 1978. "Some Empirical Evidence on Commodity Arbitrage and the Law of One Price." Journal of International Economics, 8(2): 341 – 351.

Rogoff, Kenneth. 1996. “The Purchasing Power Parity Puzzle.” Journal of Economic Literature, 34(2): pp. 647–668.

The Economist. Nov 18, 1999. “Frictions in Cyberspace.” The Economist Newspaper Limited.

Yoon, Gawon. 2010. “Do Real Exchange Rates Really Follow Threshold Autoregressive or Exponential Smooth Transition Autoregressive Models? ” Economic Modelling, 27(2): 605 – 612.

Zeileis, Achim, and Torsten Hothorn. 2002. “Diagnostic Checking in Regression Relationships.” R News, 2(3): 7–10.